MARKET SHARE CONSTRAINTS AND THE LOSS FUNCTION IN CHOICE BASED CONJOINT ANALYSIS

Timothy J. Gilbride
Assistant Professor of Marketing
315 Mendoza College of Business
University of Notre Dame
Notre Dame, IN 46556
Phone: (574) 631-9987
Fax: (574) 631-5255
tgilbrid@nd.edu

Peter J. Lenk
Associate Professor of Operations and Management Science
Associate Professor of Marketing
Ross School of Business
University of Michigan
plenk@umich.edu

Jeff D. Brazell
President/CEO
The Modellers, LLC
Salt Lake City, UT
jeffb@themodellers.com

November 30, 2006

MARKET SHARE CONSTRAINTS AND THE LOSS FUNCTION IN CHOICE BASED CONJOINT ANALYSIS

Abstract:

Choice based conjoint analysis is a popular marketing research technique to learn about consumers' preferences and to make market share forecasts under various scenarios for product offerings. Managers expect these forecasts to be "realistic" in terms of being able to replicate market shares at some pre-specified or "base case" scenario. Frequently, there is a discrepancy between the forecasted and base case market share. This paper presents a Bayesian decision theoretic approach to incorporating base case market shares into conjoint analysis via the loss function. Because defining the base case scenario typically involves a variety of management decisions, we treat the market shares as constraints on what are acceptable answers, as opposed to informative prior information. Our approach seeks to minimize the adjustment of parameters by using additive factors from a normal distribution centered at 0, with a variance as small as possible, but such that the market share constraints are satisfied. We specify an appropriate loss function and all estimates are formally derived via minimizing the posterior expected loss. We detail algorithms that provide posterior distributions of constrained and unconstrained parameters and quantities of interest. The methods are demonstrated using both multinomial logit and probit choice models with simulated data and data from a commercial market research study.

Keywords: Hierarchical Bayes, Loss Function, Posterior Risk, Bayesian Decision Theory, Conjoint Analysis

MARKET SHARE CONSTRAINTS AND THE LOSS FUNCTION IN CHOICE BASED CONJOINT ANALYSIS

1. Introduction

Discrete choice conjoint analysis has proven to be a useful tool for investigating consumer preferences in both applied and academic studies. In applied settings, results from a typical conjoint analysis are used by managers to decide on the most attractive combination of product features and price given the competitive offering, or anticipated changes to the competitive set. To facilitate "what-if" analysis, commercial market research firms frequently provide software packaged as a "market simulator." These simulators use the results from the conjoint study together with manager supplied inputs on product features and pricing to produce market share forecasts. When a manager enters a "base case scenario" consisting of a set of actual product features and prices, there is usually a discrepancy between the forecast from the simulator and the actual market share. While there are many reasons why the forecasted and actual market share may differ, when this occurs a manager may doubt the quality of the study and/or the reliability of the forecasts. The manager really has two goals from the conjoint study: to represent consumer preferences and to produce "realistic" forecasts. An interesting question is how analysts should use manager's "base case" scenario expectations. This paper presents a Bayesian decision theoretic approach to incorporating base case scenario projections into discrete choice conjoint analysis via the loss function.

Forecasted market share from a conjoint study may differ from actual market share for a variety of reasons. In an early survey on the commercial use of conjoint analysis, Cattin and Wittink (1982) list five difficulties in making market share predictions using conjoint analysis. These include the inherent difference between stated and revealed preferences, attributes present in the marketplace but excluded from the conjoint study, the inability of conjoint studies to include the effect of mass communication, distribution, competitive reaction, and other effects. Orme and Johnson (2006) reiterate many of these reasons in their summary of practitioners' methods of adjusting simulated market shares to match actual market shares. It is important to note that Orme and Johnson do not advocate adjusting the results from the market simulators and instead argue that managers should be educated on the reasons why simulated and actual market share may not agree. Allenby et al. (2005) review additional reasons why choice experiments differ from actual market choices and outline data collection procedures and new types of models which may improve the predictive accuracy of choice models.

In a Bayesian analysis, the base case market share can be modeled as part of the likelihood function, incorporated into the loss function, or used as informative prior information. Some researchers have suggested that stated and revealed preference data be incorporated into a unified model to overcome problems with conjoint studies. Louviere (2001) notes that there is a close correspondence between stated and revealed preferences. However, it may be necessary to calibrate either the location or scale parameter from analyses that only include stated preferences to account for the inherent differences between experimental and actual choice behavior. Morikawa, Ben-Akiva, and McFadden (2002) detail statistical methods of combining stated preference and revealed preference data. When available, conjoint analysis data should be augmented with actual market place choices and the full set of data modeled. However, not all studies are amenable to this solution either for logistical reasons (market data is not available for conjoint participants) or due to the lack of available models for incorporating aggregate market share results with experimental studies. Further, there is no guarantee that such analyses will necessarily produce market share projections that match a manager's base case scenario.

The base case scenario and accompanying market share used by a manager will typically require several decisions. Market shares vary over time and geography as prices change due to promotion, advertising intensity or message change, fluctuations in distribution, changes in competitive offerings, etc. Managers must decide if the base case scenario will reflect a very specific measure (e.g. one week, in one particular market) or an aggregated measure (e.g. annual market share for all markets served). If the latter, then the manager must choose whether "average values" of the product attributes will be used or if "representative values" will be selected. Selecting a point estimate to serve as the base case scenario is not trivial. Our experience suggests that managers use a combination of "known facts," aggregated, and "representative values" to arrive at the base case scenario and accompanying market share estimates or expectations.

From the analyst's perspective, the manager's base case scenario is most appropriately viewed as a constraint placed on the forecasts provided by the market simulator. Bayesian decision theory implies that these constraints should be incorporated into the loss function. Essentially, our approach to making market share forecasts is to approximate a procedure that integrates over the regions of the posterior distribution of parameters that are consistent with the market share constraints.

We have several objectives in mind when developing the loss function approach. Most explicitly, we require that the market shares at the base case be sufficiently close to the managerially specified market shares such that the choice based conjoint (CBC) analysis has face validity for the marketing manager. Clearly, this alone is not sufficient to identify the adjustment procedure, and one could imagine any number of methods to reach this objective. For example, one could adjust the preferences of a subset of subjects while leaving others untouched, or one could modify the coefficients for only a subset of attributes. Instead of being heavy-handed with the adjustments, we want the final results to be as true as possible to the CBC data. The adjustment procedure that we develop perturbs all of the estimated parameters from the CBC data as little as possible in order to leave the preference structure of the CBC data relatively intact. Our adjustment terms are additive factors with mean zero, and we specify the variance to be as small as possible while satisfying the constraints.

An attractive feature of the proposed approach is that the estimation of the CBC parameters and the adjustment process is decomposed into separate operations. The posterior distribution of the CBC parameters or simulates thereof is the input to the adjustment process. The analyst can keep the two separate and display the CBC estimates before and after the adjustment process, and he or she can give an explicit representation of the impact of the managerial constraints.

We contrast the loss function approach with several others that treat the base case scenario as an informative prior. The first problem one faces with treating the constraints as data or prior information is establishing the correct weighting between the CBC data and the base case. Because we lack revealed preference data for the subjects in the CBC data and because the base case market shares are based on a different sample than the CBC study, the relative weight to place on the prior versus the likelihood is not identified. Does one treat the base case market shares for M products as M bits of information, or does one treat them as millions of transactions? The relative weights of the CBC and market share data can be treated as a tuning parameter, however, a single study is not sufficient to identify it.

In some instances, the loss function and informative prior yield mathematically equivalent results. However, the loss function approach explicitly recognizes that the adjustments are due to external criteria or goals imposed on the analysis by the manager. We believe that this route better fits the situation where the marketing researcher is attempting to satisfy the sometimes competing goals of his or her client. While a strength of the Bayesian paradigm is its ability to introduce subjective prior information, "prior information" is handled differently than restrictions on what constitutes an acceptable answer. Both are valid additions to a decision problem from the standpoint of the analyst. This paper illustrates how the Bayesian paradigm handles market share constraints and provides practical algorithms for implementing them in discrete choice conjoint analyses.

The remainder of the paper is organized as follows. First we set-up the discrete choice analysis, review the role of the loss function in Bayesian decision theory, and propose a specific loss function that incorporates market share constraints. This loss function minimizes changes in preferences as represented in the conjoint study while being within an acceptable range of the manager's base case market share expectations. We then contrast the loss function approach with several informative prior approaches. Simulation and Markov-Chain Monte Carlo (MCMC) methods are detailed for conducting the analysis implied by the loss function and necessary for obtaining market share forecasts. These can be incorporated into the same computer program used to estimate the discrete choice model and produce posterior distributions of parameters with and with-out the market share constraint. We illustrate the proposed methods using simulated data for both the multinomial logit and correlated probit discrete choice models. Our simulations show that when the market share constraints are accurate, the proposed method yields better market share predictions for changes in product formulation outside the base case scenario. We show the results from a commercial market research study and conclude with a summary and suggestions for future research.

2. The Loss Function and Bayesian Decision Theory

2.1 Discrete Choice Model and Market Share Constraints

This section begins by stating the discrete choice model and defines various terms. We envision a commercial market research study where the data consists of a sample of respondents who have completed a CBC study and management has provided market share estimates at some base case set of product attributes for some set of the brands or products in the CBC study.

The model set-up is as follows. *N* subjects evaluate J_i choice occasions for subject *i*. Each choice occasion or tasks consists of *M* alternatives or product profiles. In the CBC design, there are *p* attribute levels, including brand intercepts. Subject *i*'s random utility for product profile j is a given by

$$
\begin{aligned}
&Y_{ij} = x'_{ij}\,\beta_i + \varepsilon_{ij} \quad \text{for subject i and choice occasion j} \\
&\varepsilon_i \sim \text{Normal or extreme value} \\
&\beta_i = \theta + \nu_{\mathrm{i}} \\
&\nu_i \sim N_p(0, \Lambda)
\end{aligned}
$$

where the observed attributes x_{ij}, the parameter vector β_i, and error terms ε_{ij} are *p*-vectors. Let *W* be the observed choices from the CBC.

Our primary objective is to use the set of $\{\beta_i\}$ in functions $g(\{\beta_i\})$ to obtain estimates of quantities of interest such as predicted market share. Let X_o represent the matrix of product attributes in the base case scenario and $S_1, \ldots, S_K$ represent the corresponding market shares. Let $\hat{S}_1, \ldots, \hat{S}_K$ represent the predicted market shares using X_o and $\{\beta_i\}$, e.g. $g(\{\beta_i\}, X_o)$. We define the discrepancy between the predicted and base case market shares as:

$$\sum_{k=1}^{K} \left| S_k - \hat{S}_k \right| = \delta \tag{1}$$

Our goal is to produce estimates of $\hat{S}_k$ that are "close" to S_k and reflect the information obtained in the conjoint study. We assume that in addition to X_o and $S_1, \ldots, S_K$, management can provide δ^t, a target level, or acceptable level of discrepancy between the predicted and base case market shares. Let C represent the set of parameters used to estimate $\hat{S}_k$ that satisfy the market share constraint $\delta \leq \delta^t$. We will use the indicator function $\chi(S) = 1$ to indicate when the constraint is satisfied and $\chi(S) = 0$ when it is not.

In the situation that we consider in this paper, the actual market behaviors of the subjects in the CBC study are not observed. Consequently, we are unable to build a model that relates the parameters for the revealed and stated preference data. In fact, the subjects in the CBC study may not have made a purchase in the product category during the period that the base market shares were computed. Another insolvable issue is the relative weighting of the market share information with the CBC data. Consider the case where a separate market based dataset, such as scanner data, is available. Should one base the weights on the number of observations in the CBC study – typically hundreds of subjects each making dozens of choices – versus the number of purchases in the market share data – which can be millions of transactions for frequently purchased consumer goods? In this "micro" approach, the market share data, derived from a very large number of purchases, will swamp the more limited CBC data. Alternatively, one could use a "macro" approach and treat the market shares for K products as K-1 observations.

Then the CBC data would dominate. Intuitively, the correct weighting is somewhere between the two extremes, but deciding where requires more than CBC data and base rates.

The approach that we propose in this paper is more direct and much simpler than developing a joint model. We propose a method to adjust the estimated parameters from the CBC study so that their implied market shares are close to the base market shares. The adjustment method is formally derived through Bayesian inference by including the base market shares into the loss function. This adjustment perturbs the CBC estimates as little as possible.

We illustrate the approach by considering a single parameter β_i. A standard Bayesian analysis will produce a posterior distribution $\pi(\beta_i|W)$ as illustrated in Panel A of Figure 1. The posterior is proportional to the prior distribution on β_i and the likelihood of β_i given the data W, or $\pi(\beta_i|W) \propto l(\beta_i|W)\pi(\beta_i)$. Within $\pi(\beta_i|W)$ there are regions, perhaps disjoint, that satisfy some exogenous constraints as indicated by the shaded regions in Panel A. Given a draw β_i^r from this posterior distribution, we generate adjustment factors $\{\alpha_i\}$ from a normal distribution $N(0, \tau^{-1})$ with density $\varphi(\alpha|\tau)$ as shown in Panel B. We consider draws of α_i that map into the shaded region of the posterior via $\alpha_i^r + \beta_i^r$ and satisfy the base case constraints. This is illustrated in Panel C of Figure 1. Note that now the{C } such that $\chi(S) = 1$ contains β and α. Formally, our approach conditions on the posterior distribution $\pi(\beta_i|W)$ and the α's are dummy variables of integration which limit the area over which posterior analyses are conducted. The α_i is not a parameter in the likelihood function, but a device we use to satisfy an external constraint. Note that in this example, the region defined by $(\beta_i + \alpha_i)$ is disjoint, so functions such as $E[(\beta_i + \alpha_i)]$ where the expectation is with respect to $\pi(\beta_i|W)$, are not particularly meaningful.

=============== Figure 1 ===================

This choice of the normal distribution for the adjustment factors is not arbitrary. We set the mean to zero with the desire to keep the adjusted estimators unbiased. We use a spherical variance structure to be indifferent about the set of attributes to adjust. We use a normal distribution because its tails decline rapidly, so that the estimate for any single attribute or subject will not be greatly modified relative to the adjustment for other parameters.

In CBC analysis we are interested in using parameters to estimate quantities of interest, such as market share. In order to enforce market share constraints, each parameter for each subject will have its own additive factor. We represent functions that use these quantities as $h(\{\beta_i + \alpha_i\})$ where $\alpha_i \sim N_p(0,\tau^{-1}I_p)$ and I_p is a p x p identity matrix. Clearly the additive factors will not be unique: for a given set of $\{\beta_i\}$ there are any number of $\{\alpha_i\}$ that will satisfy the market share constraints. Forcing each α_i to be as close to 0 as possible by making τ as large as possible, will be used to help identify the $\{\alpha_i\}$. Let B represent the set $\{\beta_i\}$ and A the set $\{\alpha_i\}$. We will use the loss function to obtain point estimates of $h(A+B)$. Next we review the basic idea of loss functions and then introduce one that reflects our goals of using parameters that match the market share constraints and minimize changes to the preferences as revealed in the CBC study.

2.2 Loss Functions

In decision analysis (c.f. DeGroot 1970), the loss function $L(\theta,d)$ quantifies the loss to the decision maker of taking action (or estimate) d when the state of nature (or parameter) is equal to θ. [Here θ is any arbitrary parameter.] Since the state of nature is not known, the Bayesian decision maker wishes to minimize the posterior expected loss, or risk function represented by:

$$\rho(\pi,\mathrm{d}) = \int_{\theta} L(\theta,d)\pi(\theta \mid x)d\theta \qquad (2)$$

where π represents the current distribution of θ, in this case the posterior. The Bayes rule is the selection of $d \in D$, the set of allowable decisions, that minimizes $\rho(\pi,\mathrm{d})$.

The statistical problem of point estimation is one application of the loss function. In this case, d is the particular estimate of θ to report and use in additional analysis. A common loss function is squared error loss represented as $L(\theta,\mathrm{d}) = (\theta - d)^2$ for a scalar parameters and $L(\theta,\mathrm{d}) = (\theta - d)'(\theta - d)$ for a vector of parameters. When a squared error loss function is used and (2) is minimized with respect to d, the resulting Bayes rule, or point estimate for θ is equal to the mean of the posterior distribution or:

$$\bar{\theta} = \int_{\theta} \theta\pi(\theta \mid x)d\theta = E(\theta \mid x) \tag{3}$$

Where $\bar{\theta}$ represents the Bayes rule. In fact, the posterior mean is frequently reported in both applied and academic studies using Bayesian methods. However, the basic machinery of Bayesian decision analysis can be applied to any loss function.

The choice of a normal distribution for the additive factors, for instance, can be motivated by considering the loss functional $T(f) = \int f(\alpha)\ln[f(\alpha)]d\alpha$, which is minus the entropy, of the density f for α. Given the constraint that the mean is zero and the precision is τ, the normal density minimizes the functional $T(f)$ (c.f. Bernardo and Smith 1994, pp 207-209.). The normal distribution maximizes the entropy for densities on the real numbers given constraints on the first two moments. In other words, given the mean and the variance, the normal density imposes less structure on the set of A than other choices for the distribution.

To enforce market share constraints on the CBC analysis, we will use a variant of the squared error loss function. The loss function is represented as:

$$L(h,d) = [h(A+B)-d]'[h(A+B)-d] \text{ for } A+B \in C \tag{4}$$

where *A* is the set of adjustment factors; *B* is the set of parameters from the CBC data; C is the constraint set defined by the market share at the base case; and *d* is the estimator of the functional *h*. Here the requirement $A + B \in C$, which means that $\alpha_i + \beta_i \in C$ for $\alpha_i \in A$ and $\beta_i \in B$, can be seen as limiting the set of allowable decisions, $d \in D$. To obtain the risk function we integrate with respect to all the unknown quantities. In this case we have the posterior distribution of *B* and the distribution of *A*, which are variables unique to the loss function.

$$\begin{gathered}\rho(\pi,\phi,d \mid \tau) = \int_C [h(A+B)-d)]'[h(A+B)-d)]\varphi(A \mid \tau)\pi(B \mid W)\,dA\,dB \\ \text{where } \varphi(A \mid \tau) = \prod_{i=1}^{N} \varphi(\alpha_i \mid \tau) \text{ and } \pi(B \mid W) = \pi(\beta_1,\ldots\beta_N \mid W).\end{gathered} \tag{5}$$

Note that the requirement $A + B \in C$ is subsumed into the area of integration. To obtain the Bayes rule d*, we differentiate (5) with respect to d and set it equal to 0:

$$d^* = \frac{\int_C h(A+B)\varphi(A \mid \tau)\pi(B \mid W)dAdB}{\int_C \varphi(A \mid \tau)\pi(B \mid W)dAdB} = P(C \mid \tau)^{-1} \int_C h(A+B)\varphi(A \mid \tau)\pi(B \mid W)dAdB$$

$$d^* = \int h(A+B)\, f(A,B \mid W,C,\tau) \qquad (6)$$

where

$$f\left(A,B \mid W,C,\tau\right) = \frac{\varphi(A \mid \tau)\pi(B \mid W)}{P(C \mid \tau)} \text{ for } A+B \in C$$

The support for $f(A, B|W, C, \tau)$ is C and the normalizing constant is $P(C|\tau)$. Thus equation (6) states that the Bayes rule for $h(A + B)$ is the expected value of the function when the values of A and B are drawn from a specific density. Also note that the Bayes rule is conditional on the value of τ; we will return to this point shortly.

The value of $\alpha_i + \beta_i$ drawn from $f(A, B|W, C, \tau)$ are dependent on all other α's and β's. However, this specific distribution is a by-product of the loss function and arriving at an estimate for $h(A+B)$: the posterior distribution for β and all other parameters in the model are obtained using standard procedures and without reference to α or the market share constraints.

Although the base case scenario attribute levels X_o will be needed in order to evaluate whether $A + B$ is in C, $h(A+B)$ may involve any attribute levels. This allows the analyst to do "what-if" analysis with a set of parameters that satisfy the market share constraints (conditional on the value of τ). However, since the $\alpha_i + \beta_i$ are dependent on each other, averaging over the posterior draws of $\{\beta_i + \alpha_i\}$ for each individual and using $\overline{(\beta_i + \alpha_i)}$ to predict market share will not necessarily satisfy the market share constraint. Saving the sets of $\{\beta_i + \alpha_i\}$ in order to do "what-if" analysis represents a change in procedure for analysts accustomed to using individual averages as inputs to market simulators. Also, as noted earlier, since the market share constraints my induce disjoint regions in the sampled values of $h(A+B)$, a histogram of the values should be inspected before calculating the expected value.

The Bayes rule is conditional on $A + B \in C$ and the value of τ. A priori it is not clear what value of τ should be specified by the analyst. If τ is too large then values of α_i from $N_p(0,\tau^{-1}I_p)$ may be too close to 0 in order to satisfy $A + B \in C$. If τ is too small then α_i will have a large variance and although we may satisfy $A + B \in C$, $A + B$ will be poorly identified and we may inadvertently alter the individual β's more than is necessary or desired. Recall that our goal is to produce market share forecasts consistent with the base case scenario while altering the β's as little as possible.

Formally, we may select the optimal value of τ by minimizing the Bayes risk. The Bayes risk as a function of τ is obtained by plugging the Bayes rule, d*, back into the risk function. It is represented by:

$$\begin{aligned} r(\tau) &= \int_C [h(A+B) - d^*]' [h(A+B) - d^*] \varphi(A \mid \tau) \pi(B \mid W) \, dA \, dB \\ &= P(C \mid \tau) \int_C [h(A+B) - d^*]' [h(A+B) - d^*] f(A, B \mid W, C, \tau) \, dA \, dB \qquad (7) \\ &= P(C \mid \tau) \sum_{m=1}^{M} \mathrm{var}[h_m(A+B) \mid W, C, \tau] \end{aligned}$$

Where M is the number of market shares we are estimating. When the market share constraints are meaningful, eg. $B \notin C$, increasing τ decreases $P(C|\tau)$ because A will be closer to 0, and will reduce $r(\tau)$. The $\mathrm{var}[h_m(A + B)|W, C, \tau]$ will also be reduced by large values of τ since $\alpha_i \sim N_p(0,\tau^{-1}I_p)$ and large τ implies less variance. This suggests that when the market share constraints are necessary, we want τ to be as large as possible as long as $A + B \in C$. We show numerical results with simulated data that reinforces this intuition. First, we present several alternative methods of incorporating market share constraints as informative priors.

3. Market Share Constraints and Informative Priors

In the loss function approach to incorporating market share constraints, the risk function incorporates the posterior distribution of *B* (and all other model parameters) and a separate distribution for *A*, the additive factors. Formally, the analysis is broken into two distinct pieces and will require draws from two different distributions. This can be represented as (focusing just on the parameter *B*):

$$\pi(B|W) \propto l(B|W)\pi(B) \qquad (8a)$$

$$f(A, B|W, C, \tau) \propto \pi(B|W)\varphi(A|\tau) \text{ for } A+B \in C \qquad (8b)$$

where (8a) is a standard hierarchical model and (8b) arises as a byproduct of estimating the Bayes rule d* for $h(A+B)$ from the loss function (4). An alternative to using a loss function is to treat the base case scenario as an informative prior; that is, the manager's expectation of the market share at specified attribute levels and the allowable discrepancy is treated as prior "data," and not as a constraint on what constitutes a correct answer. A mathematically equivalent way to represent (8a) and (8b) is:

$$\pi(B,A|W) \propto l(B|W)\pi(B)\ \varphi(A|\tau)\ for\ A \in C|B \qquad (9)$$

Note that we use the conditioning argument *C|B* so that values of *B* are drawn without reference to *A* or the market share constraint. Here the difference between the loss function and informative prior approach is purely philosophical. An alternative and perhaps more traditional way to think about the market share constraint is to represent (9) as

$$\pi(B,A|W) \propto l(B|W)\pi(B)\ \varphi(A|\tau)\ \text{for}\ A+B \in C \tag{10}$$

and draws of *B* are dependent on *A* and the market share constraints. Marginal posterior distributions of *B* from (8a) or (9) will not match those from (10).

Market share constraints can also be incorporated through informative priors without using the additive factors *A*. For instance, one can treat the forecasted market shares at X_o as a "parameter" in the model and place an informative prior on them. Consider the following transformations:

$$z_k = \ln\left(\frac{S_k}{S_K}\right) \quad \text{and} \quad \xi_k = \ln\left(\frac{\hat{S}_k}{\hat{S}_K}\right) \text{for } k = 1,\ldots,K-1 \tag{11}$$

Now let $\psi(\xi|\tau) \sim N_{K-1}(z, \tau^{-1}I_{K-1})$ and I_{K-1} is a (K-1) x (K-1) identity matrix. Now the prior parameter τ determines the tradeoff between the managerial constraints and the CBC data. Very large τ will result in posterior distributions "relatively" close to the base case, although how close will be a function of the data. An alternative approach which does not involve τ is to put a distribution on δ from equation (1). For instance we might assume that δ follows a uniform distribution between $(0, \delta^t)$. Again, in both of these approaches, draws of B are dependent on the market share constraints and marginal posterior distributions will not match those of 8(a).

In the informative prior approaches as typified in (9), (10), and (11), τ is a parameter in the prior distributions $\varphi(A|\tau)$ or $\psi(\xi|\tau)$ and should theoretically be set before the analysts sees the data, be estimated from the data, or be determined via minimizing the Bayes risk. A priori, it is difficult to preset τ. If it is estimated from the data, then for (9) and (10) the prior will have to be very informative in order to identify the model. For (11) our experience has been that the estimated value of τ mostly reflects its prior since there are only K-1 pieces of information for its estimate, and K is relatively small compared to the CBC data. It is also possible to specify a loss function, obtain the Bayes rule, and derive the Bayes risk for (9), (10), and (11) as a function of τ, analogous to our development in the previous section. This may be a fruitful area for additional research for analysts committed to an informative prior approach.

We favor the loss function approach because we feel the circumstances in most analyses are consistent with treating the base case market shares as constraints on the allowable set of answers. As noted earlier, the decision on how to measure market share, the attribute values at the base case scenario, and what constitutes a "close enough" answer are management decisions. While Bayesian analyses are perfectly amenable to incorporating subjective prior information, there is a difference between "prior information" and restrictions on the analysis. Our view is that meeting the base case scenario with the market simulator is an ancillary goal of the analysis. Contrast this with other instances in the marketing literature that use informative priors. Boatwright, McCulloch, and Rossi (1999) use truncated normal distributions to ensure price coefficients are negative in retail/market level sales response models. Allenby, Arora, and Ginter

(1995) enforce an a priori "more is better" ordering on part-worths in conjoint analysis. In these cases, economic theory informed the choice of prior distributions.

The loss function approach has been implemented in the statistics literature to capture ancillary goals of the analyses; see for instance Louis (1984) and Ghosh (1992) who use the loss function to "match" the posterior distributions of parameters to certain empirical distributions of the data. As noted by Shen and Louis (1998), a strength of the Bayesian paradigm is its ability to "structure complicated models, inferential goals, and analyses" and that "methods should be linked to an inferential goal via the loss function." The loss function approach we have outlined matches the baseline market share forecast with minimal changes to the preferences as revealed in the CBC study. The algorithm detailed in the next section allows for straightforward comparisons between the constrained and unconstrained analyses.

4. Estimation

Because B is independent of A and relies only on the CBC data, standard algorithms for estimating hierarchical Bayesian conjoint models can be modified to obtain samples of $B + A$ to use in estimating $E[h(A+B)|\ W, C, \tau)$. Alternatively, current programs can be used to obtain samples of B, and in a separate program A can be sampled such that $A+B \in C$. Well known sampling and Markov-Chain Monte Carlo (MCMC) methods are used to obtain draws from $f(A, B|W, C, \tau)$ from equation (6). As is standard, we exploit conditional independence to draw individual level parameters β_i and α_i.

The generation of the adjustment factors could either be performed inline with the analysis of the CBC model or it could be performed offline and after the MCMC for the CBC data. The inline MCMC algorithm goes through the following steps on each iteration:

1. For i=1,…,N

   1a. Draw $\beta_i|W, X, \theta, \Lambda$

   Use standard methods for either probit or MNL models

   1b. Draw $\alpha_i|\ \beta_i, \{\beta_{-i}\}, \{\alpha_{-i}\}, X_o, \tau, \delta^t$

   Detailed below

2. $\theta|\{\beta_i\}, \Lambda$

   Standard conjugate set-up

3. $\Lambda|\{\beta_i\},\theta$

   Standard conjugate set-up

For the probit model, steps are added for data augmentation and drawing the error-covariance matrix, see McCulloch and Rossi (1994). In step 1b, the notation $\{\beta_{-i}\}$and$\{\alpha_{-i}\}$ indicates the set of parameters for all respondents other than i. Thus the draw of α for person i is dependent on the current value of α and β for all other respondents. The offline algorithm treats the random draws $\{\beta_i^g\}$, for subject i and MCMC iteration g, from the MCMC algorithm as input and draws a separate α_i^g, as detailed below, from $\varphi(\alpha_i|\tau)$ given α_i^g is in $C|\beta_i^g$.

A random walk Metropolis-Hastings (Chib and Greenberg, 1995) or a weighted bootstrap (Smith and Glefand, 1992) is used to draw α_i. Recall that $A+B$ must satisfy $A+B \in C$ or $\chi(S)=1$, the market share constraint in the loss function. For each individual on each iteration, β_i is drawn from $\pi(\beta_i|W)$ and that draw is used in (6) to draw α_i from $f(A, B|W, C, \tau)$. Let $\alpha_i^{(o)}$ represent the draw of α_i from the previous iteration. When the market share constraint equation is evaluated [equation (1)] with then new β_i, the previous $\alpha_i^{(o)}$, $\{\beta_{-i}\}$and$\{\alpha_{-i}\}$, then either $\chi(S) = 1$ and the constraint is satisfied or $\chi(S) = 0$ and it is not. Since β_i is drawn without regard to $\chi(S)$, there is no guarantee that the previous draw of α_i will satisfy the market share constraint.

If the market share constraint is satisfied with $\alpha_i^{(o)}$, then a random walk Metropolis-Hastings (M-H) step is used to draw α_i. Form a candidate or new α_i as $\alpha_i^{(n)} = \alpha_i^{(o)} + \eta$ where η is drawn from $N_p(0, zI_p)$ and z is a scalar chosen to ensure a 50% rejection rate for the M-H step. Note that Λ may be used instead of I_p. The distribution $f(A, B|W, C, \tau)$ from equation (6) implies the following acceptance probability for $\alpha_i^{(n)}$:

$$Min\left[\frac{\exp\left(-\frac{\tau}{2}\alpha_i^{(n)}{}'\alpha_i^{(n)}\right)\chi(S,\alpha_i^{(n)})}{\exp\left(-\frac{\tau}{2}\alpha_i^{(o)}{}'\alpha_i^{(o)}\right)},1\right] \tag{12}$$

Note that if the new $\alpha_i^{(n)}$ does not satisfy $\chi(S)$ then the numerator is 0 and the old value of α_i is retained.

If the market share constraint is not satisfied with $\alpha_i^{(o)}$, then a weighted bootstrap is used to draw a new value of α_i. A challenge in sampling α_i from $f(A, B|W, C, \tau)$ is satisfying the market share constraint. With an appropriately selected proposal density, the weighted bootstrap facilitates this process. Let $\beta_i^{(o)}$ be the value of β_i from the previous iteration and $\beta_i^{(n)}$ be the draw from the current iteration. Define $\mu = \beta_i^{(n)} - (\beta_i^{(o)} + \alpha_i^{(o)})$; given $\{\beta_{-i}\}$and$\{\alpha_{-i}\}$, $\beta_i^{(n)} + \mu$ will satisfy the market share constraint. The target density implied by (6) is $f(\alpha) \propto \varphi(\alpha|\tau)\chi(S)$ and the proposal density for the weighted bootstrap is $g(\alpha) \sim N(\mu, \tau^{-1}I_p)$. A total of R values of $\alpha_i^{(r)}$ are drawn from $g(\alpha)$. For each draw r, calculate the weight w_r:

$$w_r = \frac{\exp\left(-\frac{\tau}{2}\left(\alpha_i^r{}'\alpha_i^r\right)\right)\chi(S,\alpha_i^r)}{\exp\left(-\frac{\tau}{2}\left(\alpha_i^r-\mu)'(\alpha_i^r-\mu\right)\right)} \tag{13}$$

The individual value $\alpha_i^{(r)}$ is then selected with probability:

$$\Pr(\alpha_i = \alpha_i^r) = \frac{w_r}{\sum_{r=1}^{R} w_r} \tag{14}$$

Note that if the market share constraint is not satisfied, $w_r = 0$ and that value of $\alpha_i^{(r)}$ cannot be selected.

The random walk or weighted bootstrap is completed for each i on each iteration of the algorithm. The chain converges to the stationary distribution implied by the posterior distribution $f(A, B|W, C, \tau)$ from the loss function. Note that the algorithm naturally produces draws of both B and $B+A$, whether performed inline or offline; this makes comparison of analyses with and without the market share constraints straightforward. As in other simulation based methods, point estimates of $E[h(A+B)|\, W, C, \tau)$ are obtained by computing $h(A+B)$ for each draw of $A+B$ and averaging over the draws.

Total computation time for the inline algorithm, as compared to models without market share constraints, will depend on the form of the error term used in the model. In order to test $\chi(S)$, market share estimates at X_o will have to be calculated. For the MNL model, choice probabilities are available in closed form and so there is no appreciable increase in computation time in order to calculate base case market share. However, for probit models which are typically estimated via data augmentation, the necessity to calculate choice probabilities at X_o will increase total computation time. The total increase in computation time for the probit model will depend on the specific simulation method used (e.g the GHK or simple frequency simulator) and the accuracy desired (e.g. the number of simulates used). Increases in computation time will also be driven by the discrepancy between the base case and unconstrained market share forecasts; the greater the discrepancy the longer the chain necessary to reach a stationary distribution of parameters that satisfy the market share constraints.

There are several other practical implementation issues that are detailed with suggestions in the appendix. Primary among these is a method for increasing τ in the distribution for $\alpha \sim N_p(0, \tau^{-1}I_p)$ as the MCMC chain progresses, subject to $A+B \in C$. In the next section we show results from simulated data sets including an importance sampling scheme used to estimate the value of the Bayes risk for different values of τ.

5. Simulation Studies

Analyses with simulated data are presented to demonstrate the efficacy of the algorithms and results when the true values of the parameters are known. Results for MNL and correlated probit models are presented. The simulated data set for each analysis consists of 300 respondents, 12 choice sets per respondent, with each choice set consisting of four brands. Each brand in each choice set was described by two randomly generated continuous covariates and one binary covariate intended to mimic a discrete product attribute.

A standard MNL model set-up is used. Choices are restricted to one of the four alternatives and well known algorithms are used to estimate the hierarchical Bayes MNL model. A base set of attributes X_o was randomly generated and the market share using the actual set of β_i

was measured. The market share constraints, e.g. $S_1, \ldots, S_K$, were then set arbitrarily, but different than market share using the true values of the parameters. The model is identified by dropping one of the brand intercepts and the MCMC chain was run for 30,000 iterations with a sample of every 10th from the last 10,000 used to describe posterior moments.

A correlated probit model was simulated with "dual-response" data. In the "dual-response" format, respondents indicate which of the four alternatives they prefer, and then in a second question, indicate whether or not they would actually purchase the item. Data augmentation methods are used to estimate the probit model. The scale of the error term is identified and the error covariance matrix is estimated using Algorithm 3 in Nobile (2000). The location is fixed by requiring the augmented variable w^* for the preferred alternative to be greater than 0 if it would actually be chosen, and less than 0 otherwise. Full details on estimating the probit model via data augmentation are available in McCulloch and Rossi (1994), Nobile (2000), and Rossi, Allenby, and McCulloch (2006). This method of identifying the model allows all four brand intercepts to be estimated. A simple "frequency simulator" with the identifying restrictions and 100 simulates was used to estimate choice probabilities and market shares at X_o. Other aspects of this simulation are similar to the MNL model.

============= Table 1 ============

============== Table 2 ============

Tables 1 and 2 present selected results from the simulations with the target level of discrepancy δ^t set at 0.10, 0.05, and 0.02. In all instances the algorithm was able to increase τ to the maximum value of 100,000 (variance = τ^{-1} = .00001) while meeting the market share constraints. Selected individual level histograms of $\beta_i + \alpha_i$ were inspected and it does not appear that $f(A, B|W, C, \tau)$ is disjoint for these data sets. We therefore present the empirical average of

$\overline{(\beta + \alpha)}$ and the standard deviation averaged across individuals and draws from the MCMC chain. These are provided as a point of comparison to the posterior mean of θ and $\sqrt{\lambda_{pp}}$, from the distribution of heterogeneity. Note that brand intercept parameters are adjusted in the expected direction given the differences between the actual and constrained market share. Figures 2 and 3 plot the individual level average $\overline{(\beta_i + \alpha_i)}$ compared to the individual level average $\overline{\beta}_i$ for selected parameters and models when $\delta^t = .02$. The plots show that individual level parameters are adjusted differently, but that adjustments are generally small and directionally consistent.

============= Figure 2 ============

============== Figure 3 ============

Taken together, these suggest that the loss function approach is capable of meeting the market share constraints with minimal influence on the *average* preference structure from just the CBC data. However, the exact amount of change needed to meet the market share constraints will depend on each data set and the discrepancy with the base case market share. Note also that the average values are not used in market share simulations; they would not necessarily match the base case scenario. The market share constraints are met by using the additive factors A to coordinate the draws of $B + A$ in each iteration of the MCMC chain.

Additional simulations and implications are explored using the MNL model with $\delta^t = .02$; the MNL model was chosen for analytical convenience. First, the Bayes risk was investigated for different values of τ. Using equation (7), the natural log of the Bayes risk was calculated for the function $h(B+A)$ that estimates the market share at the base case scenario X_o. Draws of $B + A$ from the MCMC sampler were used together with an importance sampling algorithm to estimate the log Bayes risk along a grid of values for τ, presented as τ^{-1} in Table 3. Full details on the importance sampling algorithm are available from the authors. The Bayes risk is dominated by ln[P(C|τ)] and since the market share constraints are binding, the smallest value of τ^{-1} minimizes the log Bayes risk. The lower half of Table 3 shows additional quantities calculated using the importance sampler. The $E\left[\sum_{i-1}^{N}\alpha_i'\alpha_i \mid X, C, \tau\right]$ measures the dispersion of A from 0. As expected, it reaches its minimum value at the minimum value of τ^{-1}, but the improvement is marginal beyond τ^{-1}= 0.005. The values of d* equal to E[$h(A + B)$|W, C, τ] were the same up to the third decimal place for all the different values of τ^{-1}.

============== Table 3 ==============

Since the Bayes risk involves var[$h_m(A + B)$], then the optimal value of τ will be dependent on the form of $h(A + B)$, or for purposes of market forecasts, the values of attributes used in the market simulator. A dogmatic Bayesian will determine the optimal value of τ for each different form of $h(A + B)$ and different values of attributes; an importance sampling algorithm can be used for approximating the Bayes risk with any arbitrary $h(A + B)$. However, the above analysis suggests a more pragmatic approach of conditioning on the value of τ. When A is needed, set τ as large as possible and use draws of $h(A + B)$ from f(A, B|W, C, τ) to estimate E[h(A + B)|W, C, τ]. These values can then be used directly in a market simulator (or be used in an importance sampling scheme to determine the optimal value of τ). Although it is possible to investigate the value of τ that minimizes the Bayes risk, the practical benefits of doing so are unclear.

In the next set of simulations, the actual values of β_i were systematically altered to reflect hypothesized distortions in preferences as the result of participating in a discrete choice conjoint study. Forecasted market shares using $B+A$ are then compared to the market share using the actual values of β_i. We find that when the market share constraints are accurate, the loss function approach does a remarkably good job of forecasting.

The actual values of β_i from the original MNL simulation were modified in three different ways. Let β_i^a represent the set of β_i's used in the original simulation. The first modification was to decrease the sensitivity to the 4th product attribute. This was accomplished by setting $\beta_{i4}^b = \beta_{i4}^a + 0.5$ for all i; the remaining coefficients were not changed. Attribute 4 was designed to mimic a price attribute with a true value of θ = -1.0 in the distribution of heterogeneity used to generate β_i. Some analysts believe that respondents aren't as sensitive to price in conjoint studies as they are when making actual choices; this modification was designed to reflect that view. The second modification was to scale all the coefficients by a known constant: $\beta_i^c = \beta_i^a / 0.75$. In the MNL model, this amounts to decreasing the error variance relative to the true model. Finally, the fourth set of coefficients β_i^d reflected both these biases.

Simulated choices were generated using each of the four sets of β's: β_i^a, β_i^b, β_i^c, and β_i^d. For β_i^a, standard hierarchical Bayes methods were used to obtain samples from the posterior distribution of individual level parameters. The base case scenario used a randomly generated X_o and the actual β_i^a to compute the market shares. The base case scenario and the loss function approach were then used to obtain samples of $B^b+A^b \in C$, $B^c+A^c \in C$, and $B^d+A^d \in C$. In addition to market share forecasts at the base case scenario, simulated market shares were generated for a situation where the attributes of Brand C were changed. In the "Product Change" situation, the discrete attribute (attribute 6) was added to Brand C and the value of attribute 4 was decreased by 50%. Because we use simulated data and the actual values of β_i^a are known, we can compare the results of the market simulations to the true values.

============= Table 4 ================

Table 4 shows the market share simulations for the various models using the posterior distributions of β_i's or $B+A$. In each of the three models using $B+A$, the analysis is conditioned on the value of τ^{-1}= 0.00001. Simulated market shares were estimated using a sample size of 1,000 from the appropriate distributions. The discrepancy between the actual market share and the forecasted market share is again measured by δ. Table 4 shows that the loss function approach not only forecasted the base case market share very accurately, but was very accurate when predicting changes to the base case scenario. For instance, in Panel 3 when all coefficients were divided by 0.75, using the posterior of the unadjusted β_i^c the forecasted market share at the base case yielded a discrepancy measure δ = 0.082 whereas the adjusted B^c+A^c yielded a δ = 0.004. When the attributes of Brand C were changed, the forecasted market share using B^c+A^c was very close to the market share using the actual β_i^a, δ = 0.018. The value of δ for the modified values of β_i when using the loss function approach in the "Product Change" situation ranged from 0.024 to 0.012. This compares to δ = 0.024 when the CBC data were generated using β_i^a and standard methods of estimating the model were used. When the managerial base case market shares are accurate, the loss function approach is able to improve market share predictions even for product configurations outside the base case.

6. Empirical Example

This section presents the results of a commercial market research study that involved CBC and managerially supplied market share constraints. Due to the proprietary nature of the dataset, the specific product and the attributes have been disguised. The product category involved a durable consumer electronic device that is typically used in conjunction with another consumer durable. The product is currently available in the market, but management was interested in measuring demand for products that included many new features that were recently developed. Although competitive brands are available in the market, the nature of how the product is purchased made including brand unfeasible in the current study.

The CBC study included 425 respondents who each provided dual response data on 15 choice sets. Each choice set consisted of three alternatives that were uniquely described by 20 binary attributes and the price; as noted above, brand name was not an element in the design matrix. Price was entered as the natural log of price in the response function. Respondents were asked to indicate which of the three alternatives they preferred, and then in a follow-up question, whether they would actually purchase the alternative if it were available in the market. An uncorrelated, dual response probit model was used to represent the likelihood function with a standard hierarchical structure to represent heterogeneity; standard conjugate, but non-informative priors we used to complete the hierarchy.

The base case scenario provided by the study sponsors did not include market share for competing brands. Because of the nature of the product category, management was only able to provide information on the proportion of customers who chose a representative base product, versus the "none" option. The base product was described by the attributes it included and its price. Management provided the choice share for two a priori market segments that were identified by socio-demographic variables. These variables were also available from the CBC study participants. Although the loss function approach was developed assuming that competing brands would make-up the base case scenario, it is straight forward to adopt it to a situation involving a "buy/no buy" choice set with different market segments.

The loss function approach was used to obtain draws from the posterior distributions of $\pi(B|W)$ and $f(A, B|W, C, \tau)$ with $\delta^t = 0.01$. A single analysis that produced draws from both distributions was performed. The algorithm was run for 20,000 iterations. The target δ^t was met at about iteration 3,000 and τ^{-1} met its pre-specified minimum of 0.00001 at about iteration 4,000. A sample of every 10th from the last 10,000 iterations was used to compute summary statistics. Selected individual level histograms of $\beta_i + \alpha_i$ were inspected and it does not appear that $f(A, B|W, C, \tau)$ is disjoint for this data set. Table 5 contains the summary statistics for the constrained and unconstrained parameters. For attributes included in the "base case product profile", the constrained estimates all increased in importance. For attributes not included in the "base case product profile", parameter values generally decreased or stayed the same. The constrained parameters exhibited somewhat larger measures of heterogeneity.

=============== Table 5 ===============

=============== Table 6 ===============

Changes in the parameter estimates make sense given the discrepancy between the base case choice share and that obtained using the unconstrained parameters. All forecasts are based on a sample of 1,000 draws from the individual level posterior distribution of β_i or $\beta_i + \alpha_i$. Table 6 shows that there was a sizable difference between the forecast using the unconstrained parameters and the managerial base case. Using the unconstrained parameters, respondents were forecasted to be much less likely to choose the base product. For market segment #1, the base case was 11.4% choosing the representative product versus a managerial expectation of 43.6%. Thus, it makes sense that attributes included in the base case product would increase in their relative importance in the adjusted parameters. The loss function approach was able to match the base case choice share to within the pre-specified level of accuracy despite the relatively large discrepancy between it and the forecasts using the unadjusted parameters.

Figure 4 plots the individual level average $\overline{(\beta_i + \alpha_i)}$ compared to the individual level average $\overline{\beta_i}$ for selected coefficients. Coefficient #3 was selected because the difference between the posterior mean of θ = 0.570 and the posterior mean of $\overline{(\beta + \alpha)}$ = 0.740 from Table 5 was about average; coefficient #5 was selected because it had the largest difference, θ = -0.317 and $\overline{(\beta + \alpha)}$=-0.002. Table 5 and Figure 4 show that relatively small changes in the β_i's were sufficient in order to meet the choice share constraint. Although the loss function approach is designed to minimize changes to the unconstrained CBC estimates, the exact amount of change needed to individual level parameters will depend on factors such as the number of attributes and the discrepancy between the data and the base case. Since the method produces estimates of both *B* and *B*+*A*, the analysis and changes necessary to meet the market share constraints are completely transparent to both analysts and decision makers.

========== Figure 4 =============

This example shows that the loss function approach is able to meet market share constraints with relatively modest adjustments to individual level parameters using real data, even when there is a big difference between the base case and unconstrained forecast. Further, the method is sufficiently flexible to adapt to "base case scenarios" that differ from the "K-brands" set-up used earlier to define the loss function approach. The algorithm performed as expected, but we anticipate additional research will provide areas for improving its implementation and for further defining the proper boundaries between the likelihood function, the loss function, and prior information in the Bayesian paradigm.

7. Conclusion

In this paper we take the perspective of the marketing research analyst conducting a CBC study who is presented information from the client on current market conditions. In addition to representing the preferences of the study participants, the client expects the analysis will be able to replicate the market results. How should the analyst incorporate this information into his/her analysis? This paper presents a Bayesian decision theoretic approach that uses the loss function to capture the various goals of the decision maker.

Formally, we introduce an additive factor into the loss function that maps draws from the posterior distribution $\pi(B|W)$ from the CBC data into the set that satisfies the market share constraints at the base case scenario. The set of additive factors A are variables from a normal distribution. These are variables in the loss function and not parameters in the likelihood or priors for parameters in the model. We derive the Bayes rule for any function $h(B+A)$ via the risk function; this gives rise to the probability distribution $f(A, B|W, C, \tau)$. Because of the market share constraint, there is a natural dependency in draws of the individual level additive factors α_i. Nonetheless, straightforward methods for drawing samples from $f(A, B|W, C, \tau)$ are available and detailed. Draws of $B+A$ satisfy the market share constraint and can be used by analysts in market share simulators. Conceptually, the loss function approach approximates a procedure that integrates over those regions of the posterior $\pi(B|W)$ that are consistent with the market share constraints.

Simulated and real data sets are used to illustrate the proposed approach. Results from the simulated data show that the *average* representation of preferences changes relatively little using the loss function approach. That is, individual average values of β_i from the posterior $\pi(\beta_i|W)$ are relatively close to the individual average values of $\beta_i + \alpha_i$ from $f(\alpha_i, \beta_i|W, C, \tau)$. We obtained similar results using data from a commercial market research study. The use of a normal distribution with mean 0 minimizes the adjustments at the individual level, and it is simple to illustrate the differences between the constrained and unconstrained analysis. Simulated data is also used to show that it is possible to complete a full Bayesian analysis and select the value of τ that minimizes the Bayes risk, rather than conditioning on a value of τ in the calculation of the Bayes rule

There will naturally be debate on the appropriateness of the loss function approach and whether an analyst should "adjust" the results from a marketing research study to meet the expectations of management. In the case of conjoint analysis, there are compelling theoretical and empirical rationales as to why results may not match actual market place behavior. Given the current state of the field, however, there is little guidance on how to overcome these difficulties. The approach we have outlined will not turn a "bad" conjoint study into a "good" conjoint study. However, rational decision making requires the use of all available information: at issue is how the base case scenario should be incorporated into the analysis. From the analyst's standpoint, the loss function approach truthfully describes the managerial goals as constraints on the final estimates and not as “data” with the same standing as the CBC experiment. As in all analyses, the analyst must fully disclose the methods and all assumptions used. The approach outlined in this paper has the advantage of providing a sample from the posterior distribution from a standard Bayesian analysis, and a sample from a supplemental distribution of parameters that condition on the posterior, but satisfy the market share constraints.

This paper highlights the need for additional research in a number of areas. First, models are needed that link market data to marketing research data when the actual behavior of research participants is not observed. Multiple observations of "base case scenarios and market share" will likely enhance the loss function approach described here, and perhaps enable modeling the data through the likelihood. As the proposed method is adopted and used on multiple data sets, we anticipate that improvements to the algorithm and or better guidance on implementation issues will be identified. Experience with multiple data sets may also help to identify empirical generalizations that can then be used to create theories and models linking stated and revealed preference data. Additionally, this paper proposed a loss function that linked the goals of the decision maker to the statistical analysis of CBC data; different loss functions for this situation are possible and should be explored. Although Bayesian methods have been used in marketing to produce posterior distributions of quantities of interest, such as parameters, expected profit, expected market share, etc, there is little published research that completes the Bayesian decision theoretic approach and formally incorporates the loss function facing the decision maker. We hope this paper encourages additional research on the use of loss functions and Bayesian decision theory in marketing decision making.

APPENDIX: *Implementation Issues*

Several practical issues must be addressed in order to implement the algorithm described in the paper. Our suggestions take advantage of the "burn-in" period associated with MCMC chains to set initial conditions and to make τ as large as possible. First, it would be difficult to choose initial conditions to satisfy $\chi(S)$ given δ^t. We therefore choose a convenient initial value of δ^t, δ^{int} and decrease δ^{int} at a steady rate until $\delta^{int} \leq \delta^t$. We set the initial value equal to that obtained when B=A=0. Recall that the value of δ^{int} is used on each iteration to determine if $A+B \in C$ or $\chi(S)$=1. In our algorithms we decrease δ^{int} by .001 every 25 iterations.

The weighted bootstrap may not draw any candidate values $\alpha_i^{(r)}$ that satisfy the market share constraint. This may happen on iterations when the value of δ^{int} decreases and/or on any iteration because the values of $\alpha_i^{(r)}$ are drawn stochastically. One way to deal with this is to make R, the number of simulates drawn, sufficiently large. However, making R very large slows down the program. Instead, we start with a moderate sized R_1 of approximately 100. If R_1 produces no $\alpha_i^{(r)}$ that satisfies $\chi(S)$, we increase R by a factor of 3 and redo the bootstrap with R_2. We repeat this process up to m* times. If $\chi(S)$ is still not satisfied, we increase the current value of δ^{int} by an arbitrary amount and repeat the process of drawing $\alpha_i^{(r)}$, but now with a different $\chi(S)$. In our algorithms, if after approximately 35,000 draws of $\alpha_i^{(r)}$ we do not satisfy $\chi(S)$, we increase the current value of δ^{int} by 1/N.

The value of τ will effect the performance of the MCMC sampler at different points in the chain. If τ is too large, particularly early in the chain as δ^{int} is decreasing, it may be difficult to sample α_i to satisfy the market share constraints. However, recall that the analysis of the

Bayes risk implied that we want τ to be as large as possible subject to $A+B \in C$. A practical approach is to start τ at a small value, and then to increase it at a steady rate, as long as the algorithm can draw α_i's that satisfy the market share constraint. Across iterations, if δ^{int} is not making progress toward δ^t, then decrease τ by some small amount. In our algorithms we start with $\tau \sim 1$ and we check the progress of δ^{int} every 50 iterations: if $\delta^{int} \leq \delta^{int}_{-100}$ then we increase τ, otherwise we decrease τ. The amount we increase or decrease τ depends on its current value. The maximum value τ can attain in our programs is 100,000.

In summary, we suggest:

1. Set the initial $\delta^t = \delta^{int} = \sum_{i=1}^{K} \left| S_k - 1/K \right|$ and A=B=0.

2. After each 25 iterations, decrease δ^{int} until $\delta^{int} \leq \delta^t$.

3. In the weighted bootstrap, if after R simulates, $\sum_{r=1}^{R} w_r = 0$, increase R and repeat. If unsuccessful after m* attempts, increase δ^{int}.

4. Start with τ between 1/3 and 2 and increase τ as long as δ^{int} is progressing to δ^t.

For any given dataset, the overall performance of the algorithm can be improved. For instance, resetting δ^{int} after 50 or 250 initial iterations may speed-up the progress of decreasing δ^{int} to δ^t. The effectiveness of this strategy depends on the target market share, the value of the predicted market share using the current values of B, and how fast the MCMC chain is burningoff the initial values. Also, adjusting the frequency with which δ^{int} is changed and its step size may change the performance of the algorithm. Similar comments apply for the initial value of τ and its adjustment process. Our experience suggests that large discrepancies between the base case scenario and the market share implied by β_i require a more conservative approach to checking and adjusting the values of δ^{int} and τ.

Table 1
Simulation Results for MNL

Brand	MS at Xo using $\{\beta_i\}$	MS Constraint at Xo		Actual θ
A	0.54	0.40	1)	-0.43
B	0.25	0.35	2)	0.50
C	0.05	0.15	3)	0.34
D	0.15	0.10	4)	-1.00
			5)	0.82
			6)	1.47

δ^t = 0.10

Brand	Posterior MS at Xo			Posterior θ		$\sqrt{\lambda_{pp}}$		Posterior average $(\alpha+\beta)$		$\sigma_{(\alpha+\beta)_{pp}}$
A	0.417	(0.008)	1)	-0.386	(0.09)	0.920	1)	-0.615	(0.07)	0.992
B	0.348	(0.006)	2)	0.466	(0.09)	0.998	2)	0.672	(0.07)	1.078
C	0.107	(0.005)	3)	0.309	(0.09)	0.978	3)	0.432	(0.07)	1.099
D	0.128	(0.007)	4)	-1.010	(0.08)	1.025	4)	-0.793	(0.06)	1.096
			5)	0.782	(0.07)	0.992	5)	0.789	(0.04)	1.013
	τ^{-1} = .00001		6)	1.520	(0.08)	0.994	6)	1.166	(0.05)	1.096

δ^t = 0.05

Brand	Posterior MS at Xo			Posterior θ		$\sqrt{\lambda_{pp}}$		Posterior average $(\alpha+\beta)$		$\sigma_{(\alpha+\beta)_{pp}}$
A	0.404	(0.005)	1)	-0.420	(0.09)	0.915	1)	-0.638	(0.07)	0.988
B	0.349	(0.003)	2)	0.449	(0.08)	0.986	2)	0.649	(0.06)	1.077
C	0.129	(0.003)	3)	0.291	(0.09)	0.989	3)	0.479	(0.07)	1.162
D	0.118	(0.005)	4)	-1.022	(0.08)	1.020	4)	-0.767	(0.05)	1.120
			5)	0.788	(0.07)	0.996	5)	0.792	(0.03)	1.023
	τ^{-1} = .00001		6)	1.524	(0.09)	0.993	6)	1.108	(0.06)	1.111

δ^t = 0.02

Brand	Posterior MS at Xo			Posterior θ		$\sqrt{\lambda_{pp}}$		Posterior average $(\alpha+\beta)$		$\sigma_{(\alpha+\beta)_{pp}}$
A	0.402	(0.003)	1)	-0.413	(0.09)	0.899	1)	-0.630	(0.07)	0.982
B	0.349	(0.002)	2)	0.453	(0.09)	0.989	2)	0.642	(0.06)	1.098
C	0.143	(0.002)	3)	0.296	(0.09)	0.995	3)	0.526	(0.07)	1.203
D	0.106	(0.003)	4)	-1.025	(0.09)	1.017	4)	-0.757	(0.06)	1.141
			5)	0.785	(0.07)	0.987	5)	0.806	(0.03)	1.017
	τ^{-1} = .00001		6)	1.518	(0.09)	1.003	6)	1.062	(0.05)	1.130

MS = Market Share
Xo = "Base case" design matrix

Posterior mean and (standard deviation) displayed.

Table 2
Simulation Results for Probit

Brand	MS at Xo using $\{\beta_i\}$	MS Constraint at Xo		Actual θ
A	0.40	0.30	1)	-0.05
B	0.22	0.35	2)	1.03
C	0.18	0.25	3)	0.79
D	0.20	0.10	4)	0.51
			5)	-0.91
			6)	0.92
			7)	1.49

δ =.10

Brand	Posterior MS at Xo			Posterior θ		$\sqrt{\lambda_{pp}}$		Posterior average $(\alpha + \beta)$		$\sigma_{(\alpha+\beta)_{pp}}$
A	0.304	(0.009)	1)	-0.114	(0.13)	0.994	1)	-0.211	(0.11)	1.070
B	0.332	(0.011)	2)	0.817	(0.12)	0.977	2)	1.199	(0.12)	1.197
C	0.226	(0.011)	3)	0.618	(0.13)	0.980	3)	0.848	(0.12)	1.237
D	0.137	(0.008)	4)	0.451	(0.12)	0.915	4)	0.384	(0.12)	1.046
			5)	-0.879	(0.08)	0.882	5)	-0.858	(0.07)	0.892
			6)	0.877	(0.08)	0.972	6)	0.982	(0.06)	1.000
	τ^{-1} = .00001		7)	1.408	(0.11)	1.007	7)	1.127	(0.11)	1.070

δ =.05

Brand	Posterior MS at Xo			Posterior θ		$\sqrt{\lambda_{pp}}$		Posterior average $(\alpha + \beta)$		$\sigma_{(\alpha+\beta)_{pp}}$
A	0.302	(0.006)	1)	-0.079	(0.13)	0.984	1)	-0.136	(0.12)	1.090
B	0.342	(0.008)	2)	0.856	(0.13)	0.959	2)	1.280	(0.14)	1.222
C	0.239	(0.007)	3)	0.662	(0.13)	0.981	3)	0.963	(0.13)	1.272
D	0.117	(0.005)	4)	0.494	(0.13)	0.900	4)	0.397	(0.14)	1.074
			5)	-0.866	(0.08)	0.876	5)	-0.875	(0.07)	0.913
			6)	0.859	(0.07)	0.956	6)	0.995	(0.06)	1.035
	τ^{-1} = .00001		7)	1.377	(0.10)	0.989	7)	1.066	(0.08)	1.075

δ =.02

Brand	Posterior MS at Xo			Posterior θ		$\sqrt{\lambda_{pp}}$		Posterior average $(\alpha + \beta)$		$\sigma_{(\alpha+\beta)_{pp}}$
A	0.301	(0.004)	1)	-0.068	(0.13)	0.975	1)	-0.112	(0.11)	1.135
B	0.347	(0.004)	2)	0.833	(0.13)	0.949	2)	1.292	(0.14)	1.275
C	0.247	(0.004)	3)	0.648	(0.13)	0.954	3)	0.974	(0.12)	1.326
D	0.105	(0.003)	4)	0.483	(0.12)	0.882	4)	0.371	(0.12)	1.128
			5)	-0.854	(0.08)	0.858	5)	-0.869	(0.07)	0.943
			6)	0.845	(0.08)	0.934	6)	0.993	(0.08)	1.014
	τ^{-1} = .00001		7)	1.347	(0.11)	0.962	7)	1.019	(0.10)	1.098

MS = Market Share
Xo = "Base case" design matrix

Posterior mean and (standard deviation) displayed.

Table 3
Importance-Sampling Results for MNL Model

Estimates of Log Bayes Risk and its Components

τ^{-1}	$\ln[r(\tau)]$	$\ln[P(C \mid \tau)]$	$\ln\left[\sum_{m=1}^{M} \mathrm{var}\left[h_m(A+B) \mid X, C, \tau\right]\right]$
0.1	-2,630.13	-2,625.64	-4.48
0.05	-6,000.95	-5,996.46	-4.49
0.01	-33,440.36	-33,435.87	-4.49
0.005	-68,188.66	-68,184.17	-4.49
0.001	-346,274.81	-346,270.32	-4.49
0.0001	-3,474,688.49	-3,474,684.00	-4.49
0.00001	-34,759,079.73	-34,759,075.24	-4.49

Estimates of Bayes Rule for Various Quantities

τ^{-1}	d* *Brand A*	d* *Brand B*	d* *Brand C*	d* *Brand D*	$E\left[\sum_{i=1}^{N} \alpha_i' \alpha_i \mid X, C, \tau\right]$
0.1	0.400	0.348	0.146	0.106	719.882
0.05	0.401	0.349	0.145	0.106	685.926
0.01	0.402	0.349	0.144	0.106	650.930
0.005	0.402	0.349	0.143	0.106	645.793
0.001	0.402	0.349	0.143	0.106	641.802
0.0001	0.402	0.349	0.143	0.106	640.861
0.00001	0.402	0.349	0.143	0.106	640.857

Table 4
Market Share Simulation: Adding a Product Feature and Decreasing x_4

	Base Design Matrix						*Product Change Design Matrix*					
	x_1	x_2	x_3	x_4	x_5	x_6	x_1	x_2	x_3	x_4	x_5	x_6
Brand A	1	0	0	0.0008	1.3599	1	1	0	0	0.0008	1.3599	1
Brand B	0	1	0	0.583	1.3992	0	0	1	0	0.583	1.3992	0
Brand C	0	0	1	1.4078	-0.457	0	**0**	**0**	**1**	**0.7039**	**-0.457**	**1**
Brand D	0	0	0	0.8926	-0.796	1	0	0	0	0.8926	-0.796	1

Panel 1

No Changes: β_i^a

Using ***actual*** β_i^a

Base Case Brand	Mkt Share	**Product Change** Brand	Mkt Share
A	0.54	A	0.476
B	0.26	B	0.231
C	0.05	C	0.187
D	0.15	D	0.106

Using posterior β_i^a

Base Case Brand	Mkt Share	**Product Change** Brand	Mkt Share
A	0.551	A	0.484
B	0.245	B	0.221
C	0.054	C	0.191
D	0.150	D	0.104
$\delta =$	0.030	$\delta =$	0.024

Panel 2

$\beta_i^b = \beta_i^a$ and $\beta_{4i}^b = \beta_{4i}^a + 0.5$

Using posterior β_i^b

Base Case Brand	Mkt Share	**Product Change** Brand	Mkt Share
A	0.503	A	0.441
B	0.245	B	0.220
C	0.070	C	0.209
D	0.181	D	0.130
$\delta =$	0.103	$\delta =$	0.092

Using Posterior $B^b + A^b$

Base Case Brand	Mkt Share	**Product Change** Brand	Mkt Share
A	0.538	A	0.479
B	0.258	B	0.236
C	0.053	C	0.175
D	0.150	D	0.110
$\delta =$	0.007	$\delta =$	0.024

Panel 3

$\beta_i^c = \left(\beta_i^a / 0.75\right)$

Using posterior β_i^c

Base Case Brand	Mkt Share	**Product Change** Brand	Mkt Share
A	0.577	A	0.506
B	0.231	B	0.209
C	0.038	C	0.186
D	0.154	D	0.098
$\delta =$	0.082	$\delta =$	0.061

Using Posterior $B^c + A^c$

Base Case Brand	Mkt Share	**Product Change** Brand	Mkt Share
A	0.541	A	0.476
B	0.260	B	0.237
C	0.048	C	0.190
D	0.151	D	0.097
$\delta =$	0.004	$\delta =$	0.018

Panel 4

$\beta_i^d = \left(\beta_i^a / 0.75\right)$ and $\beta_{4i}^d = \left(\beta_{4i}^a + 0.5 / 0.75\right)$

Using posterior β_i^d

Base Case Brand	Mkt Share	**Product Change** Brand	Mkt Share
A	0.530	A	0.462
B	0.241	B	0.219
C	0.062	C	0.209
D	0.168	D	0.110
$\delta =$	0.059	$\delta =$	0.052

Using Posterior $B^d + A^d$

Base Case Brand	Mkt Share	**Product Change** Brand	Mkt Share
A	0.539	A	0.475
B	0.259	B	0.237
C	0.052	C	0.187
D	0.151	D	0.101
$\delta =$	0.005	$\delta =$	0.012

Table 5
Selected Summary Statistics
Consumer Electronics Study

Attribute	Posterior θ		$\sqrt{\lambda_{pp}}$	Posterior average (α + β)		$\sigma_{(\alpha+\beta)_{pp}}$	Attribute included in base case?
1	-2.827	(0.19)	2.940	-2.752	(0.13)	3.011	Yes
2	0.368	(0.07)	1.128	0.601	(0.06)	1.411	Yes
3	0.570	(0.07)	1.045	0.740	(0.06)	1.386	Yes
4	0.777	(0.07)	0.943	1.045	(0.07)	1.425	Yes
5	-0.317	(0.06)	0.790	-0.002	(0.06)	1.281	Yes
6	0.291	(0.06)	0.882	0.236	(0.06)	1.207	No
7	0.643	(0.06)	0.766	0.504	(0.06)	1.219	No
8	0.187	(0.06)	0.831	0.139	(0.07)	1.267	No
9	0.610	(0.06)	1.020	0.825	(0.06)	1.384	Yes
10	0.481	(0.06)	0.844	0.434	(0.07)	1.272	No
11	0.367	(0.06)	0.859	0.128	(0.06)	1.254	No
12	-0.025	(0.06)	0.853	-0.010	(0.06)	1.246	No
13	-0.010	(0.06)	0.814	-0.136	(0.06)	1.253	No
14	0.181	(0.05)	0.704	0.207	(0.06)	1.173	No
15	0.305	(0.05)	0.720	0.181	(0.06)	1.210	No
16	0.090	(0.06)	0.793	0.069	(0.05)	1.207	No
17	0.122	(0.06)	0.764	0.339	(0.06)	1.275	Yes
18	0.256	(0.07)	0.766	0.092	(0.07)	1.330	No
19	-0.182	(0.07)	0.748	-0.433	(0.07)	1.309	No
20	0.873	(0.08)	1.059	0.758	(0.07)	1.382	No
ln(price)	-0.334	(0.04)	0.639	-0.202	(0.04)	0.759	Yes

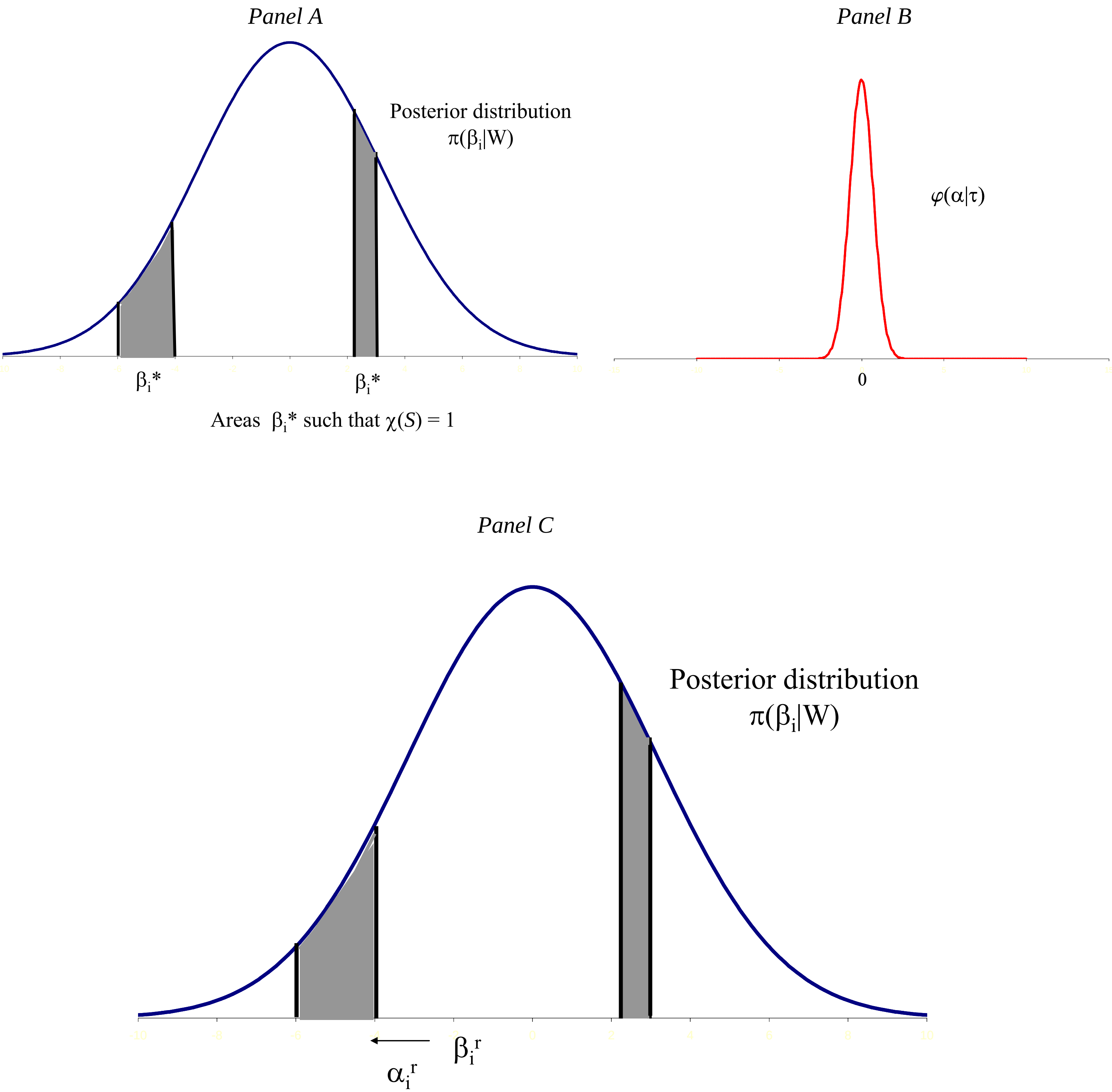
Posterior mean and (standard deviation) displayed.

Table 6
Choice Share Constraints
Consumer Electronics Study

	Forecasted Base Case Choice Share using $\{\beta_{hi}\}$	*Managerial Base Case Choice Share*	Forecasted Base Case Choice Share using $\{\beta_{hi} + \alpha_{hi}\}$
Market Segment #1	11.40%	*43.6%*	43.58%
Market segment #2	12.18%	*55.6%*	55.51%
δ Compared to Managerial Base Case	0.756		0.001

Posterior mean for forecasts displayed

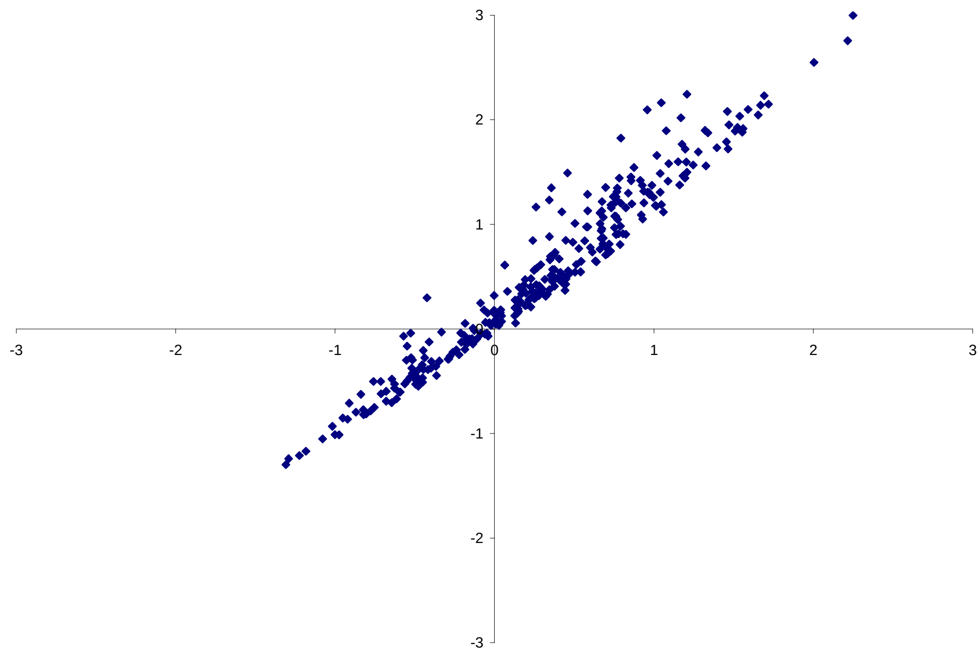
Figure 1
Mapping into Constrained Area of the Posterior Distribution



For any random draw β_i^r from $\pi(\beta_i|W)$, map into area satisfying $\chi(S)$ via $(\beta_i^r + \alpha_i^r)$ with α_i^r drawn from $\varphi(\alpha|\tau)$.

Figure 2
Simulated MNL Model: Posterior Mean of Individual Level Parameters

$\overline{\beta}_{i3}$ vs. $\overline{\beta_{i3} + \alpha_{i3}}$



$\overline{\beta}_{i5}$ vs. $\overline{\beta_{i5} + \alpha_{i5}}$

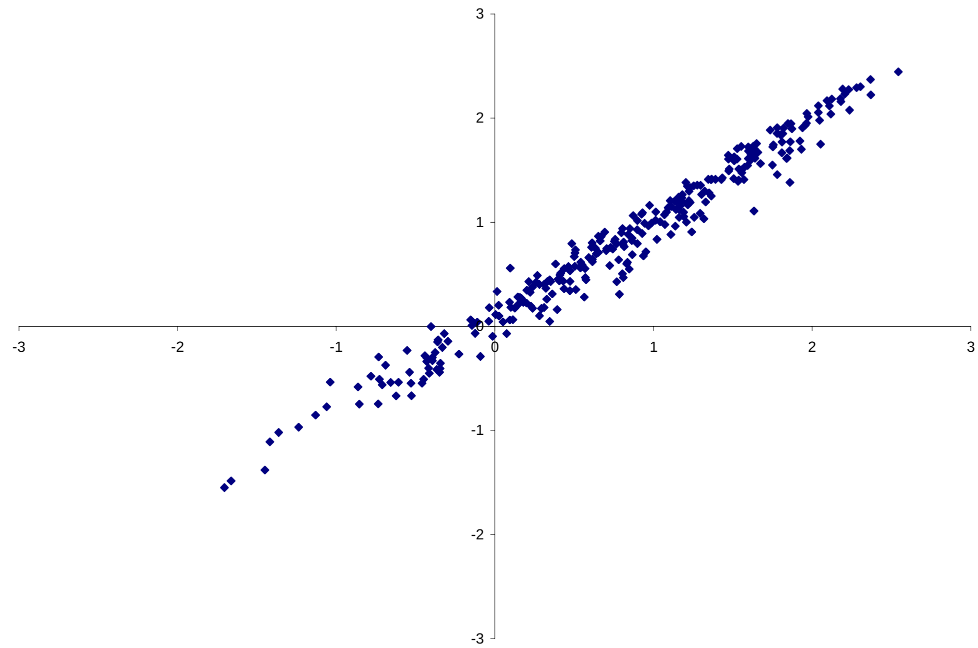
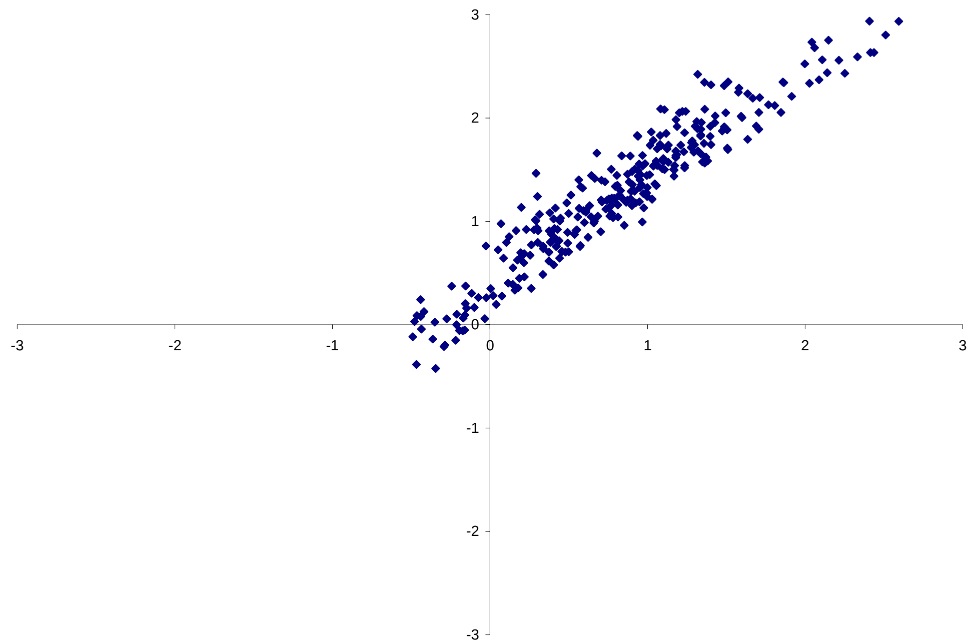


Figure 3
Simulated Probit Model: Posterior Mean of Individual Level Parameters

$\overline{\beta}_{i2}$ vs. $\overline{\beta_{i2} + \alpha_{i2}}$



$\overline{\beta}_{i7}$ vs. $\overline{\beta_{i7} + \alpha_{i7}}$

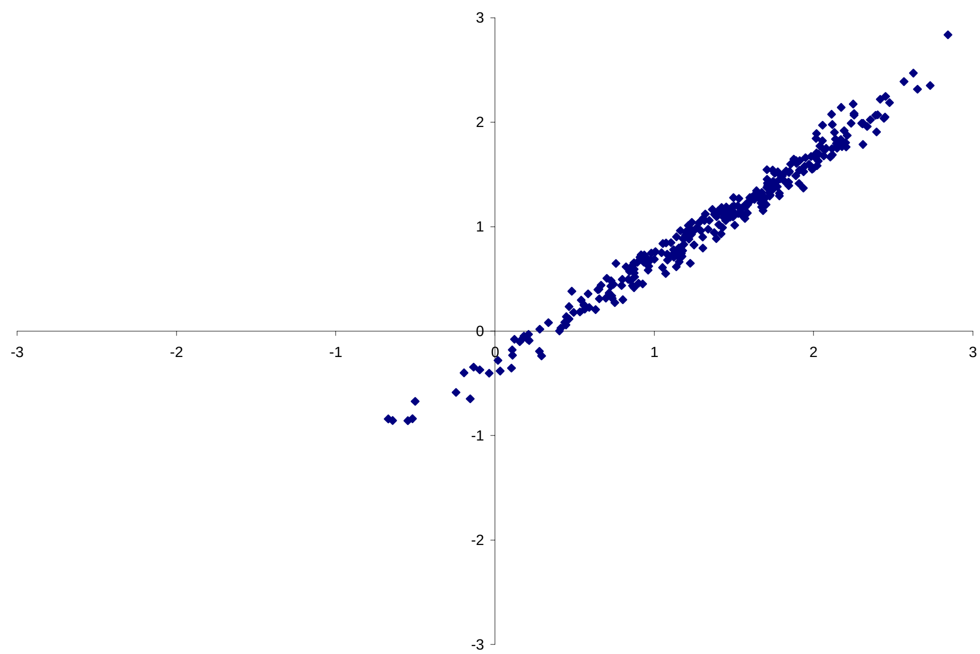
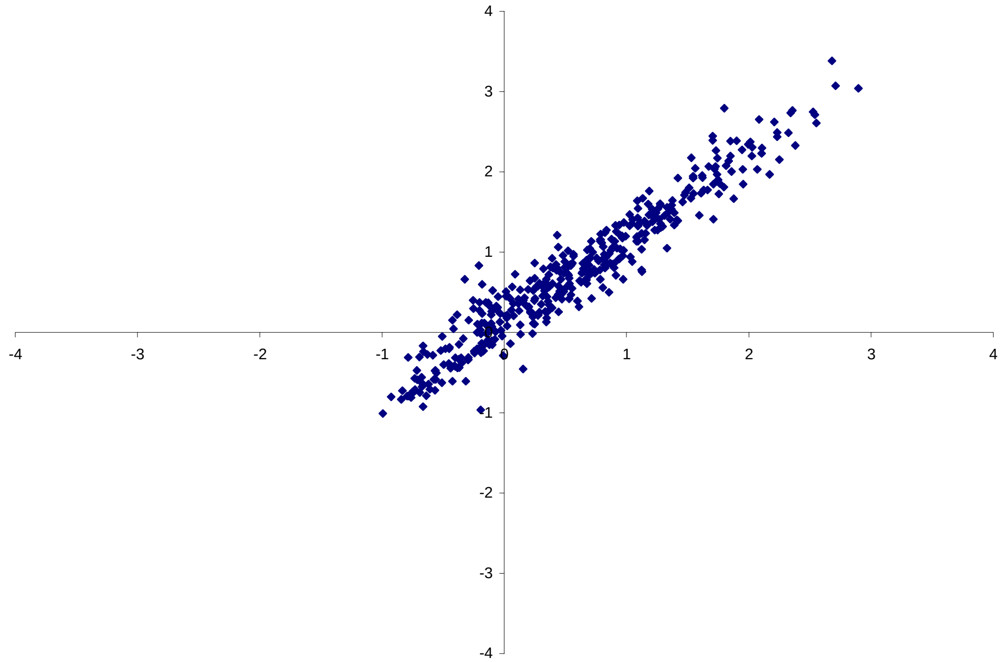
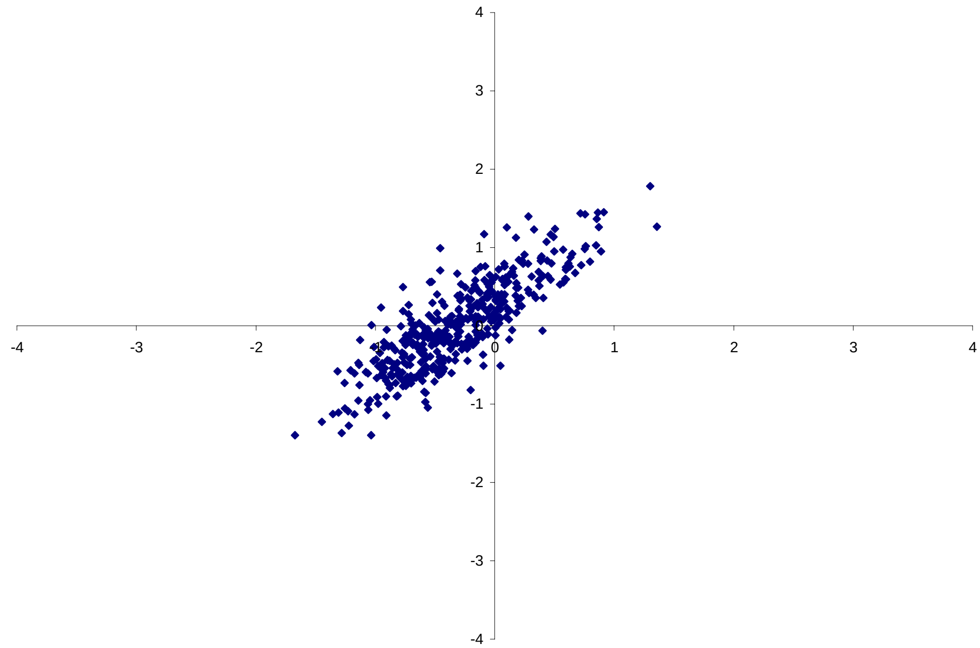


Figure 4
Consumer Electronics Study: Posterior Mean of Individual Level Parameters

$\overline{\beta}_{i3}$ vs. $\overline{\beta_{i3} + \alpha_{i3}}$



$\overline{\beta}_{i5}$ vs. $\overline{\beta_{i5} + \alpha_{i5}}$



References

Allenby, Greg M., Neeraj Arora, and James L. Ginter (1995), "Incorporating Prior Knowledge into the Analysis of Conjoint Studies," *Journal of Marketing Research*, 32, (May), 152-162.

Allenby, Greg, Geraldine Fennell, Joel Huber, Thomas Eagle, Tim Gilbride, Dan Horsky, Jaehwan Kim, Peter Lenk, Rich Johnson, Elie Olfek, Brian Orme, Joan Walker, (2005) "Adjusting Choice Models to Better Predict Market Behavior", *Marketing Letters*, 16 (3/4) pp. 197-208.

Bernardo, Jose M. and Adrian F. M. Smith (2000), *Bayesian Theory*. John Wiley & Sons, Ltd. West Sussex, England.

Boatwright, Peter, Robert McCulloch, and Peter Rossi (1999), "Account-Level Modeling for Trade Promotion: An Application of a Constrained Parameter Hierarchical Model," *Journal of the American Statistical Association*, 94, (December), 1063-1073.

Cattin, Philippe and Dick R. Wittink (1982), "Commercial Use of Conjoint Analysis: A Survey," *Journal of Marketing Research*, 46, (Summer), 44 – 53.

Chib, Siddhartha and Edward Greenberg (1995), "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, Vol. 49, No. 4 (November), 327-335.

DeGroot , Morris (1970), *Optimal Statistical Decision*. McGraw-Hill, New York.

Ghosh, Malay (1992), "Constrained Bayes Estimation with Applications," *Journal of the American Statistical Association*, 87, (June), 533-540.

Louis, Thomas A. (1984), "Estimating a Population of Parameter Values Using Bayes and Empirical Bayes Methods," *Journal of the American Statistical Association*, 79, (June), 393-398.

Louviere, Jordan J. (2002), "What if Consumer Experiments Impact Variances as well as Means? Response Variability as a Behavioral Phenomenon," *Journal of Consumer Research*, 28, (December), 506 – 511.

McCulloch, Robert and Peter E. Rossi (1994), "An Exact Likelihood Analysis of the Multinomial Probit Model," *Journal of Econometrics*, 64, 207 – 240.

Morikawa, Taka, Moshe Ben-Akiva, and Daniel McFadden (2002) "Discrete Choice Models Incorporating Revealed Preferences and Psychometric Data" in *Econometric Models in Marketing*, Vol. 16, 29-55.

Nobile, Agostino (2000), "Comment: Bayesian Multinomial Probit Models with a Normalization Constraint," *Journal of Econometrics*, 99, 335 – 345.

Orme, Bryan and Rich Johnson (2006), "External Effect Adjustments in Conjoint Analysis", Working paper, Sawtooth Software.

Rossi, Peter E., Greg M. Allenby, and Robert McCulloch (2005), *Bayesian Statistics and Marketing.* West Sussex, England: John Wiley & Sons, Ltd.

Shen, Wei and Thomas A. Louis (1998), "Triple Goal Estimates in Two-Stage Hierarchical Models," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 60, (2), 455-471.

Smith, A. F. M. and A. E. Gelfand (1992), "Bayesian Statistics Without Tears: A Sampling-Resampling Perspective," *The American Statistician*, 46, (May), 84-88.