Marketing Science

Market Share Constraints and the Loss Function in Choice-Based Conjoint Analysis

Timothy J. Gilbride, Peter J. Lenk, Jeff D. Brazell,

Please scroll down for article—it is on subsequent pages

Market Share Constraints and the Loss Function in Choice-Based Conjoint Analysis

Timothy J. Gilbride

Mendoza College of Business, University of Notre Dame, Notre Dame, Indiana 46556, tgilbrid@nd.edu

Peter J. Lenk

Stephen M. Ross School of Business, University of Michigan, Ann Arbor, Michigan 48109, plenk@umich.edu

Jeff D. Brazell

The Modellers, LLC, Salt Lake City, Utah 84047, jeff.brazell@themodellers.com

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 prespecified or "base-case" scenario. Frequently, there is a discrepancy between the recovered 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 discrete choice models with simulated data and data from a commercial market research study. These studies indicate that the method recovers base-case market shares without systematically distorting the preference structure from the conjoint experiment.

Key words: Bayesian decision theory; conjoint analysis; constrained optimization; cross-validation; hierarchical Bayes; loss function; market share prediction; penalized maximum likelihood; posterior risk
History: This paper was received December 1, 2006, and was with the authors 4 months for 2 revisions; processed by John Hauser. Published online in *Articles in Advance* July 31, 2008.

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. Although there are many reasons why the forecasted and actual market shares 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 managers' "base-case" scenario expectations. This paper presents a Bayesian decision theoretic approach to incorporating base-case scenario projections into choice-based 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, and 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 market simulators and instead argue that managers should be educated on the reasons why simulated and actual market shares 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. Ding et al. (2005) propose incentive aligned conjoint experiments to improve out-of-sample predictions.

In a Bayesian analysis, the base-case market share can be modeled as part of the likelihood function, as informative prior information, or through the loss function. 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 et al. (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 (i.e., market data are 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 due to changes in prices, 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. These choices are unique to the manager and will largely be dictated by the specific decision or forecasting context.

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. The second objective is that 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 to leave the preference structure of the CBC data relatively intact. The main purpose of most conjoint studies is to project beyond current market assumptions, and our method maintains fidelity of the CBC preferences structure as the assumptions move away from the base case. Our adjustment terms are additive factors with mean zero, and we specify the variance to be as small as possible while satisfying the constraints. By minimizing the adjustments to the CBC data we explicitly assume the conjoint study yields useful information; thus, the loss function approach will not turn a poorly designed or executed study into a "good" study. Finally, we separate the estimation of the CBC parameters from the process of perturbing them to match the constraints. This allows the analyst to judge the degree of adjustment and to maintain the integrity of the CBC results.

We contrast the loss function approach with several others including treating 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. Should the analyst treat the base-case market shares for M products as M bits of information, or should he or she treat them as millions of transactions? In the former case, the CBC data will overwhelm the market information, and one will not recover the base-case market shares. In the later case, the CBC data, which are based on a few thousand hypothetical choices, will be overshadowed by the market data and will be less useful for projecting new product ideas and positioning. 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.

Although Bayesian analyses are perfectly amenable to incorporating subjective prior information, there is a difference between "prior information" and base-case market share 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 to reflect essential features of the model. Boatwright et al. (1999) use truncated normal distributions to ensure price coefficients are negative in retail/market-level sales response models, and Allenby et al. (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 outline matches the baseline market share forecast with minimal changes to the preferences as revealed in the CBC study.

Our solution dovetails nicely with hierarchical Bayes (HB) inference, which is commonly used to analyze conjoint studies to obtain individual-level estimates. However, constrained estimation also has a rich tradition in non-Bayesian statistics (cf. Aitchison and Silvey 1958, Dupacova and Wets 1988, Liew 1976, Wang 1996). These papers not only solve the constrained optimization problem, but also describe the statistical properties of the solution. We illustrate the use of market share constraints in non-Bayesian models by adding them to the mathematical programming procedure of Evgeniou et al. (2007), which also produces individual-level estimates in CBC. The underlying HB and mathematical programming models are nearly the same, but the estimation method differs: Markov chain Monte Carlo (MCMC) versus constrained optimization. Consequently, the results of the two procedures are nearly identical.

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 with approaches using informative priors and a non-Bayesian method. Simulation and 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 without the market share constraint. We illustrate the proposed methods using simulated data for the multinomial logit (MNL) discrete choice model and examine in-sample and out-of-sample performance. 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 exact form of the model is not important to our loss function adjustment procedure, but to fix ideas the model set-up is as follows. N subjects evaluate J_i choice occasions for subject i. Each choice occasion or task consists of M alternatives or product profiles. In the CBC design, there are p attribute levels, including brand intercepts. Subject i's latent random utility for product profile j is a given by

$$Y_{ij} = x'_{ij}\beta_i + \varepsilon_{ij}, \quad \text{where } \varepsilon_i \sim \text{normal or extreme value}$$

$$\beta_i = \theta + \nu_i, \quad \text{where } \nu_i \sim N_p(0, \Lambda),$$

where the observed attributes x_{ij} and the parameter vector β_i are p-vectors, and ε_{ij} are error terms. 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 let $S_1, \ldots, S_K$ represent the corresponding market shares. Let $\hat{S}_1, \ldots, \hat{S}_K$ represent the estimated market shares using X_o and $\{\beta_i\}$, e.g., $g(\{\beta_i\}, X_o)$. We define the discrepancy between the estimated and base-case market shares as

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

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 $\delta \leq \delta^t$.

2.2. Loss Functions

In decision analysis (cf. 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.) Savage (1954) derived Bayesian loss functions from the axioms of rational preferences and developed a decision-theoretic approach to inference. Loss functions are closely related to utility functions and therefore represent the individual circumstances of the decision context and the decision maker, and not necessarily universal objective functions. Because the state of nature is not known, the Bayesian decision maker wishes to minimize the posterior expected loss represented by

$$\rho(\pi, d) = \int_\theta L(\theta, d) \pi(\theta \mid W) \, d\theta, \tag{2}$$

where π represents the current distribution of θ given the observed data W, in this case the posterior. The Bayes rule is the selection of $d \in D$, the set of allowable decisions, that minimizes $\rho(\pi, d)$. The Bayes risk integrates (2) over the sample space for W. The value of d that minimizes (2) also minimizes the Bayes risk (cf. Casella and Berger 1990, p. 474). The overall risk to the decision maker is the weighted average of how likely an event is to occur and the loss to the decision maker for taking action d if it does occur.

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 $L(\theta, d) = (\theta - d)^2$ for a scalar parameter and $L(\theta, d) = (\theta - d)'(\theta - d)$ 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

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

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, Bayesian decision analysis can be applied to any loss function. For example, absolute error loss results in the Bayes rule being the median of the posterior distribution (Berger 1985, p. 162), and 0/1 loss gives the posterior mode: the value of θ that maximizes the posterior density (Berger 1985, p. 162). The expected posterior loss at the Bayes rules quantifies estimation risk. Under squared error loss, this measure is the posterior variance of θ.

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

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

where $B = \{\beta_i\}$ 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 g. The functional g can be of any form and will not generally involve X_o. Here, the requirement that only coefficients that satisfy the market share constraints are used is represented as $B \in C$, which means that $\beta_i \in C$ for $\beta_i \in B$. This can be seen as limiting the set of allowable decisions, $d \in D$. This loss function says that the decision maker gets no utility (e.g., the loss function is ∞) from any forecast if the analysis lacks face validity, where "face validity" is defined in terms of the base-case forecast being close to the manager's definition of the base-case market share. We use the loss function because "face validity" is defined by the individual decision maker and is specific to each decision context.

In addition to meeting the market share constraints, we want to make sure that the changes to the preference structure are, in some sense, minimal and we want to explicitly separate the estimated $\{\beta_i\}$ from the conjoint study from the constrained estimates. To do so, we introduce additive factors $\{\alpha_i\}$ for each subject where each α_i is a p-vector, and use the adjusted parameters $\{\alpha_i + \beta_i\}$ in computing market share forecasts. 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. To regularize the problem, we assume that $\alpha_i \sim N_p(0, \tau^{-1} I_p)$, where I_p is a $p \times p$ identity matrix. Our procedure then forces each α_i to be as close to 0 as possible by making τ as large as possible; this will be used to help identify the $\{\alpha_i\}$.

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 as close as possible to the unadjusted estimators. 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 addition, the normal distribution maximizes entropy, a measure of uncertainty, given that the mean is zero and the precision is τ (cf. Bernardo and Smith 1994, pp. 207–209). In other words, given the mean and the variance, the normal density imposes less structure on the $\{\alpha_i\}$ than other choices for the distribution.

We illustrate the approach by considering a single parameter β_i. Panel A of Figure 1 illustrates a posterior distribution $\pi(\beta_i \mid W)$; within $\pi(\beta_i \mid W)$ there are regions, perhaps disjoint, that satisfy some exogenous constraints as indicated by the shaded regions. Given a draw β_i^r from this posterior distribution, we generate adjustment factors $\{\alpha_i\}$ as shown in Panel B from a normal distribution $N(0, \tau^{-1})$ with density $\varphi(\alpha \mid \tau)$, where the precision parameter τ is one over the variance. 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. 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 $\pi(\beta_i \mid W)$, are not particularly meaningful.

Next we develop the loss function that corresponds to this procedure. Let $A = \{\alpha_i\}$ be the set of additive factors; let $B = \{\beta_i\}$ be the set of conjoint parameters; and define $A + B = \{\alpha_i + \beta_i\}$. To obtain point estimates of market share forecasts or for any function $h(A+B)$, we use the loss function:

$$L(h, d) = [h(A + B) - d]'[h(A + B) - d]$$
$$\text{for } A + B \in C \quad (5)$$

where C, the market share constraints at the base case X_0, limits the set of allowable decisions, $d \in D$, to those that have face validity for the manager. The posterior expected loss, which includes these adjustment factors, becomes

$$\rho(\pi, \varphi, d \mid \tau) = \int_C [h(A + B) - d]'[h(A + B) - d]$$
$$\cdot \varphi(A \mid \tau)\pi(B \mid W)\, dA\, dB, \quad \text{where} \quad (6)$$

$$\varphi(A \mid \tau) = \prod_{i=1}^{N} \varphi(\alpha_i \mid \tau) \text{ and}$$

$$\pi(B \mid W) = \pi(\beta_1, \ldots, \beta_N \mid W).$$

Figure 1 **Mapping into Constrained Area of the Posterior Distribution**
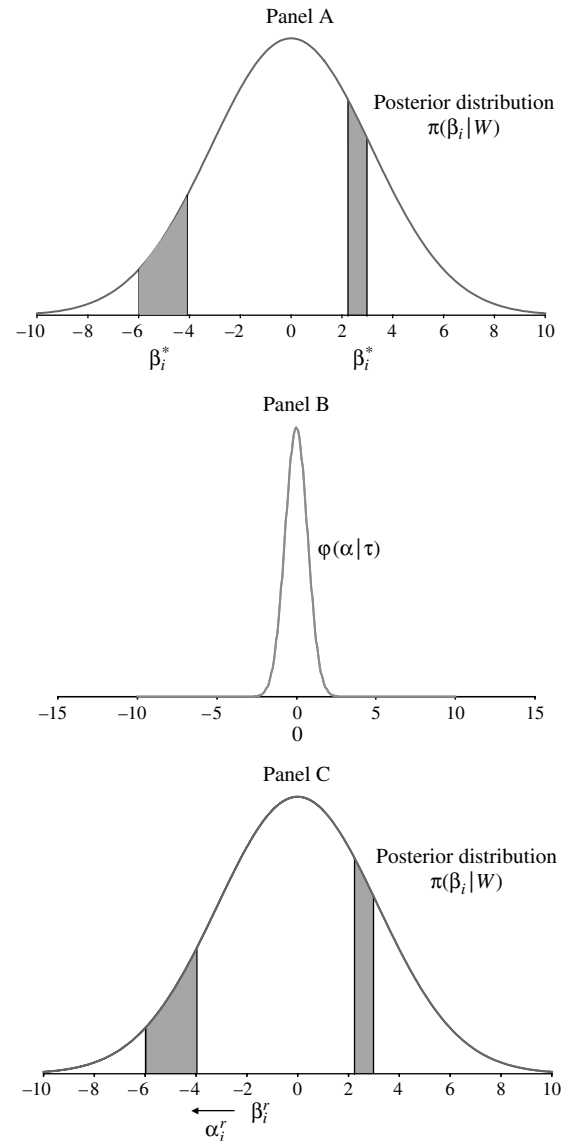
Notes. Panel A: Areas β_i^* such that external constraint is satisfied. Panel B: Normal density for adjustment factors. Panel C: For any random draw β_i^r from $\pi(\beta_i \mid W)$, map into area satisfying the constraint via $(\beta_i^r + \alpha_i^r)$ with α_i^r drawn from $\varphi(\alpha \mid \tau)$.

The requirement $A + B \in C$ is subsumed into the area of integration. To obtain the Bayes rule d^*, we differentiate (6) with respect to d and set it equal to 0:

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

where

$$f(A, B \mid W, C, \tau) = \frac{\varphi(A \mid \tau)\pi(B \mid W)}{P(C \mid W, \tau)} \quad \text{for } A + B \in C, \quad (7)$$

$$P(C \mid W, \tau) = \int_C \varphi(A \mid \tau)\pi(B \mid W)\, dA\, dB.$$

The support for the density $f(A, B \mid W, C, \tau)$ is C, and the normalizing constant is $P(C \mid W, \tau)$. Thus Equation (7) 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. The values of $\alpha_i + \beta_i$ drawn from $f(A, B \mid W, C, \tau)$ in (7) are dependent on all other αs and βs. 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.

The Bayes rule is conditional on meeting the market share constraints and the value of τ; τ determines "how close" the adjustment factors are to 0 and thus how much the CBC parameters are adjusted. If τ is too large, then values of α_i from $N_p(0, \tau^{-1} I_p)$ may be too close to 0 to satisfy $A + B \in C$, the market share-constraint. If τ is too small, then α_i will have a large variance and although we may satisfy the constraint, $A + B$ will be poorly identified and we may inadvertently alter the individual βs more than is necessary or desired.

Formally, we select the value of τ by minimizing the Bayes risk. In this case, the Bayes risk can be minimized by substituting the Bayes rule d^* into posterior expected loss and minimizing it with respect to τ. This is represented by

$$
\begin{aligned}
r(\tau) &= \int_C [h(A + B) - d^*]'[h(A + B) - d^*] \\
&\quad \cdot \varphi(A \mid \tau)\pi(B \mid W)\, dA\, dB \\
&= P(C \mid W, \tau) \int_C [h(A + B) - d^*]'[h(A + B) - d^*] \\
&\quad \cdot f(A, B \mid W, C, \tau)\, dA\, dB \\
&= P(C \mid W, \tau) \sum_{m=1}^{M} \mathrm{var}[h_m(A + B) \mid W, C, \tau], \quad (8)
\end{aligned}
$$

where M is the number of market shares we are estimating, and the variance is with respect to the density f. When the market share constraints are meaningful, e.g. they cannot be satisfied by the unadjusted parameters alone, increasing τ decreases the probability of meeting the constraint, $P(C \mid W, \tau)$, because A will be closer to 0; this will reduce $r(\tau)$. The $\mathrm{var}[h_m(A + B) \mid W, C, \tau]$ will also be reduced by large values of τ because $\alpha_i \sim N_p(0, \tau^{-1} I_p)$ and large τ implies less variance. This suggests that we want τ to be as large as possible as long as the market share constraints are satisfied. We show numerical results with simulated data that reinforce this intuition. First, we present several alternative methods of incorporating market share constraints into CBC.

3. Informative Priors and Non-Bayesian Approaches

3.1. Informative Priors
An alternative to the loss function approach is to treat the base-case scenario as informative prior information about the CBC parameters. This section compares and contrasts the loss function and prior information approaches. The crux of the problem with using market share information in the prior distribution for the CBC parameters is getting the "correct" weight between the prior and CBC data. If too much weight is placed on the CBC data, the simulated market shares at the base case will not match the actual market shares. If too much weight is placed on the market share information, then the prior overwhelms the CBC data. One could fine-tune the weight, but that would violate setting the prior *before* analyzing the data.

A prior formulation that leads to the Bayes rule d^* in Equation (7) is to treat both A and B as parameters where $\varphi(A \mid \tau)$ is the prior for A with the restriction that A is restricted to C given B. The joint posterior is

$$
\pi(B, A \mid W) \propto l(B \mid W)\pi(B)\varphi(A \mid \tau) \quad \text{for } A \in C \mid B. \quad (9)
$$

Note that we use the conditioning argument $C \mid B$ so that values of B are drawn without reference to A or the market share constraint, and draws of A depend on B. An alternative and perhaps more traditional way to think about the market share constraint as an informative prior is to represent (9) as

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

Then draws of B are dependent on A and the market share constraints. This model confounds the CBC data with the analyst's desire to satisfy the inferential goals of the manager. Marginal posterior distributions of B from (9) will not match those from (10) and will not match the analysis of the CBC data without market share constraints.

Market share constraints can also be incorporated through informative priors without using the additive factor 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:

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

Let $\psi(\xi \mid \tau) \sim N_{K-1}(z, \tau^{-1} I_{K-1})$. The prior parameter τ determines the trade-off 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, which, as previously noted, violates the rules of Bayesian inference.

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 and not data nor conjoint issues.

3.2. A Non-Bayesian Approach

Evgeniou et al. (2007) proposed a mathematical programming approach to individual-level estimation in CBC.[1] They use the following convex optimization program to obtain individual-level coefficients:

Problem 3.2.1

$$\gamma^* = \arg\min_{\gamma} \quad Cross\text{-}Validation(\gamma) \text{ and}$$

$$(\{w_i^*\}, w_0^*, D^*)$$

$$= \arg\min_{\{w_i\}, w_0, D} \left\{ -\frac{1}{\gamma^*} \sum_{i=1}^{N} \sum_{j=1}^{J_i} \log \frac{e^{x_{ijm^*} w_i}}{\sum_{m=1}^{M} e^{x_{ijm} w_i}} \right.$$

$$\left. + \sum_{i=1}^{N} (w_i - w_0)' D^{-1}(w_i - w_0) \right\} \qquad (12)$$

subject to D is a positive semi-definite matrix scaled to have trace 1,

where w_i and w_0 are vectors of length p, D is a $p \times p$ matrix, m^* indicates the chosen alternative, and all other notation follows earlier definitions. Multiplying Equation (12) by $-\gamma^*$ gives the following maximization problem:

$$\sum_{i=1}^{N} \sum_{j=1}^{J_i} \log \frac{e^{x_{ijm^*} w_i}}{\sum_{m=1}^{M} e^{x_{ijm} w_i}} - \gamma^* \sum_{i=1}^{N} (w_i - w_0)' D^{-1}(w_i - w_0).$$

This can be viewed as penalized maximum likelihood estimation (PMLE), where the first term is the standard log-likelihood function for logistic regression in CBC, and the second term is a penalty function.

PMLE is a regularization technique for ill-posed problems, such as the estimation in a large dimensional parameter space. Kimeldorf and Wahba (1970) and Good and Gaskins (1971) introduced roughness penalties to control complexity in nonparametric regression and density estimation, respectively. PMLE balances fit to the data, measured by the log-likelihood function, and complexity, expressed by a penalty term. A critical feature of PMLE is judging the relative weight, expressed here by γ, between fit and complexity. Predictive cross-validation (CV) is commonly used to select γ by minimizing holdout prediction error. Stone (1974) and Geisser (1975) introduced

CV in the statistical literature. Craven and Wahba (1979), Good and Gaskins (1980), Golub et al. (1979), and Wahba (1983) used CV in PMLE. Evgeniou et al. (2000) provide an excellent review of these procedures and their applications to vector support machines.

Subject-level parameter estimation in CBC is an ill-posed problem: the MLE does not exist or is instable when the number of individual-level observations is small relative to the number of individual-level parameters, which frequently occurs in practice. Hierarchical Bayes methods shrink individual-level estimates to an aggregate estimate where the amount of shrinkage depends on the relative precisions and degree of heterogeneity in the data (cf. Lenk et al. 1996). Bayesian models provide an alternative regularization method in both nonparametric density estimation (cf. Lenk 1988, 2003) and regression (cf. Lenk 1999). In essence, Bayesian models incorporate the penalty term into the model formulation and replace the CV step by using prior distributions for γ and updating them.

A novel aspect of the Evgeniou et al. (2007) application to conjoint analysis is using PMLE and CV to regularize the estimation of individual-level parameters. (Also, see Anderson and Blair 1982.) Evgeniou et al. (2007) refer to Equation (12) as a "loss function," where the first term reflects fit and the second term reflects shrinkage to the mean (or complexity control). For sake of clarity, we will refer to Equation (12) as an "objective function" because we will reserve "loss function" for the Bayesian loss function. The trade-off between fit and complexity is governed by the value of γ^* that is determined by CV.

As noted by many of the previously cited papers, PMLE has a Bayesian interpretation. Setting aside the cross-validation term and the restrictions on D, the convex optimization problem represents the solution to an analogous Bayesian problem. The Bayesian problem specifies a multinomial logit likelihood, $w_i \sim N(w_0, D)$ for heterogeneity, and improper priors for w_0 and D such that the normalizing constants for the distribution of heterogeneity cancel out. Finally, the 0/1 Bayesian loss function results in point estimates that are the maximum of the posterior distribution; this would give identical results to minimizing a corresponding objective function (analogous to Problem 3.2.1). Hence, the underling models for HB and PMLE are nearly the same, and they should produce similar results, which we confirm with an empirical example in §5.3.

One of the computational advantages of the approached outlined in Problem 3.2.1 is that conditional on γ, w_0, and D, the problem can be approached by solving N separate optimization problems, that is, one optimization problem for each respondent. Evgeniou et al. (2007) present an iterative procedure

---

[1] We thank the associate editor for drawing our attention to this approach.

for obtaining optimal values of all the variables in the objective function. However, market share constraints pose constraints across respondents that necessitate different solution procedures.

The structure of Problem 3.2.1 and the loss function approach of conditioning on the posterior distribution of the β_hs suggest an alternative, two-step problem. Given the set $\{w_{i*}\}$ from solving Problem 3.2.1, solve:

Problem 3.2.2

$$\min_{\{a_i\}} \quad \frac{1}{\rho}\sum_{k=1}^{K}(S_k - \hat{S}_k)^2 + \sum_{i=1}^{N}(a_i' a_i)$$

$$\text{subject to } \sum_{k=1}^{K}|S_k - \hat{S}_k| \le \delta^t,$$

where a_i are the p vectors of adjustment factors, S_k is the managerially supplied market share at the base case for brand k, and $\hat{S}_k$ is the forecasted market share using the adjusted coefficients $\{w_i + a_i\}$. Here, ρ controls the trade-off between meeting the market share constraint and ensuring that the adjustment factors a_i are close to 0. By adjusting ρ so that $\sum_{k=1}^{K}|S_k - \hat{S}_k| \le \delta^t$ is minimally satisfied, this puts as much emphasis as possible on keeping the adjustment factors close to 0. We use squared error instead of absolute error of the market shares in the objective function because our optimization algorithm uses gradients, and absolute error is not differentiable everywhere.

4. MCMC 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 $E[h(A + B) \mid W, C, \tau]$. 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 MCMC methods are used to obtain draws from $f(A, B \mid W, C, \tau)$ from Equation (7) with the challenge that the market share constraints create dependencies among subject-level parameters.

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:

Step 1. For $i = 1, \ldots, N$;
(a) Draw $\beta_i \mid W, X, \theta, \Lambda$: Use standard methods for either probit or MNL models;
(b) Draw $\alpha_i \mid \beta_i, \{\beta_{-i}\}, \{\alpha_{-i}\}, X_o, \tau, \delta^t$: Detailed in the appendix;

Step 2. $\theta \mid \{\beta_i\}, \Lambda$: Standard conjugate set-up;
Step 3. $\Lambda \mid \{\beta_i\}, \theta$: 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 1(b), the notations $\{\beta_{-i}\}$ and $\{\alpha_{-i}\}$ indicate the sets 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 in the appendix, from $\varphi(\alpha_i \mid \tau)$ given α_i^g is in $C \mid \beta_i^g$.

A random walk Metropolis-Hastings (Chib and Greenberg 1995) or a weighted bootstrap (Smith and Gelfand 1992) is used to draw α_i for each i on each iteration of the algorithm. The chain converges to the stationary distribution implied by the posterior distribution $f(A, B \mid W, C, \tau)$ 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 $E[h(A + B) \mid W, C, \tau]$ are obtained by computing $h(A + B)$ for each draw of $A + B$ and averaging over the draws.

There are several other practical implementation issues that are detailed with suggestions in the Technical Appendix, found at http://mktsci.pubs.informs.org. 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 posterior expected loss for different values of τ.

5. Simulation Studies

5.1. In-Sample Results
It is not surprising that one can use additive factors to adjust the CBC parameter estimates to recover base-case market shares. The purpose of this simulation study is to quantify the distortion of the CBC preference structure by using the proposed method. The simulation shows that relatively small additive factors are needed to meet the constraints in an MNL model. Additional results with a correlated probit model are available from the first author. The simulated data set 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 the market share using the true values of the parameters. The model is identified by dropping one of the brand intercepts (Brand D). 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.

Table 1 presents 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

Table 1   Simulation Results for MNL Bayesian Approach

Brand	MS at X_o using $\{\beta_i\}$	MS constraint at X_o	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

Brand	Posterior MS at X_o	Posterior θ		Posterior average $(\alpha + \beta)$	
		$\delta^t = 0.10$			
A	0.417 (0.008)	(1)	−0.386 (0.09)	(1)	−0.615 (0.07)
B	0.348 (0.006)	(2)	0.466 (0.09)	(2)	0.672 (0.07)
C	0.107 (0.005)	(3)	0.309 (0.09)	(3)	0.432 (0.07)
D	0.128 (0.007)	(4)	−1.010 (0.08)	(4)	−0.793 (0.06)
		(5)	0.782 (0.07)	(5)	0.789 (0.04)
		(6)	1.520 (0.08)	(6)	1.166 (0.05)
$\tau^{-1} = 0.00001$				MAD $= 0.189$	
		$\delta^t = 0.05$			
A	0.404 (0.005)	(1)	−0.420 (0.09)	(1)	−0.638 (0.07)
B	0.349 (0.003)	(2)	0.449 (0.08)	(2)	0.649 (0.06)
C	0.129 (0.003)	(3)	0.291 (0.09)	(3)	0.479 (0.07)
D	0.118 (0.005)	(4)	−1.022 (0.08)	(4)	−0.767 (0.05)
		(5)	0.788 (0.07)	(5)	0.792 (0.03)
		(6)	1.524 (0.09)	6)	1.108 (0.06)
$\tau^{-1} = 0.00001$				MAD $= 0.213$	
		$\delta^t = 0.02$			
A	0.402 (0.003)	(1)	−0.413 (0.09)	(1)	−0.630 (0.07)
B	0.349 (0.002)	(2)	0.453 (0.09)	(2)	0.642 (0.06)
C	0.143 (0.002)	(3)	0.296 (0.09)	(3)	0.526 (0.07)
D	0.106 (0.003)	(4)	−1.025 (0.09)	(4)	−0.757 (0.06)
		(5)	0.785 (0.07)	(5)	0.806 (0.03)
		(6)	1.518 (0.09)	(6)	1.062 (0.05)
$\tau^{-1} = 0.00001$				MAD $= 0.230$	

Notes. MS, market share. X_o, "Base-case" design matrix. Posterior mean and (standard deviation) displayed. MAD, mean absolute difference between posterior θ and posterior average $(\alpha + \beta)$.

to increase τ, the precision of the additive factors A, to the maximum value of 100,000 (variance $= \tau^{-1} = 0.00001$) while meeting the market share constraints. The posterior mean of θ is from the CBC data and is close to the true θ. If one were to average the posterior means of β_i (not reported in the table), they would be close to the posterior mean of θ. Selected individual-level histograms of $\beta_i + \alpha_i$ were inspected and it does not appear that $f(A, B \mid W, C, \tau)$ is disjoint for these data sets. We therefore present the empirical average of $\overline{(\beta + \alpha)}$ averaged across individuals and draws from the MCMC chain. These are provided as a point of comparison to the posterior mean of θ 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. For example, the market share constraint 0.15 for brand C at X_o is greater than the market share 0.05 using only the CBC data. Consequently, the average of $\overline{(\beta + \alpha)}$ for brand C is greater than the posterior mean of θ for brand C, thus increasing the preference for brand C to satisfy the market share constraints. Figure 2 plots the individual-level average $\overline{(\beta_i + \alpha_i)}$ compared to the individual-level average $\bar{\beta}_i$ for selected parameters when $\delta^t = 0.02$. The top panel is the partworth for Brand C. The loss function approach tends to increase the preference for subjects with moderate partworths, while leaving subjects with extreme partworths untouched. The plots show that individual-level parameters are adjusted differently, but that adjustments are generally small and directionally consistent.

Taken together, these results 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.

There is a trade-off between the target level of discrepancy δ^t and the size of the adjustment factors α_i. As δ^t gets smaller (e.g., recovered market shares must be closer to the managerial base case), the adjustment factors tend to be larger. This can be seen in Table 1 by comparing the columns "Posterior θ" to "Posterior average $(\alpha + \beta)$" across the different levels of δ^t, or more directly by looking at the value of the mean absolute difference (MAD) between the two columns. At $\delta^t = 0.10$ the MAD is 0.189 as compared to 0.230 at $\delta^t = 0.02$.

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

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

$\overline{\beta}_{i6}$ vs. $\overline{\beta_{i6} + \alpha_{i6}}$
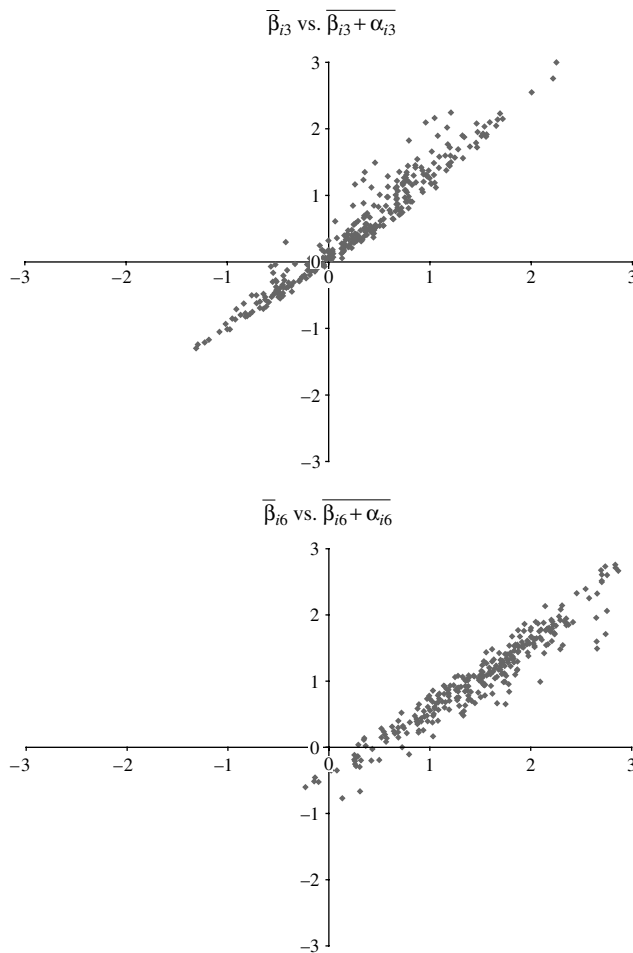
The posterior expected loss was investigated for different values of τ with $\delta^t = 0.02$. Recall that the value of τ that minimizes the posterior expected loss also minimizes the Bayes risk. Using Equation (7), the natural log of the posterior expected loss was calculated for the function $h(B + A)$ 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 function along a grid of values for τ, presented as τ^{-1} in Table 2. Full details on the importance sampling algorithm are available from the authors. The posterior expected loss is dominated by $\ln[P(C \mid W, \tau)]$ and because the market share constraints are binding, the smallest value of τ^{-1} minimizes the function. The lower half of Table 2 shows additional quantities calculated using the importance sampler. As expected, the value of $E[\sum_{i=1}^{N} \alpha_i' \alpha_i \mid X, C, \tau]$, which measures the dispersion of the adjustment factors from 0, reaches its minimum value at the minimum value of τ^{-1}, but the improvement is marginal beyond $\tau^{-1} = 0.005$. The values of d^* equal to $E[h(A + B) \mid W, C, \tau]$ were the

Table 2   Importance Sampling Results for MNL Model: Simulated Data

τ^{-1}	$\ln[r(\tau)]$	$\ln[P(C \mid \tau)]$	$\ln[\sum_{m=1}^{M} \text{var}[h_m(A+B) \mid X, C, \tau]]$
	Estimates of log posterior expected loss and its components		
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

τ^{-1}	d^* Brand A	d^* Brand B	d^* Brand C	d^* Brand D	$E[\sum_{i=1}^{N} \alpha_i' \alpha_i \mid X, C, \tau]$
	Estimates of Bayes rule for various quantities				
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

same up to the third decimal place for all the different values of τ^{-1}.

Because the posterior expected loss and hence the Bayes risk involves $\text{var}[h_m(A + B)]$, then the minimizing 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 value of τ that minimizes the posterior expected loss for each different form of $h(A + B)$ and different values of attributes. However, the above analysis suggests a more pragmatic approach of conditioning on the value of τ. When adjustment factors are needed, set τ as large as possible and use draws of $h(A + B)$ from $f(A, B \mid W, C, \tau)$ to estimate $E[h(A + B) \mid W, C, \tau]$. Although it is possible to investigate the value of τ that minimizes the posterior expected loss, the practical benefits of doing so are unclear.

5.2.   Out-of-Sample Results
Conjoint is frequently used to project market shares when attributes are different than existing product offerings. Here we want to know if imposing market share constraints results in "poor" predictions as compared to the unconstrained analysis. This section investigates the effect of the market share constraints on market share simulators when the attributes are "near" and "far" from the base-case scenario. The results show that the additive factors do not greatly impact market shares for scenarios other than the base case.

The conjoint design was adapted from an actual study used in a commercial market research application. Details of the simulations are provided in Table 3 and a full report is available from the first author. The simulation involved 300 "respondents" and 10 brands described by three binary attributes, three continuous attributes, and nine brand-specific dummies. The

Table 3   Out-of-Sample Predictions: Simulation Details

Conjoint data
- 10 brands, 3 binary attributes, 3 continuous attributes (one representing price)
- 10 choice sets per respondent
- Each experimental choice set contained 6 brands
- 300 "respondents"

Distortions
- "Additive"—for each respondent, randomly select up to 5 coefficients, add random term from $U(1, 3)$ distribution
- "Scale"—for each respondent, divide all coefficients by random term from a $U(0.50, 0.75)$ distribution
- "Missing Attributes"—include one binary and one continuous attribute in "base-Case" and "out-of-sample" choice designs that are not included in the conjoint data

Base case
- All 10 brands included in base case
- One binary attribute set $= 0$ for all brands; represents "new to market" feature
- All other attribute levels set at random

Out-of-sample prediction tasks

"Near" the base case
- 30 new choice sets
- Each choice set includes all 10 brands
- Randomly add "new" feature to some brands and randomly adjust price $+/-50\%$
- All other attributes the same as base case (including "missing attributes")

"Far" from the base case
- 30 new choice sets
- Each choice set includes all 10 brands
- All attribute levels randomly set for all brands and attributes (including "missing attributes")

Simulations
- Conjoint choice data generated using the "distorted" coefficients
- Market share constraints and prediction results measured for: All 10 brands; 5 "high" market share brands; 5 "low" market share brands; 2 arbitrary brands
- Prediction accuracy measured using δ^p where the "known" market share is estimated using the "undistorted" coefficients and "missing attributes"

---

choice data consisted of 10 choice sets per respondents, where 6 of the 10 brands were included in each choice set. We introduced several distortions between the conjoint data and the observed market shares. These represent possible reasons why forecasted and base-case or out-of-sample predictions do not match. The "additive" distortion assumes that consumers are much more sensitive to some attributes in the conjoint exercise than they are when making actual choices. The "scale" distortion assumes that there is more variance in actual market-place choices than revealed in the conjoint data; that is, our consumers are more consistent when making hypothetical choices. Finally, we assumed that there were attributes present in the market that were not included in the conjoint design. This condition is labeled "missing attributes."

Design matrices were generated for a "base case" and for a relatively large number of out-of-sample choice sets. The out-of-sample choice sets represent the possible "what-if" scenarios of interest to managers. The base case consisted of all ten brands. Attribute levels were chosen randomly[2] except for one binary attribute that was set to 0 for all alternatives; this binary attribute represents a "new to the market" attribute. Thirty out-of-sample choice sets were

generated to minimize the chance that the predictive results were being driven by one or two well-chosen (or poorly chosen) choice sets. Two separate predictive exercises were considered. "Near" the base case was intended to mimic marginal changes to the current product configuration: add the new attribute to randomly selected alternatives and change price $+/-50\%$. In the "far" condition all attributes were randomly generated and is intended to encompass a wider array of new product configurations.

Brand managers may not have market share information on all brands and may be interested in out-of-sample predictions for only their own products and/or a subset of competitors. To investigate this situation, market share constraints were applied to different subsets of brands: all 10 brands, the 5 "high" market share brands, the 5 "low" market share brands, and 2 arbitrary brands. Separate results were estimated for each of the different market share constraints. The target level of δ for the base case was set to 0.02; the lowest level of τ^{-1} was set at 0.00001. The targeted level of δ and the lowest value of τ^{-1} were reached for all the different types of distortions and different market share constraints.

Out-of-sample predictions were calculated by integrating over a sample of 1,000 observations from the posterior distribution for each respondent, with constraints and without, for each of the 30 choice

---

[2] Additional simulations showed that the results were not sensitive to the level of the attributes in the base case.

Table 4 Out-of-Sample Prediction Results; Percentage Improvement in Absolute Error of the Loss Function Adjustments Compared to Standard Market Shares from Conjoint Simulators

	Type of distortion	
	"Scale" "Additive" (%)	"Missing attributes" "Scale" "Additive" (%)
"Near" the base case		
All 10 brands	**52**	**69**
5 "high" mkt share brands	**49**	**76**
5 "low" mkt share brands	**47**	**40**
2 brands	**49**	**20**
"Far" from the base case		
All 10 brands	**23**	7
5 "high" mkt share brands	9	6
5 "low" mkt share brands	19	−3
2 brands	3	−2

Notes. Numbers in bold indicate that there is less than 5% posterior probability that the two methods produce identical estimates; i.e., bold numbers indicate a "significant" difference between the adjusted and unadjusted estimates. Out-of-sample predictions based on 30 choice sets with 10 brands in each choice set.

sets. The discrepancy between the predicted market share and the actual market share was measured using δ^p from Equation (1), again for the constrained and unconstrained model. δ^p was calculated for each choice set and then averaged across the 30 out-of-sample choice sets included in each simulation.

In the simulations, the "additive" and "scale" distortions are combined and presented in one set of results, and these are combined with the "missing attributes" distortion in a separate analysis. The results are contained in Table 4. The chart shows the relative performance of the loss function approach compared to the standard method. Positive numbers represent the percentage improvement in δ^p, the absolute error between the actual and predicted market shares, for the loss function approach, whereas negative numbers represent a decrease in predictive accuracy. Numbers in bold indicate that the 95% highest posterior density of the mean difference in δ between the two measures does not include 0; i.e., the results are "statistically significant."

In the area around the base case, the loss function approach produces significantly more accurate predictions than the standard approach. However, the changes induced by the loss function do not result in significantly worse predictions even over a broad range of out-of-sample data.[3] Thus, the more accurate *localized* predictions do not come at the expense of less accurate predictions in the range of the attribute

levels of interest to brand managers. In the "far" condition, the adjusted and unadjusted coefficients produced very similar market share forecasts. The only substantive difference in the predictive accuracy in the "far" condition was for the "scale" and "additive" distortions for "all 10 brands": the adjusted coefficients improved predictive accuracy by 23% over the unadjusted coefficients. However, for this condition, the mean absolute difference in market share forecasts between the two methods was only 1.35 percentage points across all brands in all choice sets.

5.3. PMLE Results

In this section we investigate the mathematical programming approaches outlined in §3.2. Because published studies using mathematical programming techniques have not incorporated market share constraints, we restrict ourselves to an illustrative example. We anticipate future research will result in better and more robust mathematical programming techniques. Nonetheless, our results suggest that the mathematical programming approach is feasible.

For this illustration we return to the simulated data set used in §5.1. These data consist of 300 respondents, 12 choice sets per respondents, with four alternatives per choice set. Covariates consisted of binary brand indicators (three), two continuous covariates, and a binary covariate. We used the procedure detailed by Evgeniou et al. (2007) to solve Problem 3.2.1 and obtain individual-level coefficients $\{w_i^*\}$.[4] Table 5 contains results. The average values of $\{w_i^*\}$ are somewhat smaller but are comparable to the actual θ and the posterior estimates of θ from the Bayesian approach in Table 1. We note that using $\{w_i^*\}$ we obtained a market share forecast at X_o very close to the actual result using the true $\{\beta_i\}$.

Market share constraints were incorporated via solving Problem 3.2.2 with the squared error objective function. We used a standard numeric optimization routine.[5] To obtain $\{w_i^*\}$ we (repeatedly) solved 300 problems with six coefficients; by comparison, to obtain $\{w_i^* + a^*\}$ we (repeatedly) solved one problem with $300 \times 6 = 1,800$ coefficients. The algorithm starts with a relatively large ρ, obtains a solution, decreases ρ, and uses the values from the previous solution as a starting guess to obtain a new solution. ρ is decreased until the market share constraint is minimally satisfied.

We find that using mathematical programming with the squared error objective function is a viable

[3] We used "accurate" market share constraints. An experiment with inaccurate market share constraints resulted in less accurate predictions near the base case, but no difference in the broader range of the out-of-sample data. This is consistent with the results in Table 4.

[4] To solve the optimization problem for each respondent, we used a standard IMSL routine that uses a quasi-Newton method with finite-difference gradient. The same routine was used in the cross-validation and final steps.

[5] Again, we used an IMSL routine that uses a quasi-Newton method with finite-difference gradient.

Table 5 Simulation Results for MNL: Mathematical Programming Approach

Brand	MS at X_o using $\{\beta_i\}$	MS at X_o using $\{w_i^*\}$	MS constraint at X_o	Actual θ	
A	0.54	0.55	0.40	(1)	−0.43
B	0.25	0.24	0.35	(2)	0.50
C	0.05	0.05	0.15	(3)	0.34
D	0.15	0.16	0.10	(4)	−1.00
				(5)	0.82
				(6)	1.47

γ^* from cross-validation $= 0.128$

Brand	MS at X_o	Average w_i^*		Average $(w_i^* + a_i^*)$	
		$\delta^t = 0.10$			
A	0.420	(1)	−0.371	(1)	−0.539
B	0.340	(2)	0.366	(2)	0.518
C	0.110	(3)	0.234	(3)	0.372
D	0.130	(4)	−0.852	(4)	−0.678
		(5)	0.649	(5)	0.669
		(6)	1.282	(6)	0.990
$\rho = 0.00009$				MAD $= 0.157$	
		$\delta^t = 0.05$			
A	0.407	(1)	−0.371	(1)	−0.544
B	0.346	(2)	0.366	(2)	0.524
C	0.133	(3)	0.234	(3)	0.425
D	0.114	(4)	−0.852	(4)	−0.648
		(5)	0.649	(5)	0.687
		(6)	1.282	(6)	0.933
$\rho = 0.00003$				MAD $= 0.186$	
		$\delta^t = 0.02$			
A	0.403	(1)	−0.371	(1)	−0.544
B	0.348	(2)	0.366	(2)	0.526
C	0.142	(3)	0.234	(3)	0.445
D	0.107	(4)	−0.852	(4)	−0.638
		(5)	0.649	(5)	0.698
		(6)	1.282	(6)	0.911
$\rho = 0.000013$				MAD $= 0.196$	

Notes. MS, market share. MAD, mean absolute difference between adjusted and unadjusted coefficients. X_o, "Base-case" design matrix.

approach. Table 5 shows that we were able to obtain solutions for $\delta^t = 0.10$, $\delta^t = 0.05$, and $\delta^t = 0.02$. Consistent with the Bayesian results in Table 1, the mean absolute difference between the adjusted and unadjusted coefficients increases as δ^t decreases. Strictly speaking, the MAD and adjusted coefficients are not comparable between the Bayesian and the mathematical programming approach because the adjustments are made to different bases: point estimates of $\{w_i^*\}$ versus the posterior distribution $\pi(B \mid W)$. However, we note that the average adjusted coefficients $\{w_i^* + a^*\}$ are comparable to the posterior average of $\{\beta_i + \alpha_i\}$ from the Bayes solution but that the MAD is smaller in the mathematical programming approach. This may be a result of directly minimizing $\sum a_i' a_i$ in the objective function 3.2.2; the loss function penalizes

large α_i indirectly through the posterior expected loss (8) and the parameter τ.

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 data set, 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.

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 one 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 noninformative priors were 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 straightforward 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 $\pi(B \mid W)$ and $f(A, B \mid W, C, \tau)$ 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 prespecified minimum of 0.00001

Table 6   Selected Summary Statistics: Consumer Electronics Study

Attribute	Posterior θ	$\sqrt{\lambda_{pp}}$	Posterior average $(\alpha + \beta)$	$\sigma_{(\alpha+\beta)pp}$	Attribute included in base case?
1	−1.111 (0.20)	3.231	−1.032 (0.12)	3.341	Yes
2	0.776 (0.08)	1.176	0.972 (0.07)	1.454	Yes
3	1.008 (0.08)	1.216	1.176 (0.07)	1.504	Yes
4	0.700 (0.07)	1.009	0.915 (0.08)	1.429	Yes
5	0.266 (0.07)	0.784	0.654 (0.07)	1.367	Yes
6	0.370 (0.06)	0.838	0.330 (0.07)	1.222	No
7	0.489 (0.06)	0.773	0.403 (0.08)	1.274	No
8	0.440 (0.06)	0.866	0.214 (0.07)	1.318	No
9	0.432 (0.07)	0.980	0.646 (0.06)	1.336	Yes
10	0.426 (0.06)	0.708	0.269 (0.06)	1.295	No
11	0.463 (0.06)	0.841	0.139 (0.06)	1.345	No
12	0.235 (0.06)	0.936	0.252 (0.06)	1.301	No
13	0.313 (0.06)	0.760	0.221 (0.06)	1.279	No
14	0.413 (0.06)	0.766	0.336 (0.07)	1.255	No
15	0.176 (0.06)	0.705	−0.071 (0.06)	1.268	No
16	0.352 (0.06)	0.701	0.197 (0.06)	1.283	No
17	0.204 (0.06)	0.774	0.449 (0.06)	1.301	Yes
18	−0.118 (0.07)	0.786	−0.344 (0.08)	1.469	No
19	0.043 (0.07)	0.849	−0.154 (0.09)	1.419	No
20	0.768 (0.07)	1.042	0.631 (0.07)	1.434	No
ln(price)	−0.872 (0.05)	0.691	−0.798 (0.04)	0.843	Yes

Note. Posterior mean and (standard deviation) displayed.

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 $f(A, B \mid W, C, \tau)$ is disjoint for this data set. Table 6 contains the summary statistics for the constrained and unconstrained parameters including the standard deviation for β_i from the distribution of heterogeneity $\sqrt{\lambda_{pp}}$, compared to the standard deviation $\sigma_{(\alpha+\beta)pp}$ from the empirical distribution of $(\alpha_i + \beta_i)$. 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 (see below for explanation). The constrained parameters exhibited somewhat larger measures of heterogeneity.

The top section of Table 7 shows the managerial base-case market shares and estimated market shares using $\{\beta_i\}$ and $\{\beta_i + \alpha_i\}$. All estimates are based on a sample of 1,000 draws from the individual-level posterior distribution of β_i or $\beta_i + \alpha_i$. Table 7 shows that there was a sizable difference between the estimate using the unconstrained parameters and the managerial base case. Using the unconstrained parameters, respondents were estimated to be much less likely to choose the base product. For market segment #1, the base case was 18.9% choosing the representative product versus a managerial expectation of 43.6%. To increase the attractiveness of the base product over the "none" option, the relative attractiveness of attributes included in the base-case product are increased, thus increasing their coefficients in

Table 6. The loss function approach was able to match the base-case choice share to within the prespecified level of accuracy despite the relatively large discrepancy between it and the forecasts using the unadjusted parameters.

The bottom panel of Table 7 shows the actual and estimated choice shares averaged across choice sets. Recall that in this study, respondents not only indicated which of the three options they preferred, but

Table 7   Choice Share Results: Consumer Electronics Study

	Base-case choice share using $\{\beta_i\}$	Managerial base-case choice share	Base-case choice share using $\{\beta_i + \alpha_i\}$
Market segment #1 (%)	18.9	*43.6*	43.6
Market segment #2 (%)	23.1	*55.6*	55.5
δ Compared to managerial base case	0.573		0.001

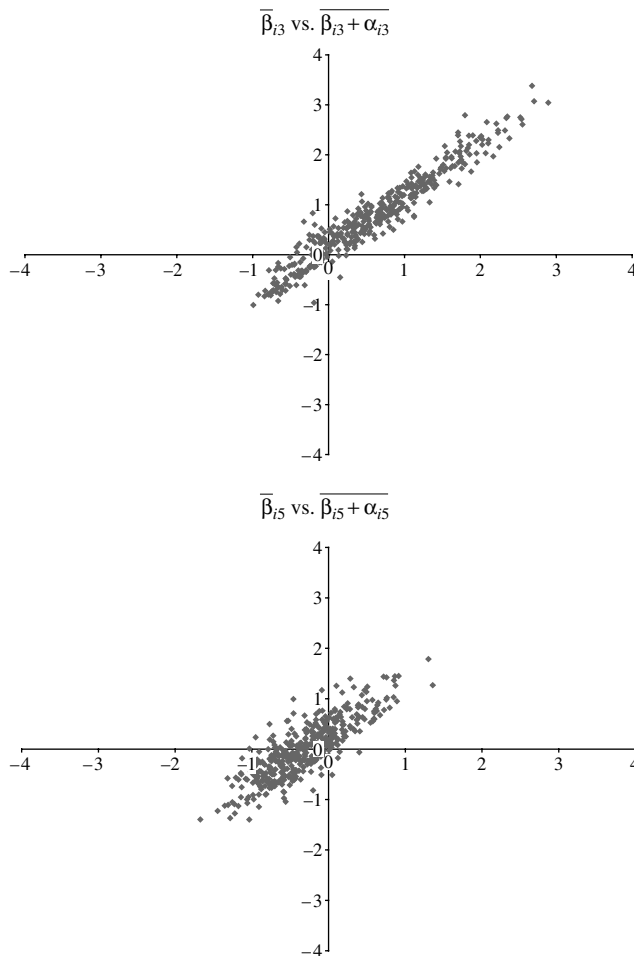|	Raw data	Estimated choice share using $\{\beta_i\}$	Estimated choice share using $\{\beta_i + \alpha_i\}$
Option 1 (%)	29.1	30.0	31.5
Option 2 (%)	33.1	31.8	31.7
Option 3 (%)	37.8	38.2	36.8
With the "none" option			
Option 1 (%)	11.7	12.3	16.5
Option 2 (%)	14.1	13.7	16.6
Option 3 (%)	17.2	16.9	19.9
None (%)	57.1	57.1	47.0

Note. Posterior mean displayed.

also whether or not they would actually purchase it, e.g., a dual response format. This allows us to estimate and compare choice shares with and without the "none" option. We can see compared to the raw data and estimates based on $\{\beta_{hi}\}$, when "none" is included in the choice set the estimates based on $\{\beta_{hi} + \alpha_{hi}\}$ decrease the relative frequency of "none" and increase the relative frequency of the other options. This is consistent with the market share constraint. However, when the choice set is restricted to the relative preference between the three options, the estimated choice shares are all reasonably close. This is another example of the localized nature of the loss function approach.

Figure 3 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 $\theta = 1.008$ and the posterior mean of $\overline{(\beta + \alpha)} = 1.176$ from Table 5 was about average. Coefficient #5 was selected because it had the largest difference, $\theta = 0.266$ and $\overline{(\beta + \alpha)} = 0.654$. Table 5 and Figure 3 show that relatively small changes in the β_is were sufficient to

Figure 3   **Consumer Electronics Study: Posterior Mean of Individual-Level Parameters**

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

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

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. Because 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.

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 estimates. 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, which are external to the CBC study. In addition to representing the preferences of the study participants, the client expects the analysis will be able to replicate "base-case" 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. In addition, we illustrate a similar procedure for non-Bayesian models. The Bayesian and non-Bayesian approaches provide similar results because the underlying models are similar: the main differences are the numerical procedures.

Formally, we introduce additive factors into the loss function that map draws from the posterior distribution into the set that satisfies the market share constraints at the base-case scenario. The set of additive factors 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 of the model parameters and additive factors via the posterior expected loss and provide an algorithm for simulating the additive factors to satisfy the market share constraints. Conceptually, the loss function approach approximates a procedure that integrates

over those regions of the posterior that are consistent with the market share constraints.

Simulated and real data sets are used to illustrate the proposed approach. These empirical studies demonstrate that the *average* representation of preferences changes relatively little using the loss function approach. The use of a normal distribution with mean 0 for the adjustment factors minimizes the adjustments at the individual level, while recovering the managerially supplied base-case market shares without distorting the preference structure from the CBC experiment. The algorithm tends to seek opportunities where a small adjustment in the estimated parameters will change choices for the base-case scenario. When predicting scenarios far from the base case, which is often the objective of conjoint studies, the adjusted and unadjusted market shares are similar.

This paper highlights the need for additional research in a number of areas. First, by adjusting the CBC results as little as possible we assume they accurately capture preferences under the new market conditions represented by the new products, attributes, attribute levels, etc., in the conjoint study. The base-case market share constraint serves to anchor any changes in preferences to a point considered to have face validity by management. If the discrepancy between revealed and stated preferences is purely a methodological artifact, then one would want to model this systematically. Second, and along these same lines, the loss function approach allows all of the coefficients to be adjusted. However, if one assumed that only brand awareness or distribution was responsible for the discrepancy, it is easy to set up the loss function such that only brand intercepts are adjusted; or only the price coefficient; or all coefficients are adjusted uniformly representing a change in scale; or that only certain respondent's coefficients are adjusted. Additional empirical and/or theoretical evidence may suggest that only a subset of coefficients should be adjusted via the loss function or provide insight on how to systematically model any adjustments.

In this research we assumed only one "base-case" observation was available consistent with our experience in applied settings. However, it is interesting to consider a situation in which multiple base-case observations may be available. Multiple base cases in the loss function should improve prediction. However, the more base cases that are added to the loss function, the greater the possible deviance from CBC preference structure, in which case a more elaborate adjustment scheme than the one described in this paper may be required. For example, it may be necessary to introduce multiple adjustment factors for

different base cases. We leave this topic for future research.

Finally, 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.

Appendix. Drawing α_i

The following steps detail how to draw $\alpha_i \mid \beta_i, \{\beta_{-i}\}, \{\alpha_{-i}\}, X_o, \tau, \delta^t$. Recall that $A + B$ must satisfy $A + B \in C$ or $\chi(S) = 1$, the market share constraint in the loss function. In this appendix we will use the indicator function $\chi(S) = 1$ to indicate when the constraint is satisfied and $\chi(S) = 0$ when it is not. For each individual on each iteration, β_i is drawn from $\pi(\beta_i \mid W)$ and that draw is used in (6) to draw α_i from $f(A, B \mid W, C, \tau)$. Let $\alpha_i^{(o)}$ represent the draw of α_i from the previous iteration.

Step 1. If $\delta^{(o)} \leq \delta^t$, then $\chi(S, \alpha_i^{(o)}) = 1$, else $\chi(S, \alpha_i^{(o)}) = 0$;
  (a) Calculate each $\hat{S}_k^{(o)}$ using $\alpha_i^{(o)}, \beta_i, \{\beta_{-i}\}, \{\alpha_{-i}\}$, and X_o;
  (b) Calculate $\sum_{k=1}^{K} |S_k - \hat{S}_k^{(o)}| = \delta^{(o)}$.

Step 2. If $\chi(S, \alpha_i^{(o)}) = 1$, 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 $N_p(0, z I_p)$ 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 $f(A, B \mid W, C, \tau)$ from Equation (6) implies the following acceptance probability for $\alpha_i^{(n)}$:

$$\min\left[\frac{\exp(-(\tau/2)\alpha_i^{(n)\prime}\alpha_i^{(n)})\chi(S, \alpha_i^{(n)})}{\exp(-(\tau/2)\alpha_i^{(o)\prime}\alpha_i^{(o)})}, 1\right].$$

$\chi(S, \alpha_i^{(n)})$ is calculated analogous to Step 1. 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.

Step 3. If $\chi(S, \alpha_i^{(o)}) = 0$, then a weighted bootstrap is used to draw a new value of α_i. A challenge in sampling α_i from $f(A, B \mid W, C, \tau)$ 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 let $\beta_i^{(n)}$ be the draw from the current iteration.

  (a) 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.

  (i) For $r = 1$, to R draw $\alpha_i^{(r)}$ from $g(\alpha)$ where $g(\alpha) \sim N(\mu, \tau^{-1} I_p)$. The target density implied by (6) is $f(\alpha) \propto \varphi(\alpha \mid \tau)\chi(S)$ and the proposal density for the weighted bootstrap is $g(\alpha) \sim N(\mu, \tau^{-1} I_p)$.

  (ii) For each draw r, calculate the weight w_r:

$$w_r = \frac{\exp(-(\tau/2)(\alpha_i^{(r)\prime}\alpha_i^{(r)}))\chi(S, \alpha_i^{(r)})}{\exp(-(\tau/2)(\alpha_i^{(r)} - \mu)\prime(\alpha_i^{(r)} - \mu))}.$$

(b) Select individual value $\alpha_i^{(r)}$ with probability:

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

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

References

Aitchison, J., S. D. Silvey. 1958. Maximum-likelihood estimation of parameters subject to restraints. *Ann. Math. Statist.* **29**(3) 813–828.

Allenby, G. M., N. Arora, J. L. Ginter. 1995. Incorporating prior knowledge into the analysis of conjoint studies. *J. Marketing Res.* **32**(May) 152–162.

Allenby, G., G. Fennell, J. Huber, T. Eagle, T. Gilbride, D. Horsky, J. Kim, P. Lenk, R. Johnson, E. Olfek, B. Orme, J. Walker. 2005. Adjusting choice models to better predict market behavior. *Marketing Lett.* **16**(3/4) 197–208.

Anderson, J. A., V. Blair. 1982. Penalized maximum likelihood estimation in logistic regression and discrimination. *Biometrika* **69**(1) 123–136.

Berger, J. O. 1985. *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer-Verlag, New York.

Bernardo, J. M., A. F. M. Smith. 1994. *Bayesian Theory*. John Wiley & Sons, New York.

Boatwright, P., R. McCulloch, P. Rossi. 1999. Account-level modeling for trade promotion: An application of a constrained parameter hierarchical model. *J. Amer. Statist. Assoc.* **94**(448) 1063–1073.

Casella, G., R. L. Berger. 1990. *Statistical Inference*. Duxbury Press, Belmont, CA.

Cattin, P., D. R. Wittink. 1982. Commercial use of conjoint analysis: A survey. *J. Marketing Res.* **46**(Summer) 44–53.

Chib, S., E. Greenberg. 1995. Understanding the metropolis-hastings algorithm. *Amer. Statist.* **49**(November) 327–335.

Craven, P., G. Wahba. 1979. Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik* **31** 377–403.

DeGroot, M. 1970. *Optimal Statistical Decision*. McGraw-Hill, New York.

Ding, M., R. Grewal, J. Liechty. 2005. Incentive-aligned conjoint analysis. *J. Marketing Res.* **42**(February) 67–82.

Dupacova, J., R. Wets. 1988. Asymptotic-behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *Ann. Statist.* **16**(4) 1517–1549.

Evgeniou, T., M. Pontil, T. Poggio. 2000. Regularization networks and support vector machines. *Adv. Comput. Math.* **13** 1–50.

Evgeniou, T., M. Pontil, O. Toubia. 2007. A convex optimization approach to modeling consumer heterogeneity in conjoint estimation. *Marketing Sci.* **26**(6) 805–818.

Geisser, S. 1975. Predictive sample reuse method with applications. *J. Amer. Statist. Assoc.* **70**(350) 320–328.

Ghosh, M. 1992. Constrained Bayes estimation with applications. *J. Amer. Statist. Assoc.* **87**(418) 533–540.

Golub, G. H., M. Heath, G. Wahba. 1979. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* **21**(2) 215–223.

Good, I. J., R. A. Gaskins. 1971. Nonparametric roughness penalties for probability densities. *Biometrika* **58**(2) 255–271.

Good, I. J., R. A. Gaskins. 1980. Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering and meteorite data. *J. Amer. Statist. Assoc.* **75**(369) 42–56.

Kimeldorf, G. S., G. Wahba. 1970. A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Statist.* **41**(2) 495–501.

Lenk, P. 1988. The logistic normal distribution for Bayesian, nonparametric, predictive densities. *J. Amer. Statist. Assoc.* **83**(402) 509–516.

Lenk, P. 1999. Bayesian inference of semiparametric regression. *J. Roy. Statist. Soc. Ser. B* **61** 863–879.

Lenk, P. 2003. Bayesian semiparametric density estimation and model verification using a logistic-gaussian process. *J. Comput. Graphical Statist.* **12**(3) 548–565.

Lenk, P., D. Wayne, G. Paul, Y. Martin. 1996. Hierarchical Bayes conjoint analysis: Recovery of partworth heterogeneity from reduced experimental designs. *Marketing Sci.* **15**(2) 173–191.

Liew, C. K. 1976. Inequality constrained least-squares estimation. *J. Amer. Statist. Assoc.* **71**(355) 746–751.

Louis, T. A. 1984. Estimating a population of parameter values using Bayes and empirical Bayes methods. *J. Amer. Statist. Assoc.* **79**(386) 393–398.

Louviere, J. J. 2001. What if consumer experiments impact variances as well as means? Response variability as a behavioral phenomenon. *J. Consumer Res.* **28**(3) 506–511.

McCulloch, R., P. E. Rossi. 1994. An exact likelihood analysis of the multinomial probit model. *J. Econometrics* **64** 207–240.

Morikawa, T., M. Ben-Akiva, D. McFadden. 2002. Discrete choice models incorporating revealed preferences and psychometric data. *Econom. Models Marketing* **16** 29–55.

Orme, B., R. Johnson. 2006. External effect adjustments in conjoint analysis. Working paper, Sawtooth Software, Sequim, WA. http://www.sawtoothsoftware.com/download/techpop/externaleffects.pdf.

Savage, L. J. 1954. *The Foundations of Statistics*. John Wiley & Sons, New York.

Shen, W., T. A. Louis. 1998. Triple goal estimates in two-stage hierarchical models. *J. Roy. Statist. Soc. Ser. B* **60**(2) 455–471.

Smith, A. F. M., A. E. Gelfand. 1992. Bayesian statistics without tears: A sampling-resampling perspective. *Amer. Statist.* **46**(May) 84–88.

Stone, M. 1974. Cross-validatory choice and assessment of statistical prediction. *J. Roy. Statist. Soc. Ser. B* **36**(2) 111–147.

Wahba, G. 1983. Bayesian "confidence intervals" for the cross-validated smoothing spline. *J. Roy. Statist. Soc. Ser. B* **45**(1) 133–150.

Wang, J. D. 1996. Asymptotics of least-squares estimators for constrained nonlinear regression. *Ann. Statist.* **24**(3) 1316–1326.