A Unified Approach to Conjoint Analysis Models

Pablo MARSHALL and Eric T. BRADLOW

We present a unified approach to conjoint analysis models using a Bayesian framework. One data source is used to form a prior distribution for the partworths, whereas full-profile evaluations under a rating scale, ranking, discrete choice, or constant-sum scale constitute the likelihood data ("one model fits all"). Standard existing models for conjoint analysis, considered in the literature, become particular cases of the proposed specification, and explicit formulas for the gains of using multiple sources of data are presented. We demonstrate our method on a conjoint analysis dataset containing both self-explicated evaluations and constant-sum profile data on new automobiles originally collected and described by Krieger, Green, and Umesh. Our empirical findings are "mixed" in that for some out-of-sample predictive measures our Bayesian approach is superior to using profile-only or self-explicated–only data, and for other measures it is not. Our findings suggest that the primary determinant as to whether self-explicated data add information above and beyond the profile data is the degree of incongruity between the calibration and validation data formats. Specifically, when the same type of data are collected for both sources, self-explicated data add less, and vice versa. A further contribution of our work (and one that is easily implemented, given the general nature of our approach) is that we take our data and fit constant sum, ranking, and binary choice models to it, allowing us to infer the "change" in information when taking data and transforming its scale (a common practice). A simulation study indicates the viability of this approach. A simple Gibbs sampler simulation scheme adapted to the form of the outcome measure, using data augmentation and Metropolis sampling, is considered for inference under the model.

KEY WORDS: Conjoint analysis, Gibbs sampler, Hierarchical Bayesian methods.

1. INTRODUCTION

Over the last 20 years, conjoint analysis has evolved as the primary marketing research technique for measuring consumer trade-offs between multiattributed products and services. Introduced into marketing by Green and Rao (1971), conjoint analysis examines how consumers develop overall preferences for goods and services by assuming that they take individual utilities or partworths (e.g., regression coefficients denoting each attribute level) and sum them to yield an overall utility value. The utilities are then mapped, via a chosen link function, to an outcome domain (a given scale).

The use of conjoint studies to define products or services with optimal combinations of attributes has been extensive (Wittink and Cattin 1989; Green, Krieger, and Wind 2000). That is, by measuring the relative contribution of each attribute level to the overall product evaluation, conjoint studies have been used to develop new products, to determine optimal prices, to predict market shares, to identify market segments, and to define market opportunities. Furthermore, recent surveys on conjoint analysis and its applications have been presented by Green and Srinivasan (1990) and Carroll and Green (1995).

Despite the ubiquitous nature of conjoint studies, however, there is debate about the method by which the outcome measure(s) should be collected. For instance, current practice ranges from collecting binary choice data ("pick the one out of the pair that you prefer"), to rating products ("rate each of the following on a 1–10 scale"), to performing constant-sum allocation ("assign a total of 100 chips to the following products"), to ordinal ranking ("rank the following from most to least preferred"), to a self-explicated approach (Green 1984) ("what utility do you have for each attribute level and how important is each attribute?"). In addition to these differences, there is also disparity in the "block size" in which the products are shown (i.e., pick one out of two, out of three, etc.). Moreover, in many cases, data in multiple formats are collected (e.g., rating data and self-explicated data).

With this variety in outcome, block size, and data sources, it would be advantageous to have a unified, conjoint approach that accommodates current practices as special cases. In this article we present a unified approach structured by a Bayesian model. This is the primary contribution of this research.

Specifically, the contribution of this research is to present a *unified* approach to conjoint analysis by means of a Bayesian model. Without loss of generality, we describe the model in which two data sources are observed: (1) a set of full profile evaluations (which can be any of choice, rating, ranking, or constant-sum) and (2) self-explicated ratings, which form the prior. Several of the models considered in the conjoint analysis literature, like self-explicated conjoint data only and full-profile conjoint data only, become particular cases of the proposed model. However, this research is by no means the first application of Bayesian methods to combining data sources in conjoint models. The commercially available ACA/HB software (*www.sawtoothsoftware.com*) describes the use of hierarchical Bayes (HB) methods to combine choice based conjoint data and self-explicated ratings. Our contribution in this area is the unified Bayesian approach to handling all forms of outcome data.

When full-profile rating data are observed, estimation can be carried out by means of the Gibbs sampler (Gelfand and Smith 1990). An extension of the basic model in which ranking– or binary choice–based data is observed has

Pablo Marshall is Associate Professor of Marketing and Statistics, Escuela de Administracion, Pontificia Universidad Catolica de Chile, Santiago, Chile (E-mail: *pmarshal@faceapuc.cl*). Eric T. Bradlow is Associate Professor of Marketing and Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104 (E-mail: *ebradlow@wharton.upenn.edu*). The authors thank the editor, the associate editor, and three anonymous reviewers for suggestions that vastly improved the quality of this manuscript, and Paul E. Green and Abba M. Krieger for the use of their automobile data.

Journal of the American Statistical Association
September 2002, Vol. 97, No. 459, Applications and Case Studies
DOI 10.1198/016214502388618410

also been considered, as proposed by Chapman and Staelin (1982) for ranking data and by Allenby, Arora, and Ginter (1995) for binary choice data. A minor extension for estimation, based on the Metropolis–Hastings algorithm (Chib and Greenberg 1995), is implemented in both the ranking and the binary choice case (as in Sawtooth's ACA/HB module). This unified computational approach is a second contribution of this research.

A third contribution of this research is primarily empirical. The dataset used here, a study of new automobile attributes for the purpose of product design, initially analyzed by Krieger, Green, and Umesh (1998), hereafter denoted as KGU98, allows us to fit each of a constant-sum, choice, ranking, and self-explicated conjoint model. That is, we can assess the change in forecast accuracy for in-sample (calibration) and out-of-sample (validation) tasks that accompanies the change in outcome rating scale. Our unified model allows the fitting of all of these models in one framework. To the best of our knowledge, this is the first study to comprehensively fit a model to all of the outcome rating scales and to contrast and compare their results.

The article is organized as follows. Section 2 presents a description of earlier related research. Section 3 gives our general model, assuming rating data, and describes extensions for the cases of ranking, choice-based, and constant-sum data. Section 4 presents an application to the new automobile conjoint data and Section 5 presents a simulation study demonstrating the viability of our approach. Finally, Section 6 gives conclusions and directions for future research.

2. EARLIER RELATED RESEARCH

Originally, in traditional conjoint studies using the full profile approach, respondents either rated or ranked all possible profiles according to an experimental design (e.g., Plackett and Burman 1946). For rated evaluations, analysis of variance or regression models are used to estimate the partworths, whereas the monotone analysis of variance procedure (Kruskal 1965) is typically used for outcomes that are ranks. However, the popularity of conjoint studies has increased the demand on this methodology for solving increasingly larger problems (i.e., more attributes and levels). One problem associated with this trend is that it increases the number of products that the decision maker must evaluate in the full profile procedure. For example, in a relatively small study with 4 attributes at 3 levels each, the respondents must evaluate $3^4 = 81$ full profiles. Even for the most involved decision maker, this task is excessively demanding, time-consuming, and frustrating (Malhotra 1986).

The usual solution has been to consider fractional experimental designs in which all interaction terms of the utility model are assumed equal to 0. As an alternative and complement to fractional designs, several models have been developed in which self-explicated responses to attribute-level desirability and attribute importance are obtained in addition to traditional profile evaluations, the so-called hybrid approach (Green and Krieger 1996; ACA/HB 1999). This is the closest link to our research and motivates the description of our model in Section 3; specifically, self-explicated responses form the prior and profile data (of any format) form the likelihood.

Self-explicated responses are typically obtained in two stages. At the first stage, the decision makers are asked to rate the desirability of each attribute level; at the second stage, they rate the importance of each attribute. Partworths are then obtained by multiplying these desirabilities and importances (i.e., differentially stretching/contracting the desirabilities). The motivation for collecting both self-explicated and full-profile evaluations is to increase precision, especially for large-scale problems, without substantially increasing the data collection tasks. To this category of models belong the popular adaptive conjoint analysis (ACA) methodology of Johnson (1987), the ACA/HB model, the hybrid models of Green and Krieger (1996), and the combining procedure developed by Srinivasan and Park (1997). We consider the model of Srinivasan and Park in Section 4 as a baseline for comparison.

The primary issue addressed by these hybrid (data-merging) models is the appropriate amount of "weight" to give the profile evaluations in contrast to the self-explicated data. Choices used in practice range from ad hoc weighting procedures to a "stacked" regression approach. (The self-explicated and full-profile data matrices are stacked to form the full design matrix.) Our proposed Bayesian procedure directly handles this issue.

3. A GENERAL MODEL

3.1 Data and Model

Consider the case in which individual self-explicated data are collected and partworths are obtained by multiplication of stated importances and desirabilities. Let μ_i denote the $(k \times 1)$ unobserved vector of true partworths for individual i and let u_i denote the observed partworths obtained via self-explication. Then, assuming normality, our model proposes that

$$[\mu_i | u_i, \Omega] \sim N_k(u_i; \Omega), \qquad i = 1, \ldots, n, \tag{1}$$

where k is the number of estimable (identified) attribute levels (the sum of the number of attribute levels minus the number of attributes plus 1), Ω is the partworth variance-covariance matrix, and $N_k(a; b)$ denotes a k-variate normal distribution with mean vector a and variance-covariance matrix b. For example, a product composed of two attributes, one with three levels and the other with four levels, has $k = (3-1) + (4-1) + 1 = 6$ estimable parameters; this is akin to the number of dummy variables estimable in a regression model plus an intercept. That is, the model in (1) specifies that the true unobserved partworths are equal to the observed vector of self-explicated partworths plus a Gaussian error. This is a common way in which covariate information (in this case u_i) is brought into a Bayesian model via the prior mean (Gelman, Carlin, Stern, and Rubin 1995, pp. 366–383). The variance-covariance matrix Ω is assumed to be the same for all individuals but unknown. If the data used for the prior were of a different type, then it would be necessary to change the prior distribution accordingly (e.g., a uniform distribution for rank data).

Consider now the case in which each respondent provides m additional *rating* evaluations, that is, each individual gives Likert ratings to m constructed products. Let y_i denote the $(m \times 1)$ observed full-profile evaluations. Thus we assume that

$$[y_i | \mu_i, \sigma^2] \sim N_m(X_i \mu_i; \sigma^2 I_m), \qquad i = 1, \ldots, n, \tag{2}$$

where X_i is an $(m \times k)$ design matrix that may be less than full rank in fractional experiments or may be independent of the subscript i if the evaluated products or services are the same for all n individuals, μ_i is the vector of unobserved partworths as in (1), σ^2 is a residual variance, and I_m is an $m \times m$ identity matrix. The model defined by (2) is the traditional full-profile conjoint analysis regression model. Note that the model given by (1) and (2) assumes that y_i and u_i are measured on the same scale. We relax this assumption when we generalize the model in Section 3.3.

A Bayesian approach to the model is obtained by combining the prior distribution for μ_i defined by (1), with the profile rating model defined by (2). Given σ^2, the posterior distribution of μ_i depends only on the evaluations of the ith individual and is given by standard formulas for a hierarchical normal-normal model,

$$[\mu_i | y_i, \sigma^2, \Omega] \sim N\left(A_i^{-1}\Omega^{-1}u_i + A_i^{-1}\frac{X_i' y_i}{\sigma^2}; A_i^{-1}\right), \qquad i = 1, \ldots, n, \quad (3)$$

and

$$A_i^{-1} = \left[\Omega^{-1} + \frac{X_i' X_i}{\sigma^2}\right]^{-1}$$

(Zellner 1971).

Therefore, the posterior mean in the distribution of the partworths μ_i combines the two data sources, one source obtained by the self-explicated method, u_i, and the other by full-profile evaluations, y_i, as a weighted average. The weights given to the two measures are defined according to the relative values of the variance-covariance matrices of these quantities. Thus our approach eliminates any ad hockery in the relative weighting of full profile and self-explicated data sources; however, it does so at the expense of the assumption that both data sources are measuring the "same" partworths.

As a second advantage, if the prior for μ_i is noninformative such that $\Omega^{-1} \to 0$, then the posterior mean reduces to the classical regression estimator obtained from full-profile evaluations. Furthermore, if we let the precision of the likelihood go to 0 (i.e., $\sigma^{-2} \to 0$), then our approach contains the standard self-explicated approach as a special case. Our approach, therefore, is more comprehensive than methods that do not combine multiple data sources.

A third benefit is that the estimator of μ_i from (3) is more precise than that obtained from either the self-explicated or the full-profile approach; the variance-covariance matrix in (3) is smaller. That is, the precisions for the two sources of data (profile evaluations and self-explicated) are summed, yielding estimates that are overall more precise.

A final advantage is that the matrix X_i' does not need to be of full rank. Given the prior information on the partworths, proper posteriors can be formed even with a small number of full-profile evaluations. A thorough discussion of this last advantage has been described in detail by Lenk, Desarbo, Green, and Young (1996).

To complete the model specification given by (1) and (2), we assume standard conjugate hyperpriors for the parameter distributions. Conjugate hyperpriors are used for both computational simplicity and to reflect our general lack of knowledge regarding their distribution. We assume that the distribution for Ω and σ^2 are independent, where σ^2 is assumed to follow an inverse chi-squared distribution with g_0 degrees of freedom and Ω has an inverse Wishart distribution with d_0 degrees of freedom and precision parameter D_0. This so-called hierarchical normal-Wishart structure has been applied in many Bayesian applications (e.g., Hoch, Kim, Montgoemery, and Rossi 1995) for a variety of different domains.

3.2 Computation

The model structure allows for computationally intensive but straightforward inference by obtaining posterior samples via a Markov chain Monte Carlo (MCMC) sampler. That is, starting from an initial value for model parameters $\Lambda^{(0)} = (\mu_i^{(0)}, \sigma^{2(0)}, \text{ and } \Omega^{(0)})$, we iterate at each step t of the sampler in turn from the set of full conditional distributions $[\mu_i | y, u_i, \Lambda_{-\mu_i}^{(t)}]$, $[\sigma^2 | y, u_i, \Lambda_{-\sigma^2}^{(t)}]$, and $[\Omega | y, u_i, \Lambda_{-\Omega}^{(t)}]$, where Λ_{-x} denotes all parameters in Λ excluding x. We then iterate the sampler for Q cycles until convergence, then use draws after convergence for inference. Our approach to convergence diagnosis, described more fully in Section 4, is based on observing multiple independent chains from different starting values (Gelman and Rubin 1992).

The full conditional distribution for μ_i is given in (3). The full-conditional distributions for Ω and σ^2, given the conjugate structure used, are inverse Wishart given by

$$[\sigma^2 | y, \mu, \Omega] \sim IW(g, G), \qquad g = g_0 + nm, \qquad G = G_0 + \textstyle\sum_{i=1}^{n}(y_i - X_i\mu_i)'(y_i - X_i\mu_i) \quad (4)$$

and

$$[\Omega | y, \mu, \sigma^2] \sim IW(d, D), \qquad d = d_0 + n, \qquad D = D_0 + \textstyle\sum_{i=1}^{n}(\mu_i - u_i)(\mu_i - u_i)', \quad (5)$$

where g and d are degrees of freedom parameters and G and D are scale matrices. Diffuse prior specifications for Ω and σ^2 are obtained if $d_0 = 1$ and $D_0 = 0$ or $g_0 = 1$ and $G_0 = 0$.

3.3 The Observations

The model described in Section 3 assumes a continuous outcome rating score with ratings and self-explicated scores on the same scale. In practice, the preferences of individuals can be measured in different formats (as described in Sec. 1). The most usual formats are continuous ratings, sum to 1, rankings, and discrete choice. In this section we discuss the application of the model to these data measures by describing the link functions that yield the alternative outcome forms. However, in all cases the base model is as described earlier. In this manner, we have a "unified" Bayesian conjoint model.

3.3.1 Ratings and Self-Explicated Data Collected on Different Scales. To generalize the model, let us assume instead that y_i (the profile ratings) along with self-explicated preferences measures u_i are observed but *not* measured on the same scale. For instance, the rating data may be on 0–100 scale and

the self-explicated data on a 1–10 scale. Then (1) is easily extended to handle this by letting

$$[\mu_i|\alpha, \Omega] \sim N_k(Z_i\alpha; \Omega), \qquad i = 1, \ldots, n,$$

where $Z_i = (\iota, u_i)$, ι is a $(k \times 1)$ vector with all elements equal to 1 and α is a (2×1) vector of unknown scale coefficients (i.e., α contains an intercept and shift parameter allowing for linear rescaling). Previous conditional distributions are still valid, conditional on α, with obvious changes (we insert $Z_i\alpha$ for u_i, because the prior mean for μ_i is $Z_i\alpha$ instead of u_i).

On the other hand, the conditional distribution of α must be rederived because given μ, α is independent of the other model parameters. Assuming a Gaussian prior distribution for α, with mean α_0 and variance Σ_0, we obtain

$$[\alpha|\mu, \Omega] \sim N\left(B^{-1}\Sigma_0^{-1}\alpha_0 + B^{-1}\left(\sum_{i=1}^{n} Z_i'\Omega^{-1}\mu_i\right); B^{-1}\right) \quad (6)$$

and

$$B^{-1} = \left[\Sigma_0^{-1} + \sum_{i=1}^{n} Z_i'\Omega^{-1}Z_i'\right]^{-1}$$

That is, (1) is replaced by (6), and the other equations are modified accordingly.

3.3.2 Constant-Sum Outcome Data. Constant-sum data are often collected in conjoint studies because their richness (strength of preference) and ability to force respondents to make trade-offs are desired. On the down side, however, it is often hard for respondents to accurately assess the variability in their preferences, and hence constant-sum ratings are thought to not "work well" if predicting choice is desired. We will partly address this issue in Section 4.

Assume that the m profiles evaluated by each individual are presented in b blocks of p profiles each, such that $(m = b * p)$, and assume that y_{ijk} corresponds to the latent utility of the kth element of the jth block presented to individual i. Instead of observing a continuous outcome rating score (as in Section 3) that which has $p = 1$ and $b = m$ for full-rating data, we observe

$$y_{ijk}^* = \frac{y_{ijk}}{\sum_{k=1}^{p} y_{ijk}}, \qquad i = 1, \ldots, n, \ j = 1, \ldots, b, \ k = 1, \ldots, p, \quad (7)$$

the fraction of total "worth" associated with the jkth element within its block. Note that the underlying model for utility is assumed to be the same as in (2). What has changed is that the identity link function applicable for rating data no longer applies; it is simply replaced by a link function equal to the fraction of total worth.

The model of Section 3.1 can be extended to consider these observations. First, the y_i is now unobserved (y_i^* is observed) and must be estimated from the model. Second, given y_i, the conditional distributions of the parameters in the previous section are independent of y_i^*, and hence they remain the same. The only conditional distribution that changes is that of y_i, because now this distribution must be computed, given y_i^*. To generate draws of y_i given y_i^* and the other model parameters Λ, we first generate for each block draws of $\sum_{k=1}^{p} y_{ijk}$ from (2) and then compute y_{ijk} from (7), a data augmentation approach (Tanner and Wong 1987). The draws of $\sum_{k=1}^{p} y_{ijk}$ are restricted so that each element is positive. This can be done by pure rejection of draws that do not satisfy the constraint. Finally, the variance of the unobserved components, y_i, is not completely identified (because we now are taking ratios of values), and thus we impose the restriction $\sigma^2 = 1$.

3.3.3 Ranking Data. Assume that the m total profiles are presented in b blocks of p profiles each, and assume again that y_{ijk} corresponds to the utility of the jkth element of y_i. Instead of y_{ijk}, we observe the ranking of the profile within each block, y_{ijk}^*. To model the probability, a given object gets ranked zth in its block ($1 \le z \le p$, z an integer), where lower z indicates a more preferred object, we follow the exploding logit model of Chapman and Staelin (1982). This model states that

$$P(y_{ijk}^* = z) = \frac{e^{y_{ijz}}}{\sum_{k=z}^{p} e^{y_{ijk}}}, \qquad i = 1, \ldots, n, \ j = 1, \ldots, b, \ k = 1, \ldots, p, \quad (8)$$

where y_{ijk} is the utility of the profile ranked zth in block j for person i with utility (in our case) given by the specification in (2).

To generate draws of y_i given y_i^*, and the other model parameters Λ, we first generate for each block draws of the minimum of the values $y_{ijk}, k = 1, \ldots, p$ from (2). Then we draw the second smallest value from (2) restricted to be greater than the previous value, and continue until the p evaluations are simulated. These restricted draws can be generated by pure rejection of draws or by generating draws from a uniform distribution on a restricted interval and then applying the inverse of the normal distribution (Allenby et al. 1995).

3.3.4 Discrete Choice Data. Assume that the m total profiles are presented in b blocks of p profiles each, and that instead of y_{ijk} we observe

$$y_{ijk}^* = \begin{cases} 1 & \text{if profile } k \text{ is the most preferred in block } j \\ 0 & \text{otherwise.} \end{cases}$$

Discrete choice data in blocks of size 2 is certainly very common in conjoint studies; it is pairwise conjoint and is a special case of this more general model.

To model $P(y_{ijk}^* = 1)$, we follow the logistic approach of Allenby et al. (1995), which yields

$$P(y_{ijk}^* = 1) = \frac{e^{y_{ijk}}}{\sum_{k=1}^{p} e^{y_{ijk}}}, \qquad i = 1, \ldots, n, \ j = 1, \ldots, b, \ k = 1, \ldots, p, \quad (9)$$

where y_{ijk} is given by the specification in (2).

In this case $[y_i^* \mid y_i, \Lambda]$ is given by a product of Bernoulli probabilities (Albert and Chib 1993), the probability of the binary choices given by (9), and does not lend itself to straightforward MCMC sampling. However, well-known procedures, such as the Metropolis–Hastings algorithm (Hastings 1970), can be used to obtain realizations of y_i, given y_i^*.

4. AN APPLICATION

4.1 New Automobile Conjoint Data

We applied our unified Bayesian conjoint model to data originally collected and described by KGU98. The KGU98 conjoint data were obtained from 128 respondents enrolled in a large private Northeastern university on preferences among new automobiles. The study group was 65% male and had an average age of 24 years old, and 83% were either single or divorced. Furthermore, 65% indicated that they had recently been involved in a new or used car purchase, indicating their degree of knowledge about the product category.

Each automobile was described in terms of six attributes: make and model (4 levels), miles per gallon (3 levels), base price (3 levels), exterior color (4 levels), sound system (3 levels), and warranty (3 levels). Details of the levels of each attribute are given in table 1 of KGU98.

A block design (not fully crossed) was implemented as follows. A master set of 64 profiles was constructed from an orthogonal main effects plan such that each level of each attribute appeared at least once in each block. Blocks were constructed to contain $p = 4$ profiles (cars). The 64 master profiles were grouped into 16 blocks of 4, and further split into 2 orthogonal halves of $b = 8$ blocks. Each respondent then received one of the two sets of eight blocks for *both* the calibration and validation tasks; that is, each respondent received the *same* set of eight blocks twice. Block order and profile order within blocks were randomized as a further attempt to reduce the effects of showing the same blocks twice. The ratings obtained the first time the set of eight blocks were shown are used for validation, and those obtained the second time the set was shown are used for calibration, as in KGU98.

The format of the data are as follows. For the validation data, respondents were shown each of the 8 blocks and asked to allocate 100 points across the 4 profiles (i.e., constant-sum data). The second time through the eight blocks of the calibration data, respondents were asked to rate each profile on a 0–100 scale. We note that KGU98's intent for this design was to assess whether constant sum data could be "replaced" by ratings data rescaled to sum to 1. Although not the primary interest here, later we address the replacability issue in what we call the "pure error" model.

All respondents also received a self-explicated rating task in which they were asked for 0 to 10 ratings for all 20 attribute levels. Then a constant sum importance task followed in which they were asked to allocate a total of 100 points across the 6 automobile attributes. These were then multiplied together to yield observed self-explicated partworths, u_i, as in Section 3.

As a second baseline, we also computed self-explicated partworths using the methodology of Srinivasan and Park (1997). They proposed a two-stage approach in which at the first stage each respondent is asked to identify attribute levels that are totally unacceptable. These acceptable levels are considered at the second stage, and self-explicated partworths are obtained as the product of preferences ratings and importances. The importance of each attribute is computed from the difference between the original preference rating for the most desirable and least desirable levels. The first stage of the Srinivasan and Park (1997) procedure was not considered in the computation here. Thus, for each attribute, all of the levels were used to form the self-explicated scores.

Although the KGU98 data are atypical (blocks of four, Likert scale, converted to constant sum), they are very rich in other ways that justified their use. First, respondents answered the same blocks twice, providing a unique out-of-sample validation task. Second, the constant sum data gives us strength of preference data that can be rescaled to choice or ranking data (see Section 4.2). Finally, the data contained two sources (constant-sum and self-explicated), allowing us to fit our unified conjoint model.

Therefore, we present the results of three exercises with this dataset. In analysis 1 we treated the calibration profile data as constant sum (their "natural" state). In analysis 2 we took the calibration data and computed the rankings within each block. We note that much research has suggested that simply taking strength of preference data and converting it to ranks will *not* yield consistent rankings. Finally, in analysis 3 we took the calibration data and set the choice indicator for that profile within a block that received the highest rating equal to 1; all others were set to 0 (i.e., we transformed the data to choice data). An area for future research is collection of data naturally in each of these formats, rather than the transformation done here. A comparison of results using our Bayesian model for the alternative forms of likelihood data, as well as using profile data only and self-explicated data only, follows.

4.2 Results and Implications

For each analysis, we obtained 10,000 draws from an MCMC sampler. Covergence was deemed to have occurred within the first 2,500 draws, and hence all reported inferences are from the last 7,500 draws. The hyperprior distribution for Ω was chosen as diffuse.

A first set of inferences was made as follows. For each individual $i = 1, \ldots, 128$, we computed $\bar{\mu}_i$, the posterior mean partworth with corresponding mean predicted scores $\hat{y}_i = X_i \bar{\mu}_i$. These predicted scores were then scaled to sum to 1 and compared with the observed validation data that were constant-sum data. To compare the fit of the predictions, $\hat{y}_i$, to the observed outcomes y_i, we used the following statistics: root mean square error (RMSE), mean absolute deviation (MAD), correlation coefficient, proportion of correct first choice, proportion of correct ranking, and MAD of the ranking vectors (within-block and then averaged across block and person). For comparison, the fit statistics were also computed for three models: the self-explicated model, which assumes that $\sigma^{-2} = 0$; the profile data–only model, which assumes $\Omega^{-1} = 0$, and the fully Bayesian model. We also computed the fit statistics for the case that sets $\hat{y}_i = y_i$. That is, because the eight blocks of four products were shown to each individual twice, we simply used the observed calibration constant-sum data to predict the validation data. We call this a "pure error"

Table 1. Out-of-Sample Goodness-of-Fit Statistics for the Sum to 1 Data

Model	RMSE	MAD	Correlation	First choice	Ranking	MAD ranking
Pure error	.1177	.0903	.7209	.6719	.5962	.5308
Self-explicated model*	.1464	.1211	.5954	.5664	.4524	.7378
Profile model	.1167	.0926	.7338	.6357	.5417	.5771
Bayesian model	.1166	.0925	.7344	.6377	.5442	.5732

*Statistically different from the Bayesian model at the 5% level.
NOTE: The pure error corresponds to the case of a perfect fit. The self explicated model sets $\sigma^{-2} = 0$ and the profile model sets $\Omega^{-1} = 0$. Estimates are based on 7,500 replications.

model, because the only variation is the amount by which persons respond differently when confronted with the same profile blocks a second time. The results of these analyses are presented in Table 1.

4.2.1 Sum to 1 Predicting Sum to 1. From Table 1, we observe that (a) the profile model and Bayesian model do not perform statistically different on any of the fit statistics (although the Bayesian model fits slightly better on all), (b) the self-explicated model performs significantly worse than either the profile or the Bayesian model on all statistics, and (c) the profile and Bayesian model perform equivalently to the pure error model. Although these findings are clearly specific to this dataset, some implications are warranted. First, both the profile and Bayesian models perform equally well when compared to the pure error model, which is somewhat remarkable. It suggests that the profile and Bayesian approaches do as well as measuring the data twice. Second, the self-explicated data alone do not do as well for the constant-sum data. On the surface, this seems at odds with existing research suggesting the superiority, in some cases, of self-explicated data. However, given that the calibration data and validation data are the same blocks (i.e., each person saw them twice) and collected under the same format, whereas the self-explicated data are collected differently, these findings are as expected. In Section 4.2.4 we address the benefit of the self-explicated data when the formats of the calibration and validation data are incongruous, that is, the model "needs" the self-explicated data.

4.2.2 Ranking Data Predicting Sum to 1. The second analysis considers transforming the calibration data to rankings within blocks and using them to predict the validation sum to 1 data. As before, we computed the average $\bar{\mu}_i, i = 1, \ldots, 128$ and predicted scores $\hat{y}_i = X_i\bar{\mu}_i$. Now, to obtain the predicted proportions, we compute the probability of purchase within each block as

$$\hat{\pi}_{ijk} = \frac{e^{\hat{y}_{ijk}}}{\sum_{k=1}^{4} e^{\hat{y}_{ijk}}}, \qquad i = 1, \ldots, 128,\ j = 1, \ldots, 8,\ k = 1, \ldots, 4.$$

We then compared these probabilities with the validation scores exactly as in the first analysis.

From Table 2 we observe that the Bayesian model performs better than the profile data–only model for the RMSE, MAD, and correlation statistics with no statistical differences for first choice, ranking, and MAD. Hence the self-explicated data add value here but are limited due in part to the transformation from the original strength of preference (rating) data to ordinal (ranking) data.

Table 2. Out-of-Sample Goodness-of-Fit Statistics for the Ranking Data

Model	RMSE	MAD	Correlation	First choice	Ranking	MAD ranking
Profile model	.3460*	.2725*	.5586*	.6348	.5530	.5806
Bayesian model	.3322	.2617	.5870	.6484	.5596	.5654

*Statistically different from the Bayesian model at the 5% level.
NOTE: The profile model sets $\Omega^{-1} = 0$. Estimates are based on 7,500 replications.

4.2.3 Choice Data Predicting Sum to 1. In the third set of analyses run, the profile scores were turned into choice data, and predictions $\hat{y}_i$ and $\hat{\pi}_{ijk}$ were then computed. The results using the same fit statistics are reported in Table 3.

From Table 3 we observe that when the profile rating data are turned into choice data, the profile model performs worse than the Bayesian model on five of the six measures. Notice now that the data format of the profile data and validation data are different, so that the self-explicated data reduces the errors in terms of predictive performance. This suggests that the loss of information differentially affects the different models. We see that when the data are turned into choice data, additional benefit comes from including the self-explicated scores along with the profile data. These findings emphasize our point that the benefit of multiple data sources certainly depends on the format of the outcome collected and the outcome type to which one is forecasting.

4.2.4 Comparison of Model Fit Across Data Formats. One last set of implications that can be derived entails comparing the magnitude of results from Tables 1, 2, and 3. First, the Bayesian model for constant-sum data performs statistically better in three of the six measures when compared to the model fit using ranking data and in five of the six measures when compared to the model for choice data. Therefore, as before, rescaling the data down to a less-rich format can hurt model prediction. In the Bayesian model, the self-explicated data make up for some of the loss of information due to the transformation.

Second, we notice that the RMSE and MAD are higher and the correlation is smaller for the models for ranking and first choice data, indicating the loss of information by dichotomizing the information. Thus, one thing that is clear is that for these data there appears to be significant variation due to the choice of data format. This is consistent with other literature

Table 3. Out-of-Sample Goodness-of-Fit Statistics for the Discrete Choice Data

Model	RMSE	MAD	Correlation	First choice	Ranking	MAD ranking
Profile model	.3559*	.2815*	.5694*	.6738	.4270*	.7935*
Bayesian model	.3400	.2682	.5778	.6631	.4675	.7075

*Statistically different from the Bayesian model at the 5% level.
NOTE: The profile model sets $\Omega^{-1} = 0$. Estimates are based on 7,500 replications.

Table 4. Out-of-Sample Goodness-of-Fit Statistics for the Self-Explicated Data

Method	RMSE	MAD	Correlation	First choice	Ranking	MAD ranking
Srinivasan and Park (1997)	.1474	.1207	.5064*	.5000*	.4065*	.8379*
Krieger, et al. (1998)	.1464	.1211	.5954	.5664	.4524	.7378

*Statistically different from the method of Krieger et al. (1998) at the 5% level.

that suggests the close performance of different conjoint models, but the form of the data themselves matters. Most studies do not perform analysis using both constant-sum and a binary transformed version of the same dataset, and thus this comparison has been examined much less often. Hence this is a further contribution of our research.

4.2.5 Self-Explicated Model Results. Reported in Table 4, we consider a comparison between two methodologies for computing the self-explicated partworths (Srinivasan and Park 1997; KGU98). Our results indicate that the approach of Srinivasan and Park performs worse in five of the six statistical measures that we report. Although we did not implement the first stage of the Srinivasan and Park procedure (in which totally unacceptable attribute levels are deleted), the results show that we obtain better results using the self-explicated scores when we consider explicitly the importance of attributes than in their approach, which infers them. Again, this result may not hold in general, but is of interest in this application.

5. SIMULATION EXAMPLE

To test the accuracy of our approach under a variety of conditions, we designed a $3 \times 3 \times 3 + 1$ simulation example similar to the KGU98 automobile data. The factors in the simulation were type of data, with four levels (rating, sum to 1, ranking, and discrete choice); amount of error variance σ^2, with three levels $(0.1, 1, 10)$; and estimation approach with three levels (Bayesian, profile only, and self-explicated only). We note that the self-explicated model, as before, is done only as a rating score model. This simulation covers data types, amount of shrinkage of the profile data to the self-explicated data (via σ^2 and Ω), and estimation approaches described in this research. We used 16 (4 blocks of 4) profiles for model estimation and 16 (different 4 blocks of 4) for out-of-sample validation. The number of partworths per profile is 8, and the number of individuals per simulated dataset is 100. For each simulation condition, 7,500 datasets were generated and used in the estimation. The simulated data were generated using the following procedures. Individual-level self-explicated partworths were obtained from

$$u_i \sim N(\mu, 10I_8), \qquad i = 1, \ldots, 100,$$

where the partworth parameters are defined as

$$\mu = (20, -3, 0, 3, -1, 1, -5, 5).$$

The components of μ were set so as to have a large variation in attribute importances. Individual partworths for the profiles were obtained, centered around the self-explicated scores, from

$$\mu_i \sim N(u_i, \Omega), \qquad i = 1, \ldots, 100,$$

with $diag(\Omega) = \{1, 16, 16, 16, 4, 4, 25, 25\}$ and $[\Omega]_{23} = [\Omega]_{24} = [\Omega]_{34} = [\Omega]_{32} = [\Omega]_{42} = [\Omega]_{43} = 12$. Thus partworths 2, 3, and 4 had a moderate degree of positive covariation, and other partworths had no covariation. The rating scale observations were then obtained from

$$y_i \sim N(X\mu_i, \sigma^2 I_{32}), \qquad i = 1, \ldots, 100,$$

where X is the design matrix and the parameter σ^2 has three possible values as described earlier.

The simulation results are given in Table 5. We summarize our findings as follows. First, there is a reduced ability to predict profile scores as one degrades the calibration data from the actual profile scores to the sum to 1 scores, to the ranking scores, and then to the discrete choice scores. This ordering of actual to sum to 1 to ranking to discrete choice is consistent with previous studies in the literature as well as with our automobile data findings. Even more, it seems that this ability to predict scores as one degrades the data does not depend on the parameter σ^2. Second, the Bayesian estimation approach is always better than the profile approach as well as the self-explicated approach. The gains of the Bayesian approach are more important in the ranking and discrete choice data than in the rating or sum to 1 scores data. Third, the ability to forecast out-of-sample profile ratings improves as σ^2 decreases. This is not surprising as the amount of individual variation decreases. All six measures of statistical accuracy have significant variation due to both σ^2 and method, suggesting that the accuracy of findings and their implications depend on the data. Finally, the Bayesian approach works; particularly when σ^2 is small, it has very high forecast accuracy and outperforms the other two approaches.

We also note that we tested the accuracy of each of the approaches to obtain the actual partworths. The pattern of results is virtually identical to that shown in Table 5 for the profile scores themselves.

6. CONCLUSIONS

We have provided a single model that can use multiple forms of calibration conjoint data in a coherent framework. The Bayesian model presented here provides a natural framework that appropriately weights each data source according to the information contained. Accordingly, this solves (to an initial degree) one open issue in conjoint analyses. In addition, we have demonstrated how this model can be used to fit rating score, constant-sum, ranking, and binary choice data. Our application of these models to the automobile data of KGU98 demonstrates the viability of our approach, as well as some implications for information loss by transforming the data. Because we analyzed the same dataset using two different models, we were able to compare the relative importance of changing data formats versus changing the underlying models.

*Table 5. Out-of-Sample Profile Scores Goodness-of-Fit Statistics**

Data type	σ^2	Model	RMSE	MAD	Correlation	First choice	Ranking	MAD ranking
Rating	10	Profile	4.489	3.685	.812	.580	.565	.541
Rating	10	Bayesian	**3.945**	**3.229**	**.850**	.620	**.601**	**.475**
Sum to 1	10	Profile	12.330	11.022	.797	.585	.551	.556
Sum to 1	10	Bayesian	11.649	10.307	**.801**	.583	.554	.549
Ranking	10	Profile	23.710	20.557	.710	.528	.506	.640
Ranking	10	Bayesian	**17.463**	**16.802**	**.825**	.578	**.581**	**.511**
Discrete choice	10	Profile	26.816	23.112	.594	.425	.451	.749
Discrete choice	10	Bayesian	**17.883**	**17.054**	**.704**	**.528**	.495	**.653**
	10	Self Explicated	6.237	5.187	.771	.568	.566	.554
Rating	1	Profile	1.419	1.165	.976	.850	.829	.184
Rating	1	Bayesian	**1.340**	**1.097**	**.979**	.860	**.848**	**.163**
Sum to 1	1	Profile	10.940	10.244	.974	.828	.813	.204
Sum to 1	1	Bayesian	**7.414**	**6.845**	.974	**.853**	.818	.195
Ranking	1	Profile	19.190	18.155	.864	.683	.642	.418
Ranking	1	Bayesian	**16.520**	**15.771**	**.950**	**.763**	**.748**	**.283**
Discrete choice	1	Profile	22.310	19.430	.680	.468	.482	.670
Discrete choice	1	Bayesian	**17.901**	**17.361**	**.868**	**.573**	**.592**	**.475**
	1	Self Explicated	5.397	4.598	.846	.703	.669	.415
Rating	.1	Profile	.449	.369	.998	.943	.948	.053
Rating	.1	Bayesian	**.441**	**.363**	**.998**	**.943**	.948	.053
Sum to 1	.1	Profile	13.201	12.527	.997	.943	.941	.059
Sum to 1	.1	Bayesian	**6.680**	**6.266**	**.994**	.938	.924	.078
Ranking	.1	Profile	19.218	18.474	.880	.745	.707	.336
Ranking	.1	Bayesian	**17.269**	**16.573**	**.962**	**.828**	**.803**	**.210**
Discrete choice	.1	Profile	22.002	19.278	.694	.493	.476	.649
Discrete choice	.1	Bayesian	**17.590**	**16.992**	**.879**	**.580**	**.603**	**.454**
	.1	Self Explicated	5.301	4.516	.855	.725	.683	.390

*Numbers in bold face correspond to cases where the Bayesian model is statistically different from the profile model at the 5% level.
NOTE: The self-explicated model sets $\sigma^{-2} = 0$, and the profile model sets $\Omega^{-1} = 0$. Estimates are based on 7,500 replications.

So what is next? Certainly, one obvious step is to apply this model to datasets from other domains to examine the robustness of the empirical findings. From these studies, researchers may be able to make informed trade-off decisions between profile and self-explicated data to yield equally precise partworth estimates, for a fixed amount of respondent burden.

In summary, as Bayesian "friendly" software (e.g., the BUGS system; Spiegelhalter, Thomas, Best, and Gilks 1996) becomes more commonplace in the marketing research toolkit, we expect that our model, which contains both the self-explicated–only and profile-only models as special cases, will gain popularity. Because the models considered here fall under the general rubric of Bayesian hierarchic regressions, they are implementable via packages such as BUGS. Of course, a corresponding increase in computing speed also will make these models more practical.

[Received May 2002. Revised May 2002.]

REFERENCES

Albert, J. H., and Chib, S. (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88, 669–679.

Allenby, G. M., Arora, N., and Ginter, J. L. (1995), "Incorporating Prior Knowledge Into the Analysis of Conjoint Studies," *Journal of Marketing Research*, 32, 152–162.

Carroll, J. D., and Green, P. E. (1995), "Psychometric Methods in Marketing Research: Part I, Conjoint Analysis," *Journal of Marketing Research*, 32, 385–391.

Chapman, R. G., and Staelin, R. (1982), "Exploiting Rank-Ordered Choice Set Data Within the Stochastic Utility Model," *Journal of Marketing Research*, 19, 288–301.

Chib, S., and Greenberg, E. (1995), "Understanding the Metropolis–Hastings Algorithm," *The American Statistician*, 49, 327–335.

Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of The American Statistical Association*, 85, 398–409.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995), *Bayesian Data Analysis*, London: Chapman and Hall.

Gelman, A., and Rubin, D. B. (1992), "Inference for Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457–511.

Green, P. E. (1984), "Hybrid Models for Conjoint Analysis: An Expository Review," *Journal of Marketing Research*, 21, 155–159.

Green, P. E., and Krieger, A. M. (1996), "Individualized Hybrid Models for Conjoint Analysis," *Management Science*, 42, 850–867.

Green, P. E., Krieger, A. M., and Wind, Y. (2000), "Thirty Years of Conjoint Analysis: Reflections and Prospects." Technical report, Wharton School.

Green, P. E., and Rao, V. R. (1971), "Conjoint Measurement for Quantifying Judgmental Data," *Journal of Marketing Research*, 8, 355–363.

Green, P. E., and Srinivasan, V. (1990), "Conjoint Analysis in Marketing Research: New Developments and Directions," *Journal of Marketing*, 54, 3–19.

Hastings, W. K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, 57, 97–109.

Hoch, S. J, Kim, B. D., Montgomery, A. L., and Rossi, P. E. (1995), "Determinants of Store Level Price Elasticity," *Journal of Marketing Research*, 32, 17–29.

Johnson, R. M. (1987), "Adaptive Conjoint Analysis Software," in *Sawtooth Software Conference Perceptual Mapping, Conjoint Analysis and Computer Interviewing*, Ketchum, ID: Sawtooth Software, pp. 253–265.

Krieger, A. M., Green, P. E., and Umesh, U. N. (1998), "Effect of Level of Disaggregation on Conjoint Cross Validations: Some Comparative Findings," *Decision Sciences*, Fall, 29, 1049–1059.

Kruskal, J. B. (1965), "Analysis of Factorial Experiments by Estimating Monotone Transformations of the Data," *Journal of The Royal Statistical Society*, Ser. B, 27, 251–263.

Lenk, P. J., DeSarbo, W. S., Green, P. E., and Young, M. R. (1996), "Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity From Reduced Experimental Designs," *Marketing Science*, 15, 173–191.

Malhotra, N. K. (1986), "An Approach to the Measurement of Consumer Preferences Using Limited Information," *Journal of Marketing Research*, 23, 33–49.

Plackett, R. L., and Burman, J. P. (1946), "The Design of Optimum Multifactorial Experiments," *Biometrika*, 33, 305–325.

Sawtooth Software (1999), *www.sawtoothsoftware.com,* ACA/hierarchical Bayes technical paper.

Spiegelhalter, D. J., Thomas, A., Best, N. G., and Gilks, W. R. (1996), *BUGS: Bayesian Inference Using Gibbs Sampling*, Version 0.5, (http://www.mrc-bsu.cam.ac.uk/bugs).

Srinivasan, V., and Park, C. (1997), "Surprising Robustness of the Self-Explicated Approach to Customer Preference Structure Measurement," *Journal of Marketing Research*, 34, 286–292.

Tanner, M. A., and Wong, W. H. (1987), "The Calculation of Posterior Distributions via Data Augmentation," *Journal of the American Statistical Association*, 82, 528–540.

Wittink, D. R., and Cattin, P. (1989), "Commercial Use of Conjoint Analysis: An Update," *Journal of Marketing*, 53, 91–96.

Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, New York: Wiley.