Seoul Journal of Business
Volume 19, Number 1 (June 2013)

Identifying Subject-Specific Relevant Explanatory Variables in Choice-Based Conjoint Studies

JIN GYO KIM*
Seoul National University
Seoul, Korea

Abstract

It is customary in conjoint studies to introduce the same set of potential explanatory variables for each subject, so as best to allow any possible trade-offs to be made. However, this presumption can mask the possibility of some subjects' considering only a subset of the presented attributes. Moreover, such subsets of relevant attributes can vary considerably across the population. This paper presents a model which allows researchers to identify relevant explanatory variables for each subject separately. This is accomplished via a solution to the well-known variable selection problem in the context of discrete choice models; the proposed solution can be widely applied throughout choice studies and in fact to other response types, such as ratings, direct paired comparisons, and ranks, with appropriate changes in likelihood function.

When estimated on a choice-based conjoint data for dial-readout scale products, the proposed model is strongly preferred to the traditional random-effect specification for choice-based conjoint. A sizeable group of subjects, approximately 63%, were found to consider proper subsets of all attributes presented. There was a great deal of heterogeneity in attributes deemed relevant across subjects: the proportion of subjects who did not consider a given attribute among the six used in the study ranged from 17.4% to 41.3%. For those who did consider a given attribute, estimated attribute level part-worths were essentially identical for the proposed model and the traditional random-effect conjoint model; but this was not the case for non-considered attributes. In fact, the traditional model was found to suffer from systematic biases in aggregate part-worth magnitudes. Finally, and most important for marketing practice, allowing for the possibility that some subject may not consider particular attributes can lead to substantial design and revenue differences in supposedly 'optimal' products, at both the individual- and the aggregate-level.

---

* Graduate School of Business, Seoul National University, Shillim-Dong, Kwanak-Ku, Seoul 151-916, Korea.

Keywords: Model Specification, Variable Selection, Model Selection, Conjoint, Choice Models, Heterogeneity, Bayesian Methods, Markov chain Monte Carlo

1. INTRODUCTION

Traditional choice-based conjoint analysis has, for reasons of parsimoniousness, operated as if all subjects used an identical set of attributes, usually all that are available, to enact trade-offs. While the weights they assign to various explanatory variables and variable interactions can vary across subjects, the base set of variables typically is identical. Parsimony aside, there is no compelling theoretical justification for this presumption. In fact, the opposite is more plausible on its face: that each subject has a 'relevant set' of attributes that are considered in deciding amongst presented alternatives. For many subjects, this set can be the same; in fact, the set may be the full set presented in the course of the conjoint task. It is difficult to argue, however, that this is invariably the case for all subjects, and there may in fact be a good deal of heterogeneity in relevant attributes across them.

It is instructive to consider a deliberately oversimplified example, illustrating how relevant attributes may require interpretations differing from extant ones. Consider a conjoint experiment on personal computers. Assume that there are two PC users – a "CPU-intensive computation" user and a "novice internet surfing" user – and also that attributes planned for the conjoint experiment are manufacturer (IBM, Dell, HP), CPU clock speed (2GZ, 3GHZ), RAM size (1G, 2G), and technical support (good, poor). We expect the CPU-intensive user to consider such performance-related attributes as CPU clock speed and RAM size, but to place substantially less emphasis, or none at all, on the others (throughout, we use *consider* as a shorthand for which variables affect subject-specific choices, not to suggest specific cognitive strategies or the formation of consideration sets). However, the novice internet user may place greater weight on manufacturer and technical support, but be indifferent to the more performance-related attributes, owing perhaps to complete lack of familiarity. In this case, traditional conjoint models can allow non-zero part-worths for irrelevant or non-considered attributes. When researchers try, as is their wont, to optimize product designs *conditional* on those part-worths, resulting product designs can be driven by variations across them. If the true part-worth variation is literally zero – as it would be whenever someone truly does not distinguish among levels of a given attribute, as the novice internet surfing user treats CPU clock speed – resulting optimal product designs can go awry. Moreover, this can occur even at the aggregate level. For example, the means of part-worths for CPU clock speeds across two users will under-estimate the true part-worths for the advanced user, but over-estimate them for the novice user.

While this is, again, a deliberately idealized scenario, it nonetheless underscores that the assumption of *fixed relevant attributes* – that is, all subjects consider all attributes – can in some cases lead to biased inferences about the most basic drivers of purchase intent: individual- and aggregate-level part-worths. Consequently, purportedly optimal product designs may be anything but optimal. Models that preclude heterogeneity in relevant attribute configurations from the outset may therefore attribute choice observations to overtly-expressed attribute levels to a greater, or smaller, extent than they should.

To avoid such potential mis-attribution problems, this paper aims to develop a model that explicitly incorporates heterogeneity in relevant attribute configurations across subjects. At its core, this problem is related to the *model specification* issue, which is among the important unsolved problems in the conjoint research (cf., Green et al. 2001). The model specification issue refers to the identification of relevant explanatory variables *for each subject* among all candidate variables. It is not difficult to see that this problem is closely related to the *variable-selection* problem in statistics.

Most previous studies on the variable selection problem in statistics literature have focused mainly on (linear) regression models. Forward selection, backward elimination, or a combination thereof are among the most popular techniques using maximum likelihood methods, and have been built into common software packages. This class of methods, however, is inapplicable to choice-based conjoint studies since, in choice models, the utilities are *unknown* random variables. Other popular selection methods in the maximum-likelihood approach utilize penalty functions such as C_p criterion (Mallows 1973), AIC (Akaike 1974) and BIC (Shibata 1984), all of which have been widely applied throughout choice modeling studies. Unfortunately, using such penalty criteria to identify subject-specific relevant explanatory variables is enormously cumbersome, even when combinatorially feasible, since numerous models with different explanatory variable configurations must be estimated for each subject. There is also a large body of Bayesian literature on the variable selection problem in regression models (cf. Andrieu et al. 2001; Bernardo and Smith 1994; Laud and Ibrahim 1995; Lindley 1968; Mitchell and Beauchamp 1988). Probabilistic fit in the form of latent mixture modelling has also drawn considerable attention (Brown et al. 1998; George and McCulloch 1993; Smith and Kohn 1996). Despite this sizeable body of variable selection literature in the linear regression setting, we are unaware of any results which can identify *individual-specific* sets of relevant explanatory variables for choice models. In short, variable selection for choice-based conjoint is very much an open problem.

As discussed before, the presumption that explanatory variables are the same for all subjects may lead to biased part-worth estimates, and consequently sub-optimal product designs based upon the estimates. For example, if a consumer truly does not consider a particular explanatory variable, any changes in the variable cannot, by definition, affect his/her choice decision; therefore, that variable's true regression coefficient is zero for the consumer in question. The prevalent approach in prior literature (using choice-based conjoint models with fixed explanatory variables) is to have coefficients stem from some distribution, as in the popular hierarchical Bayesian specification, and hope that truly irrelevant variables will wind up with regression coefficients close to zero. This is well and good when the variable with which the coefficient is multiplied is relatively small. However, when modelers engage in optimization *conditional* on the estimated coefficients, small coefficients can be blown up considerably, whereas precisely zero coefficients cannot. Simply put, the difference between coefficients being "almost zero" and "exactly zero" is one not of degree, but of kind, when post hoc optimization is involved. Any such potential errors can be avoided entirely by identifying irrelevant explanatory variables for each subject separately, and setting the associated coefficients exactly to zero. This simply is not an option within existing schemes for coefficient heterogeneity.

This study aims to develop a model that enables researchers to identify a set of explanatory variables relevant to choice decisions for each subject. Specifically, subject-specific relevant explanatory variables are identified by stochastically exploring possible configurations of all subsets of candidate explanatory variables. The proposed procedure is illustrated using real choice-based conjoint data set for a small durable. The application will help answer the following questions about the usefulness of identifying subject-specific relevant explanatory variables:

1. Are all explanatory variables relevant for all individuals? Do all individuals consider all attributes?
2. If not, can researchers identify which explanatory variables appear to drive choice for each subject?
3. To what degree does relevancy vary across subjects for each potential explanatory variable?
4. How can knowing which variables are relevant to specific subjects help marketers enact better policies?

The remainder of the paper is organized as follows. We first formulate a very general choice model setting, through which it is possible to tag certain subsets of variables as having non-zero coefficients. This formulation involves complex estimation and inference challenges, which are taken up at length. We then present a choice-based conjoint data set for dial-readout bathroom scales, and show how the model can be applied to it. Finally, we address each of the questions above, both in terms of the scale data set, and in terms of academic and managerial extensions.

2. A MULTINOMIAL PROBIT MODEL WITH HETEROGENEOUS RELEVANT EXPLANATORY VARIABLES

This section presents a model that allows researchers to estimate relevant explanatory variables for each subject in a probit model setting.

2.1 Model Specification

Let $y_{st} = j$ denote the event that subject s ($s = 1,\ldots, S$) chooses alternative j ($j = 1, \ldots., J$) on choice task t ($t = 1, \ldots, T_s$), where $j = J$ indicates the " no choice" option. Let $A = \{a_1, \ldots, a_Q\}$ be a set of candidate explanatory variables, consisting of main effects and possibly interaction effects of attributes; Φ denote a set containing all possible subsets of Q elements in A, plus the null set, { }; and $R = (R_1, \ldots, R_K)'$ denote a K–dimensional design vector given all elements in A. For example, suppose there are three attributes, $\{N_1, N_2, N_3\}$, each with three levels, and also that both main and interaction terms are considered, with interaction terms allowed even without corresponding main effect terms. If the design vectors for attribute levels are coded so that part-worths across levels for each attribute sum to zero, then: A = $\{N_1, N_2, N_3, N_1 \times N_2, N_1 \times N_3, N_2 \times N_3, N_1 \times N_2 \times N_3\}$, $Q = 7$ size (Φ) = 128, and $K = 26$. Note that size (Φ) grows quickly with the number of attributes and levels, and that this growth constitutes the main challenge in both estimation and inference.

Let $g_s \in \Phi$ denote a set of $q_s \in \{0, 1, \ldots, Q\}$ relevant explanatory variables for subject s and $x_{sjt}^{(k_s)}$ denote a k_s–dimensional design vector given the set of relevant explanatory variables, g_s, for subject s, alternative j, and choice task t. Note that k_s depends upon the number of attribute levels, and that k_s does not equal g_s since all levels of an attribute enter into the design vector when the attribute is among relevant explanatory variables. Therefore, irrelevant explanatory variables, $g_s^c = (g_s \cap A)^c$, are ones that subject s are indifferent across *all* levels of the variables allowed in the conjoint study. Let $x_{st}^{(k_s)} = (x_{s1t}^{(k_s)}, \ldots, x_{s,J-1,t}^{(k_s)})'$ denote a $(J-1) \times k_s$ design matrix given g_s for subject s, and choice task t. Finally, utilities for subject s, alternative j, and choice task t are assumed to be Gaussian random variables:

$$\begin{aligned} u_{sjt} &= \beta_{s0} + x_{st}^{(k_s)'}\beta_s^{(k_s)} + \varepsilon_{sjt}, \ j = 1, \ldots, J-1, \text{and} \\ u_{hJt} &\equiv 0, \end{aligned} \tag{1}$$

where β_{s0} is a common intercept for subject s, $\beta_s^{(k_s)}$ is a regression coefficient vector for $x_{st}^{(k_s)}$, and ε_{sjt} is a Gaussian error structure. Note that only *relevant* explanatory variables, $x_{st}^{(k_s)}$, enter the utility function and therefore $x_{st}^{(k_s)}$ is treated as a *random* variable.

In order to capture heterogeneity, β_{s0} and $\beta_s^{(k_s)}$ are specified as normal random effects:

$$\beta_{s0} \sim N(\mu_0, \sigma_0^2), \ \text{for } \forall s,$$

$$\beta_{s,i}^{(k_s)} \sim N(\mu_l, \sigma_l^2), \text{for } \forall s \text{ and } i = 1, \dots, k_s, \tag{3}$$

where $N(\mu, \sigma^2)$ is a univariate normal distribution with mean μ and variance σ^2, and $l \in \{1, \ldots, K\}$ indicates that the i–th column of $x_{st}^{(k_s)}$, $x_{st,i}^{(k_s)}$, corresponds to the i–th element in R, R_l.

Then, choice probabilities follow the multinomial probit model:

$$\begin{aligned} p_{sjt} &= p(y_{st} = j \mid \beta_{s0}, \beta_s^{(k_s)}, x_{st}^{(k_s)}) \\ &= \int_{\dot{A}_{s1t}} \cdots \int_{\dot{A}_{s,J-1,t}} N_{J-1}(u_{st} \mid 1_{J-1}\beta_{s0} + x_{st}^{(k_s)}\beta_s^{(k_s)}, I_{J-1}) du_{st}, \end{aligned} \tag{4}$$

where us = $(u_{s1t}, \ldots, u_{s,J-1,t})'$ is a $(J - 1)$ – dimensional latent utility vector for subject s and choice task t, 1_{J-1} is a column vector of J – 1 ones and $N_m(a, B)$ denotes an m–variate normal distribution with mean vector a and variance-covariance matrix B. Note that we fix the covariance matrix of $\{\varepsilon_{sjt}\}$ to an identity matrix, since it is far from clear how to introduce a general correlation or covariance matrix for numerous potential product profiles which vary across choice tasks (for a full discussion of this issue, see Haaijer et al. 1998); it is well-known that this assumption is considerably less problematic when heterogeneity on regression coefficients is accounted for in a suitably general manner (cf., Orme 1998), as in (2) and (3). When choice $y_{ht} = j < J$, the interval $\dot{A}_{sit}$ is given by

$$\dot{A}_{sit} = \begin{cases} (0, \infty), & \text{if } i = j \\ (-\infty, u_{sjt}), & \text{if } i = 1, \ldots, j-1, j+1, \ldots, J-1. \end{cases} \tag{5}$$

Eq. (4) can be simplified by linearly transforming the $(J - 1)$-dimensional vector u_{st} using a $(J - 1) \times (J - 1)$ matrix $W_j = \{w_{mn}\}$:

$$\begin{aligned} &w_{jj} = -1, \\ &w_{ii} = 1 \text{ and } w_{ij} = -1, \text{ for } i = 1, \ldots, j-1, j+1, \ldots J-1, \end{aligned}$$

and all other elements of W_j equal to 0. The resulting $(J - 1)$-dimensional vector is therefore $\tilde{u}_{st} = W_j u_{st}$ and (4) can be rewritten as

$$p_{sjt} = \int_{-\infty}^{0} \cdots \int_{-\infty}^{0} N_{J-1}(\tilde{u}_{st} \mid W_j(1_{J-1}\beta_{s0} + x_{st}^{(k_s)}\beta_s^{(k_s)}, W_j W_j') d\tilde{u}_{st}. \tag{6}$$

For the “no choice” option, i.e., $y_{st} = J$, the matrix W_J must be the identity, $W_J = I_{J-1}$. This completes the formal model specification, and discussions on methods for its estimation using discrete choice data follow.

2.2 Estimation of Heterogeneous Relevant Explanatory Variables and Their Regression Coefficients

At the heart of the proposed model is a methodology for estimating relevant explanatory variables for each subject, g_s and their regression coefficients, $\beta_s^{(k_s)}$. Note that g_s determines the dimension of design vector, k_s, and the design matrix, $x_{st}^{(k_s)}$. In order to identify g_s, it is critical to explore its sample space, Φ; .this can be implemented by using an MCMC method. Popular methods such as the Gibbs sampling or the standard Metropolis-Hasting algorithm, however, are not applicable to this problem because they each require that the number of unknown quantities be fixed. But g_s – and, consequently, k_s $x_{st}^{(k_s)}$, and $\beta_s^{(k_s)}$– needs to be allowed to change across iterations. So, alternative sampling methods must be used.

Several approaches have been proposed in the context of variable selection for regression models (cf., George and McCulloch 1993; Smith and Kohn 1996). George and McCulloch (1993) first put priors on all regression coefficients of the full design vector given *all* candidate explanatory variables. They then tested if each of these regression coefficients was zero by introducing latent binary indicators: one for each regression coefficient. This approach, however, may not be efficient because the resulting combinatorial problem grows exponentially with the dimension of the design vector given all candidate explanatory variables. In addition, in order to compute integrated likelihoods for model comparison, their approach does not completely remove regression coefficients from the model even when their binary indicators have zero values. Their approach is implemented by setting the regression coefficients to values arbitrarily close to zero if their binary indicators possess zero values. The approaches of Brown and Vannucci (1998) and Smith and Kohn (1996) completely remove regression coefficients if their binary indicators have zero values, but they also suffer from a similar combinatorial problem. As stated earlier, successfully navigating the very high-dimensional space of Φ bedevils variable selection even when explanatory variable configurations are not individual-specific.

However, there is an MCMC algorithm that is applicable when the number of unknown quantities changes over iterations – Green's (1995) reversible jump Metropolis-Hastings algorithm. This study employs this algorithm to estimate subject-specific set of relevant explanatory variables, g_s and regression coefficients, $\beta_s^{(k_s)}$ Note that there are three parameters that must vary across iterations: (1) the number of relevant explanatory variables, q_s (equivalently k_s); (2) the set of relevant explanatory variables, g_s (equivalently $x_s^{(k_s)}$); and (3) the regression coefficients *given* the explanatory variable configuration, $\beta_s^{(k_s)}$. With the reversible jump Metropolis-Hastings algorithm, it is possible to explore possible configurations of explanatory variables for each subject, under some mild distributional assumptions. Specifically, we write:

$$p(q_s, x_s^{(k_s)}, \beta_s^{(k_s)}) \propto p(\beta_s^{(k_s)} \mid x_s^{(k_s)}, q_s, \mu, \Sigma) p(x_s^{(k_s)} \mid q_s) p(q_s),$$

where $\mu = \{\mu_l\}_{l=1}^{K}$ and $\Sigma = \{\sigma_l^2\}_{l=1}^{K}$. The expression $p(q_s)$ denotes a prior distribution for $q_s \in \{0, 1, \ldots, Q\}$. Note that we allow a null set for g_s when $q_s = 0$. Given q_s, a set of q_s explanatory variables, g_s, – or, equivalently, a $(J-1) \times k_s$ design matrix $x_{st}^{(k_s)}$ – is sampled from Φ. Finally, the estimates of $\beta_s^{(k_s)}$ can be obtained conditional on the design matrix, $x_{st}^{(k_s)}$.

The question remains as to how to efficiently explore q_s and g_s. Simple uniform sampling of q_s and g_s may be dramatically less efficient for large Q. To address this problem, a scheme using four possible moves for each subject is employed: (1) adding one more explanatory variable (Birth); (2) deleting one among the currently present explanatory variables (Death); (3) replacing one currently present explanatory variable by one of the non-present ones (Replacement); and (4) updating $\beta_s^{(k_s)}$ without any change in q_s and $x_s^{(k_s)}$ (Update). These four moves suffice to explore the space of configurations of q_s, $x_s^{(k_s)}$, and $\beta_s^{(k_s)}$. Technical details are given in Appendix.

2.3 Priors

To complete the model specification, the following priors are introduced:

$$\mu_0 \sim N(m_0, v_0), \ for \ \forall s, \tag{7}$$

$$\sigma_0^2 \sim IG(a_0, b_0),\ \ for\ \forall s, \tag{8}$$

$$\mu_l \sim N(m_l, v_l),\ for\ \forall s\ \text{ and }\ l = 1, \ldots, K, \tag{9}$$

$$\sigma_l^2 \sim IG(a_l, b_l),\ \ for\ \forall s\ \text{ and }\ l = 1, \ldots, K, \tag{10}$$

where the expression $IG(a, b)$ denotes an inverse gamma distribution with shape parameter a and scale parameter b. Finally, the prior distributions relating to the possible configurations of relevant explanatory variables are as follows.

1. The prior distribution for q_s, the *number* of explanatory variables, is

$$p(q_s) = Po(\lambda), \forall s, \tag{11}$$

where Po(λ) is a Poisson distribution with a specified mean λ, truncated at Q, the number of all candidate explanatory variables.

2. The prior distribution for the set of explanatory variables, $p(x^{(k_s)} \mid q_s)$, is a constant,

$$p(x_s^{(k_s)} \mid q_s) = \frac{1}{V},$$

where $v = \begin{pmatrix} Q \\ q_s \end{pmatrix}$ is the number of possible sets consisting of q_s explanatory variables among Q candidate ones.

3. EMPIRICAL APPLICATION

3.1 Dial-readout Bathroom Scale Conjoint Data

The proposed model was fitted to a choice-based conjoint data set for a small durable product, specifically dial-readout bathroom scales.[1)] In the data set, 184 students in a variety of courses each

1) The author would like to thank the Optimal Design Engineering Laboratory at the University of Michigan for making this data set publicly available. Further information can be found at: http://ode.engin.umich.edu.

Table 1. Attribute levels in the bathroom scale conjoint data set

Attributes	Levels
Weight capacity (C)	200lb, 250lb, 300lb, 350lb, 400lb
Platform aspect ratio (R; length/width)	0.75, 0.875, 1, 1.143, 1.333
Platform area (A)	100 sq. in., 100 sq. in., 120 sq. in., 130 sq. in., 140 sq. in.
Interval mark gap (G)	0.063 in., 0.094 in., 0.125 in., 0.156 in., 0.188 in.
Size of number (S)	0.75 in., 1 in., 1.25 in., 1.5 in., 1.75 in.
Price (P)	\$10, \$15, \$20, \$25, \$30

completed 50 on-line choice tasks. In each task, subjects were presented with three product profiles, described by six attributes, and a "no-choice" option (see Table 0). Product profiles were presented both in tabular form and graphically, so that subjects got a sense of the dimensions and appearance of each scale, and could readily indicate which, if any, they most preferred. Therefore, S = 184, J = 4, and T_s = 50 for $\forall s$.

For the simplicity of illustration and subsequent interpretation, only main effects for the six attributes were considered; the size of Φ was therefore 64 (= 2^6) and Q = 6 (see Table 2). Note that interaction terms can easily be handled in the proposed model, should researchers wish to; all algorithms would be identical, although a larger number of quantities would need to be sampled. In addition, the design vector given all six attributes, R, was defined so that part-worths *within* each attribute sum to zero for identification purposes; therefore, K = 24.

In addition, the data set was divided into two subsets: training data, $\bar{y}$, and hold-out data, $\tilde{y}$. The hold-out data were the last 10 choice tasks for each subject. Finally, the chosen values for priors were:

$$m_0 = 0, \upsilon_0 = 20, a_0 = 3, b_0 = 40,$$

$$m_l = 0, \upsilon_l = 20, a_l = 3, b_l = 40, l = 1, \dots, K, \text{ and}$$

$$\lambda = 6.$$

Table 2. Candidate relevant attribute configurations

No. of Attributes	Set of Attributes
0	{}
1	{C}, {R}, {A}, {G}, {S}, {P}
2	{C,R}, {C,A}, {C,G}, {C,S}, {C,P}, {R,A}, {R,G}, {R,S}, {R,P}, {A,G}, {A,S}. {A,P}, {G,S}, {G,P}, {S,P}
3	{C,R,A}, {C,R,G}, {C,R,S}, {CRP}, {C,A,G}, {C,A,S}, {C,A,P}, {C,G,S}, {C,G,P}, {C,S,P}, {R,A,G}, {R,A,S}, {R,A,P}, {R,G,S}, {R,G,P}, {R,S,P}, {A,G,S}, {A,G,P}, {A,S,P}, {G,S,P}
4	{C,R,A,G}, {C,R,A,S}, {C,R,A,P}, {C,R,G,S}, {C,R,G,P}, {C,R,S,P}, {C,A,G,S}, {C,A,G,P}, {C,A,S,P}, {C,G,S,P}, {R,A,G,S}, {R,A,G,P}, {R,A,S,P}, {R,G,S,P}, {A,G,S,P}
5	{C,R,A,G,S}, {C,R,A,G,P}, {C,R,A,S,P}, {C,R,G,S,P}, {C,A,G,S,P}, {R,A,G,S,P}
6	{C,R,A,G,S,P}

3.2 Model Comparison

Two models, the proposed model M and a base-line model M_0, were estimated to training data, $\bar{y}$. The model M_0 is the standard random effect probit model, wherein subjects are assumed to consider all attributes, i.e., g_s = {C, R, A, G, S, P} for $\forall s$. This model was estimated by allowing only the update move for $\beta_s^{(k_s)}$, after setting $q_s = Q$, $k_s = K$, and $x_{st}^{(k_s)} = R$ for $\forall s$. The burn-in period was 10,000 iterations for both models, and all inferences were based on the next 30,000 iterations. Figure 1 presents trace plots of the number of relevant attributes for a representative subject (s = 175); attribute configurations were apparently well mixed.

A decision regarding which of the models M and M_0 is preferred can be made by computing the Bayes factor, the ratio of two competing models' integrated likelihoods. A family of promising thermodynamic integration methods for computing integrated likelihoods have been discussed under the name of bridge and path sampling (Gelman and Meng 1998) and annealed importance sampling (Neal 2001). These thermodynamic integration methods have found to be among the most accurate methods for the computation of the integrated likelihoods. Consequently, this study

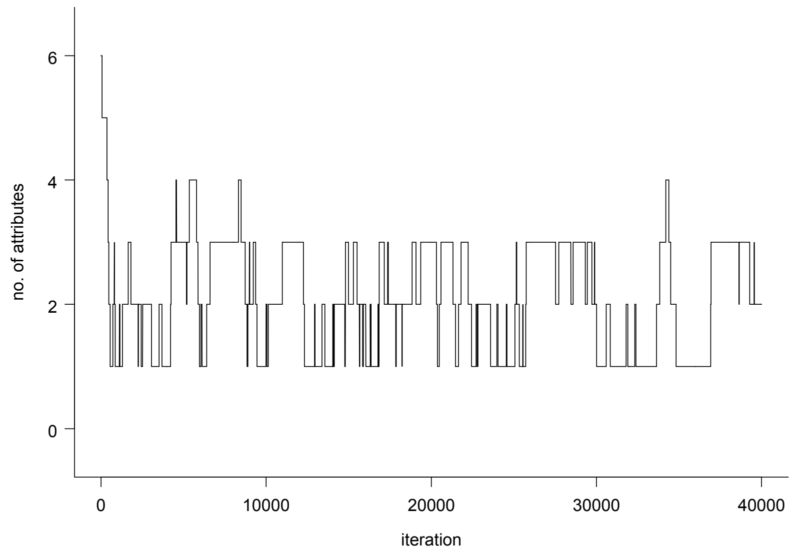

Figure 1. Trace plots: No. of relevant attributes for *s* = 175

used the annealed importance sampling method.

In the annealed importance sampling method, all unknown parameters of a given model, θ, are sampled sequentially from N consecutive annealed distributions; each distribution is formulated as

$$f_n(\theta) = f_M(\theta)^{\alpha_n} f_0(\theta)^{1-\alpha_n}, n = 1, \ldots, N, \tag{12}$$

where $0 = \alpha_1 < \alpha_2 < \ldots \alpha_N = 1$ are importance weights. For the computation of the integrated likelihood, $f_0(\theta)$ is set to be the prior distribution and $f_M(\theta)$ is set to be the product of prior and likelihood. Then, the estimate of the integrated likelihood becomes:

$$k = \frac{1}{M} \sum_{m=1}^{M} \sum_{n=2}^{N} f_n(\theta^{(n)})^{\alpha_n - \alpha_{n-1}}, \tag{13}$$

where M is the total number of iterations and $\theta^{(n)}$ is the value of θ at the n-th annealed step. Note that this method does *not* require an importance sampling distribution, which presents the practical problem of needing to be close enough to the true posterior distribution. It is also critical that $\{\alpha_n\}_{n=1}^{N}$ must be spaced smoothly

Table 4. Model comparison result

Model	Log of Integrated Likelihood	Hold-out data	
		Posterior Predictive Likelihood	Hit Rate
M_0	–7361.67	–3766.0	60.07%
M	–6641.89	–3291.7	61.26%

over an interval [0, 1] so that $\alpha_n - \alpha_{n-1}$ goes approximately in inverse proportion to n, and that $\theta^{(n)}$ must be sampled from $f_M(\theta)^{\alpha_n} f_0(\theta)$ rather than from $f_M(\theta)f_0(\theta)$. We used 201 α_n, which were spaced geometrically from 0 to 1.

Table 4 presents estimates of integrated likelihoods for both models obtained by the annealed importance sampling method. The resulting logarithm of the Bayes factor was –719.78. This is fairly strong evidence supporting heterogeneity in relevant attribute configurations across subjects.

In addition, prediction tasks for hold-out data, $\tilde{y}$, were conducted. Specifically, two quantities were computed: (1) posterior predictive likelihood, $p(\tilde{y} \mid \bar{y}) = \int p(\tilde{y} \mid \theta) p(\theta \mid \bar{y}) d\theta$, and (2) the proportion of correctly predicted choices for the given model (the so-called 'hit rate'). The posterior predictive likelihood is an average of conditional predictions for the hold-out data over the posterior distribution of θ obtained from the training data across MCMC iterations. The hit rate was obtained by first sampling latent utilities for each choice observation in $\tilde{y}$ given θ in each MCMC iteration, computing the proportion of all choices for which the actual product profile chosen equals the product profile of the highest utility value in each MCMC iteration, and then taking an average of the proportions across iterations. As given in Table 4, the proposed model offered better predictive performance for both tasks. Both training and hold-out data favor the proposed model over the baseline model. Based on this evidence, results of the proposed model, M, are further elaborated.

3.3 Estimation Result of the Proposed Model

3.3.1 Estimates of μ_0 and σ_0^2

The estimates of μ_0 and σ_0^2 were –1.2084 and 7.0778 with standard deviations 0.2071 and 0.9418, respectively. The [5 percentile, 95 percentile] intervals of μ_0 and σ_0^2 were [–1.5481, –0.8674] and

[5.7282, 8.7502], respectively, implying statistical difference from zero in the first case, and also that subjects tended to have weak intrinsic preferences towards scale products. The estimates of μ_0 and σ_0^2 under the base-line model, M_0, were –1.0560 and 6.4747 with standard deviations 0.2008 and 0.9602, respectively. A statistic to test differences in estimates of μ_0 and σ_0^2 across two models is w = (differences in posterior means)/ $\sqrt{\text{summation of posterior variance}}$; these were –0.5283 and 0.4484, respectively, suggesting that estimates of μ_0 and σ_0^2 across two models were essentially the same.

3.3.2 Estimates of Subject-specific Relevant Explanatory Variables

The presented model allows for posterior inference regarding relevant attribute configurations for each subject. In order to do so, researchers need to only store subject-specific variable configurations across MCMC iterations. Figure 5 depicts the histograms of relevant attribute configurations for subjects s = 2 and 139, who were fairly representative of the sample. For both subjects, the full set of candidate attributes, {C, R, A, G, S, P}, was not the most preferred among possible 64 subsets of six available attributes. In fact, the posterior proportions for the full set were 0.9% and 0% for s = 2 and 139, respectively. The modal relevant attribute configurations were {R, G, P} and {S} for s = 2 and 139, respectively. It is also noticeable that the number size attribute was included in all non-null sets of relevant attributes for s = 139, implying that the attribute is among most important attributes for the subject, irrespective of the coefficient associated with it. Similarly, subject s = 2 appeared to consider price to a greater extent than other attributes.

Figure 2 suggests that there is substantial variation in the number of attributes considered by different subjects. Specifically, many subjects might not consider the full set of six. To examine this further, the modal relevant attribute configuration was found for each subject by examining proportions of attribute configurations across iterations. Figure 3 presents a histogram of the size of such modal attribute configuration sets across subjects. Among the 184 subjects, 68 (37%) were found to consider all six attributes. The remaining 116 subjects (63%), however, evidently considered subsets of the six available attributes. In particular, five subjects (2.7%) were found to consider *none* of six attributes; it may well be that the task was unmotivating for them, or they lacked any relevant

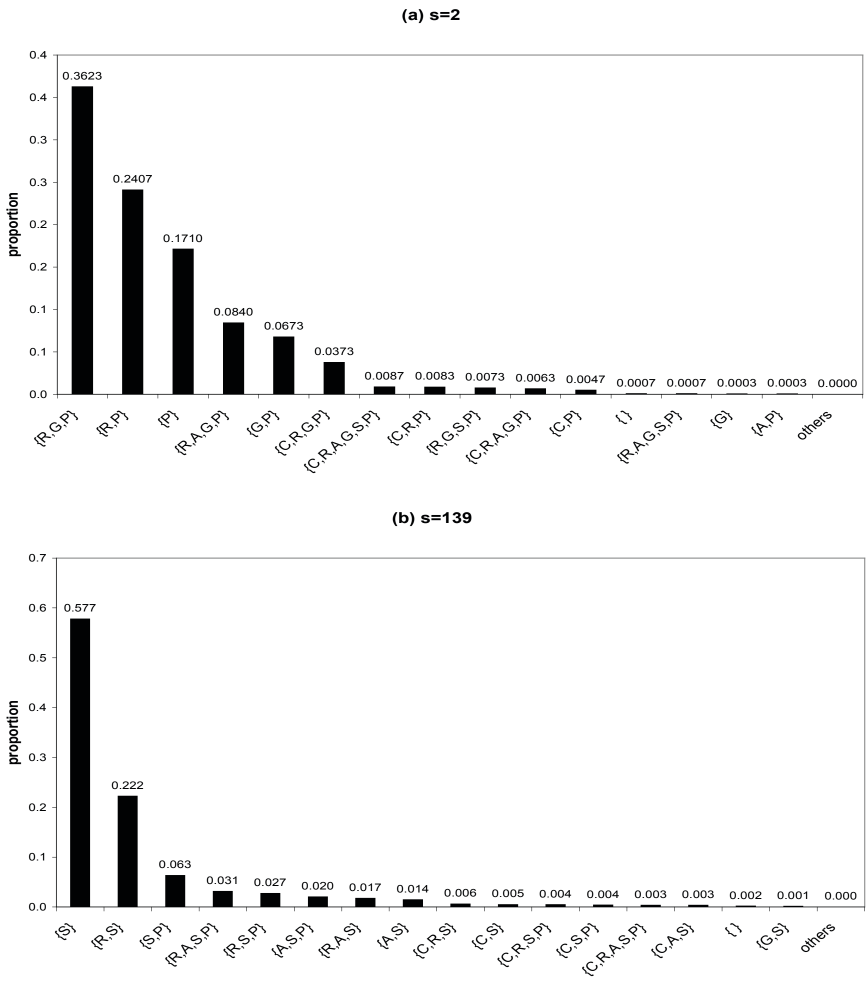

Figure 2. Histogram of relevant attribute configurations

experience in the product class. Still, it is reassuring that so small a group of subjects fell into this no-variable category. Figure 7 also gives a histogram of the number of subjects whose modal relevant attribute configurations include a given attribute. The proportions ranged from 58.7% to 82.6% across the six attributes. The interval mark gap attribute – which determines readability – was found to be most frequently considered while the number size attribute was least frequently considered. In summary, Figures 3 and 4 together offer fairly detailed evidence that there exists a substantial variation

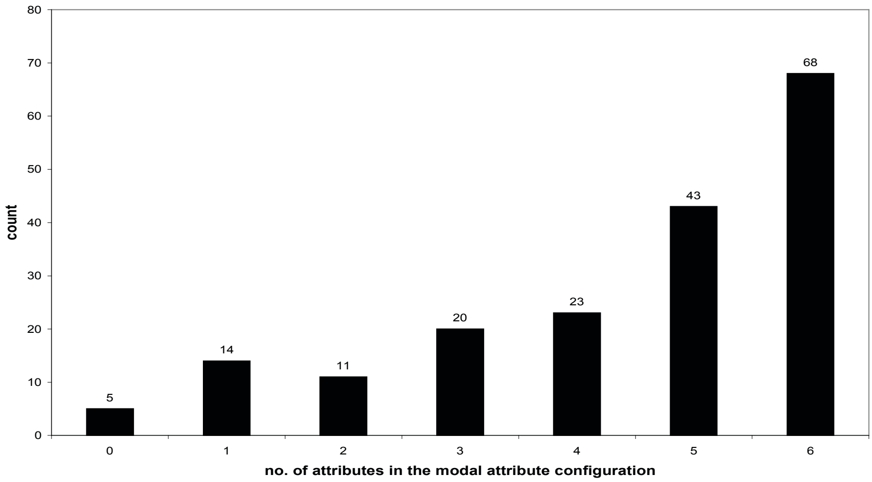

Figure 3. Histogram of the size of the modal relevant attribute configurations

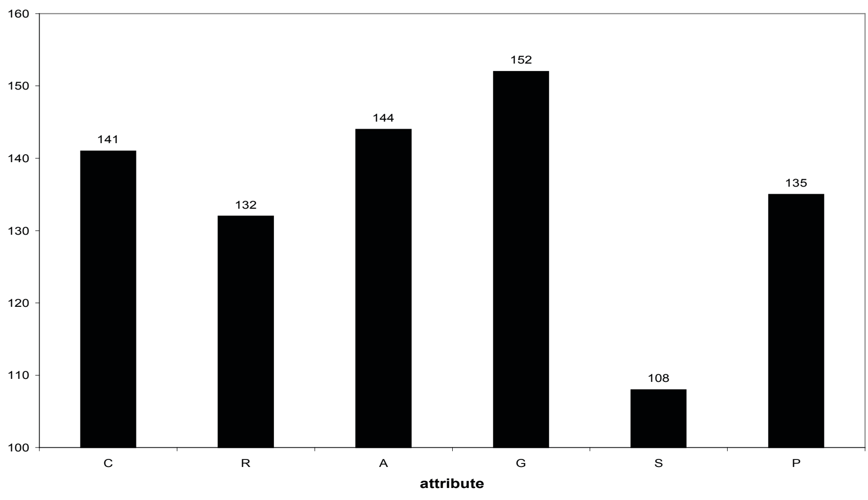

Figure 4. Histogram of the number of subjects whose modal relevant attribute configurations include a given attribute

in terms of relevant attribute configurations across subjects.

Finally, Figure 5 presents a histogram of the modal relevant attribute configurations across subjects. The most preferred attribute configuration among those who do not consider all six

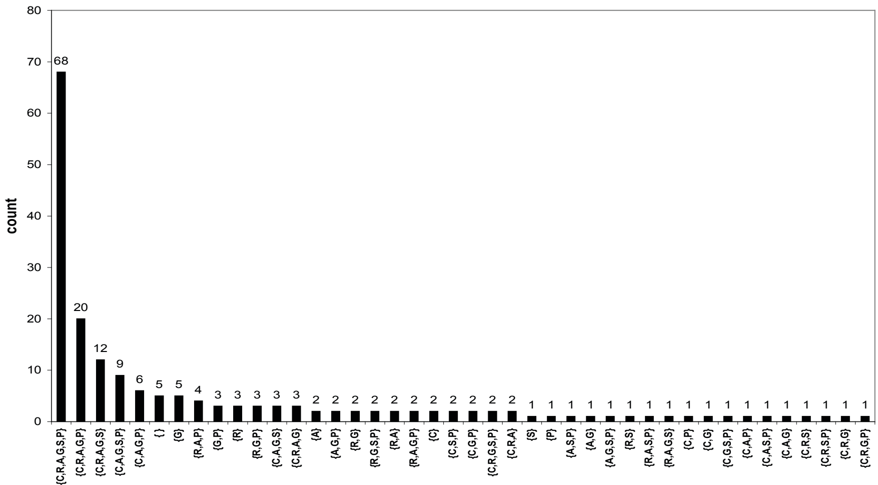

Figure 5. Histogram of the modal relevant attribute configurations across subjects

attributes was {C, R, A, G, P}, the configuration lacking the number size attribute (S). Since S was the least common attribute across subjects' modal attribute configurations, this would be expected. It is interesting to note that the fact that five-element sets are most common among all subsets of six attributes (as shown in Figure 3) does *not* imply that all five-element sets are very likely. In fact, only three of the six possible five-element sets were somewhat common, and two of them were not indicated for even a single subject. Even the empty set and the singleton containing just {G} were more common than several of the five-element subsets. The pattern of subject-specific modal attribute configurations are quite a bit more complex than a multinomial model built, for example, on the baseline frequencies of each of the attributes' posterior proportions (recall, 58.7% to 82.6%) alone would indicate. In short, a good deal of heterogeneity is indicated, and it is not of an easily predictable sort.

In addition, a type of face validity for the subject-specific relevant attribute configurations was checked. The original data set also included information on subjects' weights, heights, and ages. The first of these is obviously relevant to scale choice, and one might expect that, when all things equal, heavier subjects would prefer greater weight capacities. The 184 subjects therefore were divided into two groups: those who do not have the weight attribute in their modal attribute configurations (A), and those who do (B). The size of groups A and B were 43 and 141, respectively. The average weights were 141.3lb and 161.5lb for groups A and B, respectively, a strong difference (t_{182} = –3.31). While this is an ad hoc test, two points are worth noting: weight information was not used anywhere in the experiment or the estimation; and nowhere was the size of the associated coefficient for weight capacity used, only whether it appeared in the modal attribute configuration at all.

3.3.3 Estimates of Subject-specific Part-worths

There are two ways to obtain subject-specific part-worths given subject-specific attribute configurations: (1) obtain estimates of part-worths conditional on the most preferred attribute configurations under M; and (2) take averages of part-worths of attribute levels across MCMC iterations under M. In the former approach, the estimate of part-worth associated with the most preferred attribute configuration for a given subject can be obtained from MCMC iterations by taking averages of part-worths across iterations where simulated attribute configuration equals the most preferred attribute configuration. By doing so, the estimated part-worths for irrelevant attribute levels are set to zeros. However, we follow the latter approach since the former does not take the uncertainty on relevant attribute configurations into account. Note that the part-worths of irrelevant attribute levels at a given MCMC iteration are all set to zero by definition.

In addition, given subjects' modal attribute configuration,.we divided all subjects into two groups: (1) those who have an attribute i (i = 1, ..., 6) in their modal attribute configurations ($S_i^{(1)}$) and (2) those who do not have an attribute i (i = 1, ..., 6) in their modal attribute configuration ($S_i^{(0)}$). The sizes of $S_1^{(1)}$, ..., $S_6^{(1)}$ were given in Figure 4.

Figure 6 gives plots of subject-specific part-worths for six randomly selected subjects, one for each attribute, as representative examples. In the figure, the left and the right columns consist of subjects from $S_i^{(1)}$ and $S_i^{(0)}$, respectively. Solid and thin lines represent estimated part-worths under model M and M_0, respectively. Dotted lines indicate the [estimate ± 1.64× (std. dev.)] intervals. In addition, w, the statistic to examine whether two estimated part-worts at a given attribute level are statistically meaningfully different from each other, are given in the figure.

For those in $S_i^{(1)}$, the estimated part-worths under M and M_0 were essentially the same in terms of magnitude and curve shape. However, it is noticeable that for subjects in the right column, those from $S_i^{(0)}$, there existed statistically meaningful differences in the estimated part-worths across two models. For those in $S_i^{(0)}$, the part-worths of attribute levels under M are essentially zeros, but M_0 exhibits fairly non-linear part-worth curves. Note that at some attribute levels, the part-worth estimates under M_0 are statistically meaningfully different from zeros. For example, for s = 43, the part-worth for the third level of the "platform aspect ratio attribute", which was different from zero, was over-estimated by M_0 compared to M. Note that the part-worth curves for those in $S_i^{(0)}$ under M are flat at zero, since the attribute i was not among attributes considered. Clearly, for those in $S_i^{(0)}$, the part-worth estimates under M_0 tend to almost suspiciously vary around zero across attribute levels, and to have pronounced non-linearities in their part-worth curves.

Figure 6 suggests that part-worths estimated under M_0 can be biased for subjects in $S_i^{(0)}$ if heterogeneity in relevant attribute configurations is not accounted for. To examine this further, we computed the proportion of subjects whose estimated part-worths under M differ statistically meaningfully from those under M_0 for each attribute level across $S_i^{(0)}$ and $S_i^{(1)}$. Figure 7 gives a plot of such proportions across attribute levels. For $S_i^{(1)}$, estimated subject-specific part-worths across the two models appeared to be the same for all attribute levels, except the first and fifth levels of the "platform aspect ratio" attribute, the first level of the "platform area attribute", and the second level of the "price" attribute. For $S_i^{(0)}$, there were, however, a sizeable portion of subjects whose estimated part-worths differed across the two models for all attribute levels (except the fifth level of the " interval mark gap" attribute). This pattern of results, therefore, seems to indicate that part-worth estimates under M_0 can be more biased for the subjects in $S_i^{(0)}$.

3.3.4 Estimates of Aggregate-level Part-worths

In the model M, we classified all subjects into two groups for each attribute: $S_i^{(0)}$ and $S_i^{(1)}$, i = 1, ..., 6. In this section, we examine the effect of the existence of these two groups on aggregate-level part-worths. As a representative example, let us focus on the number size

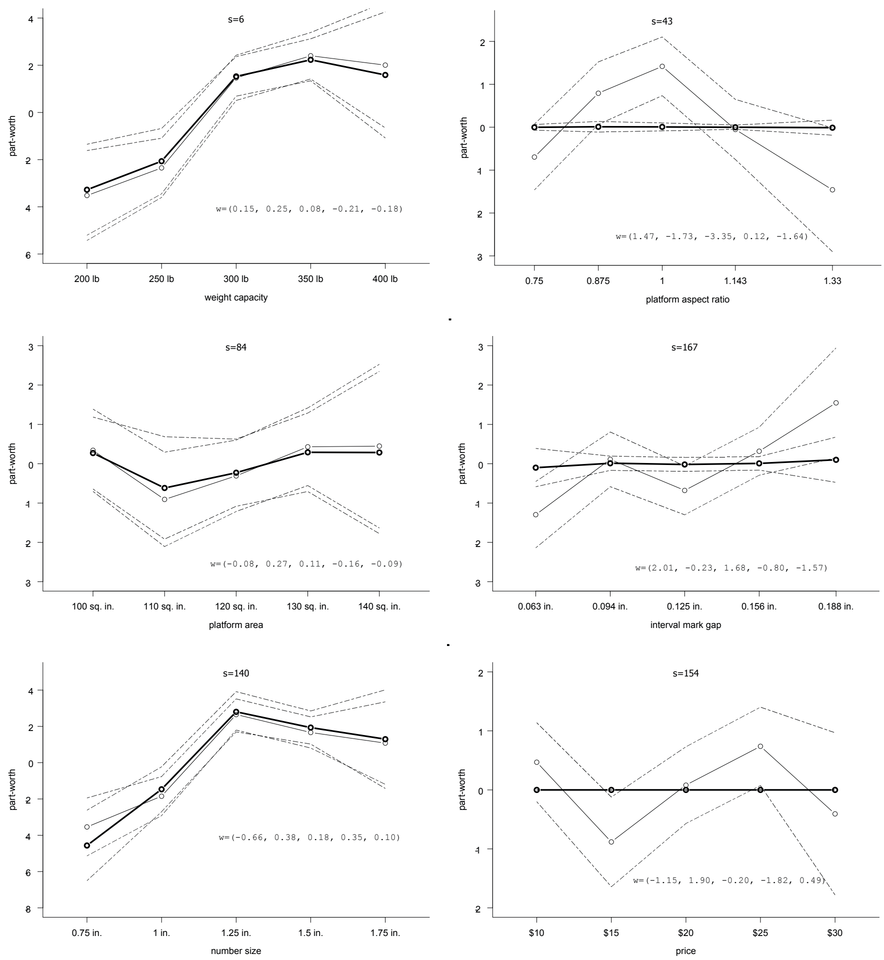

Figure 6. Plots of subject-specific part-worths

attribute (i = 5), although the ensuing procedure and discussions are equally valid for the other five attributes.

First, we computed the aggregate part-worth curves by taking averages of subjects' part-worths across iterations. Figure 7 gives plots of the estimates and the [estimate ± (1.64 × std.dev.)] intervals of the aggregate part-worth curves for the number size attribute across (a) all subjects under M_0, (b) the subjects in $S_5^{(0)}$ under M, and (c) the subjects in $S_5^{(1)}$ under M. Statistically meaningful differences were found for all three pairs of $S_5^{(0)}$, $S_5^{(1)}$, and M_0. implying that the

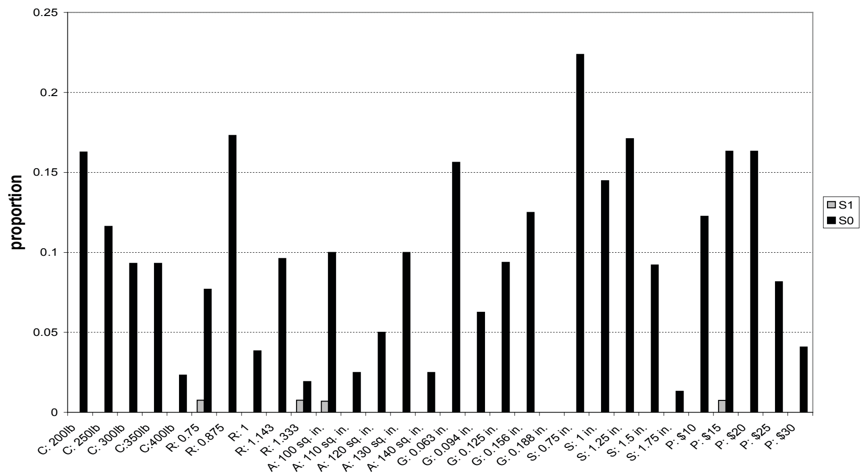

Figure 7. Proportion of subjects whose estimated part-worths of a given attribute level are statistically meaningfully different across M and M_0

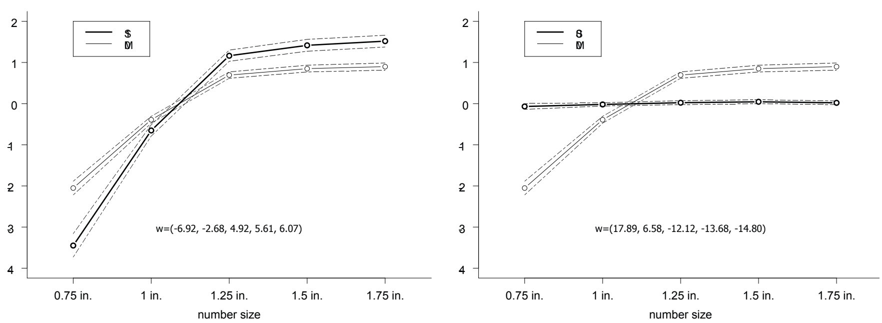

Figure 8. Plots of aggregate part-worth curves for the number size attribute

aggregate part-worth curve for $S_5^{(0)}$ was different from that for $S_5^{(1)}$ and that the aggregate part-worth curve under M_0 can be biased for both $S_5^{(0)}$ and $S_5^{(1)}$. Furthermore, it is clear from Figure 8 that aggregate part-worths under M_0 tended to under-estimate those for $S_5^{(1)}$ and over-estimate those for $S_5^{(0)}$ in magnitude. Let us examine this curious fact further.

Figure 9 gives probability density plots of part-worths for the first and the fourth levels of the number size attribute, 0.75 in. and 1.5 in., across subjects for M_0, $S_5^{(0)}$ and $S_5^{(1)}$. Note that the densities for

$S_5^{(0)}$ and $S_5^{(1)}$ were normalized so that the integrals of two density functions have value 1. The figure clearly shows that, under M, there are two distinct modes: (i) $S_5^{(0)}$ shows highly concentrated distributions with modal value zero for both levels, and (ii) $S_5^{(1)}$ shows distributions deviating somewhat from zero, with modal values –4.27 and 2.05 for 0.75 in. and 1.5 in., respectively. One can also see some uncertainty regarding inferences about grouping subjects into the two classes, $S_5^{(0)}$ and $S_5^{(1)}$, since the probability densities for these two groups overlap one another about zero. Note that part-worths can take positive or negative values for subjects in $S_5^{(1)}$. In order to examine the degree of uncertainty, we first found the [5 percentile, 95 percentile] intervals of part-worths across subjects in $S_5^{(0)}$; the resulting intervals were [–0.4515, 0.0001] and [–0.0043, 0.2621] for 0.75 in. and 1.5 in., respectively. Then, we computed the proportion of subjects in $S_5^{(1)}$ whose part-worths are located within these intervals. The computed proportions were 2, 2% and 2.7% for 0.75 in. and 1.5 in., respectively, suggesting that the distributions of part-worths between $S_5^{(0)}$ and $S_5^{(1)}$ are indeed different from one another, and further that all subjects were fairly sharply classified into two groups under M.

In addition, it is apparent that the modes of the distributions of part-worths under M_0 are located between the modes of the distributions for $S_5^{(0)}$ and $S_5^{(1)}$. This is consistent with the previously-discussed under- and over-estimation phenomenon under M_0, in terms of part-worth magnitudes. Since the proposed model M allows for two mixing components, one for $S_5^{(0)}$ and the other for $S_5^{(1)}$, we can compute the weighted sample mean and standard deviation for the part-worth of the l-th level of attribute i, across all subjects under M, μ_{il} and σ_{il}, as follows:

$$\mu_{il} = \pi_i^{(0)}\mu_{il}^{(0)} + (1 - \pi_i^{(0)})\mu_{il}^{(1)}, \tag{14}$$

$$\sigma_{il} = \sqrt{\left(\pi_i^{(0)}\right)^2 \left(\sigma_{il}^{(0)}\right)^2 + \left(1 - \pi_i^{(0)}\right)^2 \left(\sigma_{il}^{(1)}\right)^2}, \tag{15}$$

Here $\pi_i^{(0)} \in [0, 1]$ is the proportion of subjects in $S_i^{(0)}$, and $\mu_{il}^{(n)}$ and $\sigma_{il}^{(n)}$ are mean and standard deviation of part-worths of the l-th level of an attribute i across subjects in $S_i^{(n)}$, n = 0, 1. For consistency, we computed the weighted sample means and standard deviations of the part-worths for 0.75 in. and 1.5 in. of the number size attribute,

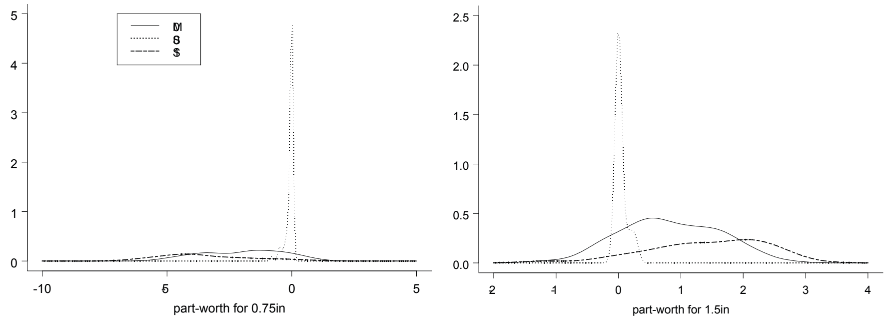

Figure 9. Density plots of part-worths for 0.75in and 1.5in levels of the number size attribute across subjects

Table 4. Summary of subjects' part-worth distribution for the first and the fourth levels of the number size attribute

Sample	0.75 in.	1.5 in.
All subjects under M_0	–1.9776* (1.6216)**	0.8226 (0.7898)
Subjects in $S_5^{(0)}$ under M	–0.0695 (0.1519)	0.0456 (0.1142)
Subjects in $S_5^{(1)}$ under M	–3.4476 (1.8063)	1.4170 (0.9048)

Note: * mean; ** standard deviation

μ_{51}, μ_{54} σ_{51} and σ_{54}. The proportion of $S_5^{(0)}$, $\pi_5^{(0)}$, was 0.41 (= 76/184). The sample means and standard deviations across different samples are given in Table 9; These values result in μ_{51} = –2.0523, μ_{54} = 0.8506 σ_{51} = 1.0621 and σ_{54} = 0.5542.

First, it is noticeable that, for both levels, the sample standard deviations of part-worths under M_0 are somewhat larger than the weighted sample standard deviations of part-worths under M. This is not surprising since M_0 allows part-worths for subjects in $S_5^{(0)}$ to be deviated from zero. To check this further, for all attribute levels, we examined differences between sample standard deviations of part-worths under M_0 and weighted sample standard deviations under M; the differences ranged from 0.0641 to 10.0401, showing that the sample standard deviations of part-worths under M_0 were 10.2% 12664.5% larger than the weighted sample standard deviations of part-worths under M. This result implies that the sample standard deviations of part-worths under M_0 may be systematically over-estimated since M_0 ignores the existence of $S_i^{(0)}$.

In addition, the weighted sample means of part-worths under

M were found to be essentially equal to the sample means of part-worths under M_0. For example, for the 0.75 in. level, the difference between the sample mean under M_0 and the weighted sample mean under *M* was 0.0747, with resulting statistic $w = 0.04$. By the same token, we examined the difference between the sample mean under M_0 and the weighted sample mean under *M* for each of 30 attribute levels; the difference ranged from –0.5153 to 0.3098, and the resulting statistic w ranged from –0.15 to 0.15. Therefore, the weighted sample means under *M* and the sample means under M_0 were essentially identical for all 30 attribute levels.

Taken together, these examinations suggest that ignoring the existence of $S_i^{(0)}$ leads to under-estimation of aggregate part-worths for subjects in $S_i^{(1)}$ and over-estimation of aggregate part-worths for subjects in $S_i^{(0)}$, since the aggregate part-worths under M_0 are essentially the weighted average of aggregate part-worths across $S_i^{(0)}$ and $S_i^{(1)}$. As (14) implies, the under-estimation bias for the aggregate part-worths for $S_i^{(1)}$ is likely to become ever more severe as $\pi_i^{(0)}$ increases. Therefore, when M_0 is fitted to all subjects *without* controlling heterogeneity in relevant attribute configurations (even though there exists a sizeable proportion of subjects who do not consider all attribute), the estimate of aggregate-level part-worths under M_0 can be biased for both $S_i^{(1)}$ and $S_i^{(0)}$, since they were compromised by the existence of $S_i^{(0)}$.

3.3.5 Inferences About Optimal Product Design

Figure 6 implies that marketers may reach different conclusions about optimal product designs depending on which of the two models, *M* and M_0, is followed. In particular, for those in $S_i^{(0)}$, *M* gives flat part-worth curves (at zero), while M_0 produces part-worth curves varying around zero. In order to find optimal product designs after taking subjects' relevant attributes into account, we checked whether part-worths for levels of irrelevant attributes are zero by examining whether zero values are located within the [estimate ± 1.64 × std.dev.] intervals of part-worths for all five levels for an attribute i ($i = 1, \ldots, 6$) for each subject in $S_i^{(0)}$; all subjects in $S_i^{(0)}$ were found to have zero part-worths for all levels of the attribute, suggesting that the part-worths of all levels of the attribute can safely be set to zero for all subjects in $S_i^{(0)}$. After setting the part-worths of all levels of attribute i to zero for subjects in $S_i^{(0)}$, we identified the optimal scale product for each subject. Similarly, under M_0, we identified the optimal scale product for each subject given the estimated subject-specific part-worths. Note that we do this in the absence of a cost model for the producer – which would typically be required to trade-off demand with the various levels of price and other attributes – and instead focus on which set of attribute levels is most preferred by each consumer.

As an example, let us consider subject s = 20. For the subject, price was never included in attribute configurations across MCMC iterations and his/her modal attribute configuration was {*C*, *R*, *A*, *G*, *S*}. Note that there were five 'optimal' product designs for s = 20 under M: a combination of {300lb, 1.143, 110 sq. in., 0.188 in., 1.25 in.} for non-price attributes and each of all five price levels. The optimal product design for s = 20 under M_0 was {300lb, 1.143, 110 sq. in., 0.188 in., 1.25 in., \$20}, which is one of five optimal product designs identified by M. Note that, for M_0, part-worths across price levels for s = 20 were (0.45, 0.70, 1.02, –0.92, –1.25).

Similarly, we identified optimal product designs for all subjects and found that there was a great deal of heterogeneity in terms of optimal product designs across subjects. In M_0, among 184 optimal product designs, one for each subject, only four product designs were most preferred by two subjects and others were most preferred by only one subject. The four modal product designs were {350lb, 0.75, 140 sq. in., 0.125in, 1.75in, \$10}, {400lb, 0.875, 100 sq. in., 0.188in, 1.25in, \$15}, {400lb, 0.875, 140 sq. in., 0.125in, 1.75in, \$10} and {400lb, 1, 120 sq. in., 0.094in, 1.75in, \$10}. In M, there were 15,625 different product designs across subjects, out of a possible 5^6 = 15,625 designs. Note that some subjects could have multiple optimal product designs under M if they do not consider all six attributes. Among 15,626 product designs, the modal product design was {250lb, 1, 100 sq. in., 0.156 in., 1.75 in., \$10}, which was most preferred by 18 subjects. The main point of these comparisons is this: if a marketer chooses an optimal product design by examining modal values across subjects, the two models lead to very different answers. To further support this point, we calculated the number of subjects whose identified optimal product designs under M_0 were not among those under M; the resulting count was 63 (34.2%) again suggesting that inferences about optimal product designs differ substantially across M and M_0.

As mentioned above, thus far optimal product designs were those maximizing subjects' utilities, not those reflecting the firm's overall objectives, usually some form of profitability, sales or market share. Clearly, for subject s = 20, {300lb, 1.143, 110 sq. in., 0.188 in., 1.25 in., \$20} cannot be optimal in terms of profit for the firm: since s = 20 was indifferent across a given price range between \$10 and \$30, the firm could charge the subject \$30 rather than \$20 Since we do not have cost information, we examined revenues given optimal product designs across subjects. Given subject-specific optimal product designs under M_0, the total revenue across the 184 subjects was \$2,565. The total revenue for subjects in $S_6^{(1)}$ given optimal product designs under M was \$1,695. For subjects in $S_6^{(0)}$, we assumed that \$30 will be charged to those who do not have the price attribute in their modal attribute configurations and the total revenue from them was \$1,470. The resulting total revenue from $S_6^{(0)}$ and $S_6^{(1)}$ together was therefore \$3,165, which is 23.4% greater than that suggested under M_0. Such a difference is non-trivial. We must caution that our results do not suggest that subjects like s = 20 were totally indifferent to *all* prices greater than \$30, just those used in the conjoint study. Indeed, we believe a fertile avenue for future research is using the method presented here to help identify subject-specific relevant *ranges*, ones which are relevant for each subject for every attribute in the study. Although the levels used in the scale-design study were chosen to cover over 90% of the scales on the market for each attribute, apparently a goodly proportion of subjects found themselves indifferent across those ranges for every one of the attributes. While this may be surprising for price, one could argue that even the highest level, \$30, was considered reasonable for a small durable, and some subjects simply did not respond especially strongly to reductions from this level.

4. CONCLUSION

Previous studies have, for reasons of tractability and parsimony, presumed that all subjects in a conjoint experiment consider each of the presented attributes in enacting trade-offs across product profiles. This assumption, though attractive and expedient, lacks strong theoretical and empirical justification; consider, for example, that wealthy subjects may be completely indifferent across prices for certain inexpensive products, or that PC power users may focus almost entirely on performance attributes and ignore others (at least within the ranges on offer in the marketplace). Therefore, the possibility remains that some subjects may consider a subset of all attributes, and that these considered attributes may substantially vary across subjects. This is closely related to one of the unsolved issues in the conjoint model literature, the so-called *Model Specification* problem: which attributes are relevant to observed overall attractiveness scores for product profiles, whether they be typically measured by discrete choices, ratings or ranking orders, for each subject? Previous studies have been silent on the model specification issue, implicitly taking part-worths of irrelevant attributes for trade-off evaluation tasks be close to zero. This paper presented a comprehensive methodology to address the model specification issue, explicitly identifying relevant attributes for each subject.

The proposed model, as fitted to choice data on dial-readout bathroom scales, performed better than the traditional random-parameters choice-based conjoint model in both training and prediction data. Our major findings can be summarized as follows:

1. Evidence was strong that a sizeable group of subjects in the experiment considered a proper subset of all attributes. Among 184 subjects, only 68 (37%) were found to consider all six attributes; the remaining 116 subjects (63%) evidently considered a smaller number of attributes, deeming at least one of them essentially irrelevant.
2. There was a great deal of heterogeneity in attributes deemed relevant across subjects: the proportion of subjects who did not consider a given attribute ranged from 17.4% to 41.3% across the six attributes. Moreover, the 116 subjects who appeared to consider fewer than six attributes exhibited a great deal of variation in which were relevant to them.
3. For those who do consider a given attribute, estimated attribute level part-worths were essentially identical for the proposed model and the traditional random-parameters conjoint model.
4. For those who do *not* consider an attribute, we found a statistically meaningful difference in part-worths under the traditional and the proposed model. In particular, the traditional model yielded part-worth curves hovering suspiciously around zero across attribute levels.
5. When the traditional model is fitted to all subjects without

controlling heterogeneity in relevant attribute configurations – even though there exists a sizeable proportion of subjects who consider subsets of all attributes – the estimate of aggregate-level part-worths for a given attribute under the traditional model tends to under-estimate average part-worths of the attribute across those who consider the attribute; this comes about because the aggregate-level part-worths are apparently compromised by the existence of a group of subjects who do not consider that attribute. Similarly, for those who do not consider the attribute, the traditional conjoint model is likely to over-estimate part-worths. This over- and under-estimation phenomenon resulted from the fact that the aggregate part-worths under M_0 would merely be weighted averages of part-worths across these two distinct groups of subjects.

6. Incorporating the possibility that every subject may not consider all attributes led to quite a substantial difference in inferences regarding optimal product designs at both the individual- and the aggregate-level. This difference could well be non-trivial in terms of revenue (or, presumably, profit) impact.

The proposed model was developed in a choice-based conjoint analysis setting. However, the general methodology underlying it can easily be applied to other measurement methods used conjoint studies, such as rating, pair-wise comparison and ranking, with appropriate changes in likelihoods (cf., Marshall and Bradlow 2002). For example, for the ranking data, the likelihood must be changed to the ordered probit model (Albert and Chib 1993) but the reversible Metropolis-Hastings algorithm given in Appendix is not changed. The proposed model was also designed to identify relevant explanatory variables among *a priori* fixed attributes. Therefore, if the number of attributes and levels is not unmanageable, the proposed model is likely to be relatively efficient, at least compared with alternative methods of sampling the space of possible variable configurations.

Although we have not investigated this in detail, we hope that the model might be applied to a variety of conjoint data sets, and that some consensus emerges regarding its efficiency in real-world settings. When the number of attributes and levels is rather large, the proposed model may not be efficient, even though it is still applicable if subjects each engage in a sufficiently large number of choice tasks. In this case, it may be wise to enact an informal procedure which allows researchers to customize experiments for each subject in order to find a set of key attributes among a large number of candidate attributes for each subject; indeed such methods are not uncommon in commercial conjoint programs, particularly those using adaptive methods. The methodology presented in this paper may serve as a starting point for additional future research along those lines.

REFERENCES

Akaike, H. (1974), "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, 19, 716-723.

Albert, J. and S. Chib (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88, 422, 669-679.

Andrieu, A., P. M. Djurić, and A. Doucet (2001), "Model Selection by MCMC Computation," *Signal Processing*, 81, 19-37.

Bernardo, J. M. and A. F. M. Smith (1994), *Bayesian Theory*, Chichester: Wiley.

Brown, P. J., M. Vannucci and T. Fearn (1998), "Multivariate Bayesian Variable Selection and Prediction." *Journal of the Royal Statistical Society*. Series B, 60, 3, 627-641.

Gelman, A. and X.-L. Meng (1998), "Simulating Normalizing Constants: From Importance Sampling to Bridge Sampling to Path Sampling," *Statistical Science*, 13, 163-185.

George, E. and R. E. McCulloch (1993), "Variable Selection via Gibbs Sampling," *Journal of the American Statistical Association*, 88, 423, 881-889.

Green, P. E., A. M. Krieger, and Y. J. Wind (2001), "Thirty Years of Conjoint Analysis: Reflections and Prospects," *Interfaces*, 31:3, Part 2 of 2, S56-S73.

Haaijer, M.E., M. Wedel, M., Vriens and T.J. Wansbeek (1998), "Utility Covariances and Context Effects in Conjoint MNP Models," *Marketing Science*, 17, 236-252.

Laud, P. W. and J. G. Ibrahim (1995), "Predictive Model Selection," *Journal of the Royal Statistical Society*. Series B, 57, 247-262.

Lindley, D. V. (1968), "The Choice of Variables in Multuple Regression (with Discussion)," *Journal of the Royal Statistical Society*. Series B, 30, 31-66.

Mallows, C. L. (1973), "Some Comments on C_p," *Technometrics*, 15, 661-675.

Marshall, P. and E. T. Bradlow (2002), "A Unified Approach to Conjoint Analysis Models," *Journal of the American Statistical Association*, 97, 459, 674-682.

Mitchell, T. J. and J. J. Beauchamp (1988), "Bayesian Variable Selection in Linear Regression (with discussions)," *Journal of the American Statistical Association*, 83, 404, 1023-1032.

Neal, R. M. (2001), "Annealed Importance Sampling," *Statistics and Computing*, 11, 125-139.

______ (2003), "Slice sampling" (with discussion), *The Annals of Statistics*, 31, 705-767.

Orme, Bryan K. (1998), "The Benefits of Accounting for Respondent Heterogeneity in Choice Modeling," Sawtooth Software Research Paper Series, Sawtooth Software, Inc.

Shibata, R. (1984), "An Optimal Selection of Regression Variables," *Biometrika*, 71, 43-49.

Smith, M. and R. Kohn (1996), "Nonparametric Regression Using Bayesian Variable Selection," *Journal of Econometrics*, 75, 317-343.

Zheng, X. and W. Loh (1995), "Consistent Variable Selection in Linear Models," *Journal of the American Statistical Association*, 90, 429, 151-156.

APPENDIX: ESTIMATION ALGORITHM

The MCMC sampler discussed here is designed to estimate subject-specific configurations of relevant explanatory variables. Let $y_s = (y_{s1}, \ldots, y_{sT_h})$, $y = (y_1, \ldots, y_S)'$; $q = (q_1, \ldots, q_S)$; $\beta_0 = (\beta_{10}, \ldots, \beta_{S0})'$; $\beta^{(q)} = (\beta_1^{(k_s)\prime}, \ldots, \beta_S^{(k_S)\prime})'$; $u_s = (u_{s1}, \ldots, u_{sT_S})'$, $u = \{u_s\}_{s=1}^{S}$; $x_s^{(k_S)} = \{x_{st}^{(k_S)}\}_{t=1}^{T_S}$, $x^{(q)} = \{x_s^{(k_S)}\}_{s=1}^{S}$; $S_l = \{s \mid R_l \in x_s^{(k_S)}, s = 1, \ldots, S\}$, $l = 1, \ldots, K$. Then, the full posterior distribution is

$$
\begin{aligned}
& p(u, \beta_0, \beta^{(q)}, x^{(q)}, q, \mu_0, \sigma_0^2, \mu, \Sigma \mid y) \propto \\
& p(y \mid u, \beta_0, \beta^{(q)}, x^{(q)}) p(u \mid \beta_0, \beta^{(q)}, x^{(q)}) \times \\
& p(\beta_0 \mid \mu_0, \sigma_0^2) p(\beta^{(q)} \mid x^{(q)}, q, \mu, \Sigma) p(x^{(q)} \mid q) \times \\
& p(q) p(\mu_0) p(\sigma_0^2) p(\mu) p(\Sigma),
\end{aligned}
$$

where $p(y \mid u, \beta_0, \beta^{(q)}, x^{(q)}) p(u \mid \beta_0, \beta^{(q)}, x^{(q)}) \propto {}_{s=1}^{S}{}_{t=1}^{T_s} (N_{J-1}(u_{st} \mid 1_{J-1}\beta_{s0} + x_{st}^{(k_s)}\beta_s^{(k_s)}, \mathbf{I}_{J-1}) I_{u_{ht} \in \dot{A}_{st}})$, $\dot{A}_{st} = \dot{A}_{s1t} \times \ldots \times \dot{A}_{s,J-1,t}$ is the sample space of u_{st} given the mapping in ((5)), $\mathbf{I}_{J-1}$ is a $(J-1) \times (J-1)$ identity matrix, and I_a is the usual indicator function for the event a. To evaluate ((16)), we use the Markov chain Monte Carlo method by sampling all unknown quantities in a sequence as follows:

1. Sampling from $p(u \mid \beta_0, \beta^{(q)}, x^{(q)}, q, \mu_0, \sigma_0{}^2, \mu, \Sigma, y)$: for each $s = 1, \ldots, S$ and $t = 1, \ldots, T_s$,
   $p(u_{st} \mid \beta_{s0}, x_{st}^{(k_s)}, \beta_s^{(k_s)}, y_{st}) \propto N_{J-1}(u_{st} \mid 1_{J-1}\beta_{s0} + x_{st}^{(k_s)}\beta_s^{(k_s)}, \mathbf{I}_{J-1}) I_{u_{st} \in \dot{A}_{st}}$,
   which can be sampled by the multivariate slice sampling method (Neal 2003).
2. Sampling from $p(\beta_0 \mid u, \beta^{(q)}, x^{(q)}, q, \mu_0, \sigma_0^2, \mu, \Sigma, y)$: for each $s = 1, \ldots, S$, $p(\beta_{s0} \mid u_s, \mu_0, \sigma_0^2, x_s^{(k_s)}, \beta_s^{(k_s)}) \propto N(\bar{\mu}_0, \bar{\sigma}_0^2)$,
   where $\bar{\mu}_0 = \{\mu_0 + \sigma_0^2 {}_{t=1}^{T_s}{}_{j=1}^{J-1}(u_{sjt} - x_{sjt}^{(k_s)\prime}\beta_s^{(k_s)})\}/(1 + \sigma_0^2 T_s(J-1))$ and
   $\bar{\sigma}_0^2 = \sigma_0^2/(1 + \sigma_0^2 T_s(J-1))$.
3. Sampling from $p(\mu_0 \mid u, \beta_0, \beta^{(q)}, x^{(q)}, q, \sigma_0^2, \mu, \Sigma, y)$:
   $p(\mu_0 \mid \beta_0, \sigma_0^2) \propto N(\bar{m}_0, \bar{\upsilon}_0)$, where $\bar{m}_0 = (\sigma_0^2 m_0 + \upsilon_0 {}_{s=1}^{S}\beta_{s0})/(\sigma_0^2 + S\upsilon_0)$
   and $\bar{\upsilon}_0 = \upsilon_0\sigma_0^2/(\sigma_0^2 + S\upsilon_0)$.
4. Sampling from $p(\sigma_0^2 \mid u, \beta_0, \beta^{(q)}, x^{(q)}, q, \mu_0, \mu, \Sigma, y)$:
   $p(\sigma_0^2 \mid \beta_0, \mu_0) \propto IG(\bar{a}_0, \bar{b}_0)$, where $\bar{a}_0 = a_0 + 0.5S$ and
   $\bar{b}_0 = b_0 + 0.5 {}_{s=1}^{S}(\beta_{s0} - \mu_0)^2$.

5. Sampling from $p(\mu \mid u, \beta_0, \beta^{(q)}, x^{(q)}, q, \mu_0, \sigma_0^2, \Sigma, y)$: for each $l = 1, \ldots, K$, $p(\mu_l \mid \beta^{(q)}, x^{(q)}, \Sigma) \propto N(\bar{m}_l, \bar{\upsilon}_l)$, where $\bar{m}_l = (\sigma_l^2 m_l + \upsilon_l \mathcal{S}_l \beta_{s,i_s}^{(k_s)}) / (\sigma_l^2 + \#(\mathcal{S}_l)\upsilon_l)$ and $\bar{\upsilon}_l = \upsilon_l \sigma_l^2 / (\sigma_l^2 + \#(\mathcal{S}_l)\upsilon_l)$. The $\#(D)$ denotes the size of a set D and the subscript i_s ($i_s = 1, \ldots, k_s$) indicates the column of $x_s^{(k_s)}$ corresponding to R_l.
6. Sampling from $p(\Sigma \mid u, \beta_0, \beta^{(q)}, x^{(q)}, q, \mu_0, \sigma_0^2, \mu, y)$: For each $l = 1, \ldots, K$, $p(\sigma_l^2 \mid \beta^{(q)}, \mu_l) \propto IG(\bar{a}_l, \bar{b}_l)$, where $\bar{a}_l = a_l + 0.5\#(\mathcal{S}_l)$ and $\bar{b}_l = b_l + 0.5 \mathcal{S}_l (\beta_{s,i_s}^{(k_s)} - \mu_l)^2$.
7. Sampling from $p(q, x^{(q)}, \beta^{(q)} \mid u, \beta_0, \mu_0, \sigma_0^2, \mu, \Sigma y)$: For each subject s, $p(q_s, x_s^{(k_s)}, \beta_s^{(k_s)} \mid u_s) \propto p(u_s \mid \beta_{s0}, x_s^{(k_s)}, \beta_s^{(k_s)})$ $p(\beta_s^{(k_s)} \mid x_s^{(k_s)}, q_s, \mu, \Sigma) p(x_s^{(k_c)} \mid q_s) p(q_s)$.
   Under the reversible jump MCMC method, four possible transitions are allowed: birth, death, replacement and update steps. Thus, the set of possible moves is $m \in \{U, W, 0, 1, 2, \ldots\}$, where U means an update of regression coefficient $\beta_s^{(k_s)}$, W means a random replacement of an explanatory variable, and $n = 0, 1, 2, \ldots$ refers to increasing the number of explanatory variables from $q_s = c$ to $q_s = c + 1$ or decreasing from $q_s = c + 1$ to $q_h = c$.

As mentioned at the prior distribution for q_s is a Poisson distribution with mean λ, Po(λ). Since the parameter of the Poisson, λ, is a researcher's prior belief on q_s, it is important to set λ so that all values of $q_s \leq Q$ have reasonably large prior probabilities. Then, the probabilities for these four possible transitions are:

(1) $\varsigma_{q_s} = a \min\left(1, \dfrac{p(q_s + 1)}{p(q_s)}\right)$ for a birth move;

(2) $\tau_{q_s} = a \min\left(1, \dfrac{p(q_s - 1)}{p(q_s)}\right)$ for a death move;

(3) $\psi_{q_s} = \dfrac{1}{c}(1 - \varsigma_{q_s} - \tau_{q_s})$ for an update move, and;

(4) $\upsilon_{q_s} = 1 - \varsigma_{q_s} - \tau_{q_s} - \psi_{q_s}$ for a replacement move,

where the constant a should be as large as possible subject to $\varsigma_{q_s} + \tau_{q_s} \leq 0.9$ for all $q_s = 0, 1, \ldots, Q$ to ensure $\varsigma_{q_s} p(q_s) = \tau_{q_s+1} p(q_s + 1)$. Note that a should fall in an interval [0, 0.5] since if $a > 0.5$, then the sum of the probabilities ς_{q_h} and τ_{q_h} could be greater than 1 for some values of q_s. Since the value of q_s should be in an interval [0, Q], set ς_0 = 1 and $\varsigma_Q = \upsilon_Q$ = 0. In addition, c was set to 0.5. However, any values ranging from 0 to 1 are valid for c.

After choosing one of these four moves given the four transition probabilities, move to the next step as follows:

(a) Update move: sample $\beta_h^{(k_S)}$ from a multivariate normal distribution with mean vector $\mu^{(k_S)} = \Sigma^{(k_s)}(\Sigma_\beta^{-1}\mu_\beta + \sum_{t=1}^{T_h} x_{st}^{(k_s)'} u_{st})$ and $\Sigma^{(k_s)} = (\Sigma_\beta^{-1} + \sum_{t=1}^{T_h} x_{st}^{(k_s)'} x_{st}^{(k_s)})^{-1}$, where the prior mean vector μ_β and diagonal covariance matrix Σ_β are constructed from μ and Σ so that an n-th element of $\mu_{\beta,n}$, $= \mu_l$ and n (n, n) element of Σ_β, $\Sigma_{\beta,nn} = \sigma_l^2$ when the n-th column of $x_{st}^{(k_S)}$ corresponds to the l-th element in R, where $n = 1, \ldots, k_s$ and $l = 1, \ldots, K$.

(b) Other moves

First, given explanatory variables currently present, $g_s^{(q_S)}$, generate $g_s^{(q_S+i)}$, $i = -1, 0, 1$, as follows:

i. Birth step: add one more explanatory variable a^b by uniformly choosing one of the $Q - q_s$ candidates, $\left(A \cap g_s^{(q_s)}\right)^c$,

ii. Death step: delete an explanatory variable, a^d, by uniformly choosing one of q_s currently present variables from $g_s^{(q_S)}$,

iii. Replacement step: replace one of the currently present q_s variables, a^d, uniformed chosen from $g_s^{(q_S)}$, with a currently non-present variable, a^b, uniformly chosen from $\left(A \cap g_s^{(q_s)}\right)^c$

Then, given $g_s^{(q_S+i)}$, define $x_{st}^{(\bar{k}_S)}$, where $\bar{k}_s$ is the new dimension given newly genertaed explanatory variable set. Next, propose $\beta_h^{(\bar{k}_S)}$ by sampling from $N(\alpha^{(\bar{k}_s)}, (x_s^{(\bar{k}_s)'} x_s^{(\bar{k}_s)})^{-1})$, where $\alpha^{(\bar{k}_S)}$ is the maximum likelihood estimate found by solving $u_s = x_s^{(\bar{k}_s)}\alpha^{(\bar{k}_s)} + \varepsilon, \varepsilon \sim N(0, I_{(J-1)\times T_s})$. Finally, decide whether or not to accept $(q_s + i, x_s^{(\bar{k}_s)}, \beta_h^{(\bar{k}_s)})$ by computing acceptance probability:

$$\alpha = \min\{1, (\text{likelihood ratio}) \times (\text{prior ratio}) \times (\text{proposal ratio}) \times \text{Jacobian})\}$$

$$= \min\left\{1, \frac{\left(\prod_{t=1}^{T_s} p(u_{st} \mid \beta_{s0}, x_{st}^{(\bar{k}_s)}, \beta_s^{(\bar{k}_s)})\right) p(\beta_s^{(\bar{k}_s)} \mid \mu^{(\bar{k}_s)}, \Sigma^{(\bar{k}_s)}) p(\beta_s^{(k_s)} \mid \alpha^{(k_s)}, (x_s^{(k_s)'} x_s^{(k_s)})^{-1})}{\left(\prod_{t=1}^{T_s} p(u_{st} \mid \beta_{s0}, x_{st}^{(k_s)}, \beta_s^{(k_s)})\right) p(\beta_s^{(k_s)} \mid \mu^{(k_s)}, \Sigma^{(k_s)}) p(\beta_s^{(\bar{k}_s)} \mid \alpha^{(\bar{k}_s)}, (x_s^{(\bar{k}_s)'} x_s^{(\bar{k}_s)})^{-1})}\right\}$$

Received April 23
Accepted May 08