PSYCHOMETRIKA—VOL. 81, NO. 1, 161–183
MARCH 2016
DOI: 10.1007/s11336-014-9437-6

# A BAYESIAN MULTINOMIAL PROBIT MODEL FOR THE ANALYSIS OF PANEL CHOICE DATA

DUNCAN K. H. FONG

THE PENNSYLVANIA STATE UNIVERSITY

SUNGHOON KIM

ARIZONA STATE UNIVERSITY

ZHE CHEN

GOOGLE INC.

WAYNE S. DESARBO

THE PENNSYLVANIA STATE UNIVERSITY

A new Bayesian multinomial probit model is proposed for the analysis of panel choice data. Using a parameter expansion technique, we are able to devise a Markov Chain Monte Carlo algorithm to compute our Bayesian estimates efficiently. We also show that the proposed procedure enables the estimation of individual level coefficients for the single-period multinomial probit model even when the available prior information is vague. We apply our new procedure to consumer purchase data and reanalyze a well-known scanner panel dataset that reveals new substantive insights. In addition, we delineate a number of advantageous features of our proposed procedure over several benchmark models. Finally, through a simulation analysis employing a fractional factorial design, we demonstrate that the results from our proposed model are quite robust with respect to differing factors across various conditions.

Key words: Bayesian analysis, heterogeneity, multinomial probit model, panel data, parameter expansion, marketing, consumer psychology.

## 1. Introduction

The multinomial probit (MNP) model has its roots in the psychometrics, econometrics, and transportation modeling literature (e.g., Thurstone, 1927; Hausman & Wise, 1978; Daganzo, 1980). The model is commonly used to analyze discrete choice data collected in a single period and its appeal is the relaxation of the independence of irrelevant alternatives (IIA) property commonly associated with multinomial logit models. Even though some mixed and generalized multinomial logit models have recently been developed to avoid the IIA property (see Fiebig, Keane, Louviere, & Wasi, 2010 for a review), the MNP model is still practically important and widely used in practice. In cases where observations are collected in multiple periods, e.g., panel data on choices, a modified version of the model is needed to accommodate heterogeneity. In this paper, we propose a heterogeneous Bayesian multinomial probit model which assumes individual level random coefficients for the analysis of panel choice data.

A historical challenge in applying the traditional MNP model was that high-dimensional integrations were required for estimation which limited its usage. In recent years, advances in Bayesian computation have made this model more accessible and flexible in the analysis of discrete choice data. For the single-period multinomial probit model, McCulloch and Rossi (1994) have

Correspondence should be made to Wayne S. DeSarbo, The Pennsylvania State University, University Park, PA 16802 USA. Email: wsd6@psu.edu

described a Bayesian analysis using the Gibbs sampler. The method is easy to implement, and the computation can become a routine process. However, a drawback of their approach is that the resulting posterior distribution is sensitive to the prior specification of the identified parameters (see Chib, Greenberg, & Chen 1998). As pointed out in Albert and Chib (1993), there is an identifiability issue with the MNP model and there is a maximum allowable number of identifiable parameters for the covariance matrix of the latent utilities. Various Bayesian papers have been written on identification strategies and imposing priors on the identifiable parameters of the MNP model. For example, see Nobile (1998, 2000), McCulloch, Polson, and Rossi (2000), Imai and van Dyk (2005), Burgette and Nordheim (2012). In the psychometric literature, Tsai (2000, 2003) and Maydeu-Olivares and Hernandez (2007) have studied identifiability issues with the related Thurstonian models for paired comparisons data.

Bayesian multi-period multinomial probit models for the analysis of panel data on choices have found important applications in the Marketing, Economics, and Consumer Psychology literatures. Rossi, McCulloch, and Allenby (1996) (referred to as RMA here after) proposed an individual level, random coefficient multinomial choice model to assess the information content of various datasets available for direct marketing purposes. The authors assume a diagonal covariance matrix for latent utilities in their model and fix the first diagonal element at one to solve the MNP identifiability problem. Here, we build on their work to propose a more general Bayesian multinomial probit model to analyze panel data on choices. We assume a correlation matrix for latent utilities in our model that leads to a general covariance matrix for *relative utilities*, which are used by many authors to construct the MNP model (e.g., McCulloch & Rossi, 1994). Note, in a conjoint analysis setting, Dotson et al. (2010) considered unit variances in their probit model with structured covariances. The assumption of a correlation matrix for latent utilities has also been made in the related literature on Bayesian multivariate probit models (Chib & Greenberg, 1998; Liu, 2001; Liu & Daniels, 2006). However, we are the first to show that in the multinomial probit setting, the correlation matrix formulation is one with the maximum allowable number of parameters where identifiability is achieved and that it yields a general model. In addition, a Markov Chain Monte Carlo algorithm employing parameter expansion techniques has been developed to compute our Bayesian estimates. Furthermore, we provide a proof that the proposed procedure enables the estimation of individual level coefficients for the single-period multinomial probit model which extends the traditional MNP model.

Through a simulation analysis employing a fractional factorial design, we demonstrate that results from our proposed model are quite robust with respect to differing factors across various conditions. As to be shown later in this manuscript, we reanalyze the scanner panel dataset in Rossi, McCulloch, and Allenby (1996) using the proposed model and obtained new substantive insights. Here, we considered three benchmark models for comparison purposes. In addition to the one in Rossi et al. (1996), we included two additional benchmark models as extensions of those in McCulloch and Rossi (1994) and McCulloch et al. (2000). Results from the three benchmark models are mixed as conclusions from each model are sensitive to the choice of vague prior specifications. We will also discuss some convergence issues associated with these procedures.

The paper is organized as follows. Section 2 presents our proposed general Bayesian multinomial probit model together with a description of the computational algorithm utilized to implement our procedure. Descriptions of the benchmark models are also given in this section. Section 3 reanalyzes the RMA (1996) canned tuna scanner purchase data and compares our results with those using the three benchmark models. In Section 4, we perform an orthogonal factorial Monte Carlo analysis to examine robustness of performance of the proposed model across diverse conditions including the number of observations and prior specifications. Finally, we provide our summary, conclusions, limitations, and directions for future research in Section 5.

## 2. The Heterogeneous Bayesian Multinomial Probit Model

### 2.1. The Proposed Model

Let us assume there are $m$ choice alternatives and $H$ subjects in a panel dataset containing the choice history. Let $I_{ht}$ denote the choice for household $h$ at time $t$, where $h = 1, \ldots, H$ and $t = 1, \ldots, T_h$, where $T_h \geq 1$. Subject $h$ will choose alternative $j$, $j = 1, \ldots, m$, if the corresponding utility is the largest amongst the $m$ alternatives. More specifically, let $\boldsymbol{y}_{ht}$ be an $m$-dimensional vector representing the latent utilities; then

$$I_{ht} = j, \quad \text{if} \quad \boldsymbol{y}_{ht[j]} > \max(\boldsymbol{y}_{ht[-j]}), \tag{1}$$

where $\boldsymbol{y}_{ht[j]}$ is the $j^{th}$ component of $\boldsymbol{y}_{ht}$ and $\max(\boldsymbol{y}_{ht[-j]})$ represents the maximum value of all components in $\boldsymbol{y}_{ht}$ excluding $\boldsymbol{y}_{ht[j]}$. The latent utility $\boldsymbol{y}_{ht}$ is assumed to follow a multivariate regression as specified below:

$$\boldsymbol{y}_{ht} = \boldsymbol{X}_{ht}\boldsymbol{\beta}_h + \boldsymbol{\varepsilon}_{ht}, \quad \boldsymbol{\varepsilon}_{ht} \sim N(\mathbf{0}, \boldsymbol{\Lambda}), \tag{2}$$

where $\boldsymbol{X}_{ht}$ is an $m \times k$ matrix of various attributes and characteristics, and includes an intercept term for each of the alternatives, $\boldsymbol{\beta}_h$ is a $k$-dimensional vector representing subject $h$'s coefficients, and $\boldsymbol{\varepsilon}_{ht}$ is an error term that follows a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Lambda}$.

Note, the parameters $\boldsymbol{\beta}_h$ and $\boldsymbol{\Lambda}$ in (2) are not likelihood identified because the observed choices do not change with a transformation of the latent utilities such as $\boldsymbol{y}^*_{\boldsymbol{ht}} = a\mathbf{1}_m + b\boldsymbol{y}_{ht}$, which can be rewritten as $\boldsymbol{y}^*_{\boldsymbol{ht}} = \boldsymbol{X}_{\boldsymbol{ht}}\boldsymbol{\beta}^*_{\boldsymbol{h}} + \boldsymbol{\varepsilon}^*_{\boldsymbol{ht}}$, where $\mathbf{1}_m$ is an $m$-dimensional vector of 1's, $b > 0$, $\boldsymbol{\beta}^*_{\boldsymbol{h}} = b\boldsymbol{\beta}_h + [a0\ldots0]'$, $\boldsymbol{\varepsilon}^*_{\boldsymbol{ht}} \sim N\left(0, \ \boldsymbol{\Lambda}^*\right)$ and $\boldsymbol{\Lambda}^* = b^2\boldsymbol{\Lambda}$. Thus, given a set of observations $(I_{ht}, \boldsymbol{X}_{ht})$, the likelihood $\mathrm{L}(\boldsymbol{\beta}_h, h = 1, \ldots, H, \boldsymbol{\Lambda})$ is equal to $\mathrm{L}(\boldsymbol{\beta}^*_h, h = 1, \ldots, H, \boldsymbol{\Lambda}^*)$. To solve this identification problem, one may fix the intercept parameter of the first alternative at 0 and set the first diagonal element of $\boldsymbol{\Lambda}$ to 1. However, a covariance matrix $\boldsymbol{\Lambda}$ with unit first diagonal element is still not likelihood identified. If we denote $\boldsymbol{w}_{ht}$ as the $(m-1)$-dimensional vector of relative utilities with the jth component given by $\boldsymbol{w}_{ht[j]} = \boldsymbol{y}_{ht[j]} - \boldsymbol{y}_{ht[m],} j = 1, \ldots, m-1$, then $I_{ht} = \mathrm{m}$ when $\max(\boldsymbol{w}_{ht}) < 0$, and $I_{ht} = \mathrm{j}$ when $\boldsymbol{w}_{ht[j]} = \max(\boldsymbol{w}_{ht}) > 0$. Since the observed choices $I_{ht}$ are dependent only on the lower dimensional latent utility vectors $\boldsymbol{w}_{ht}$ (the corresponding covariance matrix is a $(m-1) \times (m-1)$ matrix), it is clear that, for identifiability reasons, $\boldsymbol{\Lambda}$ should be parameterized in terms of a parameter vector of dimension not exceeding $m(m-1)/2$. Here we assume a correlation matrix denoted by $\boldsymbol{R} = (r_{ij})$ in (2) which has exactly $\frac{m(m-1)}{2}$ unique parameters. Note that our correlation matrix assumption is not restrictive as it yields a general covariance matrix for the relative utilities $\boldsymbol{w}_{ht}$, which can be used (instead of $\boldsymbol{y}_{ht}$; see below) to construct the proposed multinomial probit model.

For the identified parameters (with $\boldsymbol{\Lambda} = \mathbf{R}$) in (2), we first assume a multivariate regression model for $\boldsymbol{\beta}_h$ using specified background (e.g., demographic) variables as independent variables. Since we set the first element of $\boldsymbol{\beta}_h$ to be 0, there are only $(k-1)$ effective parameters. For simplicity, we will still use $\boldsymbol{\beta}_h$ to represent this reduced coefficient vector henceforth. Let $\mathbf{Z}_h$ be an $l$-dimensional vector with an intercept and $(l-1)$ background variables for subject $h$. Our model for $\boldsymbol{\beta}_h$ is

$$\boldsymbol{\beta}_h = \boldsymbol{\Delta}\mathbf{Z}_h + \boldsymbol{\delta}_h, \ \ \boldsymbol{\delta}_h \sim N\left(\mathbf{0}, \boldsymbol{V}_\beta\right), \ \ h = 1, \ldots, H, \tag{3}$$

where the $(k-1) \times l$ matrix $\boldsymbol{\Delta}$ measures the impact of background variables on the $\boldsymbol{\beta}_h$ coefficients. The error term $(\boldsymbol{\delta}_h)$ follows a $(k-1)$-variate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{V}_\beta$. The following conjugate priors are assumed for $\boldsymbol{\Delta}$ and $\boldsymbol{V}_\beta$:

$$\boldsymbol{\Delta}|\boldsymbol{V}_\beta \sim MN(\boldsymbol{\Delta}_0, \boldsymbol{V}_\beta, \boldsymbol{A}_d^{-1}); \;\; \boldsymbol{V}_\beta^{-1} \sim W(\boldsymbol{v}, \boldsymbol{V}), \tag{4}$$

where $MN(\boldsymbol{\Delta}_0, \boldsymbol{V}_\beta, \boldsymbol{A}_d^{-1})$ denotes a matrix normal distribution with mean matrix $\boldsymbol{\Delta}_0$, row covariance matrix $\boldsymbol{V}_\beta$, and column covariance matrix $\boldsymbol{A}_d^{-1}$ (cf., Dawid, 1981; Gupta & Nagar, 2000), and $W(\boldsymbol{v}, \boldsymbol{V})$ is a Wishart distribution with $\boldsymbol{v}$ degrees of freedom and mean matrix $\boldsymbol{v}\boldsymbol{V}$. If $\boldsymbol{A}_d$ is a constant (c) multiple of an identity matrix and $\boldsymbol{V}$ is an identity matrix, then a small c (so 1/c is large) and a small $\boldsymbol{v}$ will denote vague prior information in (4). Second, for the correlation matrix $\boldsymbol{R}$, we assume the locally uniform prior:

$$\pi(\boldsymbol{R}) \propto 1, \quad \boldsymbol{R} \in \beth \tag{5}$$

where the correlation matrix space $\beth$ is convex and compact (cf., Rousseeuw & Molenberghs, 1994). Note that the uniform prior specified in (5) is a proper prior (a probability distribution integrating to one; see Barnard, McCulloch, & Meng, 2000).

### *2.2. The Full Conditional Distributions for the Proposed Model*

For Bayesian computation, we generate an approximate sample from the joint posterior distribution by drawing random deviates iteratively and recursively from the following full conditional distributions. The derivations of the respective full conditional distributions are given in Appendix 1. Here,

(i)

$$\boldsymbol{y}_{ht}|\text{all others} \sim TN(\tilde{\boldsymbol{X}}_{ht}\boldsymbol{\beta}_h, \boldsymbol{R}), \tag{6}$$

where $\boldsymbol{y}_{ht[I_{ht}]} > \max(\boldsymbol{y}_{ht[-I_{ht}]})$ and $\tilde{\boldsymbol{X}}_{ht}$ represents the reduced $m \times (k-1)$ matrix of $\boldsymbol{X}_{ht}$ by deleting its first column;

(ii)

$$\boldsymbol{\beta}_h|\text{all others} \sim N\left(\boldsymbol{V}_\beta^h \left[\sum\nolimits_t (\tilde{\boldsymbol{X}}_{ht}' \boldsymbol{R}^{-1}\boldsymbol{y}_{ht}) + \boldsymbol{V}_\beta^{-1}\boldsymbol{\Delta} Z_h\right], \boldsymbol{V}_\beta^h\right), \tag{7}$$

where $\boldsymbol{V}_\beta^h = [\sum_t (\tilde{\boldsymbol{X}}_{ht}'\boldsymbol{R}^{-1}\tilde{\boldsymbol{X}}_{ht}) + \boldsymbol{V}_\beta^{-1}]^{-1}$;

(iii)

$$\boldsymbol{\Delta}|\text{all others} \sim MN((\boldsymbol{B}\boldsymbol{Z}' + \boldsymbol{\Delta}_0\boldsymbol{A}_d)(\boldsymbol{Z}\boldsymbol{Z}' + \boldsymbol{A}_d)^{-1}, \boldsymbol{V}_\beta, (\boldsymbol{Z}\boldsymbol{Z}' + \boldsymbol{A}_d)^{-1}), \tag{8}$$

where $\boldsymbol{B} = [\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_H]$ and $\boldsymbol{Z} = [\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_H]$;

(iv)

$$\begin{aligned}\boldsymbol{V}_\beta^{-1}|\text{all others} \sim W(v + \mathrm{H} + l, [(\boldsymbol{B} - \boldsymbol{\Delta}\boldsymbol{Z})(\boldsymbol{B} - \boldsymbol{\Delta}\boldsymbol{Z})' \\ +(\boldsymbol{\Delta} - \boldsymbol{\Delta}_0)\boldsymbol{A}_d(\boldsymbol{\Delta} - \boldsymbol{\Delta}_0)' + \boldsymbol{V}^{-1}]^{-1}),\end{aligned} \tag{9}$$

and

(v) the full conditional distribution of $\boldsymbol{R}, \pi(\boldsymbol{R}|\text{all others})$, is not a standard probability distribution. Sampling from this distribution is discussed in the next Section 2.3.

### 2.3. *Sampling $\boldsymbol{R}$ from its full conditional for the Proposed Model*

Unlike the covariance matrix, it is difficult to find a conjugate prior for the correlation matrix **R** so that the full conditional distribution becomes a standard probability distribution. To deal with the correlation matrix, some researchers have proposed sampling methods which are quite computationally intensive. Chib and Greenberg (1998) assume that the off-diagonal components of **R** follow a multivariate normal distribution truncated to the space of correlation matrix, and then use the Metropolis–Hastings algorithm to sample new correlation matrices. However, values sampled from their proposal density are not guaranteed to be positive definite, and the algorithm has the potential problem of slow mixing. The approach proposed by Liechty, Liechty, and Muller (2004) also uses the Metropolis–Hastings algorithm, but they simulate all the components of the correlation matrix one by one through introduction of a latent variable. In our proposed approach, we employ a parameter expansion technique (Liu & Wu, 1999; Liu & Daniels, 2006; Hobert & Marchev, 2008) to generate random correlation matrices. To implement this method, we first construct a one-to-one mapping between two sets of quantities, one includes the covariance matrix and the other the correlation matrix. Then, one uses an inverse Wishart distribution to randomly draw a covariance matrix which can be readily transformed into a correlation matrix **R**, which is accepted based on a Metropolis–Hastings acceptance probability. (Details of this simulation strategy are given in Appendix 2.) Below is a summary of the proposed MCMC sampling algorithm:

(1) Initialization of $\boldsymbol{y}_{ht},\ \boldsymbol{\beta}_h,\ \boldsymbol{\Delta},\ \boldsymbol{V}_\beta^{-1},\ \boldsymbol{R}$ .
(2) Generate $\boldsymbol{y}_{ht[j]}$ from a truncated normal distribution.
(3) Generate $\boldsymbol{\beta}_h$ from a multivariate normal distribution.
(4) Generate $\boldsymbol{\Delta}$ from a matrix normal distribution.
(5) Generate $\boldsymbol{V}_\beta^{-1}$ from a Wishart distribution.
(6) Generate a covariance matrix from an inverse Wishart distribution and compute the corresponding correlation matrix. Accept the candidate $\boldsymbol{R}$ with an appropriate acceptance probability.
(7) Save $\boldsymbol{y}_{ht},\ \boldsymbol{\beta}_h,\ \boldsymbol{\Delta},\ \boldsymbol{V}_\beta^{-1},\ \boldsymbol{R}$ per MCMC iteration.

### 2.4. *The Single Observation Case—Extending the Multinomial Probit Model*

For cross-sectional data involving a single observation per response unit, Fong, Ebbes, and DeSarbo (2012) propose a Bayesian regression model that enables the estimation of individual level regression coefficients. Here, we show that it is possible to obtain estimates of individual level coefficients for the multinomial probit model when there is only a single observed choice per response unit. It is clear that the full conditional distributions for our model in Section 2.2 remain true in this case ($T_h = 1$) and so estimates of individual level coefficients can be obtained as long as the joint posterior distribution is proper. Since a proper prior leads to a proper posterior, we only need to show that the joint posterior distribution is proper when an improper prior representing vague information is assumed.

**Theorem 1.** *In the case $T_h = 1$, when $\boldsymbol{A}_d$ in the prior distribution given in Equation (4) becomes a zero matrix in the limit to represent vague information, the posterior distribution from our proposed model is still proper. The proof of the theorem is given in Appendix 3.*

Note, when $T_h > 1$, the posterior distribution from our proposed model is proper as the posterior distribution from the single observation case becomes the proper prior distribution (Theorem 1) in the analysis of remaining data.

### 2.5. Comparative Benchmark Models

We consider several existing Bayesian MNP models or their extensions as comparative benchmark models i n the paper.

*2.5.1. Benchmark Model I (RMA)* The first benchmark model is the one proposed in Rossi et al. (1996), which restricts $\boldsymbol{\Lambda}$ in (2) to be a diagonal matrix with positive diagonal elements $(1, \sigma_2^2, \ldots, \sigma_m^2)$. The assumption was made to simplify the calculation of the choice probabilities. An argument can also be made that heterogeneity is taken into account via the hierarchy, and so the errors in the unit-level probit are much less correlated (Rossi, Allenby, & McCulloch, 2005, p. 181). So, RMA assume the same model equations (1) to (4), but the prior in (5) is replaced by the following prior on variances:

$$\sigma_j^{-2} \sim \text{Gamma}\left(\nu,\ \upsilon_j\right), j = 2, \ldots, m, \text{independently}, \tag{10}$$

where small $\nu$ and large $\upsilon_j$ represent vague prior information in (10). (Note that the expression for the full conditional distribution of $\boldsymbol{V}_\beta^{-1}$ in RMA is inaccurate.)

Although the diagonal covariance matrix assumption leads to simpler Bayesian computation (no Metropolis–Hastings step within the Gibbs sampler), this constraint has its own drawbacks. First, we show that RMA is less flexible compared to our proposed model. Since the observed choices $I_{ht}$ are dependent only on the latent relative utility vectors $\boldsymbol{w}_{ht}$ (see Section 2.1), we shall focus on the distribution of $\boldsymbol{w}_{ht}$. Without loss of generality, for the RMA benchmark model, we assume that the alternative with the largest variance is the first alternative so that $\sigma_j^2 < 1$ for all $j > 1$ (we can achieve this by suitably permuting the index of alternatives). Then, conditional on $\boldsymbol{\beta}_h$, the variance of $\boldsymbol{w}_{ht[j]}$ is $\sigma_j^2 + \sigma_m^2$, which lies between 0 and 2, and the covariance of $\boldsymbol{w}_{ht[j]}$ and $\boldsymbol{w}_{ht[k]}$ $(j \neq k)$ is $\sigma_m^2 > 0$. By comparison, for our proposed model, the variance of $\boldsymbol{w}_{ht[j]}$ is $2 - 2r_{jm}$, which takes on a wider range of values (between 0 and 4 since $-1 \leq r_{jm} \leq 1$), and the covariance of $\boldsymbol{w}_{ht[j]}$ and $\boldsymbol{w}_{ht[k]}$ is $1 - r_{jm} - r_{km} + r_{jk}$, which allows for both positive and negative values. Hence, the distribution for $\boldsymbol{w}_{ht}$ based on our model is more general than that based on RMA. Second, by allowing uncorrelated latent utilities (conditional on $\boldsymbol{\beta}_h$), in the case of a single observation per response unit, the joint posterior distribution can be improper when an improper prior representing vague information is assumed in RMA:

**Theorem 2.** *In the case of single observation per household ($T_h = 1$), if we assume an improper prior by letting $\nu \to 0$ and $\upsilon_j \to \infty$ in (10), the joint posterior distribution from RMA will be improper. The proof of the theorem is given in Appendix 4.*

Note, since improper priors are commonly used in the literature to represent vague prior information, it is important to check that a proper posterior distribution can be obtained in a Bayesian analysis or otherwise no formal probability based statements can be made.

*2.5.2. Benchmark Model II (MR)* McCulloch and Rossi (1994) presented a Bayesian random coefficients model to analyze panel data. However, they did not include background variables $\boldsymbol{Z}_h$ to model the coefficients $\boldsymbol{\beta}_h$. Here we extend their model (MR henceforth) to incorporate such variables in the analysis. We define

$$I_{ht} = \text{m when } \max\left(\boldsymbol{w}_{ht}\right) < 0 \text{ and } I_{ht} = \text{j when } \boldsymbol{w}_{ht[j]} = \max\left(\boldsymbol{w}_{ht}\right) > 0, \tag{11}$$

$$\boldsymbol{w}_{ht} = \ddot{\boldsymbol{X}}_{ht}\boldsymbol{\beta}_h + \boldsymbol{e}_{ht}, \quad \boldsymbol{e}_{ht} \sim N(\mathbf{0}, \boldsymbol{\Sigma}), \tag{12}$$

$$\boldsymbol{\beta}_h = \boldsymbol{\Delta}\boldsymbol{Z}_h + \boldsymbol{\delta}_h, \quad \boldsymbol{\delta}_h, \sim N\left(0, \boldsymbol{V}_\beta\right), \quad \text{h} = 1, \ldots, \text{H}, \tag{13}$$

$$\boldsymbol{\Delta}|\boldsymbol{V}_\beta \sim MN(\boldsymbol{\Delta}_0, \boldsymbol{V}_\beta, \boldsymbol{A}_d^{-1});\ \ \boldsymbol{V}_\beta^{-1} \sim W(v, \boldsymbol{V}), \tag{14}$$

$$\boldsymbol{\Sigma}^{-1} \sim W(v_0, \boldsymbol{V}_0), \tag{15}$$

where $\ddot{\boldsymbol{X}}$ denotes the $(m-1) \times (k-1)$ matrix obtained from $\tilde{\boldsymbol{X}}_{ht}$ by subtracting the $m$th row from the first $(m-1)$ rows and $\boldsymbol{\Sigma}$ is a $(m-1) \times (m-1)$ covariance matrix. Note that the parameters $(\boldsymbol{\beta}_{h,} \boldsymbol{\Sigma})$ are not identified here because the likelihood $\mathrm{L}(\boldsymbol{\beta}_{h,} h = 1, \ldots, H, \boldsymbol{\Sigma})$ is equal to $\mathrm{L}(b\boldsymbol{\beta}_{h,} h = 1, \ldots, H, b^2\boldsymbol{\Sigma})$ for any b > 0. If $\sigma_{11}$ is the first diagonal element of $\boldsymbol{\Sigma}$, identification can be achieved by setting $\sigma_{11}$ equal to 1.

We follow the approach in McCulloch and Rossi (1994) to implement a Gibbs sampler to generate random parameter deviates as if the general covariance matrix $\boldsymbol{\Sigma}$ were estimable, while performing posterior estimation in terms of identified parameters only (i.e., $\frac{\boldsymbol{\beta}_h}{\sqrt{\sigma_{11}}}, \frac{\boldsymbol{\Sigma}}{\sigma_{11}}$). The full conditional distribution of $\boldsymbol{w}_{ht}$ is truncated normal, $\boldsymbol{\beta}_h$ is normal, $\boldsymbol{\Delta}$ is matrix normal, and $\boldsymbol{V}_\beta^{-1}$ and $\boldsymbol{\Sigma}^{-1}$ are each Wishart. However, as noted in the literature, there are several problems with this approach. First, it is very sensitive to the starting values of the Markov chain (see Nobile, 1998). Second, the posteriors are sensitive to the prior specification of the identified parameters (see Chib et al., 1998). Third, neither improper priors nor very informative priors can be used (McCulloch & Rossi, 1994).

*2.5.3. Benchmark Model III (MPR)* McCulloch et al. (2000) propose placing a prior directly on the identified $\boldsymbol{\Sigma}$ (with $\sigma_{11} = 1$) in McCulloch and Rossi (1994). Using the following expression of $\boldsymbol{\Sigma}$, they specify priors on $\boldsymbol{\gamma}$ and $\boldsymbol{\Phi}$:

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \boldsymbol{\gamma}' \\ \boldsymbol{\gamma} & \boldsymbol{\Phi} + \boldsymbol{\gamma}\boldsymbol{\gamma}' \end{bmatrix}, \tag{16}$$

$$\boldsymbol{\gamma} \sim N(\bar{\boldsymbol{\gamma}},\ \boldsymbol{B}^{-1}), \tag{17}$$

$$\boldsymbol{\Phi}^{-1} \sim W(\kappa, \mathbf{C}). \tag{18}$$

Again, we extend their model (MPR henceforth) to incorporate background variables in the analysis. So, similar to MR, MPR assumes (11) to (14) but (15) is replaced by (16) to (18). Since the full conditional distribution of $\boldsymbol{w}_{ht}$ is truncated normal, $\boldsymbol{\beta}_h$ is normal, $\boldsymbol{\Delta}$ is matrix normal, $\boldsymbol{V}_\beta^{-1}$ is Wishart, $\boldsymbol{\gamma}$ is normal, and $\boldsymbol{\Phi}^{-1}$ is Wishart, a Gibbs sampler is used to generate random deviates from the joint posterior distribution. We note that, as pointed out by Nobile (2000), the algorithm in McCulloch et al. (2000) can be much slower to converge than either the procedure of McCulloch and Rossi (1994) or of Nobile (1998).

## 3. The Application: Tuna Scanner Data

In this section, we reanalyze the scanner panel dataset of canned tuna purchases considered in Rossi et al. (1996). We compare results from our proposed model with those from the three benchmark models: RMA, MR, and MPR.

### 3.1. The Data

We start with a brief description of the dataset which is an A.C. Nielsen scanner panel dataset of tuna purchases in Springfield, Missouri. The dataset contains purchase information of 400 households on five brands of canned tuna: Chicken of the Sea (Water), Starkist (Water), House Brand (Water), Chicken of the Sea (Oil), and Starkist (Oil). These households make, on average, 13 purchases from this set of five brands with a range of between 1 and 61 purchases. Three purchasing characteristics are also available: the price, the existence of in-store displays, and feature advertisements at the moment of purchase. As in RMA, price is recorded in its logarithm form, and display and feature are coded as dummy variables. In addition, the dataset also provides five household demographic (background) characteristics: household income, family size, if the head of household is retired, if the head of household is employed, and if the head of household is a single mom. So, the panel data consist of 400 rows and as many columns as purchase points plus extra columns for the purchasing characteristics and demographics.

As in our simulation study, we divide the dataset into two parts: the calibration set and the validation set to evaluate the models' predictive validity. Here, we save the last three purchases of each household to form the validation set, and the earlier observations are used in the calibration set for parameter estimation.

### 3.2. *Prior Settings*

To investigate the robustness of the procedures, two sets of vague priors are considered across the four competing models:

For the RMA model, (i) for *diffuse proper priors I* (identical values utilized in Rossi et al. 1996), we set $\boldsymbol{\Delta_0}$ to a zero matrix, $\boldsymbol{A_d} = 0.01\boldsymbol{E_6}$, $\upsilon = 11$, $\boldsymbol{V} = \boldsymbol{E}_7$, $\boldsymbol{\nu} = 3$, and $\upsilon_j = 1$, where $\boldsymbol{E_l}$ represents an $l \times l$ identity matrix; and (ii) for *diffuse proper priors II* (with even larger variances), we set $\boldsymbol{\Delta_0}$ to a zero matrix, $\boldsymbol{A_d} = 0.001\boldsymbol{E_6}$, $\upsilon = 11$, $\boldsymbol{V} = \boldsymbol{E}_7$, $\nu = 0.06$, and $\upsilon_j = 50$.

For the MR model, (i) for *diffuse proper priors I*, we set $\boldsymbol{\Delta_0}$ to a zero matrix, $\boldsymbol{A_d} = 0.01\boldsymbol{E_6}$, $\upsilon = 11$, $\boldsymbol{V} = \boldsymbol{E_7}$, $\upsilon_0 = 8$, and $V_0 = 0.001\boldsymbol{E_4}$; (ii) for *diffuse proper priors II*, we set $\boldsymbol{\Delta_0}$ to a zero matrix, $\boldsymbol{A_d} = 0.001\boldsymbol{E_6}$, $\upsilon = 11$, $\boldsymbol{V} = \boldsymbol{E_7}$, $\upsilon_0 = 4$, and $\boldsymbol{V}_0 = 0.0001\boldsymbol{E_4}$.

For the MPR model, (i) for *diffuse proper priors I*, we set $\boldsymbol{\Delta_0}$ to a zero matrix, $\boldsymbol{A_d} = 0.01\boldsymbol{E}_6$, $\upsilon = 11$, $\boldsymbol{V} = \boldsymbol{E}_7$, $\kappa = 7$, $\mathbf{C} = \boldsymbol{E}_3$, $\overline{\boldsymbol{\gamma}}$ a zero vector, $\tau = 0.1$, and $\mathbf{B}^{-1} = \tau\boldsymbol{E_3}$; (ii) for *diffuse proper priors II*, we set $\boldsymbol{\Delta_0}$ to a zero matrix, $\boldsymbol{A_d} = 0.001\boldsymbol{E_6}$, $\upsilon = 11$, $\boldsymbol{V} = \boldsymbol{E_7}$, $\kappa = 7$, $\mathbf{C} = \boldsymbol{E}_3$, $\overline{\boldsymbol{\gamma}}$ a zero vector, and $\tau = 0.001$.

For our proposed model, (i) for *diffuse proper priors I*, we set $\boldsymbol{\Delta_0}$ to a zero matrix, $\boldsymbol{A_d} = 0.01\boldsymbol{E_6}$, $\upsilon = 11$, $\boldsymbol{V} = \boldsymbol{E_7}$; (ii) for *diffuse proper priors II*, we set $\boldsymbol{\Delta_0}$ to a zero matrix, $\boldsymbol{A_d} = 0.001\boldsymbol{E_6}$, $\upsilon = 11$, $\boldsymbol{V} = \boldsymbol{E_7}$.

### 3.3. *Parameter Estimation Results*

One of the goals in Rossi et al. (1996) was to investigate the relative value of demographic versus purchase history information. These authors have evaluated the amount of variability of the household-specific parameters that can be explained by the demographic variables, as opposed to unobservable heterogeneity, which is given by the square root of the diagonal elements of $\boldsymbol{V}_\beta$. (A large size of the unobservable heterogeneity suggests that the demographic information may have limited value in predicting many of the key parameters.) Also, for each of the regression coefficients in $\boldsymbol{\beta}_h$, the following measure is used:

$$\rho^2 = 1 - \frac{\mathrm{Var(unobservable\ component)}}{\mathrm{Var(Total)}}. \tag{19}$$

This measure $\rho^2$ is similar to the coefficient of determination utilized in classical regression: a large value implies that the demographic variables are useful in predicting that parameter. In addition to reporting the Bayesian estimates of the parameters of interest, we will follow RMA to provide the posterior probability that a coefficient is negative or positive (depending on the sign of the estimate), which is a measure one may use to decide whether a coefficient is zero or

TABLE 1.
Posterior estimates of **Δ** coefficients from the RMA (1996) benchmark model using diffuse proper priors I for the calibration set of empirical tuna purchases.

| | Cons | ln(Inc) | ln(Fam Size) | Retire | Unemp HH | Single Mom | Unobs. Hetero | $\rho^2$ |
|---|---|---|---|---|---|---|---|---|
| Starkist Water Int | 0.00 | 0.07 | −0.03 | 0.02 | **0.64** | −0.11 | 0.76 | 0.10 |
| *Probability* | *0.53* | *0.73* | *0.58* | *0.53* | *0.99* | *0.70* | | |
| Private Label Int | −**3.46** | −**1.09** | 0.40 | 0.08 | **1.98** | −0.19 | 2.00 | 0.33 |
| *Probability* | *1.00* | *1.00* | *0.85* | *0.56* | *1.00* | *0.63* | | |
| C–O–S Oil Int | −1.34 | −0.14 | 0.05 | 0.14 | 0.41 | −0.01 | 1.69 | 0.01 |
| *Probability* | *1.00* | *0.77* | *0.58* | *0.67* | *0.83* | *0.52* | | |
| Starkist Oil Int | −**1.80** | −0.19 | −0.10 | 0.42 | **1.32** | −0.35 | 2.10 | 0.05 |
| *Probability* | *1.00* | *78.00* | *0.63* | *0.85* | *0.99* | *0.76* | | |
| Price coef | −**6.24** | −0.38 | −0.58 | −**1.45** | −0.63 | −0.24 | 2.90 | 0.08 |
| *Probability* | *1.00* | *0.83* | *0.86* | *0.98* | *0.75* | *0.61* | | |
| Display coeff | **0.18** | −0.02 | −0.12 | −**0.35** | −**0.49** | 0.10 | 0.30 | 0.72 |
| *Probability* | *1.00* | *0.54* | *0.80* | *0.95* | *0.98* | *0.67* | | |
| Feature coef | **0.32** | 0.05 | 0.04 | **0.47** | 0.16 | 0.09 | 0.30 | 0.71 |
| *Probability* | *1.00* | *0.71* | *0.62* | *1.00* | *0.78* | *0.68* | | |

**Underlined cells** indicate the cells of probability that coefficients are positive or negative exceeds 0.9.

not. For the proposed model, convergence based on trace plots and Geweke's diagnostic appeared to be achieved after 5,000 iterations for both sets of priors. Thus, we discarded the first 5,000 iterations as a burn-in period and collected the following 20,000 iterations for sampling. We did the same (i.e., running 25,000 iterations with the first 5,000 as burn-in) for the three benchmark models for diffuse priors I. However, for diffuse priors II, it appeared that the MCMC chains took much longer to converge for the benchmark models and so we increased the number of burn-in to 20,000 and collected the following 20,000 iterations for sampling for all benchmark models.

First, we run the RMA benchmark model using the diffuse proper prior specifications (the same as *diffuse proper priors I*) as stated in Rossi et al. (1996) on the calibration data. The results are shown in Table 1, which are very similar to those obtained using all available data. Thus, retaining part of the tuna purchase data for validation does not affect the model's performance here. Table 2 shows results from RMA when *diffuse proper priors II* are assumed, which are substantially different (e.g., $\rho^2$) from those in Table 1. We note that the results are still substantially different even when 50,000 iterations are used in the burn-in period. Because of this lack of robustness, it is difficult to make definitive conclusions based on the results of the RMA benchmark model.

MR is the next benchmark model to be examined. We run an analysis using the two sets of diffuse proper prior specifications on the calibration dataset. As shown in Table 3, the MR results from *diffuse proper priors I* look somewhat different from the RMA results in Table 1, where more demographic variables, namely, 'ln(Inc),' 'ln(Fam Size),' 'Retire' and 'Single Mom,' are now considered significant (Underlined cells). When comparing the results from Table 3 to the results from Table 4 (with *diffuse proper priors II*), there are non-trivial variations in parameter estimates as well as parameter significance from using the two different prior specifications. For example, when we compare 'Starkist Water Int' variable, 'C–O–S Oil Int' variable, 'Display Coef' variable, and 'Feature Coef' variable between Tables 3 and 4, we observe only 50 % or less matching results in terms of significant variables. In addition, the corresponding values of $\rho^2$ for the 'Display Coef' are inconsistent between Tables 3 and 4, and $\rho^2$ values of 'Display Coef' variable and 'Feature Coef' variable are much smaller than 0.5, which implies that the demographic characteristics are not useful in estimating marketing characteristic sensitivities according to the MR model.

TABLE 2.
Posterior estimates of $\boldsymbol{\Delta}$ coefficients from the RMA (1996) benchmark model using diffuse proper priors II for the calibration set of empirical tuna purchases.

| | Cons | ln(Inc) | ln(Fam Size) | Retire | Unemp HH | Single Mom | Unobs. Hetero | $\rho^2$ |
|---|---|---|---|---|---|---|---|---|
| Starkist Water Int | **1398.20** | 11.87 | 9.22 | −102.61 | −**613.25** | −35.65 | 1346.80 | 0.02 |
| *Probability* | *1.00* | *0.54* | *0.52* | *0.67* | *0.97* | *0.55* | | |
| Private Label Int | −**1182.30** | −**236.07** | **231.71** | 93.10 | **719.43** | 58.31 | 1161.12 | 0.06 |
| *Probability* | *1.0* | *0.98* | *0.93* | *0.67* | *1.00* | *0.59* | | |
| C–O–S Oil Int | −**1481.19** | 119.90 | −141.94 | 148.50 | 196.69 | 8.65 | 1705.25 | 0.01 |
| *Probability* | *1.00* | *0.75* | *0.73* | *0.69* | *0.68* | *0.51* | | |
| Starkist Oil Int | −**2700.22** | −51.62 | 54.00 | 316.27 | **739.81** | −271.23 | 2486.23 | 0.01 |
| *Probability* | *1.00* | *0.58* | *0.56* | *0.76* | *0.90* | *0.69* | | |
| Price coef | −**653.41** | 187.99 | 338.90 | **642.65** | 309.18 | 489.36 | 2596.64 | 0.01 |
| *Probability* | *1.00* | *0.80* | *0.86* | *0.95* | *0.72* | *0.85* | | |
| Display coeff | 75.04 | 114.88 | −88.43 | −55.31 | **589.10** | **378.52** | 1208.33 | 0.03 |
| *Probability* | *0.87* | *0.83* | *0.71* | *0.61* | *0.99* | *0.94* | | |
| Feature Coef | 45.51 | **161.26** | **231.96** | **392.47** | −111.85 | 247.11 | 1461.92 | 0.02 |
| *Probability* | *0.73* | *0.90* | *0.91* | *0.96* | *0.64* | *0.82* | | |

**Underlined cells** indicate the cells of probability that coefficients are positive or negative exceeds 0.9.

TABLE 3.
Posterior estimates of $\boldsymbol{\Delta}$ coefficients from the MR (1994) benchmark model using diffuse proper priors I for the calibration set of empirical tuna purchases.

| | Cons | ln(Inc) | ln(Fam Size) | Retire | Unemp HH | Single Mom | Unobs. Hetero | $\rho^2$ |
|---|---|---|---|---|---|---|---|---|
| Starkist Water Int | −**0.11** | **0.18** | 0.09 | 0.08 | **0.66** | −0.14 | 0.86 | 0.07 |
| *Probability* | *0.98* | *0.97* | *0.75* | *0.68* | *1.00* | *0.77* | | |
| Private Label Int | −**2.88** | −**0.82** | **0.62** | **0.66** | **2.15** | 0.18 | 1.91 | 0.18 |
| *Probability* | *1.00* | *1.00* | *1.00* | *1.00* | *1.00* | *0.84* | | |
| C–O–S Oil Int | −**1.56** | −**0.23** | −0.14 | **0.37** | **0.72** | −0.20 | 2.05 | 0.03 |
| *Probability* | *1.00* | *0.99* | *0.87* | *0.99* | *1.00* | *0.84* | | |
| Starkist Oil Int | −**2.29** | −**0.34** | −**0.31** | **0.59** | **1.23** | −**0.25** | 2.50 | 0.05 |
| *Probability* | *1.00* | *1.00* | *0.99* | *1.00* | *1.00* | *0.90* | | |
| Price Coef | −**5.92** | −**0.31** | 0.05 | −**0.86** | −**0.54** | −**0.35** | 2.88 | 0.02 |
| *Probability* | *1.00* | *0.99* | *0.60* | *1.00* | *0.95* | *0.93* | | |
| Display Coef | **0.18** | −0.02 | −0.05 | −0.14 | −**0.23** | −0.01 | 0.53 | 0.04 |
| *Probability* | *1.00* | *0.62* | *0.63* | *0.74* | *0.90* | *0.52* | | |
| Feature Coef | **0.39** | **0.10** | **0.20** | **0.47** | 0.17 | 0.08 | 0.50 | 0.26 |
| *Probability* | *1.00* | *0.92* | *0.96* | *1.00* | *0.77* | *0.70* | | |

**Underlined cells** indicate the cells of probability that coefficients are positive or negative exceeds 0.9.

MPR is the last benchmark model that we used to analyze the calibration data. Table 5 presents the MPR results assuming *diffuse proper priors I* which are somewhat similar to the RMA results in Table 1. Table 6 shows the MPR results when *diffuse proper priors II* are assumed and they are substantially different from those in Table 5. Thus, results of the MPR are not robust to prior specifications, and so it is again difficult to make definitive conclusions based on results from this benchmark model when the available information is vague.

TABLE 4.
Posterior estimates of **Δ** coefficients from the MR (1994) benchmark model using diffuse proper priors II for the calibration set of empirical tuna purchases.

| | Cons | ln(Inc) | ln(Fam Size) | Retire | Unemp HH | Single Mom | Unobs. Hetero | $\rho^2$ |
|---|---|---|---|---|---|---|---|---|
| Starkist Water Int | **0.13** | 0.02 | −0.12 | 0.07 | **0.86** | −**0.40** | 0.76 | 0.16 |
| *Probability* | *1.00* | *0.56* | *0.77* | *0.68* | *1.00* | *0.99* | | |
| Private Label Int | −**2.91** | −**1.07** | **0.34** | **0.35** | **2.04** | −0.19 | 1.93 | 0.22 |
| *Probability* | *1.00* | *1.00* | *0.99* | *0.94* | *1.00* | *0.70* | | |
| C–O–S Oil Int | −**2.02** | −0.19 | −0.08 | 0.23 | **0.84** | −0.48 | 2.38 | 0.02 |
| *Probability* | *1.00* | *0.89* | *0.67* | *0.83* | *1.00* | *0.80* | | |
| Starkist Oil Int | −**1.82** | −**0.22** | −**0.21** | **0.38** | **1.80** | −0.40 | 2.36 | 0.06 |
| *Probability* | *1.00* | *0.96* | *0.90* | *0.96* | *1.00* | *0.74* | | |
| Price coef | −**5.76** | −**0.43** | −**0.64** | −**1.15** | −**0.62** | 0.05 | 2.76 | 0.05 |
| *Probability* | *1.00* | *1.00* | *1.00* | *1.00* | *0.99* | *0.57* | | |
| Display coef | **0.19** | −0.19 | −0.06 | −**0.34** | −0.34 | −0.10 | 0.63 | 0.25 |
| *Probability* | *0.98* | *0.85* | *0.61* | *0.93* | *0.79* | *0.62* | | |
| Feature coef | **0.28** | 0.02 | 0.06 | **0.36** | 0.06 | 0.17 | 0.45 | 0.26 |
| *Probability* | *1.00* | *0.59* | *0.72* | *0.99* | *0.62* | *0.86* | | |

**Underlined cells** indicate the cells of probability that coefficients are positive or negative exceeds 0.9.

TABLE 5.
Posterior estimates of **Δ** coefficients from the MPR (2000) benchmark model using diffuse proper priors I for the calibration set of empirical tuna purchases.

| | Cons | ln(Inc) | ln(Fam Size) | Retire | Unemp HH | Single Mom | Unobs. Hetero | $\rho^2$ |
|---|---|---|---|---|---|---|---|---|
| Starkist Water Int | −0.01 | 0.10 | −0.03 | 0.02 | **0.54** | −0.08 | 0.75 | 0.08 |
| *Probability* | *0.57* | *0.83* | *0.61* | *0.54* | *0.97* | *0.65* | | |
| Private Label Int | −**3.48** | −**1.15** | 0.33 | 0.01 | **2.06** | −0.15 | 2.06 | 0.33 |
| *Probability* | *1.00* | *1.00* | *0.87* | *0.48* | *1.00* | *0.61* | | |
| C–O–S Oil Int | −**1.89** | −**0.25** | 0.09 | 0.23 | **0.61** | 0.04 | 2.31 | 0.02 |
| *Probability* | *1.00* | *0.95* | *0.67* | *0.81* | *0.96* | *0.56* | | |
| Starkist Oil Int | −**2.31** | −**0.27** | −0.14 | **0.54** | **1.48** | −0.30 | 2.65 | 0.05 |
| *Probability* | *1.00* | *0.91* | *0.69* | *0.96* | *1.00* | *0.80* | | |
| Price Coef | −**6.20** | −0.25 | −**0.75** | −**2.06** | −**0.95** | −0.14 | 3.08 | 0.11 |
| *Probability* | *1.00* | *0.81* | *0.98* | *1.00* | *0.93* | *0.56* | | |
| Display Coef | **0.18** | −0.05 | −**0.23** | −**0.39** | −**0.43** | −0.03 | 0.37 | 0.68 |
| *Probability* | *0.99* | *0.64* | *0.92* | *0.94* | *0.93* | *0.54* | | |
| Feature Coef | **0.39** | 0.02 | 0.05 | **0.51** | 0.12 | 0.06 | 0.33 | 0.73 |
| *Probability* | *1.00* | *0.59* | *0.65* | *1.00* | *0.69* | *0.64* | | |

**Underlined cells** indicate the cells of probability that coefficients are positive or negative exceeds 0.9.

We now apply our proposed model to the calibration data using both *diffuse proper priors I* and *II*, and the corresponding results are shown in Tables 7 and 8. If we compare Table 7 with Tables 1 and 5, it might be argued that results from the different models are quite similar. However, there is a significant difference in that our model yields consistent estimates with respect to the two sets of proper vague priors as shown in Tables 7 and 8. As summary statistics, the root mean squared difference of estimated coefficients from the proposed model is 0.01 whereas the RMA (between Tables 1 and 2) has the largest difference of 499.59, followed by the MPR (between

TABLE 6.
Posterior estimates of $\boldsymbol{\Delta}$ coefficients from the MPR (2000) benchmark model using diffuse proper priors II for the calibration set of empirical tuna purchases.

| | Cons | ln(Inc) | ln(Fam Size) | Retire | Unemp HH | Single Mom | Unobs. Hetero | $\rho^2$ |
|---|---|---|---|---|---|---|---|---|
| Starkist Water Int | −0.05 | 0.11 | 0.12 | 0.05 | **0.85** | −0.03 | 1.21 | 0.07 |
| *Probability* | *0.72* | *0.70* | *0.69* | *0.57* | *1.00* | *0.52* | | |
| Private Label Int | −**16.64** | −**6.16** | **1.47** | 1.44 | **11.74** | −0.88 | 16.49 | 0.11 |
| *Probability* | *1.00* | *1.00* | *0.91* | *0.81* | *1.00* | *0.67* | | |
| C–O–S Oil Int | −**14.65** | −**3.01** | −0.31 | **2.50** | **6.98** | −0.85 | 19.84 | 0.03 |
| *Probability* | *1.00* | *1.00* | *0.59* | *0.97* | *1.00* | *0.66* | | |
| Starkist Oil Int | −**17.47** | −**3.37** | −1.17 | **2.98** | **9.84** | −1.92 | 21.45 | 0.04 |
| *Probability* | *1.00* | *1.00* | *0.78* | *0.98* | *1.00* | *0.77* | | |
| Price Coef | −**14.14** | −**1.34** | −**0.98** | −**6.50** | −**8.34** | −0.30 | 8.46 | 0.16 |
| *Probability* | *1.00* | *0.94* | *0.92* | *1.00* | *1.00* | *0.52* | | |
| Display Coef | **0.45** | −0.20 | −0.24 | 0.16 | 0.61 | −0.54 | 0.76 | 0.42 |
| *Probability* | *1.00* | *0.72* | *0.73* | *0.69* | *0.79* | *0.83* | | |
| Feature Coef | **0.86** | 0.08 | 0.18 | **0.44** | 0.25 | −0.28 | 0.66 | 0.19 |
| *Probability* | *1.00* | *0.69* | *0.78* | *0.91* | *0.65* | *0.79* | | |

**Underlined cells** indicate the cells of probability that coefficients are positive or negative exceeds 0.9.

TABLE 7.
Posterior estimates of $\boldsymbol{\Delta}$ coefficients from the proposed model using diffuse proper priors I for the calibration set of empirical tuna purchases.

| | Cons | ln(Inc) | ln(Fam Size) | Retire | Unemp HH | Single Mom | Unobs. Hetero | $\rho^2$ |
|---|---|---|---|---|---|---|---|---|
| Starkist Water Int | −0.09 | 0.12 | −0.04 | 0.02 | **0.66** | −0.12 | 0.79 | 0.12 |
| *Probability* | *0.88* | *0.82* | *0.60* | *0.57* | *0.98* | *0.68* | | |
| Private Label Int | −**4.14** | −**1.31** | 0.47 | 0.20 | **2.35** | −0.22 | 2.30 | 0.34 |
| *Probability* | *1.00* | *1.00* | *0.86* | *0.63* | *1.00* | *0.62* | | |
| C–O–S Oil Int | −**1.33** | −0.16 | 0.05 | 0.13 | 0.41 | −0.04 | 1.72 | 0.01 |
| *Probability* | *1.00* | *0.79* | *0.58* | *0.65* | *0.83* | *0.54* | | |
| Starkist Oil Int | −**1.98** | −0.21 | −0.15 | 0.43 | **1.36** | −0.32 | 2.20 | 0.05 |
| *Probability* | *1.00* | *0.80* | *0.67* | *0.84* | *0.99* | *0.74* | | |
| Price coef | −**6.82** | −0.33 | −0.69 | −**1.33** | −0.71 | −0.08 | 3.21 | 0.06 |
| *Probability* | *1.00* | *0.78* | *0.87* | *0.94* | *0.74* | *0.53* | | |
| Display coeff | **0.24** | −0.01 | −0.09 | −**0.42** | −**0.56** | 0.13 | 0.31 | 0.80 |
| *Probability* | *1.00* | *0.53* | *0.70* | *0.98* | *0.98* | *0.69* | | |
| Feature coef | **0.33** | 0.08 | 0.07 | **0.53** | 0.24 | 0.16 | 0.33 | 0.69 |
| *Probability* | *1.00* | *0.76* | *0.69* | *1.00* | *0.84* | *0.78* | | |

**Underlined cells** indicate the cells of probability that coefficients are positive or negative exceeds 0.9.

Tables 5 and 6) of 3.08 and then the MR (between Tables 3 and 4) of 0.17. Since the $\rho^2$ of "Display" and "Feature" coefficients in Tables 7 and 8 are always greater than 0.5 and the corresponding unobservable heterogeneity coefficients are around 0.3, based on the criterion given in RMA, some demographic variables are useful in predicting brand preferences and estimating marketing characteristic sensitivities. This conclusion is different from what has been reported in the literature (e.g., Rossi et al., 1996).

TABLE 8.
Posterior estimates of **Δ** coefficients from the proposed model using diffuse proper priors II for the calibration set of empirical tuna purchases.

| | Cons | ln(Inc) | ln(Fam Size) | Retire | Unemp HH | Single Mom | Unobs. Hetero | $\rho^2$ |
|---|---|---|---|---|---|---|---|---|
| Starkist Water Int | −0.09 | 0.12 | −0.04 | 0.02 | **0.65** | −0.11 | 0.79 | 0.11 |
| *Probability* | *0.88* | *0.82* | *0.60* | *0.54* | *0.98* | *0.66* | | |
| Private Label Int | −**4.18** | −**1.31** | 0.45 | 0.17 | **2.37** | −0.23 | 2.33 | 0.33 |
| *Probability* | *1.00* | *1.00* | *0.85* | *0.62* | *1.00* | *0.63* | | |
| C–O–S Oil Int | −**1.33** | −0.15 | 0.06 | 0.12 | 0.41 | −0.02 | 1.72 | 0.01 |
| *Probability* | *1.00* | *0.79* | *0.59* | *0.64* | *0.82* | *0.53* | | |
| Starkist Oil Int | −**1.98** | −0.20 | −0.16 | 0.43 | **1.36** | −0.31 | 2.20 | 0.05 |
| *Probability* | *1.00* | *0.80* | *0.69* | *0.85* | *0.99* | *0.74* | | |
| Price coef | −**6.82** | −0.35 | −0.70 | −**1.27** | −0.65 | −0.14 | 3.22 | 0.06 |
| *Probability* | *1.00* | *0.79* | *0.88* | *0.95* | *0.73* | *0.56* | | |
| Display coeff | **0.24** | −0.01 | −0.09 | −**0.44** | −**0.57** | 0.12 | 0.31 | 0.81 |
| *Probability* | *1.00* | *0.52* | *0.71* | *0.98* | *0.98* | *0.68* | | |
| Feature coef | **0.33** | 0.08 | 0.07 | **0.54** | 0.25 | 0.17 | 0.33 | 0.71 |
| *Probability* | *1.00* | *0.76* | *0.71* | *1.00* | *0.85* | *0.78* | | |

**Underlined cells** indicate the cells of probability that coefficients are positive or negative exceeds 0.9.

TABLE 9.
Holdout sample prediction accuracy comparison between three benchmark models and the proposed model for the validation set of empirical tuna purchases.

| Priors specification | RMA benchmark model (%) | MR benchmark model (%) | MPR benchmark model (%) | The proposed model (%) |
|---|---|---|---|---|
| Diffuse proper priors I | 76 | 75 | 76 | 77 |
| Diffuse proper priors II | 35 | 72 | 71 | 76 |

### *3.4. Holdout Sample Prediction Results*

To compare these four models in choice prediction performance, we evaluate predicted choices from the four models for observations in the validation dataset and compute the proportion of correct predictions. The corresponding results are shown in Table 9. When *diffuse proper priors I* are used, out-of-sample prediction accuracies from the four models are quite similar (i.e., 76 % for RMA, 75 % for MR, 76 % for MPR, and 77 % for our proposed model). However, when *diffuse proper priors II* are used, the proposed model outperforms all three benchmark models with about 117, 7, and 6 % increases in prediction accuracy over RMA, MPR, and MR, respectively. These results are consistent with earlier estimation results showing that the proposed model, being more robust with respect to vague prior specifications, can yield consistent estimates as well as better predictions.

## 4. Monte Carlo Simulation

Here we present a simulation analysis employing a fractional factorial design to investigate the performance of our proposed model with respect to differing factors across various conditions.

TABLE 10.
Experimental design factors for the simulation.

| Factor[a] | Levels | Code |
|---|---|---|
| Number of subjects (X1) | $H = 200$ | 1 |
| | $H = 500$ | 2 |
| Number of choice alternatives (X2) | $m = 5$ | 1 |
| | $m = 10$ | 2 |
| | $m = 15$ | 3 |
| Number of observations per subject (X3) | $T$=1 | 1 |
| | $T = 10$ | 2 |
| | $T = 20$ | 3 |
| Prior specifications (X4)* | Vague proper I | 1 |
| | Vague proper II | 2 |
| Covariance specification (X5) | Diagonal | 1 |
| | Correlation | 2 |
| Number of attribute variables (X6) | $k$ =5 | 1 |
| | $k$ =10 | 2 |
| Number of background variables (X7) | $l$ =3 | 1 |
| | $l$ =6 | 2 |

[a] *Vague proper prior I* sets $\boldsymbol{A}_d$ =0.01$\boldsymbol{E}_l$, v=k+3, and $\boldsymbol{V} = \boldsymbol{E}_{k-1}$; *Vague proper prior II* sets $\boldsymbol{A}_d = 0.001\boldsymbol{E}_l$, v= k+3, and $\boldsymbol{V} = \boldsymbol{E}_{k-1}$.

### *4.1. A Fractional Factorial Design*

Table 10 presents seven experimentally manipulated factors and their levels in this Monte Carlo analysis: the number of subjects (X1: $H = 200$ and $H = 500$), the number of choice alternatives (X2: $m = 5$, $m = 10$, and $m = 15$), the number of observations per subject (X3; $T = 1$, $T = 10$, and $T = 20$), the specification of the priors (X4: *Vague Proper I, Vague Proper II*), the covariance specification (X5: *diagonal* and *correlation*), the number of attributes (e.g., marketing mix variables in the tuna application) (X6: $k = 5$ and $k = 10$), and the number of background variables (e.g., demographic variables in the tuna application) (X7: $l = 3$ and $l = 6$). These factors with their two or three levels were specified to reflect various conditions representing a variety of potential applications, as well as the flexibility of the proposed methodology in fitting a variety of different model and data specifications. Regarding the data generation procedure, for pre-specified $H$, $m$, $T$, $k$, and $l$ values, elements of $\boldsymbol{Z}_h$ and $\boldsymbol{X}_{ht}$ are randomly generated from the standard normal distribution, and the matrix $\boldsymbol{\Delta}$ is generated from a matrix normal distribution with zero mean matrix and two covariance matrices which are randomly generated from an inverse Wishart distribution or fixed at the identity matrix. Then, $\boldsymbol{\beta}_h$ are generated from a multivariate normal distribution according to Equation (3), and finally, latent utilities $\boldsymbol{y}_{ht}$ are generated according to Equation (2).

We attempted to create a variety of empirical settings which would realistically test the performance of the proposed heterogeneous Bayesian multinomial probit methodology. Naturally, we would expect that conditions with larger datasets, smaller numbers of parameters, and accurate informative priors would provide conditions where the model would be expected to show better performance. In contrast, alternative conditions with smaller datasets, larger number of parameters, and availability of vague priors only would likely be more challenging for this or any particular model.

Consistent with past literature in *Psychometrika* (e.g., DeSarbo, 1982; DeSarbo & Carroll, 1985; DeSarbo & Cron, 1988; Jedidi & DeSarbo, 1991) involving Monte Carlo testing of new proposed methods, we created a collection of synthetic datasets as generated by use of an orthogonal factorial design by manipulating independent factors reflecting different data, parameter, and model specification conditions. We employ a $2^5 3^2$ fractional factorial design (SPSS ORTHOPLAN command; Addelman, 1962) to study the main effects of each factor. To assess the predictive performance of the proposed procedure, validation datasets consist of the last 10 % of observations in the case of multiple observations per household, and 10 % of the sample in the case of single observation per household is used.

Three dependent measures are computed:

1. The root mean squared error (RMSE 1) between the actual and recovered attribute variable coefficients;
2. The root mean squared error (RMSE 2) between the actual and recovered background variable coefficients;
3. Holdout sample prediction: the proportion of correct predictions of predicted choices for observations in the holdout sample.

These dependent measures encompass three major areas of model performance including various parameter recoveries and holdout sample choice recovery. In this Monte Carlo analysis, we used a total of 10,000 iterations, and after discarding the first 5,000 iterations as burn-in, the next 5,000 iterations were collected for parameter estimation and choice prediction per trial for the proposed heterogeneous Bayesian multinomial probit model. To track convergence, we employed trace plots which exhibited consistent patterns after the burn-in period, and the average acceptance rate for the Metropolis–Hastings step of the proposed procedure was 0.58 in this Monte Carlo analysis.

### *4.2. Monte Carlo Study Results*

Consistent with DeSarbo, (1982; see also DeSarbo & Carroll, 1985; DeSarbo & Cron, 1988; Jedidi & DeSarbo, 1991), we performed multiple regression analysis on each of the three dependent variables after coding the seven independent factors as dummy variables to explore main effects. A primary indication of methodology robustness would be to find non-significance of each regression model ($F$ test) as well as independent factors ($t$ tests) suggesting that the proposed MCMC estimation procedure is not significantly affected by data, prior, or model specifications. Concerning the multiple regression results using RMSE 1 as the dependent variable and the seven factors as independent variables appropriately coded as dummy variables, the corresponding ANOVA table shows that the overall linear model is not significant ($p$ value is 0.481) which suggests that none of the independent factors are significant. The adjusted $R^2 = 0.045$. Indeed, the corresponding table of regression coefficients indicates that none of the factors' levels are significant at the 0.05 level. The same holds true in the analysis of RMSE 2. The ANOVA test there shows that the overall linear model is not significant ($p$ value is 0.282), and none of the factors or their coefficients (levels of the factors) are significant at the 0.05 level. The adjusted $R^2 = 0.276$. This demonstrates that the proposed new Bayesian methodology appears robust with respect to these seven factors and their levels when dealing with parameter recovery.

Next, we used the holdout sample prediction performance as a dependent measure, and performed a third multiple regression analysis. The ANOVA test indicates that the overall model is significant ($p$ value is 0.030; adjusted $R^2 = 0.714$), but that only the X3 factor (Number of observations per subject) in the table of coefficients is significant at the 0.05 level. Regarding the significance of X3 and its larger values/levels, the result is not surprising given the nature of this factor. Note, for the single observation case (level 1 of X3), the validation dataset consists of holdout observations from different subjects, whereas for the multiple-observation cases (level 2 and level 3 of X3), the validation dataset consists of holdout observations from the same subjects. It is expected that prediction results at level 1 of X3 will not be as good as those at levels 2 and 3.

## 5. Summary and Conclusions

The proposed model contributes to the literature with respect to three aspects. First, it is a flexible and general model. There is no loss of generality by assuming a correlation matrix for the choice alternative utilities as it yields a general covariance matrix for the relative utilities.

Second, our proposed model appears to be more robust than the three benchmark models with respect to prior specifications when only vague prior information is available. Since accurate and stable parameter estimates are important factors with respect to substantive insights, our proposed model provides a useful tool when the available prior information is vague. Also, we show that our proposed model extends the traditional multinomial probit model by enabling the estimation of individual level coefficients when there is only single observation per subject.

Third and last, although it can be computationally challenging to generate a correlation matrix due to the lack of an explicit conjugate prior, we employ a parameter expansion technique to develop an efficient algorithm to sample the correlation matrix and, and as a direct result, the MCMC procedure becomes somewhat easy to implement.

Future research could proceed in a number of interesting directions. More rigorous simulation work is needed employing full factorial designs with multiple replications with many additional synthetic datasets generated. In this same vein, comparisons across all trials with all competitive benchmark procedures would be useful. Additional applications across a variety of contexts need to be explored to examine the performance of the proposed procedure vs. benchmark methods and to ascertain the conditions under which the substantive findings are similar/different. Potential impact of demographics on predicting customer preferences and estimating marketing mix sensitivities requires more investigation. Finally, more rigorous experimentation with prior specification sensitivity is needed. We believe that the proposed general Bayesian multinomial probit model is likely to become a useful tool for practical applications across a wide spectrum of social science applications.

## Acknowledgments

We wish to thank the Editor and three anonymous reviewers for their constructive comments which helped to improve this manuscript. We also wish to thank Eric Bradlow and Greg Allenby for kindly providing the data for the application. This research was funded in part by the Smeal College of Business.

## 6. Appendix 1: Derivation of the Full Conditional Distributions for the Proposed Model

### *6.1. Proof of Equation (6)*

Since $\boldsymbol{\beta}_h$ now denotes the reduced $(k-1)$-dimensional vector of parameters, Equation (2) becomes

$$\boldsymbol{y}_{ht} = \tilde{\boldsymbol{X}}_{ht}\boldsymbol{\beta}_h + \boldsymbol{\varepsilon}_{ht},$$

where $\tilde{\boldsymbol{X}}_{ht}$ is the reduced matrix of $\boldsymbol{X}_{ht}$ by deleting its first column. Then, the full conditional distribution of $\boldsymbol{y}_{ht}$ is

$$\pi(\boldsymbol{y}_{ht}|\text{all others}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{y}_{ht} - \tilde{\boldsymbol{X}}_{ht}\boldsymbol{\beta}_h)'\boldsymbol{R}^{-1}(\boldsymbol{y}_{ht} - \tilde{\boldsymbol{X}}_{ht}\boldsymbol{\beta}_h)\right\} I(\boldsymbol{y}_{ht[I_{ht}]} > \max(\boldsymbol{y}_{ht[-I_{ht}]})).$$

So $(\boldsymbol{y}_{ht}|\text{all others})$ follows a truncated normal distribution, $TN\left(\tilde{\boldsymbol{X}}_{ht}\boldsymbol{\beta}_h, \boldsymbol{R}\right)$, where the truncation is such that $\boldsymbol{y}_{ht[I_{ht}]} > \max(\boldsymbol{y}_{ht[-I_{ht}]})$. The $j$th component of $\boldsymbol{y}_{ht}$, $\boldsymbol{y}_{ht[j]}$, has a univariate truncated normal distribution conditional on all other components of $\boldsymbol{y}_{ht}$ and other parameters. Let $\boldsymbol{D}_j$ be a matrix that switches the first and the $j$th components of $\boldsymbol{y}_{ht}$:

$$\boldsymbol{D}_j \tilde{\boldsymbol{X}}_{ht} \boldsymbol{\beta}_h \triangleq \begin{pmatrix} \mu_j \\ \mu_{-j} \end{pmatrix}, \text{ and } \boldsymbol{D}_j \boldsymbol{R} \boldsymbol{D}_{j'} \triangleq \begin{pmatrix} \mathrm{r_{jj}} & r_{j12} \\ r_{j21} & r_{j22} \end{pmatrix}.$$

Then,

$$\boldsymbol{y}_{ht[j]} | \text{all others} \sim TN(\mu_j + r_{j12} r_{j22}^{-1} (\boldsymbol{y}_{ht[-j]} - \boldsymbol{\mu}_{-j}), r_{jj} - r_{j12} r_{j22}^{-1} r_{j21}),$$

where the truncation is given by

$$\begin{cases} \boldsymbol{y}_{htj} \in \left(\max\left(\boldsymbol{y}_{ht[-j]}\right), +\infty\right), & \text{if} \quad j = I_{ht}; \\ \boldsymbol{y}_{htj} \in \left(-\infty, y_{ht[I_{ht}]}\right), & \text{if} \quad j \neq I_{ht}. \end{cases}$$

### *6.2. Proof of Equation (7)*

The full conditional distribution of $\boldsymbol{\beta}_h$ is

$$\begin{aligned} &\pi(\boldsymbol{\beta}_h | \text{all others}) \\ &\propto \exp\left\{ -\frac{1}{2} \sum_{t=1}^{T_h} \left[ (\boldsymbol{y}_{ht} - \tilde{\boldsymbol{X}}_{ht} \boldsymbol{\beta}_h)' \boldsymbol{R}^{-1} (\boldsymbol{y}_{ht} - \tilde{\boldsymbol{X}}_{ht} \boldsymbol{\beta}_h) \right] - \frac{1}{2} (\boldsymbol{\beta}_h - \boldsymbol{\Delta} \boldsymbol{Z}_h)' \boldsymbol{V}_\beta^{-1} (\boldsymbol{\beta}_h - \boldsymbol{\Delta} \boldsymbol{Z}_h) \right\} \\ &\propto \exp\left\{ -\frac{1}{2} \left[ \boldsymbol{\beta}_h' \left( \sum_t (\tilde{\boldsymbol{X}}_{ht}' \boldsymbol{R}^{-1} \tilde{\boldsymbol{X}}_{ht}) + \boldsymbol{V}_\beta^{-1} \right) \boldsymbol{\beta}_h - 2\, \boldsymbol{\beta}_h' \left( \sum_t (\tilde{\boldsymbol{X}}_{ht}' \boldsymbol{R}^{-1} \boldsymbol{y}_{ht}) + \boldsymbol{V}_\beta^{-1} \boldsymbol{\Delta} \boldsymbol{Z}_h \right) \right] \right\}. \end{aligned}$$

So, $\boldsymbol{\beta}_h | \text{all others} \sim N(\boldsymbol{\beta}_h^0, \boldsymbol{V}_\beta^h)$, where

$$\boldsymbol{V}_\beta^h = \left( \sum\nolimits_t (\tilde{\boldsymbol{X}}_{ht}' \boldsymbol{R}^{-1} \tilde{\boldsymbol{X}}_{ht}) + \boldsymbol{V}_\beta^{-1} \right)^{-1} \text{ and } \boldsymbol{\beta}_h^0 = \boldsymbol{V}_\beta^h \left( \sum\nolimits_t (\tilde{\boldsymbol{X}}_{ht}' \boldsymbol{R}^{-1} \boldsymbol{y}_{ht}) + \boldsymbol{V}_\beta^{-1} \boldsymbol{\Delta} \boldsymbol{Z}_h \right).$$

### *6.3. Proof of Equation (8)*

The full conditional distribution of $\boldsymbol{\Delta}$ is

$$\begin{aligned} &\pi(\boldsymbol{\Delta} | \text{all others}) \propto \operatorname{etr}\left\{ -\frac{1}{2} \boldsymbol{V}_\beta^{-1} [(\boldsymbol{B} - \boldsymbol{\Delta} \boldsymbol{Z})(\boldsymbol{B} - \boldsymbol{\Delta} \boldsymbol{Z})' + (\boldsymbol{\Delta} - \boldsymbol{\Delta}_0) \boldsymbol{A}_d (\boldsymbol{\Delta} - \boldsymbol{\Delta}_0)'] \right\} \\ &\propto \operatorname{etr}\{ \boldsymbol{V}_\beta^{-1} [\boldsymbol{\Delta}(\boldsymbol{Z}\boldsymbol{Z}' + \boldsymbol{A}_d) \boldsymbol{\Delta}' - 2 \boldsymbol{\Delta} (\boldsymbol{Z} \boldsymbol{B}' + \boldsymbol{A}_d \boldsymbol{\Delta}_0')] \}, \end{aligned}$$

here etr refers to an exponential function of the trace of (a matrix).
So, $\boldsymbol{\Delta} | \text{all others} \sim MN((\boldsymbol{B}\boldsymbol{Z}' + \boldsymbol{\Delta}_0 \boldsymbol{A}_d)(\boldsymbol{Z}\boldsymbol{Z}' + \boldsymbol{A}_d)^{-1}, \boldsymbol{V}_\beta, (\boldsymbol{Z}\boldsymbol{Z}' + \boldsymbol{A}_d)^{-1})$, where

$$\boldsymbol{B} = [\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_H] \quad \text{and} \quad \boldsymbol{Z} = [\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_H].$$

*6.4. Proof of Equation (9)*

The full conditional distribution of $\boldsymbol{V}_\beta^{-1}$ is

$$\pi(\boldsymbol{V}_\beta^{-1}|\text{all others}) \propto |\boldsymbol{V}_\beta^{-1}|^{H/2}\text{etr}\left\{-\frac{1}{2}\boldsymbol{V}_\beta^{-1}(\boldsymbol{B}-\boldsymbol{\Delta Z})(\boldsymbol{B}-\boldsymbol{\Delta Z})'\right\}$$
$$\times|\boldsymbol{V}_\beta^{-1}|^{l/2}\text{etr}\left\{-\frac{1}{2}\boldsymbol{V}_\beta^{-1}(\boldsymbol{\Delta}-\boldsymbol{\Delta}_0)\boldsymbol{A}_d(\boldsymbol{\Delta}-\boldsymbol{\Delta}_0)'\right\}|\boldsymbol{V}_\beta^{-1}|^{\frac{v-k}{2}}\text{etr}\left\{-\frac{1}{2}\boldsymbol{V}_\beta^{-1}\boldsymbol{V}^{-1}\right\}$$

which is proportional to

$$|\boldsymbol{V}_\beta^{-1}|^{\frac{v+H+l-k}{2}}\text{etr}\left\{-\frac{1}{2}\boldsymbol{V}_\beta^{-1}[(\boldsymbol{B}-\boldsymbol{\Delta Z})(\boldsymbol{B}-\boldsymbol{\Delta Z})'+(\boldsymbol{\Delta}-\boldsymbol{\Delta}_0)\boldsymbol{A}_d(\boldsymbol{\Delta}-\boldsymbol{\Delta}_0)'+\boldsymbol{V}^{-1}]\right\}.$$

So, $\boldsymbol{V}_\beta^{-1}|\text{all others} \sim W(v+H+l, \boldsymbol{V}_n)$, where

$$\boldsymbol{V}_n = [(\boldsymbol{B}-\boldsymbol{\Delta Z})(\boldsymbol{B}-\boldsymbol{\Delta Z})'+(\boldsymbol{\Delta}-\boldsymbol{\Delta}_0)\boldsymbol{A}_d(\boldsymbol{\Delta}-\boldsymbol{\Delta}_0)'+\boldsymbol{V}^{-1}]^{-1}.$$

## 7. Appendix 2: A Parameter Expansion Algorithm for Sampling the Correlation Matrix

*7.1. Stage I: Parameter Expanded Reparameterization*

We define the following one-to-one mapping from $\{\boldsymbol{y}_{ht}, \boldsymbol{R}\}$ to $\{\boldsymbol{y}_{ht}^*, \boldsymbol{\Sigma}\}$ to facilitate making random draws of the correlation matrix:

$$\begin{cases}\boldsymbol{y}_{ht} = \tilde{\boldsymbol{X}}_{ht}\boldsymbol{\beta}_h + \boldsymbol{D}^{-1/2}\boldsymbol{y}_{ht}^* \\ \boldsymbol{R} = \boldsymbol{D}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{D}^{-1/2}\end{cases} \quad h = 1, \ldots, \mathrm{H} \quad \text{and} \quad t = 1, \ldots, T_h \tag{20}$$

where $\boldsymbol{\Sigma}$ is a positive definite matrix, $\sum_{h=1}^{H}\sum_{t=1}^{T_h}(\boldsymbol{y}_{ht[j]}^*)^2 = 1$for any $j = 1, ..., m$, and $\boldsymbol{D}$ is a diagonal matrix. Note that the constraints, $\sum_{h=1}^{H}\sum_{t=1}^{T_h}(\boldsymbol{y}_{ht[j]}^*)^2 = 1$for any $j = 1, ..., m$, are needed to make the transformation in (20) a one-to-one mapping. Given $\boldsymbol{\beta}_h$, the step that draws $\boldsymbol{y}_{ht}$ implicitly draws $\boldsymbol{y}_{ht}^*$ and $\boldsymbol{D}$ because $\boldsymbol{D}_{jj}^{1/2} = [\sum_{h=1}^{H}\sum_{t=1}^{T_h}(\boldsymbol{y}_{ht[j]} - \tilde{\boldsymbol{X}}_{ht[j]}\boldsymbol{\beta}_h)^2]^{-1/2}$, where $\boldsymbol{D}_{jj}$ is the jth element of $\boldsymbol{D}$ and $\tilde{\boldsymbol{X}}_{ht[j]}$ is the jth row of $\tilde{\boldsymbol{X}}_{ht}$. Thus, one can derive the joint conditional distribution of ($\boldsymbol{y}_{ht}^*, \boldsymbol{\Sigma}|$all others) from that of ($\boldsymbol{y}_{ht}, \boldsymbol{R}|$all others). When the prior distribution of $\boldsymbol{R}$ is $\pi(\boldsymbol{R}) \propto |\boldsymbol{R}|^{-(m+1)/2}$, the full conditional distribution of $\boldsymbol{\Sigma}$ is

$$\pi\ (\boldsymbol{\Sigma}|\text{all others}) \propto \prod\nolimits_{ht} \pi\ \ (\boldsymbol{y}_{ht}^*, \boldsymbol{\Sigma}|\text{all others}) \propto |\boldsymbol{\Sigma}|^{-(\sum_h T_h+m+1)/2}\exp\left\{-\frac{1}{2}tr(\boldsymbol{S\Sigma}^{-1})\right\} \tag{21}$$

where $\boldsymbol{S} = \sum_{h=1}^{H}\sum_{t=1}^{T_h}\boldsymbol{y}_{ht}^*\boldsymbol{y}_{ht}^{*\prime}$. The expression in (21) is proportional to a Wishart distribution and so $\boldsymbol{\Sigma}^{-1}|$all others $\sim\ W(\sum_h \boldsymbol{T}_h,\ \ \boldsymbol{S})$.

*7.2. Stage II: Parameter Expanded Metropolis–Hastings*

After we have obtained a random deviate of $\boldsymbol{\Sigma}$ from (21) Stage I, we can transform it into a correlation matrix using $\boldsymbol{R}^* = \boldsymbol{D}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{D}^{-1/2}$. Since $\boldsymbol{R}^*$ is obtained based on the candidate prior $\pi(\boldsymbol{R}^*) \propto |\boldsymbol{R}^*|^{-(m+1)/2}$, it is accepted in a Metropolis–Hasting step with probability α, where, at iteration n+1, $\alpha = min\{1, \exp[\frac{m+1}{2}(log|\boldsymbol{R}^*| - log|\boldsymbol{R}^{(n)}|)]\}$.

## 8. Appendix 3

*8.1. Proof of Theorem 1*

Since the joint posterior distribution is

$$\pi(\boldsymbol{Y}, \boldsymbol{B}, \boldsymbol{\Delta}, \boldsymbol{R}, \boldsymbol{V}_{\boldsymbol{\beta}}|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}) = \left[\prod_{h=1}^{H} \pi(\boldsymbol{\beta_h}|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{y_h}, \boldsymbol{\Delta}, \boldsymbol{R}, \boldsymbol{V}_{\boldsymbol{\beta}})\right]\pi(\boldsymbol{\Delta}|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{V}_{\boldsymbol{\beta}})$$
$$\times\, \pi\left(\boldsymbol{Y}|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{R}, \boldsymbol{V}_{\boldsymbol{\beta}}\right)\pi\left(\boldsymbol{R}|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{V}_{\boldsymbol{\beta}}\right)\pi\left(\boldsymbol{V}_{\boldsymbol{\beta}}|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}\right),$$

it is a proper probability distribution if each of the posterior distributions on the right-hand side of the above equation is proper. To establish the result, we make use of the likelihood function of our model when $T_h = 1$:

$$L(\boldsymbol{Y}, \boldsymbol{B}, \boldsymbol{\Delta}, \boldsymbol{R}, \boldsymbol{V}_{\boldsymbol{\beta}}|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}) \propto \prod_{h=1}^{H} [I(\boldsymbol{y}_{h[I_h]} > \max(\boldsymbol{y}_{h[-I_h]}))]|\boldsymbol{R}|^{-\mathrm{H}/2}|\boldsymbol{V}_{\beta}|^{-\mathrm{H}/2}$$
$$\times \exp\left\{-\frac{1}{2}\sum_{h=1}^{H}((\boldsymbol{y}_h - \tilde{\boldsymbol{X}}_h\boldsymbol{\beta}_h)'\boldsymbol{R}^{-1}(\boldsymbol{y}_h - \tilde{\boldsymbol{X}}_h\boldsymbol{\beta}_h) + (\boldsymbol{\beta}_h - \boldsymbol{\Delta}\boldsymbol{Z}_h)'\boldsymbol{V}_{\beta}^{-1}(\boldsymbol{\beta}_h - \boldsymbol{\Delta}\boldsymbol{Z}_h))\right\},$$

where $\boldsymbol{X} = [\tilde{\boldsymbol{X}}_1, \ldots, \tilde{\boldsymbol{X}}_H]$ and $\boldsymbol{Y} = [\boldsymbol{y}_1', \ldots, \boldsymbol{y}_H']'$.

- For $h = 1, \ldots, H$,

$$\pi(\boldsymbol{\beta}_h|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{\Delta}, \boldsymbol{R}, \boldsymbol{V}_{\beta}) = \pi(\boldsymbol{\beta}_h|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{\Delta}, \boldsymbol{R}, \boldsymbol{V}_{\beta})$$
$$\propto exp\left\{-\frac{1}{2}[(\boldsymbol{y}_h - \tilde{\boldsymbol{X}}_h\boldsymbol{\beta}_h)'\boldsymbol{R}^{-1}(\boldsymbol{y}_h - \tilde{\boldsymbol{X}}_h\boldsymbol{\beta}_h) + (\boldsymbol{\beta}_h - \boldsymbol{\Delta}\boldsymbol{Z}_h)'\boldsymbol{V}_{\beta}^{-1}(\boldsymbol{\beta}_h - \boldsymbol{\Delta}\boldsymbol{Z}_h)]\right\}.$$

So, $\boldsymbol{\beta}_h|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{\Delta}, \boldsymbol{R}, \boldsymbol{V}_{\beta} \sim N(\boldsymbol{b}_h^0, \boldsymbol{V}_{\beta}^{\mathrm{new}})$, where

$$\boldsymbol{V}_{\beta}^{\mathrm{new}} = (\tilde{\boldsymbol{X}}_h'\boldsymbol{R}^{-1}\tilde{\boldsymbol{X}}_h + \boldsymbol{V}_{\beta}^{-1})^{-1} \text{and} \quad \boldsymbol{b}_h^0 = \boldsymbol{V}_{\beta}^{\mathrm{new}}(\tilde{\boldsymbol{X}}_h'\boldsymbol{R}^{-1}\boldsymbol{y}_h + \boldsymbol{V}_{\beta}^{-1}\boldsymbol{\Delta}\boldsymbol{Z}_h).$$

Thus, $\pi(\boldsymbol{\beta}_h|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{\Delta}, \boldsymbol{R}, \boldsymbol{V}_{\beta})$ is proper.

- Let $\boldsymbol{\eta}$ be the vectorization of $\boldsymbol{\Delta}$, and $\boldsymbol{\Upsilon}_h = \tilde{\boldsymbol{X}}_h(\boldsymbol{E}_{k-1} \otimes \boldsymbol{Z}_h)'$, where $\boldsymbol{E}_{k-1}$ is the identity matrix with dimension $(k-1) \times (k-1)$ and $\otimes$ denotes the Kronecker product. $\pi(\boldsymbol{\Delta}|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{V}_{\beta})$ is equivalent to $\pi(\boldsymbol{\eta}|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{V}_{\beta})$, where

$$
\begin{aligned}
&\pi(\boldsymbol{\eta}|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{V}_\beta) = \pi(\boldsymbol{\eta}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{V}_\beta) \propto L(\boldsymbol{\eta}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{V}_\beta)\pi(\boldsymbol{\eta}|\boldsymbol{V}_\beta) \\
&\propto \exp\left\{-\frac{1}{2}\sum\nolimits_{h=1}^{H} [(\boldsymbol{y}_h - \boldsymbol{\Upsilon}_h\boldsymbol{\eta})'(\tilde{\boldsymbol{X}}_h\boldsymbol{V}_\beta\tilde{\boldsymbol{X}}_h' + \boldsymbol{R})^{-1}(\boldsymbol{y}_h - \boldsymbol{\Upsilon}_h\boldsymbol{\eta})]\right\} \\
&\quad \times \exp\left\{-\frac{1}{2}(\boldsymbol{\eta} - \overline{\boldsymbol{d}})'(\boldsymbol{V}_\beta \otimes \boldsymbol{A}_d^{-1})^{-1}(\boldsymbol{\eta} - \overline{\boldsymbol{d}})\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{\eta}'((\boldsymbol{V}_\beta^{-1} \otimes \boldsymbol{A}_d) + \sum\nolimits_{h=1}^{H} [\boldsymbol{\Upsilon}_h'(\tilde{\boldsymbol{X}}_h\boldsymbol{V}_\beta\tilde{\boldsymbol{X}}_h' + \boldsymbol{R})^{-1}\boldsymbol{\Upsilon}_h])\boldsymbol{\eta}\right.\right. \\
&\quad \left.\left. -2\boldsymbol{\eta}'\left(\sum\nolimits_{h=1}^{H} [\boldsymbol{\Upsilon}_h'(\tilde{\boldsymbol{X}}_h\boldsymbol{V}_\beta\tilde{\boldsymbol{X}}_h' + \boldsymbol{R})^{-1}\boldsymbol{y}_h] + (\boldsymbol{V}_\beta^{-1} \otimes \boldsymbol{A}_d)\overline{\boldsymbol{d}}\right)\right]\right\}.
\end{aligned}
$$

Typically, one lets $\boldsymbol{A}_d$ in the prior distribution to approach a zero matrix to represent vague information. Then, in the limit, the above expression goes to

$$
\exp\{\frac{-1}{2}[\boldsymbol{\eta}'(\sum\nolimits_{h=1}^{H} [\boldsymbol{\Upsilon}_h'(\tilde{\boldsymbol{X}}_h\boldsymbol{V}_\beta\tilde{\boldsymbol{X}}_h' + \boldsymbol{R})^{-1}\boldsymbol{\Upsilon}_h])\boldsymbol{\eta} - 2\boldsymbol{\eta}'(\sum\nolimits_{h=1}^{H} [\boldsymbol{\Upsilon}_h'(\tilde{\boldsymbol{X}}_h\boldsymbol{V}_\beta\tilde{\boldsymbol{X}}_h' + \boldsymbol{R})^{-1}\boldsymbol{y}_h])]\}.
$$

Thus, $\boldsymbol{\eta}|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{V}_\beta \sim N(\boldsymbol{d}_0, \boldsymbol{V}_\eta)$, where

$$
\boldsymbol{V}_\eta = \left[\sum\nolimits_{h=1}^{H} (\boldsymbol{\Upsilon}_h'(\tilde{\boldsymbol{X}}_h\boldsymbol{V}_\beta\tilde{\boldsymbol{X}}_h' + \boldsymbol{R})^{-1}\boldsymbol{\Upsilon}_h)\right]^{-1}
$$
$$
\text{and} \quad \boldsymbol{d}_0 = \boldsymbol{V}_\eta \sum\nolimits_{h=1}^{H} (\boldsymbol{\Upsilon}_h'(\tilde{\boldsymbol{X}}_h\boldsymbol{V}_\beta\tilde{\boldsymbol{X}}_h' + \boldsymbol{R})^{-1}\boldsymbol{y}_h).
$$

So, the posterior distribution of $\boldsymbol{\eta}|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{V}_\beta$ is proper.

- Let

$$
\boldsymbol{\Upsilon} = \begin{pmatrix} \boldsymbol{\Upsilon}_1 \\ \vdots \\ \boldsymbol{\Upsilon}_H \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Psi} = \begin{pmatrix} (\tilde{\boldsymbol{X}}_1\boldsymbol{V}_\beta\tilde{\boldsymbol{X}}_1' + \boldsymbol{R})^{-1} & & \\ & \ddots & \\ & & (\tilde{\boldsymbol{X}}_H\boldsymbol{V}_\beta\tilde{\boldsymbol{X}}_H' + \boldsymbol{R})^{-1} \end{pmatrix}.
$$

Since

$$
\begin{aligned}
&\pi(\boldsymbol{Y}|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{R}, \boldsymbol{V}_\beta) \\
&\propto \exp\left\{-\frac{1}{2}\boldsymbol{Y}'(\boldsymbol{\Psi} - \boldsymbol{\Psi\Upsilon}(\boldsymbol{\Upsilon}'\boldsymbol{\Psi\Upsilon})^{-1}\boldsymbol{\Upsilon}'\boldsymbol{\Psi})\boldsymbol{Y}\right\} \prod\nolimits_h [I\{\boldsymbol{y}_{h[I_h]} > \max(\boldsymbol{y}_{h[-I_h]})\}].
\end{aligned}
$$

So, $\boldsymbol{Y}|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{R}, \boldsymbol{V}_\beta \sim TN(0, (\boldsymbol{\Psi} - \boldsymbol{\Psi\Upsilon}(\boldsymbol{\Upsilon}'\boldsymbol{\Psi\Upsilon})^{-1}\boldsymbol{\Upsilon}'\boldsymbol{\Psi})^{-1})$ with truncation restriction on $I\{\boldsymbol{y}_{h[I_h]} > \boldsymbol{y}_{h[-I_h]}\}$ for any $h = 1, \ldots, H$. Thus, the posterior distribution of $\boldsymbol{Y}|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{R}, \boldsymbol{V}_\beta$ is proper.

- Since the space of $m$-dimensional correlation matrix is convex and compact (Rousseeuw & Molenberghs, 1994), the prior distribution we specify on $\boldsymbol{R}$ is a proper distribution. Thus, the conditional posterior distribution $\pi(\boldsymbol{R}|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{V}_\beta)$ is also proper.
- Similar to the argument given above, because the prior distribution on $\boldsymbol{V}_\beta^{-1}$ is Wishart which is a proper distribution, the conditional distribution of $\boldsymbol{V}_\beta|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}$ is proper.

## 9. Appendix 4

### 9.1. Proof of Theorem 2

When $T_h = 1$, the joint posterior distribution is

$$\pi(\boldsymbol{Y}, \boldsymbol{B}, \boldsymbol{\Delta}, \boldsymbol{\Lambda}, \boldsymbol{V}_{\boldsymbol{\beta}} | \boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}) = \left[\prod_{h=1}^{H} \pi(\boldsymbol{\beta}_h | \boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{y}_{\boldsymbol{h}}, \boldsymbol{\Delta}, \boldsymbol{\Lambda}, \boldsymbol{V}_{\boldsymbol{\beta}})\right] \pi(\boldsymbol{\Delta} | \boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{\Lambda}, \boldsymbol{V}_{\boldsymbol{\beta}})$$
$$\times \pi\left(\boldsymbol{\Lambda} | \boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{V}_{\boldsymbol{\beta}}\right) \pi\left(\boldsymbol{Y} | \boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{V}_{\boldsymbol{\beta}}\right) \pi\left(\boldsymbol{V}_{\boldsymbol{\beta}} | \boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}\right).$$

We will show that the conditional distribution $\pi(\boldsymbol{\Lambda} | \boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{V}_{\beta})$ is improper. Recall that $\boldsymbol{\Lambda} = \text{diag}(1, \sigma_2^2, \ldots, \sigma_m^2)$, and we let

$$\boldsymbol{\Psi}^* = \begin{pmatrix} (\tilde{\boldsymbol{X}}_1 \boldsymbol{V}_{\beta} \tilde{\boldsymbol{X}}_{1'} + \boldsymbol{\Lambda})^{-1} & & \\ & \ddots & \\ & & (\tilde{\boldsymbol{X}}_H \boldsymbol{V}_{\beta} \tilde{\boldsymbol{X}}_{H'} + \boldsymbol{\Lambda})^{-1} \end{pmatrix}.$$

Since the conditional distribution is proportional to the prior of $\boldsymbol{\Lambda}$ multiplied by the corresponding likelihood function, we have (see the proof of Theorem 1)

$$\begin{aligned} &\pi(\boldsymbol{\Lambda} | \boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{V}_{\beta}) \\ &= \boldsymbol{C}_1 \prod_{j=2}^{m} \left[(\sigma_j^{-2})^{\nu-1} \exp\left\{-\frac{\sigma_j^{-2}}{v_j}\right\}\right] \left(\frac{|\boldsymbol{\Psi}^*|}{|\boldsymbol{\Upsilon}' \boldsymbol{\Psi}^* \boldsymbol{\Upsilon}|}\right)^{\frac{1}{2}} \\ &exp\left\{-\frac{1}{2} \boldsymbol{Y}'(\boldsymbol{\Psi}^* - \boldsymbol{\Psi}^* \boldsymbol{\Upsilon}(\boldsymbol{\Upsilon}' \boldsymbol{\Psi}^* \boldsymbol{\Upsilon})^{-1} \boldsymbol{\Upsilon}' \boldsymbol{\Psi}^*) \boldsymbol{Y}\right\}, \end{aligned}$$

where $\boldsymbol{C}_1$ is the normalizing constant. Again, we set the variance of the Gamma distribution to be large to represent vague information. So, in the limit as we let $\nu \to 0$ and $v_j \to \infty$, the above posterior distribution becomes

$$\begin{aligned} &\pi(\boldsymbol{\Lambda} | \boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{V}_{\beta}) \\ &= \boldsymbol{C}_1 \left[\prod_{j=2}^{m} (\sigma_j^{-2})^{-1}\right] \left(\frac{|\boldsymbol{\Psi}^*|}{|\boldsymbol{\Upsilon}' \boldsymbol{\Psi}^* \boldsymbol{\Upsilon}|}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2} \boldsymbol{Y}'(\boldsymbol{\Psi}^* - \boldsymbol{\Psi}^* \boldsymbol{\Upsilon}(\boldsymbol{\Upsilon}' \boldsymbol{\Psi}^* \boldsymbol{\Upsilon})^{-1} \boldsymbol{\Upsilon}' \boldsymbol{\Psi}^*) \boldsymbol{Y}\right\}. \end{aligned}$$

Let

$$\begin{aligned} f(\sigma^{-2}) &= \prod_{j=2}^{m} (\sigma_j^{-2})^{-1} \\ g(\sigma^{-2}) &= \left(\frac{|\boldsymbol{\Psi}^*|}{|\boldsymbol{\Upsilon}' \boldsymbol{\Psi}^* \boldsymbol{\Upsilon}|}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2} \boldsymbol{Y}'(\boldsymbol{\Psi}^* - \boldsymbol{\Psi}^* \boldsymbol{\Upsilon}(\boldsymbol{\Upsilon}' \boldsymbol{\Psi}^* \boldsymbol{\Upsilon})^{-1} \boldsymbol{\Upsilon}' \boldsymbol{\Psi}^*) \boldsymbol{Y}\right\}. \end{aligned}$$

Then,

$$\begin{aligned}
&\int \pi(\boldsymbol{\Lambda}|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{V}_\beta)\mathrm{d}\boldsymbol{\Lambda} \\
&= \int_0^\infty \cdots \int_0^\infty \pi(\sigma_2^{-2}, \ldots, \sigma_m^{-2}|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{V}_\beta)\mathrm{d}\sigma_2^{-2} \ldots \mathrm{d}\sigma_m^{-2} \\
&\geq \int_w^\infty \ldots \int_w^\infty \pi(\sigma_2^{-2}, \ldots, \sigma_m^{-2}|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{V}_\beta)\mathrm{d}\sigma_2^{-2} \ldots \mathrm{d}\sigma_m^{-2}, \quad (\forall w > 0) \\
&= \boldsymbol{C}_1 \int_w^\infty \ldots \int_w^\infty f(\sigma^{-2}) g(\sigma^{-2})\mathrm{d}\sigma_2^{-2} \ldots \mathrm{d}\sigma_m^{-2}.
\end{aligned}$$

Let $F$ be the region defined by $\{\sigma_j^{-2} > w, \forall j = 2, \ldots, m\}$. It is obvious that $g(\sigma^{-2}) > 0$ over $F$ with the only possible exception when $\sigma_j^{-2}$ approaches infinity. However, we show that $g(\sigma^{-2}) \neq 0$ in the limiting case. First, when $\sigma_j^{-2} \to \infty$ $(j = 2, \ldots, m)$, $(\tilde{\boldsymbol{X}}_j \boldsymbol{V}_\beta \tilde{\boldsymbol{X}}'_j + \boldsymbol{\Lambda})^{-1} \to (\tilde{\boldsymbol{X}}_j \boldsymbol{V}_\beta \tilde{\boldsymbol{X}}'_j + \ddot{\boldsymbol{E}})^{-1}$, where $\ddot{E}$ is a $m \times m$ zero matrix except that its (1, 1) component is 1. Then,

$$\boldsymbol{\Psi}^* \to \begin{pmatrix} (\tilde{\boldsymbol{X}}_{\boldsymbol{1}} \boldsymbol{V}_{\boldsymbol{\beta}} \tilde{\boldsymbol{X}}'_{\boldsymbol{1}} + \ddot{\boldsymbol{E}})^{-1} & & \\ & \ddots & \\ & & (\tilde{\boldsymbol{X}}_{\boldsymbol{H}} \boldsymbol{V}_{\boldsymbol{\beta}} \tilde{\boldsymbol{X}}'_{\boldsymbol{H}} + \ddot{\boldsymbol{E}})^{-1} \end{pmatrix} \triangleq \boldsymbol{\Psi}^*_\infty.$$

Therefore,

$$g(\sigma^{-2}) \to \left(\frac{|\boldsymbol{\Psi}^*_\infty|}{|\boldsymbol{\Upsilon}'\boldsymbol{\Psi}^*_\infty\boldsymbol{\Upsilon}|}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\boldsymbol{Y}'(\boldsymbol{\Psi}^*_\infty - \boldsymbol{\Psi}^*_\infty\boldsymbol{\Upsilon}(\boldsymbol{\Upsilon}'\boldsymbol{\Psi}^*_\infty\boldsymbol{\Upsilon})^{-1}\boldsymbol{\Upsilon}'\boldsymbol{\Psi}^*_\infty)\boldsymbol{Y}\right\} > 0.$$

Hence, $\min_{\sigma^{-2} \in F} g(\sigma^{-2}) = m_0$, where $0 < m_0 < \infty$. Thus

$$\begin{aligned}
\int \pi(\boldsymbol{\Lambda}|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{V}_\beta)d\boldsymbol{\Lambda} &\geq \boldsymbol{C}_1 \int_w^\infty \ldots \int_w^\infty f(\sigma^{-2}) g(\sigma^{-2})\mathrm{d}\sigma_2^{-2} \ldots \mathrm{d}\sigma_m^{-2} \\
&\geq m_0 \boldsymbol{C}_1 \int_w^\infty \cdots \int_w^\infty \left[\prod\nolimits_{j=2}^m (\sigma_j^{-2})^{-1}\right]\mathrm{d}\sigma_2^{-2} \ldots \mathrm{d}\sigma_m^{-2} \\
&= m_0 \boldsymbol{C}_1 \prod\nolimits_{j=2}^m [\ln\sigma_j^{-2}|_w^\infty] = \infty.
\end{aligned}$$

Because $\int \pi(\boldsymbol{\Lambda}|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{V}_\beta)d\boldsymbol{\Lambda}$ does not exist, $\pi(\boldsymbol{\Lambda}|\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{V}_\beta)$ is *not* a proper distribution, and so the joint posterior distribution is improper.

## References


Addelman, S. (1962). Orthogonal main-effect plans for asymmetrical factorial experiments. *Technometrics, 4*, 21–46.

Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, *88*(422), 669–679.

Barnard, J., McCulloch, R., & Meng, X. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with applications to shrinkage. *Statistica Sinica*, *10*, 1281–1311.

Burgette, L. F., & Nordheim, E. V. (2012). The trace restriction: An alternative identification strategy for the Bayesian multinomial probit model. *Journal of Business and Economic Statistics*, *30*(3), 404–410.

Chib, S., & Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, *85*(2), 347–361.

Chib, S., Greenberg, E., Chen, Y. (1998). *MCMC methods for fitting and comparing multinomial response models*. Working Paper, Olin School of Business, Washington University.

Daganzo, C. (1980). *Multinomial probit*. New York: Academic Press.

Dawid, A. (1981). Some matrix-variate distribution theory: Notational considerations and a Bayesian application. *Biometrika*, *68*(1), 265–274.
DeSarbo, W. S. (1982). GENNCLUS: New models for general nonhierarchical clustering analysis. *Psychometrika*, *47*(4), 449–475.
DeSarbo, W. S., & Carroll, J. D. (1985). Three-way metric unfolding via alternating weighted least squares. *Psychometrika*, *50*(3), 275–300.
DeSarbo, W. S., & Cron, W. L. (1988). A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification*, *5*(2), 249–282.
Dotson, J., Lenk, P., Brazell, J., Otter, T., MacEachern, S., Allenby G. M. (2010). A probit model with structured covariances for similarity effects and source of volume calculations. Working paper.
Fiebig, D. G., Keane, M. P., Louviere, J., & Wasi, N. (2010). The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. *Marketing Science*, *29*, 393–421.
Fong, D. K. H., Ebbes, P., & DeSarbo, W. S. (2012). A heterogeneous Bayesian regression model for cross sectional data involving a single observation per response unit. *Psychometrika*, *77*(2), 293–314.
Gupta, A. K., & Nagar, D. K. (2000). *Matrix variate distributions. Monographs and surveys in pure and applied mathematics* (Vol. 104). London: Chapman & Hall/CRC.
Hausman, J., & Wise, D. (1978). A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica*, *45*, 319–339.
Hobert, J. P., & Marchev, D. (2008). A theoretical comparison of the data augmentation, marginal augmentation and px-da algorithms. *Annals of Statistics*, *36*(2), 532–554.
Imai, K., & van Dyk, D. A. (2005). A Bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of Econometrics*, *124*(2), 311–334.
Jedidi, K., & DeSarbo, W. S. (1991). A stochastic multidimensional scaling procedure for the spatial representation of three-mode, three-way pick any/J data. *Psychometrika*, *56*(3), 471–494.
Liechty, J. C., Liechty, M. W., & Muller, P. (2004). Bayesian correlation estimation. *Biometrika*, *91*(1), 1–14.
Liu, C. (2001). Discussion on the art of data augmentation. *Journal of Computational and Graphical Statistics*, *10*(1), 75–81.
Liu, J. S., & Wu, Y. N. (1999). Parameter expansion for data augmentation. *Journal of the American Statistical Association*, *94*(448), 1264–1274.
Liu, X., & Daniels, M. (2006). A new efficient algorithm for sampling a correlation matrix based on parameter expansion and re-parameterization. *Journal of Computational and Graphical Statistics*, *15*(4), 897–914.
Maydeu-Olivares, A., & Hernández, A. (2007). Identification and small sample estimation of Thurstone's unrestricted model for paired comparisons data. *Multivariate Behavioral Research*, *42*(2), 323–347.
McCulloch, R., & Rossi, P. E. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, *64*, 207–240.
McCulloch, R. E., Polson, N. G., & Rossi, P. E. (2000). A Bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics*, *99*(1), 173–193.
Nobile, A. (1998). A hybrid Markov chain for the Bayesian analysis of the multinomial probit model. *Statistics and Computing*, *8*, 229–242.
Nobile, A. (2000). Comment: Bayesian multinomial probit models with normalization constraint. *Journal of Econometrics*, *99*(1), 335–345.
Rossi, P. E., Allenby, G. M., & McCulloch, R. E. (2005). *Bayesian statistics and marketing*. Chichester: Wiley.
Rossi, P. E., McCulloch, R. E., & Allenby, G. M. (1996). The value of purchase history data in target marketing. *Marketing Science*, *15*(4), 321–340.
Rousseeuw, P., & Molenberghs, G. (1994). The shape of correlation matrices. *The American Statistician*, *48*, 276–279.
Thurstone, L. (1927). A law of comparative judgment. *Psychological Review*, *34*, 273–286.
Tsai, R. (2000). Remarks on the identifiability of Thurstonian ranking models: Case V, case III, or neither? *Psychometrika*, *65*(2), 233–240.
Tsai, R. (2003). Remarks on the identifiability of Thurstonian paired comparison models under multiple judgment. *Psychometrika*, *68*(1), 361–372.

*Manuscript Received: 13 JULY 2013*
*Published Online Date: 10 DEC 2014*