PSYCHOMETRIKA—VOL. 67, NO. 1, 49–78
MARCH 2002

HETEROGENEOUS FACTOR ANALYSIS MODELS: A BAYESIAN APPROACH

ASIM ANSARI AND KAMEL JEDIDI

COLUMBIA UNIVERSITY

LAURETTE DUBE

MCGILL UNIVERSITY

Multilevel factor analysis models are widely used in the social sciences to account for heterogeneity in mean structures. In this paper we extend previous work on multilevel models to account for general forms of heterogeneity in confirmatory factor analysis models. We specify various models of mean and covariance heterogeneity in confirmatory factor analysis and develop Markov Chain Monte Carlo (MCMC) procedures to perform Bayesian inference, model checking, and model comparison.

We test our methodology using synthetic data and data from a consumption emotion study. The results from synthetic data show that our Bayesian model perform well in recovering the true parameters and selecting the appropriate model. More importantly, the results clearly illustrate the consequences of ignoring heterogeneity. Specifically, we find that ignoring heterogeneity can lead to sign reversals of the factor covariances, inflation of factor variances and underappreciation of uncertainty in parameter estimates. The results from the emotion study show that subjects vary both in means and covariances. Thus traditional psychometric methods cannot fully capture the heterogeneity in our data.

Key words: confirmatory factor analysis, multilevel models, random coefficient models, MCMC methods, Gibbs sampling, Metropolis-Hastings.

1. Introduction

Confirmatory factor analysis is widely used by researchers in the social sciences to identify constructs and to model the relationship between these latent constructs and observed variables. Many applications of confirmatory factor analysis assume a common model for all individuals and therefore implicitly ignore the influence of unobserved heterogeneity although it is unlikely that all individuals in the sample have the same set of parameters. It is well known that ignoring to account for unobserved heterogeneity can lead to biased parameter estimates and therefore can yield distorted results. Using a hypothetical example where groups of subjects have common measurement parameters but have different variable means, Muthén (1989) shows that ignoring this type of heterogeneity leads to inflated measurement reliability. His result is consistent with the classic results of Lord and Novick (1968, pp. 129–131) on effects of group heterogeneity and selection on test reliability. Other types of distortions are possible depending upon the nature and extent of heterogeneity in data.

Several extensions of the basic confirmatory factor analysis model have appeared in the psychometric literature for treating population heterogeneity. Jöreskog (1971) and Sörbom (1981) developed the multiple group confirmatory factor analysis model to treat the situation where data come from several a priori identified groups that are heterogeneous in their factor structure. Using a finite mixture approach, Bladfield (1980) and more recently, Jedidi, Jagpal, and DeSarbo (1997a, 1997b), Yung (1997), Arminger and Stein (1997), Arminger, Stein and Wittenberg (1999) generalized the multigroup model to handle the case where the groups are unobserved a priori. The multigroup and finite mixture approaches work well with few groups and typically require a large number of observations per group. Contrary to these fixed effects approaches, a number of researchers (Goldstein & McDonald, 1988; Kaplan & Elliot, 1997;

Requests for reprints should be sent to Asim Ansari, 517 Uris Hall, Columbia University, 3022 Broadway, New York, NY, 10027. E-Mail: maa48@columbia.edu

0033-3123/2002-1/2000-0790-A $00.75/0

Longford & Muthén, 1992; McDonald & Goldstein, 1989; Muthén, 1989, 1994; Muthén & Satorra, 1989) have championed multilevel covariance structure models for the random effects analysis of hierarchical data (e.g., achievement data obtained from students sampled from within classrooms and classrooms sampled from within schools). As we show later in the paper, unlike the multigroup model and its finite mixture extensions, multilevel models capture heterogeneity only in mean structures (i.e., measurement intercepts) but not in covariance structures (i.e., the variances and covariances of the factors and measurement errors).

In this paper, we extend previous work on multilevel models by describing general forms of heterogeneity in confirmatory factor analysis models. Specifically, we describe how mean and covariance heterogeneity can be implemented in confirmatory factor analysis models using a hierarchical Bayesian framework. We develop Markov Chain Monte Carlo (MCMC) procedures for sampling based Bayesian inference in such models. The hierarchical Bayesian approach allows for appropriate pooling of the data while taking into account heterogeneity and is particularly suitable for studies in which multilevel data, panel data or multiple observations are available for a given set of subjects or objects (e.g., schools). Such data are commonly collected in the social sciences. For example, in the educational achievement studies, data are available for several students within a classroom and for several classrooms within a school. Similarly, in consumer psychology, firms use consumer panels for tracking consumer satisfaction and perceptions over time.

The hierarchical Bayesian approach for modeling heterogeneity provides several theoretical and practical advantages. From a practical viewpoint, Bayesian methods allow the estimation of individual-specific parameters (such as factor scores) while accounting properly for the uncertainty in such estimates. Moreover, as we discuss later, by using MCMC procedures, simulation-based estimates of the parameters can be obtained. This circumvents the need for evaluating complex multidimensional integrals that are often required to implement maximum likelihood methods for heterogeneous data. From a statistical viewpoint, sampling-based Bayesian methods are appealing because they do not rely on asymptotic theory. In addition, Bayesian methods allow for the incorporation of prior information where available.

There is a rich literature on Bayesian modeling of covariance structure models. Martin and McDonald (1975) provides an early illustration of Bayesian techniques for the factor analysis model and Lee (1981) focuses on the use of different prior distributions, whereas Bartholomew (1981) and Shi and Lee (1997) illustrate the use of Bayesian analysis to derive posterior estimates of factor scores, given point estimates of the other parameters. More recently, Arminger and Muthén (1998) develop Bayesian methods for handling nonlinear models. Scheines, Hoijtink and Boomsma (1999) describe MCMC methods for covariance structure models and Ansari and Jedidi (2000) describe Bayesian multilevel models for binary data.

The rest of the paper is organized as follows. Section 2 discusses the hierarchical Bayesian confirmatory factor analysis model. Section 3 discusses the specification of the priors and describes the MCMC method for inference. Section 4 discusses procedures for model adequacy and model comparison. Section 5 reports the results of two studies involving synthetic data. Section 6 illustrates our methodology using emotion data obtained from a consumer panel. Section 7 summarizes the paper and discusses directions for future research.

2. Model

We begin by describing a general random coefficient confirmatory factor analysis model. Suppose data are available from I individuals. Let individual i provide $j = 1$ to n_i multivariate observations on a p dimensional vector $\mathbf{y}_{ij}$ of indicator variables[1]. The total number of observations across all individuals is then given by $N = \sum_i n_i$ and since n_i can vary across indi-

[1] We depart from the standard multilevel nomenclature by using the term individuals for level-two units (i.e., groups) and the term observations for level-one units.

viduals, the data are unbalanced. The variation in the observed variables $\mathbf{y}_{ij}$ can be explained in terms of m latent factors. A confirmatory factor analysis model can be written for individual i as $\mathbf{y}_{ij} = \boldsymbol{\alpha}_i + \Lambda_i \boldsymbol{\xi}_{ij} + \boldsymbol{\epsilon}_{ij}$, for $j = 1$ to n_i. The vector $\boldsymbol{\alpha}_i$ contains the p measurement intercepts, $\boldsymbol{\xi}_{ij}$ contains the m latent factors and Λ_i is a $p \times m$ matrix of factor loadings. The latent factors $\boldsymbol{\xi}_{ij}$ are assumed to be normally distributed $N(\boldsymbol{\nu}_i, \Phi_i)$, where $\boldsymbol{\nu}_i$ contains the m factor means and Φ_i is a $m \times m$ positive definite covariance matrix of the factors. Finally, the measurement errors in $\boldsymbol{\epsilon}_{ij}$ are assumed to be normal, $N(\mathbf{0}, \Theta_i)$, where Θ_i as usual is a $p \times p$ diagonal matrix containing the measurement error variances.

If a large number of observations are available per individual, then we can perform separate factor analyses for each individual. In many research contexts, this is not feasible because of the scarcity of observations. In such situations, a random coefficient approach can be used to appropriately pool information across individuals and to account for the unobserved sources of heterogeneity (Longford, 1993, Muthén 1989). In the random coefficient approach, a second stage model specifies how the parameters from the individual level factor analysis models vary across the different individuals. The second stage combines the individual level models using a population distribution over the parameters. The set of parameters $\boldsymbol{\omega}_i = \{\boldsymbol{\alpha}_i, \Lambda_i, \boldsymbol{\nu}_i, \Phi_i, \Theta_i\}$ of the factor analysis model for individual i can be assumed to come from a population distribution $f(\boldsymbol{\Omega})$. Combining the individual level models with the population model, the generic form of the random coefficient factor analysis model can then be written as follows:

$$\begin{aligned}
\mathbf{y}_{ij} &= \boldsymbol{\alpha}_i + \Lambda_i \boldsymbol{\xi}_{ij} + \boldsymbol{\epsilon}_{ij}, \\
\boldsymbol{\xi}_{ij} &= N(\boldsymbol{\nu}_i, \Phi_i) \\
\boldsymbol{\epsilon}_{ij} &= N(\mathbf{0}, \Theta_i) \\
\boldsymbol{\omega}_i &\sim f(\boldsymbol{\Omega}), \quad i = 1 \text{ to } I, \quad j = 1 \text{ to } n_i.
\end{aligned} \tag{1}$$

It is important to note that the above approach assumes that the individual specific factor analysis models are form invariant (Bollen, 1989) which implies that the collection of individual level models all share the same form (i.e., they all have the same number of factors, the same form of parameter matrices with the same dimensions and have the same locations of free, constrained and fixed parameters).

A general random coefficient model that allows across individual variation in all parameters as in (1), is not identified. Identifiability necessitates restrictions on the form and the variability of some of the parameters. In this paper, we choose the nature of these restrictions to yield a set of identified models that highlight different sources of across individual heterogeneity. Specifically, we differentiate between models that are heterogeneous in mean structures alone and those that also include heterogeneity in covariance structures.

2.1. Heterogeneity in Mean Structures

We begin by describing two models that allow heterogeneity in mean structures across the individuals.

2.1.1. Heterogeneous Factor Means

In many situations, researchers are interested in modeling individual differences in latent constructs. For example, in emotion studies, researchers are interested in profiling different individuals in terms of their average levels of different types of emotions. Such individual differences can be modeled using different factor means $\boldsymbol{\nu}_i$ for the individuals. Different factor means result in different mean structures for the individuals. As heterogeneity in mean structures is introduced using different factor means, the measurement intercepts $\boldsymbol{\alpha}_i$ can be assumed to be

invariant, that is, $\boldsymbol{\alpha}_i = \boldsymbol{\alpha}$ for $i = 1$ to I. In this particular model, as the focus is on studying variation in mean structures alone, we assume that all individuals have common factor loadings Λ, common factor covariance matrices Φ, and common measurement variances Θ. The implied mean vector and covariance matrix for an arbitrary observation for individual i is then given by $E[\mathbf{y}_{ij} \mid \boldsymbol{\nu}_i] = \boldsymbol{\alpha} + \Lambda\boldsymbol{\nu}_i$ and $V[\mathbf{y}_{ij} \mid \boldsymbol{\nu}_i] = \Lambda\Phi\Lambda' + \Theta$ respectively. The complete model for an arbitrary individual, i, can then be written as

$$\begin{aligned} \mathbf{y}_{ij} &= \boldsymbol{\alpha} + \Lambda\boldsymbol{\xi}_{ij} + \boldsymbol{\epsilon}_{ij}, \\ \boldsymbol{\xi}_{ij} &\sim N(\boldsymbol{\nu}_i, \Phi), \qquad j = 1 \text{ to } n_i \\ \boldsymbol{\epsilon}_{ij} &\sim N(\mathbf{0}, \Theta). \end{aligned} \tag{2}$$

As is usual in confirmatory factor analysis, certain restrictions are required on the parameters of the above model for identification purposes. Specifically, the loading matrix Λ usually has a patterned structure with certain elements set to zero. Given the arbitrary scaling of the latent factors, restrictions are needed either on Λ (e.g., by setting the scale of each factor to the scale of an *a priori* chosen variable), or alternatively, on Φ, by assuming it to be a correlation matrix. If restrictions are imposed on Λ then Φ is a covariance matrix.

The population distribution specifying the heterogeneity in the individual-level factor means can be written as

$$\boldsymbol{\nu}_i \sim N(\mathbf{0}, \Delta), \tag{3}$$

for $i = 1$ to I. In (3), the factor means $\boldsymbol{\nu}_i$ for each individual, come from a population normal distribution with mean zero and a covariance matrix, Δ. We assume a zero mean for this distribution in order to fix the location of the grand mean and to ensure identification of parameters. This is analogous to setting the factor means of the first group to zero in a multigroup (i.e., fixed effects) factor analysis model. The covariance matrix Δ describes the covariation in the factor means across individuals (between individuals variation). Taking into account the differences in factor means across individuals, the unconditional mean for an arbitrary observation can be written as $E[E[\mathbf{y} \mid \boldsymbol{\nu}_i]] = \boldsymbol{\alpha}$ and the unconditional covariance can be written as

$$\begin{aligned} V[\mathbf{y}_{ij}] &= E[V(\mathbf{y}_{ij} \mid \boldsymbol{\nu}_i)] + V[E(\mathbf{y}_{ij} \mid \boldsymbol{\nu}_i)] \\ &= E[\Lambda\Phi\Lambda' + \Theta] + V[\boldsymbol{\alpha} + \Lambda\boldsymbol{\nu}_i] \\ &= \Lambda(\Phi^{Agg})\Lambda' + \Theta, \end{aligned} \tag{4}$$

where $\Phi^{Agg} = \Phi + \Delta$ combines the within and between sources of variation in the factors. Thus an analysis that ignores heterogeneity may recover Λ and Θ but will fail to separate the two components of Φ^{Agg}.

In other words, ignoring the heterogeneity in factor means can yield misleading inferences regarding the factor structure of the underlying latent constructs. Usually, researchers are interested in studying the within individual factor structure represented in Φ. For example, researchers may wish to investigate how different types of emotional states covary. As mentioned earlier, if heterogeneous data are analyzed using a conventional factor analysis model that ignores heterogeneity, the estimated factor covariance matrix Φ^{Agg} from such a model will confound Φ and Δ. Consequently, two types of misleading inferences regarding the elements in Φ will result from using such an analysis. First, as the diagonal elements in Φ^{Agg} are necessarily larger than the corresponding elements in Φ, factor reliability estimates will be inflated (Lord and Novick, 1968 pp. 129–131, Muthén 1989). Second, the magnitude and the signs of the factor covariances in Φ^{Agg} will be distorted. For example, if the elements Φ_{ij} and Δ_{ij} are of the same sign, then the

magnitude of Φ_{ij}^{Agg} would be amplified. Alternatively, if Φ_{ij} and Δ_{ij} are of opposite sign, then Φ_{ij}^{Agg} may get attenuated or may indeed have the wrong sign.

2.1.2. *Heterogeneous Intercepts*

Heterogeneity in mean structures can alternatively be modeled by assuming that individuals differ in their measurement intercepts $\boldsymbol{\alpha}_i$. Such an approach is suitable when researchers are interested in modeling the differences in scale usage across individuals. As heterogeneity in mean structures is modeled in terms of the measurement intercepts, we need to restrict the individual factor means for identification. We therefore assume that the factor means for each individual are zero, that is, $\boldsymbol{\nu}_i = 0$, for $i = 1$ to I. The implied mean vector for an arbitrary observation for individual i is then given by $E[\mathbf{y}_{ij} \mid \boldsymbol{\alpha}_i] = \boldsymbol{\alpha}_i$. As in the previous model, we assume that all individuals have a common loadings matrix Λ, a common factor covariance matrix Φ, and common measurement variances Θ. The implied covariance matrix for an observation belonging to the individual is given by $V[\mathbf{y}_{ij} \mid \boldsymbol{\alpha}_i] = \Lambda\Phi\Lambda' + \Theta$. The model for an arbitrary individual i, can then be written as

$$\begin{aligned} \mathbf{y}_{ij} &= \boldsymbol{\alpha}_i + \Lambda\boldsymbol{\xi}_{ij} + \boldsymbol{\epsilon}_{ij}, \\ \boldsymbol{\xi}_{ij} &\sim N(\mathbf{0}, \Phi), \\ \boldsymbol{\epsilon}_{ij} &\sim N(\mathbf{0}, \Theta), \end{aligned} \tag{5}$$

for observations $j = 1$ to n_i. The usual restrictions on Λ or Φ can be placed as in the previous model for identification purposes.

The second stage population distribution that specifies the heterogeneity in the individual-level intercept parameters can be written as

$$\boldsymbol{\alpha}_i \sim N(\boldsymbol{\mu}, \Sigma_b). \tag{6}$$

The intercepts for the different individuals are assumed to originate from a multivariate normal distribution with population mean $\boldsymbol{\mu}$ and a $(p \times p)$ covariance matrix Σ_b. If interest is in studying the variation across individuals then a factor analytic structure can be further imposed on Σ_b such that $\Sigma_b = \Lambda_b\Phi_b\Lambda_b' + \Theta_b$. Such a model is known as the multilevel model in the psychometric literature (Longford & Muthén, 1992).

Taking into account the measurement intercepts across individuals, the unconditional mean for an arbitrary observation can be written as $E[E[\mathbf{y}_{ii} \mid \boldsymbol{\alpha}_i]] = \boldsymbol{\mu}$ and the unconditional covariance is given by $V[\mathbf{y}_{ij}] = E[V(\mathbf{y}_{ij} \mid \boldsymbol{\alpha}_i)] + V[E(\mathbf{y}_{ij} \mid \boldsymbol{\alpha}_i)] = E[\Lambda\Phi\Lambda' + \Theta] + V[\boldsymbol{\alpha}_i] = \Lambda\Phi\Lambda' + \Theta + \Sigma_b$. The form of the unconditional covariance matrix indicates that if we analyze data originating from such a model using traditional methods that ignore heterogeneity, then misleading inferences regarding Λ, Φ and Θ may result. For example, if $\Sigma_b = \Lambda\Phi_b\Lambda' + \Theta_b$ (i.e., common loadings across levels) then $V[\mathbf{y}_{ij}] = \Lambda(\Phi + \Phi_b)\Lambda' + \Theta + \Theta_b$. Thus, in this special case, conventional confirmatory factor analysis will confound both the factor and the measurement error covariance matrices.

2.2. *Heterogeneity in Covariance Structures*

The first two models focussed solely on the variation in means and assumed invariant covariance structures across all individuals. Specifically, we assumed that the factor loading matrices Λ, the factor covariance matrices Φ and the measurement variances Θ are invariant across individuals. Here we relax these assumptions and focus on characterizing heterogeneity by accounting for the variation in the factor structures and measurement variances. We describe two forms of heterogeneous factor structures. In each case we assume that the measurement variances Θ_i and the mean structures are different across individuals.

2.2.1. *Heterogeneous Factor Loadings*

In characterizing heterogeneity in factor structures, researchers may be interested in specifying different factor to variable transformation mechanisms for different individuals (see Yung, 1997). Even if two individuals have the same true scores on the latent factors, they may have different scores on the manifest variables due to differences in their loadings for the factors. To model such situations, the loadings matrices, Λ_i, can be allowed to vary across individuals. To ensure appropriate comparisons across individuals and to limit attention to variation in factor loading matrices, variance of each factor can be fixed to one and a common factor correlation matrix Φ for each individual can be assumed. The heterogeneity in the means can be captured by either allowing the factor means to vary or by allowing the measurement intercepts to differ across individuals. For example, choosing the former variation in means, the implied mean vector for an arbitrary observation for individual i is given as $E[\mathbf{y}_{ij} \mid \Lambda_i, \boldsymbol{\nu}_i] = \boldsymbol{\alpha} + \Lambda_i \boldsymbol{\nu}_i$ and the implied covariance matrix is given by $V[\mathbf{y}_{ij} \mid \Lambda_i, \Theta_i] = \Lambda_i \Phi \Lambda_i' + \Theta_i$. The resulting model for individual i can be represented by the following set of equations:

$$\begin{aligned} \mathbf{y}_{ij} &= \boldsymbol{\alpha} + \Lambda_i \boldsymbol{\xi}_{ij} + \boldsymbol{\epsilon}_{ij}, \\ \boldsymbol{\xi}_{ij} &\sim N(\boldsymbol{\nu}_i, \Phi), \qquad j = 1 \text{ to } n_i \\ \boldsymbol{\epsilon}_{ij} &\sim N(\mathbf{0}, \Theta_i). \end{aligned} \tag{7}$$

The population distribution that describes the variation in the individual level parameters can be specified in terms of the product of independent population distributions over $\boldsymbol{\nu}_i$, Λ_i and Θ_i. The population distributions can be written as

$$\begin{aligned} \boldsymbol{\nu}_i &\sim N(\mathbf{0}, \Delta) \\ \Lambda_i &\sim \prod_{km} N(\lambda_{km}, \kappa_{km}), \qquad i = 1 \text{ to } I \\ \Theta_i &\sim \prod_{k=1}^{p} IG(a_k, b_k). \end{aligned} \tag{8}$$

The population distribution over $\boldsymbol{\nu}_i$ is as in (3). The population distribution for Λ_i specifies that each nonzero element (row k, column m) of the loadings matrix is independently distributed univariate normal with mean λ_{km} and variance κ_{km}. The third population distribution specifies how the p measurement error variances vary in the population of individuals. We choose the product of independent inverse gamma $IG(a, b)$ distributions for this population distribution.

Taking into account the variation across individuals, the unconditional mean structure can be written as $E[\mathbf{y}_{ii}] = E[E[\mathbf{y}_{ii} \mid \Lambda_i, \boldsymbol{\nu}_i]] = \boldsymbol{\alpha}$, and the unconditional variance can be written as $V[\mathbf{y}_{ii}] = E[\Lambda_i(\Phi + \Delta)\Lambda_i'] + E[\Theta_i]$. $E[\Theta_i]$ is a diagonal matrix whose k-th diagonal element is obtained from the mean $1/[b_k(a_k - 1)]$, of the corresponding inverse gamma distribution $IG(a_k, b_k)$. The term $E[\Lambda_i(\Phi + \Delta)\Lambda_i]$ depends upon the specific form of the Λ_i matrix and can be easily computed given our population distribution. It is clear from the implied variance that if the true data generating process involves heterogeneous covariance structures, then misleading inferences pertaining to factor loadings, factor covariances and the measurement variances may result when models ignoring heterogeneity are used.

2.2.2. *Heterogeneous Factor Covariances*

An alternative characterization of heterogeneity in factor structures can be made by specifying different covariation patterns, Φ_i, of the latent factors across different individuals. For example, individuals can differ in the extent to which the different types of emotional states covary with each other. For some individuals, the latent factors may be near orthogonal, whereas

for others the same factors may be highly correlated. In order to make meaningful comparisons across the individuals, the factor loading matrices can be assumed to be invariant. The heterogeneity in mean levels can be captured either through the factor means or through heterogeneous measurement intercepts. Here, we model heterogeneous factor means. The heterogeneous factor covariances model with heterogeneous intercepts is described in Appendix 2. The implied mean vector for an arbitrary observation for individual i is then given by $E[\mathbf{y}_{ij} \mid \boldsymbol{\nu}_i] = \boldsymbol{\alpha} + \Lambda\boldsymbol{\nu}_i$, and the implied covariance structure is given by $V[\mathbf{y}_{ij} \mid \Phi_i, \Theta_i] = \Lambda\Phi_i\Lambda' + \Theta_i$. The resulting model for the individual can then be written as

$$\begin{aligned} &\mathbf{y}_{ij} = \boldsymbol{\alpha} + \Lambda\boldsymbol{\xi}_{ij} + \boldsymbol{\epsilon}_{ij}, \\ &\boldsymbol{\xi}_{ij} \sim N(\boldsymbol{\nu}_i, \Phi_i), \qquad j = 1 \text{ to } n_i \\ &\boldsymbol{\epsilon}_{ij} \sim N(\mathbf{0}, \Theta_i). \end{aligned} \tag{9}$$

The population distribution can be specified by choosing independent population distributions over the individual sets of parameters. These distributions can be written as

$$\begin{aligned} &\boldsymbol{\nu}_i \sim N(\mathbf{0}, \Delta) \\ &\Phi_i \sim IW(\rho, R) \qquad i = 1 \quad \text{to} \quad I \\ &\Theta_i \sim \prod_{k=1}^{p} IG(a_k, b_k). \end{aligned} \tag{10}$$

The population distribution for the factor covariance matrices, Φ_i, is assumed to be an inverse Wishart distribution with scale matrix R, and a positive shape parameter ρ. The other population distributions are as in the previous models. The unconditional mean for an arbitrary observation is given by $E[\mathbf{y}_{ij}] = \boldsymbol{\alpha}$. The unconditional variance is $V[\mathbf{y}_{ij}] = \Lambda(\Delta + E[\Phi_i])\Lambda' + E[\Theta_i]$ where the expectation $E[\Phi_i]$ is given from the inverse Wishart distribution as $R/(\rho - 2p - 2)$ and $E[\Theta_i]$ is the same as in the previous model. As is the case for the previous models, the implied covariance structure indicates that ignoring heterogeneity can provide misleading inferences regarding parameters.

While it is conceptually possible to specify and provide Bayesian estimation procedures for models with heterogeneous covariance structures, we wish to caution that in many empirical applications, data may not provide enough information to allow reliable estimation of all parameters in such complex models. Researchers therefore need to impose reasonable levels of heterogeneity so as to obtain valid inferences for quantities of interest. In addition, as is the case in consumer psychology studies, researchers may have a priori hypotheses regarding the role of moderating variables. For example, researchers may be interested in knowing how different groups of individuals (males versus females) differ in their emotional structures. In such situations, instead of allowing $\boldsymbol{\nu}_i$, Φ_i and Θ_i to vary for each individual, a hybrid approach (random coefficient + multigroup) that allows complete variation in the factor means and measurement variances, but restricts the Φ matrices to be group specific may be more fruitful. This can be achieved by reparametrizing $\boldsymbol{\nu}_i$ as a function of the hypothesized moderator variables and by estimating separate Φ matrices for each group.

In summary, notice that the first two models that focus only on the heterogeneity in mean structures can be considered as special cases of the last two models which allow for more general patterns of heterogeneity. In the main body of the paper, we focus on the heterogeneous factor means and heterogeneous factor covariances model. The algorithmic details pertaining to the estimation of the heterogeneous factor loadings model are described in Appendix 1 and the details for the heterogeneous factor covariance model with heterogeneous intercepts are provided in Appendix 2.

3. Inference

Bayesian inference requires specification of priors for all parameters in a model. We describe the priors associated with the heterogeneous factor covariances model.

3.1. Priors

The unknown parameters for the model can be collected in the set $\boldsymbol{\beta} = \{\boldsymbol{\alpha}, \Lambda, \Delta, \rho, R, \{a_k\}, \{b_k\}\}$. Lee (1981) and Arminger and Muthén (1998) discuss different forms of prior distributions for factor analysis models. In this paper we specify the prior distribution over $\boldsymbol{\beta}$ as a product of independent priors.

We use proper, but diffuse priors over all model parameters. The prior for the overall mean $\boldsymbol{\alpha}$ can be chosen to be multivariate normal $N(\boldsymbol{\eta}, C)$. The covariance matrix C for this prior can be assumed to be diagonal and the diagonal elements (variances) can be set to large values to represent vague knowledge. The exact location of this distribution is no longer critical once C is set to be large and therefore, $\boldsymbol{\eta}$ can be set to zero.

The matrix of factor loadings, Λ, has a patterned structure. We therefore specify independent multivariate normal priors over the free elements within each row of the matrix. We have for row k a prior $N(\mathbf{g}_k, H_k)$. The covariance matrix H_k can be assumed diagonal with large variances to ensure a diffuse prior. The prior over Λ then is a product of the independent priors associated with the p rows.

The precision matrix Δ^{-1} associated with the population distribution of the factor means $\boldsymbol{\nu}_i \sim N(0, \Delta)$, is a $m \times m$, positive definite matrix. In keeping with standard Bayesian analysis of linear models, we assume a Wishart prior $W(\delta, (\delta\Omega)^{-1})$ where Ω^{-1} can be considered the expected prior precision of the $\boldsymbol{\nu}_i$s. Smaller δ values correspond to vaguer prior distributions.

The Wishart population distribution $W(\rho, R)$ of the individual specific factor precision matrices, Φ_i^{-1}, has two unknown parameters, ρ and R. The scale parameter R is a $m \times m$ symmetric positive definite matrix. A conjugate prior for R^{-1} can be assumed to facilitate further analysis. We therefore choose a Wishart prior, $R^{-1} \sim W(\boldsymbol{\gamma}, (\boldsymbol{\gamma}S)^{-1})$. The shape parameter, ρ, is a positive scalar quantity. The Wishart distribution $W(\rho, R)$ is a proper density only if $\rho \geq m$. We therefore choose a truncated univariate normal prior over $\rho' = \log(\rho)$. Specifically, we assume $\rho' \sim tN(0, \tau)$ where τ can be assumed to be large enough to represent vague knowledge.

There are p different population distributions $IG(a_k, b_k)$, $k = 1$ to p, for the p measurement error variances contained in Θ_i. We therefore need to specify priors over the set of unknowns, $\{\{a_k\}, \{b_k\}\}$. We choose independent conjugate inverse gamma priors, $b_k \sim IG(g_k, h_k)$, for $k = 1$ to p, to represent the prior uncertainty about the scale parameters. Finally, we assume independent univariate normal $N(0, \tau_{ak})$ priors over $\log(a_k)$, for $k = 1$ to p. The prior variance, τ_{ak}, can be set to a large value to ensure a diffuse prior.

3.2. Inference Procedure

Inference in the Bayesian framework involves summarizing the joint posterior of all unknowns. We use simulation based methods to obtain random draws from the posterior density as this density is not available in closed form for our models. Inference is based on the empirical distribution of the draws. The complexity of the posterior density also implies that we cannot use direct methods for obtaining these draws. We therefore use Markov Chain Monte Carlo (MCMC) methods involving data augmentation (Tanner & Wong, 1987) Metropolis-Hastings methods (Hastings 1970; Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953) and Gibbs sampling (Geman & Geman 1984) to obtain these draws. The MCMC methods involve sampling parameter estimates from the full conditional distribution of each block of parameters. If the full conditional is known only up to a normalizing constant, then a Metropolis Hastings step can be used, otherwise, if it is completely known, a Gibbs sampling step can be employed. In the context of the heterogeneous factor covariances model, we need to generate random draws

for $\{\boldsymbol{\alpha}, \Lambda, \{\boldsymbol{\xi}_{ij}\}, \{\boldsymbol{\nu}_i\}, \{\Phi_i\}, \Delta^{-1}, \rho, R^{-1}, \{\Theta_i\}, \{a_k\}, \{b_k\}\}$. Each iteration of the MCMC sampler involves sequentially sampling from the full conditional distributions associated with each block of parameters. The MCMC procedure also provides samples of the factor scores $\{\xi_{ij}\}$ via data augmentation, and therefore enable posterior inference about factor scores. As these factor scores are obtained as part of the MCMC output, a proper accounting of the uncertainty is possible in their estimation. We now describe the simulation steps involved in each iteration. The $(t+1)$-th iteration of the sampling algorithm involves generating random draws from the following full conditional distributions:

1. The overall mean $\boldsymbol{\alpha}$ has a normal prior. From standard Bayesian theory, the full conditional distribution can be written as a multivariate normal distribution given by

$$p(\boldsymbol{\alpha} \mid \Lambda, \{\boldsymbol{\xi}_{ij}\}, \{\Theta_i\}, \{\mathbf{y}_{ij}\}) = N(\hat{\boldsymbol{\alpha}}, V_\alpha) \tag{11}$$

where

$$V_\alpha^{-1} = C^{-1} + \sum_{i=1}^{I} n_i \Theta_i^{-1} \quad \text{and} \quad \hat{\boldsymbol{\alpha}} = V_\alpha \left(C^{-1}\boldsymbol{\eta} + \sum_{i=1}^{I} \Theta_i^{-1} \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \Lambda \boldsymbol{\xi}_{ij}) \right).$$

2. The loading matrix Λ is a patterned matrix containing both fixed and free elements. Of the fixed elements, some are fixed to zero, while others are fixed at one to impose identifiability constraints. The full conditional distribution for the free elements in a row of the matrix Λ is multivariate normal. Given the choice of the prior distributions, the full conditionals pertaining to the different rows are independent. Therefore, the rows can be handled sequentially. Let $\boldsymbol{\lambda}_k$ be the vector of free elements in row k. The prior for $\boldsymbol{\lambda}_k$ is given by $p(\boldsymbol{\lambda}_k) = N(\mathbf{g}_k, H_k)$.

   Let $\tilde{\boldsymbol{\xi}}_{ijk}$ be the vector of factor scores corresponding to the elements in row k of Λ that are set to one and let $\boldsymbol{\xi}_{-ijk}$ contain the remaining factor scores from $\boldsymbol{\xi}_{ij}$. Form the adjusted variable $\tilde{y}_{ijk} = y_{ijk} - \boldsymbol{\iota}' \tilde{\boldsymbol{\xi}}_{ijk} - \alpha_k$, where $\boldsymbol{\iota}$ is a vector of ones. Given the prior, the vector $\boldsymbol{\lambda}_k$ can be sampled from the full conditional distribution given by

$$p(\boldsymbol{\lambda}_k \mid \{\tilde{y}_{ijk}\}, \{\boldsymbol{\xi}_{-ijk}\}, \theta_{i,k}) = N\left(D_k \left[\sum_{i=1}^{I} \sum_{j=1}^{n_i} \theta_{i,k}^{-1} \boldsymbol{\xi}_{-ijk} \tilde{y}_{ijk} + H_k^{-1} \mathbf{g}_k \right], D_k \right) \tag{12}$$

where

$$D_k^{-1} = \sum_{i=1}^{I} \sum_{j=1}^{n_i} \theta_{i,k}^{-1} \boldsymbol{\xi}_{-ijk} \boldsymbol{\xi}'_{-ijk} + H_k^{-1},$$

and $\theta_{i,k}$ is the (k,k)-th element of Θ_i.

3. The factor scores can be obtained in a data augmentation step. Given the multivariate normal prior $N(\boldsymbol{\nu}_i, \Phi_i)$, for the factor scores, the full conditional distribution for the factor scores $\boldsymbol{\xi}_{ij}$ for each observation is multivariate normal and is given by

$$p(\boldsymbol{\xi}_{ij} \mid \mathbf{y}_{ij}, \boldsymbol{\alpha}, \Lambda, \Theta_i, \boldsymbol{\nu}_i, \Phi_i) = N\left(\hat{\boldsymbol{\xi}}_{ij}, V_{\xi_{ij}}\right) \tag{13}$$

where

$$V_{\xi_{ij}}^{-1} = \Phi_i^{-1} + \Lambda' \Theta_i^{-1} \Lambda \quad \text{and} \quad \hat{\boldsymbol{\xi}}_{ij} = V_{\xi_{ij}} \left(\Phi_i^{-1} \boldsymbol{\nu}_i + \Lambda' \Theta_i^{-1} (\mathbf{y}_{ij} - \boldsymbol{\alpha}) \right).$$

4. The full conditional distribution for the individual level factor means $\boldsymbol{\nu}_i$ is a multivariate normal distribution that can be written as

$$p(\boldsymbol{\nu}_i \mid \{\boldsymbol{\xi}_{ij}\}, \Phi_i, \Delta) = N(\hat{\boldsymbol{\nu}}_i, V_{\nu_i}), \tag{14}$$

where

$$V_{\nu_i}^{-1} = \Delta^{-1} + n_i \Phi_i^{-1} \quad \text{and} \quad \hat{\boldsymbol{\nu}}_i = V_{\nu_i} \Phi_i^{-1} \sum_{j=1}^{n_i} \boldsymbol{\xi}_{ij}.$$

5. The normal sampling distribution for the factor scores when combined with the Wishart prior for the individual specific factor precision matrices, $\Phi_i^{-1} \sim W(\rho, R)$, yields a Wishart full conditional distribution for Φ_i^{-1} which can be written as

$$p(\Phi_i^{-1} \mid \{\boldsymbol{\xi}_{ij}\}_{j=1}^{n_i}, \rho, R) = W(\rho_{pos}, R_{pos}) \tag{15}$$

where

$$\rho_{pos} = \rho + n_i \quad \text{and} \quad R_{pos} = \left[\sum_{j=1}^{n_i} (\boldsymbol{\xi}_{ij} - \boldsymbol{\nu}_i)(\boldsymbol{\xi}_{ij} - \boldsymbol{\nu}_i)' + R^{-1} \right]^{-1}.$$

6. Similarly, the full conditional distribution for the precision matrix Δ^{-1} of the individual specific factor means $\boldsymbol{\nu}_i$, is a Wishart distribution. Given the prior $\Delta^{-1} \sim W(\delta, (\delta\Omega)^{-1})$, this full conditional distribution can be written as

$$p(\Delta^{-1} \mid \{\boldsymbol{\nu}_i\}) = W\left(\delta + I, \left[\sum_{i=1}^{I} \boldsymbol{\nu}_i \boldsymbol{\nu}_i^T + \delta\Omega \right]^{-1} \right). \tag{16}$$

7. The full conditional distribution for the hyperparameter R^{-1} is a Wishart Distribution. The likelihood associated with this full conditional distribution is a product of Wishart distributions. As the population distribution for the factor precisions Φ_i is given by $\Phi_i^{-1} \sim W(\exp(\rho'), R)$, the likelihood can be written as

$$L(\{\Phi_i^{-1}\} \mid \rho', R) = \frac{\exp\left(-\frac{1}{2}\text{tr}\left(R^{-1}\sum_{i=1}^{I}\Phi_i^{-1}\right)\right)\prod_{i=1}^{I}\left|\Phi_i^{-1}\right|^{\frac{\exp(\rho')-p-1}{2}}}{|R|^{\frac{\exp(\rho')I}{2}} 2^{\frac{p\exp(\rho')}{2}} \prod_{j=1}^{m}\Gamma\left(\frac{\exp(\rho')+1-j}{2}\right)}. \tag{17}$$

The above likelihood when combined with the conjugate Wishart prior, $R^{-1} \sim W(\gamma, (\gamma S)^{-1})$ yields a full conditional distribution that can be written as

$$p(R^{-1} | \{\Phi_i^{-1}\}, \rho, \gamma, S) = W\left(\gamma + I\rho, \left(\gamma S + \sum \Phi_i^{-1}\right)^{-1}\right). \tag{18}$$

8. The full conditional for the hyperparameter $\rho' = \log(\rho)$ of the Wishart population distribution $W(\rho, R)$ cannot be written in closed form. This full conditional is proportional to product of the likelihood expression specified in (17), and the prior density of ρ' which is a truncated univariate normal $p(\rho') = N(0, \tau)$. We therefore use a random walk Metropolis Hastings step to generate random draws of ρ'. We use a univariate normal proposal density $N(\rho'^{(t)}, t)$, that is centered on the old value of $\rho'^{(t)}$ to generate a candidate ρ'^{c}. The tuning constant t in the proposal density is chosen to facilitate rapid mixing and to avoid excessive rejections of the candidate draws. The generated candidate ρ'^{c} is accepted with the following acceptance probability

$$\min\left\{ \frac{L(\rho'^{c} \mid \{\Phi_i^{-1}\}, R)p(\rho'^{c})}{L(\rho'^{(t)} \mid \{\Phi_i^{-1}\}, R)p(\rho'^{(t)})}, 1 \right\} \tag{19}$$

If the candidate is accepted then $\rho'^{(t+1)} = \rho'^{c}$, otherwise $\rho'^{(t+1)} = \rho'(t)$.

9. The full conditional distributions for the diagonal elements of the individual specific measurement error variances in Θ_i, that is, $\theta_{i,k}$, $k = 1$ to p, are independent inverse gamma distributions. These follow from standard Bayesian theory and can be written as

$$p(\theta_{i,kk} \mid \boldsymbol{\lambda}_k, \alpha_k, \{\boldsymbol{\xi}_{ij}\}_{j=1}^{n_i}, a_k, b_k) = IG\left(\frac{n_i}{2} + a_k, \left[\frac{\sum_{j=1}^{n_i}(y_{ijk} - \alpha_k - \boldsymbol{\lambda}_k'\boldsymbol{\xi}_{ij})^2}{2} + b_k^{-1}\right]^{-1}\right) \tag{20}$$

where $\boldsymbol{\lambda}_k$ contains the elements of row k of Λ.

10. The full conditional for the hyperparameter b_k of the inverse gamma population distribution over the k-th measurement error variance, $\theta_{i,k}$, is also an inverse gamma distribution as the inverse gamma prior, $IG(g_k, h_k)$ is a conjugate distribution. The full conditional can be written as

$$p(b_k \mid \{\theta_{i,k}\}, a_k, g_k, h_k) = IG\left(Ia_k + g_k, \left[h_k^{-1} + \sum_{i=1}^{I}\theta_{i,k}\right]^{-1}\right). \tag{21}$$

The parameter draws from the full conditional distributions of each b_k, $k = 1$ to p can be made in sequence.

11. The full conditional for the hyperparameter $a_k' = \log(a_k)$ of the inverse gamma population distribution for the k-th measurement error variance, $\theta_{i,k}$, cannot be written in closed form. The likelihood of the "data" can be written as

$$L(\{\theta_{i,k}\}|a_k', b_k) = \prod_{i=1}^{I} \frac{\theta_{i,k}^{exp(a_k')-1}\exp(-b_k^{-1}\theta_{i,k})}{\Gamma(exp(a_k'))b_k^{\exp(a_k')}}. \tag{22}$$

The prior density of a_k' is a univariate normal $p(a_k') = N(0, \tau_a)$. The full conditional is proportional to the product of the likelihood and the prior. We use a random walk Metropolis-Hastings step to generate random draws of a_k'. We use a univariate normal proposal density $N(a_k'^{(t)}, t_a)$, that is centered on the old value of $a_k'^{(t)}$ to generate a candidate $a_k'^{c}$. The tuning constant t_a in the proposal density is chosen to allow rapid mixing and to avoid excessive rejections of the candidates. The generated candidate $a_k'^{c}$ is accepted with the following acceptance probability

$$\min\left\{\frac{L(a_k'^{c} \mid \{\theta_{i,k}^{-1}\}, b_k)p(a_k'^{c})}{L(a_k'^{(t)} \mid \{\theta_{i,k}^{-1}\}, b_k)p(a_k'^{(t)})}, 1\right\}. \tag{23}$$

If the candidate is accepted then $a_k'^{(t+1)} = a_k'^{c}$, otherwise $a_k'^{(t+1)} = a_k'(t)$. Parameter draws can be sequentially made for each a_k', $k = 1$ to p.

The MCMC sampler is run for a large number of iterations. This iterative scheme of sequential draws generates a Markov chain that converges in distribution to the joint posterior under fairly general conditions (Tierney, 1994). After discarding the initial draws (burn-in draws) the subsequent draws from the chain can be used as a sample from the posterior distribution. A large sample of draws can be obtained to approximate the posterior distribution to any desired level of accuracy.

4. Model Assessment

4.1. Model Adequacy

The adequacy of a Bayesian model can be assessed using posterior predictive model checking (Gelman, Carlin, Stern, & Rubin, 1996; Yao & Böckenholt, 1999). Let y^{obs} be the observed data, β be the vector of all unknowns and d be the number of posterior draws that are used for adequacy assessment. The sample of parameter draws $\beta_1, \beta_2, \ldots \beta_d$ available from the MCMC algorithm can be used along with the appropriate sampling distribution $p(y \mid \beta)$ to generate hypothetical replicated multilevel data sets $y_1^{rep}, y_2^{rep}, \ldots, y_d^{rep}$. The actual data set can be compared with the replicated data sets using test quantities $T(y, \beta)$ involving either the data alone or both data and parameters. These test quantities are chosen to measure departures of the observed data from the assumed model. They can be omnibus goodness of fit measures or could be chosen specifically to highlight substantive aspects of the application of interest. If the replicated data sets differ systematically from the actual data on some test quantities, then we can ascertain that the model does not adequately capture the data generation process on those aspects that are captured by the test quantities. A posterior predictive p-value given by

$$p(y) = P[T(y^{rep}, \beta) \geq T(y^{obs}, \beta) \mid y^{obs}] \tag{24}$$

can be used to detect model inadequacies. This p-value can be approximated easily from the MCMC sequence of draws using

$$p(y) = \frac{1}{d} \sum_{t=1}^{d} I(T(y_t^{rep}, \beta_t) \geq T(y^{obs}, \beta_t)), \tag{25}$$

where I is an indicator function. The expression in (25) estimates the p-value as the proportion of the d replications in which the simulated discrepancy variable exceeds the realized value. A p-value close to zero or one indicates that the model is inadequate for the aspects measured by the discrepancy variable T.

In heterogeneous factor analysis models, we suggest test statistics based on the within- and between-individual covariance matrices, $\mathbf{\Sigma}_W$ and $\mathbf{\Sigma}_B$, respectively. Let $\overline{y}_k$ denote the grand mean for indicator y_k and $\overline{y}_{ik}$ denote the mean for individual i. Then for a given data set y^t, these test quantities are computed as

$$T(s_{k,l}^w) = \frac{1}{N} \sum_{i=1}^{I} \sum_{j=1}^{n_i} (y_{ijk} - \overline{y}_{ik})(y_{ijl} - \overline{y}_{il}), \tag{26}$$

and

$$T(s_{k,l}^b) = \frac{1}{N} \sum_{i=1}^{I} n_i (\overline{y}_{ik} - \overline{y}_k)(\overline{y}_{il} - \overline{y}_l), \tag{27}$$

where $s_{k,l}^w$ and $s_{k,l}^b$ denote the within- and between-individual covariances, respectively. If a model consistently overpredicts or underpredicts a within- (between) individual covariance then we can conclude that the within (between) covariance structure implied by the model fails in replicating that covariance in the actual data. In the synthetic data applications to follow, we illustrate the diagnostic potential of such test statistics.

The above suggested test statistics are by no means exhaustive. Further research can investigate the specification and empirical performance of other test statistics and discrepancy measures. In addition to using posterior predictive checks, researchers can also use Bayesian residuals based on a fitted model and validation data to see how well the model fits the data.

Bayesian residuals are available easily as a by-product of the MCMC simulation. Various summary measures of these residuals can be utilized to assess model adequacy. For example, Q-Q plots can be utilized to test the normality assumptions of the measurement errors.

4.2. Model Comparison

In Bayesian analysis, Bayes factors (Kass & Raftery, 1995) have traditionally been used to compare two models. However, Bayes factors are difficult to compute for complex models such as ours. We therefore use the pseudo-Bayes factor (PsBF; Geisser & Eddy, 1979; Gelfand, 1996) as a surrogate for the Bayes factor. While the Bayes factor uses the prior predictive density to compute the marginal likelihood, the PsBF is based on the cross-validation predictive density of the data. It can therefore be used even with improper priors. Moreover, it can be very conveniently computed for structural equation models using the MCMC draws.

Let $\mathbf{y}$ be the observed data and let $\mathbf{y}_{(ijk)}$ represent the data with the kth variable of observation j from individual i deleted. The cross-validation predictive density can then be written as

$$\pi(y_{ijk} \mid \mathbf{y}_{(ijk)}) = \int \pi(y_{ijk} \mid \beta, \mathbf{y}_{(ijk)})\pi(\beta \mid \mathbf{y}_{(ijk)})d\beta \tag{28}$$

where β is the vector of all parameters in the model. The PsBF for comparing two models (M1 and M2) is expressed in terms of the product of cross-validation predictive densities and can be written as

$$\text{PsBF} = \prod_{i=1}^{I}\prod_{j=1}^{n_i}\prod_{k=1}^{p}\frac{\pi(y_{ijk} \mid \mathbf{y}_{(ijk)}, M1)}{\pi(y_{ijk} \mid \mathbf{y}_{(ijk)}, M2)}. \tag{29}$$

The PsBF summarizes the evidence provided by the data for M1 against M2 and its value can be interpreted as the number of times model M1 is more (or less) probable than model M2.

The PsBF for our model can be calculated easily from a sample of d MCMC draws $\{\beta_1, \ldots, \beta_d\}$. As β is the vector of all parameters, including the factor scores, the responses y_{ijk}, $i = 1$ to I, $j = 1$ to n_i and $k = 1$ to p, are conditionally independent given β (i.e., $y_{ijk} \sim N(\alpha_k + \lambda_k \boldsymbol{\nu}_{ij}, \theta_{i,k})$). In such a situation, a Monte Carlo estimate of $\pi(y_{ijk} \mid \mathbf{y}_{(ijk)})$ can be obtained as

$$\hat{\pi}(y_{ijk} \mid \mathbf{y}_{(ijk)}) = \left[\frac{1}{d}\sum_{t=1}^{d}\frac{1}{f(y_{ijk}; \beta^{(t)})}\right]^{-1}. \tag{30}$$

where the univariate normal density $f(y_{ijk}; \beta^{(t)})$ is evaluated at t-th draw, $\beta^{(t)}$, from the MCMC sampler. Gelfand (1996) provides the derivation for the above equation. In practice, we can calculate the logarithms of the numerator and denominator of the PsBF. These can be considered as a surrogate for the log-marginal data likelihoods log(Pr(D)) from the models.

5. Synthetic Data Applications

We investigate the MCMC procedures described above using two synthetic data studies. The first study illustrates the consequences of ignoring heterogeneity and examines the performance of the algorithms in recovering the true parameters. The second study examines the performance of the different criteria for model assessment.

5.1. Study 1: The Consequences of Ignoring Heterogeneity

In this section we illustrate the application of the MCMC procedures on synthetic data. The aim is to highlight that simpler methods can yield misleading inferences when data come from a

heterogeneous model. We used a balanced 6 variate data set with 100 groups and 30 observations within each group according to the heterogeneous factor covariances model described in (9) and (10). We set $\boldsymbol{\alpha} = 0$,

$$\Lambda' = \begin{pmatrix} 1 & 0.8 & 0.6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.6 & 0.8 \end{pmatrix}, \tag{31}$$

$$E[\Phi_i] = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}, \quad \Delta = \begin{pmatrix} 0.6 & -0.4 \\ -0.4 & 0.6 \end{pmatrix}, \tag{32}$$

and $E[\Theta_i] = 0.4$ to generate the data.

We used SAS and our MCMC procedures to estimate three models on the data. The first model (NH) is a nonhierarchical model that ignores the nature of the clustering of the observations and is given by

$$\mathbf{y}_{ij} \sim N(\boldsymbol{\alpha}, \Sigma)$$

where $\Sigma = \Lambda\Phi\Lambda' + \Theta$. We estimated this model using Proc CALIS in the SAS software. The second model (ML) is the multilevel model with heterogeneous factor means described by (2) and (3). Finally, our third model is the true model, that is, the Heterogeneous Factor Covariances (HC) model represented by (9) and (10). In estimating these models we use priors that are similar to those outlined in section 3.1. The prior for $\boldsymbol{\alpha}$ is assumed to be $p(\boldsymbol{\alpha}) = N(0, 100I)$. The measurement variances in the first two models have independent inverse gamma priors each given by $p(\theta_{kk}) = IG(0.001, 1000), k = 1, \ldots p$. We use $R^{-1} \sim W(3, 3I)$ and $\log(\rho) \sim tN(0, 100)$ for the hyperparameters associated with the factor covariances. For the mean factor covariances we use $\Delta^{-1} \sim W(3, 3I)$ and finally we assume independent univariate normal $N(0, 100)$ priors over the individual elements in Λ.

The estimates for the models are based on 10,000 draws from the joint posterior distribution. These draws were obtained after discarding 2,000 draws from the initial transient portion of the chain. Convergence was assessed using a variety of diagnostics detailed in the CODA package (Best, Cowles & Vines, 1995) and by using time series plots to graphically assess the quality of the mixing of the chain.

Table 1 reports the estimated factor loadings and the associated 95% posterior intervals for the three models. It is clear from the table that all three models yield almost identical estimates of the factor loadings. This is not surprising as the data generating mechanism assumed identical Λ for all individuals.

Table 2 reports the within factor covariance matrix Φ for the first two models and the population expectation of the Φ_i matrices, $E(\Phi_i)$ for the heterogenous factor covariance model (HC).

TABLE 1.
Factor loadings

Parameter	True	NH	ML	HC
λ_{21}	0.8	0.79 (0.76, 0.82)	0.79 (0.76, 0.82)	0.78 (0.75, 0.82)
λ_{31}	0.6	0.61 (0.58, 0.64)	0.61 (0.58, 0.63)	0.61 (0.59, 0.63)
λ_{52}	0.6	0.60 (0.57, 0.63)	0.60 (0.57, 0.63)	0.60 (0.58, 0.63)
λ_{62}	0.8	0.80 (0.76, 0.83)	0.79 (0.75, 0.82)	0.79 (0.76, 0.83)

TABLE 2.
Factor covariance matrices

Parameter	True	NH	ML	HC
Φ_{11}	1.0	1.37	0.78	0.98
		(1.28, 1.47)	(0.72, 0.84)	(0.84, 1.17)
Φ_{12}	0.2	−0.14	0.19	0.22
		(−0.19, −0.08)	(0.15, 0.22)	(0.15, 0.29)
Φ_{22}	1.0	1.27	0.81	0.97
		(1.18, 1.37)	(0.75, 0.87)	(0.83, 1.13)
Δ_{11}	0.6		0.59	0.60
		–	(0.43, 0.77)	(0.44, 0.80)
Δ_{12}	−0.4		−0.33	−0.34
		–	(−0.47, −0.21)	(−0.48, −0.22)
Δ_{22}	0.6		0.48	0.48
		–	(0.36, 0.66)	(0.35, 0.66)

It also reports the across individual covariance matrix Δ of the mean factor scores $\boldsymbol{\nu}_i$ for the two models that incorporate heterogeneity in factor means. It is clear from the estimates obtained for the nonhierarchical model (NH) that ignoring heterogeneity leads to a confounding of the within and between covariance matrices of the factor scores. We see that the variance terms Φ_{11} and Φ_{22} are inflated when compared to their true values and the covariance Φ_{12} has the wrong sign because of this confounding. The estimates from the true model (HC) are close to the true values as expected. However, it is interesting to note that the estimates of Φ from the multilevel model (ML), are misleading. While the ML covariance estimate Φ_{12} is of the proper sign, the magnitudes of the variances differ from the true values and the 95% posterior intervals do not cover the true values. Finally, the estimates of Δ obtained from the two models are close to the true values.

Table 3 reports the estimated measurement error variances Θ for the NH and ML models and the population expectation of Θ_i for the HC model. The magnitude of the estimates are similar across all the models. A closer look at the estimates associated with Θ_{33} and Θ_{55} reveals that the 95% posterior intervals for the first two models do not cover the true value of 0.4. The posterior intervals for the full model (HC) are wider as they properly reflect the uncertainty arising due to heterogeneity and cover the true value for all parameters.

TABLE 3.
Measurement error variances

Parameter	True	NH	ML	HC $E(\Theta_i)$
Θ_{11}	0.4	0.41	0.41	0.40
		(0.37, 0.46)	(0.37, 0.46)	(0.34, 0.47)
Θ_{22}	0.4	0.39	0.39	0.41
		(0.36, 0.42)	(0.36, 0.42)	(0.36, 0.47)
Θ_{33}	0.4	0.35	0.35	0.35
		(0.32, 0.37)	(0.32, 0.37)	(0.3, 0.40)
Θ_{44}	0.4	0.43	0.41	0.42
		(0.38, 0.48)	(0.37, 0.45)	(0.36, 0.48)
Θ_{55}	0.4	0.35	0.35	0.35
		(0.32, 0.37)	(0.32, 0.37)	(0.31, 0.40)
Θ_{66}	0.4	0.45	0.46	0.46
		(0.42, 0.49)	(0.42, 0.49)	(0.4, 0.52)

In summary, the results from this simulated example indicate that our MCMC algorithm does well in recovering the true parameters. More importantly, the example clearly demonstrates that ignoring heterogeneity can lead to misleading inferences. Ignoring heterogeneity altogether can lead to sign reversals and inflation of factor variances. Ignoring heterogeneity in covariance structures can lead to misleading estimates of parameter values and an underappreciation of uncertainty in these estimates.

5.2. Study 2: Model Assessment

To assess the performance of the model adequacy test quantities described in section 4.1 and to investigate the performance of the pseudo Bayes factor for model comparison, we generated data according to a standard multilevel model (i.e., the heterogeneous intercept model in (5) and (6)) and estimated two models. The first model (Model 1) is the true model while the second (Model 2) is a mis-specified heterogenous factor means model described in (2) and (3).

We specified a two-factor structure at both levels with three indicators per factor. We set $\boldsymbol{\alpha} = 0$,

$$\Lambda' = \Lambda_b' = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \tag{33}$$

$$\Phi = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}, \quad \Phi_b = \begin{pmatrix} 1 & -0.2 \\ -0.2 & 1 \end{pmatrix} \tag{34}$$

$$\Theta = \text{diag}(0.2), \quad \Theta_b = \text{diag}(0.1) \tag{35}$$

to generate a balanced 6 variate data set with 300 groups and 30 observations within each group. We then computed the posterior predictive p-values for the within- and between covariances (see (26) and (27)) and the PsBF. Note that Model 2 is mis-specified since it assumes constant measurement intercepts but heterogeneous factor means. In contrast, the correctly specified model (Model 1) assumes heterogeneous measurement intercepts but common factor means.

Columns 2 and 3 in Table 4 report the results of the posterior predictive p-values associated with the $p(p+1)/2$ nonredundant elements of the within individual covariance matrix and Columns 7 and 8 report the corresponding p-values for the between individual covariance matrix for the two models. The p-values were computed based on 1500 replicated data sets. It is clear from columns 2 and 7 that the p-values for Model 1 are all near 0.5 for all the within and between covariance elements. This is expected as Model 1 is correctly specified. Most of the p values for Model 2 (columns 3 and 8), however, have extreme values. At first glance, this clearly indicates that a heterogeneous factor means model does not adequately capture the covariance structure implied by a standard multilevel model (i.e., heterogeneous intercepts).

Assuming common factor loadings across levels, recall that the within- and between-individual covariance matrices implied by the standard multilevel model (Model 1) are, respectively, $\Sigma_W^1 = \Lambda\Phi\Lambda' + \Theta$ and $\Sigma_B^1 = \Lambda\Phi_b\Lambda' + \Theta_b$ (see (5) and (6)). The corresponding quantities implied by the heterogeneous factor means model (Model 2) are $\Sigma_W^2 = \Lambda\Phi\Lambda' + \Theta$ and $\Sigma_B^2 = \Lambda\Delta\Lambda'$ (see (2) and (3)). Examining these quantities, it is clear that the heterogenous factor model is likely to over (under) estimate the within- (between) group variances but would be adequate in estimating the covariance elements. A re-examination of columns 3 and 8 confirms this conjecture as the p-values associated with the within and between variance elements are extreme (see numbers in bold). In fact, these p-values are near 1 in column 3 and near 0 in column 8, thus clearly indicating that for the replicated data sets, the within covariance is inflated and the between covariance is underestimated. Not all the p-values associated with the covariances, however, are close to 0.5 (e.g., p-value for $s_{1,2}^b$ is 1).

As a further check, we computed the mean absolute deviations $|s_{act} - s_{rep}|$ between the covariances from the actual data s_{act} and the covariances from the replicated data sets s_{rep}. The

TABLE 4.
Model adequacy check

	Within-Subject Covariance					Between-Subject Covariance			
	p-values		MAD			p-values		MAD	
	Model 1	Model 2	Model 1	Model 2		Model 1	Model 2	Model 1	Model 2
$s^w_{1,1}$	0.473	**1.000**	0.0111	**0.0761**	$s^b_{1,1}$	0.518	**0.000**	0.0114	**0.0776**
$s^w_{1,2}$	0.513	0.000	0.0080	0.0419	$s^b_{1,2}$	0.493	1.000	0.0081	0.0399
$s^w_{1,3}$	0.473	0.018	0.0079	0.0273	$s^b_{1,3}$	0.495	0.992	0.0083	0.0301
$s^w_{1,4}$	0.591	0.614	0.0064	0.0080	$s^b_{1,4}$	0.487	0.236	0.0075	0.0103
$s^w_{1,5}$	0.487	0.506	0.0065	0.0075	$s^b_{1,5}$	0.535	0.510	0.0075	0.0078
$s^w_{1,6}$	0.369	0.346	0.0068	0.0092	$s^b_{1,6}$	0.509	0.162	0.0072	0.0118
$s^w_{2,2}$	0.512	**1.000**	0.0117	**0.0580**	$s^b_{2,2}$	0.502	**0.000**	0.0115	**0.0579**
$s^w_{2,3}$	0.501	0.000	0.0082	0.0353	$s^b_{2,3}$	0.490	1.000	0.0085	0.0366
$s^w_{2,4}$	0.484	0.486	0.0064	0.0085	$s^b_{2,4}$	0.526	0.670	0.0077	0.0083
$s^w_{2,5}$	0.437	0.432	0.0063	0.0081	$s^b_{2,5}$	0.537	0.934	0.0078	0.0155
$s^w_{2,6}$	0.432	0.346	0.0066	0.0095	$s^b_{2,6}$	0.599	0.998	0.0075	0.0275
$s^w_{3,3}$	0.516	**1.000**	0.0109	**0.0739**	$s^b_{3,3}$	0.529	**0.000**	0.0119	**0.0707**
$s^w_{3,4}$	0.716	0.762	0.0072	0.0093	$s^b_{3,4}$	0.439	0.016	0.0076	0.0210
$s^w_{3,5}$	0.399	0.516	0.0066	0.0079	$s^b_{3,5}$	0.535	0.470	0.0075	0.0074
$s^w_{3,6}$	0.558	0.572	0.0058	0.0082	$s^b_{3,6}$	0.483	0.182	0.0077	0.0104
$s^w_{4,4}$	0.526	**1.000**	0.0104	**0.0857**	$s^b_{4,4}$	0.519	**0.000**	0.0102	**0.0813**
$s^w_{4,5}$	0.492	0.010	0.0079	0.0272	$s^b_{4,5}$	0.507	0.996	0.0073	0.0282
$s^w_{4,6}$	0.515	0.000	0.0073	0.0376	$s^b_{4,6}$	0.470	1.000	0.0071	0.0407
$s^w_{5,5}$	0.489	**1.000**	0.0113	**0.0558**	$s^b_{5,5}$	0.543	**0.000**	0.0108	**0.0604**
$s^w_{5,6}$	0.503	0.000	0.0080	0.0414	$s^b_{5,6}$	0.471	1.000	0.0072	0.0364
$s^w_{6,6}$	0.505	**0.998**	0.0110	**0.0640**	$s^b_{6,6}$	0.498	**0.000**	0.0104	**0.0636**

fourth and the ninth columns in Table 4 show that the mean absolute deviations for Model 1 are very close to zero indicating perfect recovery of the within and between-group covariance matrices. Columns 5 and 10, (Model 2) show that the mean absolute deviations associated with the variance elements are of larger magnitude than those corresponding to the covariance elements (see numbers in bold). The mean absolute deviations of the covariance elements, however, are still larger than those from Model 1. Thus combining the results from the posterior predictive checking (i.e., p-values) and the mean absolute deviations clearly shows that Model 2 is misspecified.[2]

To complete our model assessment, we computed the Psuedo Bayes Factor for Model 1 versus Model 2 and found $PsBF = \exp(10, 351)$. This value clearly provides strong support for the true model (Model 1) over the mis-specified model (Model 2).

In summary, our synthetic data results indicate that our MCMC algorithm does well in recovering the true parameters. In addition, our model adequacy test function and the pseudo Bayes factor measure are effective in model diagnosis and selection.

[2]Test statistics based on the total covariance matrix are not diagnostic. Comparing the sum of the implied within- and between-individual covariance matrices (i.e., $\boldsymbol{\Sigma}^l_W + \boldsymbol{\Sigma}^l_B, l = 1, 2$) for both models suggests that the heterogeneous factor means model has an identical total covariance structure as the standard multilevel model but confounds $\boldsymbol{\Theta}$ and $\boldsymbol{\Theta}_b$. We computed p-values and mean absolute deviations based on the total covariance matrix and found that both models are adequate in recovering such a test statistics.

6. Application: The Structure of Consumption Emotions

We now report the application of the above procedures on a data set involving the structure of consumption emotions. Consumption emotions are affective states associated with product consumption. An understanding of the structure of consumption emotions is important as past research has shown a significant relationship between such emotions and consumer satisfaction (Oliver 1993) and other evaluative judgements. Bagozzi, Gopinath and Nyer (1999) present a comprehensive review of emotions and their role in marketing.

Consumption emotions have been conceptualized in three ways. At one extreme, they are represented in a rich taxonomy of discrete emotions such as joy, happiness, anger, and guilt (see Richins, 1997). At the other extreme, consumption emotions have been structured into a limited number of basic dimensions such as pleasure/arousal (e.g., Mano & Oliver, 1993) or positive/negative affect (e.g., Dube & Morgan, 1996). Between these two extremes, an attribution-based conceptualization has emerged. In this representation, negative emotions are further differentiated into "Other-attributed emotions", "Situation-attributed emotions" and "Self-attributed emotions" on the basis of their locus of attribution. Other-attributed emotions (e.g, anger and frustration) are those negative emotions that are attributed to a specific external agent who failed to manage controllable circumstances. Situation-attributed emotions (e.g., anxiety and sadness) result from uncontrollable circumstances. Finally, Self-attributed emotions (e.g., guilt, shame) are attributed to own actions and such emotions rarely emerges in a consumption context (Westbrook & Oliver, 1991). There is considerable empirical evidence that supports an attribution-based structure of consumption emotions in a diversity of industries (see Oliver, 1993; also Dube, Belanger, & Trudeau, 1996).

We used our Bayesian approach to estimate a confirmatory factor analysis model of consumption emotions using a data set in which a panel of 54 customers of a college dining meal plan reported the consumption emotions they experienced during each of a sequence of 39 consecutive dinner episodes. The original data were collected to examine the relationship between consumption emotions and consumer satisfaction. For this paper, we obtained the portion of the original data that pertains to consumption emotions to illustrate our methodology. On average, each customer reported emotions for 29 episodes. Respondents were asked to indicate how intensely they felt each emotion during dinner on a 7-point scale. The emotions items for positive emotions (PE) were happy, proud, and elated. The other-attributed emotion (OAN) items included hostile, irritated, and upset. The items for situation-attributed emotion (SAN) were anxiety, fear, and sadness. Figure 1 presents the consumption emotion structure that we specified.

We specified three models to examine the nature of customer heterogeneity in consumption emotion structure The first model (M1) is the traditional confirmatory factor analysis model that assumes no heterogeneity. The second model (M2), described in section 2.1.1, captures heterogeneity in mean structure by assuming that subjects differ in their factor means. The third model (M3) generalizes M1 and M2 by specifying heterogeneity in both mean and covariance structures.[3] Specifically, it allows subjects to have different factor means and different measurement error variances. However, to make the factor scores comparable across subjects, M3 constrains the factor loadings to be invariant (see Yung, 1997). For identification, we imposed the following constraints on the parameters. In all three models, to fix the scale of the factors, we set the loadings of one of the indicator variables pertaining to each factor to unity. In models M1 and M2, as discussed in section 2, we set the population mean of the subject-specific factor means to zero to fix the origin of the factors.

In estimating these models we use priors that are similar to those outlined in section 3.1. The prior for $\boldsymbol{\alpha}$ is assumed to be $p(\boldsymbol{\alpha}) = N(0, 100I)$. The measurement variances in the first

[3] As noted by an anonymous reviewer, the repeated measurement nature of our study necessitates that we also account for time dependence. As our model does not handle time series, we leave this issue for future research. We note, however, that our model does imply some correlation structure between the observations of the same individual.

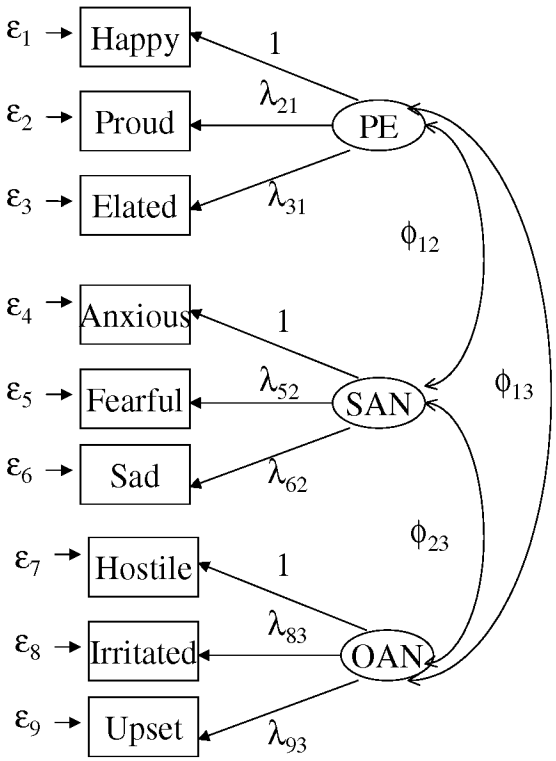

FIGURE 1.
The factor analysis model for consumption emotions.

two models (M1 and M2) have independent inverse gamma priors each given by $p(\theta_{kk}) = IG(0.001, 1000)$, $k = 1, \ldots p$. We use $R^{-1} \sim W(3, 3I)$ and $\log(\rho) \sim tN(0, 100)$ for the hyperparameters associated with the factor covariances. For the mean factor covariances we use $\Delta^{-1} \sim W(3, 3I)$ and finally we assume independent univariate normal $N(0, 100)$ priors over the individual elements in Λ.

We use MCMC procedures for parameter inference. The MCMC procedures involve iterative sampling from the full conditionals pertaining to the model unknowns. The full conditionals for the parameters of the restricted models can be easily adapted from those given in section 3. For example, if a particular parameter is assumed to be invariant in a restricted model, then the hyper-parameters (i.e., parameters specific to the associated population distribution) are fully specified and are not treated as random. Given the hierarchical nature of the specifications, this involves minimal changes in the full conditionals described in section 3. Given priors and appropriate starting values, the MCMC sampler invloves iterative drawing from the full conditional distributions. The parameter estimates that we report for the three models are based on 20,000 draws after discarding the initial 5,000 draws as burn-in. Convergence was assessed using a variety of diagnostics implemented in CODA (Gilks, Richardson, & Spiegelhalter, 1996).

6.1. Model Selection

To select the model that best captures heterogeneity in consumption emotion structure, we computed the log-marginal likelihoods based on the cross validation density (LML; see Gelfand, 1996)) for the three models from the simulated parameter draws. The LML for models M1, M2

and M3, respectively, are −17669.03, −17419.46, and −14215.02. The improvements in LML from M1 to M3 and M2 to M3 are very large suggesting that the data strongly support M3. Thus, subjects are clearly different in both the mean and covariance structures.

6.2. *Parameter Estimates*

We now describe the parameter estimates of the three models. Table 5 reports the estimated factor loadings. The loading estimates from models M1 and M2 are similar. However, the estimates from M3 which allows heterogeneity in factor means and in covariance structures are different from those of M1 and M2. Specifically, except for the "Upset" item, the factor loadings from M3 are lower in magnitude than those from M1 and M2. As the reliability of a measure is proportional to its squared loading, this result clearly indicates that ignoring heterogeneity leads to inflated measurement reliability. Finally, all the factor loadings from M3 are significant. This suggests that the indicators variables we used are reliable measures for the underlying factors. Comparing the magnitude of the factor loadings reveals that the "fearful" and "sad" measures are the least reliable.

Table 6 reports the measurement error variances. First, note that all the measurement error variances from all three models are positive (i.e., no heywood cases). Second, the variance estimates from M1 and M2 are slightly different for the SAN and OAN emotion items but deviate significantly for the PE items. Third, column 5 reports the mean measurement error variances from M3. Comparing these estimates with those from M1, we see that ignoring both mean and covariance heterogeneity results in more amplified bias. Finally, the last column in Table 6 reports the across-individual variation, $Std(\theta_i)$, of the measurement error variances. Clearly, there is considerable level of heterogeneity in the measurement error variances. This means that our sample subjects responded to questions with different degree of accuracy.

TABLE 5.
Factor loadings

Factor	Indicator	M1	M2	M3
	Happy	1	1	1
Positive Emotion (PE)	Proud	0.873 (0.041)	0.997 (0.055)	0.882 (0.051)
	Elated	1.04 (0.057)	0.877 (0.036)	0.840 (0.042)
	Anxious	1	1	1
Negative Emotion (SAN)	Fearful	0.925 (0.026)	0.890 (0.025)	0.517 (0.037)
	Sad	0.727 (0.032)	0.723 (0.033)	0.490 (0.04)
	Hostile	1	1	1
Negative Emotion (OAN)	Irritated	0.798 (0.029)	0.773 (0.028)	0.597 (0.025)
	Upset	1.181 (0.041)	1.118 (0.041)	1.143 (0.040)

TABLE 6.
Measurement error variances

Factor	Indicator	M1	M2	M3	
				$E(\Theta_i)$	$Std(\Theta_i)$
	Happy	1.080 (0.059)	1.039 (0.058)	1.144 (0.140)	1.006
Positive Emotion (PE)	Proud	1.046 (0.050)	0.784 (0.054)	0.744 (0.091)	0.165
	Elated	0.331 (0.048)	0.581 (0.036)	0.360 (0.047)	0.185
	Anxious	0.341 (0.020)	0.315 (0.019)	0.303 (0.03)	3.637
Negative Emotion (SAN)	Fearful	0.240 (0.016)	0.267 (0.015)	0.068 (0.010)	2.678
	Sad	0.765 (0.030)	0.752 (0.030)	0.77 (0.012)	1.064
	Hostile	0.571 (0.028)	0.519 (0.027)	0.472 (0.057)	1.489
Negative Emotion (OAN)	Irritated	0.416 (0.019)	0.411 (0.019)	0.207 (0.025)	0.631
	Upset	0.504 (0.031)	0.549 (0.030)	0.387 (0.053)	1.361

Table 7 reports the estimated factor covariance matrices Φ from the three models. For M2 and M3, the table also reports the covariance matrix Δ of the mean factor scores $\boldsymbol{\nu}_i$ across individuals. In section 2.1.1, we show that ignoring heterogeneity in factor means leads to an aggregate factor covariance matrix that confounds Φ and Δ (i.e., $\Phi^{Agg} = \Phi + \Delta$). We also show that this confound could lead to a sign reversal of the factor covariances under some conditions. Comparing the factor covariance estimates from M1 (column 2) with the sum of the estimated within (column 3) and between factor covariance (column 4) matrices from M2 clearly confirms this theoretical relationship. For example, Φ_{PE-PE} from M1 is 0.922. This value is approximately equal to the sum of the estimates $\Phi_{PE-PE} = 0.383$ and $\Delta_{PE-PE} = 0.583$ obtained from M2. Most importantly, note that the factor covariance estimates Φ_{PE-SAN} and Φ_{PE-OAN} from M1 have the opposite sign. This result of a positive relationship between positive and negative emotions is clearly misleading. The same comparison between M1 and M3 confirms that the aggregate model confounds both sources of factor variability and leads to a sign reversal of the covariances between the positive and the negative emotion factors. However, the relationship $\Phi^{Agg} = \Phi + \Delta$ does not hold as closely. This is because, unlike M2, M3 accounts for both mean and covariance heterogeneity. Thus, if factor mean and/or covariance heterogeneity are ignored, the results from a conventional factor analysis model (M1) will always overestimate the factor variances and therefore inflate the measurement reliability. They could also lead to sign reversal of the factor covariances. Finally, comparing the factor covariance estimates from M2 and M3 in Table 7, we see that the heterogeneous factor means model (M2) generally produced larger Φ and lower Δ estimates. This suggests that the conventional multilevel factor analysis model may not fully correct for heterogeneity and can also lead to inflated measurement reliability.

TABLE 7.
Variation in factors scores

	Φ			Δ	
Parameter	M1	M2	M3	M2	M3
PE-PE	0.922	0.383	0.378	0.583	0.716
	(0.075)	(0.041)	(0.042)	(0.123)	(0.162)
PE-SAN	0.173	-0.048	-0.041	0.182	0.257
	(0.028)	(0.018)	(0.012)	(0.071)	(0.097)
PE-OAN	0.001	-0.207	-0.153	0.157	0.202
	(0.030)	(0.024)	(0.021)	(0.061)	(0.072)
SAN-SAN	0.780	0.452	0.115	0.357	0.501
	(0.041)	(0.025)	(0.017)	(0.075)	(0.108)
SAN-OAN	0.571	0.334	0.209	0.257	0.330
	(0.031)	(0.020)	(0.019)	(0.059)	(0.074)
OAN-OAN	0.777	0.583	0.473	0.256	0.261
	(0.048)	(0.036)	(0.034)	(0.056)	(0.059)

We now focus on the Φ and Δ estimates from the selected model M3. As expected, we see that the negative emotion factors SAN and OAN are positively correlated with each other but negatively correlated with the positive emotion factor, PE. The Δ estimates show that all the factor means are positively correlated across individuals. In addition, the large magnitude of the diagonal elements of Δ show that there is considerable heterogeneity in factor means across individuals. Figure 2 provides a plot of the individual mean factor scores.

In summary, this empirical example clearly illustrates the consequences of ignoring heterogeneity. Specifically, our empirical results show that ignoring both mean and covariance heterogeneity leads to sign reversal in the factor covariance matrix and inflated measurement reliability. Our results also show that multilevel factor analysis models that only capture heterogeneity in means are likely to understate the factor means variability in the population and to overstate the factor variances. Thus, multilevel factor model can also lead to inflated measurement reliability.

7. Summary and Conclusions

This paper develops and tests a hierarchical Bayesian framework for handling mean and covariance heterogeneity in confirmatory factor analysis. We develop Markov Chain Monte Carlo (MCMC) procedures for sampling based Bayesian inference. The hierarchical Bayesian approach allows for appropriate pooling of the data while taking into account heterogeneity and is particularly suitable for studies in which multilevel data, panel data or multiple observations are available for a given set of subjects. The Bayesian procedures we developed in this paper circumvent the need for complex multidimensional integration which is necessary for maximum likelihood estimation. An important feature of our Bayesian approach is that it automatically provides individual-level estimates of model parameters and factor scores while accounting for the uncertainty in such estimates.

Our analysis of the simulated data sets indicate that our MCMC algorithm does well in recovering the true parameters. In addition, our model adequacy test function and the pseudo Bayes factor measure are effective in model diagnosis and selection. More importantly, our analysis clearly illustrates the consequences of ignoring heterogeneity. Specifically, we find that ignoring heterogeneity altogether can lead to sign reversals and inflation of factor variances. Ignoring

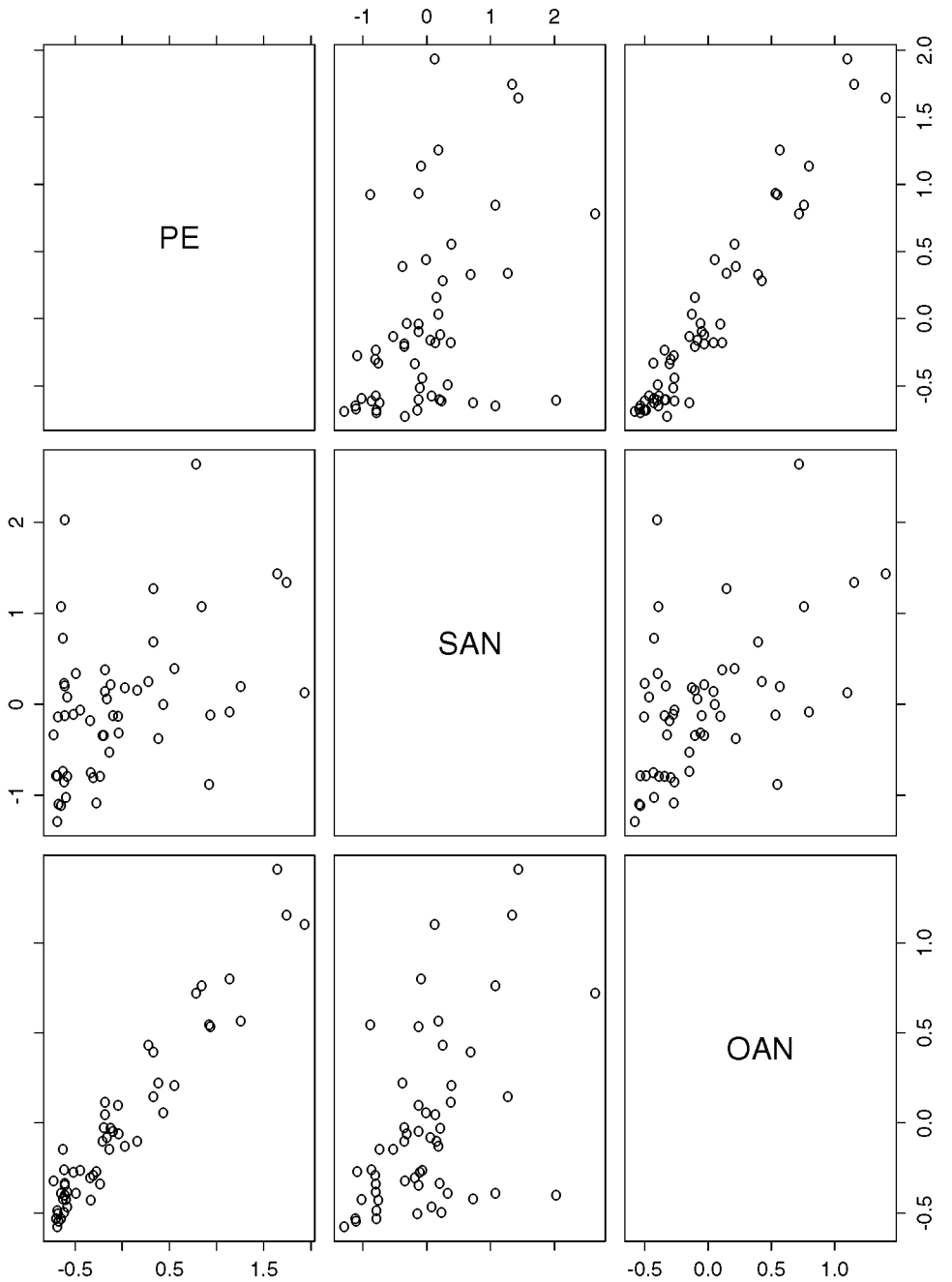

FIGURE 2.
Mean factor scores for individuals.

heterogeneity in covariance structures, however, can lead to misleading estimates of parameter values and an underappreciation of uncertainty in these estimates.

We also tested our Bayesian methodology using data from a consumption emotion study. The results show that both factor means and measurement error variances vary significantly in the population. The results also show that conventional confirmatory factor analysis that totally ignores heterogeneity leads to sign reversal in the factor covariance matrix and inflated measure-

ment reliability. Our results also indicate that multilevel factor analysis models that only capture heterogeneity in means is likely to understate the factor means variability in the population and to overstate the factor variances. Thus accounting for heterogeneity in both mean and covariance structures is important for obtaining proper inferences.

In this paper, we concentrated on confirmatory factor analysis but our procedures can be readily used for exploratory factor analysis models given proper constraints to control for the rotational indeterminacies of the factor solution. Bock and Gibbons (1996) suggest one type of constraints that can be applied. Similarly, although our procedures were presented in the context of metric data, they can be generalized to accommodate nonmetric (e.g., binary or ordinal) data situations. Our algorithms can also be naturally extended to data structures with multiple levels of nesting and can also be modified easily to include regressors at all levels of the hierarchy. Finally, future research should generalize our model to account for time dependence in situations where the level-one units are repeated measures.

A. Appendix 1: Full Conditional Distributions for the Heterogeneous Factor Loadings Model

We need to generate random draws for $\{\boldsymbol{\alpha}, \Lambda_i, \{\lambda_{km}\}, \{\kappa_{km}\}, \{\boldsymbol{\xi}_{ij}\}, \{\boldsymbol{\nu}_i\}, \Phi, \Delta^{-1}, \{\Theta_i\}, \{a_k\}, \{b_k\}\}$. Each iteration of the MCMC sampler involves sequentially sampling from the full conditional distributions associated with each block of parameters. We describe the appropriate priors while detailing the full conditionals

1. The full conditional for $\boldsymbol{\alpha}$ is multivariate normal $N(\hat{\boldsymbol{\alpha}}, V_\alpha)$, where $V_\alpha^{-1} = C^{-1} + \sum_{i=1}^{I} n_i \Theta_i^{-1}$ and $\hat{\boldsymbol{\alpha}} = V_\alpha[C^{-1}\boldsymbol{\eta} + \sum_{i=1}^{I} \Theta_i^{-1} \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \Lambda_i \boldsymbol{\xi}_{ij})]$.
2. The free elements of Λ_i can be drawn in sequence from univariate normal distributions. The population distribution in Equation (8) acts as a prior for the elements of Λ_i. For sake of notational simplicity, we drop subscripts and let λ_i be a free element of Λ_i. Also let the associated population distribution be $N(\lambda, \kappa)$. The full conditional can be written as $N(\hat{\lambda}_i, v_{\lambda_i})$, where $v_{\lambda_i}^{-1} = \kappa^{-1} + \theta_i^{-1}\boldsymbol{\xi}'\boldsymbol{\xi}$ and $\hat{\lambda}_i = v_{\lambda_i}[\kappa^{-1}\lambda + \theta_i^{-1}\boldsymbol{\xi}'\tilde{\mathbf{y}}_i]$, where θ_i is the appropriate measurement error variance, $\boldsymbol{\xi}$ is the vector of the relevant factor scores for individual i and $\tilde{\mathbf{y}}_i$ is a vector of adjusted scores for the manifest variable associated with λ_i for the ith individual.
3. The full conditional for the mean λ_{km} of the population distribution associated with the (k, m) element of Λ_i is a univariate normal distribution. Given a common diffuse prior $\lambda_{i,km} \sim N(g, h)$, the full conditional is given by $N(\hat{\lambda}_{km}, v_{\lambda_{km}})$, where

$$v_{\lambda_{km}}^{-1} = h^{-1} + I\kappa_{km}^{-1} \quad \text{and} \quad \hat{\lambda}_{km} = v_{\lambda_{km}}\left[h^{-1}g + \kappa_{km}^{-1}\sum \lambda_{i,km}\right].$$

4. The full conditional for κ_{km} is given by

$$IG\left(I/2 + c, \left[\sum_{i=1}^{I}(\lambda_{i,km} - \lambda_{km})^2/2 + d^{-1}\right]^{-1}\right),$$

where $IG(c, d)$ is the prior for κ_{km}.
5. The full conditional distribution for the factor scores $\boldsymbol{\xi}_{ij}$ for each observation is multivariate normal $N(\hat{\boldsymbol{\xi}}_{ij}, V_{\xi_{ij}})$ where

$$V_{\xi_{ij}}^{-1} = (\Phi^{-1} + \Lambda_i'\Theta_i^{-1}\Lambda_i) \quad \text{and} \quad \hat{\boldsymbol{\xi}}_{ij} = V_{\xi_{ij}}(\Phi^{-1}\boldsymbol{\nu}_i + \Lambda_i'\Theta_i^{-1}(\mathbf{y}_{ij} - \boldsymbol{\alpha})).$$

6. The full conditional distribution for the individual level factor means $\boldsymbol{\nu}_i$ is a multivariate normal distribution $N(\hat{\boldsymbol{\nu}}_i, V_{\nu_i})$, where

$$V_{\nu_i}^{-1} = \Delta^{-1} + n_i \Phi^{-1} \quad \text{and} \quad \hat{\boldsymbol{\nu}}_i = V_{\nu_i} \Phi^{-1} \sum_{j=1}^{n_i} \boldsymbol{\xi}_{ij}.$$

7. The correlation matrix Φ can be drawn using a Metropolis Hit and Run algorithm (Ansari & Jedidi, 2000; Chen & Schmeiser, 1993; Dey & Chen, 1998). If the prior distribution for the nonredundant and free elements of Φ that are contained in the vector vec(Φ) is given by $\pi(\text{vec}(\Phi) \mid \Phi_0, \mathbf{G}_0)$, then the full conditional of Φ is proportional to the product of the likelihood $L(\Phi|.)$ and the prior. Here $L(.)$ is the conditional likelihood of observing the "data" and is proportional to

$$|\Theta_i|^{-\frac{n_i}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij} - \boldsymbol{\alpha} - \Lambda_i \boldsymbol{\xi}_{ij})' \Theta_i^{-1} (y_{ij} - \boldsymbol{\alpha} - \Lambda_i \boldsymbol{\xi}_{ij})\right\}.$$

Direct methods for sampling from this full conditional distribution are not available so we generate Φ using a Metropolis Hit-and-Run algorithm. If $\Phi^{(t)}$ is the current value of the correlation matrix, then in the $(t+1)$-th step, a candidate matrix Φ^c is generated by specifying a random walk chain $\Phi^c = \Phi^{(t)} + H$, where $H = (h_{ij})$ is an increment matrix with $E(h_{ij}) = 0$ and $h_{ii} = 0$, for all i and j. Let γ be the smallest eigenvalue of $\Phi^{(t)}$. Then the elements of the increment matrix H can be generated using the Hit-and-Run algorithm which involves the following steps:

(a) generate a sequence of iid standard normal deviates $z_{12}, z_{13}, \ldots, z_{(m-1)m}$, of length $m(m-1)/2$,
(b) generate a deviate d from $N(0, \sigma_d^2)$ which is truncated to the interval $(-\frac{\gamma}{\sqrt{2}}, \frac{\gamma}{\sqrt{2}})$,
(c) formulate the elements

$$h_{ij} = \frac{dz_{ij}}{\left(\sum_{j=1}^{J-1}\sum_{l=j+1}^{J} z_{jl}^2\right)^{(1/2)}}$$

for $i < j$, $h_{ii} = 0$, and $h_{ij} = h_{ji}$ for $i > j$.

Here σ_d^2 is a tuning constant that needs to be chosen such that candidates are not rejected disproportionately. If γ^c is the smallest eigenvalue of the candidate matrix, then once a candidate is generated, it is accepted or rejected based on the following Metropolis-Hastings acceptance probability

$$\min\left\{\frac{L(\Phi^c \mid .)p(\text{vec}(\Phi^c))\left(\Phi\left(\frac{\gamma^c}{\sqrt{2}\sigma_d}\right) - \Phi\left(\frac{-\gamma^c}{\sqrt{2}\sigma_d}\right)\right)}{L(\Phi^{(m)} \mid .)p(\text{vec}(\Phi^{(t)}))\left(\Phi\left(\frac{\gamma}{\sqrt{2}\sigma_d}\right) - \Phi\left(\frac{-\gamma}{\sqrt{2}\sigma_d}\right)\right)}, 1\right\}$$

where $\Phi(.)$ is the standard normal cumulative distribution function. If the candidate is accepted then $\Phi^{(t+1)} = \Phi^{(c)}$, otherwise $\Phi^{(t+1)} = \Phi^{(t)}$.

8. The full conditional distribution for the precision matrix Δ^{-1} of the individual specific factor means $\boldsymbol{\nu}_i$, is a Wishart distribution,

$$W\left(\delta + I, \left[\sum_{i=1}^{I} \boldsymbol{\nu}_i \boldsymbol{\nu}_i^T + \delta\Omega\right]^{-1}\right)$$

9. The full conditional distributions for the diagonal elements of the individual specific measurement error variances in Θ_i i.e., $\theta_{i,k}$, $k = 1$ to p, are independent inverse gamma distributions

$$IG\left(\frac{n_i}{2}+a_k,\left[\frac{\sum_{j=1}^{n_i}(y_{ijk}-\alpha_k-\boldsymbol{\lambda}_{ik}'\boldsymbol{\xi}_{ij})^2}{2}+b_k^{-1}\right]^{-1}\right),$$

where $\boldsymbol{\lambda}_{ik}$ contains the elements of row k of Λ_i.

10. The full conditional for the hyperparameter b_k is the same as for the heterogeneous covariance model and is given by Equation (21) of the paper.
11. The full conditional for the hyperparameter $a_k' = \log(a_k)$ is described in Item 11 in section 3.2.

A. Appendix 2: Heterogeneous Intercepts and Heterogeneous Covariance Model

In this appendix we describe the model that allows for heterogeneous intercepts, heterogeneous factor covariances and heterogeneous measurement errors across individuals. The model for an arbitrary individual i, can be written as

$$\begin{aligned}\mathbf{y}_{ij} &= \boldsymbol{\alpha}_i + \Lambda\boldsymbol{\xi}_{ij} + \boldsymbol{\epsilon}_{ij},\\ \boldsymbol{\xi}_{ij} &\sim N(\mathbf{0}, \Phi_i),\\ \boldsymbol{\epsilon}_{ij} &\sim N(\mathbf{0}, \Theta_i),\end{aligned} \tag{A1}$$

for observations $j = 1$ to n_i. The second stage population distribution that specifies the heterogeneity in the individual-level intercept parameters can be written as:

$$\boldsymbol{\alpha}_i \sim N(\boldsymbol{\mu}, \Sigma_b) \tag{A2}$$

The intercepts for the different individuals are assumed to originate from a multivariate normal distribution with population mean $\boldsymbol{\mu}$ and a $(p \times p)$ covariance matrix Σ_b. A factor analytic structure can be further imposed on Σ_b such that $\Sigma_b = \Lambda_b\Phi_b\Lambda_b' + \Theta_b$. This is akin to representing the intercepts $\boldsymbol{\alpha}_i$ by the equation $\boldsymbol{\alpha}_i = \boldsymbol{\mu} + \Lambda_b\boldsymbol{\delta}_i + \mathbf{e}_i$, where $\boldsymbol{\delta}_i$ represents the level-two factor score and $\mathbf{e}_i \sim N(\mathbf{0}, \Theta_b)$. In addition, the heterogeneity in the covariance parameters can be represented by the population distributions

$$\begin{aligned}\Phi_i &\sim IW(\rho, R)\\ \Theta_i &\sim \prod_{k=1}^{p} IG(a_k, b_k) \qquad i = 1 \quad \text{to} \quad I.\end{aligned} \tag{A3}$$

These population distributions are assumed to be mutually independent. Notice that when the individual-specific factor covariances are all the same and the measurement error variances are identical across individuals, we obtain the traditional multilevel model (Longford & Muthén, 1992; McDonald & Goldstein, 1989)

Priors and Full conditionals: We describe succinctly all the required full conditionals and the priors associated with the unknown parameters .

1. The overall mean $\boldsymbol{\mu}$ has a normal prior $N(\boldsymbol{\eta}, C)$. Therefore the full conditional distribution can be written as a multivariate normal distribution given by $p(\boldsymbol{\mu} \mid \{\boldsymbol{\alpha}_i\}, \Sigma_b) = N(\tilde{\boldsymbol{\mu}}, V_\mu)$ where

$$V_\mu^{-1} = C^{-1} + I\Sigma_b^{-1} \quad \text{and} \quad \tilde{\boldsymbol{\mu}} = V_\mu\left(C^{-1}\boldsymbol{\eta} + \sum_{i=1}^{I}\Sigma_b^{-1}\boldsymbol{\alpha}_i\right).$$

2. The full conditional for the individual specific intercept $\boldsymbol{\alpha}_i$ is normal $N(\tilde{\boldsymbol{\alpha}}_i, V_{\alpha_i})$, where

$$V_{\alpha_i} = \Sigma_b^{-1} + n_i \Sigma_i^{-1} \quad \text{and} \quad \tilde{\boldsymbol{\alpha}}_i = V_{\alpha_i} \left[\Sigma_b^{-1} \boldsymbol{\mu} + \sum_{j=1}^{n_i} \Sigma_i^{-1} \mathbf{y}_{ij} \right].$$

The covariance matrix $\Sigma_i = \Lambda \Phi_i \Lambda' + \Theta_i$.

3. The full conditional distribution for the free elements in a row of the matrix Λ_b is multivariate normal. Let $\boldsymbol{\lambda}_{b,k}$ be the vector of free elements in row k. The prior for $\boldsymbol{\lambda}_{b,k}$ is given by $p(\boldsymbol{\lambda}_{b,k}) = N(\mathbf{g}_{b,k}, H_{b,k})$. Let $\tilde{\boldsymbol{\delta}}_{ik}$ be the vector of factor scores corresponding to the elements in row k of $\boldsymbol{\Lambda}_b$ that are set to one and let $\boldsymbol{\delta}_{-ik}$ contain the remaining factor scores from $\boldsymbol{\delta}_i$. Form the adjusted variable $\tilde{\alpha}_{ik} = \alpha_{ik} - \boldsymbol{\iota}' \tilde{\boldsymbol{\delta}}_{ik} - \mu_k$, where $\boldsymbol{\iota}$ is a vector of ones. Given the prior, the vector $\boldsymbol{\lambda}_{b,k}$ can be sampled from the full conditional distribution given by

$$p(\boldsymbol{\lambda}_{b,k} \mid \{\tilde{\alpha}_{ik}\}, \{\boldsymbol{\delta}_{-ik}\}, \theta_{b,k}) = N\left(D_k \left[\sum_{i=1}^{I} \theta_{b,k}^{-1} \boldsymbol{\delta}_{-ik} \tilde{\alpha}_{ik} + H_{b,k}^{-1} \mathbf{g}_{b,k} \right], D_k \right)$$

where $D_k^{-1} = \sum_{i=1}^{I} \theta_{b,k}^{-1} \boldsymbol{\delta}_{-ik} \boldsymbol{\delta}_{-ik}' + H_{b,k}^{-1}$.

4. Given a prior $\Phi_b^{-1} \sim W(\rho_b, (\rho_b R_b)^{-1})$, the full conditional for Φ_b^{-1} can be written as

$$p(\Phi_b^{-1} | \{\boldsymbol{\delta}_i\}) = W\left(\rho_b + I, \left[\sum_{i=1}^{I} \boldsymbol{\delta}_i \boldsymbol{\delta}_i' + \rho_b R_b \right]^{-1} \right).$$

5. Let the priors for the diagonal elements of the level-two measurement error variance matrix Θ_b be independent inverse Gamma distributions $IG(r_k, s_k)$, $k = 1$ to p. The full conditional distributions for the diagonal elements of Θ_b, i.e., $\theta_{b,k}$, $k = 1$ to p are then independent inverse gamma distributions given by

$$p(\theta_{b,k} | \boldsymbol{\mu}, \{\boldsymbol{\alpha}_i\}, \{\boldsymbol{\delta}_i\}, \Lambda_b) = IG\left(I/2 + r_k, \left[\sum_{i=1}^{I} (\alpha_{ik} - \mu_k - \boldsymbol{\lambda}_{b,k}' \boldsymbol{\delta}_i)^2/2 + s_k^{-1} \right]^{-1} \right).$$

6. The full conditional for Φ_i^{-1} is a Wishart $W(\rho + n_i, R_{pos})$, where

$$R_{pos} = \left(\sum_{j=1}^{n_i} \xi_{ij} \xi_{ij}' + R^{-1} \right)^{-1}.$$

7. Given a conjugate prior $R^{-1} \sim W(\gamma, (\gamma S)^{-1})$, the full conditional for R^{-1} is identical to that in (18) of the paper.
8. Given a truncated univariate normal prior $\rho' \sim tN(0, \tau)$, the full conditional for $\rho' = \log(\rho)$ is the same as in Item (8) of section 3.2 in the paper.
9. The full conditional for the diagonal elements in Θ_i are independent inverse gamma distributions

$$p(\theta_{i,k} \mid \boldsymbol{\lambda}_k, \alpha_{ik}, \{\boldsymbol{\xi}_{ij}\}_{j=1}^{n_i}, a_k, b_k) = IG\left(n_i/2 + a_k, \left[\sum_{j=1}^{n_i} (y_{ijk} - \alpha_k - \boldsymbol{\lambda}_k' \boldsymbol{\xi}_{ij})^2/2 + b_k^{-1} \right]^{-1} \right)$$

where $\boldsymbol{\lambda}_k$ contains the elements of row k of Λ.

10. The full conditional for the kth hyperparameter, b_k, $k = 1$ to p of the inverse gamma population distribution of the measurement error variances is the same as given in Item (10) in section 3.2.

11. The full conditional for the hyperparameter $a'_k = \log(a_k)$ is the same as described in Item (11) in section 3.2 of the paper.
12. The full conditional distribution for the free elements in a row of the matrix $\mathbf{\Lambda}$ is multivariate normal. Let $\boldsymbol{\lambda}_k$ be the vector of free elements in row k. The prior for $\boldsymbol{\lambda}_k$ is given by $p(\boldsymbol{\lambda}_k) = N(\mathbf{g}_k, H_k)$. Let $\tilde{\boldsymbol{\xi}}_{ijk}$ be the vector of factor scores corresponding to the elements in row k of $\mathbf{\Lambda}$ that are set to one and let $\boldsymbol{\xi}_{-ijk}$ contain the remaining factor scores from $\boldsymbol{\xi}_{ij}$. Form the adjusted variable $\tilde{y}_{ijk} = y_{ijk} - \boldsymbol{\iota}'\tilde{\boldsymbol{\xi}}_{ijk} - \alpha_{ik}$, where $\boldsymbol{\iota}$ is a vector of ones. Given the prior, the vector $\boldsymbol{\lambda}_k$ can be sampled from the full conditional distribution given by

$$p(\boldsymbol{\lambda}_k \mid \{\tilde{y}_{ijk}\}, \{\boldsymbol{\xi}_{-ijk}\}, \theta_{i,k}) = N\left(D_k\left[\sum_{i=1}^{I}\sum_{j=1}^{n_i}\theta_{i,k}^{-1}\boldsymbol{\xi}_{-ijk}\tilde{y}_{ijk} + H_k^{-1}\mathbf{g}_k\right], D_k\right)$$

where

$$D_k^{-1} = \sum_{i=1}^{I}\sum_{j=1}^{n_i}\theta_{i,k}^{-1}\boldsymbol{\xi}_{-ijk}\boldsymbol{\xi}'_{-ijk} + H_k^{-1}.$$

References

Ansari, A., & Jedidi, K. (2000). Bayesian factor analysis for multilevel binary observations. *Psychometrika, 65*, 475–496.

Arminger, G., & Muthén, B. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika, 63*, 271–300.

Arminger, G., & Stein, P. (1997). Finite mixtures of covariance structure models with regressors: Loglikelihood function, minimum distance estimation, fit indices, and a complex example. *Sociological Methods & Research, 26*, 148–182.

Arminger, G., Stein, P., & Wittenberg, J. (1998). Mixtures of conditional mean- and covariance structure models. *Psychometrika, 64*, 475–494.

Bagozzi, R.P., Gopinath, M., & Nyer P.U. (1999). The role of emotions in marketing. *Journal of the Academy of Marketing Science, 27*, 184–206.

Bartholomew, D.J. (1981). Posterior analysis of the factor model. *British Journal of Mathematical and Statistical Psychology, 34*, 93–99.

Best, N.G., Cowles, M.K. & Vines, S.K. (1995) *CODA: Convergence diagnostics and output analysis software for Gibbs sampler output, Version 0.3.* (Software Manual). Cambridge, U.K.: Institute of Public Health, Biostatistics Unit-MRC.

Bladfield, E. (1980). *Clustering of observations from finite mixtures with structural information* (Jyvaskyla Studies in Computer Science, Economics and Statistics, No. 2). Jyvaskyla, Finland: Jyvaskyla University.

Bock, R.D., & Gibbons, R.D., (1996) High-dimensional multivariate probit analysis. *Biometrics, 52*, 1183–1194.

Bollen, K.A. (1989). *Structural equations with latent variables*. New York, NY: Wiley Interscience.

Chen, Ming-Hui, & Schmeiser, B.W. (1993). Performance of the Gibbs, hit-and-run, and Metropolis samplers. *Journal of Computational and Graphical Statistics, 2*, 251–272.

Dey, Dipak K., & Chen, M.H. (1998). Bayesian analysis of correlated binary data models. *Sankhya, Series A, 60*, 322–343.

Dube, L., Belanger, M., & Trudeau, E. (1996). The role of emotions in health care satisfaction. *Journal of Health Care Marketing, 16*, 45–51.

Dube, L., & Morgan, M.S. (1996). Trend effects and gender differences in retrospective judgments of consumption emotions. *Journal of Consumer Research, 23*, 156–162.

Geisser, S., & Eddy, E. (1979). A predictive approach to model selection. *Journal of the American Statistical Association, 74*, 153–60.

Gelfand, A.E. (1996). Model determination using sampling-based methods. In W.R. Gilks, S. Richardson & D.J. Spiegelhalter (Eds.), *Markov chain Monte Carlo in practice* (pp. 145–161). London, U.K.: Chapman & Hall.

Gelman, A., Carlin, J.B., Stern, H.S., & Rubin, D.B. (1996). Posterior Predictive Assessment of Model Fitness (with discussion). *Statistica Sinica 6*, 733–807.

Geman S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions of Pattern Analysis and Machine Intelligence, 6*, 721–741.

Gilks, W.R., Richardson, S., & Spiegelhalter, D.J. (1996). *Markov chain Monte Carlo in practice*. London, U.K.: Chapman & Hall.

Goldstein, H., & McDonald, R.P. (1988). A general model for the analysis of multilevel data. *Psychometrika, 53*, 455–467.

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika, 57*, 97–109.

Jedidi, K., Jagpal, H.S., & DeSarbo, W.S. (1997a), Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science, 16*, 39–59.

Jedidi, K., Jagpal, H.S., & DeSarbo, W.S. (1997b), STEMM: A general finite-mixture structural equation model. *Journal of Classification, 14*, 23–50.
Jöreskog, K.G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*, 409–426.
Kaplan, D., & Elliot, P.E. (1997). A Didactic example of multilevel structural equation modeling applicable to the study of organizations. *Structural Equation Modeling: A Multidisciplinary Journal, 4*, 1–25.
Kass, R.E., & Raftery, A.E. (1995). Bayes factors. *Journal of American Statistical Association, 90*, 773–795.
Lee, S.-Y (1981). A Bayesian approach to confirmatory factor analysis. *Psychometrika, 46*, 153–160.
Longford, N.T. (1993). *Random coefficient models*, New York, NY: Oxford University Press.
Longford, N.T., & Muthén, B. (1992) Factor analysis for clustered observations. *Psychometrika, 57*, 581–597.
Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addision-Wesley.
Mano, H., & Oliver, R.L. (1993). Assessing the dimensionality and structure of the consumption experience: Evaluation, feeling and satisfaction. *Journal of Consumer Research, 20*(3), 451–466.
Martin, J.K., & McDonald, R.P. (1975) Bayesian estimation in unrestricted factor analysis: A treatment for Heywood cases. *Psychometrika, 40*, 505–517.
McDonald, R.P., & Goldstein, H. (1989). Balanced versus unbalanced designs for linear structural relations in two-level data. *British Journal of Mathematical and Statistical Psychology, 42*, 214–232.
Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., & Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics, 21*, 1087–1091.
Muthén, B. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika, 54*, 557–585.
Muthén, B. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research, 22*, 376–398.
Muthén, B. & Satorra, A. (1989). Multilevel aspects of varying parameters in structural models. In R.D. Bock (Ed.), *Multilevel analysis of educational data* (pp. 87–99). New York, NY: Academic Press.
Oliver, R.L. (1993). Cognitive, affective, and attribute bases of the satisfaction response. *Journal of Consumer Research, 20*(December), 418–430.
Richins, M. (1997). Measuring emotions in the consumption experience. *Journal of Consumer Research, 24*(2), 127–146.
Scheines, R., Hoijtink, H., & Boomsma A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika, 64*, 37–52.
Shi, J., & Lee, S. (1997). A Bayesian estimation of factor score in confirmatory factor model with polytomous, censored or truncated data. *Psychometrika, 62*, 29–50.
Sörbom, D. (1981). Structural equation models with structured means. In K.G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, and prediction* (pp. 183–195). Amsterdam, Netherlands: North Holland.
Tanner, M.A., & Wong, W.H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of American Statistical Association, 82*, 528–550.
Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics, 22*, 1701–1762.
Westbrook, R.A., & Oliver, R.L. (1991). The dimensionality of consumption emotion patterns and consumer satisfaction. *Journal of Consumer Research, 18*, 184–191.
Yao, G., & Böckenholt, U. (1999). Bayesian estimation of Thurstonian ranking models based on the Gibbs sampler. *British Journal of Mathematical and Statistical Psychology, 52*, 79–92.
Yung, Y.-F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika, 62*, 297–330.

Manuscript received 24 JAN 2000
Final version received 1 DEC 2000