

# A Hierarchical Bayesian Methodology for Treating Heterogeneity in Structural Equation Models

Asim Ansari • Kamel Jedidi • Sharan Jagpal

*Graduate School of Business, Columbia University, Uris Hall, 3022 Broadway,  
New York, New York 10027-6902*

*maa48@columbia.edu • kj7@columbia.edu*

*Faculty of Management, Rutgers University, Newark, NJ, jagpal007@gateway.net*

## Abstract

Structural equation models are widely used in marketing and psychometric literature to model relationships between unobserved constructs and manifest variables and to control for measurement error. Most applications of structural equation models assume that data come from a homogeneous population. This assumption may be unrealistic, as individuals are likely to be heterogeneous in their perceptions and evaluations of unobserved constructs. In addition, individuals may exhibit different measurement reliabilities. It is well-known in statistical literature that failure to account for unobserved sources of individual differences can result in misleading inferences and incorrect conclusions.

We develop a hierarchical Bayesian framework for modeling general forms of heterogeneity in partially recursive structural equation models. Our framework elucidates the motivations for accommodating heterogeneity and illustrates theoretically the types of misleading inferences that can result when unobserved heterogeneity is ignored. We describe in detail the choices that researchers can make in incorporating different forms of measurement and structural heterogeneity. Current random-coefficient models in psychometric literature can accommodate heterogeneity solely in mean structures. We extend these models by allowing for heterogeneity both in mean and covariance structures. Specifically, in addition to heterogeneity in measurement intercepts and factor means, we account for heterogeneity in factor covariance structure, measurement error, and structural parameters. Models such as random-coefficient factor analysis, random-coefficient second-order factor analysis, and random-coefficient, partially recursive simultaneous equation models are special cases of our proposed framework. We also develop Markov Chain Monte Carlo (MCMC) procedures to perform Bayesian inference in partially recursive, random-coefficient structural equation models. These procedures provide individual-specific estimates of the factor scores, structural coefficients, and other model parameters.

We illustrate our approach using two applications. The

first application illustrates our methods on synthetic data, whereas the second application uses consumer satisfaction data involving measurements on satisfaction, expectation disconfirmation, and performance variables obtained from a panel of subjects. Our results from the synthetic data application show that our Bayesian procedures perform well in recovering the true parameters. More importantly, we find that models that ignore heterogeneity can yield a severely distorted picture of the nature of associations among variables and can therefore generate misleading inferences. Specifically, we find that ignoring heterogeneity can result in inflated estimates of measurement reliability, wrong signs of factor covariances, and can yield attenuated model fit and standard errors. The results from the consumer satisfaction study show that individuals vary both in means and covariances and indicate that conventional psychometric methods are not appropriate for our data. In addition, we find that heterogeneous models outperform the standard structural equation model in predictive ability. Managerially, we show how one can use the individual-level factor scores and structural parameter estimates from the Bayesian approach to perform quadrant analysis and refine marketing policy (e.g., develop a one-on-one marketing policy).

The framework introduced in this paper and the inference procedures we describe should be of interest to researchers in a wide range of disciplines in which measurement error and unobserved heterogeneity are problematic. In particular, our approach is suitable for studies in which panel data or multiple observations are available for a given set of respondents or objects (e.g., firms, organizations, markets). At a practical level, our procedures can be used by managers and other policymakers to customize marketing activities or policies.

Future research should extend our procedures to deal with the general nonrecursive structural equation model and to handle binary and ordinal data situations.

*(Structural Equation Models; Heterogeneity; Hierarchical Bayes; MCMC Procedures; Metropolis-Hastings Algorithm; Gibbs Sampling; Customer Satisfaction)*

## 1. Introduction

Structural equation models are widely used in marketing and psychometric literature to model relationships among unobserved constructs and manifest variables, and to control for measurement error. Most applications of structural equation models assume that the data come from a homogeneous population, and hence implicitly ignore the influence of unobserved sources of heterogeneity. This approach is problematic: As the vast literature on heterogeneity in regression and discrete choice models clearly shows, a failure to adequately account for unobserved sources of heterogeneity is likely to result in misleading inferences.

Many different sources of unobserved heterogeneity are possible in structural equation models. To illustrate these sources, consider a product positioning study that uses a structural equation approach to model the structure of consumer preferences for automobiles. Suppose theory suggests that the salient unobserved perceived benefits that affect preference are safety, reliability, and economy, and assume that a set of scales are available to operationalize these unobservable constructs. The measurement model that captures consumer perceptions of the unobserved benefits can be affected by many types of unobserved heterogeneity. For example, the mean perceptual scores (i.e., factor means for safety, reliability, and economy) may vary in the population. The correlational structure of benefits can also vary across consumers. Thus, some consumers may exhibit a halo effect; for example, safety and reliability may be positively correlated for one type of consumers (e.g., novices) but may be orthogonal for another type (e.g., experts). Finally, consumers may provide information with different degrees of accuracy (e.g., variability in measurement error for any given scale item could differ across individuals). The structural model may also be affected by unobserved heterogeneity. For example, consumers may differ in their importance weights (i.e., the structural parameters) for perceived benefits. As we show in this paper, when such general types of measurement and structural heterogeneity are present, analyzing the pooled data using a conventional structural equation model that ignores the sources of individual differences can result in misleading inferences and incorrect conclusions. See also Muthén (1989), Jedidi et al. (1997), and

Jagpal (1999, Chapter 4) for more theoretical discussion and examples.

In marketing literature, researchers have used either a finite-mixture (Kamakura and Russell 1989) or a random-coefficient approach (Allenby and Rossi 1999, Chintagunta et al. 1991) for modeling heterogeneity. Jedidi et al. (1997) recently developed a finite-mixture approach for response-based segmentation in structural equation models. In psychometric literature on structural equations models, heterogeneity is typically handled using random-coefficient procedures (also called multilevel models; see Muthén 1989, 1994; Longford and Muthén 1992) that allow variation *only* in mean structures (i.e., in factor means or measurement intercepts). In many marketing applications (as in the product positioning example above), the data may exhibit heterogeneity not only in mean structures but also in structural parameters and covariance structures. Existing random-coefficient models cannot be used in such contexts. In fact, the classical procedures that are used for estimating these models cannot be adapted to accommodate general forms of heterogeneity. A random-coefficient approach that allows the specification and estimation of heterogeneity in both *mean* and *covariance* structures (i.e., in structural parameters, measurement error variances, and factor covariances) is clearly needed. In this paper we describe such an approach.

We develop a hierarchical Bayesian framework for modeling general forms of heterogeneity in partially recursive structural equation models. Our framework elucidates the motivations for accommodating heterogeneity and illustrates theoretically the types of misleading inferences that can result when unobserved heterogeneity is ignored. We describe in detail the choices that researchers can make in incorporating different forms of measurement and structural heterogeneity. We also develop Markov Chain Monte Carlo (MCMC) procedures to perform Bayesian inference in partially recursive, random-coefficient structural equation models.

The hierarchical Bayesian approach is eminently suitable for studies in which panel data or multiple observations are available for a given set of respondents or objects (e.g., firms). This approach allows for appropriate pooling of information while taking into

account heterogeneity. We illustrate our approach using two applications. The first application illustrates our methods on synthetic data, whereas the second application uses consumer satisfaction data involving measurements on satisfaction, expectation disconfirmation, and performance variables obtained from a panel of subjects. We find that models that ignore heterogeneity can yield a severely distorted picture of the nature of associations among variables and can therefore generate misleading inferences. Specifically, we show that ignoring heterogeneity leads to inflated estimates of measurement reliability, can generate wrong signs of factor covariances, and can result in attenuated model fit and standard errors. In addition, we find that heterogeneous models outperform the standard structural equation model in predictive ability.

The hierarchical Bayesian approach provides several theoretical and practical advantages over classical methods in treating heterogeneity in structural equation models. (See Allenby and Rossi 1999 for a lucid discussion of the advantages of the Bayesian methodology in the context of choice models.) From a practical viewpoint, Bayesian methods allow the flexible incorporation of prior information, whenever available, about model parameters. In addition, Bayesian methods allow the estimation of individual-specific estimates while accounting for the uncertainty in such estimates. Specifically, in our modeling context, the hierarchical Bayesian structural equation methodology provides individual-specific estimates of the factor scores, structural coefficients, and other model parameters. Thus, managers can use our approach to target selected individuals or groups in a highly focused manner using customized marketing strategies.

From a statistical viewpoint, sampling-based Bayesian methods are appealing because they do not rely on asymptotic theory, a practice that is known to be misleading when the sample size is small (see Scheines et al. 1999). The Bayesian approach is well-suited for dealing with unbalanced designs, a common problem in marketing studies. As discussed later, by using MCMC procedures one can obtain simulation-based estimates of the parameters of a structural equation model. This circumvents the evaluation of complex multidimensional integrals that are required to implement maximum likelihood methods on panel data.

The rest of the paper is organized as follows. Section 2 presents the hierarchical Bayesian approach for estimating structural equation models. Section 3 discusses the specification of the priors and briefly describes the MCMC method for inference. Section 4 describes the results from a synthetic data set. Section 5 illustrates our approach on customer satisfaction study, and § 6 provides a summary and discusses directions for future research.

## 2. The Hierarchical Structural Equation Model

A hierarchical structural equation model can be specified in two stages. In the first stage, structural and measurement models are specified for each individual. In the second stage, a population distribution is specified to model variation in individual-level parameters across all individuals.

Let  $i = 1$  to  $I$  represent individuals and let  $j = 1$  to  $N_i$  index the observations belonging to the  $i$ th individual. Suppose each individual provides multivariate observations on  $q$  endogenous and  $p$  exogenous indicator (manifest) variables,  $y_{ij}$  and  $x_{ij}$ , respectively. The associations among these manifest variables can be described in terms of latent constructs using a structural equation model. Let the  $(m \times 1)$  vector  $\xi_{ij}$  and the  $(n \times 1)$  vector  $\eta_{ij}$ , respectively, contain the exogenous and endogenous latent variables. As is well-known, the complete model for individual  $i$  consists of a measurement model that describes the relationship between observed and latent variables and a structural model that relates the exogenous and endogenous latent variables. Specifically, the measurement model for individual  $i$  can be written as follows:

$$\begin{aligned} x_{ij} &= \alpha_{i,x} + \Lambda_{i,x}\xi_{ij} + \delta_{ij}, \\ y_{ij} &= \alpha_{i,y} + \Lambda_{i,y}\eta_{ij} + \epsilon_{ij}, \end{aligned} \quad (1)$$

for  $j = 1$  to  $N_i$ , where  $\alpha_{i,x}$  and  $\alpha_{i,y}$  are  $p \times 1$  and  $q \times 1$  measurement intercept vectors, respectively, for the exogenous and endogenous indicator variables. The  $(p \times m)$  matrix  $\Lambda_{i,x}$  and the  $(q \times n)$  matrix  $\Lambda_{i,y}$  contain the factor loadings. The terms  $\delta_{ij} \sim N(0, \Theta_{i,x})$  and  $\epsilon_{ij} \sim N(0, \Theta_{i,y})$ , respectively, represent the vectors of measurement errors. The  $p \times p$  matrix  $\Theta_{i,x}$  and the  $q \times q$  matrix  $\Theta_{i,y}$  are diagonal and contain the measurement

error variances. The  $m$  latent factors in  $\xi_{ij}$  are assumed to be normally distributed  $N(\nu_i, \Phi_i)$ , where  $\nu_i$  is a  $(m \times 1)$  vector of factor means and  $\Phi_i$  is a  $m \times m$  covariance matrix of factor scores.

The structural model that relates the latent constructs  $\xi_{ij}$  and  $\eta_{ij}$  for each individual  $i$  is

$$\mathbf{B}_i \eta_{ij} = \gamma_{0i} + \Gamma_i \xi_{ij} + \zeta_{ij} \quad (2)$$

where  $\mathbf{B}_i$  is a  $(n \times n)$  matrix of structural parameters specifying the links among the endogenous latent variables,  $\gamma_{0i}$  is a  $(n \times 1)$  vector of structural intercept terms,  $\Gamma_i$  is a  $(n \times m)$  coefficient matrix denoting the effects of  $\xi_{ij}$  on  $\eta_{ij}$ , and  $\zeta_{ij}$  is a  $n \times 1$  vector of disturbances. The disturbances  $\zeta_{ij}$  are assumed to be uncorrelated with  $\xi_{ij}$  and are distributed  $N(0, \Psi_i)$ , where  $\Psi_i$  is a  $(n \times n)$  covariance matrix that captures the residual variation in the structural model. In our model,  $\mathbf{B}_i$  is restricted to being triangular.

When each individual provides enough observations (i.e., when  $N_i$  is large for all  $i$ ), separate models can be estimated for each individual. In most situations, however, only a limited number of observations are available for at least some individuals, thus precluding individual-level analyses. Moreover, in most substantive enquiries, researchers are primarily interested in summarizing the relationships among constructs at the population level while controlling for individual differences. A hierarchical approach is, therefore, eminently suitable in such situations, because this method appropriately pools information across all individuals to estimate the desired relationships. A hierarchical model accounts for observed and unobserved sources of heterogeneity by specifying a second-stage population distribution that describes how individual-specific parameters in Equations (1) and (2) vary in the population. Specifically, in the second stage of the hierarchical model, the individual-level parameters  $\phi_i = \{\alpha_{i,x}, \alpha_{i,y}, \Lambda_{i,x}, \Lambda_{i,y}, \nu_i, \Phi_i, \Theta_{i,x}, \Theta_{i,y}, \mathbf{B}_i, \gamma_{0i}, \Gamma_i, \Psi_i\}$  are specified as random variables drawn from a general population distribution  $h(\phi_i)$ .

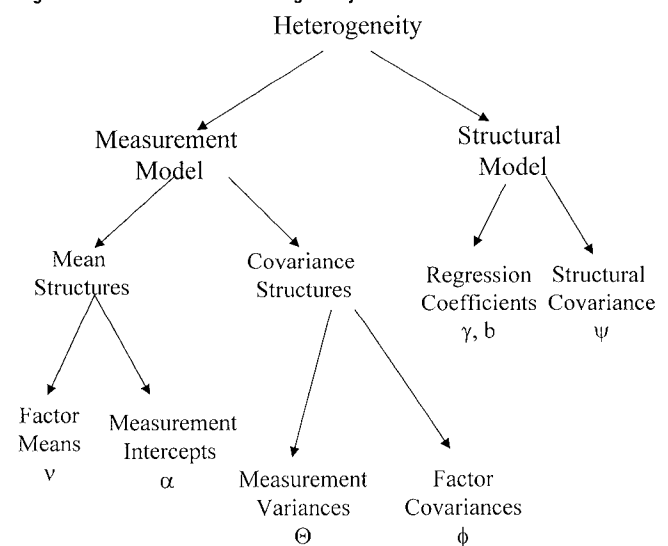
It is important to note that a very general hierarchical model that allows all parameters to vary freely across individuals is not identifiable. We need to restrict certain parameters at both the individual and population levels for identification purposes. As in conventional structural equation models, we can also

include additional restrictions based on prior theory to obtain parsimonious models. Figure 1 provides a graphical summary of the potential sources of heterogeneity in structural equation models. We discuss below how different types of heterogeneity can be specified in the measurement and structural models. We also discuss the types of restrictions that are needed for the identification of parameters.

## 2.1. Measurement Heterogeneity

Previous attempts to capture heterogeneity using random parameter formulations have focused *solely* on capturing individual differences in mean structures. In the psychometric literature on multilevel models (see Muthén 1989, 1994; Goldstein and McDonald 1988), researchers have either allowed measurement intercepts to vary across individuals or have assumed that factor means differ across individuals (see Figure 1). In many studies, researchers are primarily interested in modeling individual differences in underlying constructs. We follow this approach and allow factor means to differ across individuals. Specifically, we set  $\alpha_{i,x} = \alpha_x$  and  $\alpha_{i,y} = \alpha_y$  for all  $i$  and assume that individual-level factor means  $\nu_i$  are distributed multivariate normal  $N(0, \Delta)$  in the population. We assume a zero mean for this population distribution to fix the location of the grand mean and ensure identification. This is analogous to setting the factor means of the first group to

Figure 1 Sources of Heterogeneity



zero in multigroup structural equation models (Jöreskog 1971, Sörbom 1981).

Allowing solely for differences in mean structures results in a model that assumes that individuals have identical covariance structures. As is clear from the product positioning example in §1, the assumption implicit in multilevel models that factor covariances and measurement error variances are invariant may not hold in empirical applications. For example, experts may provide more precise measures, i.e., they may have lower measurement error variances. Moreover, experts often possess more refined knowledge structures than novices; hence experts are able to discriminate finely among latent constructs. Thus, factor covariance matrices,  $\Phi_i$ , may be nearly diagonal for experts; in contrast, factor covariance matrices for novices may exhibit high correlations because of halo effects.

The literature on multigroup structural equation models (Jöreskog 1971, Sörbom 1981) allows for differences in covariance structures across a small number of a priori (exogenously) specified groups. The multigroup model assumes a fixed-effects approach to heterogeneity and typically requires a large number of observations per group. Recent research on finite mixtures of structural equation models (Jedidi et al. 1997) and confirmatory factor models (Yung 1997) has also considered differences in covariance structures across endogenously determined groups of individuals in which the number of groups is small. In this paper, we are interested in a random-effects formulation of heterogeneity. Consequently, we allow the factor covariances  $\Phi_i$  and the measurement error variances  $\Theta_i$  to vary across individuals by specifying continuous population distributions. Specifically, we assume that the precision matrices  $\Phi_i^{-1}$  come from a common Wishart population distribution  $W(\rho, \mathbf{R})$ , where  $\rho$  is the degrees of freedom and  $\mathbf{R}$  is a  $m \times m$  positive definite scale matrix. To make the factor scores comparable across individuals (see Yung 1997) and to preserve the definition and interpretability of the constructs, we assume that the factor loading matrices are invariant across individuals, i.e.,  $\Lambda_{i,x} = \Lambda_x$  and  $\Lambda_{i,y} = \Lambda_y$  for  $i = 1$  to  $I$ . Because the scales of the latent factors are arbitrary, we can proceed by setting the appropriate elements in the loading matrices to unity. As in

other structural equation models, additional constraints must be imposed on a case-by-case basis to ensure identification. Finally, we allow the measurement error variances  $\Theta_{i,x}$  and  $\Theta_{i,y}$  to vary across individuals. Specifically, we assume that each measurement error variance comes from an independent inverse gamma population distribution  $IG(0; a, b)$ .

To understand the implications of ignoring measurement heterogeneity, we need to examine the unconditional implied covariance matrix  $\Sigma$  of the structural equation model. Using the identity  $\text{Var}[w] = E[\text{Var}[w|.] ] + \text{Var}[E[w|.]]$ , we can show that the implied aggregate covariance matrix

$$\Sigma = \begin{pmatrix} \Lambda_y \mathbf{B}^{-1} (\Gamma(E[\Phi_i] + \Lambda) \Gamma' + \Psi) \mathbf{B}^{-1'} \Lambda_y' + E[\Theta_{iy}] & \Lambda_y \mathbf{B}^{-1} \Gamma(E[\Phi_i] + \Lambda) \Lambda_x' \\ \Lambda_x (E[\Phi_i] + \Lambda) \Gamma' \mathbf{B}^{-1'} \Lambda_y' & \Lambda_x (E[\Phi_i] + \Lambda) \Lambda_x' + E[\Theta_{ix}] \end{pmatrix} \quad (3)$$

Typically, researchers are interested in understanding the within-individual factor covariance structure represented in  $E[\Phi_i]$ . It is clear from the above equation that  $E[\Phi_i]$  and  $\Lambda$  are not separately identifiable based on the aggregate covariance matrix. Hence, an aggregate model would generate  $\Phi^{\text{Agg}} = E[\Phi_i] + \Lambda$ . Thus, a conventional structural equation analysis on data containing measurement heterogeneity can generate two types of misleading inferences about the factor structure. First, as the elements of  $\Phi^{\text{Agg}}$  are necessarily larger than the corresponding elements in the population matrix  $E[\Phi_i]$ , factor reliability estimates based on  $\Phi^{\text{Agg}}$  will be inflated (Lord and Novick 1968, Muthén 1989). Secondly, the magnitude and the signs of the covariances in  $\Phi^{\text{Agg}}$  will be distorted. For example, if the elements  $E[\Phi_{i,kj}]$  and  $\Lambda_{kj}$  are of the same sign, then the magnitude of  $\Phi_{kj}^{\text{Agg}}$  will be amplified. If, on the other hand,  $E[\Phi_{i,kj}]$  and  $\Lambda_{kj}$  are of the opposite sign, then  $\Phi_{kj}^{\text{Agg}}$  may get attenuated or may have the wrong sign. Such a situation can clearly have unfortunate consequences for theory development. Other types of distortions are also possible. We report these in the applications to follow.

## 2.2. Structural Heterogeneity

The structural model is crucially important to decision makers, because it includes structural parameters that

measure the impact of key antecedent variables on outcomes. In many research studies, this relationship is of primary interest, and researchers often focus on studying individual differences in structural parameters. Previous research on multilevel models has not allowed for such forms of heterogeneity. In contrast, we allow structural coefficients to vary across individuals. Specifically, let vector  $\pi_i$  contain the terms in  $\mathbf{B}_i$ ,  $\gamma_{i0}$ , and  $\Gamma_i$ . Then  $\pi_i$  is assumed to be distributed multivariate normal  $N(\mathbf{Z}_i\vartheta, \mathbf{Y})$ , where  $\mathbf{Z}_i$  is a matrix that contains individual specific covariates (e.g., gender or age). Such a specification allows for both observed and unobserved sources of heterogeneity. Specifically, the inclusion of covariates through  $\mathbf{Z}_i$  allows a researcher to test for specific moderating effects. If individual-level covariates are not available, then  $\mathbf{Z}_i$  reduces to an identity matrix. The parameters in  $\vartheta$  explain the individual differences in the structural parameters in terms of the individual-level covariates. The  $\mathbf{Y}$  matrix captures the covariation in the structural parameters resulting from unobserved individual-level variables and, for parsimony, is assumed to be invariant across individuals. Because of the scale indeterminacy of the endogenous factors, we impose  $E(\gamma_{i0}) = 0$  for identification. Although the population mean of the structural intercepts is fixed to zero, note that the individual-level intercepts are estimable subject to this constraint. This is analogous to fixing the intercept of one group to zero in a multigroup analysis (see Jöreskog 1971, Sörbom 1981).

Taking into account the individual-level models and the heterogeneity specifications, the complete two-stage model can be written as:

**Stage 1:**

$$\begin{aligned} x_{ij} &= \alpha_x + \Lambda_x \xi_{ij} + \delta_{ij} \\ y_{ij} &= \alpha_y + \Lambda_y \eta_{ij} + \epsilon_{ij} \\ \mathbf{B}_i \eta_{ij} &= \gamma_{0i} + \Gamma_i \xi_{ij} + \zeta_{ij} \\ \delta_{ij} &\sim N(0, \Theta_{i,x}) \\ \epsilon_{ij} &\sim N(0, \Theta_{i,y}) \\ \xi_{ij} &\sim N(\nu_i, \Phi_i) \\ \zeta_{ij} &\sim N(0, \Psi) \end{aligned} \quad (4)$$

**Stage 2:**

$$\begin{aligned} \nu_i &\sim N(0, \Delta) \\ \Phi_i^{-1} &\sim W(\rho, \mathbf{R}) \\ \Theta_{ix} &\sim \prod_{k=1}^p IG(a_k, b_k) \\ \Theta_{iy} &\sim \prod_{k=1}^q IG(a_{y'k}, b_{y'k}) \\ \pi_i &\sim N(\mathbf{Z}_i\vartheta, \mathbf{Y}) \end{aligned} \quad (5)$$

Recall that, following the tradition in multigroup structural equation models, we assume that  $E(\nu_i) = 0$  and  $E(\gamma_{i0}) = 0$  to fix the origin of the factors. As is standard in structural equation models, we also restrict the loading of one indicator variable per factor to one to fix the scale of the factors. Note, however, that the identification of the structural equation model in Equation (4) requires that both the measurement and structural models are identified. Bollen (1989, pp. 88–104 and pp. 238–254) discusses the identification issues in structural equation models and provides general rules for identification that can be followed on a case-by-case basis.

The hierarchical structural equation model defined by Equations (4) and (5) subsumes a variety of models as special cases. These include hierarchical confirmatory (first-order and second-order) factor analysis, hierarchical multivariate regression, and hierarchical simultaneous equation models.

### 3. Inference

While the foregoing discussion on the different types of heterogeneity and its impact on inference is relevant for the general structural equation model, we now consider inference procedures for the case in which  $\mathbf{B}_i$  is triangular but  $\Psi_i$  is unrestricted. Let  $w_{ij} = \{y_{ij}, x_{ij}\}$  be the joint vector of manifest variables for an arbitrary observation  $j$  for individual  $i$ . According to the model represented in Equations (4) and (5), this observation comes from a multivariate normal distribution  $f_i(w_{ij}; \mu_i, \Sigma_i)$  with conditional mean vector

$$\mu_i = \begin{pmatrix} \alpha_y + \Lambda_y \mathbf{B}_i^{-1}(\gamma_{i0} + \Gamma_i \nu_i) \\ \alpha_x + \Lambda_x \nu_i \end{pmatrix} \quad (6)$$

and conditional covariance matrix

$$\Sigma_i = \begin{pmatrix} \Lambda_y \mathbf{B}_i^{-1} (\Gamma_i \Phi_i \Gamma_i' + \Psi) \mathbf{B}_i^{-1} \Lambda_y' + \Theta_{iy} & \Lambda_y \mathbf{B}_i^{-1} \Gamma_i \Phi_i \Lambda_x' \\ \Lambda_x \Phi_i \Gamma_i' \mathbf{B}_i^{-1} \Lambda_y' & \Lambda_x \Phi_i \Lambda_x' + \Theta_{ix} \end{pmatrix} \quad (7)$$

The likelihood for individual  $i$  is

$$L_i = \prod_{j=1}^{N_i} (2\pi)^{-(p+q)/2} \left| \Sigma_i \right|^{-1/2} \exp \left( -\frac{1}{2} (w_{ij} - \mu_i)' \Sigma_i^{-1} (w_{ij} - \mu_i) \right) \quad (8)$$

and the unconditional likelihood for a random sample of  $I$  individuals is given by the continuous mixture

$$L = \prod_{i=1}^I \int \cdots \int L_i(\mu_i(\varphi), \Sigma_i(\varphi)) h(\varphi) d\varphi \quad (9)$$

where  $h(\varphi)$  is the continuous population distribution that captures the heterogeneity in the parameters for the individuals. The unconditional likelihood  $L$  given in Equation (9) is a function of the parameters  $\varphi = \{\alpha_x, \alpha_y, \Lambda_x, \Lambda_y, \rho, \mathbf{R}, \Delta, \mathbf{a}_x, \mathbf{b}_x, \mathbf{a}_y, \mathbf{b}_y, \vartheta, \mathbf{Y}, \Psi\}$  and cannot be written in closed form, making maximum likelihood estimation extremely difficult. We therefore use a simulation-based Bayesian approach that uses MCMC methods to estimate the parameters.

Bayesian inference requires the specification of priors for the model parameters. Let  $\Lambda = \begin{pmatrix} \Lambda_x & 0 \\ 0 & \Lambda_y \end{pmatrix}$ ,  $\alpha = \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix}$ ,  $\Theta_i = \begin{pmatrix} \Theta_{ix} & 0 \\ 0 & \Theta_{iy} \end{pmatrix}$ ,  $\mathbf{a} = \begin{pmatrix} \mathbf{a}_x \\ \mathbf{a}_y \end{pmatrix}$ , and  $\mathbf{b} = \begin{pmatrix} \mathbf{b}_x \\ \mathbf{b}_y \end{pmatrix}$ . The unknown parameters for the model are then given by  $\varphi = \{\alpha, \Lambda, \rho, \mathbf{R}, \Delta, \mathbf{a}, \mathbf{b}, \vartheta, \mathbf{Y}, \Psi\}$ . Lee (1981) and Arminger and Muthén (1998) discuss different forms of prior distributions for factor analysis and covariance structure models. Appendix 1 describes the prior distributions over the parameters in our model.

Inference in the Bayesian framework is based on the joint posterior of all unknowns. Because this posterior density is very complex, we use simulation-based methods to obtain random draws from the posterior density (see Appendices 1 and 2 for details). Inference can then be based on the empirical distribution of the draws. The complexity of the posterior density precludes the use of direct methods for obtaining these draws. We therefore use Markov Chain Monte Carlo (MCMC) methods to obtain these draws. Specifically, our MCMC procedure involves Gibbs sampling

(Geman and Geman 1984, Gelfand and Smith 1990) and Metropolis-Hastings (Metropolis et al. 1953, Hastings 1970, Chib and Greenberg 1995) steps in tandem with data augmentation (Tanner and Wong 1987) to obtain the requisite draws. The MCMC methods require sampling parameter estimates from the full conditional distribution of each block of parameters. In the context of our model, we need to generate random draws for the blocks  $\{\alpha, \Lambda, \{\xi_{ij}\}, \{\eta_{ij}\}, \{\nu_i\}, \Phi_i, \rho, \mathbf{R}, \Delta, \Theta_i, \{a_k\}, \{b_k\}, \pi_i, \vartheta, \mathbf{Y}, \Psi\}$ . Each iteration of the MCMC procedure involves sequentially sampling from the full conditional distributions associated with each block of parameters. The MCMC procedure also provides samples of the factor scores  $f_{ij} = \{\xi_{ij}, \eta_{ij}\}$  and  $\nu_i$ , thus enabling posterior inference about these quantities. Appendix 2 describes the full conditional distributions and the simulation steps involved in each iteration of the MCMC procedure.

## 4. Synthetic Data Application

In this section we demonstrate our estimation procedures on artificial data. The aim is to illustrate the types of misleading inferences that are possible when models that ignore heterogeneity are applied to heterogeneous data. We generated a data set with 200 individuals and 30 observations per individual according to the model described in Equations (4) and (5). The data set includes four exogenous variables ( $x_1, \dots, x_4$ ) and four endogenous variables ( $y_1, \dots, y_4$ ). The association among the manifest variables is modeled in terms of two exogenous factors ( $\xi_1$  and  $\xi_2$ ) and two endogenous factors ( $\eta_1$  and  $\eta_2$ ). Figure 2 shows the model structure. We set

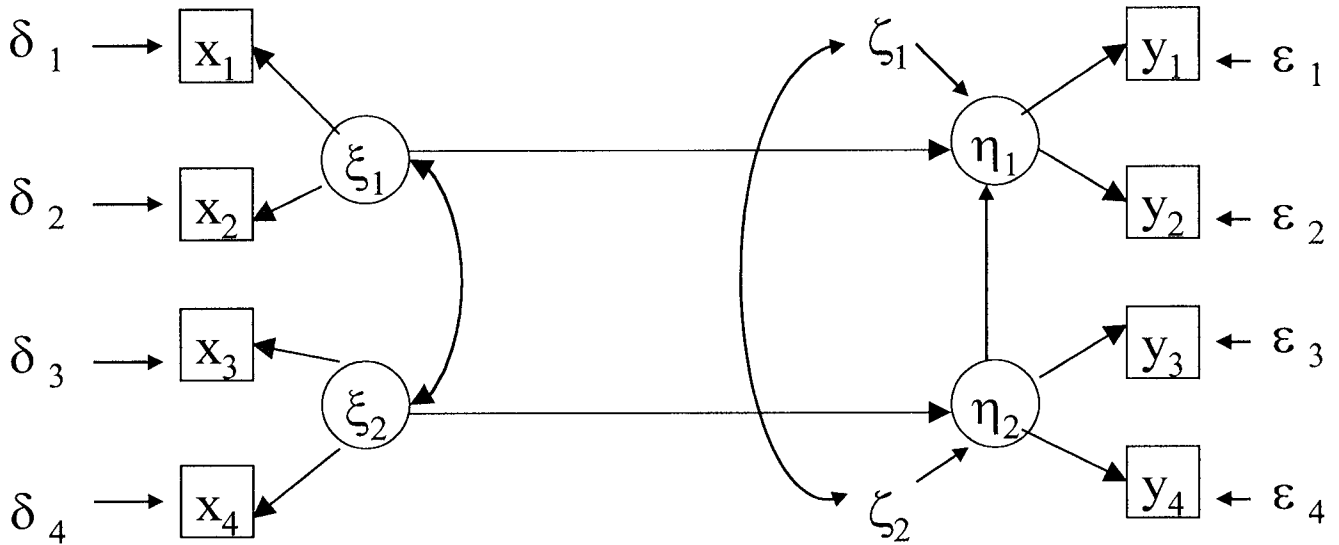
$$\Lambda_x' = \Lambda_y' = \begin{pmatrix} 1 & 0.8 & 0 & 0 \\ 0 & 0 & 0.8 & 1.0 \end{pmatrix}, \quad (10)$$

$$E[\Phi_i] = \begin{pmatrix} 1 & -0.3 \\ -0.3 & 1 \end{pmatrix}, \Delta = \begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{pmatrix}, \quad (11)$$

and  $E[\Theta_{i,x}] = E[\Theta_{i,y}] = \text{diag}(1.0)$  for the measurement model. For the structural model, we used a population distribution,  $\pi_i \sim N(\vartheta, \mathbf{Y})$  for the individual level structural coefficients in

$$\begin{aligned} \eta_1 &= b_1 \eta_2 + \gamma_{11} + \gamma_{13} \xi_1 + \xi_1, \\ \eta_2 &= \gamma_{12} + \gamma_{14} \xi_2 + \xi_2, \end{aligned} \quad (12)$$

Figure 2 Model Structure for Synthetic Data



where  $\vartheta = \{E[b_i] = -0.5, E[\gamma_{i1}] = 0, E[\gamma_{i2}] = 0, E[\gamma_{i3}] = -2, E[\gamma_{i4}] = 2\}$  and  $\mathbf{Y} = \text{diag}(0.1)$ . The structural errors  $\zeta$  come from a normal distribution  $N(\mathbf{0}, \Psi)$  where

$$\Psi = \begin{pmatrix} 0.4 & 0.2 \\ 0.2 & 0.4 \end{pmatrix}. \quad (13)$$

We used SAS and our MCMC procedures to estimate two models on the data. The first model (NH) is a nonhierarchical model that ignores the nature of the clustering of the observations and assumes that each individual has the same set of parameters. We estimated this model using Proc Calis in the SAS software. The second model is the true model represented by Equations (4) and (5). In estimating the second model we used priors that are similar to those outlined in Appendix 1. The estimates for the second model are based on 3,000 draws from the joint posterior distribution obtained after discarding 1,000 draws from the initial transient portion of the chain. Convergence was assessed using a variety of diagnostics detailed in the CODA package (Best et al. 1995) and by using time series plots to graphically assess the quality of the mixing of the chain.

Table 1 reports the within-factor covariance matrix  $\Phi$  for the first model and the population expectation of the  $\Phi_i$  matrices,  $E(\Phi_i)$  for the fully heterogeneous

model (FH). It also reports the across-individual covariance matrix  $\Delta$  of the mean factor scores  $\nu_i$  for Model FH. In §2.1, we showed that ignoring heterogeneity in factor means and measurement error leads to an aggregate factor covariance matrix  $\Phi^{\text{Agg}}$  that confounds the within-individual  $\Phi$  and between-individual  $\Delta$  factor covariance matrices. We also showed that this confound can lead to sign reversal of the factor covariances under some conditions. A comparison of the factor covariance estimates from the nonhierarchical model (NH) with the sum of the estimated within- and between-factor covariance matrices from the fully heterogeneous model (FH) clearly confirms this theoretical relationship. For example,  $\Phi_{11}$  from NH is 1.61. This value is approximately equal to the sum of  $\Phi_{11} = 1.0$  and  $\Delta_{11} = 0.58$  obtained from FH. Most importantly, we see that the covariance estimate  $\Phi_{12}$  from NH has the wrong sign because of this confounding. Finally, note that the estimates from the correct model (FH) are close to the true values as expected. Similarly, the estimates of  $\Delta$  obtained from Model (FH) are close to the true values.

Table 2 reports the structural model parameters. The top portion of the table reports the structural coefficients for the nonhierarchical model and the population expectation of the individual-level coefficients for



**Table 1** Factor Covariance Matrices

Parameter	True	NH	FH
$\Phi_{11}$	1.0	1.61 (1.51, 1.70)	1.0 (0.92, 1.08)
$\Phi_{12}$	-0.3	0.13 (0.07, 0.18)	-0.28 (-0.33, -0.23)
$\Phi_{22}$	1.0	1.60 (1.51, 1.70)	0.95 (0.87, 1.04)
$\Lambda_{11}$	0.6	—	0.58 (0.46, 0.72)
$\Lambda_{12}$	0.4	—	0.42 (0.32, 0.54)
$\Lambda_{22}$	0.6	—	0.63 (0.51, 0.78)

**Table 2** Structural Model Parameters

Population Regression Coefficients			
Parameter	Actual	NH	FH
$b$	-0.5	-0.55 (-0.57, -0.52)	-0.53 (-0.59, -0.47)
$\gamma_1$	0	0	0
$\gamma_2$	0	0	0
$\gamma_3$	-2	-1.96 (-2.02, -1.89)	-2.02 (-2.11, -1.92)
$\gamma_4$	2	1.99 (1.92, 2.05)	2.0 (1.91, 2.09)
Error Variances			
$\Psi_{11}$	0.4	1.62 (1.43, 1.81)	0.38 (0.28, 0.5)
$\Psi_{12}$	0.2	0.21 (0.10, 0.32)	0.21 (0.13, 0.26)
$\Psi_{22}$	0.4	0.62 (0.44, 0.78)	0.42 (0.32, 0.53)

the hierarchical model. Whereas the magnitudes of the coefficients are approximately the same for the two models, the confidence interval for  $b$  for Model NH does not cover the true value. A comparison of the error variances across the two models reveals that the

aggregate model provides inflated estimates of the error variances. This result is not surprising because the aggregate model (NH) forces the error terms to absorb the unobserved heterogeneity in the parameters. Jedidi et al. (1997) found the same results, based on an extensive simulation. Specifically, they found that the aggregate model captures the population means of the structural parameters very well but confounds the variance due to heterogeneity with the error variance.

Table 3 reports the estimated factor loadings and the measurement error variances for the two models. The table also shows the 95% posterior intervals associated with the parameters. It is clear from the table that both models yield almost identical estimates of the factor loadings. This is not surprising, as the data-generating mechanism assumed identical  $\Lambda$  for all individuals. Comparing the estimated measurement error variances  $\Theta$  for the NH model and the population expectation of  $\Theta$ , for the FH model, we see that the magnitude of the estimates are similar. A closer look at the variance estimates associated with  $x_1$  and  $x_2$  reveals that the confidence intervals for the first model do not

**Table 3** Measurement Model Parameters

Parameter	Factor Loadings			Measurement Error Variances		
	True	NH	FH	True	NH	FH $E(\Theta_i)$
$y_1$	1	1	1	1	1.0 (0.89, 1.10)	1.02 (0.91, 1.12)
$y_2$	0.8	0.79 (0.78, 0.81)	0.79 (0.78, 0.81)	1	1.0 (0.93, 1.07)	1.0 (0.92, 1.08)
$y_3$	0.8	0.81 (0.79, 0.82)	0.80 (0.79, 0.82)	1	1.02 (0.96, 1.08)	1.02 (0.93, 1.1)
$y_4$	1	1	1	1	0.97 (0.88, 1.05)	0.97 (0.88, 1.07)
$x_1$	1	1	1	1	0.94 (0.89, 0.99)	0.96 (0.88, 1.03)
$x_2$	0.8	0.80 (0.76, 0.82)	0.80 (0.77, 0.83)	1	0.93 (0.89, 0.97)	0.95 (0.89, 1.01)
$x_3$	0.8	0.79 (0.76, 0.82)	0.80 (0.77, 0.83)	1	1.05 (1.00, 1.09)	1.04 (0.98, 1.12)
$x_4$	1	1	1	1	1.03 (0.97, 1.08)	1.04 (0.97, 1.12)

Note: All loadings equal to 1 are fixed for identification.

cover the true value of 1.0. The posterior intervals for the full model (FH) are wider, as they properly reflect the uncertainty arising from heterogeneity and cover the true value for all parameters.

In summary, it is clear from the above example that ignoring heterogeneity can yield misleading inferences. In fact, we see that aggregate models may have wrong signs for covariance parameters, thus providing a distorted picture of the association structure of variables. In addition, aggregate models may report inflated estimates of factor variances, thus overstating the reliability of factors. Finally, models that pool data across individuals and thus ignore heterogeneity may underappreciate uncertainty in parameter estimates and provide a false sense of precision.

## 5. A Customer Satisfaction Application

Many researchers have analyzed the antecedents of customer satisfaction (see Oliver 1997 for a review). Typically, researchers have postulated that overall customer satisfaction with a product or service is affected by its perceived performance and by the extent to which a customer's expectations are met or disconfirmed (e.g., Oliver 1993 and Johnson et al. 1995).

We used the hierarchical Bayesian approach to estimate a structural equation model of satisfaction, using panel data from a study on campus dining services conducted at a large northeastern university. The data were collected using the following procedure. The population of interest consisted of students who had purchased all-inclusive meal plans for the fall semester. At the beginning of the semester during registration week, subjects from the population were recruited, using sign-up sheets circulated by experimenters in booths set up in dining outlets and dormitories at the university. Sixty individuals signed up to attend an information session; of these, 55 agreed to participate in the panel. Each subject who participated in the study was paid \$50 and was required to complete a daily diary for 39 consecutive days, using the following procedure. As soon as possible after dinner every day, subjects were required to record their degree of satisfaction with the dining service and their perception of the service provider's performance along

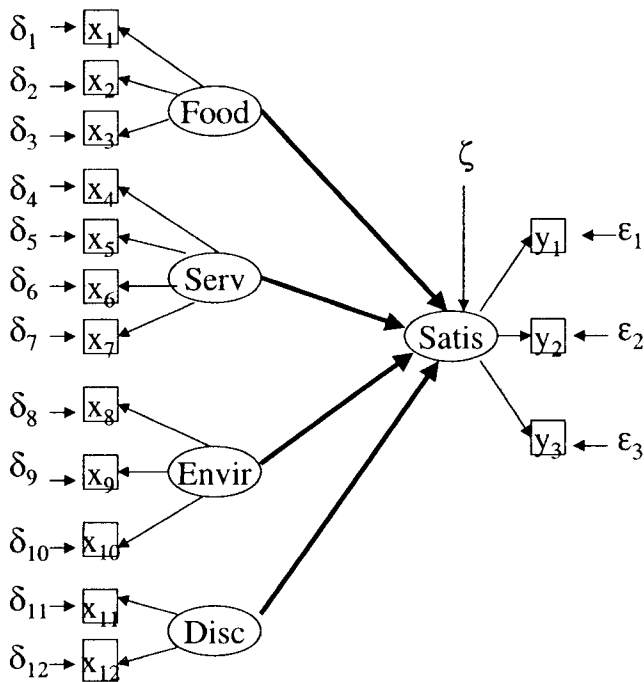
key dimensions, including food, service, and the dining environment. Subjects also recorded their expectancy disconfirmation, i.e., the degree to which their expectations of performance were met or disconfirmed. All the observable variables (items) were measured using seven-point scales. We excluded two subjects from the analysis because of limited data for those individuals. Because of missing data for some subjects, our final data set contains 1,542 observations from the remaining 53 subjects. We have therefore, an average of 29 observations per subject.

We specified a model in which customers' satisfaction ( $\eta$ ) with the dining service depends on perception of the dining service's performance on food ( $\xi_1$ ), service ( $\xi_2$ ), and the dining environment ( $\xi_3$ ). In addition, satisfaction depends on disconfirmation ( $\xi_4$ ). Satisfaction for each service episode was measured using the following three items: very dissatisfied to very satisfied ( $y_1$ ); felt terrible to delighted ( $y_2$ ); liked very little to liked very much ( $y_3$ ). Food performance was measured using the following three observable indicators: unpalatable to palatable ( $x_1$ ), bad taste to good taste ( $x_2$ ), and not nutritious to nutritious ( $x_3$ ). Service performance was measured on four items: indifferent to responsive ( $x_4$ ), unfriendly to friendly ( $x_5$ ), inefficient to efficient ( $x_6$ ), and uncaring to caring ( $x_7$ ). Dining environment was measured using three items: unpleasant to pleasant ( $x_8$ ), dirty to clean ( $x_9$ ), and stressful to relaxing ( $x_{10}$ ). Finally, disconfirmation was measured using two items: much better than expected to much worse than expected ( $x_{11}$  and  $x_{12}$ ). Figure 3 presents a graphical summary of the structure of the satisfaction model. Note that to avoid clutter, Figure 3 does not show the covariances among the factors.

As discussed in §2, the data can contain several forms of heterogeneity. For example, in the measurement model, the factor means  $\nu_i$ , the factor covariances,  $\Phi_i$ , and the measurement error variances  $\Theta_i$  can differ across individuals. In the structural model, the structural coefficients (i.e., importance weights of the antecedents of satisfaction)  $\gamma_i$  can differ across respondents.

To understand the nature of heterogeneity in the customer satisfaction process, we specified and estimated six models. These model specifications are described in the table below.

Figure 3 The Structural Model for Satisfaction



Model Specifications

Parameter	Description
Model NH	No Heterogeneity
Model ML	Heterogeneous Factor Means – Multilevel Model
Model HS	ML + Heterogeneous Structural Coefficients
Model HSHC	HS + Heterogeneous Factor Covariances
Model HSHV	HS + Heterogeneous Measurement Variances
Model FH	HSHV + Heterogeneous Factor Covariances

The base model, Model NH, is the conventional structural equation model that assumes that the data do not contain any measurement or structural heterogeneity. The other models add increasing levels of heterogeneity. Model FH, for instance, allows heterogeneity in factor means, factor covariances, measurement error variances, and structural coefficients and, therefore, is the most general model.

In estimating the models, we imposed appropriate restrictions on parameters to accommodate the indeterminacy in the scales and origins of the factors. For example, in all the models, for each factor we set the loading of one indicator variable to unity. In addition, following the usual practice in structural equation modeling, we set the mean of each factor to zero for

Model NH. In contrast, for the other models, we assume that the factor means  $\nu_i$  vary across individuals with a grand mean  $E(\nu_i) = 0$ . As discussed earlier, we also set the population mean of the structural intercept to zero, i.e.,  $E(\gamma_{i0}) = 0$ . Note that the measurement parameters in the models are identified because each factor has at least two indicators and the factors are allowed to covary (see Bollen 1989, pp. 238–251). The structural parameters are also identified, because the structural model is a multiple regression equation.

We used the MCMC procedure described in § 3 and in Appendix 2 to estimate the models. The MCMC algorithm was coded using the C language. Furthermore, we estimated Model NH, using Proc Calis in SAS to compare our Bayesian estimates with those obtained from the maximum likelihood estimator (MLE). Appendix 1 reports the prior distributions we used. The parameter estimates for all models are based on 20,000 draws obtained after discarding the first 10,000 iterations. Convergence was assessed using a variety of diagnostics detailed in the CODA package (Best et al. 1995) and by using time series plots to graphically assess the quality of the mixing of the chain.

## 5.1. Model Assessment

In this section we compare model performance using marginal likelihood values computed from the simulation output. We also compute the out-of-sample performance on a holdout data set to assess whether accounting for unobserved heterogeneity leads to improvements in predictive ability.

**5.1.1. Model Comparison.** We computed the log-marginal likelihood using the importance sampling method illustrated in Newton and Raftery (1994) to compare the six models. A model with a higher marginal likelihood is the preferred model. The differences in the marginal likelihood values from those of a base model (Model NH) can be used to assess model improvements. The differences in log-marginal likelihood for the models are (a) Model ML: 103.51, (b) Model HS: 146.6, (c) Model HSHC: 248.69, (d) Model HSHV: 1603.15, and (e) Model FH: 1741.2. These values clearly suggest that accounting for unobserved heterogeneity is important in this application. Among all the models that account for some form of heterogeneity,

the full model (Model FH) shows the greatest improvement in log-marginal likelihood and, hence, is the most preferred model. As the models incorporate increasing levels of heterogeneity, the magnitudes of the improvements can be used to assess the importance of particular types of heterogeneity. A comparison across the different heterogeneity specifications reveals that adding covariance heterogeneity (Model HSHC and Model HSHV) results in considerable improvement in marginal likelihoods. Of the two forms of covariance heterogeneity that are possible within the measurement model, adding heterogeneity in measurement variances (Model HSHV) results in much greater model improvement for our data. Improvements arising from heterogeneous factor covariances (Model HSHC) are modest in comparison. These results clearly show that multilevel models do not fully capture the heterogeneity in our data. Specifically, multilevel models allow heterogeneity only in mean structures. Our model selection results, however, suggest that individuals vary both in means and covariances.

**5.1.2. Predictive Ability.** To further assess the benefits of accounting for unobserved heterogeneity, we compared the predictive ability of the conventional structural equation model (Model NH) with that of the full model (Model FH). To compute the out-of-sample performance, we created a holdout data set consisting of 427 observations from our original calibration sample. This data set was created by randomly selecting approximately 8 observations from each of the 53 subjects in the sample. We estimated the NH and FH models, using the remaining 1,115 observations and used these estimates to compute measures of predictive performance from the holdout data set. We assessed how well the models perform in recovering the mean and covariance structure of the data set. Table 4 reports the Mean Absolute Deviations (MAD) in predicting the means of the exogenous and endogenous manifest variables. It is clear from the table that the model that incorporates heterogeneity outperforms the aggregate model in predicting the means across all of the manifest variables.

As structural equation models capture the covariance structure of the variables, we also investigated

**Table 4** Predictive Ability: MAD for Variable Means

Factor	Indicator	Model NH	Model FH
Food	$X_1$	1.134	0.955
	$X_2$	1.207	1.095
	$X_3$	1.087	0.986
Service	$X_4$	0.959	0.912
	$X_5$	0.968	0.910
	$X_6$	1.033	0.953
	$X_7$	1.063	0.952
Envir.	$X_8$	1.096	1.018
	$X_9$	1.127	0.964
	$X_{10}$	1.121	0.968
Disc.	$X_{11}$	1.193	1.114
	$X_{12}$	1.317	1.248
Satis.	$Y_1$	1.176	1.097
	$Y_2$	1.147	1.044
	$Y_3$	1.146	1.036

how well the two models did in recovering the within-individual covariance matrix  $\Sigma_W$ , the between-individual  $\Sigma_B$  covariance matrix, and the total  $\Sigma_T$  covariance matrix of the data. Table 5 reports two measures of fit (MAD and RMR) for these covariance matrices. The MAD represents the mean absolute deviation between the elements of the actual and the predicted matrices. The RMR (Bollen 1989) represents the square root of the mean squared deviation between the elements of actual and predicted matrices. In computing these measures, only the nonredundant elements of the covariance matrices are used. Table 5 shows that the hierarchical model clearly outperforms the simple model in recovering both the within- and between-covariance matrices. It is also interesting to note that the two models do not differ in their ability to recover the total covariance matrix. This result is consistent with the insight that aggregate models do well in recovering the aggregate aspects of a data set but do

**Table 5** Predictive Ability: Covariance Matrices

Measure	Model NH			Model FH		
	$\Sigma_W$	$\Sigma_B$	$\Sigma_T$	$\Sigma_W$	$\Sigma_B$	$\Sigma_T$
RMR	0.379	0.367	0.138	0.153	0.135	0.138
MAD	0.361	0.356	0.107	0.135	0.118	0.107

poorly in representing the association structure at the microlevel. This result is also consistent with the simulation results of Jedidi et al. (1997), who found that fit measures that are based on the aggregate covariance matrix (e.g., GFI and RMR) are not useful for detecting heterogeneity.

In summary, the marginal likelihood values and the predictive comparisons unambiguously suggest the importance of accounting for both mean and covariance heterogeneity in the satisfaction data.

## 5.2. Parameter Estimates

We now examine the parameter estimates from our models. These estimates are based on the entire calibration sample consisting of 1,542 observations from 53 individuals. To avoid clutter, we report the estimates for only three models. These models include the simple nonhierarchical model (Model NH), the multi-level model (Model ML), and the full model (Model FH). For the NH model, we report both the Bayesian and the maximum likelihood (MLE) estimates that we obtained using SAS Proc Calis. We first discuss the results from the measurement model and then analyze the structural model estimates. Table 6 shows the factor loadings from the measurement model. The MLE factor loadings for Model NH, reported in Column 3, are virtually identical to those obtained from the Bayesian analysis of the model (see Column 4). The similarity of the estimates is expected, because we used diffuse priors and our sample size is large (1,542 observations). It is also clear from Table 6 that the loading estimates differ only slightly across all three models. This finding is not surprising, because we assumed common factor loadings for all individuals. Overall, all factor loadings are significant, suggesting that the indicators are reliable measures for the underlying factors.

Table 7 reports the measurement error variances. First, as in Table 6, the maximum likelihood estimates (MLE) of Model NH are almost identical to those obtained from the Bayesian analysis. Second, the estimates from Model NH and Model ML are essentially similar, as both of these models assume identical measurement error variances across individuals. The population mean estimates from Model FH, however,

**Table 6** Measurement Model Results: Factor Loadings

Factor	Indicator	Model NH (MLE)	Model NH (Bayes)	Model ML	Model FH
Food	$x_1$	1	1	1	1
	$x_2$	1.106 (0.023)	1.109 (0.023)	1.067 (0.023)	1.053 (0.018)
	$x_3$	0.717 (0.026)	0.719 (0.026)	0.723 (0.026)	0.711 (0.024)
	$x_4$	1	1	1	1
Service	$x_5$	1.028 (0.025)	1.028 (0.024)	1.031 (0.025)	1.021 (0.019)
	$x_6$	0.994 (0.028)	0.996 (0.029)	1.009 (0.030)	0.953 (0.025)
	$x_7$	1.015 (0.027)	1.016 (0.028)	1.027 (0.028)	0.995 (0.023)
	$x_8$	1	1	1	1
Envir.	$x_9$	1.220 (0.029)	1.216 (0.030)	1.234 (0.029)	1.120 (0.027)
	$x_{10}$	1.016 (0.029)	1.018 (0.029)	1.027 (0.029)	1.040 (0.027)
	$x_{11}$	1	1	1	1
Disc.	$x_{12}$	0.984 (0.024)	0.984 (0.024)	0.987 (0.024)	0.982 (0.023)
Satis.	$y_1$	1	1	1	1
	$y_2$	1.038 (0.019)	1.040 (0.019)	1.042 (0.019)	1.020 (0.016)
	$y_3$	1.080 (0.019)	1.082 (0.019)	1.083 (0.019)	1.046 (0.016)
	$y_4$	1.080 (0.019)	1.082 (0.019)	1.083 (0.019)	1.046 (0.016)

Note: Posterior standard deviations are shown in parentheses.

show considerable differences from the estimates obtained for Models NH and ML. In addition, it is important to note that the standard deviation of the population means are much higher for Model FH, as should be expected in a model that incorporates heterogeneity. Thus, conventional models can provide distortions in inference when applied to heterogeneous data. Finally, the last column of Table 7 reports the across-individual variation,  $Std(\theta_i)$ , in the measurement error variances.<sup>1</sup> Clearly, there is considerable heterogeneity in the measurement error variances in our sample. That is, subjects respond to questions with different degrees of accuracy.

<sup>1</sup>As  $\theta_{ik} \sim IG(a, b_k)$ ,  $E(\theta_{ik}) = 1/(b_k(a - 1))$  and  $Std(\theta_{ik}) = \sqrt{1/(b_k^2(a - 1)^2(a - 2))}$ .

**Table 7 Measurement Model Results: Measurement Error Variances**

Factor	Indicator	Model NH (MLE)	Model NH (Bayes)	Model ML	Model FH	
					$E(\theta_i)$	$Std(\theta_i)$
Food	$x_1$	0.458 (0.021)	0.460 (0.025)	0.403 (0.026)	0.332 (0.037)	0.293 (0.047)
	$x_2$	0.345 (0.027)	0.348 (0.026)	0.420 (0.029)	0.385 (0.042)	0.339 (0.054)
	$x_3$	1.217 (0.046)	1.213 (0.045)	1.173 (0.047)	1.186 (0.110)	1.048 (0.166)
Service	$x_4$	0.469 (0.023)	0.471 (0.025)	0.478 (0.025)	0.424 (0.044)	0.374 (0.059)
	$x_5$	0.428 (0.023)	0.429 (0.024)	0.437 (0.023)	0.351 (0.039)	0.310 (0.049)
	$x_6$	0.697 (0.030)	0.697 (0.032)	0.681 (0.032)	0.751 (0.074)	0.663 (0.108)
	$x_7$	0.609 (0.028)	0.609 (0.030)	0.597 (0.027)	0.629 (0.063)	0.556 (0.089)
Envir.	$x_8$	0.703 (0.032)	0.696 (0.032)	0.720 (0.032)	0.727 (0.071)	0.643 (0.106)
	$x_9$	0.217 (0.028)	0.235 (0.029)	0.203 (0.027)	0.232 (0.032)	0.204 (0.037)
	$x_{10}$	0.838 (0.036)	0.836 (0.034)	0.828 (0.037)	0.882 (0.084)	0.779 (0.123)
Disc.	$x_{11}$	0.568 (0.034)	0.570 (0.036)	0.594 (0.034)	0.748 (0.08)	0.662 (0.113)
	$x_{12}$	0.561 (0.033)	0.562 (0.034)	0.584 (0.032)	0.761 (0.078)	0.672 (0.110)
Satis.	$y_1$	0.487 (0.021)	0.486 (0.021)	0.487 (0.021)	0.629 (0.081)	2.19 (3.7)
	$y_2$	0.294 (0.016)	0.294 (0.016)	0.293 (0.015)	0.267 (0.035)	0.92 (1.315)
	$y_3$	0.258 (0.015)	0.256 (0.015)	0.258 (0.015)	0.275 (0.036)	0.951 (1.568)

*Notes:* Posterior standard deviations are shown in parentheses.  $E(\theta_i)$  and  $Std(\theta_i)$  denote the population and standard deviation of the measurement variances.

Table 8 reports the estimated covariance matrix  $\Phi$  of the exogenous factors,  $\xi_{ij}$ , for Models NH and ML and the population expectation of the  $\Phi_i$  matrices,  $E(\Phi_i)$ , for Model FH. For Models ML and FH, the table also reports the covariance matrix  $\Delta$  of mean factor scores  $\nu_i$  across individuals. As in the previous tables, the MLE estimates are very similar to those obtained from

the Bayesian analysis of Model NH. Recall that the multilevel model (Model ML) only captures heterogeneity in the factor means,  $\nu_i$ . As we saw in the previous synthetic data application, ignoring heterogeneity in means results in a confounding of the within- and between-covariance matrices of the factors, i.e.,  $\Phi_{ModelNH} = \Phi_{ModelML} + \Delta_{ModelML}$ . The results for Model NH and Model ML confirm this theoretical relationship. For example, the first row of Table 8 reveals that  $\Phi_{Food-Food}$  for Model NH is 1.569. This value is approximately equal to the sum of the estimates  $\Phi_{Food-Food} = 1.165$  and  $\Delta_{Food-Food} = 0.426$  obtained from Model ML. Thus, if the factor means are heterogeneous (Model ML), the results from a conventional model that ignores heterogeneity will always underestimate the factor variances and therefore inflate the measurement reliability.

Recall that Model FH allows for heterogeneous  $\Phi_i$  matrices across individuals by assuming that  $\Phi_i^{-1} \sim W(\rho, R)$ . Column 5 of Table 8 reports the population means of the  $\Phi_i$ 's. The last column of the table reports the estimated  $\Delta$ , which captures the heterogeneity in mean factor scores ( $\nu_i$ ) across individuals. When Model NH is used to analyze data originating from a process that involves general forms of measurement and structural heterogeneity (e.g., as in Model FH), the estimation bias manifests itself in a complex manner across different sets of parameters. The results show that the parameter estimates are indeed different across the three models.

Focusing on the estimates for  $\Phi$  for Model FH, we see that all the exogenous factors are positively correlated. The estimates of  $\Delta$  show that the mean factor scores are also positively correlated across individuals. In addition, the large magnitudes of the diagonal elements of  $\Delta$  show that there is considerable heterogeneity in factor means across individuals.

Table 9 reports the regression coefficients from the structural model. In interpreting the table, recall that Model FH captures heterogeneity in the impact of the antecedent constructs on satisfaction by assuming that the individual-specific coefficients come from a multivariate normal population distribution  $N(\vartheta, Y)$ . The population mean,  $\vartheta$ , for the structural coefficients and the standard deviation of the individual specific coefficients  $\sqrt{Y_{kk}}$  are reported in Columns 5 and 6, respec-

Table 8 Estimated Covariance Structure of Antecedent Factors

Parameter	$\Phi$				$\Delta$	
	Model NH (MLE)	Model NH (Bayes)	Model ML	Model FH	Model ML	Model FH
Food-Food	1.569 (0.074)	1.562 (0.074)	1.165 (0.055)	1.388 (0.155)	0.426 (0.093)	0.433 (0.095)
Food-Service	0.617 (0.043)	0.618 (0.043)	0.399 (0.032)	0.416 (0.083)	0.216 (0.063)	0.224 (0.067)
Food-Envir.	0.764 (0.048)	0.765 (0.047)	0.414 (0.032)	0.457 (0.079)	0.331 (0.080)	0.343 (0.085)
Food-Disc.	1.232 (0.061)	1.230 (0.061)	0.997 (0.050)	1.296 (0.159)	0.229 (0.062)	0.230 (0.063)
Service-Service	1.170 (0.059)	1.170 (0.060)	0.881 (0.046)	1.131 (0.131)	0.285 (0.064)	0.305 (0.070)
Service-Envir.	0.691 (0.042)	0.694 (0.044)	0.409 (0.029)	0.417 (0.073)	0.266 (0.066)	0.289 (0.072)
Service-Disc.	0.602 (0.046)	0.603 (0.046)	0.428 (0.037)	0.457 (0.099)	0.161 (0.049)	0.173 (0.051)
Envir.-Envir.	1.297 (0.070)	1.297 (0.070)	0.852 (0.046)	0.963 (0.113)	0.426 (0.093)	0.453 (0.100)
Envir.-Disc.	0.704 (0.050)	0.707 (0.050)	0.462 (0.038)	0.596 (0.101)	0.218 (0.059)	0.234 (0.063)
Disc.-Disc.	1.801 (0.087)	1.799 (0.087)	1.534 (0.079)	2.143 (0.252)	0.231 (0.056)	0.222 (0.054)

Notes: Posterior standard deviations are in parentheses.  $\Phi$  = Unduplicated elements in covariance matrix of exogenous factors.  $\Delta$  = Unduplicated elements in covariance matrix of mean factor scores for exogenous factors.

tively, of the table. The mean estimates for Model FH show that satisfaction is significantly affected by perceived performance on food, service, and the environment. In addition, the positive coefficient for disconfirmation confirms the previous findings in the literature that better-than-expected performance increases satisfaction. The magnitudes of the across-individual standard deviations (see the last column), are large and confirm that the importance weights of the antecedent dimensions on satisfaction vary significantly across subjects.

The differences in the magnitudes of the coefficients across the models show that ignoring heterogeneity can yield misleading inferences about the structural parameters. Furthermore, it appears that Model NH

Table 9 Structural Model: Regression Coefficients

Exogenous Factor	Model NH (MLE)	Model NH (Bayes)	Model ML	Model FH	
				$\vartheta$	$\sqrt{Y_{kk}}$
Food	0.311 (0.027)	0.311 (0.028)	0.258 (0.028)	0.366 (0.052)	0.236 (0.031)
Service	0.108 (0.022)	0.108 (0.023)	0.110 (0.022)	0.079 (0.039)	0.260 (0.040)
Envir.	0.058 (0.022)	0.057 (0.023)	0.053 (0.021)	0.094 (0.041)	0.206 (0.028)
Disc.	0.536 (0.027)	0.536 (0.027)	0.588 (0.028)	0.503 (0.048)	0.246 (0.036)
$\Psi$	0.285 (0.019)	0.285 (0.019)	0.266 (0.019)	0.168 (0.015)	

Notes: Posterior standard deviations are shown in parentheses.  $\vartheta$  denotes the population mean of the structural coefficients.  $\sqrt{Y_{kk}}$  denotes the population standard deviation of the structural coefficients.

and Model ML seriously overstate statistical significance, because they understate the posterior standard deviations of all structural parameter estimates. In addition, the last row of Table 9 shows that by failing to allow for structural heterogeneity, Models NH and ML understate the goodness-of-fit of the structural model. This result is not surprising, because the unaccounted heterogeneity is absorbed by the structural error term.

To illustrate how the model can be used for managerial purposes, we performed a "quadrant analysis" (see Figure 4). For each antecedent of satisfaction, we plotted for each individual the mean factor scores (which are automatically produced by the Bayesian approach) and structural parameter estimates; in addition, we partitioned the appropriate space into four quadrants, based on the corresponding mean factor scores and mean importance weights in the sample. (Recall that the grand mean for each antecedent factor was set to zero.) Thus, in Figure 4, individuals corresponding to points in the upper-right quadrant have above-average scores on the food factor; in addition, for these individuals, food has an above-average effect on overall satisfaction.

Using this partitioning into quadrants, managers selectively target consumers and develop marketing pol-

**Figure 4**      **Quadrant Analysis: Perceived Performance (Factor Scores) vs. Importance**



targeting. The extent of uncertainty associated with an individual's estimates depends intricately on the number of observations available from that individual and also on the extent of shrinkage induced by the choice of priors. In fact, Bayesian methods allow a proper assessment of uncertainty. For example, the figure shows the 95% confidence ellipsoids associated with Individual 9. The figure shows that there is considerable uncertainty about the individual-level parameters. This is not surprising, as a highly parametrized model is being fitted to relatively few observations from a person. The confidence ellipsoids also show that the uncertainty is higher for the importance weights when compared to the uncertainty in the perceptual scores. This is possible because the importance weights link two sets of latent variables (i.e., factor scores) and are therefore farther removed from the observed data. Such assessments need to be used for sensitivity analysis in decision making.

## 6. Summary and Conclusions

This paper develops and tests a hierarchical Bayesian framework for handling general forms of unobserved heterogeneity in structural equation models. Our methodology is appropriate when the heterogeneity in the population can be measured on a continuum and multiple observations are available for each individual. An important feature of our method is that it automatically provides individual-specific estimates of model parameters and factor scores.

We tested the Bayesian methodology, using both synthetic data and data from a satisfaction study. The results from the synthetic data application show that the nonheterogeneous (aggregate) model confounds random variability (i.e., measurement and structural errors) with variability because of heterogeneity. In the measurement model, this confounding led to inflated indicator reliabilities and resulted in sign reversal of the factor covariances. In the structural model, the confounding led to lower model fit. In addition, the nonheterogeneous model understated the standard errors of the measurement and structural parameter estimates.

The results from the satisfaction study show that the structural parameters vary significantly in the population (i.e., it is incorrect to analyze the pooled data set

using a conventional structural equation model). As in the case of the synthetic data example, the estimates obtained from a conventional structural equation model (i.e., a model that does not allow for unobservable heterogeneity) were misleading, even though the model was comparable to the other models on prediction of aggregate quantities such as the total covariance matrix  $\Sigma_T$ . In particular, the nonhierarchical model understated the goodness-of-fit for the structural equation and provided biased estimates of the structural parameters. Perhaps most importantly, the nonhierarchical model seriously understated the standard errors of the structural parameters; that is, models that ignore heterogeneity overstate statistical significance. Finally, we showed that the traditional psychometric methods (e.g., multilevel models) do not fully capture the heterogeneity in our data. Thus, accounting for heterogeneity in both mean and covariance structures is important for obtaining proper inferences.

From a managerial viewpoint, we showed how one can use the individual-level factor scores and structural parameter estimates from the Bayesian approach to perform quadrant analysis and refine marketing policy (e.g., develop a one-on-one marketing policy). From a substantive viewpoint, our Bayesian procedure should be applied to a broad range of marketing problems in which the data contain unobservable heterogeneity and structural equation modeling is appropriate (e.g., instances involving measurement error and latent constructs). From a methodological viewpoint, we designed our MCMC procedures to deal with the partially recursive structural model and assumed interval-scaled data. Future research should extend our procedures to deal with the general nonrecursive structural equation model and to accommodate ordinal data (Rossi et al. 1999) and binary data situations (Ansari and Jedidi 2000).<sup>2</sup>

## Appendix 1: Prior Distributions

In this paper, we specify the prior distribution over  $\phi$  as a product of independent priors. We use proper but diffuse priors over all model parameters. The priors for the measurement intercepts  $\alpha$  are

<sup>2</sup>Please address all correspondence to Asim Ansari, 517 Uris Hall, Columbia University, 3022 Broadway, NY, NY 10027. The authors thank Laurette Dube for kindly providing the data for the empirical study and Don Lehmann for his constructive comments.

assumed to be multivariate normal  $N(\kappa, A)$ . The covariance matrix  $A$  can be specified to be diagonal with the elements (variances) set to large values to represent vague knowledge. The exact location of this distribution is no longer critical once, a large variance has been assumed; therefore, without a loss of generality  $\kappa$  can be set to zero. For the application we used  $N(0, 100I)$ , where  $I$  is an identity matrix with appropriate dimensionality.

The combined factor loading matrix  $\Lambda$  has a patterned structure owing to the identifiability restrictions that require setting some of the elements of  $\Lambda$  to zero or to one. We therefore specified independent multivariate normal priors over the free elements within each row of the matrix. We have for row  $k$  a prior  $N(g_k, H_k)$ . The covariance matrix  $H_k$  is specified to be diagonal with large variances to ensure a diffuse prior. Thus, the prior over the loading matrix is the product of the independent priors associated with the rows of  $\Lambda$ . In the application, we specified the prior distribution for each row to be  $N(0, 100I)$ .

The  $m \times m$  matrix  $R^{-1}$  associated with the Wishart population distribution,  $W(p, R)$ , is assumed to come from conjugate Wishart prior  $W(p_R, (\rho_R S)^{-1})$ . Smaller values of  $\rho_R$  correspond to more vague prior distributions. For the application, we used a prior of  $W(5, 5I)$ . We assumed a univariate truncated normal  $N(0, \tau_p)$  prior over  $\log(p)$ . For the population Wishart distribution  $W(p, R)$  to be proper, the truncation must be such that  $p$  is greater than  $p$ . The prior variance,  $\tau_p$ , can be set to a large value to ensure a diffuse prior. For the application we set  $\tau_p = 100$ . Alternatively, a truncated gamma prior can be used, and its parameters can be chosen appropriately to ensure minimal influence.

The precision matrix  $\Lambda^{-1}$  associated with the population distribution  $\nu_i \sim N(0, \Lambda)$  is a  $p \times p$  positive definite matrix. In keeping with standard Bayesian analysis of linear models, we assumed a Wishart prior  $W(\delta, (\delta\Omega)^{-1})$ , where  $\Omega^{-1}$  can be considered the expected prior precision of the  $\nu$ 's. Smaller  $\delta$  values correspond to vaguer prior distributions. For the application we set  $\delta = 5$  and  $\Omega = I$  to ensure a proper prior.

There are  $r = (p + q)$  different population distributions  $IG(a_k, b_k)$ ,  $k = 1$  to  $r$ , for the measurement error variances contained in  $\Theta_i$ . We therefore need to specify priors over the set of unknowns  $\{a_k\}, \{b_k\}$ . We chose independent conjugate inverse gamma priors,  $b_k \sim IG(g_k, h_k)$ , for  $k = 1$  to  $r$ , to quantify prior uncertainty about the scale parameters. In the application we assumed  $a_k = a$  and  $g_k = 2$  and  $b_k = 2$  for all  $k$ . Finally, we assumed independent univariate normal  $N(0, \tau_k)$  priors over  $\log(a_k)$  for  $k = 1$  to  $r$ . The prior variance,  $\tau_k$ , must be set to a large value to ensure a diffuse prior. In the application we set  $\tau_k = 100$ .

The prior for the vector  $\vartheta$  is specified to be multivariate normal  $N(c, C)$ . The covariance matrix  $C$  can be assumed to be diagonal and its entries chosen to represent vague knowledge. The exact location vector  $c$  is no longer critical once a large variance has been assumed; hence, without loss of generality we set  $c$  to zero. In the application we therefore set  $c = 0$  and  $C = 100I$ .

We assumed a Wishart prior  $W(\rho_{\psi}, (\rho_{\psi} R_{\psi})^{-1})$  over the precision matrix  $\Psi^{-1}$  associated with the structural parameters  $\pi_i$ . Smaller  $\rho_{\psi}$  values correspond to vaguer prior distributions. We therefore set  $\rho_{\psi}$

$= 5$  and  $R_{\psi} = 0.1I$ . Finally, we assumed a Wishart prior  $W(\rho_{\psi}, (\rho_{\psi} R_{\psi})^{-1})$  over the precision matrix  $\Psi^{-1}$  of the structural model. If the structural model consists of a single equation, then an inverse gamma prior over the single structural error variance parameter can be assumed instead. In the application we chose  $IG(3, 1000)$  for the prior over the univariate  $\psi$ .

## Appendix 2: Full Conditional Distributions

The  $(m + 1)$ th iteration of the MCMC algorithm requires us to generate random draws from the following full conditional distributions:

(a) The measurement intercepts  $\alpha$  are generated from the multivariate normal full conditional distribution given by

$$p(\alpha | \Lambda, \{f_{ij}\}, \{\Theta_i\}, \{w_{ij}\}) = N(\hat{\alpha}, V_{\alpha}) \quad (A.1)$$

where  $V_{\alpha}^{-1} = A^{-1} + \sum_{i=1}^I N_i \Theta_i^{-1}$  and  $\hat{\alpha} = V_{\alpha} [A^{-1} \kappa + \sum_{i=1}^I \Theta_i^{-1} \sum_{j=1}^{N_i} (w_{ij} - \Lambda f_{ij})]$ .

(b) The loading matrix  $\Lambda$  is a patterned matrix containing both fixed and free elements. Of the fixed elements, some are fixed to zero, whereas others are fixed at one to impose identifiability constraints. The full conditional distribution for the free elements in a row of the matrix  $\Lambda$  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  $\lambda_k$  be the vector of free elements in row  $k$ . The prior for  $\lambda_k$  is given by  $p(\lambda_k) = N(g_k, H_k)$ . Let  $\tilde{f}_{ijk}$  be the vector of factor scores corresponding to the elements in row  $k$  of  $\Lambda$  that are set to one and let  $f_{ijk}$  contain the remaining factor scores from  $f_{ij}$ . Form the adjusted variable  $\tilde{w}_{ijk} = w_{ijk} - \mathbf{1}' \tilde{f}_{ijk} - \alpha_k$ , where  $\mathbf{1}$  is a vector of ones. Given the prior, the vector  $\lambda_k$  can be sampled from the full conditional distribution given by

$$p(\lambda_k | \{\tilde{w}_{ijk}\}, \{f_{-ijk}\}, \{b_k\}) = N \left( D_k \left[ \sum_{i=1}^I \sum_{j=1}^{N_i} \Theta_{i,k}^{-1} f_{-ijk} \tilde{w}_{ijk} + H_k^{-1} g_k \right], D_k \right) \quad (A.2)$$

where  $D_k^{-1} = \sum_{i=1}^I \sum_{j=1}^{N_i} \Theta_{i,k}^{-1} f_{-ijk} f_{-ijk}' + H_k^{-1}$ .

(c) The full conditional distribution for the factor scores  $\xi_{ij}$  is a multivariate normal distribution. The mean and variance of this distribution can be obtained by considering the prior  $\xi_{ij} \sim N(\nu_{ij}, \Phi_i)$  and the two data sources for  $\xi_{ij}$ . The first "data" we consider are  $\eta_{ij}$ . Consider the structural equation Model,  $B_i \eta_{ij} = \gamma_{i0} + \Gamma_i \xi_{ij} + \zeta_{ij}$ . Define  $\tilde{\eta}_{ij} = B_i \eta_{ij} - \gamma_{i0}$ . An intermediate posterior distribution for  $\xi_{ij}$  can be written as  $N(\tilde{\xi}_{1,ij}, V_{1,ij})$ , where  $V_{1,ij}^{-1} = \Phi_i^{-1} + \Gamma_i' \Psi^{-1} \Gamma_i$  and  $\tilde{\xi}_{1,ij} = V_{1,ij} [\Phi_i^{-1} \nu_i + \Gamma_i' \Psi^{-1} \tilde{\eta}_{ij}]$ . This intermediate posterior acts as a prior for the other data source  $x_{ij}$ . Taking into account the measurement equation,  $x_{ij} = \alpha_x + \Lambda_x \xi_{ij} + \delta_{ij}$ , the full conditional for  $\xi_{ij}$  can be written as  $N(\tilde{\xi}_{2,ij}, V_{2,ij})$ , where  $V_{2,ij}^{-1} = V_{1,ij}^{-1} + \Lambda_x' \Theta_{ix}^{-1} \Lambda_x$  and  $\tilde{\xi}_{2,ij} = V_{2,ij} [V_{1,ij}^{-1} \tilde{\xi}_{1,ij} + \Lambda_x' \Theta_{ix}^{-1} (x_{ij} - \alpha_x)]$ .

(d) The full conditional distribution for the individual-level factor means  $\nu_i$  is a multivariate normal distribution that can be written as

$$p(\nu_i | \{\xi_{ij}\}, \Phi_i, \Lambda) = N(\hat{\nu}_i, V_{\nu_i}) \quad (A.3)$$

where  $V_{\nu_i}^{-1} = \Lambda^{-1} + N_i \Phi_i^{-1}$  and  $\hat{\nu}_i = V_{\nu_i} [\Lambda^{-1} \sum_{j=1}^{N_i} \xi_{ij}]$ .

(c) The full conditional distribution for the precision matrix  $\Phi_i^{-1}$  of the individual-level factor scores is Wishart and is given by

$$p(\Phi_i^{-1} | \{\xi_{ij}\}, \{v_i\}) = W(\rho_{\text{pos}}, R_{\text{pos}}) \quad (\text{A.4})$$

where  $\rho_{\text{pos}} = \rho + N_i$  and  $R_{\text{pos}} = [\sum_{j=1}^{N_i} (\xi_{ij} - v_i)(\xi_{ij} - v_i)' + R^{-1}]^{-1}$ .

(f) The full conditional for the precision matrix  $\Delta^{-1}$  of the factor means is a Wishart distribution. Given the prior  $W(\delta, (\delta\Omega)^{-1})$ , the full conditional can be written as

$$p(\Delta^{-1} | \{v_i\}) = W\left(\delta + I, \left[\sum_{i=1}^I v_i v_i' + \delta\Omega\right]^{-1}\right). \quad (\text{A.5})$$

(g) The full conditional distributions for the diagonal elements of the measurement error variances  $\Theta_i$  for each individual, i.e.,  $\theta_{i,k}$ ,  $k = 1$  to  $r$ , are independent inverse gamma distributions. These distributions follow from standard Bayesian theory and can be written as

$$p(\theta_{i,k} | \lambda_k, \alpha_k, \{f_{ij}\}) = IG\left(\frac{N_i}{2} + a_k, \left[\frac{\sum_{j=1}^{N_i} (w_{ijk} - \alpha_k - \lambda_k' f_{ij})^2}{2} + b_k^{-1}\right]^{-1}\right) \quad (\text{A.6})$$

where  $\lambda_k$  is a vector containing the elements of row  $k$  of  $\Lambda$ .

(h) 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. Given the conjugate inverse gamma prior,  $IG(g_k, h_k)$ , the full conditional can be written as

$$p(b_k | \{\theta_{i,k}\}, a_k) = IG\left(la_k + g_k, \left[h_k^{-1} + \sum_{i=1}^I \theta_{i,k}^{-1}\right]^{-1}\right) \quad (\text{A.7})$$

Parameter draws can be sequentially made from the full conditional distributions of each  $b_k$ ,  $k = 1$  to  $r$ .

(i) The full conditional for the hyperparameter  $la_k = \log(a_k)$  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

$$L(\{\theta_{i,k}\} | la_k, b_k) = \prod_{i=1}^I \frac{(\theta_{i,k}^{-1})^{\exp(la_k)+1} \exp(-b_k^{-1} \theta_{i,k}^{-1})}{\Gamma(\exp(la_k)) b_k^{\exp(la_k)}} \quad (\text{A.8})$$

The prior density of  $la_k$  is univariate normal  $p(la_k) = N(0, \tau_k)$ . 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  $la_k$ . To generate a candidate  $la_k^c$ , we use a univariate normal proposal density,  $N(la_k^{(m)}, t_a)$ , that is centered on the old value of  $la_k^{(m)}$  from iteration  $m$ . The tuning constant  $t_a$  in the proposal density is chosen to allow rapid mixing and to avoid excessively frequent rejections of the candidates. The generated candidate  $la_k^c$  is accepted with the following acceptance probability

$$\min\left\{\frac{L(la_k^c | \{\theta_{i,k}\}, b_k) p(la_k^c)}{L(la_k^{(m)} | \{\theta_{i,k}\}, b_k) p(la_k^{(m)})}, 1\right\} \quad (\text{A.9})$$

If the candidate is accepted, then  $la_k^{(m+1)} = la_k^c$ , otherwise,  $la_k^{(m+1)} =$

$la_k^{(m)}$ . Parameter draws can be sequentially made for each  $la_k$ ,  $k = 1$  to  $r$ .

(j) The full conditional for the factor scores  $\eta_{ij}$  is a multivariate normal distribution. Consider the reduced form for the structural model  $\eta_{ij} = B_i^{-1}(\gamma_{i0} + \Gamma_i \xi_{ij} + \zeta_{ij})$ . Let  $\Omega = B_i^{-1} \Psi B_i^{-1'}$ . Then the full conditional is given by  $N(\tilde{\eta}_{ij}, V_{ij})$ , where  $V_{ij}^{-1} = \Omega^{-1} + \Lambda_y' \Theta_{ij}^{-1} \Lambda_y$  and  $\tilde{\eta}_{ij} = V_{ij}[\Omega^{-1}(B_i^{-1} \gamma_{i0} + B_i^{-1} \Gamma_i \xi_{ij}) + \Lambda_y' \Theta_{ij}^{-1} (y_i - \alpha_y)]$ .

(k) The full conditional for the structural parameters  $\pi_i = \{B_i, \gamma_{i0}, \Gamma_i\}$  associated with individual  $i$  is a multivariate normal distribution. For a recursive system of simultaneous equations, the structural model  $B_i \eta_{ij} = \gamma_{i0} + \Gamma_i \xi_{ij} + \zeta_{ij}$  is a general triangular system (see Zellner 1971, p. 252). This system can be recast as  $\eta_{ij} = \Xi_{ij} \pi_i + \zeta_{ij}$ , where  $\Xi_{ij}$  is an appropriately dimensioned matrix containing both the exogenous and endogenous factor scores. The key feature of a triangular system is that the determinant of the matrix of the coefficients of the endogenous variables  $\eta$  vanishes. Thus we can treat the system  $\eta_{ij} = \Xi_{ij} \pi_i + \zeta_{ij}$  as a seemingly unrelated regression (SUR) system. Given the prior  $p(\pi_i) = N(Z_i \vartheta, Y)$ , the full conditional is given by  $N(\tilde{\pi}_i, V_{\pi})$ , where  $V_{\pi}^{-1} = Y^{-1} + \sum_{j=1}^{N_i} \Xi_{ij}' \Psi^{-1} \Xi_{ij}$  and  $\tilde{\pi}_i = V_{\pi} [Y^{-1} Z_i \vartheta + \sum_{j=1}^{N_i} \Xi_{ij}' \Psi^{-1} \eta_{ij}]$ .

(l) The full conditional distribution for the precision matrix  $\Psi^{-1}$  of the structural model is Wishart and is given by

$$p(\Psi^{-1} | \{\xi_{ij}\}, \{\eta_{ij}\}, \{\gamma_{i0}\}, \Gamma_i, B_i) = W(\rho_{\text{pos}}, R_{\text{pos}}) \quad (\text{A.10})$$

where  $\rho_{\text{pos}} = \rho_{\Psi} + N$  and  $R_{\text{pos}} = [\sum_{i=1}^I \sum_{j=1}^{N_i} \tilde{\eta}_{ij} \tilde{\eta}_{ij}' + \rho_{\Psi} R_{\Psi}]^{-1}$  and  $\tilde{\eta}_{ij} = B_i \eta_{ij} - \gamma_{i0} - \Gamma_i \xi_{ij}$ .

(m) The full conditional for the population structural parameters  $\vartheta$  is a multivariate normal distribution and can be written as

$$p(\vartheta | \{\pi_i\}, Y) = N(\hat{\vartheta}, V_{\vartheta}) \quad (\text{A.11})$$

where  $V_{\vartheta}^{-1} = C^{-1} + \sum_{i=1}^I Z_i' Y^{-1} Z_i$  and  $\hat{\vartheta} = V_{\vartheta}(C^{-1} c + \sum_{i=1}^I Z_i' Y^{-1} \pi_i)$ .

(n) The full conditional for the precision matrix  $Y^{-1}$  of the individual-level structural parameters  $\pi_i$  is a Wishart distribution. Given the prior  $W(\rho_{\Psi}, (\rho_{\Psi} R_{\Psi})^{-1})$ , the full conditional can be written as

$$p(Y^{-1} | \{\pi_i\}) = W\left(\rho_{\Psi} + I, \left[\sum_{i=1}^I (\pi_i - Z_i \vartheta)(\pi_i - Z_i \vartheta)' + \rho_{\Psi} R_{\Psi}\right]^{-1}\right). \quad (\text{A.12})$$

(o) 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  $\Phi_i$  is given by  $\Phi_i^{-1} \sim W(\exp(\rho'), R)$ , the likelihood can be written as

$$L(\{\Phi_i^{-1}\} | \rho', R) = \frac{\exp\left(-\frac{1}{2} \text{tr}\left(R^{-1} \sum_{i=1}^I \Phi_i^{-1}\right)\right) \prod_{i=1}^I |\Phi_i^{-1}|^{(\exp(\rho') - p - 1)/2}}{|R|^{(\exp(\rho')I)/2} 2^{\exp(\rho')/2} \prod_{j=1}^p \Gamma\left(\frac{\exp(\rho') + 1 - j}{2}\right)}. \quad (\text{A.13})$$

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

$$p(R^{-1} | \{\Phi_i^{-1}\}, \rho, \gamma, S) = W(\gamma + I\rho, (\gamma S + \sum \Phi_i^{-1})^{-1}). \quad (A.14)$$

(p) The full conditional for the hyperparameter  $\rho' = \log(\rho)$  of the Wishart distribution over  $\Phi_i^{-1}$  cannot be written in closed form. This full conditional is proportional to the product of the likelihood expression specified above, and the prior density of  $\rho'$ , which is a univariate normal  $p(\rho') = N(0, \tau_{mm})$ . We therefore use a random-walk Metropolis-Hastings step to generate random draws of  $\rho'$ . To generate a candidate  $\rho'^c$ , we use a univariate normal proposal density  $N(\rho'^{(m)}, t)$  that is centered on the old value of  $\rho'^{(m)}$ . The tuning constant  $t$  in the proposal density is chosen to allow rapid mixing and to avoid excessively frequent rejections of the candidates. The generated candidate  $\rho'^c$  is accepted with the following acceptance probability:

$$\min \left\{ \frac{L(\rho'^c | \{\Phi_i^{-1}\}, R)p(\rho'^c)}{L(\rho'^{(m)} | \{\Phi_i^{-1}\}, R)p(\rho'^{(m)})}, 1 \right\} \quad (A.15)$$

If the candidate is accepted, then  $\rho'^{(m+1)} = \rho'^c$ ; otherwise,  $\rho'^{(m+1)} = \rho'^{(m)}$ .

## References

- Allenby, G. M., J. L. Ginter. 1995. Using extremes to design products and segment markets. *J. Marketing Res.* **32** 392–403.
- , P. E. Rossi. 1999. Marketing models of consumer heterogeneity. Special Issue: Marketing and Econometrics. T. Wansbeek and M. Wedel, eds. *J. Econometrics* **89** (1–2) 57–78.
- Ansari, A., K. Jedidi. 2000. Bayesian factor analysis for multilevel binary observations. *Psychometrika*, forthcoming.
- Arminger, G., B. Muthén. 1998. A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika* **63** 271–300.
- Best, N. G., M. K. Cowles, S. K. Vines. 1995. CODA: *Convergence Diagnostics and Output Analysis Software for Gibbs Sampler Output: Version 0.3*. Technical report, Biostatistics Unit-MRC, Cambridge, U.K.
- Bollen, K. A. 1989. *Structural Equations with Latent Variables*. Wiley Interscience, New York.
- Chib, S., E. Greenberg. 1995. Understanding the Metropolis-Hastings algorithm. *Amer. Statist.* **49** 327–335.
- Chintagunta, P., D. C. Jain, N. J. Vilcassim. 1991. Investigating heterogeneity in brand preferences in logit models for panel data. *J. Marketing Res.* **28** 417–428.
- Gelfand, A. E. 1996. Model determination using sampling-based methods. W. R. Gilks, S. Richardson, D. J. Spiegelhalter, eds. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London, 145–161.
- , A. F. M. Smith. 1990. Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** 972–985.
- Geman, S., D. Geman. 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intelligence* **6** 721–741.
- Goldstein, H., R. P. McDonald. 1988. A general model for the analysis of multilevel data. *Psychometrika* **553** 455–467.
- Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109.
- Jagpal, Sharan. 1999. *Marketing Strategy and Uncertainty*. Oxford University Press, New York.
- Jedidi, K., H. S. Jagpal, W. S. DeSarbo. 1997. Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Sci.* **16** (1) 39–59.
- Johnson, M. D., E. W. Anderson, C. Fornell. 1995. Rational and adaptive performance expectations in a customer satisfaction framework. *J. Consumer Res.* **21** 695–707.
- Jöreskog, K. G. 1971. Simultaneous factor analysis in several populations. *Psychometrika* **36** 409–426.
- Kamakura, W. A., G. J. Russell. 1989. A probabilistic choice model for market segmentation and elasticity structure. *J. Marketing Res.* **26** (November) 379–390.
- Lee, S.-Y. 1981. A Bayesian approach to confirmatory factor analysis. *Psychometrika* **46** 153–160.
- Longford, N. T., B. Muthén. 1992. Factor analysis for clustered observations. *Psychometrika* **57** 581–597.
- Lord, F. M., M. R. Novick. 1968. *Statistical Theories of Mental Test Scores*. Addison-Wesley, Reading, MA. 129–131.
- Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller. 1953. Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1091.
- Muthén, B. 1989. Latent variable modeling in heterogeneous populations. *Psychometrika* **54** 557–585.
- . 1994. Multilevel covariance structure analysis. *Sociol. Methods Res.* **22** 376–398.
- Newton, M. A., A. E. Raftery. 1994. Approximate Bayesian inference by the weighted likelihood bootstrap (with discussion). *J. Roy. Statist. Soc. Ser. B.* **56** 3–18.
- Oliver, R. L. 1993. Cognitive, affective, and attribute bases of the satisfaction response. *J. Consumer Res.* **20** 418–430.
- . 1997. *Satisfaction: A Behavioral Perspective on the Consumer*. McGraw-Hill, New York.
- Rossi, P., Z. Gilula, G. Allenby. 1999. Overcoming scale heterogeneity. Working paper, University of Chicago, Chicago, IL.
- Scheines, R., H. Hoijtink, A. Boomsma. 1999. Bayesian estimation and testing of structural equation models. *Psychometrika* **64** 37–52.
- Sörbom, D. 1981. Structural equation models with structured means. K. G. Jöreskog, H. Wold, eds. *Systems Under Indirect Observation: Causality, Structure, And Prediction*. North Holland, Amsterdam, Netherlands. 183–195.
- Tanner, M. A., W. H. Wong. 1987. The calculation of posterior distributions by data augmentation (with Discussion). *J. Amer. Statist. Assoc.* **82** 528–550.
- Yung, Yiu-Fai. 1997. Finite mixtures in confirmatory factor-analysis models. *Psychometrika* **62** (3) 297–330.
- Zellner, Arnold. 1971. *An Introduction to Bayesian Inference in Econometrics*. Wiley, New York.

This paper was received August 5, 1999, and has been with the authors 6 months for 3 revisions; processed by Greg Allenby.