Bayesian Analysis (2008) **3**, Number 1, pp. 197–224

A Bayesian Structural Equations Model for Multilevel Data with Missing Responses and Missing Covariates

Sonali Das*, Ming-Hui Chen†, Sungduk Kim‡ and Nicholas Warren§

Abstract. Motivated by a large multilevel survey conducted by the US Veterans Health Administration (VHA), we propose a structural equations model which involves a set of latent variables to capture dependence between different responses, a set of facility level random effects to capture facility heterogeneity and dependence between individuals within the same facility, and a set of covariates to account for individual heterogeneity. Identifiability associated with structural equations modeling is addressed and properties of the proposed model are carefully examined. An effective and practically useful modeling strategy is developed to deal with missing responses and to model missing covariates in the structural equations framework. Markov chain Monte Carlo sampling is used to carry out Bayesian posterior computation. Several variations of the proposed model are considered and compared via the deviance information criterion. A detailed analysis of the VHA all employee survey data is presented to illustrate the proposed methodology.

Keywords: DIC, Latent variable, Markov chain Monte Carlo, missing at random, random effects, VHA all employee survey data

1 Introduction

It is challenging to model a large scale survey, conducted to evaluate system dynamics. Often, such modeling requires the implementation of structural equations modeling (SEM) as they are a powerful multivariate regression technique to handle scenarios in which the predictor and outcome variables can both be either latent or observed. This feature of SEM has been appealing to many researchers, especially in the fields of behavioral sciences (Bollen (1989), Bentler and Wu (2002), Jöreskog and Sörbom (1996), and Sánchez et al. (2005)). SEM has been extensively used to model survey data arising in the fields of sociology, psychology, health and economics – with increasing applications where self assessment questionnaires are the means to collect data. A very comprehensive application of latent variables in psychology and social sciences is available in Bollen (2002).

The data we investigate are from the Veterans Health Administration (VHA), which is the largest integrated health care provider in the United States of America and has

*Logistics and Quantitative Methods, CSIR BE, Pretoria, South Africa, mailto:sdas@csir.co.za
†Department of Statistics, University of Connecticut, Storrs, CT, mailto:mhchen@stat.uconn.edu
‡Division of Epidemiology, Statistics and Prevention Research,National Institute of Child Health and Human Development, Rockville, MD, mailto:kims2@mail.nih.gov
§University of Connecticut Health Center, Farmington, CT, mailto:warren@nso.uchc.edu

 ba0009

initiated substantive efforts to improve quality and safety of patient care over the last decade, with considerable success (Jha et al. (2003), Greenfield and Kaplan (2004), and Asch et al. (2004)). The focus of the VHA All Employee Survey (AES) was to ascertain areas of organizational climate and performance that need attention in the workplace, with the larger objective being to improve the quality of service provided by the VHA. In this paper, we propose a structural equations model to analyze and assess an all employees survey conducted by the VHA in 1997 to evaluate organizational climate and performance that need attention with primary focus on 3 outcome variables. The main objective is to develop an individual level model that captures the association between a set of variables of interest, via a set of latent variables, taking into account the natural pattern of heterogeneity in the sample that is due to the nesting of individuals into their respective higher hospital cluster.

The rest of the paper is organized as follows. Section 2 provides a detailed description of the VHA AES 1997 data. The development of the structural equations model for such survey data is given in Section 3. The likelihood, the prior, and the posterior based on the proposed model are discussed in Section 4. Models are assessed using the Bayesian Deviance Information Criterion (DIC), and the appropriate deviance function is derived in Section 5. A comprehensive data analysis of the AES 1997 data is given in Section 6. We conclude the paper with a discussion of the various issues encountered in this development in Section 7. Details of the computational algorithm to sample from the posterior distributions are given in the Appendix.

2 Data

To maintain and improve quality of service, the VHA needs the understanding of employee working conditions, and to that effect the first such Veterans Health Administration All Employee Survey was conducted in 1997. The survey was conducted via questionnaires that participants had to self report. Here the workplaces are the various hospitals under the VHA, called facilities, and the target participants were all employees. Based on the findings, regional networks and individual facilities are expected to undertake intervention measures to improve areas of weakness. Of course, not all employees participated from each of the facilities. Moreover, as in most large scale surveys, there was a lot of missing data even in questionnaires of those who did participate. One feature of this survey is that due to the sensitive nature of many questions, a labor/management confidentiality agreement was reached where the identification of individual respondents was not recorded. However, the identification of the hospital (facility) to which a respondent reported was traceable. Each facility has a distinct character, and ignoring this natural heterogeneity due to the clustering of individual employees into facilities can lead to misleading conclusions. In fact, we can also assume that there exists an interaction effect between facility and response to question.

With the overall objective being to capture the association between variables that characterize organizational climate at the individual level, our primary focus is on 3 outcome variables, viz., — Customer satisfaction, Employee satisfaction and Quality, via their relationship to 3 important and intrinsic workplace characteristics or traits, viz., — Leadership, Support and Resources. These 3 traits are theoretical constructs, or latent variables, and as such are not directly measurable. Hence, we identify 21 other observed variables that can be considered as manifestations of the 3 latent variables, and thus, by combining the 3 variable types – outcome, observed and latent, via structural equations modeling, as we will elaborate in Sections 3 and 4, we aim to capture the objective of this investigation.

The part of the AES 1997 data we work with has 111,249 individuals belonging to one of 154 facilities. The variables under consideration in the data include a facility identifier, 3 outcome variables, and 21 other manifest variables. We also consider 3 demographic variables as covariates for each individual, viz., — Age, Gender and Years of service with the VA. For this subset of the AES 1997 data, the 21 observed manifest variables were composite or aggregate variables, which can be considered continuous. Once again, the survey has its share of missing data in both the observed response variables as well as the covariates. Brief descriptive statistics comprising mean, standard deviation (SD) and the missing percentage for each of the variables considered are presented in Tables 1 and 2.

Variable	Mean	SD	Missing	Variable	Mean	SD	Missing
REW_IND	2.69	1.05	1159	LEAD_INV	2.86	0.96	112
REW_SERV	2.80	1.08	1220	RES_SKLL	2.78	1.03	255
DIV_ZERO	3.30	1.05	551	RES_SUPP	2.65	0.99	569
DIV_DIFF	3.77	0.96	6628	RES_SAFE	3.34	0.87	406
CUST_NDS	3.36	1.06	3966	WORK_FAM	3.04	0.93	1801
CUST_INF	3.49	0.91	5225	COOP_TEAM	2.99	1.02	218
PAY_SAT	3.00	1.06	747	PLAN_EVL	3.25	0.79	59
EMP_DEVL	3.18	0.96	17	DIV_MGRS	3.32	1.01	5568
INNOV	2.86	0.91	18	SUP_SUPT	2.94	1.06	1053
LEAD_GOAL	3.00	0.96	258	CUST_SAT	3.45	0.97	2193
COOP_RES	3.65	1.12	1982	EMP_SAT	3.42	1.06	1971
CONF_RES	2.69	1.13	3582	QUALITY	3.96	0.91	1114

Table 1: Descriptive Statistics of Manifest and Outcome Variables of AES 1997

Among those who responded, a majority of 57% were females, about two thirds were 49 years or younger, and about 70% were with the VA for more than 5 years. There were 21078 (18.9 %) individuals with at least one manifest information missing, 6383 (5.7 %) individuals with at least one covariate missing, and overall, there were 24723 (22.2 %) individuals with at least one of manifest or covariate information missing. Thus, there were about 2.4% of respondents with one response from both manifest variables and covariate information missing. By considering *complete-response* (CR) over *all-cases* (AC) we loose about 18.9% of the data, and by considering *complete-cases* (CC) only, we loose a further 2.4% of the data.

Covariate	Category	Frequency	Percentage
Gender	Female	62819	57%
	Male	43574	39%
	Missing	4856	4%
Age	$\leq$ 49 years	72914	65%
	$>$ 49 years	34197	31%
	Missing	4138	4%
Years in VA	$\leq$ 5 years	28258	25%
	$>$ 5 years	79468	71%
	Missing	3523	4%

Table 2: Descriptive Statistics of the 3 Covariates Considered in the AES 1997

3 Model

Latent variable modeling has gained increased attention recently (Sammel and Ryan (1996) and Muthén (2002)). Latent variable models for multilevel data that account for heterogeneity are often appropriate because many experimental and survey data are nested. A unified framework for generalized multilevel structural equations modeling is introduced and discussed in Rabe-Hesketh et al. (2004). Bayesian methods to handle heterogeneity in SEM have shown that ignoring heterogeneity can result in inflated estimates of measurement reliability, wrong signs of factor covariances, and can yield inappropriate model fit and standard errors (Ansari et al. (2000)). Modeling the effect of covariates on latent variables or on outcome variables is a recent development (Sammel et al. (1997) and Arminger and Küsters (1988)). Bayesian methods for analyzing raw data with clustered structure via SEM are investigated by Dunson (2000), Ansari et al. (2000), and Dunson and Perreault (2001). Computationally, Lee and Shi (2001) and Lee and Song (2004a) developed the efficient Monte Carlo EM algorithms for maximum likelihood estimation for multilevel latent variable models with mixed continuous and discrete data. Issues of Bayesian estimation via SEM or nonlinear SEM are discussed in Scheines et al. (1999) and Song and Lee (2004). Furthermore, Lee and Song (2004b) carried out Bayesian model comparison via Bayes factor of nonlinear SEM with missing response data.

Capturing heterogeneity in SEM can be handled in different ways depending on the focus of investigation. If the focus is on the aggregate level, latent variable modeling can be devised to investigate differences between populations with respect to the within and between group covariance matrix. However, in our investigation, the focus is on individual level modeling, along with accounting for any facility level effects, and all inferences thereof are at the individual level. There are two major issues in an analysis of survey data: first, most models considered for comparison are nested, and second, there is a problem of missing data. Missing data can arise because of missing response, missing covariate, or both. Moreover, as latent variables are important in SEM, they may also be viewed as missing data. We proceed to address all these issues in the development of the model.

To capture the association structure of a set of latent variables and a set of responses of interest, we need to identify a set of response variables that can be considered as reasonable manifestations of the latent variables. We recall that latent variables represent the constructs we *want* to study, and are forced to do so via a set of observable variables we *can* study. Let y_{ijk} be the k^{th} response by the j^{th} individual belonging to the i^{th} facility for $i = 1, 2, \ldots, I$, $j = 1, 2, \ldots, n_i$, and $k = 1, 2, \ldots, K$, where I denotes the total number of the facilities, n_i is the number of individuals within the i^{th} facility, and K is the total number of responses considered. We propose the measurement model part of the SEM as follows

$$y_{ijk} = \mu_k + \tau_i + \tau_{ik} + \boldsymbol{\beta}_k' \boldsymbol{\omega}_k \boldsymbol{\eta}_{ij} + \boldsymbol{\phi}_k' Z_{ij} + \epsilon_{ijk}, \tag{1}$$

where $\epsilon_{ijk} \overset{iid}{\sim} \mathcal{N}(0, \sigma_k^2)$ for $k = 1, 2, \ldots, K$, μ_k is the mean effect due to response k, $\boldsymbol{\beta}_k$ is a p_k-dimensional column vector of coefficients loading on $\boldsymbol{\eta}_{ij}$, a r-dimensional vector of zero mean latent variables, $\boldsymbol{\omega}_k$ is the $p_k \times r$ fixed loading matrix that controls the dependence of response k on the set of latent variables, and $\boldsymbol{\phi}_k$ is the q-dimensional vector of regression coefficients corresponding to Z_{ij}, a q-dimensional vector of covariates. In (1), τ_i and τ_{ik} are random components to capture facility effect due to an individual j belonging to facility i, and the facility i and response k interaction, respectively. We assume that $\tau_i \overset{iid}{\sim} N(0, \sigma_\tau^2)$, $\tau_{ik} \overset{iid}{\sim} N(0, \sigma_{\tau^*}^2)$, and τ_i, τ_{ik}, and ϵ_{ijk} are mutually independent for $i = 1, \ldots, I$, $j = 1, 2, \ldots, n_i$, and $k = 1, \ldots, K$.

Based on the SEM given by (1), the mean and variance of y_{ijk} conditional on $(\mu_k, \boldsymbol{\beta}_k, \boldsymbol{\phi}_k, \mathbf{z}_{ij})$ are

$$\mu_{ijk} = E(y_{ijk} | \mu_k, \boldsymbol{\beta}_k, \boldsymbol{\phi}_k, \mathbf{z}_{ij}) = \mu_k + \boldsymbol{\phi}_k' \mathbf{z}_{ij} \tag{2}$$

and

$$\sigma_{ijk}^2 = \mathrm{Var}(y_{ijk} | \mu_k, \boldsymbol{\beta}_k, \boldsymbol{\phi}_k, \mathbf{z}_{ij}) = \boldsymbol{\beta}_k' \boldsymbol{\omega}_k \mathrm{Var}(\boldsymbol{\eta}_{ij}) \boldsymbol{\omega}_k' \boldsymbol{\beta}_k + \sigma_\tau^2 + \sigma_{\tau^*}^2 + \sigma_k^2.$$

For the same individual indexed by i and j, the covariance between answering different questions can be quantified as

$$\mathrm{Cov}(y_{ijk}, y_{ijk'}) = \boldsymbol{\beta}_k' \boldsymbol{\omega}_k \mathrm{Var}(\boldsymbol{\eta}_{ij}) \boldsymbol{\omega}_{k'}' \boldsymbol{\beta}_{k'} + \sigma_\tau^2$$

for $k \neq k'$; for different individuals in the same facility i answering different questions, this covariance equals $\mathrm{Cov}(y_{ijk}, y_{ij'k'}) = \sigma_\tau^2$ for $j \neq j'$ and $k \neq k'$; while the covariance between two individuals in the same facility i answering the same question is $\mathrm{Cov}(y_{ijk}, y_{ij'k}) = \sigma_\tau^2 + \sigma_{\tau^*}^2$ for $j \neq j'$. Observe that the covariance structure reflects the response pattern of different individuals belonging to the same facility who answer different questions, as well as different individuals within the same facility who answer the same question. The variability in response in the former is solely due to the random effect due to facility effect. However, in the latter, the variability is accounted for by the facility randomness as well as an additional component from the variability due to individual effect as they respond to the same question. While in the former the covariation between responding to different questions by the same individual in a particular facility is accounted for by the facility effect variability, in the latter the variability is accounted for by the structural dependency as well as the facility effect.

The structural part of the model is given by

$$\boldsymbol{\eta}_{ij} = \Gamma\boldsymbol{\eta}_{ij} + \boldsymbol{\xi}_{ij}, \tag{3}$$

where $\boldsymbol{\xi}_{ij} = (\xi_{ij1}, \xi_{ij2}, \ldots, \xi_{ijr})' \sim \mathcal{N}(0, \mathrm{diag}(\sigma^2_{\eta_1}, \ldots, \sigma^2_{\eta_r}))$, and $\boldsymbol{\xi}_{ij}$ is assumed to be independent of ϵ_{ijk}, τ_i, and τ_{ik}. In writing the structural part as in (3), we deviate a little from the conventional representation by considering a single loading matrix for endogenous and exogenous latent variables together as Γ, $\boldsymbol{\eta}_{ij}$ is the vector of latent variables associated with individual j in facility i and is of length r. In (3), Γ is the $r \times r$ loading matrix. The special feature of Γ is that the diagonal elements are all zero, and the rows corresponding to the exogenous latent variables are all zero since a latent variable cannot define itself, and an exogenous variable does not depend on any other variables within the system. We assume that $I - \Gamma$ is invertible.

Now we return to discuss further the nature of the covariance matrix associated with $\boldsymbol{\eta}$. Let $\mathrm{Var}(\boldsymbol{\eta}_{ij}) = V_\eta$. In (3), we run into the scaling problem between the variance of $\boldsymbol{\eta}_{ij}$ and the scale of $\boldsymbol{\beta}_k$. Specifically, if the variances of η_{ijr}'s become large, then $\boldsymbol{\beta}_k$ in (1) becomes smaller. To circumvent this problem, we set the variances of all components of the latent variables $\boldsymbol{\eta}_{ij}$ to 1, with the aim to cut this circular dependence between $\boldsymbol{\beta}$'s and the elements of Γ. To this end, we assume that we have a structural model (3) in terms of $\boldsymbol{\eta}^*$ and $\boldsymbol{\xi}^*$ as

$$\boldsymbol{\eta}^*_{ij} = \Gamma^*\boldsymbol{\eta}^*_{ij} + \boldsymbol{\xi}^*_{ij} \quad \text{or} \quad \boldsymbol{\eta}^*_{ij}(I - \Gamma^*) = \boldsymbol{\xi}^*_{ij}, \tag{4}$$

where $\boldsymbol{\xi}^*_{ij} \sim \mathcal{N}(0, I_r)$. Assume that the matrix $(I - \Gamma^*)$ is invertible. Then, from (4), we obtain

$$\mathrm{Var}(\boldsymbol{\eta}^*_{ij}) = (I - \Gamma^*)^{-1}\left[(I - \Gamma^*)'\right]^{-1}. \tag{5}$$

We define a diagonal matrix D whose r^{*th} diagonal element is the inverse of the r^{*th} diagonal element of (5) for $r^* = 1, 2, \ldots, r$. We use this diagonal matrix D to re-parameterize (4) by pre-multiplying (4) by $D^{\frac{1}{2}}$ on both sides to get

$$D^{\frac{1}{2}}\boldsymbol{\eta}^*_{ij} = D^{\frac{1}{2}}\Gamma^*\boldsymbol{\eta}^*_{ij} + D^{\frac{1}{2}}\boldsymbol{\xi}^*_{ij}. \tag{6}$$

We write $D^{\frac{1}{2}}\boldsymbol{\eta}^*_{ij} = \boldsymbol{\eta}_{ij}$ and $D^{\frac{1}{2}}\boldsymbol{\xi}^*_{ij} = \boldsymbol{\xi}_{ij}$. Then (6) can be written as

$$\boldsymbol{\eta}_{ij} = D^{\frac{1}{2}}\Gamma^*\boldsymbol{\eta}^*_{ij} + \boldsymbol{\xi}_{ij} = D^{\frac{1}{2}}\Gamma^* D^{-\frac{1}{2}}\boldsymbol{\eta}_{ij} + \boldsymbol{\xi}_{ij} = \Gamma\boldsymbol{\eta}_{ij} + \boldsymbol{\xi}_{ij},$$

where $\Gamma = D^{\frac{1}{2}}\Gamma^* D^{-\frac{1}{2}}$. Thus, $\boldsymbol{\xi}_{ij} \sim \mathcal{N}(0, D)$ and the variance of the latent variables $\boldsymbol{\eta}_{ij}$

can be evaluated as

$$\begin{aligned} V_{\boldsymbol{\eta}} &= \mathrm{Var}(\boldsymbol{\eta}_{ij}) = \left(I - D^{\frac{1}{2}}\Gamma^* D^{-\frac{1}{2}}\right)^{-1} D\left\{\left(I - D^{\frac{1}{2}}\Gamma^* D^{-\frac{1}{2}}\right)'\right\}^{-1} \\ &= D^{\frac{1}{2}}(I-\Gamma^*)^{-1}\left[(I-\Gamma^*)'\right]^{-1} D^{\frac{1}{2}}, \end{aligned} \tag{7}$$

which by the way we defined D ensures that the diagonal elements of (7) are exactly all equal to unity. In addition, we have $\mathrm{diag}(\sigma^2_{\eta_1}, \ldots, \sigma^2_{\eta_r})) = D$.

For the AES 1997 data, the loading of the 21 manifest variables and 3 outcome variables in the measurement model, as well as the structural model of the SEM is illustrated in Figure 1. We note that (3) is a conventional representation of the structural part of the SEM, which is used in SAS PROC CALIS and also discussed in detail in Hatcher (2000).

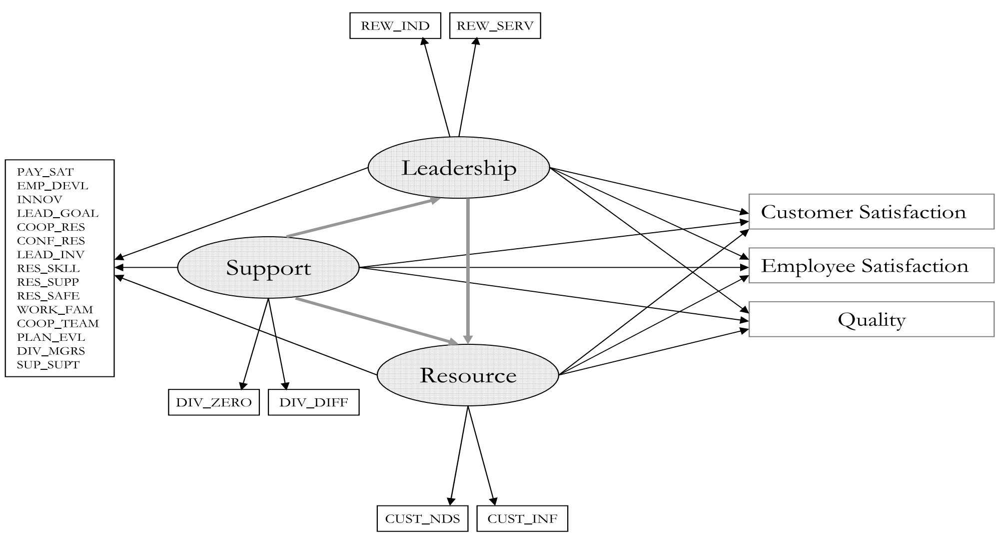

Figure 1: Path diagram for the AES 1997.

We illustrate the re-parameterization for the structural part of the SEM in Figure 1 as

$$\begin{aligned} \eta^*_{ij1} &= \gamma_1\eta^*_{ij2} + \xi^*_{ij1}, \\ \eta^*_{ij2} &= \xi^*_{ij2}, \\ \eta^*_{ij3} &= \gamma_2\eta^*_{ij2} + \gamma_3\eta^*_{ij1} + \xi^*_{ij3}, \end{aligned} \tag{8}$$

where $\boldsymbol{\xi}^*_{ij} \sim \mathcal{N}(0, I_3)$. From (8) we have

$$\Gamma^* = \begin{pmatrix} 0 & \gamma_1 & 0 \\ 0 & 0 & 0 \\ \gamma_3 & \gamma_2 & 0 \end{pmatrix} \text{ and } (I - \Gamma^*) = \begin{pmatrix} 1 & -\gamma_1 & 0 \\ 0 & 1 & 0 \\ -\gamma_3 & -\gamma_2 & 1 \end{pmatrix}$$
$$\Longrightarrow D = \begin{pmatrix} \frac{1}{1+\gamma_1^2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{1+\gamma_3^2+(\gamma_2+\gamma_1\gamma_3)^2} \end{pmatrix}.$$

After reparameterization, we have

$$\boldsymbol{\eta}_{ij} = D^{\frac{1}{2}} \boldsymbol{\eta}^*_{ij} \sim \mathcal{N}(0, V_\eta), \quad \boldsymbol{\xi}_{ij} = D^{\frac{1}{2}} \boldsymbol{\xi}^*_{ij} \sim \mathcal{N}(0, D),$$

and

$$\begin{aligned} V_\eta &= D^{1/2} \times \begin{pmatrix} 1+\gamma_1^2 & \gamma_1 & \gamma_3 + \gamma_1(\gamma_2+\gamma_1\gamma_3) \\ \gamma_1 & 1 & \gamma_2 + \gamma_1\gamma_3 \\ \gamma_3 + \gamma_1(\gamma_2+\gamma_1\gamma_3) & \gamma_2+\gamma_1\gamma_3 & \gamma_3^2 + (\gamma_2+\gamma_1\gamma_3)^2 + 1 \end{pmatrix} \times D^{1/2} \\ &= \begin{pmatrix} 1 & \frac{\gamma_1}{\sqrt{1+\gamma_1^2}} & \frac{\gamma_1\gamma_2+(1+\gamma_1^2)\gamma_3}{\sqrt{1+\gamma_1^2}\sqrt{1+\gamma_3^2+(\gamma_2+\gamma_1\gamma_3)^2}} \\ \frac{\gamma_1}{\sqrt{1+\gamma_1^2}} & 1 & \frac{\gamma_2+\gamma_1\gamma_3}{\sqrt{1+\gamma_3^2+(\gamma_2+\gamma_1\gamma_3)^2}} \\ \frac{\gamma_1\gamma_2+(1+\gamma_1^2)\gamma_3}{\sqrt{1+\gamma_1^2}\sqrt{1+\gamma_3^2+(\gamma_2+\gamma_1\gamma_3)^2}} & \frac{\gamma_2+\gamma_1\gamma_3}{\sqrt{1+\gamma_3^2+(\gamma_2+\gamma_1\gamma_3)^2}} & 1 \end{pmatrix}. \end{aligned}$$

Now, the structural model of the SEM can be explicitly written as

$$\begin{aligned} \eta_{ij1} &= \tfrac{\gamma_1}{\sqrt{1+\gamma_1^2}} \eta_{ij2} + \xi_{ij1}, \\ \eta_{ij2} &= \xi_{ij2}, \\ \eta_{ij3} &= \tfrac{\gamma_2}{\sqrt{1+\gamma_3^2+(\gamma_2+\gamma_1\gamma_3)^2}} \eta_{ij2} + \tfrac{\sqrt{(1+\gamma_1^2)}\gamma_3}{\sqrt{1+\gamma_3^2+(\gamma_2+\gamma_1\gamma_3)^2}} \eta_{ij1} + \xi_{ij3}, \end{aligned} \tag{9}$$

where $\xi_{ij1} \overset{iid}{\sim} \mathcal{N}\left(0, \frac{1}{1+\gamma_1^2}\right)$, $\xi_{ij2} \overset{iid}{\sim} \mathcal{N}(0, 1)$, and $\xi_{ij3} \overset{iid}{\sim} \mathcal{N}\left(0, \frac{1}{1+\gamma_3^2+(\gamma_2+\gamma_1\gamma_3)^2}\right)$. Observe that the diagonal elements of V_η above all turn out to equal unity.

As discussed in Section 2, there were missing values in both response variables y_{ijk}'s and covariates Z_{ijq}'s. With the protection of the labor/management confidentiality agreement, it is unlikely that missing responses are due to the sensitive nature of questions as the identification of individual respondents was not recorded. In addition, it does not appear that there were any apparent systematic patterns in the missing values of y_{ijk}'s or Z_{ijq}'s in the AES 1997 data. Therefore, it is reasonable to assume that any missingness in y_{ijk} and covariates Z_{ijq} is missing at random (MAR) (Rubin (1976) and Little and Rubin (2002)). Letting $\boldsymbol{y}_{ij} = (y_{ij1}, y_{ij2}, \ldots, y_{ijK})'$, we thus partition the response vector $\boldsymbol{y}'_{ij}$ into $(\boldsymbol{y}'_{ij,obs}, \boldsymbol{y}'_{ij,mis})$, and the vector of corresponding covariates Z'_{ij} as $(Z'_{ij,obs}, Z'_{ij,mis})$, where suffix *obs* corresponds to the collection of observed elements of the respective vectors, and suffix *mis* corresponds to the collection of missing elements of the respective vectors. Further, we introduce an indicator δ_{ijk} for the observed and missing response in the data as

$$\delta_{ijk} = \begin{cases} 1 \text{ if } y_{ijk} \text{ is observed} \\ 0 \text{ if } y_{ijk} \text{ is missing} \end{cases}.$$

As discussed in Ibrahim et al. (1999) and Ibrahim et al. (2005), for MAR missing response y_{ijk} we do not need to model the missing data mechanism. We write

$$\begin{aligned} & f(y_{ijk}|\mu_k, \tau_i, \tau_{ik}, \boldsymbol{\beta}_k, \boldsymbol{\eta}_{ij}, \boldsymbol{\phi}_k, \sigma_k^2, Z_{ij}) \\ = \quad & \frac{1}{\sqrt{2\pi}\sigma_k} \exp\Big\{ -\frac{1}{2\sigma_k^2}[y_{ijk} - (\mu_k + \tau_i + \tau_{ik} + \boldsymbol{\beta}_k'\boldsymbol{\omega}_k\boldsymbol{\eta}_{ij} + \boldsymbol{\phi}_k' Z_{ij})]^2 \Big\}. \end{aligned} \quad (10)$$

Let $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_K)'$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \ldots, \boldsymbol{\beta}_K')'$, $\boldsymbol{\phi} = (\boldsymbol{\phi}_1', \boldsymbol{\phi}_2', \ldots, \boldsymbol{\phi}_K')'$, $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \ldots, \sigma_K^2)'$, and $\boldsymbol{\tau}_i^* = (\tau_{i1}, \tau_{i2}, \ldots, \tau_{iK})'$. Given $\boldsymbol{\mu}, \tau_i, \boldsymbol{\tau}_i^*, \boldsymbol{\beta}, \boldsymbol{\eta}_{ij}, \boldsymbol{\phi}, \boldsymbol{\sigma}^2$, and Z_{ij}, y_{ij1}, y_{ij2}, $\ldots$, y_{ijK} are independent. Thus, we have

$$\int f(\boldsymbol{y}_{ij,mis}|\boldsymbol{\mu}, \tau_i, \boldsymbol{\tau}_i^*, \boldsymbol{\beta}, \boldsymbol{\eta}_{ij}, \boldsymbol{\phi}, \boldsymbol{\sigma}^2, Z_{ij}) d\boldsymbol{y}_{ij,mis} = 1,$$

where $f(\boldsymbol{y}_{ij,mis}|\boldsymbol{\mu}, \tau_i, \boldsymbol{\tau}_i^*, \boldsymbol{\beta}, \boldsymbol{\eta}_{ij}, \boldsymbol{\phi}, \boldsymbol{\sigma}^2, Z_{ij})$ is the joint conditional probability density function of $\boldsymbol{y}_{ij,mis}$ given $\boldsymbol{\mu}, \tau_i, \boldsymbol{\tau}_i^*, \boldsymbol{\beta}, \boldsymbol{\eta}_{ij}, \boldsymbol{\phi}, \boldsymbol{\sigma}^2$, and Z_{ij}.

We denote the distribution of the covariates Z_{ij} by $f(Z_{ij}|\boldsymbol{\alpha})$, which may be specified by a sequence of q one-dimensional conditional distributions proposed by Lipsitz and Ibrahim (1996) and Chen et al. (2008). Specifically, we write the distribution of the q-dimensional covariate vector $Z_{ij} = (z_{ij1}, z_{ij2}, \ldots, z_{ijq})'$ as

$$\begin{aligned} & f(z_{ij1}, z_{ij2}, \ldots, z_{ijq}|\boldsymbol{\alpha}) \\ = \quad & f(z_{ijq}|z_{ij1}, \ldots, z_{ij,q-1}, \boldsymbol{\alpha}_q) f(z_{ij,q-1}|z_{ij1}, \ldots, z_{ij,q-2}, \boldsymbol{\alpha}_{q-1}) \ldots f(z_{ij1}|\boldsymbol{\alpha}_1), \end{aligned} \quad (11)$$

where $\boldsymbol{\alpha}_l$ is a vector of parameters for the l^{th} conditional distribution, the $\boldsymbol{\alpha}_l$'s are distinct, and moreover, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1', \boldsymbol{\alpha}_2', \ldots, \boldsymbol{\alpha}_q')'$. Lipsitz and Ibrahim (1996) have shown that the covariate model described as above closely approximates the joint log-linear model for the covariates, and that the ordering of the covariates has no significant effect on the parameter estimation in the sampling scheme. It is important to note that such a model needs to be specified only for the AC and the CR scenarios. In the Bayesian framework, the specification of the covariate distribution in (11) has other attractive features, such as easing the prior elicitation of the $\boldsymbol{\alpha}$, since Gaussian priors are often suitable. The other advantage of writing the joint covariate distribution as in (11) is that it eases the computational burden on the Gibbs scheme since at each instance only a one-dimensional conditional covariate distribution is considered. Furthermore, if each conditional distribution belongs to the exponential family, we have the desired log-concave property of the posterior distributions as long as the priors are log-concave.

4 Prior and Posterior Distributions

Our parameters of interest are $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\sigma}^2, \sigma_\tau^2, \sigma_{\tau^*}^2, \boldsymbol{\alpha})'$, where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_r)'$ is a r-dimensional vector of the parameters in (3). Let $\boldsymbol{\eta} = (\boldsymbol{\eta}'_{ij},\ j = 1, 2, \ldots, n_i,\ i = 1, 2, \ldots, I)'$, $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_I)'$, and $\boldsymbol{\tau}^* = ((\boldsymbol{\tau}_1^*)', (\boldsymbol{\tau}_2^*)', \ldots, (\boldsymbol{\tau}_I^*)')'$. Also let $D_{obs} = (\boldsymbol{y}_{obs}, Z_{obs})$ denote the observed data, where $\boldsymbol{y}_{obs} = (\boldsymbol{y}'_{ij,obs},\ j = 1, 2, \ldots, n_i,\ i = 1, 2, \ldots, I)'$ and $Z_{obs} = (Z_{ij,obs},\ j = 1, 2, \ldots, n_i,\ i = 1, 2, \ldots, I)'$. Then, based on the model given in Section 3, the likelihood function for $\boldsymbol{\theta}$, $\boldsymbol{\eta}$, $\boldsymbol{\tau}$, and $\boldsymbol{\tau}^*$ given D_{obs} is

$$\begin{aligned} & L(\boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\tau}, \boldsymbol{\tau}^* | D_{obs}) \\ = & \int \prod_{i=1}^{I} \Big\{ \prod_{j=1}^{n_i} \Big[\prod_{k=1}^{K} f^{\delta_{ijk}} (y_{ijk} | \mu_k, \tau_i, \tau_{ik}, \boldsymbol{\beta}_k, \boldsymbol{\eta}_{ij}, \boldsymbol{\phi}_k, \sigma_k^2, Z_{ij}) \Big] f(\boldsymbol{\eta}_{ij} | \gamma) \\ & \times f(Z_{ij,obs}, Z_{ij,mis} | \boldsymbol{\alpha}) \Big\} \Big[\prod_{k=1}^{K} f(\tau_{ik} | \sigma_{\tau^*}^2) \Big] f(\tau_i | \sigma_\tau^2) dZ_{mis}, \end{aligned} \tag{12}$$

where $Z_{mis} = (Z_{ij,mis},\ j = 1, 2, \ldots, n_i,\ i = 1, 2, \ldots, I)'$, $f(y_{ijk} | \mu_k, \tau_i, \tau_{ik}, \boldsymbol{\beta}_k, \boldsymbol{\eta}_{ij}, \boldsymbol{\phi}_k, \sigma_k^2, Z_{ij})$ is given by (10), $f(\tau_i | \sigma_\tau^2) = \frac{1}{\sqrt{2\pi}\sigma_\tau} \exp\big\{ -\frac{\tau_i^2}{2\sigma_\tau^2} \big\}$, $f(\tau_{ik} | \sigma_{\tau^*}^2) = \frac{1}{\sqrt{2\pi}\sigma_{\tau^*}} \exp\big\{ -\frac{\tau_{ik}^2}{2\sigma_{\tau^*}^2} \big\}$, $f(\boldsymbol{\eta}_{ij} | \boldsymbol{\gamma}) = \frac{1}{(2\pi)^{\frac{r}{2}} |V_\eta|^{1/2}} \exp\big\{ -\frac{1}{2} \boldsymbol{\eta}'_{ij} V_\eta^{-1} \boldsymbol{\eta}_{ij} \big\}$, and $f(Z_{ij,obs}, Z_{ij,mis} | \boldsymbol{\alpha})$ is given in (11).

We take the joint prior for $\boldsymbol{\theta}$ as

$$\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\mu}) \pi(\boldsymbol{\beta}) \pi(\boldsymbol{\phi}) \pi(\boldsymbol{\gamma}) \pi(\boldsymbol{\sigma}^2) \pi(\sigma_\tau^2) \pi(\sigma_{\tau^*}^2) \pi(\boldsymbol{\alpha}). \tag{13}$$

The detailed specification for each prior on the right hand side of (13) is given as follows. For location parameters, we take $\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \sigma_0^2)$, $k = 1, 2, \ldots, K$, for $\pi(\boldsymbol{\mu})$, $\boldsymbol{\beta}_k \overset{iid}{\sim} \mathcal{N}(\beta_0, \Sigma_{\beta_0})$, $k = 1, 2, \ldots, K$, for $\pi(\boldsymbol{\beta})$, $\boldsymbol{\phi}_k \overset{iid}{\sim} \mathcal{N}(\phi_0, \Sigma_{\phi_0})$, $k = 1, 2, \ldots, K$, for $\pi(\boldsymbol{\phi})$, and $\gamma_l \overset{iid}{\sim} N(\gamma_0, \sigma_{0\gamma}^2)$, $l = 1, 2, \ldots, r$, for $\pi(\boldsymbol{\gamma})$. For the scale parameters, we assume inverse gamma priors as follows: $\sigma_k^2 \overset{iid}{\sim} \mathcal{IG}(a_0, b_0)$, $k = 1, 2, \ldots, K$, for $\pi(\boldsymbol{\sigma}^2)$, $\sigma_\tau^2 \sim \mathcal{IG}(a_1, b_1)$ for $\pi(\sigma_\tau^2)$, and $\sigma_{\tau^*}^2 \sim \mathcal{IG}(a_2, b_2)$ for $\pi(\sigma_{\tau^*}^2)$. For $\pi(\boldsymbol{\alpha})$, we assume $\pi(\boldsymbol{\alpha}) = \prod_{l=1}^{q} \pi(\boldsymbol{\alpha}_l)$, where the prior specification for each $\pi(\boldsymbol{\alpha}_l)$ depends on the form of the one-dimensional conditional distribution for z_{ijl} in (11).

From (12) and (13), the joint posterior distribution of $\boldsymbol{\theta}$, $\boldsymbol{\eta}$, $\boldsymbol{\tau}$, and $\boldsymbol{\tau}^*$ given D_{obs} is thus given by

$$\pi(\boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\tau}, \boldsymbol{\tau}^* | D_{obs}) \propto L(\boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\tau}, \boldsymbol{\tau}^* | D_{obs}) \pi(\boldsymbol{\theta}). \tag{14}$$

Although the analytical evaluation of the posterior distribution in (14) is not possible due to the complexity of the structural equation model as well as the presence of missing responses and covariates, an efficient computational algorithm via Markov chain Monte Carlo (MCMC) sampling can be easily developed. The detailed steps of the MCMC sampling algorithm to sample from $\pi(\boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\tau}, \boldsymbol{\tau}^* | D_{obs})$ is given in the Appendix.

5 Model Assessment

The model developed in Section 3 accounts for both facility effects as well as covariate effects. We refer to this as the saturated model and denote it by $\mathcal{M}_1$. It is of great practical interest to investigate how the inclusion of the two random terms introduced in $\mathcal{M}_1$ contributes to the fit of the data. To do so, we consider three other models, which are sub-models of $\mathcal{M}_1$. Specifically, we will investigate sub-models resulting from excluding facility effects, covariate effects and both effects from model $\mathcal{M}_1$. The forms of these models are given as follows:

$\mathcal{M}_1$ (Facility and covariates effects): $y_{ijk} = \mu_k + \tau_i + \tau_{ik} + \boldsymbol{\beta}_k'\boldsymbol{\omega}_k\boldsymbol{\eta}_{ij} + \boldsymbol{\phi}_k' Z_{ij} + \epsilon_{ijk}$;

$\mathcal{M}_2$ (Facility effect, no covariates): $y_{ijk} = \mu_k + \tau_i + \tau_{ik} + \boldsymbol{\beta}_k'\boldsymbol{\omega}_k\boldsymbol{\eta}_{ij} + \epsilon_{ijk}$;

$\mathcal{M}_3$ (No facility effect, but covariates effect): $y_{ijk} = \mu_k + \boldsymbol{\beta}_k'\boldsymbol{\omega}_k\boldsymbol{\eta}_{ij} + \boldsymbol{\phi}_k' Z_{ij} + \epsilon_{ijk}$; and

$\mathcal{M}_4$ (Neither facility nor covariates effects): $y_{ijk} = \mu_k + \boldsymbol{\beta}_k'\boldsymbol{\omega}_k\boldsymbol{\eta}_{ij} + \epsilon_{ijk}$.

We carry out a formal comparison of these competing SEMs via the Bayesian Deviance Information Criteria (DIC) proposed by Spiegelhalter et al. (2002). The use of DIC for missing data models has been discussed in detail in Celeux et al. (2006). The DIC is defined as follows

$$\mathrm{DIC} = D(\bar{\boldsymbol{\theta}}^*) + 2p_{_D}, \tag{15}$$

where $\boldsymbol{\theta}^*$ is the vector of all model parameters, $D(\boldsymbol{\theta}^*)$ is a deviance function and $\bar{\boldsymbol{\theta}}^* = E[\boldsymbol{\theta}^*|D_{obs}]$ is the posterior mean of $\boldsymbol{\theta}^*$. In (15), $p_{_D}$ is the effective number of model parameters, which is calculated as $p_{_D} = \overline{D(\boldsymbol{\theta}^*)} - D(\bar{\boldsymbol{\theta}}^*)$, where $\overline{D(\boldsymbol{\theta}^*)} = E[D(\boldsymbol{\theta}^*)|D_{obs}]$.

Since the responses y_{ijk}'s are of primary interest in the SEM, we define the deviance function as follows. For model $\mathcal{M}_1$, we treat all facility effects, τ_i and τ_{ik}, and missing covariates, $Z_{ij,mis}$, as parameters. Thus, we define $\boldsymbol{\theta}^* = (\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\sigma}^2, \sigma_\tau^2, \sigma_{\tau^*}^2, \boldsymbol{\tau}, \boldsymbol{\tau}^*, Z_{mis})$ and $D(\boldsymbol{\theta}^*) = -2\log\Big[\prod_{i=1}^{I}\prod_{j=1}^{n_i} L_{ij}(\boldsymbol{\theta}^*)\Big]$, where $L_{ij}(\boldsymbol{\theta}^*|D_{obs}) = \int \Big\{\prod_{k=1}^{K}\Big[f(y_{ijk}|\mu_k, \tau_i, \tau_{ik}, \boldsymbol{\beta}_k, \boldsymbol{\eta}_{ij}, \boldsymbol{\phi}_k, \sigma_k^2, Z_{ij})\Big]^{\delta_{ijk}}\Big\} f(\boldsymbol{\eta}_{ij}|\gamma) d\boldsymbol{\eta}_{ij}$. After some algebra, we obtain

$$\begin{aligned}
& L_{ij}(\boldsymbol{\theta}^*|D_{obs}) \\
= \; & \Big[\prod_{k=1}^{K}\Big(\frac{1}{\sqrt{2\pi\sigma_k^2}}\Big)^{\delta_{ijk}}\Big] \times |V_\eta|^{-1/2}\Big|V_\eta^{-1} + \sum_{k=1}^{K}\frac{\delta_{ijk}}{\sigma_k^2}\boldsymbol{\omega}_k'\boldsymbol{\beta}_k\boldsymbol{\beta}_k'\boldsymbol{\omega}_k\Big|^{-\frac{1}{2}} \\
& \times \exp\Big\{-\frac{1}{2}\sum_{k=1}^{K}\frac{\delta_{ijk}}{\sigma_k^2}\Big(y_{ijk} - \mu_k - \tau_i - \tau_{ik} - \boldsymbol{\phi}_k' Z_{ij}\Big)^2\Big\} \\
& \times \exp\Big\{-\frac{1}{2}\Big(\sum_{k=1}^{K}\frac{\delta_{ijk}}{\sigma_k^2}(y_{ijk} - \mu_k - \tau_i - \tau_{ik} - \boldsymbol{\phi}_k' Z_{ij})\boldsymbol{\beta}_k'\boldsymbol{\omega}_k\Big) \\
& \Big(V_\eta^{-1} + \sum_{k=1}^{K}\frac{\delta_{ijk}}{\sigma_k^2}\boldsymbol{\omega}_k'\boldsymbol{\beta}_k\boldsymbol{\beta}_k'\boldsymbol{\omega}_k\Big)^{-1}\Big(\sum_{k=1}^{K}\frac{\delta_{ijk}}{\sigma_k^2}(y_{ijk} - \mu_k - \tau_i - \tau_{ik} - \boldsymbol{\phi}_k' Z_{ij})\Big)\boldsymbol{\omega}_k'\boldsymbol{\beta}_k\Big\}.
\end{aligned}$$

Thus, the deviance function for model $\mathcal{M}_1$ is

$$
\begin{aligned}
D(\boldsymbol{\theta}^*) &= -2\log\Big[\prod_{i=1}^{I}\prod_{j=1}^{n_i} L_{ij}(\boldsymbol{\theta}^*|D_{obs})\Big] \\
&= \sum_{i=1}^{I}\sum_{j=1}^{n_i}\Big\{\sum_{k=1}^{K}\delta_{ijk}\log(2\pi\sigma_k^2)+\log|V_\eta|+\sum_{k=1}^{K}\frac{\delta_{ijk}}{\sigma_k^2}(y_{ijk}-\mu_k-\tau_i-\tau_{ik}-\boldsymbol{\phi}_k' Z_{ij})^2 \\
&\quad +\log|V_\eta^{-1}+\sum_{k=1}^{K}\frac{\delta_{ijk}}{\sigma_k^2}\boldsymbol{\omega}_k'\boldsymbol{\beta}_k\boldsymbol{\beta}_k'\boldsymbol{\omega}_k| + \Big[\sum_{k}\frac{\delta_{ijk}}{\sigma_k^2}(y_{ijk}-\mu_k-\tau_i-\tau_{ik}-\boldsymbol{\phi}_k' Z_{ij})\boldsymbol{\beta}_k'\boldsymbol{\omega}_k\Big] \\
&\quad \times\big(V_\eta^{-1}+\sum_{k}\frac{\delta_{ijk}}{\sigma_k^2}\boldsymbol{\omega}_k'\boldsymbol{\beta}_k\boldsymbol{\beta}_k'\boldsymbol{\omega}_k\big)^{-1}\Big[\sum_{k=1}^{K}\frac{\delta_{ijk}}{\sigma_k^2}(y_{ijk}-\mu_k-\tau_i-\tau_{ik}-\boldsymbol{\phi}_k' Z_{ij})\boldsymbol{\omega}_k'\boldsymbol{\beta}_k\Big]\Big\}.
\end{aligned}
$$

Following Huang et al. (2005), in $D(\bar{\boldsymbol{\theta}}^*)$, instead of $(E[\boldsymbol{\phi}_k|D_{obs}])'E[Z_{ij}|D_{obs}]$, we compute $E[\boldsymbol{\phi}_k' Z_{ij}|D_{obs}]$ in the presence of missing covariates, which yields a more appropriate dimensional penalty p_D.

Similarly, the deviance function for model $\mathcal{M}_2$ is

$$
\begin{aligned}
D(\boldsymbol{\theta}^*) &= \sum_{i=1}^{I}\sum_{j=1}^{n_i}\Big\{\sum_{k=1}^{K}\delta_{ijk}\log(2\pi\sigma_k^2)+\log|V_\eta|+\sum_{k=1}^{K}\frac{\delta_{ijk}}{\sigma_k^2}(y_{ijk}-\mu_k-\tau_i-\tau_{ik})^2 \\
&\quad +\log|V_\eta^{-1}+\sum_{k=1}^{K}\frac{\delta_{ijk}}{\sigma_k^2}\boldsymbol{\omega}_k'\boldsymbol{\beta}_k\boldsymbol{\beta}_k'\boldsymbol{\omega}_k| + \Big[\sum_{k=1}^{K}\frac{\delta_{ijk}}{\sigma_k^2}(y_{ijk}-\mu_k-\tau_i-\tau_{ik})\boldsymbol{\beta}_k'\boldsymbol{\omega}_k\Big] \\
&\quad \big(V_\eta^{-1}+\sum_{k}\frac{\delta_{ijk}}{\sigma_k^2}\boldsymbol{\omega}_k'\boldsymbol{\beta}_k\boldsymbol{\beta}_k'\boldsymbol{\omega}_k\big)^{-1}\Big[\sum_{k}\frac{\delta_{ijk}}{\sigma_k^2}(y_{ijk}-\mu_k-\tau_i-\tau_{ik})\boldsymbol{\omega}_k'\boldsymbol{\beta}_k\Big]\Big\},
\end{aligned}
$$

where $\boldsymbol{\theta}^* = (\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\sigma}^2,\sigma_\tau^2,\sigma_{\tau^*}^2,\boldsymbol{\tau},\boldsymbol{\tau}^*)$. The deviance function for model $\mathcal{M}_3$ is given by

$$
\begin{aligned}
D(\boldsymbol{\theta}^*) &= \sum_{i=1}^{I}\sum_{j=1}^{n_i}\Big\{\sum_{k=1}^{K}\delta_{ijk}\log(2\pi\sigma_k^2)+\log|V_\eta|+\sum_{k=1}^{K}\frac{\delta_{ijk}}{\sigma_k^2}(y_{ijk}-\mu_k-\boldsymbol{\phi}_k' Z_{ij})^2 \\
&\quad +\log|V_\eta^{-1}+\sum_{k}\frac{\delta_{ijk}}{\sigma_k^2}\boldsymbol{\omega}_k'\boldsymbol{\beta}_k\boldsymbol{\beta}_k'\boldsymbol{\omega}_k| + \Big[\sum_{k=1}^{K}\frac{\delta_{ijk}}{\sigma_k^2}(y_{ijk}-\mu_k-\boldsymbol{\phi}_k' Z_{ij})\boldsymbol{\beta}_k'\boldsymbol{\omega}_k\Big] \\
&\quad \times\big(V_\eta^{-1}+\sum_{k=1}^{K}\frac{\delta_{y_{ijk}}}{\sigma_k^2}\boldsymbol{\omega}_k'\boldsymbol{\beta}_k\boldsymbol{\beta}_k'\boldsymbol{\omega}_k\big)^{-1}\Big[\sum_{k=1}^{K}\frac{\delta_{ijk}}{\sigma_k^2}(y_{ijk}-\mu_k-\boldsymbol{\phi}_k' Z_{ij})\boldsymbol{\omega}_k'\boldsymbol{\beta}_k\Big]\Big\},
\end{aligned}
$$

where $\boldsymbol{\theta}^* = (\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\sigma}^2, Z_{mis})$. Finally, the deviance function for model $\mathcal{M}_4$ is

$$\begin{aligned} D(\boldsymbol{\theta}^*) \quad = \quad & \sum_{i=1}^{I}\sum_{j=1}^{n_i}\Big\{\sum_{k=1}^{K}\delta_{ijk}\log(2\pi\sigma_k^2) + \log|V_\eta| + \sum_{k=1}^{K}\frac{\delta_{ijk}}{\sigma_k^2}(y_{ijk}-\mu_k)^2 \\ & + \log|V_\eta^{-1} + \sum_k \frac{\delta_{y_{ijk}}}{\sigma_k^2}\boldsymbol{\omega}'_k\boldsymbol{\beta}_k\boldsymbol{\beta}'_k\boldsymbol{\omega}_k| + \Big[\sum_k \frac{\delta_{y_{ijk}}}{\sigma_k^2}(y_{ijk}-\mu_k)\boldsymbol{\beta}'_k\boldsymbol{\omega}_k\Big] \\ & \big(V_\eta^{-1} + \sum_{k=1}^{K}\frac{\delta_{ijk}}{\sigma_k^2}\boldsymbol{\omega}'_k\boldsymbol{\beta}_k\boldsymbol{\beta}'_k\boldsymbol{\omega}_k\big)^{-1}\Big[\sum_k \frac{\delta_{y_{ijk}}}{\sigma_k^2}(y_{ijk}-\mu_k)\boldsymbol{\omega}'_k\boldsymbol{\beta}_k\Big]\Big\}, \end{aligned}$$

where $\boldsymbol{\theta}^* = (\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\sigma}^2)$.

The DIC defined above is a Bayesian measure of predictive model performance, decomposed into a measure of fit and a measure of model complexity (p_D). The smaller the value the better the model will predict new observations generated in the same way as the data. The other properties of the DIC can be found in Spiegelhalter et al. (2002). Note that it is important to integrate out all latent variables $\boldsymbol{\eta}_{ij}$ in the deviance calculation as this yields a more appropriate dimensional penalty p_D, since they add a nuisance dimension penalty of the order $(\sum_i n_i) \times r$.

6 Analysis of the AES data

The AES 1997 was an all employees survey that was conducted via self reported questionnaire to evaluate organizational climate and locate intervention points. The main focus was on the three outcome variables investigated via three work place traits. These traits are latent constructs, and thus, observed responses from the questionnaire were identified that were considered as manifestations of the latent constructs. The specific dimension of all the response (outcome + manifest) variables is $K = 3 + 21 = 24$, the number of latent traits is $r = 3$, and the number of facilities is $I = 154$.

The variables, Customer satisfaction, Employee satisfaction and Quality, are the primary responses of interest, and they are theoretically motivated to load on all the three system latent variables, namely, Leadership, Support and Resource. Besides the outcome variables, 21 other manifest variables were identified to be the manifestations of the three latent variables. The loading of these 21 was motivated by a number of reasons. First, we do not load all the the 21 manifest variables on all the three latent variables simply because the clear definition of each latent trait will vanish. Second, we do not load three disjoint subsets of the 21 manifest variables to different latent variables simply because this will make the definition of the latent variables too tight. Thus, we load only two manifest variables singly on their most closely matched latent variable definition, and then load the remaining 15 ($= 21 - 3\times 2$) manifest on all the three latent variables. This is a more moderate stand than the two extremes mentioned before. This strategy also matches the feedback from the field investigators in the VHA that only a few manifest variables can be clearly tagged as effect of the latent traits versus the others. The identification of the three specific pairs of manifest variables were also clearly identified via a preliminary factor analytic EDA. The schematic representation has been laid out in Figure 1.

6.1 Specification of the Model and Priors

The dimensions of the vectors involved are as follows: for the single loading manifest variables $p_k = 1,\ k = 1, 2, \ldots, 6$. For the 15 manifest variables and three outcome variables $p_k = 3,\ k = 7, 8, \ldots, 24$. The corresponding loading matrices are: $\boldsymbol{\omega}_1 = \boldsymbol{\omega}_2 = (1, 0, 0)'$; $\boldsymbol{\omega}_3 = \boldsymbol{\omega}_4 = (0, 1, 0)'$; $\boldsymbol{\omega}_5 = \boldsymbol{\omega}_6 = (0, 0, 1)'$ and $\boldsymbol{\omega}_k = I_3,\ k = 7, 8, \ldots, 24$. In the structural part of the model, the dimension of the latent variable vector is $r = 3$. The proposed measurement model (1) is essentially a confirmatory factor analysis (CFA) model discussed in Bollen (1989). The above specification of loading matrices $\boldsymbol{\omega}_k$'s along with the unity diagonal elements of V_η for the latent vector $\boldsymbol{\eta}_{ij}$ ensures the measurement model (1) identifiable as our resulting CFA model satisfies the t-rule as well as the two-indicator rule of Bollen (1989). Note that under the two-indicator rule discussed in Bollen (1989), one of these $\boldsymbol{\beta}_k$'s is fixed while the variance of $\boldsymbol{\eta}_{ijr^*}$ is set to be free. Equivalently, we set $\boldsymbol{\beta}_k$ to be free but set the variance of $\boldsymbol{\eta}_{ijr^*}$ to be 1.

The other prior distributions have been described in Section 4. We specify the hyperparameters as follows. A $\mathcal{N}(0, 1000)$ or $\mathcal{N}(\mathbf{0}, 1000I)$ prior is used for all location parameters including $\mu_k, \boldsymbol{\beta}_k, \boldsymbol{\phi}_k, \gamma_1, \gamma_2$, and γ_3. For the scale parameters, the hyperparameters are $a_0 = a_1 = a_2 = 1$ and $b_0 = b_1 = b_2 = 0.001$. As mentioned in Section 3, if the response is missing, we do not need to model the missing data mechanism. On the other hand, for missing covariates we can model them via a sequence of one-dimensional conditional distributions. Let $(Z_{ij1}, Z_{ij2}, Z_{ij3})$ correspond to the covariates 'gender', 'age', and 'years in VA'. Since the Z_{ijs}'s are all dichotomized, we specify

$$f(z_{ij1}|\alpha_1) = \alpha_1^{z_{ij1}}(1-\alpha_1)^{1-z_{ij1}},\ f(z_{ij2}|z_{ij1}, \alpha_{21}, \alpha_{22}) = \frac{\exp\{z_{ij2}(\alpha_{21} + \alpha_{22} z_{ij1})\}}{1 + \exp(\alpha_{21} + \alpha_{22} z_{ij1})},$$

$$f(z_{ij3}|z_{ij1}, z_{ij2}, \alpha_{31}, \alpha_{32}, \alpha_{33}) = \frac{\exp\{z_{ij3}(\alpha_{31} + \alpha_{32} z_{ij1} + \alpha_{33} z_{ij2})\}}{1 + \exp(\alpha_{31} + \alpha_{32} z_{ij1} + \alpha_{33} z_{ij2})}.$$

The prior distribution for α_1 is $\text{Beta}(0.001, 0.001)$, and the priors distributions of α_{21}, α_{22}, α_{31}, α_{32}, and α_{33} are all $\mathcal{N}(0, 1000)$ independently.

6.2 Posterior Computation

In all the computations below, we used 20,000 Gibbs samples, after a burn-in of 1000 for each model, to compute all the posterior estimates, including posterior means, posterior standard deviations (SDs), 95% highest posterior density (HPD) intervals and DICs. Codes were written for the FORTRAN 95 compiler, and we used IMSL subroutines with double precision accuracy. The convergence of the Gibbs sampler was checked using several diagnostic procedures as recommended by Cowles and Carlin (1996). The trace plots and auto-correlation plots for $\beta_{24,1}$, γ_1, σ^2_τ, and $\sigma^2_{\tau^*}$ under model $\mathcal{M}_1$ are displayed in Figure 2, which indicate good convergence and mixing for those parameters.

6.3 Model Comparison

We calculated the DICs defined in Section 5 as in Huang et al. (2005). Table 3 shows the DIC values for the four models under consideration.

Model	$\overline{D(\theta)}$	p_D	DIC
$\mathcal{M}_1$	6032315.7	3548.3	6035864.1
$\mathcal{M}_2$	6041745.6	3332.4	6045077.7
$\mathcal{M}_3$	6063755.4	332.6	6064088.5
$\mathcal{M}_4$	6073386.4	110.6	6073496.8

Table 3: DIC Values for Different Models for the *All-Cases* Data

The clear trend in the DIC values suggests that with the *all cases* data, i.e., all the 111,249 cases, $\mathcal{M}_1$ is the best in spite of it having the largest penalty for dimension. This indicates that the inclusion of the facility level effects, the facility-response interaction effects, and the covariates improves the predictive performance of the model. Between the facility effect and the facility-response interaction on one hand ($\mathcal{M}_2$) and the just covariates on the other ($\mathcal{M}_3$), the former has a smaller DIC value, again in spite of the higher penalty for larger dimension of the parameters. The model with the largest DIC among those considered is model $\mathcal{M}_4$ that does not consider either of the effects. It is interesting to note here that in organizational parlance, there is a concept of 'plant effect' that is attributed to the differences between plants, here facilities, and our analysis corroborates this fact. Hence, in spite of not having any covariates to characterize the facilities, the introduction of the random effect terms did capture this heterogeneity.

The penalty due to dimension p_D can be readily appreciated via model $\mathcal{M}_4$ where we have, with η_{ijk}'s integrated out, 6 β coefficients for the single loading manifest variables, 18×3 β coefficients for manifest and outcome variables that load on all the three latent variables, 24 σ^2_k's corresponding to the 24 measurement equations in the SEM, 3 γ's from the structural part of the SEM, and 24 μ's that add up to 111 parameters. From the DIC results, this is 110.6. We also subjected the best model $\mathcal{M}_1$ to the *complete-responses* (CR) data and the *complete-cases* (CC) data (Tables 5 and 6).

In addition, we consider the quantity

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\sum_{k=1}^{K}\delta_{ijk}(y_{ijk}-\hat{y}_{ijk})^2,$$

where $\hat{y}_{ijk}$ is the fitted value of y_{ijk}, which is $E[\mu_{ijk}|D_{obs}]$ and μ_{ijk} is defined in (2), and $N = \sum_{i=1}^{I}\sum_{j=1}^{n_i}\sum_{k=1}^{K}\delta_{ijk}$. The quantity, MSE, measure the absolute goodness-of-fit of a given model. For the AES 1997 data, we obtained that the MSEs are 0.46233,

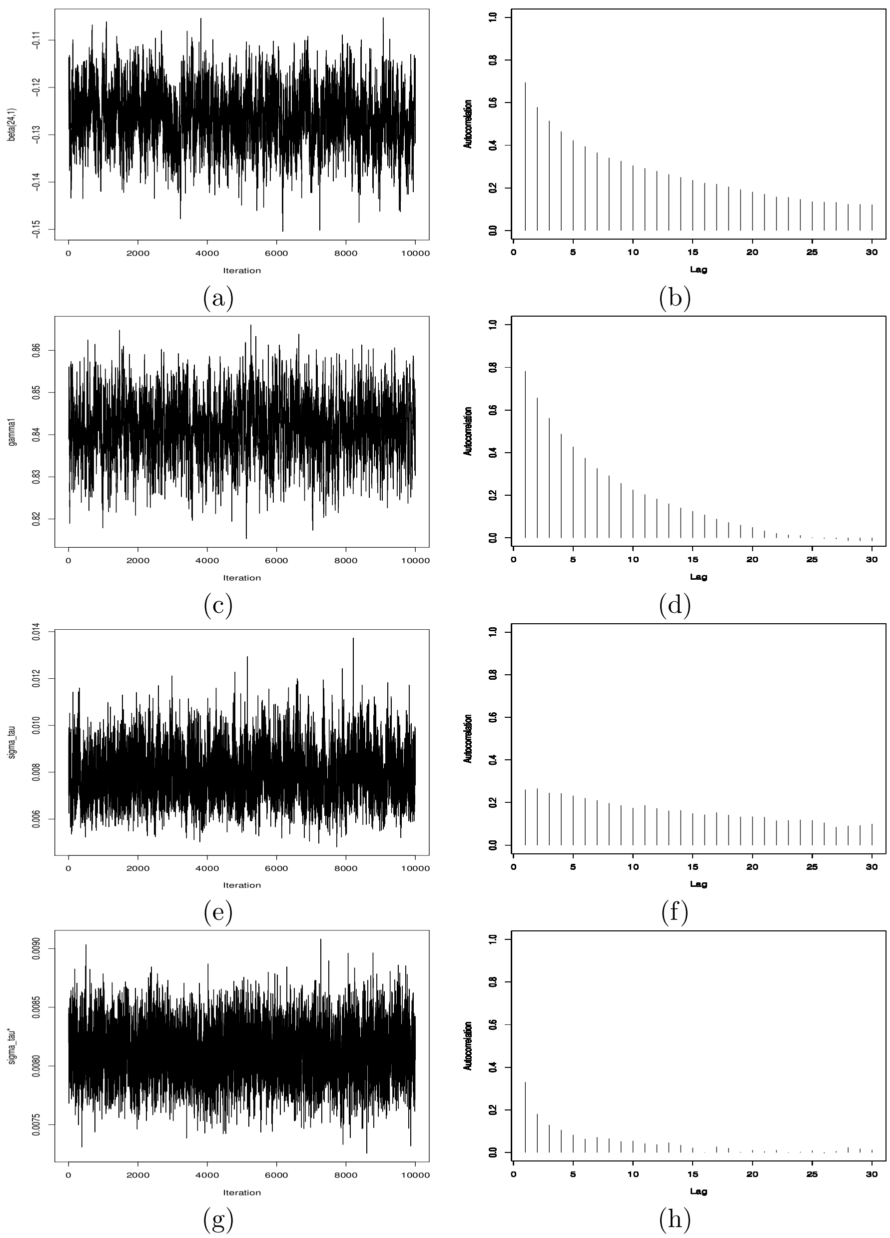

Figure 2: Trace and autocorrelation plots of $\beta_{24,1}$ in (a) and (b), γ_1 in (c) and (d), σ_τ^2 in (e) and (f), and $\sigma_{\tau^*}^2$ in (g) and (h) under the model $\mathcal{M}_1$.

0.46374, 0.46758, and 0.46903 corresponding to models $\mathcal{M}_1$ to $\mathcal{M}_4$, respectively. These four values are indeed quite different as $N = 2,629,304$ is large. These results clearly indicate that $\mathcal{M}_1$ yields the fitted values closest to the observed data among the four models under consideration. It is interesting to mention that these results are very consistent with the DICs shown in Table 3.

6.4 Posterior Estimates

There are a number of ways to describe the association between two variables in the structural equations framework. The simplest is the *direct effect*, that is the direct path from one variable to the other. The other is the *total effect* measure that takes into account the direction and all possible path coefficients between two variables in the system (Bollen (1987)). Using (9), the corresponding *total effect* between y_{ijk} and $\boldsymbol{\eta}_{ij}$ is calculated as follow:

$$
\begin{aligned}
\boldsymbol{\beta}^T_{y_{ijk},\boldsymbol{\eta}_{ij}} &= \boldsymbol{\beta}'\big(I_3 + \Gamma + \Gamma^2 + \dots\big) \\
&= \big(\beta_{k1}, \beta_{k2}, \beta_{k3}\big) I_3 + \big(\beta_{k1}, \beta_{k2}, \beta_{k3}\big) \begin{pmatrix} 0 & \frac{\gamma_1}{a} & 0 \\ 0 & 0 & 0 \\ \frac{a\gamma_3}{b} & \frac{\gamma_2+\gamma_1\gamma_3}{b} & 0 \end{pmatrix} \\
&= \begin{pmatrix} \beta_{k1} + \beta_{k3}\frac{a\gamma_3}{b} \\ \beta_{k1}\frac{\gamma_1}{a} + \beta_{k2} + \beta_{k3}\frac{\gamma_2+\gamma_1\gamma_3}{b} \\ \beta_{k3} \end{pmatrix},
\end{aligned}
$$

where $a = \sqrt{1+\gamma_1^2}, b = \sqrt{1+\gamma_3^2 + (\gamma_2+\gamma_1\gamma_3)^2}$, since $\Gamma^g = \mathbf{0}$ for $g \geq 3$ and $k = 1, 2, \dots, 24$.

We note that if neither is exogenous, then total effect fails to capture the upstream effect due to other connected relationships in the model. Another measure is the so-called *superbeta* measure defined as:

$$
\beta^{sb}_{D^*,I^*} = \text{Superbeta}(D^*, I^*) = \frac{\text{Cov}(D^*, I^*)}{\text{Var}(I^*)}, \tag{16}
$$

where I^* is the cause variable (either endogenous or exogenous) and D^* is the effect variable (endogenous) in the relationship. The *superbeta* measure does take into account not only the direction of the relationship between the variables but also any upstream relationship that the I^* variable may have as part of the structural equation system, and thus measures the overall association between the two variables. Thus, if I^* itself is an endogenous variable, then it is incorporated unless the final variable is exogenous. In fact, the *total effect* is the covariance between the dependent variable D^* and the root error term associated with the I^* variable, taking into account all the path coefficients. If I^* is exogenous, then the *superbeta* equals the corresponding total effect. Below a superbeta calculation is shown between y_{ijk} and η_{ijr^*}, $r^* = 1, 2, 3$:

$$\begin{aligned}\beta^{sb}_{(y_{ijk},\eta_{ij1})} &= \mathrm{Cov}(y_{ijk},\eta_{ij1})/\mathrm{Var}(\eta_{ij1}) = \mathrm{Cov}(\beta_{k1}\eta_{ij1}+\beta_{k2}\eta_{ij2}+\beta_{k3}\eta_{ij3}+\epsilon_{ijk},\eta_{ij1}) \\ &= \beta_{k1}+\beta_{k2}\frac{\gamma_1}{a}+\beta_{k3}\left[\frac{\gamma_1\gamma_2}{ab}+\frac{\gamma_3\ a}{b}\right],\end{aligned}$$

$$\begin{aligned}\beta^{sb}_{(y_{ijk},\eta_{ij2})} &= \mathrm{Cov}(y_{ijk},\eta_{ij2})/\mathrm{Var}(\eta_{ij2}) \\ &= \mathrm{Cov}(\beta_{k1}a\eta_{ij1}+\beta_{k2}\eta_{ij2}+\beta_{k3}\eta_{ij3}+\epsilon_{ijk},\eta_{ij2}) \\ &= \beta_{k1}\frac{\gamma_1}{a}+\beta_{k2}+\beta_{k3}\left[\frac{\gamma_2}{b}+\frac{\gamma_1\gamma_3}{b}\right],\end{aligned}$$

and

$$\begin{aligned}\beta^{sb}_{(y_{ijk},\eta_{ij3})} &= \mathrm{Cov}(y_{ijk},\eta_3)/\mathrm{Var}(\eta_{ij3}) = \mathrm{Cov}(\beta_{k1}\eta_{ij1}+\beta_{k2}\eta_{ij2}+\beta_{k3}\eta_{ij3}+\epsilon_{ijk},\eta_{ij3}) \\ &= \beta_{k1}\left[\frac{\gamma_1\gamma_2}{ab}+\frac{\gamma_3 a}{b}\right]+\beta_{k2}\left[\frac{\gamma_2}{b}+\frac{\gamma_1\gamma_3}{b}\right]+\beta_{k3}\end{aligned}$$

for $k = 1, 2, \ldots, 24$.

In Table 4, we present the posterior estimates of *direct effect*, *total effect*, *superbeta*, μ_k, $\boldsymbol{\phi}_k$, and σ^2_k corresponding to outcome variables y_{ijk}, $k = 22, 23, 24$, $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, and σ^2_τ and $\sigma^2_{\tau^*}$ under model $\mathcal{M}_1$ as it fits the data much better than the others. Also symbols used in Table 4 are as follow — $\boldsymbol{\beta}_k$: direct effect; $\boldsymbol{\beta}^T_k$: total effect; and $\boldsymbol{\beta}^{sb}_k$: superbeta.

There are some interesting findings from the results presented in Table 4. For outcome variable Customer satisfaction, the superbeta measure was strongest with Resources, followed by Support and Leadership. This suggests that when a respondent answers questions regarding their perception of service received, the responses to questions regarding resources in the facility are more important over other intrinsic traits like support in the work place and the leadership qualities in the workplace. For the outcome variable Employee satisfaction, based on the superbeta measure, when the respondent answers questions regarding their experience as employees within the system, Resources in the system are valued highest, followed by Leadership and Support. This reflects that as a worker in the VHA system, an employee gives higher priority to traits like resources available to them and leadership in the workplace over the general support in the workplace. For the outcome variable Quality, based on the superbeta measure, respondents most highly valued the available resources, followed by the support in the system, and finally leadership. The above observations reflect an expected behavior. That is, in answering questions relating to the outcome variables Customer satisfaction and Quality, the respondent's perception is from the perspective of a service receiver, and hence traits like Resources and Support, that affect quality of service, are more important than Leadership. On the other hand, when the respondent answers questions relating to their being an employee, or service provider, he or she values the traits that affect their immediate work environment. Hence questions relating to Resources and Leadership that affect their immediate working conditions are more important over questions relating to Support. Focusing on the total-effect measure (β^T) except for one

Variable	Mean	SD	95% HPD	Variable	Mean	SD	95% HPD
Customer satisfaction							
$\beta_{22,1}$	-0.115	0.0061	(-0.126, -0.103)	$\phi_{22,1}$	0.128	0.0064	(0.116, 0.1408)
$\beta^T_{22,1}$	0.120	0.0046	(0.111, 0.129)	$\phi_{22,2}$	0.021	0.0060	(0.009, 0.0322)
$\beta^{sb}_{22,1}$	0.335	0.0031	(0.329, 0.342)	$\phi_{22,3}$	0.025	0.0067	(0.012, 0.0384)
$\beta_{22,2}$	0.143	0.0052	(0.132, 0.153)	$\phi_{23,1}$	0.189	0.0070	(0.175, 0.2025)
$\beta^T_{22,2}$	0.412	0.0031	(0.406, 0.418)	$\phi_{23,2}$	-0.072	0.0067	(-0.085, -0.0587)
$\beta^{sb}_{22,2}$	0.412	0.0031	(0.406, 0.418)	$\phi_{23,3}$	0.002	0.0073	(-0.012, 0.0163)
$\beta_{22,3}$	0.554	0.0057	(0.543, 0.565)	$\phi_{24,1}$	0.098	0.0060	(0.086, 0.1098)
$\beta^T_{22,3}$	0.554	0.0057	(0.543, 0.565)	$\phi_{24,2}$	-0.056	0.0056	(-0.066, -0.0442)
$\beta^{sb}_{22,3}$	0.568	0.0030	(0.562, 0.574)	$\phi_{24,3}$	0.065	0.0063	(0.053, 0.0773)
Employee satisfaction				μ_{22}	3.398	0.0120	(3.374, 3.421)
$\beta_{23,1}$	0.177	0.0067	(0.164, 0.190)	μ_{23}	3.390	0.0122	(3.367, 3.415)
$\beta^T_{23,1}$	0.352	0.0049	(0.343, 0.362)	μ_{24}	3.934	0.0116	(3.911, 3.956)
$\beta^{sb}_{23,1}$	0.533	0.0033	(0.527, 0.540)	σ^2_{22}	0.572	0.0029	(0.566, 0.578)
$\beta_{23,2}$	0.138	0.0055	(0.127, 0.148)	σ^2_{23}	0.688	0.0032	(0.681, 0.694)
$\beta^T_{23,2}$	0.508	0.0034	(0.501, 0.515)	σ^2_{24}	0.611	0.0030	(0.605, 0.617)
$\beta^{sb}_{23,2}$	0.508	0.0034	(0.501, 0.515)	γ_1	0.841	0.0072	(0.827, 0.855)
$\beta_{23,3}$	0.413	0.0066	(0.400, 0.426)	γ_2	0.486	0.0106	(0.464, 0.506)
$\beta^T_{23,3}$	0.413	0.0066	(0.400, 0.426)	γ_3	0.453	0.0089	(0.435, 0.469)
$\beta^{sb}_{23,3}$	0.613	0.0037	(0.609, 0.620)	σ^2_τ	0.008	0.0011	(0.006, 0.010)
Quality				$\sigma^2_{\tau^*}$	0.008	0.0897	(0.008, 0.009)
$\beta_{24,1}$	-0.126	0.0062	(-0.138, -0.114)	α_{11}	0.401	0.0015	(0.398, 0.404)
$\beta^T_{24,1}$	0.009	0.0047	(-0.001, 0.018)	α_{21}	-0.899	0.0086	(-0.915, -0.882)
$\beta^{sb}_{24,1}$	0.255	0.0030	(0.249, 0.261)	α_{22}	0.244	0.0133	(0.217, 0.269)
$\beta_{24,2}$	0.272	0.0052	(0.262, 0.282)	α_{31}	0.915	0.0097	(0.895, 0.933)
$\beta^T_{24,2}$	0.388	0.0029	(0.382, 0.394)	α_{32}	0.981	0.0175	(0.947, 1.015)
$\beta^{sb}_{24,2}$	0.388	0.0029	(0.382, 0.394)	α_{33}	-0.252	0.0144	(-0.280, -0.224)
$\beta_{24,3}$	0.319	0.0059	(0.307, 0.330)				
$\beta^T_{24,3}$	0.319	0.0059	(0.307, 0.330)				
$\beta^{sb}_{24,3}$	0.405	0.0031	(0.399, 0.412)				

Table 4: Posterior Estimates for the AES 1997 Data for $\mathcal{M}_1$ - All Cases

outcome-latent combination, all other are positive and significant as their 95% HPD intervals do not contain 0. Specifically, the total-effect between outcome variable Quality and Leadership is insignificant as its corresponding 95% HPD interval contains 0. This suggests that ignoring the heterogeneity and/or covariates gives different conclusions based on the total-effect measure.

Also from Table 4, we see that for outcome variable Customer satisfaction, all the 3 covariate coefficients are significantly positive, suggesting that male respondents, senior respondents and longer serving respondents were more satisfied customers. For outcome variable Employee satisfaction — coefficient corresponding to Gender is significantly negative (0 corresponded to female; 1 corresponded to male), suggesting that as an employee, females are more satisfied than males (it is also interesting that more female employees participated in the survey); where as for covariate Age, the coefficient was positive suggesting that as an employee, senior respondents were more satisfied than recent recruits; for covariate Years in VA, the coefficient was not significant. For outcome variable Quality — Age and Years in VA were significantly positive suggesting that senior respondents with longer service rated overall quality of service higher; while for covariate Gender, the coefficient was significantly negative suggesting that female respondents rated overall quality of service higher than male respondents. The variability associated with Employee satisfaction was largest among the 3 outcome variables, suggesting that as service providers respondents were more heterogeneous, which in turn could be reflecting the natural heterogeneity due to clustering of individuals into facilities.

In the structural model part of all the four SEMs, all the coefficients were positive. The coefficient between Leadership and Support was strongest. The posterior estimates of $\text{Corr}(\eta_{ij1}, \eta_{ij3})$ are the strongest among all the 3 pairwise correlations between the elements of $\boldsymbol{\eta}_{ij}$. Specifically, they were 0.647 (0.004), 0.654 (0.005), 0.659 (0.004) and 0.667 (0.004) for the models $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$ and $\mathcal{M}_4$, respectively. The figures in parenthesis are the corresponding posterior standard deviations. As we strip random effect terms and covariate effect terms from model $\mathcal{M}_1$ sequentially, observe that the $\text{Corr}(\eta_{ij1}, \eta_{ij3})$ increases, suggesting that keeping the terms helps to separate the identities of the latent traits. In fact, the corresponding 95% HPD intervals of model $\mathcal{M}_1$ and $\mathcal{M}_4$ are completely disjoint.

In Tables 5 and 6 we present the posterior estimates of the parameters of interest for $\mathcal{M}_1$ for the CR and CC cases. Comparing the estimates of the full model $\mathcal{M}_1$ under the AC with the CR and CC data, we observe the following. First, comparing superbeta's, all the β^{sb} coefficients in CR and CC data sets are consistently higher than the AC data. This indicates that ignoring response with missing data results in an inflation of the associations between the outcome variables and the latent variables. Second, the total effect between Quality and Support in AC was negative, while it is positive for the CR and CC case. While all the superbeta measures were in the same direction. This indicates that total-effect measure can give inconsistent associations even under the same model. Third, in the structural part of the SEM, the coefficients are stronger in the CR and CC case compared to the AC case. Fourth, between the CR and CC cases, barring one, all the coefficients are very close and in the same direction. The one exception being for the coefficient between Employee satisfaction and Years in VA — which is insignificant for the CR data, while it is positive in the CC data.

Finally, we conducted a sensitivity analysis of the posterior estimates of model parameters on the choice of the hyperparameters (b_0, b_1, b_2) under the best model $\mathcal{M}_1$. Specifically, we considered $b_0 = b_1 = b_2 = 0.01$ and $b_0 = b_1 = b_2 = 0.1$. We compared the resulting posterior estimates under these two choices to those under $b_0 = b_1 = b_2 = 0.001$ shown in Table 4. We found that the posterior estimates are quite robust to these choices of (b_0, b_1, b_2). For example, when $b_0 = b_1 = b_2 = 0.01$, the posterior means, the posterior standard deviations, and the 95% HPD intervals are 0.2547, 0.0031, and (0.2487, 0.2606) for $\beta^{sb}_{24,1}$; 0.3878, 0.0030, and (0.3821, 0.3938) for $\beta^{sb}_{24,2}$; 0.4054, 0.0031, and (0.3994, 0.4117) for $\beta^{sb}_{24,3}$; and 0.6112, 0.0030, and (0.6055, 0.6171) for σ^2_{24}. When $b_0 = b_1 = b_2 = 0.1$, the posterior means, the posterior standard deviations, and the 95% HPD intervals are 0.2547, 0.0031, and (0.2488, 0.2608) for $\beta^{sb}_{24,1}$;

Variable	Mean	SD	95% HPD	Variable	Mean	SD	95% HPD
Customer satisfaction							
$\beta_{22,1}$	-0.112	0.0069	(-0.125, -0.098)	$\phi_{22,1}$	0.133	0.0070	(0.119, 0.147)
$\beta^T_{22,1}$	0.128	0.0051	(0.118, 0.138)	$\phi_{22,2}$	0.020	0.0066	(0.007, 0.032)
$\beta^{sb}_{22,1}$	0.347	0.0035	(0.341, 0.354)	$\phi_{22,3}$	0.035	0.0079	(0.019, 0.050)
$\beta_{22,2}$	0.131	0.0058	(0.120, 0.142)	$\phi_{23,1}$	0.193	0.0077	(0.178, 0.208)
$\beta^T_{22,2}$	0.419	0.0034	(0.412, 0.425)	$\phi_{23,2}$	-0.073	0.0072	(-0.087, -0.059)
$\beta^{sb}_{22,2}$	0.419	0.0034	(0.412, 0.425)	$\phi_{23,3}$	0.014	0.0087	(-0.003, 0.031)
$\beta_{22,3}$	0.567	0.0064	(0.554, 0.579)	$\phi_{24,1}$	0.101	0.0066	(0.088, 0.114)
$\beta^T_{22,3}$	0.567	0.0064	(0.554, 0.579)	$\phi_{24,2}$	-0.060	0.0062	(-0.072, 0.048)
$\beta^{sb}_{22,3}$	0.577	0.0032	(0.570, 0.583)	$\phi_{24,3}$	0.077	0.0073	(0.063, 0.091)
Employee satisfaction				μ_{22}	3.387	0.0141	(3.360, 3.415)
$\beta_{23,1}$	0.183	0.0074	(0.169, 0.198)	μ_{23}	3.381	0.0142	(3.354, 3.409)
$\beta^T_{23,1}$	0.360	0.0054	(0.350, 0.371)	μ_{24}	3.927	0.0130	(3.900, 3.952)
$\beta^{sb}_{23,1}$	0.541	0.0037	(0.534, 0.548)	σ^2_{22}	0.565	0.0032	(0.559, 0.572)
$\beta_{23,2}$	0.124	0.0061	(0.112, 0.136)	σ^2_{23}	0.680	0.0034	(0.674, 0.687)
$\beta^T_{23,2}$	0.511	0.0037	(0.504, 0.518)	σ^2_{24}	0.602	0.0032	(0.595, 0.608)
$\beta^{sb}_{23,2}$	0.511	0.0037	(0.504, 0.518)	γ_1	0.869	0.0079	(0.854, 0.885)
$\beta_{23,3}$	0.419	0.0072	(0.405, 0.433)	γ_2	0.512	0.0116	(0.490, 0.535)
$\beta^T_{23,3}$	0.419	0.0072	(0.405, 0.433)	γ_3	0.454	0.0096	(0.435, 0.473)
$\beta^{sb}_{23,3}$	0.619	0.0040	(0.611, 0.627)	σ^2_τ	0.008	0.0011	(0.006, 0.010)
Quality				$\sigma^2_{\tau*}$	0.008	0.0892	(0.008, 0.008)
$\beta_{24,1}$	-0.124	0.0068	(-0.138, -0.111)	α_{11}	0.408	0.0017	(0.405, 0.412)
$\beta^T_{24,1}$	0.015	0.0051	(0.005, 0.025)	α_{21}	-0.878	0.0096	(-0.897, -0.860)
$\beta^{sb}_{24,1}$	0.263	0.0034	(0.256, 0.269)	α_{22}	0.226	0.0147	(0.198, 0.255)
$\beta_{24,2}$	0.259	0.0056	(0.248, 0.270)	α_{31}	0.956	0.0109	(0.935, 0.977)
$\beta^T_{24,2}$	0.388	0.0032	(0.381, 0.394)	α_{32}	1.002	0.0199	(0.962, 1.040)
$\beta^{sb}_{24,2}$	0.388	0.0032	(0.381, 0.394)	α_{33}	-0.198	0.0159	(-0.229, -0.166)
$\beta_{24,3}$	0.329	0.0065	(0.317, 0.342)				
$\beta^T_{24,3}$	0.329	0.0065	(0.317, 0.342)				
$\beta^{sb}_{24,3}$	0.413	0.0034	(0.406, 0.419)				

Table 5: Posterior Estimates for the AES 1997 Data for $\mathcal{M}_1$ - Complete Response (CR)

0.3878, 0.0030, and (0.3820, 0.3937) for $\beta^{sb}_{24,2}$; 0.4054, 0.0031, and (0.3994, 0.4116) for $\beta^{sb}_{24,3}$; and 0.6112, 0.0030, and (0.6053, 0.6168) for σ^2_{24}. Also, the posterior means, the posterior standard deviations, and the 95% HPD intervals for $\sigma^2_{\tau^*}$ are 0.0081, 0.0898, and (0.0077, 0.0086) when $b_0 = b_1 = b_2 = 0.01$; and 0.0082, 0.0902, and (0.0077, 0.0087) when $b_0 = b_1 = b_2 = 0.1$. These values are similar to those given in Table 4. We note that other priors for σ^2_k, σ^2_τ, and $\sigma^2_{\tau^*}$ like those discussed by Gelman (2006) can also be considered. However, we do not expect that there are any substantial changes in the posterior estimates due to large sample size.

7 Discussion

From the investigation of the 4 models we observe that including both facility level random effect terms as well as individual covariates gives the best predictive performance

Variable	Mean	SD	95% HPD	Variable	Mean	SD	95% HPD
Customer satisfaction							
$\beta_{22,1}$	-0.112	0.0070	(-0.125, -0.097)	$\phi_{22,1}$	0.127	0.0071	(0.113, 0.140)
$\beta^T_{22,1}$	0.128	0.0053	(0.117, 0.138)	$\phi_{22,2}$	0.017	0.0066	(0.004, 0.030)
$\beta^{sb}_{22,1}$	0.347	0.0036	(0.340, 0.354)	$\phi_{22,3}$	0.037	0.0077	(0.022, 0.052)
$\beta_{22,2}$	0.131	0.0058	(0.120, 0.143)	$\phi_{23,1}$	0.186	0.0078	(0.170, 0.201)
$\beta^T_{22,2}$	0.418	0.0035	(0.411, 0.425)	$\phi_{23,2}$	-0.076	0.0073	(-0.090, -0.062)
$\beta^{sb}_{22,2}$	0.418	0.0035	(0.411, 0.425)	$\phi_{23,3}$	0.018	0.0085	(0.001, 0.034)
$\beta_{22,3}$	0.566	0.0064	(0.553, 0.578)	$\phi_{24,1}$	0.097	0.0067	(0.084, 0.110)
$\beta^T_{22,3}$	0.566	0.0064	(0.553, 0.578)	$\phi_{24,2}$	-0.062	0.0062	(-0.074, -0.049)
$\beta^{sb}_{22,3}$	0.577	0.0033	(0.570, 0.583)	$\phi_{24,3}$	0.080	0.0074	(0.065, 0.094)
Employee satisfaction				μ_{22}	3.394	0.0121	(3.371, 3.418)
$\beta_{23,1}$	0.189	0.0075	(0.175, 0.204)	μ_{23}	3.391	0.0125	(3.365, 3.415)
$\beta^T_{23,1}$	0.365	0.0056	(0.354, 0.376)	μ_{24}	3.935	0.0118	(3.912, 3.958)
$\beta^{sb}_{23,1}$	0.544	0.0038	(0.537, 0.552)	σ^2_{22}	0.561	0.0032	(0.555, 0.567)
$\beta_{23,2}$	0.123	0.0063	(0.111, 0.135)	σ^2_{23}	0.676	0.0035	(0.669, 0.683)
$\beta^T_{23,2}$	0.512	0.0038	(0.505, 0.520)	σ^2_{24}	0.597	0.0032	(0.591, 0.604)
$\beta^{sb}_{23,2}$	0.512	0.0038	(0.505, 0.520)	γ_1	0.870	0.0082	(0.853, 0.885)
$\beta_{23,3}$	0.417	0.0073	(0.403, 0.432)	γ_2	0.508	0.0118	(0.485, 0.531)
$\beta^T_{23,3}$	0.417	0.0073	(0.403, 0.432)	γ_3	0.452	0.0099	(0.433, 0.472)
$\beta^{sb}_{23,3}$	0.620	0.0041	(0.612, 0.628)	σ^2_{τ}	0.008	0.0011	(0.006, 0.010)
Quality				$\sigma^2_{\tau^*}$	0.008	0.0889	(0.007, 0.008)
$\beta_{24,1}$	-0.121	0.0069	(-0.135, -0.108)				
$\beta^T_{24,1}$	0.016	0.0052	(0.007, 0.027)				
$\beta^{sb}_{24,1}$	0.262	0.0034	(0.256, 0.269)				
$\beta_{24,2}$	0.258	0.0057	(0.247, 0.269)				
$\beta^T_{24,2}$	0.386	0.0033	(0.379, 0.392)				
$\beta^{sb}_{24,2}$	0.386	0.0033	(0.379, 0.392)				
$\beta_{24,3}$	0.326	0.0066	(0.313, 0.339)				
$\beta^T_{24,3}$	0.326	0.0066	(0.313, 0.339)				
$\beta^{sb}_{24,3}$	0.410	0.0035	(0.404, 0.417)				

Table 6: Posterior Estimates for the AES 1997 Data for $\mathcal{M}_1$ - Complete Case (CC)

in terms of the smallest DIC even though it has the largest penalty for dimensionality. Also, including facility random effect terms is more important than including just covariate information, implying that variability within facilities is more pronounced than variability in terms of demographic variables within individuals. This confirms, as is well-known, that facility characteristics differ not only in size but also in the available service resources, and response by individuals belonging to different facilities are bound to reflect that. In fact, including facility random effects captures the natural heterogeneity because of the clustering of individuals into specific facilities. Excluding both the effects gives the worst assessments in terms of DIC, even though it has the smallest penalty for dimension. Also, subjecting the AC, CR and CC data sets to model $\mathcal{M}_1$ revealed that there is indeed no pattern in the missing scheme as posterior estimates are all very close.

From the VHA point of view this may be an important administrative result as this is a concept akin to the 'plant effect' in occupational epidemiology that 'corrects for' unmeasured aspects of an organization – its culture, its work organization, its economic position, and the like – anything that might be the root cause of exposure levels experienced on the shop floor. Researchers have noticed that, when one controls for all reasonable covariates (age, gender, years in occupation, et cetera) at the individual level, the exposure-outcome relationships can still be different in similar companies, and this has often been conveniently grouped into a grab-bag variable called 'plant effect' (plant meaning a building or location in manufacturing, in our case the hospital/facility) (Banker et al. (1993)). This motivates a further investigation into actually examining the components of a 'plant effect' that varies by hospital.

The computational and inferential issues encountered in this paper lead us to delve on some aspects of modeling and inference, and computation. On the inference front, the model assessment criterion DIC used for Bayesian model assessment has many nuances to it – especially for comparing performance of nested models with random effects and latent variables. Because latent variables are unobserved and in fact, computationally can be treated as parameters, it is imperative to take care of issues like whether to construct DIC from the marginal distribution $\int f(y|\boldsymbol{\eta}, \boldsymbol{\theta}) d\boldsymbol{\eta}$ or to compute it from the conditional distribution $f(y|\boldsymbol{\eta}, \boldsymbol{\theta})$. The latter consideration is similar to the one proposed by Huang et al. (2005) in the setting of missing covariates data models. This critical issue for the use of DIC for missing data models is also extensively discussed in Celeux et al. (2006). The second issue is computation. The survey data we have subjected to a Bayesian analysis has a very high dimension, both in terms of sample size as well in terms of variables investigated, consisting of over a hundred thousand individuals with 27 variables. While on the one hand it is a blessing to have such a rich data set to work with, on the other hand implementing Bayesian methods is a huge computational challenge. Thus, it is essential to have the support of very powerful computing machines with very large capacity memory that can handle high dimensional parameter space, especially since we also have latent variables to sample.

Appendix: Computational Development

In this appendix, we only discuss how to sample from the posterior distribution under model $\mathcal{M}_1$ in (14), as the extension to the other models considered in Section 5 is straightforward. To this end, we propose to use the Gibbs sampling algorithm, which requires to sample from the following full conditional distributions in turn:

(i) $[\boldsymbol{\mu}|\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{\tau}, \boldsymbol{\tau}^*, \boldsymbol{\eta}, Z_{mis}, D_{obs}]$;

(ii) $[\boldsymbol{\sigma}^2|\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\tau}, \boldsymbol{\tau}^*, \boldsymbol{\eta}, Z_{mis}, D_{obs}]$;

(iii) $[\boldsymbol{\beta}|\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{\tau}, \boldsymbol{\tau}^*, \boldsymbol{\eta}, Z_{mis}, D_{obs}]$;

(iv) $[\boldsymbol{\phi}|\boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{\tau}, \boldsymbol{\tau}^*, \boldsymbol{\eta}, Z_{mis}, D_{obs}]$;

(v) $[\sigma^2_{\tau}|\boldsymbol{\tau}, D_{obs}]$;

(vi) $[\sigma^2_{\tau^*}|\boldsymbol{\tau}^*, D_{obs}]$;

(vii) $[\boldsymbol{\eta}|\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\gamma},\boldsymbol{\sigma}^2,\boldsymbol{\tau},\boldsymbol{\tau}^*,Z_{mis},D_{obs}]$;

(viii) $[\boldsymbol{\tau}|\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\sigma}^2,\sigma_\tau^2,\boldsymbol{\tau}^*,\boldsymbol{\eta},Z_{mis},D_{obs}]$;

(ix) $[\boldsymbol{\tau}^*|\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\sigma}^2,\sigma_{\tau^*}^2,\boldsymbol{\tau},\boldsymbol{\eta},Z_{mis},D_{obs}]$;

(x) $[\boldsymbol{\gamma}|\boldsymbol{\eta},D_{obs}]$;

(xi) $[\boldsymbol{\alpha}|Z_{mis},D_{obs}]$; and

(xii) $[Z_{mis}|\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\tau},\boldsymbol{\tau}^*,\boldsymbol{\eta},\boldsymbol{\alpha},D_{obs}]$.

We briefly discuss how we sample from each of the above posterior conditional distributions. For (i), given $\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\sigma}^2,\boldsymbol{\tau},\boldsymbol{\tau}^*,\boldsymbol{\eta},Z_{mis},D_{obs}$, the μ_k are conditionally independent and

$$\begin{aligned}&\mu_k|\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\sigma}^2,\boldsymbol{\tau},\boldsymbol{\tau}^*,\boldsymbol{\eta},Z_{mis},D_{obs}\\ \sim\;&\mathcal{N}\left(\frac{\frac{\sum_{i=1}^I\sum_{j=1}^{n_i}\delta_{ijk}(y_{ijk}-\tau_i-\tau_{ik}-\boldsymbol{\beta}_k'\boldsymbol{\omega}_k\boldsymbol{\eta}_{ij}-\boldsymbol{\phi}_k'Z_{ij})}{\sigma_k^2}+\frac{\mu_0}{\sigma_0^2}}{\frac{\sum_{i=1}^I\sum_{j=1}^{n_i}\delta_{ijk}}{\sigma_k^2}+\frac{1}{\sigma_0^2}},\ \frac{1}{\frac{\sum_{i=1}^I\sum_{j=1}^{n_i}\delta_{ijk}}{\sigma_k^2}+\frac{1}{\sigma_0^2}}\right),\end{aligned}$$

for $k=1,2,\ldots,K$. For (ii), again the σ_k^2 are conditionally independent and distributed as

$$\sigma_k^2|\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\tau},\boldsymbol{\tau}^*,\boldsymbol{\eta},Z_{mis},D_{obs}\ \sim\ \mathcal{IG}\Big(\frac{\sum_{i=1}^I\sum_{j=1}^{n_i}\delta_{ijk}}{2}+a_0,\ b_\sigma\Big),$$

where $b_\sigma=\frac{1}{2}\Big\{\sum_{i=1}^I\sum_{j=1}^{n_i}\delta_{ijk}[y_{ijk}-(\mu_k+\tau_i+\tau_{ik}+\boldsymbol{\beta}_k'\boldsymbol{\omega}_k\boldsymbol{\eta}_{ij}+\boldsymbol{\phi}_k'Z_{ij})]^2\Big\}+b_0$. For (iii), we have

$$\boldsymbol{\beta}_k\mid\boldsymbol{\mu},\boldsymbol{\phi},\boldsymbol{\sigma}^2,\boldsymbol{\tau},\boldsymbol{\tau}^*,\boldsymbol{\eta},Z_{mis}\ \sim\ N_{p_k}(B_{\boldsymbol{\beta}_k}^{-1}A_{\boldsymbol{\beta}_k},B_{\boldsymbol{\beta}_k}^{-1})$$

for $k=1,2,\ldots,K$, where $A_{\boldsymbol{\beta}_k}=\frac{1}{\sigma_k^2}\sum_{i=1}^I\sum_{j=1}^{n_i}\boldsymbol{\omega}_k\boldsymbol{\eta}_{ij}\delta_{ijk}(y_{ijk}-\mu_k-\tau_i-\tau_{ik}-\boldsymbol{\phi}_k'\boldsymbol{z}_{ij})+\Sigma_0^{-1}\boldsymbol{\beta}_0$ and $B_{\boldsymbol{\beta}_k}=\frac{1}{\sigma_k^2}\boldsymbol{\omega}_k\Big[\sum_{i=1}^I\sum_{j=1}^{n_i}\boldsymbol{\eta}_{ij}\boldsymbol{\eta}_{ij}'\Big]\boldsymbol{\omega}_k'+\Sigma_0^{-1}$. For (iv),

$$\boldsymbol{\phi}_k\mid\boldsymbol{\beta},\boldsymbol{\sigma}^2,\boldsymbol{\tau},\boldsymbol{\tau}^*,\boldsymbol{\eta},Z_{mis},D_{obs}\ \sim\ N_q(B_{\boldsymbol{\phi}_k}^{-1}A_{\boldsymbol{\phi}_k},B_{\boldsymbol{\phi}_k}^{-1}),$$

where $A_{\boldsymbol{\phi}_k}=\frac{1}{\sigma_k^2}\sum_{i=1}^I\sum_{j=1}^{n_i}(y_{ijk}-\mu_k-\tau_i-\tau_{ik}-\boldsymbol{\beta}_k'\boldsymbol{\omega}_k\boldsymbol{\eta}_{ij})Z_{ij}+\Sigma_{\phi_0}^{-1}\boldsymbol{\phi}_0$ and $B_{\boldsymbol{\phi}_k}=\frac{1}{\sigma_k^2}\sum_{i=1}^I\sum_{j=1}^{n_i}\boldsymbol{z}_{ij}\boldsymbol{z}_{ij}'+\Sigma_{\phi_0}^{-1}$. For (v)

$$\sigma_\tau^2\mid\boldsymbol{\tau},D_{obs}\sim\mathcal{IG}\Big(a_1+\frac{I}{2},\ \frac{1}{2}\sum_{i=1}^I\tau_i^2+b_1\Big),$$

and for (vi)

$$\sigma_{\tau*}^2|\boldsymbol{\tau}^*,D_{obs}\sim\mathcal{IG}\Big(\frac{IK}{2}+a_2,\frac{\sum_i\sum_k\tau_{ik}^2}{2}+b_2\Big).$$

For the latent variables in (vii),

$$\boldsymbol{\eta}_{ij}|\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\gamma},\boldsymbol{\sigma}^2,\boldsymbol{\tau},\boldsymbol{\tau}^*,Z_{mis},D_{obs} \sim N_r(B_{\boldsymbol{\eta}_{ij}}^{-1}A_{\boldsymbol{\eta}_{ij}},B_{\boldsymbol{\eta}_{ij}}^{-1}),$$

where $A_{\boldsymbol{\eta}_{ij}}=\sum_{k=1}^K\Big[\frac{1}{\sigma_k^2}\boldsymbol{\omega}_k'\boldsymbol{\beta}_k\delta_{ijk}(y_{ijk}-\mu_k-\tau_i-\tau_{ik}-\boldsymbol{\phi}_k'\boldsymbol{z}_{ij})\Big]$ and $B_{\boldsymbol{\eta}_{ij}}=\sum_{k=1}^K\Big[\frac{\delta_{ijk}}{\sigma_k^2}\times\boldsymbol{\omega}_k'\boldsymbol{\beta}_k\boldsymbol{\beta}_k'\boldsymbol{\omega}_k\Big]+V_\eta^{-1}$. For (viii), the τ_i are conditionally independent and distributed as

$$\begin{aligned}&\tau_i \mid \boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\sigma}^2,\sigma_\tau^2,\boldsymbol{\tau}^*,\boldsymbol{\eta},Z_{mis},D_{obs}\\ \sim\ & N\Big(\frac{\sum_{j=1}^{n_i}\sum_{k=1}^K\delta_{ijk}(y_{ijk}-\mu_k-\tau_{ik}-\boldsymbol{\beta}_k'\boldsymbol{\omega}_k\boldsymbol{\eta}_{ij}-\boldsymbol{\phi}_k'\boldsymbol{z}_{ij})/\sigma_k^2}{\sum_{j=1}^{n_i}\sum_{k=1}^K\frac{\delta_{ijk}}{\sigma_k^2}+\frac{1}{\sigma_\tau^2}},b_\tau\Big),\end{aligned}$$

where $b_\tau=\Big(\sum_{j=1}^{n_i}\sum_{k=1}^K\frac{\delta_{ijk}}{\sigma_k^2}+\frac{1}{\sigma_\tau^2}\Big)^{-1}$, and for (ix),

$$\begin{aligned}&\tau_{ik} \mid \boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\sigma}^2,\sigma_{\tau^*}^2,\boldsymbol{\tau},\boldsymbol{\eta},Z_{mis},D_{obs}\\ \sim\ & \mathcal{N}\left(\frac{\sum_{j=1}^{n_i}\delta_{ijk}(y_{ijk}-\mu_k-\tau_i-\boldsymbol{\beta}_k'\boldsymbol{\omega}_k\boldsymbol{\eta}_{ij}-\boldsymbol{\phi}_k'\boldsymbol{z}_{ij})}{\sigma_k^2(\frac{\sum_{j=1}^{n_i}\delta_{ijk}}{\sigma_k^2}+\frac{1}{\sigma_{\tau*}^2})},\ \frac{1}{\frac{\sum_{j=1}^{n_i}\delta_{ijk}}{\sigma_k^2}+\frac{1}{\sigma_{\tau*}^2}}\right).\end{aligned}$$

The conditional distributions for (i) to (ix) are either normal or inverse gamma distributions and therefore, sampling from each of these distributions is straightforward.

For (x), we use the localized Metropolis algorithm (Chapter 2 of Chen et al. (2000)) to sample $\boldsymbol{\gamma}$ from $[\boldsymbol{\gamma}|\boldsymbol{\eta},D_{obs}]$. Let

$$\pi^*(\boldsymbol{\gamma}|\boldsymbol{\eta},D_{obs})=\Big[\prod_{i=1}^I\prod_{j=1}^{n_i}|V_\eta|^{-1/2}\exp\Big\{-\frac{1}{2}\boldsymbol{\eta}_{ij}'V_\eta^{-1}\boldsymbol{\eta}_{ij}\Big\}\Big]\pi(\boldsymbol{\gamma}),$$

where $\pi(\boldsymbol{\gamma})$ is the prior for $\boldsymbol{\gamma}$. We compute

$$\widehat{\boldsymbol{\gamma}}=\mathrm{argmax}_{\gamma_1,\gamma_2,\gamma_3}\log\pi^*(\boldsymbol{\gamma}|\boldsymbol{\eta},D_{obs})\ \text{ and }\ \widehat{\Sigma}=\Big[-\frac{\partial^2\log\pi^*(\boldsymbol{\gamma}|\boldsymbol{\eta},D_{obs})}{\partial\gamma_i\partial\gamma_j}\Big|_{\boldsymbol{\gamma}=\widehat{\boldsymbol{\gamma}}}\Big]^{-1}.$$

We use $N(\widehat{\boldsymbol{\gamma}},c^*\widehat{\Sigma})$ as the proposed density for the localized Metropolis algorithm, where c^* is a tuning parameter.

For (xi), $\pi(\boldsymbol{\alpha}|Z_{mis},D_{obs})\propto\Big[\prod_{i=1}^I\prod_{j=1}^{n-i}f(Z_{ij,obs},Z_{ij,mis}|\boldsymbol{\alpha})\Big]\pi(\boldsymbol{\alpha})$. For various covariate distributions specified through a series of one dimensional conditional distributions, sampling $\boldsymbol{\alpha}$ is straightforward. For example, in Section 6, the conditional posterior distribution for α_1 is a beta distribution, which is very easy to sample from; while for $\boldsymbol{\alpha}_2$ and $\boldsymbol{\alpha}_3$, the conditional posterior distributions are log-concave, and hence we can sample these $\boldsymbol{\alpha}_j$'s via the adaptive rejection algorithm of Gilks and Wild (1992). For (xii), given $\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\sigma}^2,\boldsymbol{\tau},\boldsymbol{\tau}^*,\boldsymbol{\eta},\boldsymbol{\alpha}$, and D_{obs}, $Z_{ij,mis}$'s are independent across all i

and j, and the conditional distribution for $Z_{ij,mis}$ is

$$\begin{aligned}&\pi(Z_{ij,mis}|\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\sigma}^2,\boldsymbol{\tau},\boldsymbol{\tau}^*,\boldsymbol{\eta},\boldsymbol{\alpha},D_{obs})\\ \propto\ &\prod_{k=1}^{K}\exp\Big\{-\frac{\delta_{ijk}}{2\sigma_k^2}[y_{ijk}-(\mu_k+\tau_i+\tau_{ik}+\boldsymbol{\beta}_k'\boldsymbol{\omega}_k\boldsymbol{\eta}_{ij}+\boldsymbol{\phi}_k'Z_{ij})]^2\Big\}\\ &\times f(Z_{ij,obs},Z_{ij,mis}|\boldsymbol{\alpha}),\end{aligned}$$

where $f(Z_{ij,obs},Z_{ij,mis}|\boldsymbol{\alpha})$ is given by (11). Thus, sampling Z_{mis} depends on the form of $f(Z_{ij,obs},Z_{ij,mis}|\boldsymbol{\alpha})$. In Section 6, for the AES 1997 data, the conditional posterior distribution of $Z_{ij,mis}$ is simply a multinomial distribution, which is easy to sample from.

References

Ansari, A., Jedidi, K., and Jagpal, S. (2000). "A Hierarchical Bayesian Methodology for Treating Heterogeneity in Structural Equation Models." *Marketing Science*, 19: 328–347. 200

Arminger, G. and Küsters, U. (1988). "Latent Trait Models Indicators of Mixed Measurement Level." In Langeheine, R. and Rost, J. (eds.), *Latent Trait and Latent Class Models*, 51–73. New York: Plenum Press. 200

Asch, S. M., McGlynn, E. A., Hogan, M. M., Hayward, R. A., Shekelle, D. P., Rubenstein, L., Keesey, J., Adams, J., and Kerr, E. A. (2004). "Comparison of Quality of Care for Patients in the Veterans Health Administration and Patients in a National Sample." *Annals of Internal Medicine*, 141: 938–945. 198

Banker, R. D., Potter, G., and Schroeder, R. G. (1993). "Reporting Manufacturing Performance Measures to Workers: An Empirical Study." *Journal of Management Accounting Research*, 5: 33–55. 219

Bentler, P. M. and Wu, E. J. C. (2002). *EQS6 for Windows*. Encino, CA: Multivariate Software. 197

Bollen, K. A. (1987). "Total, Direct, and Indirect Effects in Structural Equation Models." *Sociological Methodology*, 17: 37–69. 213

— (1989). *Structural Equations with Latent Variables*. New York: Wiley. 197, 210

— (2002). "Latent Variables in Psychology and the Social Sciences." *Annual Review of Psychology*, 53: 605–634. 197

Celeux, G., Forbes, F., Robert, C. P., and Titterington, D. M. (2006). "Deviance Information Criteria for Missing Data Models (with Discussion)." *Bayesian Analysis*, 1: 651–674. 207, 219

Chen, M.-H., Shao, Q.-M., and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer-Verlag. 221

Chen, Q., Ibrahim, J. G., Chen, M.-H., and Senchaudhuri, P. (2008). "Theory and Inference for Regression Models with Missing Response and Covariates." *Journal of Multivariate Analysis*, 99: in press. 205

Cowles, C. and Carlin, B. P. (1996). "Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review." *Journal of the American Statistical Association*, 91: 883–904. 210

Dunson, D. B. (2000). "Bayesian Latent Variable Models for Clustered Mixed Outcomes." *Journal of the Royal Statistical Society, Series B*, 62: 355–366. 200

Dunson, D. B. and Perreault, S. D. (2001). "Factor Analytic Models of Clustered Multivariate Data with Information Censoring." *Biometrics*, 57: 302–308. 200

Gelman, A. (2006). "Prior Distributions for Variance Parameters in Hierarchical Models (Comment on Article by Browne and Draper)." *Bayesian Analysis*, 1: 515–534. 217

Gilks, W. R. and Wild, P. (1992). "Adaptive Rejection Sampling for Gibbs Sampling." *Applied Statistics*, 41: 337–348. 221

Greenfield, S. and Kaplan, S. H. (2004). "Creating a Culture of Quality: the Remarkable Transformation of the Department of Veterans Affairs Health Care System." *Annals of Internal Medicine*, 141: 316–318. 198

Hatcher, L. (2000). *A Step-by-Step Approach to Using the SAS System for Factor Analysis and Structural Equation Modeling*. Cary, NC: SAS Institute Inc. 203

Huang, L., Chen, M.-H., and Ibrahim, J. G. (2005). "Bayesian Analysis for Generalized Linear Models with Nonignorably Missing Covariates." *Biometrics*, 61: 767–780. 208, 211, 219

Ibrahim, J. G., Chen, M.-H., Lipsitz, S. R., and Herring, A. H. (2005). "Missing Data Methods in Regression Models." *Journal of the American Statistical Association*, 100: 332–346. 205

Ibrahim, J. G., Lipsitz, S. R., and Chen, M.-H. (1999). "Missing Covariates in Generalized Linear Models When the Missing Data Mechanism Is Nonignorable." *Journal of the Royal Statistical Society, Series B*, 61: 173–190. 205

Jha, A. K., Perlin, B., Kenneth, W. K., and A., D. R. (2003). "Effect of the Transformation of the Veterans Affairs Health Care System on the Quality of Care." *The New England Journal of Medicine*, 348: 2218–2227. 198

Jöreskog, K. G. and Sörbom, D. (1996). *LISREL 8: Structural Equation Modelling with SIMPLIS Command Language*. Hove and London: Scientific Software International. 197

Lee, S. and Shi, J. (2001). "Maximum Likelihood Estimation of Two-Level Latent Variable Models with Mixed Continuous and Polytomous Data." *Biometrics*, 57: 787–794. 200

Lee, S. and Song, X. (2004a). "Maximum Likelihood Analysis of a General Latent Variable Model with Hierarchically Mixed Data." *Biometrics*, 60: 624–636. 200

— (2004b). "Bayesian Model Comparison of Nonlinear Structural Equation Models with Missing Continuous and Ordinal Categorical Data." *British Journal of Mathematical and Statistical Psychology*, 57: 131–150. 200

Lipsitz, S. R. and Ibrahim, J. G. (1996). "A Conditional Model for Incomplete Covariates in Parametric Regression Models." *Biometrika*, 83: 916–922. 205

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. New York: Wiley, 2nd edition. 204

Muthén, B. (2002). "Beyond SEM: General Latent Variable Modeling." *Behaviormetrika*, 29: 81–117. 200

Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2004). "Generalized Multilevel Structural Equation Modelling." *Psychometrika*, 69: 167–190. 200

Rubin, D. B. (1976). "Inference and Missing Data." *Biometrika*, 63: 581–592. 204

Sammel, M. D. and Ryan, L. M. (1996). "Latent Variable Models with Fixed Effects." *Biometrics*, 52: 650–663. 200

Sammel, M. D., Ryan, L. M., and Legler, J. M. (1997). "Latent Variable Models for Mixed Discrete and Continuous Outcomes." *Journal of the Royal Statistical Society, Series B*, 59: 667–678. 200

Sánchez, B. N., Budtz-Jørgensen, E., Ryan, L. M., and Hu, H. (2005). "Structural Equations Models: A Review with Applications to Environment Epidemiology." *Journal of the American Statistical Association*, 100: 1443–1455. 197

Scheines, R., Hoijtink, H., and Boomsma, A. (1999). "Bayesian Estimation and Testing of Structural Equation Models." *Psychometrika*, 64: 37–52. 200

Song, X. and Lee, S. (2004). "Bayesian Analysis of Two-Level Nonlinear Structural Equation Models with Continuous and Polytomous Data." *British Journal of Mathematical and Statistical Psychology*, 57: 29–52. 200

Spiegelhalter, D. J., Best, N. G., Carlin., B. P., and van der Linde, A. (2002). "Bayesian Measures of Model Complexity and Fit." *Journal of the Royal Statistical Society, Series B*, 62: 583–640. 207, 209

Acknowledgments

The authors wish to thank the Editor-in-Chief, the Editor, the Associate Editor, and the two referees for their helpful comments and suggestions which have led to an improvement of this article. The authors also wish to acknowledge and thank the Veterans Health Administration and the National Center for Organizational Development for access to the All Employee Survey (AES) 1997 data and their substantial efforts to insure the quality and integrity of these data.