



Bayesian Estimation of MIRT Models with General and Specific Latent Traits in MATLAB

Yanyan Sheng

Southern Illinois University-Carbondale

Abstract

Multidimensional item response models have been developed to incorporate a general trait and several specific trait dimensions. Depending on the structure of these latent traits, different models can be considered. This paper provides the requisite information and description of software that implement the Gibbs sampling procedures for three such models with a normal ogive form. The software developed is written in the MATLAB package **IRTm2noHA**. The package is flexible enough to allow a user the choice to simulate binary response data with a latent structure involving general and specific traits, specify prior distributions for model parameters, check convergence of the MCMC chain, and obtain Bayesian fit statistics. Illustrative examples are provided to demonstrate and validate the use of the software package.

Keywords: IRT, two-parameter normal ogive models, hierarchical structure, additive structure, MCMC, Gibbs sampling, Gelman-Rubin R, Bayesian DIC, posterior predictive model checks.

1. Introduction

Item response theory (IRT) is a popular approach for solving various measurement problems, and has been found useful in a wide variety of applications (e.g., Bezruckzo 2005; Boomsma, van Duijn, and Snijders 2001; Chang and Reeve 2005; Embretson and Reise 2000; Feske, Kirisci, Tarter, and Plkonis 2007; Imbens 2000; Kolen and Brennan 1995; Lord 1980). For binary response data, IRT provides a collection of models that describe how persons and items interact to yield probabilistic 0/1 responses. The influence of each person and item on the response is modeled by distinct sets of parameters, namely, the person latent trait(s) θ_i and the item's characteristics ξ_j so that

$$P(y = 1) = f(\theta_i, \xi_j),$$

where f can assume a probit or a logit function, giving rise to what is called in the IRT literature a normal ogive model or a logistic model. Unidimensional IRT models assume one θ_i parameter for each person, signifying that each test item measures some facet of the unified latent trait. In some situations, it suffices to assume one latent dimension and use the unidimensional model. However, in many other situations where multiple traits are required in producing the manifest responses to an item, multidimensional IRT (MIRT; Reckase 2009) models need to be considered. A special case of the MIRT model is the multi-unidimensional IRT model (Lee 1995; Sheng 2008b; Sheng and Wikle 2007), applicable in situations where a test involves multiple traits with each item measuring one of them. The items measuring a specific latent trait can be grouped together to form a subtest, and hence the test can be viewed as one consisting of multiple such unidimensional subtests. In such situations, it is also reasonable to assume that each subtest measures some facet of an overall or general latent trait dimension and hence to consider MIRT models incorporating both the general trait and traits specific for individual subtests so that separate inferences can be made about persons for each of them. Here, the general trait signifies the unified latent dimension that involves cognitive process common for all test items, whereas the specific trait involves cognitive process needed for answering items in an individual subtest that may be related to the general trait or other specific traits. Different such MIRT models have been developed in the literature so that the general trait may be underlying the specific traits to form a hierarchical structure (Sheng and Wikle 2008), or the general and specific traits form an additive structure, i.e., each item measures a general trait and a specific trait directly (Sheng and Wikle 2009). These two types of models make different assumptions on the underlying continuous dimensions. Their latent structures can be comparable to that of second-order factor models (Schmid and Leiman 1957) and that of bifactor models (Holzinger and Swineford 1937, though bifactor models assume that the general and specific factors are orthogonal), respectively, which have been well studied in the factor analytic framework. Recent research on factor analysis indicates that the second-order and bifactor models, though with similar interpretations, are not mathematically equivalent (Gustafsson and Balke 1993; McDonald 1999; Mulaik and Quartetti 1997; Yung, Thissen, and McLeod 1999). In particular, Yung *et al.* (1999) have demonstrated that the former is actually hierarchically nested within the latter (see also Rindskopf and Rose 1988). This finding further sheds light on the relation between the two MIRT models considered in this paper.

Parameter estimation offers the basis for the theoretical advantages of IRT, and hence has been a major concern in the application of IRT models. As the influence of items and persons on the responses is modeled by distinct sets of parameters, simultaneous estimation of these parameters results in statistical complexities in the estimation task, which have made estimation procedure a primary focus of psychometric research over decades (Birnbaum 1969; Bock and Aitkin 1981; Molenaar 1995). Recent attention is focused on the fully Bayesian approach using Markov chain Monte Carlo (MCMC) simulation techniques, which are extremely general and flexible and have proved useful in practically all aspects of Bayesian inferences, such as parameter estimation or model comparisons. One of the simplest MCMC algorithms is Gibbs sampling (Casella and George 1992; Geman and Geman 1984). The method is straightforward to implement when each full conditional distribution associated with a particular multivariate posterior distribution is a known distribution that is easy to sample. Gibbs sampling has been applied to common unidimensional models (Albert 1992; Sahu 2002) and two-parameter normal ogive (2PNO) multi-unidimensional models (Lee 1995) using the data augmentation idea

of Tanner and Wong (1987). Sheng and Wikle (2008, 2009) further extended their approaches and developed the Gibbs sampling procedures for the MIRT models with a general trait and several specific trait dimensions. The model with an additive structure, in particular, has been found to provide a better description to the data in various situations compared with the other IRT models (Sheng and Wikle 2009).

This paper provides a MATLAB (The MathWorks, Inc. 2007) package that implements Gibbs sampling for the 2PNO MIRT models incorporating a general and several specific traits with the option of specifying uniform or conjugate priors for item parameters. One may note that different R packages have been developed to estimate dichotomous and/or polytomous IRT models (see e.g., Anderson, Li, and Vermunt 2007; Johnson 2007; Mair and Hatzinger 2007; Rizopoulos 2006). These packages focus on the maximum likelihood estimation of either Rasch models or models with one latent trait, and shall be differentiated from the fully Bayesian approach and/or the multidimensional dichotomous IRT models considered in this study. The remainder of the paper is organized as follows. Section 2 reviews the models and Section 3 briefly describes the MCMC algorithms implemented in the package **IRTN2noHA**. In Section 4, a brief illustration is given of Bayesian model choice or checking technique for testing the adequacy of a model. The package **IRTN2noHA** is introduced in Section 5, where a description is given of common input and output variables. In Section 6, illustrative examples are provided to demonstrate the use of the source code. Finally, a few summary remarks are given in Section 7.

This paper only considers normal ogive models given that more complicated MCMC procedures have to be adopted for the logistic form of these models (e.g., Patz and Junker 1999a,b) and that the logistic and normal ogive forms of the IRT model are essentially indistinguishable in model fit or parameter estimates given proper scaling (Birnbaum 1968; Embretson and Reise 2000).

2. IRT models with general and specific traits

Suppose a K -item test consists of m subtests, each containing k_v dichotomous (0/1) items, where $v = 1, 2, \dots, m$. Let y_{vij} denote the i th person's response to the j th item in the v th subtest, where $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, k_v$ ¹. With a probit link, the probability of a correct response can be defined as

$$P(y_{vij} = 1 | \theta_{vi}, \alpha_{vj}, \beta_{vj}) = \Phi(\alpha_{vj}\theta_{vi} - \beta_{vj}) = \int_{-\infty}^{\alpha_{vj}\theta_{vi} - \beta_{vj}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt, \quad (1)$$

where θ_{vi} is a scalar person trait parameter specific for the v th latent dimension, α_{vj} is a positive scalar slope parameter representing the item discrimination in the v th dimension, and β_{vj} is a scalar intercept parameter that is related to the location in the v th dimension where the item provides maximum information. This probability function takes the form of that of the multi-unidimensional model (see Sheng and Wikle 2007; Sheng 2008b) where m specific latent traits, θ_{vi} , are considered. We may incorporate an additional general latent

¹For conventional IRT models, it is common to use the y_{ij} notation to denote person i 's response to item j , where $i = 1, \dots, n$ and $j = 1, \dots, K$. The subscript v is introduced here because a different trait dimension, θ_v , specific for each subtest is assumed, and each item has a nonzero discriminating power with respect to θ_v . Essentially, y_{vij} can be thought of as dividing the entire $n \times K$ matrix of y_{ij} into m submatrices, with each being of size $n \times k_v$.

dimension by assuming a hierarchical structure so that each θ_{vi} either (a) is a linear function of the general trait parameter θ_{0i} ,

$$\theta_{vi} \sim N(\delta_v \theta_{0i}, 1), \quad (2)$$

or (b) linearly combines to form the general trait parameter,

$$\theta_{0i} \sim N\left(\sum_v \lambda_v \theta_{vi}, 1\right) \quad (3)$$

(Sheng and Wikle 2008), where δ_v and λ_v are positive coefficients relating each θ_{vi} with θ_{0i} . For simplicity, the two formulations are referred to as hierarchical model 1 and hierarchical model 2, respectively.

Alternatively, the general and specific latent dimensions can assume an additive structure so the probability function is defined as

$$P(y_{vij} = 1 | \theta_{0i}, \theta_{vi}, \alpha_{0vj}, \alpha_{vj}, \beta_{vj}) = \Phi(\alpha_{0vj} \theta_{0i} + \alpha_{vj} \theta_{vi} - \beta_{vj}) \quad (4)$$

(Sheng and Wikle 2009), where α_{0vj} is a positive scalar slope parameter associated with θ_{0i} . Hence, a correct response is assumed to be determined directly by two latent traits – a general one and a specific one. Due to the additive structure of the latent traits, this model is referred to as the additive model in this paper. As Figure 1 shows, it differs from the two hierarchical models in the effect of the general trait on item responses. One may note the similarity of this model with the original bifactor model proposed by Holzinger and Swineford (1937), where each item has a nonzero loading on a primary factor and a second loading on only one of several secondary factors. However, they differ in several aspects. First, the additive model is for binary response data and is hence essentially similar to the bifactor full-information factor model presented by Gibbons and Hedeker (1992); see also Li, Bolt, and Fu (2006). Second, as with other IRT models, the additive model uses “difficulties” (which are related to intercepts in our discussion) and “discriminations” instead of “thresholds” and “loadings” as with factor models. Third, in the additive model, it is assumed that the slope is positive, whereas in the bifactor model, the threshold can assume negative values. Finally, the bifactor model constrains the correlation between the general and specific dimensions to be zero, whereas the additive model does not, though it is pointed out later in the paper that large correlations tend to result in high collinearity, which creates problem in parameter estimation. Since the additive model specified in (4) is more complex than the conventional IRT models or even the hierarchical models, certain constraints need to be adopted to help identification. See Section 3.3 for a detailed illustration of the model specification.

3. MCMC algorithms

Gibbs sampling is one of the simplest MCMC algorithms that can obtain item and person parameter estimates simultaneously. The method is particularly straightforward when each full conditional distribution can be obtained in closed form. Its general underlying strategy is to iteratively sample item and person parameters from their respective posterior distributions, conditional on the sampled values of all other person and item parameters. This iterative process continues for a sufficient number of samples after the posterior distributions reach stationarity (a phase known as burn-in).

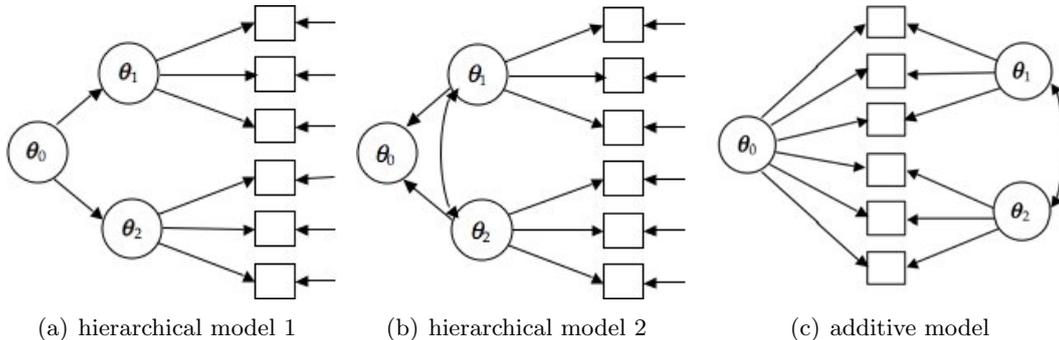


Figure 1: Graphical representations of the three MIRT models incorporating general and specific latent traits. (Circles represent latent traits and squares represent observed item responses.)

As with standard Monte Carlo, the posterior means of all the samples collected after the burn-in stage are considered as estimates of the true model parameters. Similarly, their posterior standard deviations are used to describe the statistical uncertainty. However, Monte Carlo standard errors cannot be calculated using the sample standard deviations because subsequent samples in each Markov chain are autocorrelated (e.g., [Patz and Junker 1999b](#)). Among the standard methods for estimating them ([Ripley 1987](#)), batching is said to be a crude but effective method ([Verdinelli and Wasserman 1995](#)) and hence is considered in this paper. Here, with a long chain of samples being separated into contiguous batches of equal length, the Monte Carlo standard error for each parameter is then estimated to be the standard deviation of these batch means. The Monte Carlo standard error of the estimate is hence a ratio of the Monte Carlo standard error and the square root of the number of batches. More sophisticated methods for estimating standard errors can be found in [Gelman and Rubin \(1992\)](#).

The Gibbs sampler for each MIRT model involving both general and specific latent traits is now described.

3.1. Gibbs sampling for the hierarchical model 1

To implement Gibbs sampling for the model defined in (1) with $\theta_{vi} \sim N(\delta_v \theta_{0i}, 1)$, an augmented continuous variable Z is introduced so that $Z_{vij} \sim N(\eta_{vij}, 1)$ ([Albert 1992](#); [Sheng and Wikle 2008](#); [Tanner and Wong 1987](#)), where $\eta_{vij} = \alpha_{vj} \theta_{vi} - \beta_{vj}$. With prior distributions assumed for θ_{0i} , ξ_{vj} , where $\xi_{vj} = (\alpha_{vj}, \beta_{vj})'$, and δ_v , the joint posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\xi}, \mathbf{Z}, \boldsymbol{\theta}_0, \boldsymbol{\delta})$ is then

$$p(\boldsymbol{\theta}, \boldsymbol{\xi}, \mathbf{Z}, \boldsymbol{\theta}_0, \boldsymbol{\delta}) \propto f(\mathbf{y}|\mathbf{Z})p(\mathbf{Z}|\boldsymbol{\theta}, \boldsymbol{\xi})p(\boldsymbol{\xi})p(\boldsymbol{\theta}|\boldsymbol{\theta}_0, \boldsymbol{\delta})p(\boldsymbol{\theta}_0)p(\boldsymbol{\delta}), \quad (5)$$

where $\boldsymbol{\theta}_0 = (\theta_{01}, \dots, \theta_{0n})'$, $\boldsymbol{\delta} = (\delta_1, \dots, \delta_m)'$, and

$$f(\mathbf{y}|\mathbf{Z}) = \prod_{v=1}^m \prod_{i=1}^n \prod_{j=1}^{k_v} p_{vij}^{y_{vij}} (1 - p_{vij})^{1 - y_{vij}} \quad (6)$$

is the likelihood function with p_{vij} being as defined in (1).

The implementation of the Gibbs sampler thus involves five of the sampling processes, namely, a sampling of the augmented Z parameters from

$$Z_{vij}|\bullet \sim \begin{cases} N_{(0,\infty)}(\eta_{vij}, 1), & \text{if } y_{vij} = 1 \\ N_{(-\infty,0)}(\eta_{vij}, 1), & \text{if } y_{vij} = 0 \end{cases}; \quad (7)$$

a sampling of person specific trait parameters θ from

$$\theta_i|\bullet \sim N_m((\mathbf{A}'\mathbf{A} + \mathbf{I})^{-1}(\delta\theta_{0i} + \mathbf{A}'\mathbf{B}), (\mathbf{A}'\mathbf{A} + \mathbf{I})^{-1}), \quad (8)$$

where $\theta_i = (\theta_{1i}, \dots, \theta_{mi})'$, $\mathbf{A} = \begin{bmatrix} \alpha_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \alpha_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \alpha_m \end{bmatrix}_{K \times m}$ and $\mathbf{B} = \begin{bmatrix} \mathbf{Z}_{1i} + \beta_1 \\ \mathbf{Z}_{2i} + \beta_2 \\ \vdots \\ \mathbf{Z}_{mi} + \beta_m \end{bmatrix}_{K \times 1}$, in

which $\alpha_v = (\alpha_{v1}, \dots, \alpha_{vk_v})'$, $\mathbf{Z}_{vi} = (Z_{vi1}, \dots, Z_{vik_v})'$, $\beta_v = (\beta_{v1}, \dots, \beta_{vk_v})'$; a sampling of person general trait parameters θ_0 from

$$\theta_{0i}|\bullet \sim N((\delta'\delta + 1)^{-1}(\delta'\theta_i), (\delta'\delta + 1)^{-1}); \quad (9)$$

a sampling of the item parameters ξ from

$$\xi_{vj}|\bullet \sim N_2((\mathbf{x}'_v\mathbf{x}_v)^{-1}\mathbf{x}'_v\mathbf{Z}_{vj}, (\mathbf{x}'_v\mathbf{x}_v)^{-1})I(\alpha_{vj} > 0), \quad (10)$$

where $\mathbf{x}_v = [\theta_v, -1]$, assuming noninformative uniform priors $\alpha_{vj} > 0$ and $p(\beta_{vj}) \propto 1$, or from

$$\xi_{vj}|\bullet \sim N_2((\mathbf{x}'_v\mathbf{x}_v + \Sigma_{\xi_v}^{-1})^{-1}(\mathbf{x}'_v\mathbf{Z}_{vj} + \Sigma_{\xi_v}^{-1}\boldsymbol{\mu}_{\xi_v}), (\mathbf{x}'_v\mathbf{x}_v + \Sigma_{\xi_v}^{-1})^{-1})I(\alpha_{vj} > 0), \quad (11)$$

where $\boldsymbol{\mu}_{\xi_v} = (\mu_{\alpha_v}, \mu_{\beta_v})'$ and $\Sigma_{\xi_v} = \text{diag}(\sigma_{\alpha_v}^2, \sigma_{\beta_v}^2)$, assuming conjugate normal priors $\alpha_{vj} \sim N_{(0,\infty)}(\mu_{\alpha_v}, \sigma_{\alpha_v}^2)$, $\beta_{vj} \sim N(\mu_{\beta_v}, \sigma_{\beta_v}^2)$; and a sampling of the hyperparameters δ from

$$\delta|\bullet \sim N_m((\theta'_0\theta_0\mathbf{I})^{-1}\theta'\theta_0, (\theta'_0\theta_0\mathbf{I})^{-1})I(\delta > 0) \quad (12)$$

assuming noninformative uniform priors $\delta_v > 0$, or from

$$\delta|\bullet \sim N_m((\theta'_0\theta_0\mathbf{I} + \Sigma_{\delta}^{-1})^{-1}(\theta'\theta_0 + \Sigma_{\delta}^{-1}\boldsymbol{\mu}_{\delta}), (\theta'_0\theta_0\mathbf{I} + \Sigma_{\delta}^{-1})^{-1})I(\delta > 0), \quad (13)$$

where $\boldsymbol{\mu}_{\delta} = (\mu_{\delta_1}, \dots, \mu_{\delta_m})'$ and $\Sigma_{\delta} = \text{diag}(\sigma_{\delta_1}^2, \dots, \sigma_{\delta_m}^2)$, assuming conjugate normal priors $\delta_v \sim N_{(0,\infty)}(\mu_{\delta_v}, \sigma_{\delta_v}^2)$.

The location and scale parameters for θ_{vi} are commonly assumed to be 0 and 1, respectively, to ensure unique scaling and to resolve a particular identification problem for such models (see e.g., [Albert 1992](#); [Lee 1995](#)). Given the model specification, θ_{vi} are centered around $\delta_v\theta_{0i}$, and hence need to be normalized to have a mean of 0 and a standard deviation of 1. To retain a common scale for all parameters, α_{vj} and β_{vj} are rescaled as well following [Bafumi, Gelman, Park, and Kaplan \(2005\)](#). This transformation further allows computation to be more efficient ([Gelman, Carlin, Stern, and Rubin 2004](#)).

3.2. Gibbs sampling for the hierarchical model 2

Assuming $\theta_{0i} \sim N(\sum_v \lambda_v \theta_{vi}, 1)$, the hierarchical model 2 calls for a slightly different Gibbs sampling procedure. In particular, we can assume a multivariate normal prior density for

person specific trait parameters $\boldsymbol{\theta}_i$ so that $\boldsymbol{\theta}_i \sim N_m(\mathbf{0}, \mathbf{P})$, where \mathbf{P} is a covariance matrix with the variances being fixed at 1. It follows that the off-diagonal element of \mathbf{P} is the correlation ρ_{st} between θ_{si} and θ_{ti} , $s \neq t$, $i = 1, \dots, m$. It is noted that the proper multivariate normal prior for θ_{vi} with their location and scale parameters specified to be 0 and 1 ensures unique scaling and hence is essential in resolving a particular identification problem for this model (see e.g. Lee 1995, for a description of the problem). Moreover, introduce an unconstrained covariance matrix $\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma} = [\sigma_{vv'}]_{m \times m}$, so that the correlation matrix \mathbf{P} can be readily transformed from $\boldsymbol{\Sigma}$ using

$$\rho_{st} = \frac{\sigma_{st}}{\sqrt{\sigma_{ss}\sigma_{tt}}}, \quad s \neq t \quad (14)$$

(the same approach was adopted for the multi-unidimensional model; see Lee 1995; Sheng 2008b). A noninformative prior can be assumed for $\boldsymbol{\Sigma}$ so that $p(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{m+1}{2}}$ (Lee 1995).

Hence, with prior distributions assumed for $\boldsymbol{\xi}_{vj}$ and $\boldsymbol{\lambda}$, where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)'$, the joint posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\xi}, \mathbf{Z}, \boldsymbol{\theta}_0, \boldsymbol{\lambda}, \boldsymbol{\Sigma})$ is

$$p(\boldsymbol{\theta}, \boldsymbol{\xi}, \mathbf{Z}, \boldsymbol{\theta}_0, \boldsymbol{\lambda}, \boldsymbol{\Sigma} | \mathbf{y}) \propto f(\mathbf{y} | \mathbf{Z}) p(\mathbf{Z} | \boldsymbol{\theta}, \boldsymbol{\xi}) p(\boldsymbol{\xi}) p(\boldsymbol{\theta}_0 | \boldsymbol{\theta}, \boldsymbol{\lambda}) p(\boldsymbol{\theta} | \mathbf{P}) p(\boldsymbol{\lambda}) p(\boldsymbol{\Sigma}), \quad (15)$$

where $f(\mathbf{y} | \mathbf{Z})$ is as defined in (6).

The implementation of the Gibbs sampler thus involves six processes, namely, a sampling of the augmented Z parameters from (7); a sampling of the item parameters $\boldsymbol{\xi}$ from (10) assuming uniform priors or from (11) assuming conjugate normal priors; a sampling of person specific trait parameters $\boldsymbol{\theta}$ from

$$\boldsymbol{\theta}_i | \bullet \sim N_m((\mathbf{P}^{-1} + \boldsymbol{\lambda}\boldsymbol{\lambda}' + \mathbf{A}'\mathbf{A})^{-1}(\boldsymbol{\lambda}\theta_{0i} + \mathbf{A}'\mathbf{B}), (\mathbf{P}^{-1} + \boldsymbol{\lambda}\boldsymbol{\lambda}' + \mathbf{A}'\mathbf{A})^{-1}), \quad (16)$$

where \mathbf{A} and \mathbf{B} are as defined in (8); a sampling of the person general trait parameters $\boldsymbol{\theta}_0$ from (3); a sampling of the hyperparameters $\boldsymbol{\lambda}$ from

$$\boldsymbol{\lambda} | \bullet \sim N_m((\boldsymbol{\theta}'\boldsymbol{\theta})^{-1}\boldsymbol{\theta}'\boldsymbol{\theta}_0, (\boldsymbol{\theta}'\boldsymbol{\theta})^{-1}) I(\boldsymbol{\lambda} > 0) \quad (17)$$

assuming noninformative uniform priors $\lambda_v > 0$, or from

$$\boldsymbol{\lambda} | \bullet \sim N_m((\boldsymbol{\theta}'\boldsymbol{\theta} + \boldsymbol{\Sigma}_\lambda^{-1})^{-1}(\boldsymbol{\theta}'\boldsymbol{\theta}_0 + \boldsymbol{\Sigma}_\lambda^{-1}\boldsymbol{\mu}_\lambda), (\boldsymbol{\theta}'\boldsymbol{\theta} + \boldsymbol{\Sigma}_\lambda^{-1})^{-1}) I(\boldsymbol{\lambda} > 0), \quad (18)$$

where $\boldsymbol{\mu}_\lambda = (\mu_{\lambda_1}, \dots, \mu_{\lambda_m})'$ and $\boldsymbol{\Sigma}_\lambda = \text{diag}(\sigma_{\lambda_1}^2, \dots, \sigma_{\lambda_m}^2)$, assuming conjugate normal priors $\lambda_v \sim N_{(0, \infty)}(\mu_{\lambda_v}, \sigma_{\lambda_v}^2)$; a sampling of the unconstrained covariance matrix $\boldsymbol{\Sigma}$ from

$$\boldsymbol{\Sigma} | \bullet \sim W^{-1}(\mathbf{S}^{-1}, n) \quad (19)$$

(an inverse Wishart distribution), where $\mathbf{S} = \sum_{i=1}^n (\mathbf{C}\boldsymbol{\theta}_i)(\mathbf{C}\boldsymbol{\theta}_i)'$, in which

$$\mathbf{C} = \text{diag} \left(\left(\prod_{j=1}^{k_1} \alpha_{1j} \right)^{1/k_1}, \dots, \left(\prod_{j=1}^{k_m} \alpha_{mj} \right)^{1/k_m} \right); \text{ and a transformation of } \boldsymbol{\Sigma} \text{ to } \mathbf{P} \text{ using (14).}$$

3.3. Gibbs sampling for the additive model

For the additive model, the Gibbs sampler again involves the introduction of an augmented variable Z so that $Z_{vij} \sim N(\eta_{vij}, 1)$, where $\eta_{vij} = \alpha_{0vj}\theta_{0i} + \alpha_{vj}\theta_{vi} - \beta_{vj}$. Assume a multivariate

normal prior distribution for person general and specific trait parameters $\boldsymbol{\theta}_i$, where $\boldsymbol{\theta}_i = (\theta_{0i}, \theta_{1i}, \dots, \theta_{mi})'$, so that $\boldsymbol{\theta}_i \sim N_{m+1}(\mathbf{0}, \mathbf{P})$. Here, \mathbf{P} is a constrained covariance matrix, with the diagonal element being 1 and the off-diagonal element being the correlation ρ_{st} between θ_{si} and θ_{ti} , $s \neq t$, $i = 0, \dots, m$. It is noted again that the location and scale parameters for person parameters are specified to be 0 and 1, which ensures unique scaling and helps resolve the model identification problem. Similar to the procedure described in Section 3.2, introduce an unconstrained covariance matrix $\boldsymbol{\Sigma} = [\sigma_{vv'}]_{(m+1) \times (m+1)}$ so that \mathbf{P} can be transformed from $\boldsymbol{\Sigma}$ using (14), and assume $p(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{m+2}{2}}$. Then, with prior distributions assumed for $\boldsymbol{\xi}_{vj}$, where $\boldsymbol{\xi}_{vj} = (\alpha_{0vj}, \alpha_{vj}, \beta_{vj})'$, the joint posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\xi}, \mathbf{Z}, \boldsymbol{\Sigma})$ is

$$p(\boldsymbol{\theta}, \boldsymbol{\xi}, \mathbf{Z}, \boldsymbol{\Sigma} | \mathbf{y}) \propto f(\mathbf{y} | \mathbf{Z}) p(\mathbf{Z} | \boldsymbol{\theta}, \boldsymbol{\xi}) p(\boldsymbol{\xi}) p(\boldsymbol{\theta} | \mathbf{P}) p(\boldsymbol{\Sigma}), \quad (20)$$

where $f(\mathbf{y} | \mathbf{Z}) = \prod_v \prod_i \prod_j p_{vij}^{y_{vij}} (1 - p_{vij})^{1 - y_{vij}}$ is the likelihood function with p_{vij} being as defined in (4).

The Gibbs sampler hence involves five processes: a sampling of the augmented Z parameters from (7), where $\eta_{vij} = \alpha_{0vj}\theta_{0i} + \alpha_{vj}\theta_{vi} - \beta_{vj}$; a sampling of the person trait parameters $\boldsymbol{\theta}$ from

$$\boldsymbol{\theta}_i | \bullet \sim N_{m+1}((\mathbf{A}'\mathbf{A} + \mathbf{P})^{-1}\mathbf{A}'\mathbf{B}, (\mathbf{A}'\mathbf{A} + \mathbf{P})^{-1}), \quad (21)$$

$$\text{where } \mathbf{A} = \begin{bmatrix} \alpha_{01} & \alpha_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \alpha_{02} & \mathbf{0} & \alpha_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \alpha_{0m} & \mathbf{0} & \mathbf{0} & \cdots & \alpha_m \end{bmatrix}_{K \times (m+1)} \quad \text{and } \mathbf{B} = \begin{bmatrix} \mathbf{Z}_{1i} + \beta_1 \\ \mathbf{Z}_{2i} + \beta_2 \\ \vdots \\ \mathbf{Z}_{mi} + \beta_m \end{bmatrix}_{K \times 1}, \quad \text{in which}$$

$\alpha_{0v} = (\alpha_{0v1}, \dots, \alpha_{0vk_v})'$ and $\alpha_v, \mathbf{Z}_{vi}, \beta_v$ are as defined in (8); a sampling of the item parameters $\boldsymbol{\xi}$ from

$$\boldsymbol{\xi}_{vj} | \bullet \sim N_3((\mathbf{x}'_v \mathbf{x}_v + \mathbf{I})^{-1} \mathbf{x}'_v \mathbf{Z}_{vj}, (\mathbf{x}'_v \mathbf{x}_v + \mathbf{I})^{-1}) I(\alpha_{0vj} > 0) I(\alpha_{vj} > 0), \quad (22)$$

where $\mathbf{x} = [\boldsymbol{\theta}_0, \boldsymbol{\theta}_v, -1]$, assuming uniform priors $\alpha_{0vj} > 0$, $\alpha_{vj} > 0$, and $p(\beta_{vj}) \propto 1$, or from

$$\boldsymbol{\xi}_{vj} | \bullet \sim N_3((\mathbf{x}'_v \mathbf{x}_v + \boldsymbol{\Sigma}_{\boldsymbol{\xi}_v}^{-1})^{-1} (\mathbf{x}'_v \mathbf{Z}_{vj} + \boldsymbol{\Sigma}_{\boldsymbol{\xi}_v}^{-1} \boldsymbol{\mu}_{\boldsymbol{\xi}_v}), (\mathbf{x}'_v \mathbf{x}_v + \boldsymbol{\Sigma}_{\boldsymbol{\xi}_v}^{-1})^{-1}) I(\alpha_{0vj} > 0) I(\alpha_{vj} > 0), \quad (23)$$

where $\boldsymbol{\mu}_{\boldsymbol{\xi}_v} = (\mu_{\alpha_{0v}}, \mu_{\alpha_v}, \mu_{\beta_v})'$ and $\boldsymbol{\Sigma}_{\boldsymbol{\xi}_v} = \text{diag}(\sigma_{\alpha_{0v}}^2, \sigma_{\alpha_v}^2, \sigma_{\beta_v}^2)$, assuming conjugate normal priors $\alpha_{0vj} \sim N_{(0, \infty)}(\mu_{\alpha_{0v}}, \sigma_{\alpha_{0v}}^2)$, $\alpha_{vj} \sim N_{(0, \infty)}(\mu_{\alpha_v}, \sigma_{\alpha_v}^2)$, $\beta_{vj} \sim N(\mu_{\beta_v}, \sigma_{\beta_v}^2)$; a sampling of the unconstrained covariance matrix $\boldsymbol{\Sigma}$ from (19), where $\boldsymbol{\theta}_i$ is as defined in (21) and

$$\mathbf{C} = \text{diag} \left(\left(\prod_{v=1}^m \prod_{j=1}^{k_v} \alpha_{0vj} \right)^{1/K}, \left(\prod_{j=1}^{k_1} \alpha_{1j} \right)^{1/k_1}, \dots, \left(\prod_{j=1}^{k_m} \alpha_{mj} \right)^{1/k_m} \right);$$

and a transformation of $\boldsymbol{\Sigma}$ to \mathbf{P} using (14).

Due to an additional model indeterminacy resulted from the additive nature of θ_{0i} and θ_{vi} , further constraints are adopted to help identify the model. Specifically, at each iteration, the θ_{0i} and θ_{vi} parameters are normalized to have a mean of 0 and a standard deviation of 1, with the corresponding α_{0vj} , α_{vj} and β_{vj} parameters being rescaled so that the likelihood is preserved (see Bafumi *et al.* 2005). This procedure, as mentioned in Section 3.1, also improves mixing and computational efficiency.

4. Bayesian model choice or checking

In Bayesian statistics, the adequacy of the fit of a given model is evaluated using several model choice or checking techniques, among which, Bayesian deviance and posterior predictive model checks are considered and briefly illustrated.

4.1. Bayesian deviance

The Bayesian deviance information criterion (DIC; Spiegelhalter, Best, Carlin, and van der Linde 2002) is based on the posterior distribution of the deviance. This criterion is defined as $\text{DIC} = \bar{D} + p_D$, where $\bar{D} = E(-2 \log L(\mathbf{y}|\boldsymbol{\vartheta}))$ is the posterior expectation of the deviance (with $L(\mathbf{y}|\boldsymbol{\vartheta})$ being the model likelihood function, where $\boldsymbol{\vartheta}$ denotes all model parameters) and $p_D = \bar{D} - D(\bar{\boldsymbol{\vartheta}})$ is the effective number of parameters (Carlin and Louis 2000). Further, $D(\bar{\boldsymbol{\vartheta}}) = -2 \log(L(\mathbf{y}|\bar{\boldsymbol{\vartheta}}))$, where $\bar{\boldsymbol{\vartheta}}$ is the posterior mean. To compute Bayesian DIC, MCMC samples of the parameters, $\boldsymbol{\vartheta}^{(1)}, \dots, \boldsymbol{\vartheta}^{(G)}$, can be drawn using the Gibbs sampler, then $\bar{D} = \frac{1}{G}(-2 \log \prod_{g=1}^G L(\mathbf{y}|\boldsymbol{\vartheta}^{(g)}))$. Small values of the deviance suggest a better-fitting model. Generally, more complicated models tend to provide better fit. Hence, penalizing for the number of parameters (p_D) makes DIC a more reasonable measure to use.

4.2. Posterior predictive model checks

The posterior predictive model checking (PPMC; Rubin 1984) method provides a popular Bayesian model checking technique that is intuitively appealing, simple to implement, and easy to interpret (Sinharay and Stern 2003). The basic idea is to draw replicated data \mathbf{y}^{rep} from its posterior predictive distribution $p(\mathbf{y}^{\text{rep}}|\mathbf{y}) = \int p(\mathbf{y}^{\text{rep}}|\boldsymbol{\vartheta})p(\boldsymbol{\vartheta}|\mathbf{y})d\boldsymbol{\vartheta}$, and compare them to the observed data \mathbf{y} . If the model fits, then replicated data generated under the model should look similar to the observed data. A test statistic known as the discrepancy measure $T(\mathbf{y}, \boldsymbol{\vartheta})$ has to be chosen to define the discrepancy between the model and the data. For each \mathbf{y}^{rep} drawn from the predictive distribution, the realized discrepancy $T(\mathbf{y}) = T(\mathbf{y}, \boldsymbol{\vartheta})$ can be compared with the predictive discrepancy $T(\mathbf{y}^{\text{rep}}) = T(\mathbf{y}^{\text{rep}}, \boldsymbol{\vartheta})$ by plotting the pairs on a scatter plot. Alternatively, one can obtain a quantitative measure of lack of fit by calculating the tail-area probability or the PPP-value (Sinharay, Johnson, and Stern 2006), $P(T(\mathbf{y}^{\text{rep}}) \geq T(\mathbf{y})|\mathbf{y}) = \int_{T(\mathbf{y}^{\text{rep}}) \geq T(\mathbf{y})} p(\mathbf{y}^{\text{rep}}|\boldsymbol{\vartheta})p(\boldsymbol{\vartheta}|\mathbf{y})d\mathbf{y}^{\text{rep}}d\boldsymbol{\vartheta}$.

To implement the method, one draws G samples from the posterior distribution of $\boldsymbol{\vartheta}$ using Gibbs sampling. Then for each simulated $\boldsymbol{\vartheta}^{(g)}$, a $\mathbf{y}^{\text{rep}(g)}$ can be drawn from the predictive distribution so there are G draws from the joint posterior distribution $p(\mathbf{y}^{\text{rep}}, \boldsymbol{\vartheta}|\mathbf{y})$. The predictive test statistic $T(\mathbf{y}^{\text{rep}(g)})$ and the realized test statistic $T(\mathbf{y})$ are computed and subsequently compared to provide graphical and numerical evidence about model inadequacy. Specifically, the proportion of the G simulated samples for which the replicated data could be more extreme than the observed data, i.e., $\frac{1}{G} \sum_{g=1}^G I(T(\mathbf{y}^{\text{rep}(g)}) \geq T(\mathbf{y}))$, provides an estimate of the PPP-value. Extreme PPP-values (close to 0 or 1) indicate model misfit. It is noted that although both are defined as tail-area probabilities, the PPP-value has to be differentiated from the traditional hypothesis-testing p -value in that the posterior predictive checking approach does not perform a hypothesis test (Gelman, Meng, and Stern 1996). The choice of the discrepancy measure is critical in implementing this method. Sinharay *et al.* (2006), after evaluating a number of discrepancy measures for assessing IRT models, concluded that the odds ratio for measuring associations among item pairs, $T(\mathbf{y}) = OR_{ij} = \frac{n_{11}n_{00}}{n_{01}n_{10}}$, is powerful for

detecting lack of fit when the model assumption is violated. Hence, this measure is adopted as the test statistic for the PPMC method in this paper.

5. Package IRTm2noHA

The package **IRTm2noHA** contains two major user-callable routines. A function for generating binary response data using the multidimensional model involving a general and several specific latent dimensions titled, **simm2noHA**, and a function that implements Gibbs sampler to obtain posterior samples, estimates, convergence statistics, or model choice/checking statistics, **gsm2noHA**.

The function **simm2noHA** has input arguments **n**, **kv**, **model**, **b**, **r**, and **iparm** for the number of respondents, the number of items in each subtest, the MIRT model based on which the binary response is generated (**model** = 1 for the hierarchical model 1, **model** = 2 for the hierarchical model 2, and **model** = 3 for the additive model), the actual δ or λ coefficients relating the general trait to the specific traits in the two hierarchical models, the actual correlation matrix for person specific traits in the hierarchical model 2 or for person traits in the additive model, and user-specified item parameters, respectively. The optional **iparm** argument allows the user the choice to input item parameters for the model, or randomly generate them from uniform distributions so that $\alpha_{vj} \sim U(0, 2)$, $\beta_{vj} \sim U(-2, 2)$ for the hierarchical models, or $\alpha_{0vj} \sim U(0, 2)$, $\alpha_{vj} \sim U(0, 2)$, $\beta_{vj} \sim U(-2, 2)$ for the additive model. The user can further choose to store the simulated person (**theta**) and item (**item**) parameters. It is noted that a transformation is adopted in the Gibbs sampler for the hierarchical model 1 to rescale the θ_{vi} estimates (see Section 3.1). To keep the simulated values on the same scale, θ_{vi} are also normalized to have a mean of 0 and a standard deviation of 1 for this model.

With the required user-input binary response data (**y**) and number of items in each subtest (**kv**), the function **gsm2noHA** initially reads in starting values for person and item parameters (**th0**, **item0**) and the person hyperparameters (**b0**, **sigma0**), or sets them to be $\theta_{vi}^{(0)} = 0$, $\alpha_{vj}^{(0)} = 2$, $\beta_{vj}^{(0)} = -\Phi^{-1}(\sum_i y_{vij}/n)\sqrt{5}$, (following Albert 1992), $\delta_v^{(0)} = 1$ or $\lambda_v^{(0)} = 1$, and $\mathbf{P}^{(0)} = \mathbf{I}$ for the hierarchical models, or $\theta_{0i}^{(0)} = 0$, $\theta_{vi}^{(0)} = 0$, $\alpha_{0vj}^{(0)} = 2$, $\alpha_{vj}^{(0)} = 2$, $\beta_{vj}^{(0)} = 1$, and $\mathbf{P}^{(0)} = \mathbf{I}$ for the additive model. It then implements the Gibbs sampler to a user-specified MIRT model incorporating general and specific traits (**model**) and iteratively draws random samples for the parameters and hyperparameters from their respective full conditional distributions. The prior distributions for the item parameters or the δ/λ parameters can be uniform (**unif** = 1, default) or conjugate (**unif** = 0). In the latter case, the user can specify any values of interest or use the default values, namely, $\mu_{\alpha_v} = 0$, $\mu_{\beta_v} = 0$ (**xmu**), $\sigma_{\alpha_v}^2 = 1$, $\sigma_{\beta_v}^2 = 1$ (**xvar**) for α_{vj} and β_{vj} , and $\mu_{\delta_v}/\mu_{\lambda_v} = 0$ (**bm**), $\sigma_{\delta_v}^2/\sigma_{\lambda_v}^2 = 1$ (**bvar**) for δ_v/λ_v in the hierarchical models, or $\mu_{\alpha_{0v}} = 0$, $\mu_{\alpha_v} = 0$, $\mu_{\beta_v} = 0$ (**xmu**), $\sigma_{\alpha_{0v}}^2 = 1$, $\sigma_{\alpha_v}^2 = 1$, $\sigma_{\beta_v}^2 = 1$ (**xvar**) for α_{0vj} , α_{vj} and β_{vj} in the additive model. It is noted that the prior location and scale parameters for α_{vj} , β_{vj} , or α_{0vj} can be set to be different across the m subtests. The algorithm continues until all the (**kk**) samples are simulated, with the early burn-in samples (**burnin**) being discarded, where **kk** and **burnin** can be 10,000 and **kk**/2 (default) or any values of interest. It then computes the posterior estimates, posterior standard deviations, and Monte Carlo standard errors of the person (**pparm**), item (**iparm**) or hyperparameter (**hparm**) estimates. Posterior samples after the burn-in stage of these parameters can also be stored (**samples**) for further analysis.

In addition to Monte Carlo standard errors, convergence can be evaluated using the Gelman-Rubin R statistic (Gelman *et al.* 2004) for each person or item parameter. The usual practice is using multiple Markov chains from different starting points. Alternatively, a single chain can be divided into sub-chains so that convergence is assessed by comparing the between and within sub-chain variances. Since a single chain is less wasteful in the number of iterations needed, the latter approach is adopted to compute the R statistic (`gr`) with `gsm2noHA`. The Bayesian deviance estimates, including \bar{D} , $D(\vartheta)$, p_D and DIC, can be obtained (`deviance`) to measure the relative fit of a model. Moreover, the PPMC method can be adopted using the odds ratio as the discrepancy measure so that PPP-values (`ppmc`) are obtained to evaluate model misfit. Extreme PPP-values can be further plotted using the function `ppmcplt`, in which the threshold (`crit`) can be 0.01 (default) or any level of interest so that PPP-values larger than $1 - \text{crit}/2$ are denoted using the right triangle sign and those smaller than $\text{crit}/2$ are denoted using the left triangle sign. The functions' input and output arguments are completely specified in the m-files.

6. Illustrative examples

To demonstrate the use of the **IRTm2noHA** package, simulated and real data examples are provided in this section to illustrate item parameter recovery as well as model comparisons. The code to reproduce the results of each example is provided in the m-file `v34i03.m`.

6.1. Parameter recovery

For parameter recovery, tests with two subtests were considered so that the first half measured one latent trait and the second half measured another. A 1000-by-32 (i.e., $n = 1000$, $m = 2$, $k_1 = 16$, $k_2 = 16$, and $K = 32$) dichotomous data matrix was generated from each of the three MIRT models described in Section 2. The item parameters were randomly simulated from $\alpha_{0vj} \sim U(0, 2)$, $\alpha_{vj} \sim U(0, 2)$, and $\beta_{vj} \sim U(-2, 2)$, and are shown in the first column of Tables 1 through 7. Gibbs sampling was subsequently implemented to recover model parameters assuming the noninformative uniform or informative conjugate prior distributions described previously. The posterior means and standard deviations of item parameters (α_v , β_v , and/or α_{0v}), as well as their Monte Carlo standard errors of estimates and Gelman-Rubin R statistics were obtained and are displayed in the rest of the tables. The estimation accuracy was also evaluated using the average square error between the actual and estimated item parameters, which is shown at the bottom of each table.

The Gelman-Rubin R statistic provides a numerical measure for assessing convergence for each item parameter. With values close to 1, it is determined that in the implementations of the Gibbs sampler, Markov chains reached stationarity with a run length of 10,000 iterations and a burn-in period of 5,000 iterations. The posterior estimates of the item parameters shown in the seven tables are fairly close to the true parameters, suggesting that each MIRT model incorporating a general trait and several specific trait dimensions recovers its model parameters with enough accuracy. In addition, the two sets of posterior estimates, resulted from different prior distributions, differ only slightly from each other, signifying that the posterior estimates are not sensitive to the choice of uniform or conjugate priors for the item and δ/λ parameters. This point is further supported by the small differences between the average square errors.

True	Noninformative priors				Informative priors			
	Mean	SD	MCSE	R	Mean	SD	MCSE	R
α_1								
0.9200	0.8430	0.0752	0.0043	1.0017	0.8409	0.0729	0.0034	1.0017
0.4367	0.3862	0.0714	0.0039	1.0074	0.3779	0.0721	0.0032	1.0069
0.5408	0.6060	0.0589	0.0021	1.0044	0.6090	0.0607	0.0026	1.0026
0.0189	0.0646	0.0433	0.0014	1.0002	0.0623	0.0417	0.0007	1.0000
1.6427	1.6079	0.1335	0.0162	1.0633	1.5662	0.1318	0.0143	1.0791
1.8599	2.3267	0.2827	0.0302	1.0278	2.1719	0.1968	0.0231	1.0394
0.9943	1.0113	0.0956	0.0059	1.0267	1.0129	0.1018	0.0071	1.0442
1.3165	1.2855	0.0938	0.0057	1.0032	1.2755	0.0993	0.0063	1.0118
0.6044	0.5670	0.0532	0.0020	1.0002	0.5694	0.0545	0.0015	1.0017
0.2940	0.3136	0.0488	0.0022	1.0050	0.3137	0.0494	0.0016	1.0022
0.4665	0.3432	0.0638	0.0035	1.0215	0.3507	0.0663	0.0028	1.0041
1.1079	1.2406	0.1009	0.0056	1.0205	1.2299	0.1018	0.0075	1.0208
1.4760	1.3453	0.1120	0.0072	1.0326	1.3344	0.1240	0.0141	1.0734
0.3414	0.3586	0.0486	0.0018	1.0019	0.3572	0.0479	0.0014	1.0022
0.4347	0.4671	0.0739	0.0034	1.0081	0.4628	0.0709	0.0047	1.0287
0.8841	0.8100	0.0694	0.0028	1.0094	0.8097	0.0690	0.0027	1.0041
α_2								
1.2562	1.2733	0.1032	0.0064	1.0248	1.2528	0.0986	0.0060	1.0081
0.4675	0.5060	0.0511	0.0021	1.0110	0.5087	0.0498	0.0014	1.0010
0.1689	0.2604	0.0465	0.0011	1.0002	0.2586	0.0469	0.0011	1.0025
0.7057	0.6992	0.0619	0.0023	1.0037	0.6954	0.0597	0.0017	1.0021
0.2663	0.2319	0.0657	0.0023	1.0056	0.2331	0.0671	0.0031	1.0104
0.9276	0.9670	0.0738	0.0021	1.0024	0.9605	0.0770	0.0039	1.0022
0.1900	0.2422	0.0459	0.0009	1.0003	0.2441	0.0465	0.0012	1.0012
1.4285	1.4023	0.1709	0.0260	1.0863	1.3651	0.1553	0.0235	1.1355
0.7349	0.7368	0.0731	0.0053	1.0022	0.7361	0.0723	0.0027	1.0059
1.5261	1.5850	0.1353	0.0174	1.1441	1.5484	0.1250	0.0066	1.0203
0.9510	1.0386	0.0701	0.0038	1.0173	1.0327	0.0749	0.0019	1.0014
1.1355	1.1848	0.0841	0.0038	1.0027	1.1742	0.0887	0.0070	1.0338
0.5615	0.6170	0.0687	0.0035	1.0027	0.6122	0.0666	0.0032	1.0035
0.3012	0.3260	0.0574	0.0012	1.0030	0.3254	0.0579	0.0028	1.0119
0.4418	0.5183	0.0607	0.0027	1.0033	0.5194	0.0625	0.0025	1.0059
1.3371	1.3690	0.1044	0.0058	1.0110	1.3559	0.1031	0.0048	1.0198
	Average square error = 0.0104				Average square error = 0.0067			

Table 1: Posterior mean, standard deviation (SD), Monte Carlo standard error of the estimate (MCSE) and Gelman-Rubin R statistic for α_v in the hierarchical model 1 where $\delta_1 = .6$ and $\delta_2 = .8$ (chain length = 10,000, burn-in = 5,000).

True	Noninformative priors				Informative priors			
	Mean	SD	MCSE	R	Mean	SD	MCSE	R
β_1								
-0.9783	-0.9068	0.0587	0.0025	1.0005	-0.9041	0.0572	0.0018	1.0026
-1.5911	-1.5623	0.0699	0.0023	1.0024	-1.5538	0.0712	0.0036	1.0044
0.6860	0.6906	0.0459	0.0010	1.0002	0.6900	0.0483	0.0023	1.0069
-1.2417	-1.2918	0.0544	0.0016	1.0060	-1.2868	0.0557	0.0017	1.0000
-0.5412	-0.6128	0.0639	0.0057	1.0321	-0.6059	0.0638	0.0044	1.0305
0.1742	0.3166	0.0784	0.0068	1.0127	0.2913	0.0674	0.0038	1.0100
-1.5993	-1.5149	0.0870	0.0054	1.0310	-1.5121	0.0926	0.0067	1.0447
0.1798	0.2626	0.0503	0.0019	1.0002	0.2567	0.0503	0.0023	1.0027
0.0593	0.1534	0.0419	0.0007	1.0004	0.1530	0.0407	0.0012	1.0029
0.4612	0.4931	0.0422	0.0015	1.0019	0.4909	0.0423	0.0014	1.0000
-1.3639	-1.3449	0.0598	0.0029	1.0104	-1.3459	0.0608	0.0020	0.9999
-0.9123	-0.9669	0.0705	0.0028	1.0090	-0.9598	0.0698	0.0040	1.0130
1.1756	1.0618	0.0763	0.0045	1.0243	1.0549	0.0807	0.0077	1.0522
-0.2556	-0.2636	0.0414	0.0009	1.0032	-0.2629	0.0413	0.0013	1.0037
-1.4649	-1.4735	0.0706	0.0024	1.0012	-1.4679	0.0688	0.0036	1.0208
-0.9086	-0.7995	0.0531	0.0017	1.0067	-0.7984	0.0533	0.0025	1.0145
β_2								
0.6280	0.6431	0.0594	0.0035	1.0231	0.6384	0.0580	0.0025	1.0071
0.0095	0.0167	0.0407	0.0008	1.0005	0.0174	0.0405	0.0010	1.0046
-0.4497	-0.4327	0.0414	0.0007	0.9997	-0.4311	0.0417	0.0009	1.0014
-0.5110	-0.5810	0.0471	0.0016	1.0064	-0.5790	0.0468	0.0007	0.9997
1.4806	1.4481	0.0602	0.0017	1.0050	1.4414	0.0617	0.0018	1.0025
0.5780	0.5553	0.0503	0.0015	1.0063	0.5536	0.0516	0.0023	1.0082
-0.1822	-0.1971	0.0400	0.0006	1.0008	-0.1980	0.0404	0.0010	1.0013
-1.9445	-1.9836	0.1692	0.0260	1.0908	-1.9512	0.1574	0.0229	1.1256
-1.0016	-1.0124	0.0597	0.0032	1.0002	-1.0116	0.0605	0.0021	1.0047
0.6201	0.6661	0.0693	0.0067	1.0663	0.6546	0.0653	0.0031	1.0190
0.0102	0.0238	0.0474	0.0012	1.0042	0.0228	0.0458	0.0011	1.0028
-0.3454	-0.4190	0.0508	0.0023	0.9992	-0.4151	0.0507	0.0027	1.0092
1.0597	1.0731	0.0578	0.0020	1.0001	1.0697	0.0581	0.0021	1.0038
-0.9957	-1.0509	0.0522	0.0012	1.0023	-1.0496	0.0506	0.0012	1.0003
-0.9956	-1.0600	0.0548	0.0019	1.0015	-1.0544	0.0546	0.0020	1.0010
-0.0211	-0.0081	0.0515	0.0017	1.0070	-0.0104	0.0507	0.0008	1.0016
	Average square error = 0.0034				Average square error = 0.0031			

Table 2: Posterior mean, standard deviation (SD), Monte Carlo standard error of the estimate (MCSE) and Gelman-Rubin R statistic for β_v in the hierarchical model 1 where $\delta_1 = .6$ and $\delta_2 = .8$ (chain length = 10,000, burn-in = 5,000).

True	Noninformative priors				Informative priors			
	Mean	SD	MCSE	R	Mean	SD	MCSE	R
α_1								
0.9200	0.9367	0.0794	0.0032	1.0073	0.9289	0.0851	0.0065	1.0540
0.4367	0.4709	0.0739	0.0049	1.0219	0.4633	0.0726	0.0039	1.0276
0.5408	0.5393	0.0558	0.0022	1.0129	0.5311	0.0576	0.0025	1.0096
0.0189	0.1795	0.0593	0.0015	1.0022	0.1728	0.0584	0.0023	1.0083
1.6427	1.4750	0.1143	0.0075	1.0215	1.4323	0.1143	0.0063	1.0131
1.8599	2.1168	0.2207	0.0348	1.1330	2.0524	0.2100	0.0348	1.0916
0.9943	0.9745	0.0973	0.0061	1.0289	0.9483	0.0942	0.0072	1.0234
1.3165	1.4624	0.1163	0.0083	1.0068	1.4307	0.1122	0.0064	1.0117
0.6044	0.5947	0.0567	0.0016	1.0025	0.5880	0.0560	0.0027	1.0206
0.2940	0.3290	0.0498	0.0015	1.0031	0.3227	0.0494	0.0016	1.0080
0.4665	0.5210	0.0755	0.0046	1.0194	0.5016	0.0729	0.0029	1.0056
1.1079	1.2433	0.1047	0.0082	1.0298	1.2452	0.0997	0.0067	1.0218
1.4760	1.4776	0.1265	0.0117	1.0544	1.4378	0.1267	0.0126	1.0297
0.3414	0.3244	0.0471	0.0010	0.9998	0.3223	0.0485	0.0026	1.0181
0.4347	0.4343	0.0770	0.0061	1.0277	0.4194	0.0697	0.0025	1.0058
0.8841	0.9643	0.0802	0.0041	1.0148	0.9596	0.0817	0.0044	1.0207
α_2								
1.2562	1.3276	0.1120	0.0095	1.0535	1.3173	0.1169	0.0115	1.0852
0.4675	0.4770	0.0505	0.0010	1.0010	0.4715	0.0510	0.0017	1.0001
0.1689	0.1541	0.0457	0.0015	1.0029	0.1525	0.0447	0.0010	1.0006
0.7057	0.8328	0.0683	0.0035	1.0193	0.8312	0.0677	0.0030	1.0203
0.2663	0.2081	0.0668	0.0037	1.0055	0.2073	0.0668	0.0043	1.0000
0.9276	0.9285	0.0783	0.0041	1.0064	0.9172	0.0730	0.0031	1.0112
0.1900	0.2760	0.0470	0.0021	1.0029	0.2761	0.0464	0.0009	1.0016
1.4285	1.4849	0.1632	0.0193	1.1127	1.4109	0.1424	0.0085	1.0123
0.7349	0.7410	0.0708	0.0021	1.0070	0.7323	0.0728	0.0037	1.0027
1.5261	1.4517	0.1306	0.0142	1.0840	1.4044	0.1113	0.0074	1.0067
0.9510	1.0972	0.0825	0.0062	1.0201	1.0779	0.0807	0.0025	1.0034
1.1355	1.2415	0.0982	0.0055	1.0237	1.2102	0.1030	0.0079	1.0073
0.5615	0.6145	0.0704	0.0025	1.0025	0.6010	0.0672	0.0028	1.0050
0.3012	0.2512	0.0554	0.0022	1.0054	0.2504	0.0552	0.0016	1.0010
0.4418	0.4307	0.0577	0.0022	1.0059	0.4240	0.0582	0.0014	1.0002
1.3371	1.5194	0.1254	0.0119	1.0419	1.4993	0.1247	0.0068	1.0069
	Average square error = 0.0089				Average square error = 0.0078			

Table 3: Posterior mean, standard deviation (SD), Monte Carlo standard error of the estimate (MCSE) and Gelman-Rubin R statistic for α_v in the hierarchical model 2 where $\delta_1 = .6$, $\delta_2 = .8$ and $\rho_{12} = .5$ (chain length = 10,000, burn-in = 5,000).

True	Noninformative priors				Informative priors			
	Mean	SD	MCSE	R	Mean	SD	MCSE	R
β_1								
-0.9783	-0.9065	0.0655	0.0037	1.0110	-0.9057	0.0672	0.0040	1.0250
-1.5911	-1.5377	0.0726	0.0040	1.0086	-1.5329	0.0708	0.0043	1.0313
0.6860	0.6606	0.0483	0.0019	1.0057	0.6586	0.0495	0.0020	1.0075
-1.2417	-1.2048	0.0533	0.0010	1.0012	-1.2019	0.0536	0.0019	1.0059
-0.5412	-0.5109	0.0716	0.0031	1.0100	-0.5040	0.0755	0.0034	1.0073
0.1742	0.1900	0.0925	0.0060	1.0217	0.1818	0.0896	0.0073	1.0297
-1.5993	-1.5616	0.0936	0.0050	1.0152	-1.5483	0.0896	0.0068	1.0097
0.1798	0.1134	0.0677	0.0023	1.0066	0.1093	0.0681	0.0037	1.0036
0.0593	0.0881	0.0461	0.0012	1.0012	0.0883	0.0470	0.0021	1.0093
0.4612	0.5771	0.0438	0.0014	1.0047	0.5767	0.0453	0.0013	1.0058
-1.3639	-1.3755	0.0699	0.0033	1.0205	-1.3643	0.0668	0.0025	1.0085
-0.9123	-0.9281	0.0768	0.0035	1.0017	-0.9350	0.0775	0.0039	1.0165
1.1756	1.1276	0.0941	0.0077	1.0427	1.1130	0.0943	0.0088	1.0087
-0.2556	-0.2459	0.0426	0.0010	1.0008	-0.2480	0.0420	0.0011	1.0022
-1.4649	-1.5165	0.0719	0.0053	1.0283	-1.5023	0.0674	0.0026	1.0063
-0.9086	-0.9033	0.0655	0.0035	1.0103	-0.9062	0.0673	0.0038	1.0147
β_2								
0.6280	0.6841	0.0729	0.0066	1.0594	0.6819	0.0728	0.0049	1.0395
0.0095	0.0759	0.0446	0.0019	1.0071	0.0756	0.0440	0.0015	1.0054
-0.4497	-0.3907	0.0422	0.0011	1.0055	-0.3922	0.0409	0.0006	1.0006
-0.5110	-0.4442	0.0519	0.0028	1.0083	-0.4468	0.0545	0.0016	1.0056
1.4806	1.4154	0.0611	0.0017	0.9992	1.4158	0.0600	0.0017	1.0052
0.5780	0.6841	0.0610	0.0037	1.0158	0.6798	0.0596	0.0029	1.0177
-0.1822	-0.1603	0.0411	0.0012	1.0028	-0.1600	0.0419	0.0008	0.9998
-1.9445	-1.9879	0.1548	0.0200	1.1750	-1.9277	0.1384	0.0078	1.0122
-1.0016	-0.9895	0.0604	0.0018	1.0020	-0.9865	0.0638	0.0028	1.0070
0.6201	0.6695	0.0773	0.0073	1.0707	0.6491	0.0710	0.0027	1.0050
0.0102	-0.0105	0.0552	0.0028	1.0128	-0.0148	0.0578	0.0031	1.0127
-0.3454	-0.2301	0.0585	0.0039	1.0194	-0.2304	0.0628	0.0033	1.0191
1.0597	1.1380	0.0629	0.0021	1.0019	1.1304	0.0625	0.0028	1.0127
-0.9957	-0.9963	0.0496	0.0018	1.0033	-0.9949	0.0489	0.0015	1.0029
-0.9956	-0.9101	0.0508	0.0020	1.0135	-0.9093	0.0510	0.0016	1.0047
-0.0211	0.0551	0.0681	0.0044	1.0273	0.0501	0.0688	0.0024	1.0048
	Average square error = 0.0033				Average square error = 0.0032			

Table 4: Posterior mean, standard deviation (SD), Monte Carlo standard error of the estimate (MCSE) and Gelman-Rubin R statistic for β_v in the hierarchical model 2 where $\delta_1 = .6$, $\delta_2 = .8$ and $\rho_{12} = .5$ (chain length = 10,000, burn-in = 5,000).

True	Noninformative priors				Informative priors			
	Mean	SD	MCSE	R	Mean	SD	MCSE	R
α_{01}								
0.3622	0.3950	0.1020	0.0092	1.0257	0.3849	0.1051	0.0125	1.0679
1.8023	1.7087	0.1764	0.0206	1.0069	1.6251	0.1468	0.0127	1.0211
0.8676	0.8240	0.0877	0.0089	1.0287	0.8204	0.0844	0.0080	1.0769
0.9131	1.0560	0.0988	0.0077	1.0462	1.0328	0.0926	0.0037	1.0021
0.3775	0.3376	0.1539	0.0199	1.0435	0.3227	0.1467	0.0221	1.1052
1.8610	1.8023	0.2156	0.0304	1.0632	1.7220	0.2044	0.0338	1.1313
1.5901	1.6608	0.1819	0.0228	1.0548	1.5544	0.1440	0.0206	1.1514
0.3591	0.5296	0.1330	0.0153	1.0345	0.5133	0.1377	0.0203	1.1252
0.4321	0.4001	0.0766	0.0070	1.0153	0.3904	0.0762	0.0084	1.0661
1.5851	1.5611	0.1233	0.0132	1.0315	1.5583	0.1289	0.0075	1.0311
1.0088	1.0856	0.1140	0.0062	0.9995	1.0638	0.1065	0.0076	1.0173
0.6340	0.5418	0.1166	0.0134	1.0369	0.5306	0.1137	0.0138	1.0716
0.1482	0.1711	0.1202	0.0158	1.0401	0.1692	0.1179	0.0179	1.1214
1.4812	1.6710	0.1478	0.0216	1.1396	1.6405	0.1449	0.0164	1.0898
1.9192	2.0101	0.2118	0.0260	1.0508	1.8728	0.1734	0.0263	1.0155
0.0574	0.0800	0.0669	0.0073	1.0455	0.0736	0.0609	0.0067	1.0668
α_{02}								
0.4049	0.1710	0.1175	0.0109	1.0409	0.2080	0.1374	0.0255	1.0717
0.3290	0.2495	0.0707	0.0044	1.0190	0.2591	0.0768	0.0106	1.0266
0.4107	0.4005	0.0625	0.0023	1.0016	0.4049	0.0633	0.0037	1.0054
0.6261	0.6225	0.0881	0.0063	1.0138	0.6353	0.0870	0.0118	1.0282
1.7202	1.5387	0.1537	0.0152	1.0667	1.4904	0.1547	0.0166	1.0750
1.0177	0.9394	0.1108	0.0065	1.0042	0.9453	0.1205	0.0179	1.0164
1.1805	1.3132	0.1028	0.0096	1.0273	1.3141	0.1021	0.0039	1.0106
1.1562	1.0613	0.1691	0.0177	1.0476	1.0401	0.1787	0.0348	1.0617
0.9309	0.8562	0.0978	0.0071	1.0126	0.8592	0.0979	0.0134	1.0029
0.1401	0.1090	0.0927	0.0088	1.0484	0.1343	0.1106	0.0209	1.1045
0.3970	0.3308	0.0900	0.0055	1.0168	0.3474	0.0993	0.0170	1.0592
0.8003	0.8278	0.1250	0.0106	1.0285	0.8402	0.1353	0.0231	1.0733
1.1214	1.0252	0.1047	0.0071	1.0314	1.0356	0.1084	0.0114	1.0167
0.8617	0.8807	0.0979	0.0064	1.0247	0.8846	0.0974	0.0111	1.0192
0.9304	0.7925	0.0970	0.0065	1.0202	0.7844	0.0958	0.0133	1.0531
1.8383	1.8918	0.1763	0.0171	1.0301	1.8378	0.1995	0.0279	1.0549
	Average square error = 0.0095				Average square error = 0.0097			

Table 5: Posterior mean, standard deviation (SD), Monte Carlo standard error of the estimate (MCSE) and Gelman-Rubin R statistic for α_{0v} in the additive model where $\rho_{01} = .4$, $\rho_{02} = .4$ and $\rho_{12} = .8$ (chain length = 10,000, burn-in = 5,000).

True	Noninformative priors				Informative priors			
	Mean	SD	MCSE	R	Mean	SD	MCSE	R
α_1								
0.9200	0.8992	0.0980	0.0064	1.0175	0.8842	0.1007	0.0073	1.0405
0.4367	0.4592	0.1197	0.0095	1.0388	0.4390	0.1089	0.0103	1.0674
0.5408	0.4769	0.0749	0.0060	1.0139	0.4698	0.0766	0.0037	1.0163
0.0189	0.0418	0.0371	0.0015	1.0026	0.0396	0.0339	0.0016	1.0040
1.6427	1.5697	0.1674	0.0123	1.0094	1.5432	0.1521	0.0116	1.0226
1.8599	1.5576	0.1766	0.0188	1.0279	1.4848	0.1628	0.0160	1.0284
0.9943	1.0283	0.1448	0.0186	1.0728	0.9636	0.1264	0.0066	1.0064
1.3165	1.3626	0.1256	0.0078	1.0090	1.3551	0.1315	0.0110	1.0227
0.6044	0.5889	0.0728	0.0068	1.0471	0.5922	0.0703	0.0039	1.0304
0.2940	0.2196	0.0944	0.0095	1.0224	0.2097	0.0923	0.0092	1.0513
0.4665	0.4532	0.0968	0.0067	1.0096	0.4380	0.0949	0.0084	1.0202
1.1079	1.0224	0.1078	0.0101	1.0284	1.0128	0.1039	0.0084	1.0665
1.4760	1.3613	0.1546	0.0215	1.0510	1.3524	0.1345	0.0100	1.0258
0.3414	0.4107	0.1081	0.0081	1.0259	0.3931	0.0989	0.0088	1.0473
0.4347	0.1818	0.1066	0.0109	1.0590	0.1701	0.0996	0.0096	1.0437
0.8841	0.8333	0.0871	0.0068	1.0456	0.8251	0.0818	0.0036	1.0124
α_2								
1.2562	1.5201	0.1504	0.0138	1.0218	1.5182	0.1488	0.0101	1.0157
0.4675	0.5380	0.0708	0.0029	1.0089	0.5351	0.0732	0.0059	1.0085
0.1689	0.1604	0.0619	0.0041	1.0052	0.1583	0.0631	0.0030	1.0035
0.7057	0.6430	0.0819	0.0058	1.0359	0.6299	0.0880	0.0087	1.0115
0.2663	0.3216	0.1113	0.0101	1.0241	0.3102	0.1136	0.0109	1.0344
0.9276	0.9084	0.1076	0.0075	1.0324	0.8896	0.1029	0.0100	1.0075
0.1900	0.1146	0.0715	0.0055	1.0139	0.1036	0.0658	0.0058	1.0277
1.4285	1.4796	0.1979	0.0196	1.0891	1.4045	0.1940	0.0307	1.0095
0.7349	0.6369	0.0966	0.0079	1.0487	0.6158	0.0952	0.0109	1.0261
1.5261	1.3962	0.1372	0.0088	1.0168	1.3786	0.1326	0.0108	1.0163
0.9510	0.8547	0.0887	0.0059	1.0339	0.8393	0.0902	0.0088	1.0049
1.1355	1.2105	0.1244	0.0101	1.0257	1.1865	0.1274	0.0147	1.0126
0.5615	0.5438	0.0892	0.0066	1.0138	0.5332	0.0982	0.0087	1.0157
0.3012	0.3747	0.0918	0.0067	1.0126	0.3556	0.0875	0.0101	1.0299
0.4418	0.5267	0.0910	0.0069	1.0209	0.5192	0.0911	0.0077	1.0021
1.3371	1.3294	0.1841	0.0289	1.0553	1.2450	0.1618	0.0222	1.0449
	Average square error = 0.0109				Average square error = 0.0134			

Table 6: Posterior mean, standard deviation (SD), Monte Carlo standard error of the estimate (MCSE) and Gelman-Rubin R statistic for α_v in the additive model where $\rho_{01} = .4$, $\rho_{02} = .4$ and $\rho_{12} = .8$ (chain length = 10,000, burn-in = 5,000).

True	Noninformative priors				Informative priors			
	Mean	SD	MCSE	R	Mean	SD	MCSE	R
β_1								
-0.9783	-1.0321	0.0671	0.0030	1.0065	-1.0246	0.0669	0.0033	1.0131
-1.5911	-1.4919	0.1118	0.0118	1.0090	-1.4376	0.1014	0.0066	1.0255
0.6860	0.7160	0.0541	0.0016	1.0007	0.7184	0.0535	0.0027	1.0040
-1.2417	-1.2544	0.0788	0.0057	1.0373	-1.2332	0.0749	0.0032	1.0015
-0.5412	-0.5928	0.0706	0.0034	1.0033	-0.5801	0.0653	0.0042	1.0328
0.1742	0.1034	0.0687	0.0031	1.0184	0.1190	0.0690	0.0037	1.0053
-1.5993	-1.5733	0.1298	0.0173	1.0891	-1.4889	0.1119	0.0125	1.0442
0.1798	0.1345	0.0564	0.0024	1.0088	0.1403	0.0540	0.0025	1.0080
0.0593	-0.0051	0.0442	0.0009	1.0026	-0.0039	0.0440	0.0011	1.0016
0.4612	0.3799	0.0583	0.0037	1.0165	0.3905	0.0571	0.0014	1.0048
-1.3639	-1.4328	0.0890	0.0034	1.0071	-1.4196	0.0872	0.0060	1.0204
-0.9123	-0.9047	0.0676	0.0051	1.0189	-0.8984	0.0667	0.0042	1.0325
1.1756	1.0642	0.0880	0.0102	1.0258	1.0626	0.0790	0.0036	1.0117
-0.2556	-0.2793	0.0593	0.0043	1.0268	-0.2627	0.0590	0.0038	1.0159
-1.4649	-1.4179	0.1343	0.0153	1.0400	-1.3331	0.1066	0.0149	1.0088
-0.9086	-1.0148	0.0628	0.0034	1.0281	-1.0077	0.0607	0.0024	1.0126
β_2								
0.6280	0.6330	0.0632	0.0040	1.0061	0.6420	0.0672	0.0030	1.0075
0.0095	0.0281	0.0439	0.0010	1.0003	0.0289	0.0431	0.0005	1.0001
-0.4497	-0.5023	0.0443	0.0012	1.0043	-0.5010	0.0440	0.0013	1.0006
-0.5110	-0.4396	0.0489	0.0012	1.0003	-0.4346	0.0495	0.0014	1.0025
1.4806	1.4643	0.1113	0.0118	1.0637	1.4313	0.1058	0.0138	1.0952
0.5780	0.5512	0.0614	0.0032	1.0075	0.5531	0.0578	0.0018	1.0017
-0.1822	-0.1873	0.0512	0.0016	1.0046	-0.1840	0.0517	0.0017	1.0085
-1.9445	-2.0231	0.1797	0.0225	1.1532	-1.9518	0.1738	0.0291	1.0318
-1.0016	-1.0401	0.0662	0.0039	1.0169	-1.0306	0.0663	0.0030	1.0179
0.6201	0.6398	0.0666	0.0029	1.0118	0.6400	0.0658	0.0048	1.0055
0.0102	-0.0407	0.0470	0.0015	1.0089	-0.0398	0.0462	0.0017	1.0054
-0.3454	-0.3687	0.0603	0.0027	1.0013	-0.3615	0.0569	0.0018	1.0034
1.0597	1.0360	0.0711	0.0050	1.0323	1.0355	0.0704	0.0029	1.0029
-0.9957	-1.1128	0.0687	0.0039	1.0193	-1.1034	0.0679	0.0033	1.0166
-0.9956	-0.9348	0.0648	0.0039	1.0046	-0.9229	0.0617	0.0034	1.0175
-0.0211	-0.0540	0.0724	0.0056	1.0256	-0.0318	0.0703	0.0042	1.0067
	Average square error = 0.0033				Average square error = 0.0041			

Table 7: Posterior mean, standard deviation (SD), Monte Carlo standard error of the estimate (MCSE) and Gelman-Rubin R statistic for β_v in the additive model where $\rho_{01} = .4$, $\rho_{02} = .4$ and $\rho_{12} = .8$ (chain length = 10,000, burn-in = 5,000).

6.2. Model comparison

To illustrate the use of the Bayesian model choice or checking techniques for evaluating the relative (mis)fit of a model, two 1000-by-30 dichotomous data matrices were simulated from the hierarchical model 2 and the additive model, respectively, so that 15 items measured θ_1 and another 15 items measured θ_2 . This way, the three MIRT models incorporating both general and specific latent dimensions could be evaluated and compared under the situation where the actual structure was hierarchical or additive. To generate the data matrices, item parameters were randomly drawn from the uniform distributions described in Section 5. The Gibbs sampler assuming noninformative priors for the item and δ/λ parameters was subsequently implemented for each of the three MIRT models so that 10,000 samples were simulated with the first 5,000 set to be burn-in. The Gelman-Rubin R statistics suggest that the chains converged to their stationary distributions within 10,000 iterations. Hence, the Bayesian deviance estimates, including \bar{D} , $D(\bar{\vartheta})$, p_D and DIC, were obtained from each implementation and are displayed in Table 8. It is clear from the table that both posterior deviance (column 2) and Bayesian DIC (column 5) estimates pointed to the more complicated additive model, indicating that it provided a better description of the simulated data regardless of the actual relationship between general and specific traits. A close examination of these estimates reveals that compared with the situation where the latent structure was hierarchical, the additive model resulted in a larger reduction in the deviance or DIC estimate and hence had a higher improvement in model-data fit over the hierarchical models when the actual structure was additive. Another interesting observation of Table 8 pertains to the effective number of parameters (p_D). Given the model complexity, the additive model consistently had a larger p_D than the two hierarchical models. Among the two scenarios considered, the difference in p_D between the two types of models was noticeably larger when the actual structure was additive. It has to be noted that the difference between the DIC estimates for the two hierarchical models was rather trivial. Indeed, the two models had almost identical values of \bar{D} , $D(\bar{\vartheta})$, p_D and DIC, and hence were essentially similar in model-data fit in spite of different assumptions made regarding the latent structure.

PPMC was also implemented to obtain PPP-values for the three MIRT models where odds ratios were used as the discrepancy measure. Graphical representations of extreme PPP-values are shown in Figure 2, where the upper diagonal is left blank due to symmetry. Here,

	\bar{D}	$D(\bar{\vartheta})$	p_D	DIC
with a hierarchical structure				
hierarchical model 1	23069.0215	21393.1488	1675.8727	24744.8942
hierarchical model 2	23067.2464	21392.6289	1674.6174	24741.8638
additive model	22732.8392	20969.6988	1763.1404	24495.9796
with an additive structure				
hierarchical model 1	17163.9692	15596.2537	1567.7155	18731.6848
hierarchical model 2	17163.8166	15596.5737	1567.2429	18731.0595
additive model	16020.1091	14026.8340	1993.2751	18013.3842

Table 8: Bayesian deviance estimates for the three MIRT models fit to the simulated data where the actual general and specific traits form a hierarchical structure or an additive structure.

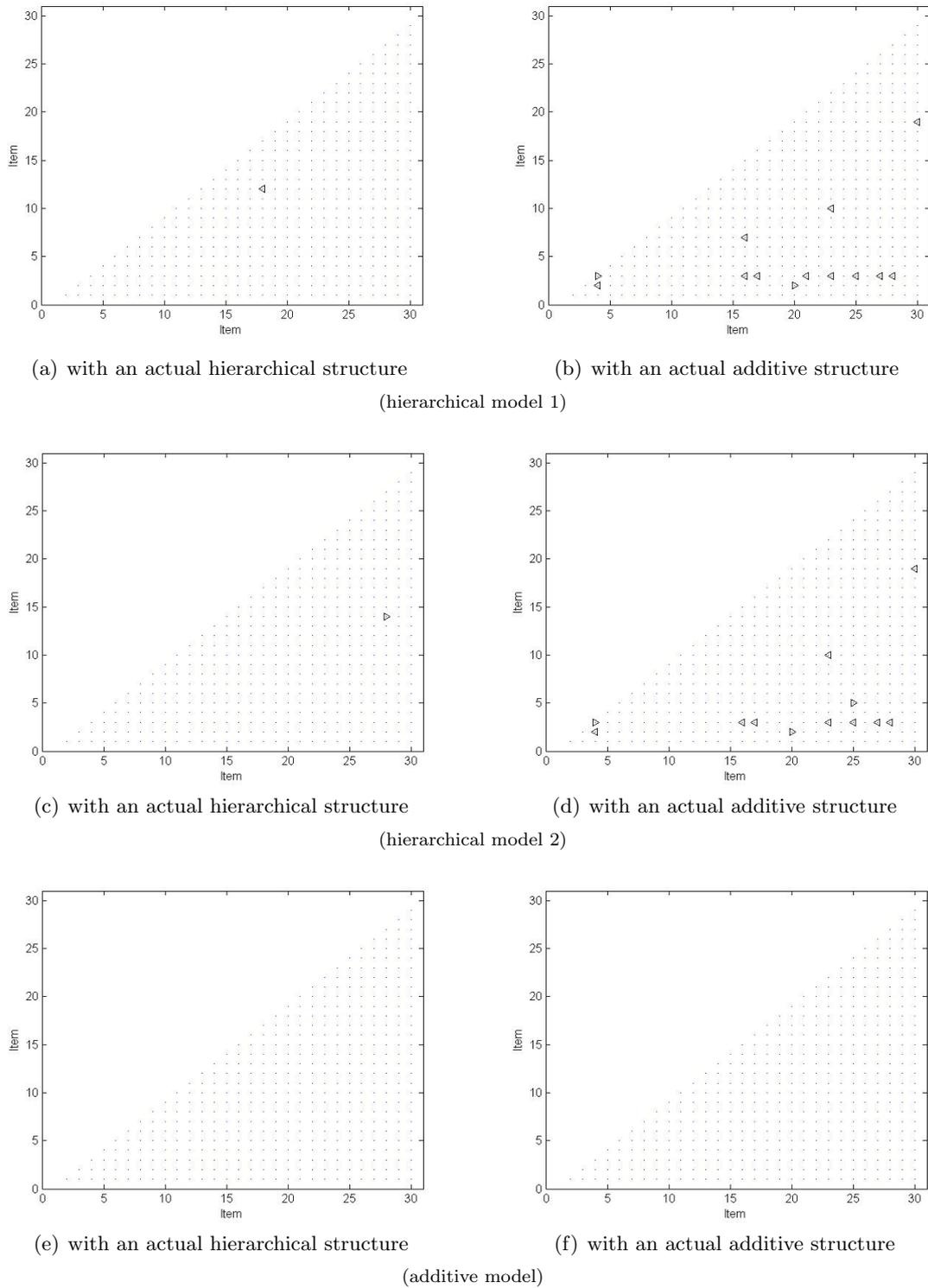


Figure 2: Plots of PPP-values for odds ratios for the three MIRT models fit to multidimensional data involving general and specific traits. (Triangles represent extreme PPP-values, where values smaller than 0.005 are denoted using left triangle signs and values larger than 0.995 are denoted using right triangle signs.)

with a threshold of 0.01, the plots indicate that the additive model had relatively fewer number of extreme predicted odds ratios and hence provided a better description than the hierarchical models in the two situations considered. On the other hand, the overwhelmingly larger numbers of extreme PPP-values clearly suggest the lack of fit of the hierarchical models when the actual structure was additive (see Figures 2(b), 2(d)). The two hierarchical models performed fairly similarly using PPMC. These results agree with those using DIC.

Hence, the model comparison results based on the Bayesian deviance and PPMC criteria suggest that the additive model performs better than the hierarchical models in a wide range of test situations and even when the actual latent structure is hierarchical. This is consistent with findings from Sheng and Wikle (2009). The fact that the additive model works better even when the hierarchical model is true can be explained by mathematically showing that the latter is a special case of the former. Indeed, we can rewrite (2) and (3) as $\theta_{vi} = \delta_v \theta_{0i} + \epsilon_{vi}$ and $\theta_{0i} = \sum \lambda_v \theta_{vi} + \epsilon_i$, where $\epsilon_{vi} \sim N(0, 1)$ and $\epsilon_i \sim N(0, 1)$; replace θ_{vi} in the hierarchical model so that the systematic component in (1) is $\alpha_{vj} \delta_v \theta_{0i} + \alpha_{vj} \epsilon_{vi} - \beta_{vj}$ or $\frac{\alpha_{vj}}{\lambda_v} (\theta_{0i} - \epsilon_i) - \frac{\alpha_{vj}}{\lambda_v} \sum_{v' \neq v} \lambda_{v'} \theta_{v'i} - \beta_{vj}$; and show that each of these two forms can be viewed as a constrained version of the additive model as defined in (4). This is similar to Yung *et al.*'s (1999) finding that the second-order factor model is hierarchically nested within the bifactor model.

6.3. Empirical example

A subset of the *College Basic Academic Subjects Examination* (CBASE; Osterlind 1997) English data was further used to illustrate the Bayesian approach for model choice or model checking. The data contain independent binary responses of 1,200 college students to a total of 41 English multiple-choice items. The English test is further organized into levels of increasing specificity by two subtests, namely, writing and reading, so that 16 items are in one subtest and 25 are in the other. One may assume that the test measures an overall English dimension with two specific trait dimensions. Model comparison is hence necessary for establishing the model that provides a relatively better representation of the data. The three MIRT models incorporating both general and specific traits were then each fit to the CBASE data using Gibbs sampling with a run length of 10,000 iterations and a burn-in period of 5,000, which was sufficient for the chains to converge. As expected, Bayesian deviance and PPMC, displayed in Table 9 and Figure 3, respectively, pointed to the more complicated additive model as it had a smaller Bayesian DIC estimate and fewer number of extreme PPP-values for odds ratios. Given the observations made from the simulation study in Section 6.2, comparisons of the DIC and p_D estimates using Bayesian deviances or that of the extreme PPP-values using PPMC suggest that the actual latent structure for the CBASE data is closer to be additive, with the general and specific traits being moderately correlated.

	\bar{D}	$D(\bar{\vartheta})$	p_D	DIC
hierarchical model 1	53444.5661	51985.0857	1459.4805	54904.0466
hierarchical model 2	53480.1684	52057.2671	1422.9013	54903.0698
additive model	52836.1005	51207.4303	1628.6702	54464.7707

Table 9: Bayesian deviance estimates for the three MIRT models with the CBASE data.

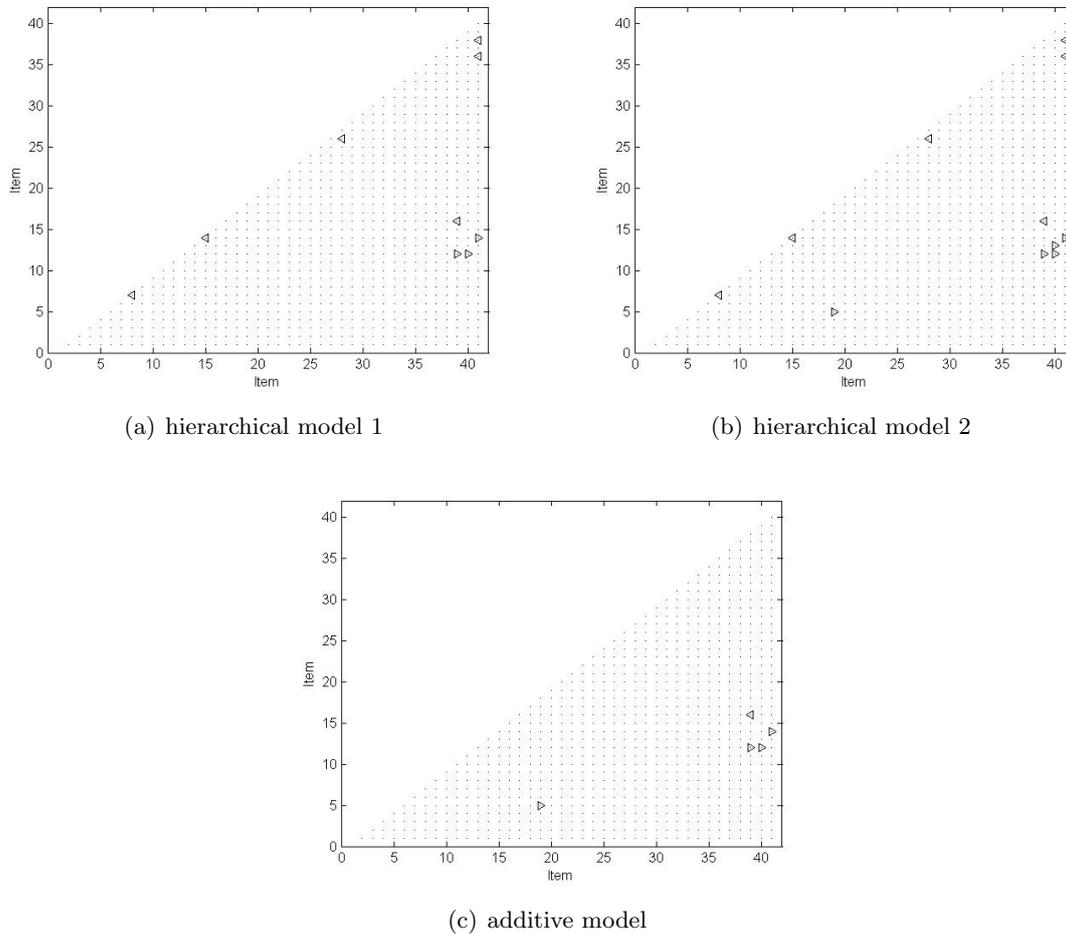


Figure 3: Plots of PPP-values for odds ratios for the three MIRT models with the *CBASE* data. (Triangles represent extreme PPP-values, where values smaller than 0.005 are denoted using left triangle signs and values larger than 0.995 are denoted using right triangle signs.)

7. Discussion

With functions for generating dichotomous response data from and implementing the Gibbs sampler for the 2PNO MIRT models that incorporate a general trait and several specific trait dimensions, **IRTm2noHA** allows the user the choice to set the number of total or burn-in samples, specify starting values or prior distributions for model parameters, check convergence of the Markov chain, as well as obtain Bayesian model choice or model checking statistics. The package leaves it to the user to choose between uniform and conjugate priors for the item parameters and certain hyperparameters. In addition, the user can choose to set the location and scale parameters for the conjugate normal priors of α_{vj} , β_{vj} , and/or α_{0vj} to reflect different prior beliefs on their distributions. For example, if there is a stronger prior opinion that the item intercept parameters in the additive model should be close to 0, a smaller $\sigma_{\beta_v}^2$ can be specified in the `gsm2noHA` function such that `xvar = [ones(m+1,1), ones(m+1,1), 0.5*ones(m+1,1)]`. If, however, this prior opinion concerns with the item intercepts in

the first subtest, not the entire test, the input argument becomes `xvar = [ones(m+1,1), ones(m+1,1), [0.5; ones(m,1)]]`. This way, different prior distributions can be specified for α_{vj} , β_{vj} , or α_{0vj} across the m subtests.

The intertrait correlations can be estimated fairly accurately using `gsm2noHA`. In the examples of parameter recovery in Section 6.1, ρ_{12} was set to be 0.5 for the hierarchical model 2, and it was estimated to be 0.5505 assuming uniform priors and 0.5554 assuming informative priors. Similarly, for the additive model, the actual ρ_{01} , ρ_{02} , and ρ_{12} parameters were 0.4, 0.4, and 0.8, respectively. Their posterior estimates were 0.3865, 0.4615, and 0.8148, respectively, assuming uniform priors, and 0.3728, 0.4758, and 0.8173, respectively, assuming informative priors. Hence, in addition to item or person parameters, the package provides a good estimation of the actual interrelations among latent traits provided that the latent structure is specified correctly. As illustrated, the additive model allows the general trait to be correlated with each specific trait. However, high correlations between them would give rise to collinearity and create problem with the Gibbs sampler.

One should note that during an implementation of the Gibbs sampler, if a Markov chain does not converge within a run length of certain iterations, additional iterations can be obtained by invoking the `gsm2noHA` function with starting values `th0`, `item0`, `sigma0`, and/or `b0` set to be their respective posterior samples drawn on the last iteration of the Markov chain (see Sheng 2008a, for a demonstration of such procedure).

The illustrative examples provided in Section 6 deal with two subtests for simplicity. On a 2.8GHz Intel dual core processor with 2GB of RAM, the execution time for the three examples is about 2-3 minutes, 5.5-6 minutes and 10.5-11 minutes, respectively. For tests with three or more subtests, `IRTm2noHA` can be used in a similar fashion with possibly increased complexity and consequently a longer computing time. In addition, Bayesian deviance and PPMC are adopted in this paper to evaluate the model-data fit of a candidate model. One may also want to consider Bayes factors, which provide more reliable and powerful results for model comparisons in the Bayesian framework. However, they are difficult to calculate due to the difficulty in exact analytic evaluation of the marginal density of the data (Kass and Raftery 1995) and hence are not considered in the paper. Finally, this paper adopts the Gelman-Rubin R statistic to assess convergence. Its multivariate extension, the Brooks-Gelman multivariate potential scale reduction factor (Brooks and Gelman 1998), may be considered as well.

Acknowledgments

The author would like to thank the two anonymous reviewers for their valuable comments and suggestions.

References

- Albert JH (1992). "Bayesian Estimation of Normal Ogive Item Response Curves Using Gibbs Sampling." *Journal of Educational Statistics*, **17**, 251–269.
- Anderson CJ, Li Z, Vermunt JK (2007). "Estimation of Models in a Rasch Family for Poly-

- tomous Items and Multiple Latent Variables.” *Journal of Statistical Software*, **20**(6), 1–36. URL <http://www.jstatsoft.org/v20/i06/>.
- Bafumi J, Gelman A, Park DK, Kaplan N (2005). “Practical Issues in Implementing and Understanding Bayesian Ideal Point Estimation.” *Political Analysis Advance Access*, **13**, 171–187.
- Bezruckzo N (ed.) (2005). *Rasch Measurement in Health Sciences*. JAM Press, Maple Grove, MN.
- Birnbaum A (1968). “The Logistic Test Model.” In F Lord, M Novick (eds.), *Statistical Theories of Mental Test Scores*, pp. 397–423. Reading, Mass: Addison-Wesley Publishing Co.
- Birnbaum A (1969). “Statistical Theory for Logistic Mental Test Models with a Prior Distribution of Ability.” *Journal of Mathematical Psychology*, **6**, 258–276.
- Bock RD, Aitkin M (1981). “Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM Algorithm.” *Psychometrika*, **46**, 443–459.
- Boomsma A, van Duijn MAJ, Snijders TAB (eds.) (2001). *Essays on Item Response Theory*. Springer-Verlag, New York.
- Brooks SP, Gelman A (1998). “General Methods for Monitoring Convergence of Iterative Simulations.” *Journal of Computational and Graphical Statistics*, **7**, 434–455.
- Carlin BP, Louis TA (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. 2nd edition. Chapman & Hall, London.
- Casella G, George EI (1992). “Explaining the Gibbs Sampler.” *The American Statistician*, **46**, 167–174.
- Chang CH, Reeve BB (2005). “Item Response Theory and Its Applications to Patient-Reported Outcomes Measurement.” *Evaluation & the Health Professions*, **28**, 264–282.
- Embretson SE, Reise SP (2000). *Item Response Theory for Psychologists*. Lawrence Erlbaum Associates, Inc, New Jersey.
- Feske U, Kirisci L, Tarter RE, Plkonis PA (2007). “An Application of Item Response Theory to the DSM-III-R Criteria for Borderline Personality Disorder.” *Journal of Personality Disorders*, **21**, 418–433.
- Gelman A, Carlin JB, Stern HS, Rubin DB (2004). *Bayesian Data Analysis*. 2nd edition. Chapman & Hall/CRC, Boca Raton.
- Gelman A, Meng XL, Stern HS (1996). “Posterior Predictive Assessment of Model Fitness via Realized Discrepancies.” *Statistica Sinica*, **6**, 733–807.
- Gelman A, Rubin DB (1992). “Inference from Iterative Simulation Using Multiple Sequences.” *Statistical Science*, **7**, 457–511.

- Geman S, Geman D (1984). “Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images.” *IEEE Trans. Pattern Analysis and Machine Intelligence*, **6**, 721–741.
- Gibbons RD, Hedeker DR (1992). “Full-Information Item Bi-Factor Analysis.” *Psychometrika*, **57**, 423–436.
- Gustafsson J, Balke G (1993). “General and Specific Abilities as Predictors of School Achievement.” *Multivariate Behavioral Research*, **28**, 407–434.
- Holzinger KJ, Swineford F (1937). “The Bifactor Method.” *Psychometrika*, **2**, 41–54.
- Imbens GW (2000). “The Role of the Propensity Score in Estimating Dose-Response Functions.” *Biometrika*, **87**, 706–710.
- Johnson MS (2007). “Marginal Maximum Likelihood Estimation of Item Response Models in R.” *Journal of Statistical Software*, **20**(10), 1–24. URL <http://www.jstatsoft.org/v20/i10/>.
- Kass RE, Raftery AE (1995). “Bayes Factors.” *Journal of the American Statistical Association*, **90**, 773–795.
- Kolen MJ, Brennan RL (eds.) (1995). *Test Equating: Methods and Practices*. Springer-Verlag, New York.
- Lee H (1995). “Markov Chain Monte Carlo Methods for Estimating Multidimensional Ability in Item Response Analysis.” Unpublished Dissertation.
- Li Y, Bolt DM, Fu J (2006). “A Comparison of Alternative Models for Testlets.” *Applied Psychological Measurement*, **30**, 3–21.
- Lord FM (1980). *Applications of Item Response Theory to Practical Testing Problems*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- Mair P, Hatzinger R (2007). “Extended Rasch Modeling: The **eRm** Package for the Application of IRT Models in R.” *Journal of Statistical Software*, **20**(9), 1–20. URL <http://www.jstatsoft.org/v20/i09/>.
- McDonald RP (1999). *Test Theory: A Unified Approach*. Lawrence Erlbaum, Mahwah, NJ.
- Molenaar IW (1995). “Estimation of Item Parameters.” In GH Fischer, IW Molenaar (eds.), *Rasch Models: Foundations, Recent Developments, and Applications*, pp. 39–51. Springer-Verlag, New York.
- Mulaik SA, Quartetti DA (1997). “First Order or Higher Order General Factor?” *Structural Equation Modeling*, **4**, 193–211.
- Osterlind S (1997). *A National Review of Scholastic Achievement in General Education: How are We Doing and Why should We Care?*, volume 25 of *ASHE-ERIC Higher Education Report*. George Washington University Graduate School of Education and Human Development, Washington, DC.

- Patz RJ, Junker BW (1999a). “Applications and Extensions of MCMC in IRT: Multiple Item Types, Missing Data, and Rated Responses.” *Journal of Educational and Behavioral Statistics*, **24**, 342–366.
- Patz RJ, Junker BW (1999b). “A Straightforward Approach to Markov Chain Monte Carlo Methods for Item Response Models.” *Journal of Educational and Behavioral Statistics*, **24**, 146–178.
- Reckase MD (ed.) (2009). *Multidimensional Item Response Theory*. Springer-Verlag, New York.
- Rindskopf D, Rose T (1988). “Some Theory and Applications of Confirmatory Second-Order Factor Analysis.” *Multivariate Behavioral Research*, **23**, 51–67.
- Ripley BD (1987). *Stochastic Simulation*. John Wiley & Sons, New York.
- Rizopoulos D (2006). “**ltm**: An R Package for Latent Variable Modeling and Item Response Analysis.” *Journal of Statistical Software*, **17**(5), 1–25. URL <http://www.jstatsoft.org/v17/i05/>.
- Rubin DB (1984). “Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician.” *The Annals of Statistics*, **12**, 1151–1172.
- Sahu SK (2002). “Bayesian Estimation and Model Choice in Item Response Models.” *Journal of Statistical Computation and Simulation*, **72**, 217–232.
- Schmid J, Leiman JM (1957). “The Development of Hierarchical Factor Solutions.” *Psychometrika*, **22**, 53–61.
- Sheng Y (2008a). “Markov Chain Monte Carlo Estimation of Normal Ogive IRT Models in MATLAB.” *Journal of Statistical Software*, **25**(8), 1–15. URL <http://www.jstatsoft.org/v25/i08/>.
- Sheng Y (2008b). “A MATLAB Package for Markov Chain Monte Carlo with a Multi-dimensional IRT Model.” *Journal of Statistical Software*, **28**(10), 1–19. URL <http://www.jstatsoft.org/v28/i10/>.
- Sheng Y, Wikle CK (2007). “Comparing Multi-dimensional and Unidimensional IRT Models.” *Educational & Psychological Measurement*, **67**, 899–919.
- Sheng Y, Wikle CK (2008). “Bayesian Multi-dimensional IRT Models with a Hierarchical Structure.” *Educational & Psychological Measurement*, **68**, 413–430.
- Sheng Y, Wikle CK (2009). “Bayesian IRT Models Incorporating General and Specific Abilities.” *Behaviormetrika*, **36**, 27–48.
- Sinharay S, Johnson MS, Stern HS (2006). “Posterior Predictive Assessment of Item Response Theory Models.” *Applied Psychological Measurement*, **30**, 298–321.
- Sinharay S, Stern HS (2003). “Posterior Predictive Model Checking in Hierarchical Models.” *Journal of Statistical Planning and Inference*, **111**, 209–221.

- Spiegelhalter D, Best N, Carlin B, van der Linde A (2002). “Bayesian Measures of Model Complexity and Fit.” *Journal of the Royal Statistical Society B*, **64**, 583–640.
- Tanner MA, Wong WH (1987). “The Calculation of Posterior Distribution by Data Augmentation.” *Journal of the American Statistical Association*, **82**, 528–550.
- The MathWorks, Inc (2007). *MATLAB – The Language of Technical Computing, Version 7.5*. Natick, Massachusetts. URL <http://www.mathworks.com/products/matlab/>.
- Verdinelli I, Wasserman L (1995). “Computing Bayes Factors Using a Generalization of the Savage-Dickey Density Ratio.” *Journal of the American Statistical Association*, **90**, 614–618.
- Yung YF, Thissen D, McLeod LD (1999). “On the Relationship Between the Higher-Order Factor Model and the Hierarchical Factor Model.” *Psychometrika*, **64**, 113–128.

Affiliation:

Yanyan Sheng
Department of Educational Psychology & Special Education
Wham 223, Mail Code 4618
Southern Illinois University-Carbondale
Carbondale, IL 62901, United States of America
E-mail: ysheng@siu.edu