Published in final edited form as:
Psychometrika. 2017 June ; 82(2): 295–307. doi:10.1007/s11336-017-9564-y.

Dealing with reflection invariance in Bayesian factor analysis

Elena A. Erosheva,
Associate Professor of Statistics and Social Work, Department of Statistics, University of Washington, Box 354320, Seattle WA 98195.

S. McKay Curtis
Research Scientist, Division of General Internal Medicine, University of Washington, Box 359780, Seattle WA 98104.

Abstract

A well-known inherent ambiguity in factor models is that factors and factor loadings can only be identified up to an orthogonal rotation. This paper is concerned with a special case of rotations – reflections – that correspond to sign changes of the columns in the matrix of loadings, and the associated implications for Bayesian factor analysis. Using a classic data set as an example, we first demonstrate that naive implementation of rotational constraints can be problematic for Bayesian inference. Specifically, constraining some loadings to be one or positive to achieve identifiability may result in nontrivial multimodality. We present a simple approach for dealing with reflection invariance in Bayesian factor analysis. We recommend fitting Bayesian factor analysis models without rotational constraints on the loadings – allowing Markov chain Monte Carlo algorithms to explore the full posterior distribution – and then using a relabeling algorithm to pick a factor solution that corresponds to one mode. We demonstrate our approach on the case of a bifactor model, however, the relabeling algorithm is straightforward to transfer to other settings, and has already been applied in the literature for Bayesian estimation of other latent variable models.

Keywords

Identifiability constraints; label-switching; Markov chain Monte Carlo; relabeling; rotation; rotational invariance

1 Introduction

Factor analysis is an indispensable method in modern psychometrics. Although maximum likelihood estimation is most common in factor analysis, Bayesian estimation methods have their contributions. For example, in exploratory factor analysis, Martin and McDonald (1975) used Bayesian methods to circumvent the problem of Heywood cases. In confirmatory factor analysis (CFA), Lee (1981) used a Bayesian approach to account for different forms of prior information. More recently, advances in computing power and methodology have enabled researchers to take advantage of the benefits of fully Bayesian

elena@stat.washington.edu.

analyses that aim to obtain a joint posterior distribution for all parameters in a model. These benefits include incorporation of multiple sources of uncertainty in a single conceptual framework, an enhanced ability to check model fit, the ability to obtain parameter estimates of complex functions, and to develop model extensions in a straightforward way, including those for latent variable models (Scheines, Hoijtink, and Boomsma, 1999; Savitsky and McCaffrey, 2014).

A well-known ambiguity in factor models is that factors and factor loadings can only be identified up to an orthogonal rotation. This paper is concerned with a special case of rotations that are due to sign invariance of latent variables. Such reflections create 2^q symmetric local modes in the likelihood for a model with q factors. Reflection invariance is common in confirmatory factor analysis (CFA) where structural constraints are typically placed in such a way that the model is identified up to all possible column sign changes of the matrix of loadings (Jennrich, 1978). Any factor analysis estimation method must address reflection invariance in some way.

In maximum likelihood (ML) framework, one can often ignore reflection invariance by relying on optimization algorithms that converge to one mode in the likelihood that is nearest to the starting values. Alternatively, when a particular reflection mode is desirable, one could impose further rotational constraints, e.g., by setting one loading per factor to a fixed value. However, the implications of this approach are not fully understood. For example, Millsap (2001) demonstrates that different choices of apparently equivalent rotational constraints may not only lead to less transparently equivalent solutions, but may also create estimation and identification problems, and even lead to solutions with widely different model fit results. Additionally, Loken (2005) demonstrates computational problems in ML estimation for factor models when some parameters are fixed at non-zero values; he indicates that a Bayesian approach could provide a solution to this problem.

Modern Bayesian estimation is done via constructing Markov chain Monte Carlo (MCMC) algorithms that allow one to obtain sample draws from posterior distributions of model parameters. When MCMC moves between 2^q equivalent reflection modes, simple summaries of parameters such as the posterior mean or posterior standard deviation will be misleading. Even though this problem is widely recognized in the context of factor analysis (Jackman, 2001; Quinn, 2004; Lopes and West, 2004; Ghosh and Dunson, 2009; Loken, 2005; Nishihara, Minka, and Tarlow, 2013), there appears to be no universal solution. Below, we review existing approaches that have been suggested to address this problem.

First, in some data analyses, no special attention may be required because MCMC chains get stuck in a local mode nearest to the starting values. This can happen when data are highly informative about latent dimensions and posterior reflection modes are well-separated; see Jackman (2001, p.232) for a graphical illustration of the phenomenon. However, when data contain little information about some of the latent dimensions, the reflection modes may not be well-separated in the posterior (Jackman, 2001). In such cases, one can still avoid dealing with reflection invariance by relying on maximum a posteriori (MAP) estimators, a penalized maximum likelihood equivalent (Lee, 1981). However, while MAP point estimators allow researchers to incorporate their prior beliefs about parameter uncertainty, this approach lacks benefits of a fully Bayesian analysis and has been criticized for its inability to reflect true uncertainty in the inference (Jasra, Holmes, and Stephens, 2005).

Second, with a fully Bayesian approach, it has been suggested to use the Geweke and Zhou (1996) restrictions, or, alternatively, constrained MCMC analyses. The Geweke and Zhou (1996) restrictions fix all upper diagonal elements of the matrix of loadings to zero and constrain all diagonal elements to be positive (Lopes and West, 2004; Ghosh and Dunson, 2008, 2009). While these restrictions remove indeterminacy due to rotational and reflection invariance in exploratory factor analysis (Muirhead, 1982, Theorem A9.8, page 592), they cannot be readily applied to CFA because of the presence of structural constraints. On the other hand, constrained MCMC approaches can be carried out for both exploratory and confirmatory factor models. These approaches follow the practice from maximum likelihood estimation and incorporate additional identifiability restrictions that constrain MCMC draws to the vicinity of only one mode of the posterior distribution (Jackman, 2001; Quinn, 2004; Congdon, 2003, 2006). We illustrate problems with these approaches to CFA estimation on an example of bifactor models.

The remainder of this paper is organized as follows. In Section 2, we establish notation for factor models and briefly review current practices for placing constraints in CFA with maximum likelihood estimation. In Section 3, we implement similar rotational identifiability constraints for Bayesian CFA of bifactor models and illustrate the problems that arise when analyzing a classic data set from Holzinger and Swineford (1939). In Section 4, observing that reflections are akin to label-switching in mixture models, we develop a simple and elegant approach for dealing with reflection invariance in Bayesian factor analysis by drawing on an analogy with mixture models (Stephens, 2000). In essence, we recommend fitting Bayesian factor analysis models without rotational constraints on the loadings – allowing MCMC chains to explore the full posterior distribution – and then post-processing MCMC draws with a relabeling algorithm to pick a factor solution that corresponds to one mode. Finally, in Section 5, we illustrate our method on the Holzinger and Swineford (1939) data.

2 Preliminary considerations

Let $\mathbf{Y}_i$ denote p-dimensional vector of observable variables, and $\boldsymbol{\xi}_i$ denote q-dimensional vector of latent variables for individual i, $i = 1,\ldots,n$. Let $\mathrm{N}_p(\boldsymbol{\mu},\boldsymbol{\Psi})$ denote the p-dimensional normal density with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Psi}$. Consider a factor analysis model

$$\begin{aligned} \mathbf{Y}_i \big| \boldsymbol{\xi}_i &\sim \mathrm{N}_p\left(\boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\xi}_i, \boldsymbol{\Psi}\right) \\ \boldsymbol{\xi}_i &\sim \mathrm{N}_q(\mathbf{0}, \boldsymbol{\Phi}) \end{aligned} \quad (1)$$

where $\boldsymbol{\Lambda}$ is a matrix of factor loadings, and $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ are covariance matrices. The factor model in equation (1) implies that the marginal distribution of $\mathbf{Y}_i$ is

$$\mathbf{Y}_i \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}). \quad (2)$$

In CFA, structural constraints are placed on $\boldsymbol{\Lambda}$ that reflect substantive knowledge about the relationship between observed variables and latent factors. Typically, these substantive constraints require certain elements of loadings matrix $\boldsymbol{\Lambda}$ to be fixed to zero or, less frequently, to non-zero values. In this paper, we deal with the orthogonal factor model that restricts the factor variance matrix $\boldsymbol{\Phi}$ and the matrix of uniquenesses $\boldsymbol{\Psi}$ to be diagonal (Johnson and Wichern, 1998).

A general approach to placing rotational constraints for identifiability in maximum likelihood estimation of CFA is to fix $q(q-1)/2$ loadings to zero, satisfying certain rank conditions (Loken, 2005). In addition to specifying a pattern of zero loadings in the loading matrix $\boldsymbol{\Lambda}$, researchers typically use one of two approaches to establish a scale for the latent variables. The first is to fix the factor variances to one; the corresponding restrictions are the default restrictions in the software Latent Gold (Vermunt and Magidson, 2005). Solutions to CFA models under these restrictions can still be reflected via column sign changes in $\boldsymbol{\Lambda}$, although this reflection invariance does not generally hinder maximum likelihood estimation. If none of the loadings are fixed to a nonzero value, the likelihood will have 2^q equivalent modes corresponding to all possible column sign changes in $\boldsymbol{\Lambda}$. In this case, it is known that the 2^q solutions have the same interpretation, and maximum likelihood estimates behave nicely as long as the likelihood modes are well separated (Dolan and Molenaar, 1991).

The second approach establishes a scale for the latent variables by setting one loading in each column of the loading matrix $\boldsymbol{\Lambda}$ to be equal to 1 and allowing factor variances to be freely estimated. Unlike the first approach, in this case the restrictions are designed to completely remove indeterminacy due to reflection invariance. These restrictions are also common; they are the default in the Mplus software (Muthén and Muthén, 2007) and in the latent variable modeling package GLLAAM (Rabe-Hesketh, Skrondal, and Pickles, 2004).

3 Bayesian Factor Analysis

In this Section, we illustrate problems that arise when MCMC draws are constrained to be in the vicinity of one mode of the posterior distribution. Below, we demonstrate the challenges that arise in MCMC estimation of a bifactor model an a classic data set from Holzinger and Swineford (1939).

3.1 MCMC with rotational constraints

We distinguish two types of rotational constraints that, together with CFA constraints, are designed to completely remove indeterminacy associated with rotational and reflection invariance (Table 1). The *fixed value constraints* set one loading per factor to 1 and allow us to freely estimate the factor variances. These constraints mirror the common ML approach. The *positivity constraints* require one loading per factor to be positive and fix factor variances to 1; this approach was suggested for Bayesian CFA by Congdon (2003, 2006).

We use the following hierarchical Bayesian CFA model specification:

$$\begin{aligned}&\mathbf{Y}_i \sim \mathrm{N}_p\left(\boldsymbol{\Lambda}\boldsymbol{\xi}_i, \boldsymbol{\Psi}_p\right),\\&\boldsymbol{\xi}_i \sim \mathrm{N}_q\left(\mathbf{0}, \boldsymbol{\Phi}_q\right),\\&\boldsymbol{\mu} \sim \mathrm{N}_p\left(m_\mu, s_\mu^2\right),\\&\psi_{jj} \sim \mathrm{InvGam}\left(a_\psi, b_\psi\right),\\&\psi_{jj'} = 0, \quad \text{for } j \neq j',\end{aligned} \tag{3}$$

where $\mathbf{0}$ is a vector of all zeroes, $\boldsymbol{\Psi}_p$, $\boldsymbol{\Phi}_q$ are covariance matrices of appropriate dimensions, and InvGam(a_ψ, b_ψ) denotes the inverse gamma distribution with location a_ψ and scale b_ψ. In addition, for the bifactor model with *fixed value constraints*, we specify

$$\begin{aligned}&\lambda_{ij} = 1, \text{ for the first nonzero loading in each column of } \boldsymbol{\Lambda},\\&\lambda_{ij} \sim \mathrm{N}\left(m_\lambda, s_\lambda^2\right), \text{ for the remaining nonzero loadings},\\&\phi_{kk} \sim \mathrm{InvGam}\left(a_\phi, b_\phi\right); \quad \phi_{kk'} = 0 \text{ for } j \neq k'.\end{aligned} \tag{4}$$

Analogously, for the bifactor model with *positivity constraints*, we specify

$$\begin{aligned}&\lambda_{ij} \sim \mathrm{N}_{(0,\infty)}\left(m_\lambda, s_\lambda^2\right), \text{ for the first nonzero loading in each column of } \boldsymbol{\Lambda},\\&\lambda_{ij} \sim \mathrm{N}\left(m_\lambda, s_\lambda^2\right), \text{ for the remaining nonzero loadings},\\&\boldsymbol{\Phi}_q = \mathbf{I}_q,\end{aligned} \tag{5}$$

where $\mathrm{N}_{(0,\infty)}\left(m_\lambda, s_\lambda^2\right)$ is a normal density truncated at zero on the left, and $\mathbf{I}_q$ is the q-dimensional identity matrix. For all analyses in this paper, we place InvGam(1,1) priors on specific and factor variances by setting $a_\psi = 1.0$, $b_\psi = 1.0$, $a_\varphi = 1.0$, and $b_\varphi = 1.0$. We place normal priors on unconstrained loadings, and normal truncated at zero on loadings constrained to be positive, with $m_\lambda = 0.0$, $s_\lambda = 10.0$. For the prior on $\boldsymbol{\mu}$, we set $m_\mu = 0$ and $s_\mu = 1000$. We assess goodness of fit using an MCMC approximation to the loglikelihood ratio fit statistic (Bishop, Fienberg, and Holland, 1975; Bollen, 1989; Cressie and Read, 1989), defined as $G^2 = -2(\log L_0 - \log L_1)$. The two log-likelihoods are

$$\log L_0 = -(n/2)\log\left|\widehat{\boldsymbol{\Psi}}\right| - (n/2)\mathrm{tr}\left(\widehat{\boldsymbol{\Psi}}^{-1} S^*\right) - np\log(2\pi)/2$$

$$\log L_1 = -(n/2)\log|S^*| - (np)/2 - np\log(2\pi)/2$$

where $\widehat{\boldsymbol{\Psi}}$ is the covariance matrix implied by the CFA model, and $S^* = \frac{(n-1)}{n} S$ is the maximum likelihood estimate of the population covariance matrix S.

3.2 Holzinger Swineford Bifactor Data

We consider the bifactor model (Holzinger and Swineford, 1937; Gibbons and Hedeker, 1992), where each of the p response variables loads onto exactly two factors: a general factor, onto which all variables load, and a secondary factor, onto which only a subset of the variables loads. Holzinger and Swineford (1939) collected data from a sample of 7th and 8th grade students in Chicago area ($n = 301$). Students completed a battery of 24 tests covering five cognitive domains: spatial, verbal, memory, speed, and mathematical deduction. Assuming a classic bifactor structure, Holzinger and Swineford (1939) eventually dropped the mathematical deduction factor for items 20–24 from the model because they concluded – based on their initial analysis – that the mathematical deduction factor was not distinct from the general factor. Similarly to Holzinger and Swineford (1939), we use the bifactor structure that contains the general factor and the four secondary factors as shown in Table 2.

3.3 Constrained Analysis

We analyze the Holzinger and Swineford data under the rotational identifiability constraints from Table 1, fixing the first loading in each column of the loading matrix – specifically, loadings $\lambda_{1,1}$, $\lambda_{1,2}$, $\lambda_{5,3}$, $\lambda_{10,4}$, and $\lambda_{14,5}$ – to be either one or positive.

We explored three strategies for setting starting values for factor loadings: random on the real line, random but constrained to be positive, and PCA-generated. We used $N(0,10)$ to generate random starting values, and we used absolute values of the random draws from $N(10,3)$ to generate positive starting values. For each combination of a starting-value strategy and a constraint from Table 1, we obtained three parallel chains of 65,000 MCMC draws. The use of multiple chains and overdispersed starting values follows a standard practice in implementing MCMC (Gelman, Carlin, Stern, and Rubin, 2004, pp. 295–296). From every chain, we discarded 5,000 iterations as burn-in and kept only every 20th iteration for a simulation size of 3,000. We used WinBUGS (Lunn, Thomas, Best, and Spiegelhalter, 2000) code to implement the Gibbs sampler for CFA models with rotational constraints.

Table 3 provides a summary of our analyses. For each scenario defined by a starting value selection strategy and a constraint type, we report the number of chains converged, the best value of the MCMC approximation to the loglikelihood ratio fit statistic (Bishop et al., 1975; Bollen, 1989) among the converged chains, and the warning signs related to the quality of the corresponding factor model solutions. We observe that at least one chain in each scenario has failed to reach convergence based on the Geweke diagnostics (Geweke, 1992). The differences in the fit statistics indicate that the factor model solutions from the converged chains lack consensus across the scenarios of the constrained analysis.

To provide a graphical illustration, we present MCMC trace plots and the corresponding density plots from three scenarios.[1] Figure 1 shows trace plots and corresponding density plots from a chain with random starting values under the *fixed value constraints* for selected loadings on factors 1, 2, and 4. The locations of trace plots and density plots in Figure 1, and

[1]To reduce clutter, we plot only selected factor loadings.

other figures that follow, schematically represent the bifactor pattern as specified in the matrix of loadings. This chain has successfully converged as indicated by the Geweke diagnostics (Geweke, 1992), however, the posterior variance estimates for factors 1, 2, and 5 appear to be very close to the boundary of the parameter space with estimated posterior means of 0.56, 1.60, and 0.70, respectively (posterior variance estimates for the other two factors are about 30 and 110). Small factor variances bring up questions about the quality of this bifactor solution, which is also reflected in the relatively poor fit as indicated by the fit statistic in Table 3. If this particular constrained estimation scenario was implemented in isolation, further analyses that address the possibility of some factors being redundant would have been necessary. In addition, small estimates for factor variances may indicate that the resulting inference is sensitive to the choice of hyperparameters in the inverse-gamma prior distribution (Gelman, 2006; Gelman and Hill, 2007). We observed a similar situation with the *fixed value constraint* when starting values were positive.

Figure 2 shows trace plots and corresponding density plots from a chain obtained under the *positivity constraints* with random starting values that has converged successfully. Here, posterior estimates of factor loadings $\lambda_{11,4}$, $\lambda_{12,4}$, $\lambda_{13,4}$ are of opposite polarity with the positively constrained $\lambda_{10,4}$. This phenomenon is problematic because there were no reversely coded items in the data. The same phenomenon was observed when positive or PCA-motivated starting values were used under the *positivity constraints* (Table 3).

Finally, Figure 3 illustrates an obvious case of a mode-switching behavior for a chain with *fixed value constraints* and positive starting values. It is clear that the modes are not symmetric with respect to the origin and, hence, are not rotationally equivalent.

To summarize, this analysis illustrates that ensuring unique identification of a CFA model by requiring selected loadings to be one or positive may not lead to a satisfactory solution. Our analysis of the Holzinger and Swineford data with rotational constraints has encountered several problems such as convergence failures, mode-switching behavior, estimated loadings of opposite polarity that are counter-intuitive for the data, and factor variance estimates that are close to zero. We note that these difficulties are not specific to the data set as we were able to replicate them on a simulation study (Erosheva and Curtis, 2011).

4 Relabeling Algorithm

4.1 MCMC without rotational constraints

We propose a two-step procedure for Bayesian inference in factor models that avoids placing rotational constraints and thereby avoids problems demonstrated in the previous section. The first step is to carry out a Bayesian CFA without additional constraints on the loadings, allowing MCMC chains to explore the full posterior distribution with 2^q equivalent modes. Here, we complete the Bayesian hierarchical CFA model from equation (3) by placing standard normal prior on unconstrained loadings with $m_\lambda = 0$ and $s_\lambda = 1$, and fixing factor variances to 1:

$$\lambda_{ij} \sim \mathrm{N}\left(m_\lambda, s_\lambda^2\right), \text{ for all nonzero loadings,} \quad (6)$$
$$\mathbf{\Phi}_q = \mathbf{I}_q .$$

This approach allows us to establish a scale for the latent factors but places no rotational constraints on factor loadings.

The next step is to use statistical decision theory to relabel signs of some of the MCMC draws of factor loadings so that all draws correspond to only one mode out of the 2^q equivalent modes in the posterior. This procedure is inspired by, and is analogous to, the use of relabeling in mixture models (Stephens, 2000). In the following, we first review basic concepts of statistical decision theory (Schervish, 1995), and then derive a relabeling algorithm for posterior draws of factor loadings from Bayesian CFA with no rotational constraints.

4.2 Statistical Decision Theory

Statistical decision theory is a formal method for choosing an optimal action or decision based on observed data. Assume a loss function $\mathscr{L}(\mathbf{a};\boldsymbol{\theta})$ assigns a penalty to each vector of actions $\mathbf{a}$ from a set of all possible actions $\mathscr{A}$ for a given true parameter vector $\boldsymbol{\theta}$. For example, an action could be a choice of a point estimate $\hat{\boldsymbol{\theta}}$ and the loss function could be a penalty for picking estimates that are different from the true value $\boldsymbol{\theta}$. Since the true vector of parameters $\boldsymbol{\theta}$ is unknown, the optimal action should consider the loss for all possible values of $\boldsymbol{\theta}$. In Bayesian decision theory, this is accomplished by computing the *posterior risk*,

$$\mathscr{R}(\mathbf{a}) = \mathrm{E}[\mathscr{L}(\mathbf{a};\boldsymbol{\theta}) \,|\, \mathbf{y}] \quad (7)$$
$$= \int \mathscr{L}(\mathbf{a};\boldsymbol{\theta}) p(\boldsymbol{\theta} \,|\, \mathbf{y}) d\boldsymbol{\theta}$$

which is the expected loss for taking action $\mathbf{a}$ over the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$. The optimal action $\mathbf{a}^*$ is the action which minimizes the posterior risk. In most practical cases, the integral in (7) is obtained via a Monte Carlo approximation

$$\mathscr{R}(\mathbf{a}) \approx \mathscr{R}_{\mathrm{MC}}(\mathbf{a}) \quad (8)$$
$$= \frac{1}{N} \sum_{t=1}^{N} \mathscr{L}\left(\mathbf{a};\boldsymbol{\theta}^{(t)}\right),$$

where $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(N)}$ are N draws from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$.

4.3 Relabeling for Confirmatory Factor Analysis

In the context of confirmatory factor analysis with q factors, allowing for reflections will result in 2^q maxima in the complete posterior distribution of the loadings that correspond to all possible column sign changes. Analogously to the relabeling algorithm introduced by Stephens (2000) for the label-switching problem in mixture models, we derive a relabeling algorithm for CFA to produce a factor solution corresponding to one mode.

Assume the parameter vector $\boldsymbol{\theta}$ is the matrix $\Lambda = \{\lambda_{ij}\}$ of true factor loadings for some fixed mode. Define the action **a** to be a choice of a vector (**m**,**s**) of means and variances of the normal densities for the loadings' posterior distribution. We need to define a loss function to penalize choosing values of (**m**, **s**) that would make it unlikely to observe Λ. However, due to possibility of reflections, for a given draw of loadings, we first need to define a vector of sign change parameters $\nu = (\nu_1, \ldots, \nu_q)$, where $\nu_k \in \{-1,1\}$ for all k, that will allow us to choose a single mode of the posterior. We define the loss function as follows

$$\mathscr{L}((\mathbf{m}, \mathbf{s}); \boldsymbol{\Lambda}) = \min_{\nu}\left\{-\sum_{i=1}^{p}\sum_{j=1}^{q} 1\left(\lambda_{ij} \neq 0\right)\log\left[f_N\left(\boldsymbol{\nu}_j\lambda_{ij}; m_{ij}, s_{ij}^2\right)\right]\right\}, \quad (9)$$

where $f_N(\cdot; m, s^2)$ is the density of a normal distribution with mean m and variance s^2, and $\mathbf{1}(\lambda_{ij} \neq 0)$ is an indicator function that equals 0 when loading λ_{ij} is constrained to be 0, and equals 1 otherwise. The loss function is the negative log of independent normal densities for unrestricted loadings.

To develop the relabeling algorithm, let $\lambda_{ij}^{(t)}$ be the t^{th} random draw from the posterior distribution of λ_{ij}, and let $\boldsymbol{\nu}^{(t)}$ be a vector of sign change parameters at that draw. Given the loss function in equation (9), the Monte Carlo approximation to the posterior risk from equation (8) can be written as

$$\begin{aligned}
\mathscr{R}_{\text{MC}}(\mathbf{a}) &= \frac{1}{N}\sum_{t=1}^{N}\mathscr{L}\left((\mathbf{m}, \mathbf{s}); \boldsymbol{\Lambda}^{(t)}\right) \quad (10)\\
&= \frac{1}{N}\sum_{t=1}^{N}\min_{\boldsymbol{\nu}(t)}\left\{-\sum_{i=1}^{p}\sum_{j=1}^{q} 1\left(\lambda_{ij}^{(t)} \neq 0\right)\log\left[f_N\left(\nu_j^{(t)}\lambda_{ij}^{(t)}; m_{ij}, s_{ij}^2\right)\right]\right\}\\
&= \min_{\boldsymbol{\nu}^{(1)}, \ldots, \boldsymbol{\nu}^{(N)}}\left\{-\frac{1}{N}\sum_{t=1}^{N}\sum_{i=1}^{p}\sum_{j=1}^{q} 1\left(\lambda_{ij}^{(t)} \neq 0\right)\log\left[f_N\left(\nu_j^{(t)}\lambda_{ij}^{(t)}; m_{ij}, s_{ij}^2\right)\right]\right\}.
\end{aligned}$$

To implement the relabeling procedure for Bayesian factor analysis, start by assigning initial values to the vectors of sign changes $\boldsymbol{\nu}^{(1)}, \ldots, \boldsymbol{\nu}^{(N)}$. For example, one could set them all to q-length vectors of 1s which corresponds to no sign changes. Next, iterate through the following two steps until values of the vectors $\boldsymbol{\nu}^{(1)}, \ldots, \boldsymbol{\nu}^{(N)}$ do not change:

1. Find the values of (**m**, **s**) that minimize $\mathscr{R}_{\text{MC}}(\mathbf{a})$ (equation (10)), conditional on the current values of the sign change parameters $\boldsymbol{\nu}^{(1)}, \ldots, \boldsymbol{\nu}^{(N)}$.

2. Find values of $\boldsymbol{\nu}^{(1)}, \ldots, \boldsymbol{\nu}^{(N)}$ that minimize $\mathscr{R}_{\mathrm{MC}}(\mathbf{a})$ (equation (10)), conditional on the values of $(\mathbf{m}, \mathbf{s})$ from the previous step.

In contrast to standard decision theory applications, we are not only interested in the optimal action ($\mathbf{m}^*$, $\mathbf{s}^*$) but also in the value of $\boldsymbol{\nu}^*$ at the optimal action. Note that the choice of the loss function as defined by equation (9) assigns higher penalties to those values of $\boldsymbol{\nu}$ that make $\{\nu_j\lambda_{ij}\}$ to be unlikely realizations from the normal distribution with the specified mean and variance parameters. As a result, the relabeling algorithm obtains a sequence of labels for MCMC draws that is likely to correspond to only one local maximum of the posterior distribution out of possible 2^q maxima that are due to reflection invariance.

The above relabeling algorithm proceeds in the same manner as in the mixture-model case (Stephens, 2000), except that the minimization over the sign change vectors $\boldsymbol{\nu}^{(1)}, \ldots, \boldsymbol{\nu}^{(N)}$ is computationally simpler than the minimization over all possible permutations of mixture component labels. For this reason, our relabeling algorithm has the same convergence properties as the relabeling algorithm for mixture models. Because the number of possible permutations for the vector of sign changes is finite and $\mathscr{R}_{\mathrm{MC}}(\mathbf{a})$ decreases at each iteration, the above algorithm is guaranteed to find a fixed point. There is no guarantee, however, that the solution attained will be the global optimal solution, as is the case with all hill climbing optimization algorithms.

We note that, in simple cases when 2^q equivalent modes are clearly separated, the relabeling of MCMC draws from different chains could be achieved by applying a sign alignment method, similar to those used in jackknife exploratory factor analysis (Pennell, 1972; Clarkson, 1979). Note that this would only be possible when each MCMC sampler visited exactly one of the 2^q equivalent modes. In such cases, while the sampler may pick out the means from the data, the failure to visit different modes of the posterior would still be an undesirable behavior as it indicates the inability to fully explore posterior distribution. Jasra et al. (2005) provide an example of an MCMC sampler stuck in one mode in case of a mixture of normals, and discuss implications of such behavior. They point out that good properties of MCMC samplers become especially important when the posterior distribution is genuinely multimodal, i.e., when there are multiple models that can not be explained by sign changes of factor loadings. In cases when posterior modes are not well-separated and mode-switching is present, MCMC draws are not amenable to a simple sign alignment. The relabeling algorithm that we propose, on the contrary, remains valid. Moreover, the algorithm is still valid when not all the items are positively scored, and the opposite polarity of scoring does not impact the algorithm's complexity.

5 Bifactor Analysis with Relabeling

We use the Holzinger and Swineford (1939) data to carry out the Bayesian CFA as described in Section 4, restricting only the factor variances to be fixed at 1 and placing no additional constraints on the loadings apart from the structural zeros. As before, we use 3 parallel chains with random starting values, with 65,000 draws per chain. Calculations were performed in a C function that was called from R (R Development Core Team, 2010). From every chain, we discard the first 5,000 iterations as burn-in and keep only every 20th iteration for a simulation size of 3,000. Finally, we apply the relabeling algorithm from Section 4 to the matrix of posterior draws of factor loadings.

Figure 4 shows posterior density plots for several selected factor loadings before and after relabeling: six loadings for the general factor, and two loadings each for the spatial and speed secondary factors.[2] Before relabeling (Figure 4, left panel), posterior density plots for loadings $\lambda_{1,2}$ and $\lambda_{4,2}$ exhibit multiple modes that are symmetric about the origin. The relabeling of the posterior draws removes the symmetric secondary modes in the marginal posterior distributions (Figure 4, right panel), making the relabeled draws amenable to standard summaries. Note also that the resulting posterior densities of the loadings may assign a non-zero probability mass to both the positive and the negative region of the real line (e.g., $\lambda_{1,2}$ in Figure 4), however, the posterior modes of all relabeled factor loadings are of the same polarity. The signs of all loadings on $q = 5$ factors are positive in our case, but could have had any combination of 2^5 sign combinations for each factor for other runs of the relabeling algorithm, depending on our specification of the starting values for sign changes. The vertical lines in the right panel of Figure 4 show point estimates of respective factor loadings obtained by maximizing the likelihood (with factor variances constrained to 1). ML estimates were obtained using the quasi-Newton algorithm as implemented in the R function optim (Byrd, Lu, Nocedal, and Zhu, 1995). Multiple starting values had to be used to ensure that the Hessian matrix at the solution was positive definite and to increase the probability of finding the global maximum.

The Geweke convergence diagnostics applied to relabeled draws indicate that all three chains have converged successfully. In addition, the effective sample sizes for the model parameters from the relabeled chains showed a substantial improvement over those obtained using the constrained chains described in Section 3.3. The MCMC approximations to the loglikelihood ratio fit statistic for the three chains were 341.070, 340.745, and 341.096, respectively. The corresponding fit statistic for the maximum likelihood solution was 340.712. This indicates that the model fit after relabeling is consistent across MCMC chains, is much better as compared to the factor solutions found in our constrained analyses (see Table 3), and is comparable to the fit of the maximum likelihood solution. By using our Bayesian estimation approach with relabeling, we were able to obtain consistent fit results across MCMC chains without the need to decide on the type of constraints or to select custom starting values.

6 Conclusion

This paper presents a simple and elegant approach for dealing with reflection invariance in Bayesian factor analysis and illustrates it on the example of a bifactor model. It is worth pointing out that while the necessity of Bayesian methods may not be obvious for simpler examples such as those discussed in this paper, Bayesian methods undoubtedly prevail when models are complex (Lee, 2007; Scheines et al., 1999). For example, with longitudinal data, the high-dimensional integrals may become intractable analytically while limited information estimation methods may not be efficient (Joe and Maydeu-Olivares, 2011). A

[2]We present two chains out of three to reduce clutter.

Bayesian framework, on the other hand, is readily extendable to more complex cases. However, any Bayesian model that contains a factor structure as a building block must still address the issue of identifiability. The relabeling algorithm that we present in this paper can be readily incorporated in Bayesian model fitting procedures for complex models.

The implications of constrained maximum likelihood estimation in factor models are not fully understood. In simple factor models, when each variable loads on exactly one factor, different sets of constraints necessarily result in solutions that are equivalent with respect to model fit. However, when some variables load on more than one factor, merely changing the choice of the variables for placing fixed values can lead to inability to obtain maximum likelihood estimates due to convergence failures or to factor solutions from different equivalence classes (Millsap, 2001). Such unforeseen impacts give a reason to avoid unnecessary constraints in Bayesian – as well as in likelihood-based – estimation.

Identifiability constraints that are aimed at resolving multimodality due to label-switching have been scrutinized in the literature on Bayesian analysis of mixture models (Celeux, Hurn, and Robert, 2000; Stephens, 2000; Jasra et al., 2005). In this paper, we have demonstrated that the practice of setting up rotational constraints – by imposing *fixed value* or *positivity constraints* on some loadings – may not provide a satisfactory solution for Bayesian CFA. This could be due to the constrained posterior becoming oddly-shaped and difficult to explore fully with standard MCMC; or due to potential appearance of local modes that correspond to parts of the suitably relabeled posterior distribution where signs of the loadings are different from their signs at the chosen mode; or due to changes in the constrained likelihood such as those observed by Millsap (2001). It is important to point out that while careful researchers may uncover problems with Bayesian inference in simpler cases, the impact of additional constraints on the posterior distribution in complex cases may not be as obvious because of potential model misspecification. This is another reason why unnecessary identifiability constraints are to be avoided.

We suggest estimate Bayesian CFA models without additional identifiability constraints on the loadings and use the relabeling algorithm for postprocessing MCMC draws. The decision-theoretic interpretation provides a solid theoretical foundation for relabeling algorithms (Stephens, 2000). The relabeling algorithm is guaranteed to find a fixed point that corresponds to a local minima of the expected loss. Intuitively, the relabeling algorithm derived in this paper obtains a sequence of sign changes for MCMC draws such that the marginal posterior distribution of the parameters is as close as possible to normal distribution, providing researchers with draws that correspond to one out of 2^q reflection modes.

Compared with maximum likelihood methods (Dolan and Molenaar, 1991), the Bayesian estimation with relabeling does not require the likelihood modes to be well separated (Jasra et al., 2005). We found the relabeling algorithm to be particularly useful for CFA models because the identifiability restrictions of Geweke and Zhou (1996), commonly employed for Bayesian estimation of exploratory factor analysis, cannot be applied in the presence of structural zeros. The relabeling algorithm, however, is not limited to CFA and can also be used for other models that exhibit multimodalities in the likelihood due to sign invariance of latent variables. For example, Savitsky and McCaffrey (2014) incorporate the relabeling algorithm in exploratory factor analysis as part of a Bayesian multivariate hierarchical model for ordered outcomes; Gruhl, Erosheva, and Crane (2013) use the relabeling algorithm in Bayesian estimation of a semiparametric bifactor model for mixed outcomes; Tanaka (2013) and Matlosz (2013) employ the relabeling algorithm with Bayesian multidimensional scaling of partial rank and ordinal preference data, respectively. The aforementioned studies reference the original technical report (Erosheva and Curtis, 2011) that contains simulation studies and the code for Bayesian analysis of bifactor models with relabeling. The code is also freely available in R package relabeLoadings.

Acknowledgments

This research was supported by grant R01 AG029672-01A1 from the National Institutes on Health. The authors are grateful to Thomas Richardson, Adrian Raftery, Peter Hoff and Jonathan Gruhl for helpful discussions, and to Terrance Savitsky for useful comments on an earlier draft of the paper and on the R code.

References

Bishop Y, Fienberg SE, and Holland P (1975), Discrete Multivariate Analysis: Theory and Practice, The MIT press.

Bollen KA (1989), Structural Equations with Latent Variables, New York: John Wiley & Sons, Inc.

Byrd RH, Lu P, Nocedal J, and Zhu C (1995),“A limited memory algorithm for bound constrained optimization,” SIAM Journal on Scientific Computing, 16, 1190–1208.

Celeux G, Hurn M, and Robert CP (2000), “Computational and inferential difficulties with mixture posterior distributions,” Journal of the American Statistical Association, 95, 957–970.

Clarkson DB (1979), “Estimating the Standard Eoors of Rotated Factor Loadings by jackknifing,” Psychometrika, 44, 297–314.

Congdon P (2003), Applied Bayesian Modelling, Wiley.

— (2006), Bayesian Statistical Modelling, John Wiley & Sons.

Cressie N and Read T (1989), “Pearson’s χ^2 and the loglikelihood ration statistic G^2 - A comparative review,” International Statistical Review, 57, 19–43.

Dolan CV and Molenaar PC (1991), “A comparison of four methods of calculating standard errors of maximum-likelihood estimates in the analysis of covariance structure,” British Journal of Mathematical and Statistical Psychology, 44, 359–368.

Erosheva EA and Curtis SM (2011), “Dealing with rotational invariance in Bayesian confirmatory factor analysis,” Tech. Rep. 589, University of Washington.

Gelman A (2006), “Prior distributions for variance parameters in hierarchical models,” Bayesian Analysis, 1, 515–533.

Gelman A, Carlin JB, Stern HS, and Rubin DB (2004), Bayesian Data Analysis, New York: Chapman & Hall/CRC.

Gelman A and Hill J (2007), Data Analysis Using Regression and Multilevel & Hierarchical Models, Cambridge.

Geweke J (1992), “Evaluating the accuracy of sampling-based approaches to calculating posterior moments,” in Bayesian Statistics 4, eds. Bernardo JM, Berger JO, Dawid AP, and Smith AFM, Oxford, UK: Clarendon Press, pp. 169–194.

Geweke J and Zhou G (1996),“Measuring the pricing error of the arbitrage pricing theory,” Review of Financial Studies, 9, 557–587.

Ghosh J and Dunson D (2008), “Bayesian model selection in factor analytic models,” in Random Effect and Latent Variable Model Selection, Springer-Verlag, pp. 151–163.

— (2009),“Default prior distributions and efficient posterior computation in Bayesian factor analysis,” Journal of Computational and Graphical Statistics, 18, 306–320. [PubMed: 23997568]

Gibbons RD and Hedeker DR (1992), "Full-information item bi-factor analysis," Psychometrika, 57, 423–436.

Gruhl J, Erosheva EA, and Crane PK (2013), "A semiparametric approach to mixed outcome latent variable models: Estimating the association between cognition and regional brain volumes," The Annals of Applied Statistics, 7, 2361–2383.

Holzinger KJ and Swineford F (1937),"The bi-factor method,"Psychometrika, 2, 41–54.

— (1939), A Study in Factor Analysis: The Stability of a Bi-factor Solution, no. 48 in Supplementary Educational Monographs, University of Chicago.

Jackman S (2001), "Multidimensional analysis of roll call data via Bayesian simulation: identification, estimation, inference, and model checking," Political Analysis, 9, 227–241.

Jasra A, Holmes C, and Stephens D (2005), "Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling," Statistical Science, 20, 50–67.

Jennrich RI (1978), "Rotational equivalence of factor laoding matrices with specified values," Psychometrika, 43, 421–426.

Joe H and Maydeu-Olivares A (2011),"A general family of limited information goodness-of-fit statistics for multinomial data," Psychometrika, 75, 393–419.

Johnson RA and Wichern DW (1998), Applied Multivariate Statistical Analysis, Prentice Hall.

Lee S-Y (1981),"A Bayesian approach to confirmatory factor analysis,"Psychometrika, 46, 153–160, 10.1007/BF02293896.

— (2007), Structural Equation Modeling: A Bayesian Approach, West Sussex, England: John Wiley & Sons.

Loken E (2005), "Identification constraints and inference in factor analysis models," Structural Equation Modeling, 12, 232–244.

Lopes HF and West M (2004),"Bayesian model assessment in factor analysis,"Statistica Sinica, 14, 41–67.

Lunn DJ, Thomas A, Best N, and Spiegelhalter D (2000), "WinBUGS–A Bayesian modelling framework: Concepts, structure, and extensibility," Statistics and Computing, 10, 325–337.

Martin J and McDonald R (1975), "Bayesian estimation in unrestricted factor analysis: A treatment for heywood cases," Psychometrika, 40, 505–517, 10.1007/BF02291552.

Matlosz K (2013),"Bayesian Multidimensional Scaling Model for Ordinal Preference Data," Ph.D. thesis, Columbia University.

Millsap RE (2001), "When trivial constraints are not trivial: The choice of uniqueness constraints in confirmatory factor analysis," Structural Equation Modeling, 8, 1–17.

Muirhead RJ (1982), Aspects of Multivariate Statistical Theory, John Wiley & Sons.

Muthén LK and Muthén BO (2007), Mplus: Statistical Analysis with Latent Variables, Los Angeles, CA.

Nishihara R, Minka T, and Tarlow D (2013),"Detecting Parameter Symmetries in Probabilistic Models," arXiv preprint arXiv:1312.5386.

Pennell R (1972), "Routinely computable confidence intervals for factor loadings using the 'Jack-knife'," British Journal of Mathematical and Statistical Psychology, 25, 107–114.

Quinn KM (2004),"Bayesian factor analysis for mixed ordinal and continuous responses," Political Analysis, 12, 338–353.

R Development Core Team (2010), R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Rabe-Hesketh S, Skrondal A, and Pickles A (2004), GLLAMM Manual.

Savitsky TD and McCaffrey DF (2014),"Bayesian hierarchical multivariate formulation with factor analysis for nested ordinal data," Psychometrika, 79, 275–302. [PubMed: 24659372]

Scheines R, Hoijtink H, and Boomsma A (1999), "Bayesian estimation and testing of structural equation models," Psychometrika, 64, 37–52.

Schervish MJ (1995), Theory of Statistics, New York: Springer.

Stephens M (2000),"Dealing with label switching in mixture models,"Journal of the Royal Statistical Society, Series B, 62, 795–809.

Tanaka K (2013),“A Bayesian Multidimensional Scaling Model for Partial Rank Preference Data,” Ph.D. thesis, Columbia University.

Vermunt JK and Magidson J (2005), Technical Guide for Latent GOLD 4.0: Basic and Advanced, Statistical Innovations Inc.

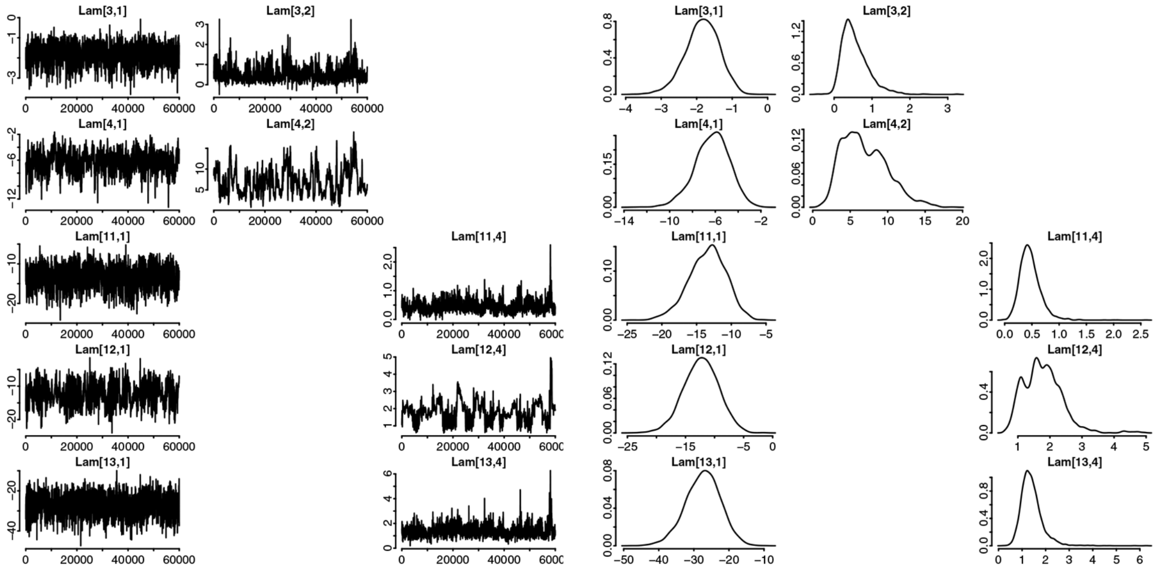

Figure 1:

MCMC trace plots (left panel) and posterior density plots (right panel) for selected factor loadings from the five-factor model with fixed value constraints. Initial values for loadings were generated randomly on the real line. This chain has passed Geweke convergence diagnostics.

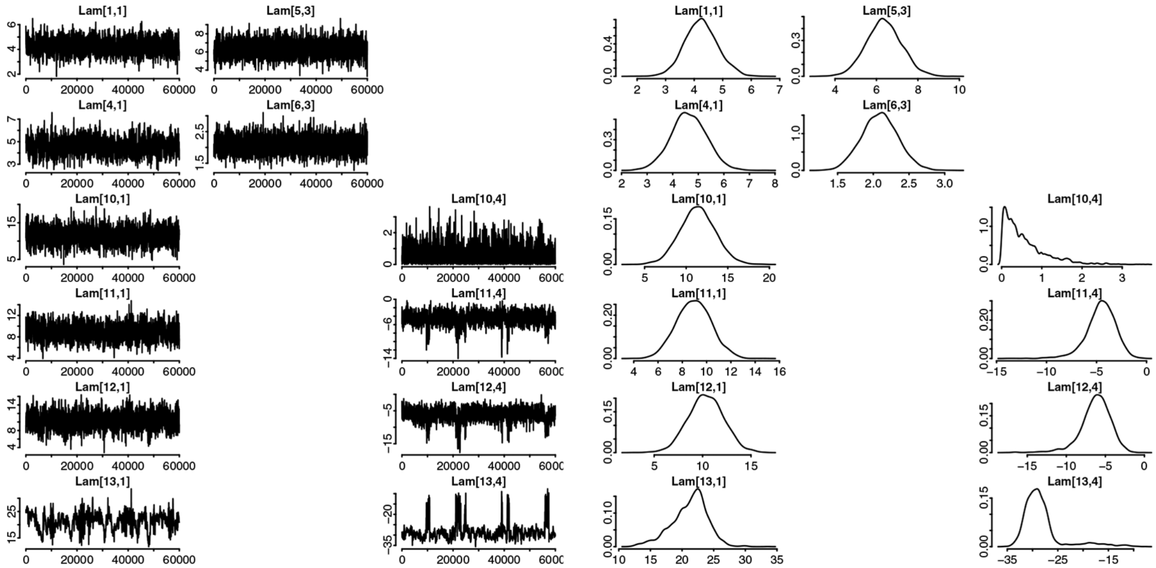

Figure 2:

MCMC trace plots (left panel) and posterior density plots (right panel) for selected factor loadings from the five-factor model with positivity constraints on the first loading in each column of the loading matrix. Initial values for loadings were generated randomly on the real line. This chain has passed Geweke convergence diagnostics.

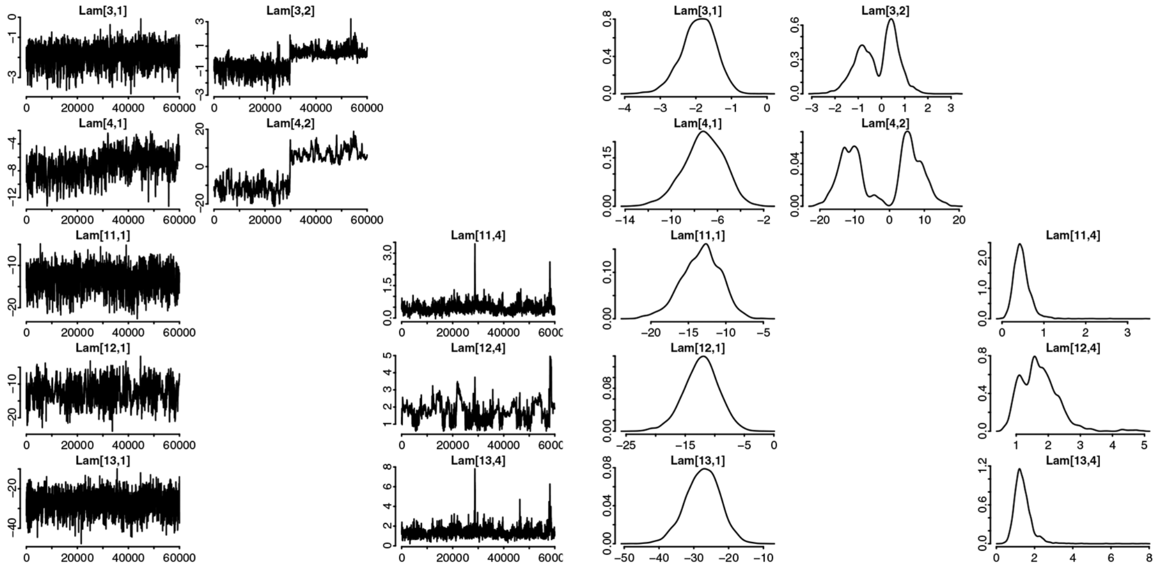

Figure 3:

MCMC trace plots (left panel) and posterior density plots (right panel) for selected factor loadings from the five-factor model with fixed-value constraints on the first loading in each column of the loading matrix. Initial values were generated randomly, but were constrained to be positive.

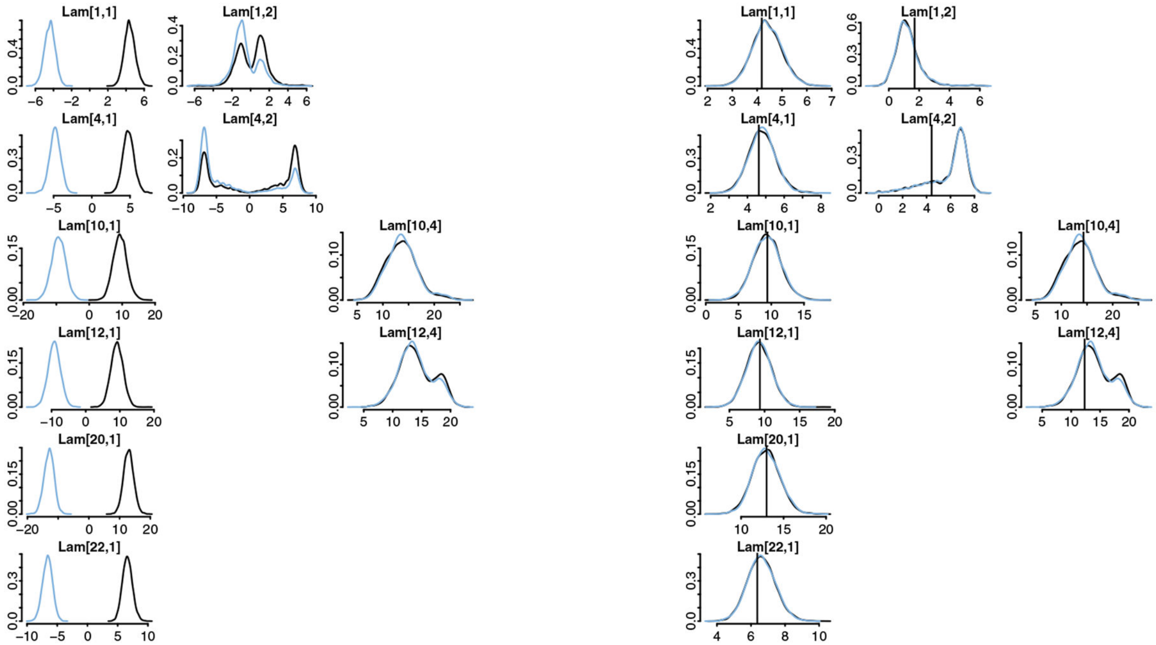

Figure 4:

Holzinger and Swineford data. Marginal posterior density plots of factor loadings for five-factor bifactor model from two parallel MCMC chains. Left panel: Selected density plots before relabeling. Right panel: Selected density plots after relabeling. Vertical lines provide the corresponding MLEs.

Table 1:

CFA rotational identifiability constraints.

	Restrictions on	
Constraints	**loadings**	**factor variances**
Fixed value	one $\lambda_{jk} = 1$ per factor	none
Positivity	one $\lambda_{jk} > 0$ per factor	=1

Table 2:

Secondary factor structure for the bifactor model of Holzinger and Swineford (1939).

Items	Secondary Factor
1–4	Spatial
5–9	Verbal
10–13	Speed
14–19	Memory
20–24	(None)

Author Manuscript

Author Manuscript

Author Manuscript

Table 3:

Summary of constrained MCMC analyses of Holzinger and Swineford data. For each scenario, 3 MCMC chains were obtained. “Converged” reports the number of chains that have converged. “Best G^2” reports the MCMC approximation to the best loglikelihood ratio fit statistic among converged chains.

Starting values	Constraint type	Converged	Best G^2	Notes
Random	Fixed value	2	4040.882	small factor variance
Positive	Fixed value	1	4040.909	small factor variance
PCA-generated	Fixed value	0	NA	NA
Random	Positivity	2	6214.451	opposite polarity
Positive	Positivity	2	2716.817	opposite polarity
PCA-generated	Positivity	2	2858.398	opposite polarity