PSYCHOMETRIKA—VOL. 82, NO. 2, 295–307
JUNE 2017
DOI: 10.1007/s11336-017-9564-y

DEALING WITH REFLECTION INVARIANCE IN BAYESIAN FACTOR ANALYSIS

ELENA A. EROSHEVA AND S. MCKAY CURTIS

UNIVERSITY OF WASHINGTON

This paper considers the reflection unidentifiability problem in confirmatory factor analysis (CFA) and the associated implications for Bayesian estimation. We note a direct analogy between the multimodality in CFA models that is due to all possible column sign changes in the matrix of loadings and the multimodality in finite mixture models that is due to all possible relabelings of the mixture components. Drawing on this analogy, we derive and present a simple approach for dealing with reflection in variance 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 generalize for handling multimodalities due to sign invariance in the likelihood in other factor analysis models.

Key words: identifiability constraints, label-switching, Markov chain Monte Carlo, relabeling, rotation, rotational invariance.

1. Introduction

This paper is concerned with a particular identifiability issue that is due to sign invariance of the latent variables in factor analysis models. Such reflection invariance creates multimodalities in the likelihood—2^q symmetric local modes for a q-factor model—that presents difficulties for Bayesian estimation via Markov chain Monte Carlo (MCMC). In this paper, we note a direct analogy between the reflection invariance in factor analysis and the relabeling invariance in finite mixture models (Stephens, 2000). Building on this analogy, we propose a relabeling approach for Bayesian factor analysis. On the example of a bifactor confirmatory factor analysis (CFA) model, we illustrate difficulties in MCMC estimation that are associated with reflection invariance and present our solution to this problem.

Bayesian methods have become increasingly popular for complex latent variable models (e.g., Scheines, Hoijtink, & Boomsma, 1999; Savitsky & McCaffrey, 2014). Bayesian factor analysis provides researchers with a vehicle that naturally allows building upon core model elements (Geweke & Zhou, 1996; Lopes & West, 2004; Rowe, 2001; Ghosh & Dunson, 2008, 2009; Peeters, 2012). The early work of Martin and McDonald (1975) introduced the use of Bayesian methods for exploratory factor analysis to circumvent the problem of Heywood cases. Lee (1981) illustrated a Bayesian approach to obtaining point estimates in CFA by specifying different forms of prior distributions.

This paper is concerned with modern Bayesian estimation which is done via constructing MCMC algorithms to obtain sample draws from posterior distributions of model parameters. Several authors have expressed warnings for using MCMC inference when likelihood surface has 2^q equivalent modes that correspond to all possible reflections of the matrix of loadings (Jackman, 2001; Quinn, 2004; Lopes & West, 2004; Ghosh & Dunson, 2009; Loken, 2005; Peeters, 2012; Nishihara, Minka, & Tarlow, 2013). When an MCMC algorithm moves between different modes

Correspondence should be made to Elena A. Erosheva, Department of Statistics, University of Washington, Box 354320, Seattle, WA 98195, USA. Email: erosheva@u.washington.edu

of such a posterior, simple summaries of parameters such as the posterior mean or posterior standard deviation will be misleading.

Several approaches have been suggested in Bayesian factor analysis to address the problem of reflection identifiability in MCMC. First, in some cases, the lack of identification due to reflection invariance may require no special attention due to MCMC chains being stuck in a local mode nearest to the starting values. This happens when data are highly informative about latent dimensions which corresponds to reflection modes in the posterior distribution being well separated; see Jackman (2001, p. 232) for a graphical illustration of the phenomenon. Such clear separation assures that the probability of an MCMC sampler jumping from one reflection mode to another is extremely low. Approaches such as careful selection of starting values (Peeters, 2012) or informative priors (Jackman, 2001; Bafumi, Gelman, Park, & Kaplan 2005) have been suggested to encourage MCMC samplers to get stuck in a well-separated region of the posterior density. These approaches, however, are not guaranteed to work (Bafumi et al., 2005).

Second, Geweke and Zhou (1996) addressed the general rotational identifiability issue in Bayesian exploratory q-factor models by fixing $q(q-1)/2$ upper diagonal elements of the matrix of loadings to zero and constraining q diagonal elements to be positive; we will refer to this set of constraints as GZ (identifiability) constraints. Constraining the upper diagonal elements to zero removes the indeterminacy due to rotational unidentifiability, while constraining the diagonal elements to be positive removes the indeterminacy due to reflection unidentifiability (Muirhead, 1982 , Theorem A9.8, p. 592). The GZ restrictions have become de facto constraints in Bayesian exploratory factor analysis applications (Lopes & West, 2004; Ghosh & Dunson, 2008, 2009). However, we note that the GZ constraints cannot be readily applied to confirmatory factor analysis (CFA) for two reasons. First, the patterns of zeros reflecting substantive considerations—such as zero loadings specified in a bifactor model—will, in most cases, prevent us from using the Geweke and Zhou (1996) estimation algorithm that does not allow for any other restrictions on the loadings apart from fixing all upper diagonal elements to zero and constraining the diagonal elements to be positive. Second, the GZ constraints can induce unintended prior information via declaring a preferred variable ordering for selecting the first q variables (Peeters, 2012, p. 30).

Finally, incorporating identifiability constraints used in maximum likelihood (ML) estimation into MCMC algorithms has been suggested as a means of resolving reflection invariance in Bayesian factor analysis (Jackman, 2001; Quinn, 2004; Congdon, 2003, 2006). Broadly speaking, such strategies are designed to limit the exploration of posterior distribution to only one mode out of 2^q possible maxima.

In this paper, we make two contributions to the literature. First, we show that borrowing restriction placement practices from ML can be problematic for Bayesian CFA estimation. Specifically, we observe that limiting MCMC to only one mode may result in nontrivial multimodality with inferior modes and mode-switching behavior in Bayesian factor analysis. Second, building on the analogy between the reflection invariance problem in factor analysis and the label-switching problem in mixture models, we propose an approach for dealing with reflection invariance in Bayesian factor analysis (Stephens, 2000). In essence, we recommend fitting Bayesian factor analysis models without rotational constraints on the loadings, which allows MCMC chains to explore the full posterior distribution, and post-processing MCMC draws with a relabeling algorithm to pick a factor solution that corresponds to one mode.

2. Relabeling Algorithm for Bayesian Factor Analysis

We consider a general CFA q-factor model for a p-dimensional outcome on individual k:

$$\mathbf{Y}_k \sim \mathsf{N}_p\left(\boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\xi}_k, \boldsymbol{\Psi}_p\right), \quad \text{for } i = 1, \ldots, K,$$

where $\boldsymbol{\mu}$ is the population mean, $\boldsymbol{\Lambda} = \{\lambda_{ij}\}$ is a $p \times q$ matrix of factor loadings, with some loadings fixed to zero according to substantive CFA constraints, $\boldsymbol{\xi}_k$ is a vector of latent factors, and $\boldsymbol{\Psi}_p$ is a p-dimensional diagonal matrix of variances. In this section, we derive a relabeling algorithm for $\{\lambda_{ij}^{(t)},\ t = 1, \ldots, N\}$ posterior draws of factor loadings from any Bayesian factor analysis MCMC sampler. The relabeling algorithm results in a sequence of MCMC draws that correspond to only one local maximum of the posterior distribution out of possible 2^q maxima that are due to reflection invariance.

For factor loadings at some fixed mode, define parameter $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_j, \ldots, \nu_q)$, where $\nu_j \in \{-1, 1\}$ is the sign change for all loadings on the jth latent factor. Let $(\mathbf{m}, \mathbf{s})$ denote a vector of means and variances of the normal densities for the loadings' posterior distribution. According to principles of Bayesian decision theory (e.g., Schervish, 1995), we define a loss function to penalize values of $(\mathbf{m}, \mathbf{s})$ that make it unlikely to observe $\boldsymbol{\Lambda}$. We choose the loss function to be the negative log of independent normal densities for unrestricted loadings:

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

where $f_N(\cdot; m, s^2)$ is the density of a normal distribution with mean m and variance s^2, and $\mathbf{1}\big(\lambda_{ij} \neq 0\big)$ is the indicator for loadings that are not substantively restricted to be zero.

Let $\lambda_{ij}^{(t)}$ be the tth 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 Eq. (1), the Monte Carlo approximation to the posterior risk is

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

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 $(\mathbf{m}, \mathbf{s})$ that minimize $\mathcal{R}_{\mathrm{MC}}((\mathbf{m}, \mathbf{s}))$ (Eq. 2), 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 $\mathcal{R}_{\mathrm{MC}}((\mathbf{m}, \mathbf{s}))$ (Eq. 2), 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 loss function as defined by Eq. (1) assigns higher penalties to those values of $\boldsymbol{\nu}$ that make relabeled loadings $\{\nu_j^{(t)} \lambda_{ij}^{(t)}\}$ to be unlikely realizations from the normal distribution with the specified mean and variance parameters, therefore encouraging unimodality.

TABLE 1.
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)

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 $\mathcal{R}_{\mathrm{MC}}((\mathbf{m}, \mathbf{s}))$ 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.

3. Bayesian CFA: Bifactor model

In this section, we illustrate our relabeling algorithm on a bifactor model using a classical data set from Holzinger and Swineford (1939). We first demonstrate that using current practices from ML estimation for placing identifiability constraints in MCMC estimation of factor analysis models can result in poor estimation. We then compare the constrained MCMC estimation results to our results with relabeling and to standard ML estimation results.

3.1. Holzinger and Swineford Bifactor Data

For demonstration purposes, we focus on a specific CFA model—the bifactor model (Holzinger & Swineford, 1937; Gibbons & Hedeker, 1992). In a bifactor model, 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 two different Chicago area schools—Grant-White Elementary School in Forest Park, Illinois ($n = 145$ students), and Pasteur Elementary School in Chicago ($n = 156$ students). Students completed a battery of 24 tests covering five cognitive domains—spatial, verbal, memory, speed, and mathematical deduction. Starting with a classic bifactor structure, Holzinger and Swineford (1939) eventually dropped the secondary 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 1. Items that do not load onto a secondary factor have the corresponding loadings set to zero in the bifactor model; we refer to these zero-set loadings as structural (or substantive) constraints.

3.2. MCMC for Bifactor Models

A general approach to placing rotational constraints for identifiability in ML estimation of CFA is to fix $q(q-1)/2$ loadings to zero, satisfying certain rank conditions (Loken, 2005). In

TABLE 2.
CFA rotational identifiability constraints.

Constraints	Restrictions on	
	Loadings	Factor variances
Fixed value	One $\lambda_{jk} = 1$ per factor	None
Positivity	One $\lambda_{jk} > 0$ per factor	$=1$

addition to specifying a pattern of zero loadings in the loading matrix $\mathbf{\Lambda}$, two approaches are typically used to establish a scale for the latent variables. The first is to fix the factor variances to one. The corresponding identifiability constraints are the default option in commercial software Latent Gold (Vermunt & Magidson, 2005). If none of the loadings is fixed to a nonzero value, the likelihood will have 2^q equivalent modes corresponding to all possible column sign changes in $\mathbf{\Lambda}$. This multimodality generally does not hinder ML estimation (Jennrich, 1978); the 2^q solutions have the same interpretation, and ML estimates behave nicely as long as the likelihood modes are well separated (Dolan & Molenaar, 1991). An analogous approach in Bayesian factor analysis estimation fixes the factor variances to one and requires MCMC draws for one loading per factor to be positive in order to constrain MCMC sampler from moving among the 2^q equivalent modes (Congdon, 2003, 2006).

The second approach used in ML estimation establishes a scale for the latent variables by setting one loading in each column of the loading matrix $\mathbf{\Lambda}$ to be equal to 1 and allowing factor variances to be freely estimated. Unlike fixing the variances as in the first approach, here the identifiability restrictions are designed to completely remove indeterminacy due to reflection invariance. These restrictions are also commonly used; they are the default in `Mplus` software (Muthén & Muthén, 2005) and in `GLLAAM`, a latent variable modeling package for `Stata` (Rabe-Hesketh & Skrondal, 2008). Analogous constraints can be used in Bayesian estimation. We summarize the two types of identifiability constraints—referred as the fixed value and the positivity constraints—in Table 2.

3.2.1. Model Specification Let $\mathbf{Y}_k$ denote a vector of p manifest variables and let $\boldsymbol{\xi}_k$ denote a vector of q latent factors for individual k. The hierarchical Bayesian CFA model is:

$$
\begin{aligned}
\mathbf{Y}_k &\sim \mathsf{N}_p\big(\boldsymbol{\mu} + \mathbf{\Lambda}\boldsymbol{\xi}_k, \mathbf{\Psi}_p\big), \quad \text{for } k = 1, \dots, K,\\
\boldsymbol{\xi}_k &\sim \mathsf{N}_q\big(\mathbf{0}, \mathbf{\Phi}_q\big),\\
\boldsymbol{\mu} &\sim \mathsf{N}_p\Big(m_\mu, s_\mu^2\Big),\\
\psi_{ii} &\sim \mathsf{InvGam}\big(a_\psi, b_\psi\big), i = 1, \dots, p,\\
\psi_{ii'} &= 0, \quad \text{for } i \neq i';\ i, i' = 1, \dots, p,
\end{aligned}
\tag{3}
$$

where $\mathbf{\Lambda}$ is a matrix of factor loadings with substantive zeros reflecting the bifactor structure from Table 1, $\mathbf{0}$ is a vector of all zeroes, and the factor variance matrix $\mathbf{\Phi}_q$ and the uniquenesses matrix $\mathbf{\Psi}_p$ are diagonal of dimensions q and p, respectively.

In addition, for the bifactor model with fixed value identifiability constraints, we specify

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

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

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

Finally, for the MCMC without additional identifiability constraints on the loadings (to be relabeled later), we complete the model from Eq. (3) with

$$\begin{aligned} &\lambda_{ij} \sim \mathsf{N}\left(m_\lambda, s_\lambda^2\right), \ \text{for all nonzero loadings,} \\ &\boldsymbol{\Phi} = \mathbf{I}. \end{aligned} \tag{6}$$

For all analyses in this paper, we place $\mathsf{InvGam}(1, 1)$ priors on specific and factor variances by setting $a_\psi = 1.0$, $b_\psi = 1.0$, $a_\phi = 1.0$, and $b_\phi = 1.0$. We chose the first nonzero loadings in each colums $\lambda_{1,1}$, $\lambda_{1,2}$, $\lambda_{5,3}$, $\lambda_{10,4}$, and $\lambda_{14,5}$ for placing identifiability constraints from Eqs. 4 and 5. 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 used WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000) code to implement the Gibbs sampler for the models with additional identifiability constraints (Eqs. 5 and 6) and implement the Gibbs sampler of Lee (2007, section 4.3.1) in a C function for the unconstrained case (Eq. 6).

We assess goodness of fit using an MCMC approximation to the loglikelihood ratio fit statistic (Bishop, Fienberg, & Holland, 1975; Bollen, 1989; Cressie & Read, 1989), defined as $G^2 = -2(\log L_0 - \log L_1)$. The two loglikelihoods are

$$\begin{aligned} \log L_0 &= -(n/2)\log|\hat{\Psi}| - (n/2)\mathrm{tr}(\hat{\Psi}^{-1}S^*) - np\log(2\pi)/2 \\ \log L_1 &= -(n/2)\log|S^*| - (np)/2 - np\log(2\pi)/2 \end{aligned}$$

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

We explored three strategies for setting starting values for factor loadings: random, random but constrained to be positive, and PCA-generated starting values. For each starting-value strategy and for each constraint from Table 2, 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, & Rubin, 2004, pp. 295–296). From every chain, we discarded 5000 iterations as burn-in and kept every twentieth iteration for a simulation size of 5000.

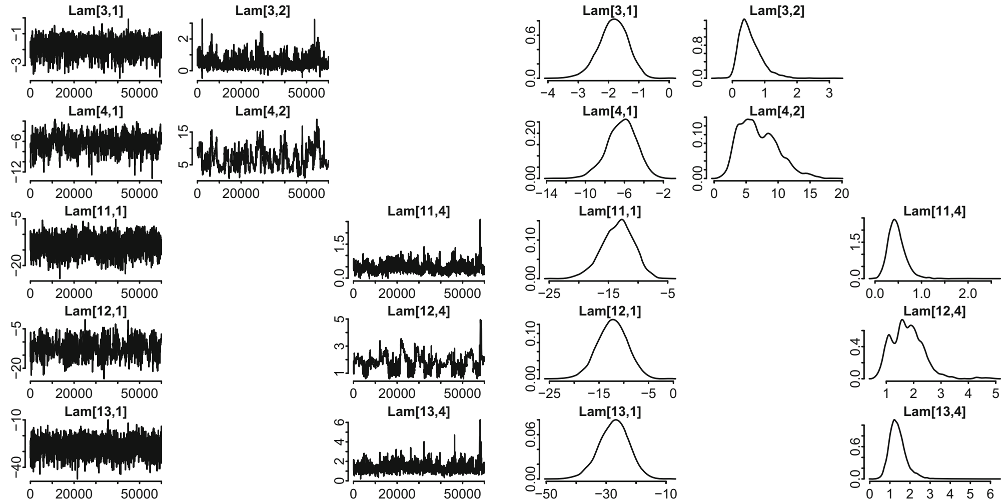

FIGURE 1.
MCMC trace plots (*left panel*) and posterior density plots (*right panel*) for selected factor loadings from the five-factor model with fixed value constraints. Initial values for loadings were generated randomly. This chain has passed Geweke convergence diagnostics.

3.3. Results: Holzinger and Swineford Bifactor Data

3.3.1. Results from Constrained MCMC Samplers We examined MCMC trace plots and the corresponding density plots for the nine scenarios. We present several illustrative figures below.[1] Figure 1 shows trace plots and corresponding density plots from a chain with random starting values under the fixed value constraint for selected loadings on factors 1, 2, and 4. The location of trace plots and density plots in this and other figures that follow schematically represents the bifactor pattern of structural zero constraints in the matrix of loadings. This chain has successfully converged as indicated by the Geweke convergence diagnostics (Geweke, 1992); however, the posterior variance estimates for factors 1, 2, and 5 are 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. If this particular constrained estimation scenario was implemented in isolation, further analyses that address the possibility of some factors being redundant given the data 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 & Hill, 2007). We observed a similar behavior for the fixed value constraint and positive starting values.

Figure 2 corresponds to 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 is problematic because there were no reversely coded items in the data. Using positive or PCA starting values also resulted in estimated loadings of opposite polarity: negative loadings on factors 1 and 3 conflict with the positively constrained loadings $\lambda_{1,1}$ and $\lambda_{5,3}$ (not shown).

Finally, Fig. 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.

[1] To reduce clutter, we plot only selected factor loadings that illustrate our points.

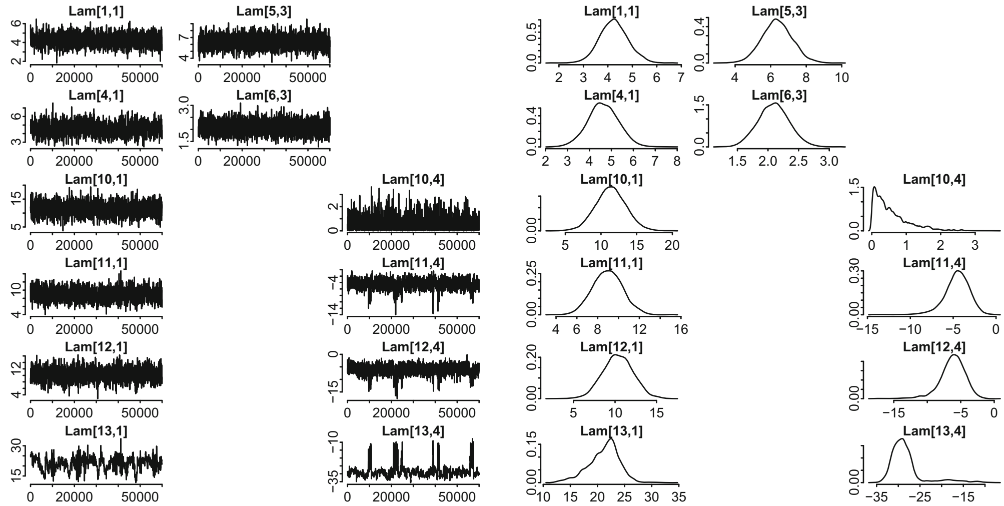

FIGURE 2.
MCMC trace plots (*left panel*) and posterior density plots (*right panel*) for selected factor loadings from the five-factor model with positivity constraints on the first loading in *each column* of the loading matrix. Initial values for loadings were generated randomly. This chain has passed Geweke convergence diagnostics.

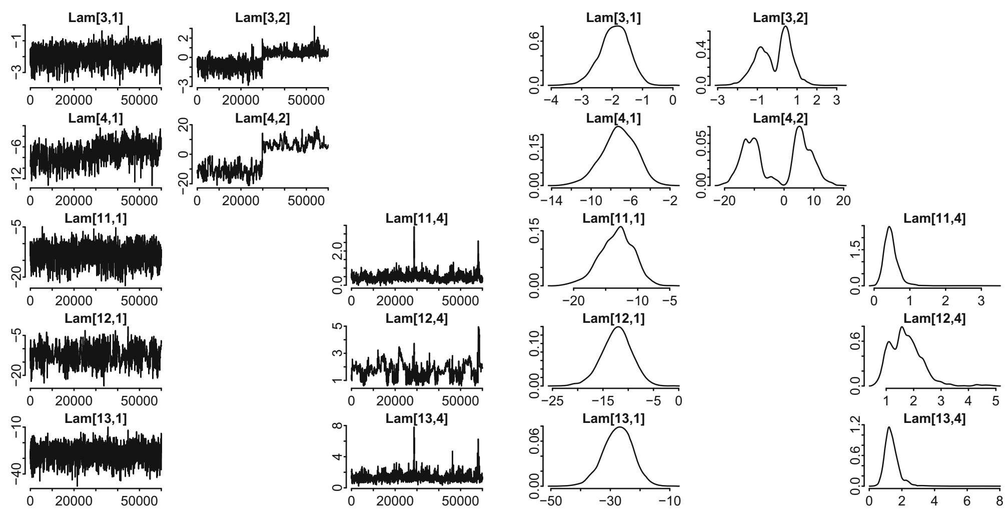

FIGURE 3.
MCMC trace plots (*left panel*) and posterior density plots (*right panel*) for selected factor loadings from the five-factor model with fixed value constraints on the first loading in *each column* of the loading matrix. Initial values were generated randomly, but were constrained to be positive.

Overall, the constrained MCMC samplers have either failed the convergence diagnostics or produced mode-switching behavior, counter-intuitive estimates of factor loadings, and estimates of factor variances that are close to zero. We note that these difficulties are not specific to the Holzinger and Swineford bifactor data as we were able to replicate them in a simulation study (Erosheva & Curtis, 2011).

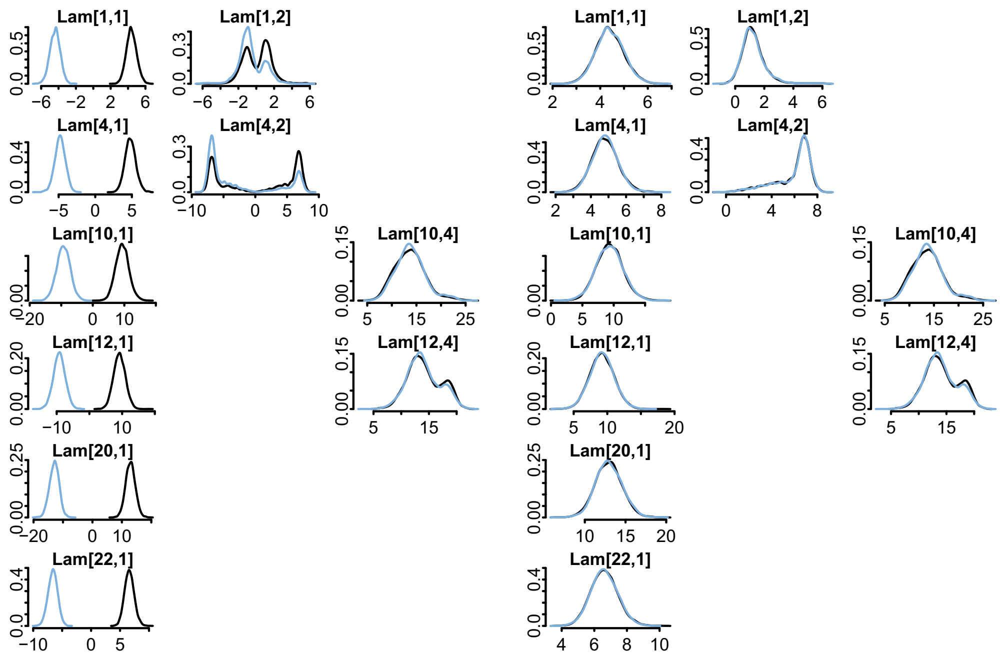

FIGURE 4.
Holzinger and Swineford data. Marginal posterior density plots of factor loadings for five-factor bifactor model from two parallel MCMC chains. *Left panel* Selected density plots before relabeling. *Right panel* Selected density plots after relabeling.

3.3.2. Results from Unconstrained MCMC Samplers We first carried out a Bayesian CFA where only the factor variances are fixed at 1, but no additional constraints are placed on the loadings apart from structural zeros (Eq. 6). This allowed MCMC chains to explore the full posterior distribution with 2^q equivalent modes. We then post-process MCMC draws with our relabeling algorithm to remove indeterminacy associated with reflection invariance.

Figure 4 shows posterior density plots for several selected factor loadings from the unconstrained MCMC sampler before relabeling (left panel) and after relabeling (right panel).[2] Before relabeling, 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 (Fig. 4, right panel), making the relabeled draws amenable to standard summary statistics. Note also that the resulting posterior densities of the loadings may assign a nonzero probability mass to both the positive and the negative region of the real line (e.g., as for $\lambda_{1,2}$ in Fig. 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 signs for each factor for other runs of the relabeling algorithm, depending on our specification of the starting values for sign changes.

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 MCMC. The loglikelihood ratio fit statistics were consistent across chains at 341.070, 340.745, and 341.096, respectively.

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

TABLE 3.
Summary of analyses of Holzinger and Swineford data.

Starting values	Estimation	Constraint type	Converged	Best G^2
Random	MCMC	Fixed value	Yes 2/3	4040.882
Positive	MCMC	Fixed value	Yes 1/3	4040.909
PCA	MCMC	Fixed value	Yes 0/3	NA
Random	MCMC	Positivity	Yes 2/3	6214.451
Positive	MCMC	Positivity	Yes 2/3	2716.817
PCA	MCMC	Positivity	Yes 2/3	2858.398
Lavaan	ML	Positivity	Yes	337.632
Lavaan	ML	Fixed value	Yes	337.632
Random	MCMC+relabel	None	Yes 3/3	340.745

For each MCMC analysis, 3 chains were obtained. "Best G^2" reports the MCMC approximation to the best loglikelihood ratio fit statistic (among converged chains for MCMC).

3.3.3. Results Summary Table 3 provides a complete 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, and the best value of the MCMC approximation to the loglikelihood ratio fit statistic (Bishop et al., 1975) among the converged chains. The fit statistics indicate that the factor model solutions from the converged chains lack consensus across the scenarios of the constrained analysis; we would expect the fit statistics to be much more similar across scenarios if the solutions were equivalent. Table 3 also provides the loglikelihood ratio fit statistic for ML estimates obtained with package `Lavaan` (Rosseel, 2012) that is freely available in `R` (R Development Core Team, 2010); by default, `Lavaan` initializes all factor loadings by setting each loading equal to one. We observe that the model fit from the unconstrained MCMC with relabeling is much better than the fit from constrained MCMC chains and is comparable to the fit of the ML solution (small differences can be attributed to Bayesian shrinkage). The loadings estimates from our approach are comparable to those obtained from `Lavaan` and to the estimates from the original analysis of Holzinger and Swineford (1939) (comparison not shown).

4. Discussion

Reflection invariance is a limited aspect of the general identification problem in factor analysis, namely invariance of factor solutions with respect to arbitrary rotations (Anderson & Rubin, 1956). This paper presents an approach for dealing with reflection invariance in Bayesian factor analysis. For a q-factor model, we suggest to obtain MCMC samplers without additional identifiability constraints and to use the relabeling algorithm for post-processing MCMC draws to align them to only one out of the 2^q reflection modes. By following our Bayesian estimation approach with relabeling, we obtain consistent results across MCMC chains that are comparable to the ML results. Most importantly, we are able to do so without the need to select custom starting values, or to decide on the type of constraints (i.e., fixed value or positivity), or to make preferential choices among variables (i.e., selecting one loading per factor to be positive) for implementation of reflection identifiability constraints. Because "[a]lmost any maximum likelihood computation can be done by some Markov chain Monte Carlo scheme" (Geyer & Thompson, 1992, p. 660) and because "[t]he geometry of models with multiple factors is still largely unknown" (Drton, 2009, p. 1003), our relabeling approach, although directly applicable only to Bayesian MCMC estimation, may be also of note to researchers who are accustomed to maximum likelihood estimation. For example, Millsap (2001) observed that different choices of variables for placement of uniqueness constraints may result in ML estimates that are not rotationally equivalent.

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, & Robert, 2000; Stephens, 2000; Jasra, Holmes, & Stephens, 2005). Similarly, our findings for Bayesian factor analysis illustrate that constraining some loadings to be one or positive may result in nontrivial (i.e., not easily identified by visual inspection) multimodality in the likelihood, and, as a consequence, in a nontrivial mode-switching behavior with Markov chain Monte Carlo samplers. Mode-switching behavior within MCMC chains, analogously to label-switching in mixture models (Stephens, 2000), presents a serious problem for Bayesian inference because it invalidates posterior estimates.[3] 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 likelihood surface may not be as obvious when models are complex and the potential for model misspecification is high. This is another reason why unnecessary identifiability constraints are to be avoided.

We note that it might be possible in some cases to obtain better results by handpicking the starting values (e.g., Peeters, 2012). However, in complex models such as CFA extensions for mixed outcomes with covariates (e.g., Gruhl, Erosheva, Crane, et al., 2013), handpicking starting values is not only difficult but also counterproductive as it could result in a failure to fully explore the posterior distribution (Gelman et al., 2004). The practice of carefully selecting starting values also prohibits analyses automation necessary for intensive model search procedures.

Similarly, in some special cases when 2^q equivalent modes are clearly separated, the relabeling could be achieved by applying a sign alignment method—similar to those used in jackknife exploratory factor analysis (Pennell, 1972; Clarkson, 1979)—to MCMC draws from different chains. However, the jackknife method would only be possible when the MCMC samplers from each chain have visited only one of the 2^q equivalent modes which, by itself, indicates the inability to fully explore posterior distribution (Gelman et al., 2004). In cases when 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 in those cases. Moreover, the algorithm is still valid when not all the items are positively scored. In addition, opposite polarity of scoring will not increase the algorithm's complexity.

While the necessity of Bayesian methods may not be obvious for simpler examples such as those discussed in this paper, Bayesian methods undoubtedly prevail for complex models (Lee, 2007; Scheines et al., 1999). The relabeling algorithm presented in this paper is not limited to CFA. It can be used for other models that exhibit multimodalities in the likelihood due to sign invariance of latent variables. For example, Gruhl et al. (2013) use the relabeling algorithm for Bayesian estimation of a semiparametric bifactor model for mixed outcomes; Savitsky and McCaffrey (2014) incorporate the relabeling algorithm in exploratory factor analysis as part of a Bayesian multivariate hierarchical model for ordered 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 & Curtis, 2011) that contains key code components and simulation studies. The relabeling algorithm from this paper is implemented in package `relabeLoadings` (Curtis & Erosheva, 2016) that is freely available in `R` (R Development Core Team, 2010).

[3]Note that the label-switching problem does not apply to maximum a posteriori or ML estimation that are not MCMC-based (e.g., Celeux, Forbes, Robert, & Titterington, 2006, p. 656).

Acknowledgments

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

References

Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of the third berkeley symposium on mathematical statistics and probability* (pp. 111–150). Oakland: University of California Press.

Bafumi, J., Gelman, A., Park, D. K., & Kaplan, N. (2005). Practical issues in implementing and understanding Bayesian ideal point estimation. *Political Analysis*, *13*, 171–187.

Bishop, Y., Fienberg, S. E., & Holland, P. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge: The MIT press.

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

Celeux, G., Forbes, F., Robert, C. P., & Titterington, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, *1*, 651–673.

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

Clarkson, D. B. (1979). Estimating the standard eoors of rotated factor loadings by jackknifing. *Psychometrika*, *44*, 297–314.

Congdon, P. (2003). *Applied Bayesian modelling*. New York: Wiley.

Congdon, P. (2006). *Bayesian statistical modelling*. New York: Wiley.

Cressie, N., & Read, T. (1989). Pearson's χ^2 and the loglikelihood ration statistic G^2—A comparative review. *International Statistical Review*, *57*, 19–43.

Curtis, S. M., & Erosheva, E. A. (2016). relabeLoadings: Relabel loadings from MCMC output for confirmatory factor analysis. R package version 1.0.

Dolan, C. V., & Molenaar, P. C. (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.

Drton, M. (2009). Likelihood ratio tests and singularities. *The Annals of Statistics, 37*, 979–1012.

Erosheva, E. A., & Curtis, S. M. (2011). Dealing with rotational invariance in Bayesian confirmatory factor analysis. Technical Report 589, University of Washington.

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

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis*. New York: Chapman & Hall/CRC.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel and hierarchical models*. Cambridge: Cambridge University Press.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian statistics 4* (pp. 169–194). Oxford: Clarendon Press.

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

Geyer, C. J., & Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society, Series B (Methodological), 54*, 657–699.

Ghosh, J., & Dunson, D. (2008). Bayesian model selection in factor analytic models. In D. B. Dunson (Ed.), *Random effect and latent variable model selection* (pp. 151–163). Berlin: Springer.

Ghosh, J., & Dunson, D. (2009). Default prior distributions and efficient posterior computation in Bayesian factor analysis. *Journal of Computational and Graphical Statistics*, *18*, 306–320.

Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika*, *57*, 423–436.

Gruhl, J., Erosheva, E. A., Crane, P. K., et al. (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, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika*, *2*, 41–54.

Holzinger, K. J., & Swineford, F. (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., & Stephens, D. (2005). Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science*, *20*, 50–67.

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

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

Lee, S.-Y. (2007). *Structural equation modeling: A Bayesian approach*. West Sussex: Wiley.

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

Lopes, H. F., & West, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica*, *14*, 41–67.

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

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

Matlosz, K. (2013). Bayesian multidimensional scaling model for ordinal preference data. Ph.D. thesis, Columbia University.

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

Muirhead, R. J. (1982). *Aspects of multivariate statistical theory*. New York: Wiley.

Muthén, L. K., & Muthén, B. O. (2005). *Mplus: Statistical analysis with latent variables: User's guide*. Los Angeles: Muthén & Muthén.

Nishihara, R., Minka, T., & Tarlow, D. (2013). Detecting parameter symmetries in probabilistic models. arXiv preprint arXiv:1312.5386.

Peeters, C. F. (2012). Bayesian exploratory and confirmatory factor analysis: Perspectives on constrained-model selection. Ph.D. thesis, Utrecht University.

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, K. M. (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*. Vienna: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and longitudinal modeling using Stata*. College Station: STATA press.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36.

Rowe, D. B. (2001). A model for Bayesian factor analysis with jointly distributed means and loadings. *Social Science Working Paper, 1108*, 1–16.

Savitsky, T. D., & McCaffrey, D. F. (2014). Bayesian hierarchical multivariate formulation with factor analysis for nested ordinal data. *Psychometrika*, *79*, 275–302.

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

Schervish, M. J. (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, J. K., & Magidson, J. (2005). *Technical guide for Latent GOLD 4.0: Basic and advanced*. Belmont: Statistical Innovations Inc.

Manuscript Received: 14 DEC 2011
Final Version Received: 29 NOV 2016
Published Online Date: 13 MAR 2017