Psychological Methods

A Partially Confirmatory Approach to Scale Development With the Bayesian Lasso

Jinsong Chen, Zhihan Guo, Lijin Zhang, and Junhao Pan

Online First Publication, July 13, 2020. http://dx.doi.org/10.1037/met0000293

American Psychological Association

Psychological Methods

ISSN: 1082-989X

http://dx.doi.org/10.1037/met0000293

A Partially Confirmatory Approach to Scale Development With the Bayesian Lasso

Jinsong Chen, Zhihan Guo, Lijin Zhang, and Junhao Pan
Sun Yat-Sen University

Abstract

The exploratory and confirmatory approaches of factor analysis lie on two ends of a continuum of substantive input for scale development. Recent advancements in Bayesian regularization methods enable more flexibility in covering a wide range of the substantive continuum. Based on the Bayesian Lasso (least absolute shrinkage and selection operator) methods for the regression model and covariance matrix, this research proposes a partially confirmatory approach to address the loading and residual structures at the same time. With at least one specified loading per item, a one-step procedure can be applied to figure out both structures simultaneously. With a few specified loadings per factor, a two-step procedure is preferred to capture the model configuration correctly. In both cases, the Bayesian hierarchical formulation is implemented using Markov Chain Monte Carlo estimation with different Lasso or regular priors. Both simulated and real data sets were analyzed to evaluate the validity, robustness, and practical usefulness of the proposed approach across different situations.

Translational Abstract

Exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) are traditional approaches for development of psychological scales, and each has pros and cons. EFA is driven by data, whereas CFA is driven by theory. However, different amounts of substantive knowledge are available to analyze the factor structure of a scale, so the approaches become two ends of a continuum. Although it is desirable to cover a wide range of the continuum in practice, there are psychometric challenges in combining EFA and CFA within one framework. Relying on Bayesian Lasso methods and sound statistical modeling, this research proposes a partially confirmatory factor analysis (PCFA) approach to address the challenges. With a two-step procedure, the proposed approach can offer more flexibility to scale development, especially when there are multiple factors and many items. Using both simulation experiments and real data sets related to mathematics learning and humor style questionnaires, this article shows that PCFA can help researchers better develop scales with both exploratory and confirmatory elements across different settings.

Keywords: factor analysis, Bayesian Lasso, partially confirmatory, Lasso loading, residual covariance

Factor analysis (FA) is a statistical technique widely used for developing psychological scales with multiple factors and many items. The major purpose is to explain item responses in term of the structure of the factorial loadings and residuals. With FA, there are two typical approaches of scale development: exploratory (EFA; Jennrich & Sampson, 1966) and confirmatory (CFA; Jöreskog, 1969). EFA is data driven with little substantive knowledge or constraint on the loading structure, whereas CFA is theory driven with strong knowledge or constraint on the structure. In general, however, different amounts of substantive input can be available, making the two approaches two ends of a continuum. To be more flexible on the substantive continuum, one traditional practice is to apply post hoc model modification (PMM) based on modification indices (Kaplan, 1990; Sörbom, 1989) to free up constraints on the basis of the confirmatory approach. However, PMM is theoretically suboptimal and will be problematic when many constraints need to be released (Bentler, 2007; MacCallum, 1986; Steiger, 1990).

With the recent development of Bayesian regularization methods such as Bayesian ridge regression (Hsiang, 1975), Bayesian least absolute shrinkage and selection operator or Lasso (Hans, 2009; Park & Casella, 2008), and spike-and-slab prior (O'Hara & Sillanpää, 2009), it is increasingly possible to achieve more flex-

Jinsong Chen, Zhihan Guo, Lijin Zhang, and Junhao Pan, Department of Psychology, Sun Yat-Sen University.

This work was supported by the Philosophy and Social Science Planning Project of Guangdong Province (GD17HJY01) and the Key Nurturing Program for Young Teachers at the Sun Yat-Sen University (18wkzd14). It was also supported by the National Natural Science Foundation of China (NSFC 31871128) and the MOE (Ministry of Education) Project of Humanities and Social Science of China (18YJA190013). The idea and methods in this article has not been disseminated at any conference or shared on any website.

Correspondence concerning this article should be addressed to Jinsong Chen or Junhao Pan, Department of Psychology, Sun Yat-Sen University, 132 Waihuan Dong Road, Higher Education Mega Center, Guangzhou 510006, China. E-mail: jinsong.chen@live.com or panjunh@mail.sysu.edu.cn

ibility on the substantive continuum in scale development using FA. Specifically, cross-loadings can be regularized via the Bayesian structural equation modeling (BSEM) approach with the ridge regression prior (B. O. Muthén & Asparouhov, 2012) or spike-and-slab prior (Lu, Chow, & Loken, 2016). Regularization of off-diagonal elements in the residual variance-covariance matrix can be conducted via the Bayesian covariance Lasso for CFA (BLCCFA) approach (Pan, Ip, & Dubé, 2017). Studies have also demonstrated the superiority of Bayesian methods over PMM for identifying significant cross-loadings or residual covariance in corresponding settings. The methods all offer variable selection with different norms of penalty on the parameters of interests, Bayesian hierarchical modeling with the Markov Chain Monte Carlo (MCMC; Gilks, Richardson, & Spiegelhalter, 1996) algorithm, and simultaneous estimation of the shrinkage parameters with proper priors.

On the basis of previous literature, this research proposes a partially confirmatory factor analysis (PCFA) approach with the Bayesian Lasso method to address the loading and residual covariance structures. Addressing both structures at the same time is challenging, and some constraints can be used. With the constraint of one specified loading per item, a one-step procedure can be applied to capture the loading and residual covariance structures simultaneously. Otherwise, an additional step with a few specified loadings per factor and diagonal residual matrix can be used to satisfy the constraints before applying the above step, making it a two-step procedure. In both cases, the MCMC estimation is implemented based on derivations with different Lasso or regular priors on the loading and residual structures. Both the Gibbs sampler (Casella & George, 1992; Geman & Geman, 1984) and Metropolis-Hastings algorithm (Hastings, 1970; Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953) are used in the MCMC estimation. Note that the one-step procedure reduces to the BLCCFA approach when all loadings are specified or fixed, whereas the first step in the two-step procedure is similar to the BSEM approach but with different regularization methods. With the two procedures, the proposed approach can cover a wide range of the substantive continuum and be flexible and useful for scale development, especially with three or more factors and many items. Both simulated and real data sets were analyzed to evaluate the validity, robustness, and practical usefulness of the proposed approach across different situations.

The Partially Confirmatory Framework

Suppose there are J items in a psychological scale with N respondents, and item responses $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N$ are independently and randomly distributed observations. In a general FA model, each $\mathbf{y}_i = (\mathbf{y}_{i1}, \mathbf{y}_{i2}, \ldots, \mathbf{y}_{iJ})^T$ satisfies the following equation:

$$\mathbf{y}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\varepsilon}_i, i = 1, 2, \ldots, N, \tag{1}$$

where model parameters include the $J \times 1$ intercept vector $\boldsymbol{\mu}$ (usually of little interest during scale development), $J \times K$ loading matrix $\boldsymbol{\Lambda} = (\lambda_{jk})$, $K \times K$ factorial covariance matrix $\boldsymbol{\Phi}$, and $J \times J$ residual covariance matrix $\boldsymbol{\Psi} = (\psi_{jj'})$, with $K \times 1$ latent factors $\boldsymbol{\omega}_i \sim N_K[\mathbf{0}, \boldsymbol{\Phi}]$ and $J \times 1$ residuals $\boldsymbol{\varepsilon}_i \sim N_J[\mathbf{0}, \boldsymbol{\Psi}]$. To determine the scale of the latent factors, one can either fix the factorial variance or fix one loading per factor. In this research, all factorial variances are fixed to one, which means $\boldsymbol{\Phi}$ is a correlation matrix. Loadings in $\boldsymbol{\Lambda}$ can be fixed to zero (based on substantive knowledge), specified as free to estimate (also based on substantive knowledge), or unspecified (learned through regularization). With different combinations of zero-fixed, specified, and unspecified loadings, different amounts of substantive input can be transformed into different structures of the loading matrix of $\boldsymbol{\Lambda}$.

In EFA, both the number of factors and structure of $\boldsymbol{\Lambda}$ are unknown and of primary interest. This means all loadings are unspecified. Moreover, $\boldsymbol{\Psi}$ is modeled as a diagonal matrix, and the accuracy of loading and $\boldsymbol{\Phi}$ estimates is secondary. The model is usually underidentified and cannot be estimated with traditional algorithms such as the maximum likelihood estimation. Instead, purely data-driven techniques such as principal component analysis, subjective criteria such as eigenvalue, and rotation are adopted to obtain a small number of factors and a simple structure for the loading matrix. In CFA, the number of factors and the structure of $\boldsymbol{\Lambda}$ are given based on theory-driven hypothesis. This means that all loadings are either zero or specified. Free elements in $\boldsymbol{\Lambda}$ and $\boldsymbol{\Phi}$ are of primary interest. $\boldsymbol{\Psi}$ can be nondiagonal, with usually a few nonzero off-diagonal elements, which are nuisance parameters imposing challenges for model estimation and fitting. Under certain conditions, the model is overidentified and can be estimated with traditional methods (see, e.g., Kaplan, 2009).

The framework becomes essentially exploratory when all loadings are unspecified, and confirmatory when all loadings are either zero-fixed or specified. More generally, developers can designate some loadings as zero or specified while keeping others as unspecified in the framework, which means the loading matrix is partially specified. A typical PCFA model with two factors and 10 items can be found in Figure 1, where there is one specified and unspecified loading per item. Note that all residuals are allowed to correlate with each other in the figure. However, two thorny issues need to be addressed before the framework can be useful: the accommodation of both specified and unspecified loadings, and the identification of nonzero off-diagonal elements or residual correlations in $\boldsymbol{\Psi}$. The Bayesian Lasso can address both issues.

The Bayesian Lasso Approach

Bayesian Regression and Covariance Lasso

The issues of both unspecified loadings and a residual covariance matrix can be addressed based on variable selection with the Bayesian Lasso. For a regression model with p predictors $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ and responses

$$\mathbf{y} = \mu\mathbf{1}_n + \mathbf{X}\boldsymbol{\beta} + \varepsilon, \tag{2}$$

the Lasso estimates are least squares estimates with L_1 norm penalty (Tibshirani, 1996):

$$\hat{\boldsymbol{\beta}} = \underset{\beta}{\operatorname{argmin}}\left[(\tilde{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta})^T(\tilde{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta}) + \delta\sum_{j=1}^{p} |\beta_j|\right] \tag{3}$$

where $\mathbf{X}$ is the standardized regressors, $\tilde{\mathbf{y}} = \mathbf{y} - \bar{y}\mathbf{1}_n$, $\boldsymbol{\varepsilon}$ is the residual term, and $\delta \geq 0$ is the tuning parameter that determines the amount of shrinkage. The Bayesian Lasso estimates accomplish the same goals as the frequentist counterpart by choosing appropriate forms of prior distributions that play the role of the penalty in Equation 3. In a fully Bayesian approach with a hier-

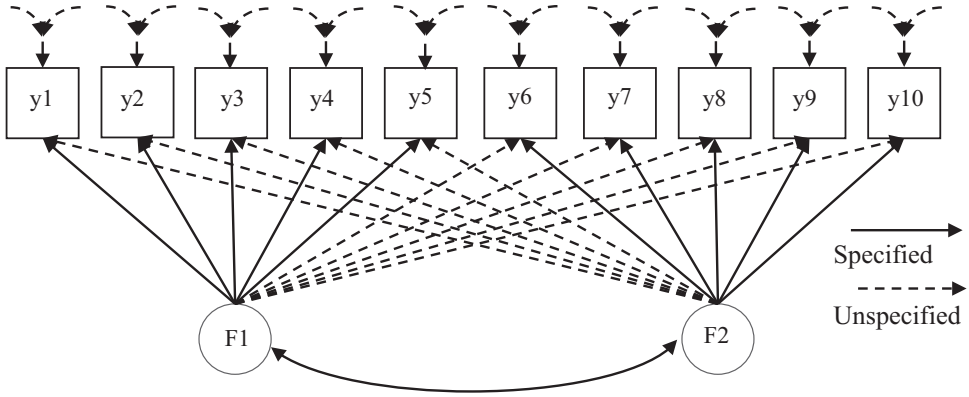

Figure 1. A C-step example in PCFA with two factors and 10 items.

archical formulation (Park & Casella, 2008), hierarchical priors can be applied and, as a result, the priors of the predictors follow the double exponential (i.e., Laplace) distributions with the shrinkage parameter. With hyperpriors and the MCMC algorithm, the shrinkage parameter can be estimated simultaneously with other parameters.

The Bayesian regression Lasso has been extended to address covariance structure with sparse off-diagonal terms bounded away from zero (Khondker, Zhu, Chu, Lin, & Ibrahim, 2013; Wang, 2012). Specifically, the L_1 norm penalty is applied to the inverse of the covariance matrix with the graphical Lasso. Double exponential priors with similar shrinkage parameters are used for the off-diagonal elements while independent exponential priors can be used for the diagonal terms. With hyperparameters for the priors and an adapted MCMC algorithm, the Bayesian covariance Lasso is successfully implemented to address a sparse residual covariance structure under the CFA context (Pan et al., 2017).

The Confirmatory Steps

In PCFA, the loadings in one item can be zero, specified, or unspecified. For a specific item, the latent factors and loading vector in Equation 1 can be rearranged so that the zero or specified loadings and unspecified loadings can be partitioned. Denote the $K \times 1$ loading vector for item j as $\mathbf{\Lambda}_j = \begin{pmatrix} \mathbf{\Lambda}_j' \\ \mathbf{\Lambda}_j'' \end{pmatrix}$, where $\mathbf{\Lambda}_j^T$ is the jth row of $\mathbf{\Lambda}$ and $j = 1, 2, \ldots, J$. The $K' \times 1$ vector $\mathbf{\Lambda}_j'$ and $K'' \times 1$ vector $\mathbf{\Lambda}_j''$ are the zero or specified and unspecified parts, respectively. All unspecified loadings are estimated with Bayesian Lasso and are called Lasso loadings in this research. Similar to Bayesian regression Lasso, different priors can be assigned to Lasso loadings in hierarchical form, as

$$\begin{aligned} \mathbf{\Lambda}_j'' &\sim N_{K''}(\mathbf{0}, \mathbf{D}_{\tau_j}), \\ \mathbf{D}_{\tau_j} &= diag(\tau_{j1}^2, \ldots, \tau_{jK''}^2), \\ \tau_{jk}^2 &\sim Gamma\left(1, \frac{\delta_l^2}{2}\right), \end{aligned} \tag{4}$$

where $k = 1, 2, \ldots, K''$ and δ_l is the shrinkage parameter. Note that an individual parameter can be set to shrink each item j, but the difference of regularization is trivial. Here one shrinkage parameter is used for simplicity of discussion. The hierarchical priors form a double exponential distribution with the L_1 norm penalty, which can maintain conjugacy of the posterior distribution. The explicit form of the full conditional distribution can be obtained, as shown in Appendix A, which can be directly sampled from using the Gibbs sampler. Hyperpriors can be assigned to δ_l as $\delta_l^2 \sim Gamma(a_l, b_l)$ and the common choices for hyperparameters are 1 for a_l and a small b_l. For specified loadings, conjugate prior distribution $\mathbf{\Lambda}_j' \sim N_{K'}(\mathbf{\Lambda}_{0j}, \mathbf{H}_{0j})$ can be assigned with $\mathbf{\Lambda}_{0j}$ and $\mathbf{H}_{0j}$ as hyperparameters, which can also lead to an explicit full conditional distribution (Lee, 2007). If needed, a conjugate prior for the intercept vector can be assigned as $\boldsymbol{\mu} \sim N_j(\boldsymbol{\mu}_0, \mathbf{H}_{\mu 0})$, with $\boldsymbol{\mu}_0$ and $\mathbf{H}_{\boldsymbol{\mu} 0}$ as hyperparameters and an explicit full conditional distribution for sampling.

The residual covariance matrix in a partially confirmatory model is not necessarily diagonal. The entire $\mathbf{\Psi}$ can be modeled as a sparse structure on off-diagonal elements with the Bayesian covariance Lasso. Denote $\mathbf{\Sigma} = \mathbf{\Psi}^{-1} = (\sigma_{jj'})_{J \times J}$. Double exponential priors can be assigned to the off-diagonal elements, which has the form $\frac{\delta_s}{2}\exp(-\delta_s |\, \sigma_{jj'} \,|)$ with $j < j'$, whereas independent exponential priors can be assigned to the diagonal elements with the form $\frac{\delta_s}{2}\exp\left(-\frac{\delta_s}{2}\sigma_{jj}\right)$ (Wang, 2012). The full conditional distribution can be obtained and directly sampled from using a block Gibbs sampler. Hyperpriors can be assigned to the shrinkage parameter δ_s as $\delta_s \sim Gamma(a_s, b_s)$ and the common choices for hyperparameters are 1 for a_s and a small b_s. The related full conditional distributions are presented in Appendix C.

The factorial correlation matrix is also estimated by adapting the relevant MCMC algorithms (Liu, 2008; Liu & Daniels, 2006). However, an explicit form of the full conditional distribution for the factorial correlation matrix $\mathbf{\Phi}$ cannot be easily obtained and directly sampled from. Instead, one can first sample from the original covariance matrix $\mathbf{\Phi}^*$ with conjugate prior $\mathbf{\Phi}^* \sim Inv\text{-}Wishart(\boldsymbol{S}_0, \boldsymbol{\nu}_0)$ and then transform the sampled $\mathbf{\Phi}^*$ into corresponding $\mathbf{\Phi}$ with the reduction function $\mathbf{\Phi} = \boldsymbol{D}^{-1}\mathbf{\Phi}^*\boldsymbol{D}^{-1}$, where $\mathbf{D}$ is an expansion parameter matrix. The provisional $\mathbf{\Phi}$ is accepted based on a Metropolis-Hastings acceptance probability, which is computed by comparing the determinant of the correlation matrix obtained from the previous draw and current draw. Liu (2008) and Liu and Daniels (2006) provide more details on how to sample the correlation matrix $\mathbf{\Phi}$.

It is challenging to estimate a partially specified loading matrix and a residual covariance matrix with off-diagonal elements si-

multaneously. The PCFA model can be especially problematic when there are too many unspecified or Lasso loadings. Specifically, we found that parameter estimates are unstable or difficult to converge if there is no specified loading for many items. This research imposed a constraint of having at least one specified loading per item, although it might be fine if there are a few items without any specified loading. The constraint is meaningful substantively since there is usually a target factor for every item during scale development. The specified loading can be the major loading under the BSEM context (B. O. Muthén & Asparouhov, 2012) or any other loading that the developer is confident to specify. With the constraint of one specified loading per item, the formulation tends to be confirmatory and is called the confirmatory step (C-step) in PCFA. Figure 1 gives an example of the C-step with two factors, 10 items, and one specified loading per item. Note that item residuals are allowed to completely correlate with each other under the Bayesian Lasso covariance estimation. When all loadings are specified, the C-step is essentially the BLCCFA (Pan et al., 2017), with a minor difference of fixing the factorial variance rather than having one item per factor for scale indeterminacy.

The Exploratory Step

When no loading can be specified for many items, the above constraint can be released by modeling $\boldsymbol{\Psi}$ as a diagonal matrix or ignoring any residual correlation. The loadings in $\boldsymbol{\Lambda}$ can be mostly unspecified given there is at least one specified loading per factor to identify the factor. For more stable performance, more specified loadings per factor are preferred. The hierarchical priors of the unspecified loadings conditional on the diagonal residual terms become:

$$\begin{aligned} \boldsymbol{\Lambda}_j^{''} | \psi_{jj} &\sim N_{K''}(\mathbf{0}, \psi_{jj}\mathbf{D}_{\tau_j}), \\ \psi_{jj}^{-1} &\sim Gamma(a_{0j}, b_{0j}), \\ \mathbf{D}_{\tau_j} &= \text{diag}(\tau_{j1}^2, \ldots, \tau_{jK''}^2), \\ \tau_{jk}^2 &\sim Gamma\left(1, \frac{\delta_l^2}{2}\right), \end{aligned} \tag{5}$$

where ψ_{jj} is the jth diagonal term of $\boldsymbol{\Psi}$, and the priors for the shrinkage parameters δ_l are similar to the C-step. The explicit full conditional distribution can be obtained (see Appendix B) and directly sampled from using the Gibbs sampler. Other parameter estimates are the same as the C-step above, except that there is no off-diagonal element in the residual covariance matrix. This is similar to the BSEM approach, but with the Lasso method rather than the ridge regression or spike-and-slab priors (Lu et al., 2016; B. O. Muthén & Asparouhov, 2012). Parameter estimates are deemed biased when the true residual structure is not diagonal, but it would suffice to identify at least one specified loading per item for the C-step. Because the formulation tends to be exploratory especially compared with the previous step, it is called the exploratory step (E-step) in PCFA. Figure 2 gives an example of the E-step with two factors, 10 items, and two specified loadings per factor. Note that there is no off-diagonal element in the residual structure.

Markov Chain Monte Carlo Estimation

Graphical representations of the hierarchical structure of the C-step and E-step are presented in Figures 3 and 4, respectively. The figures show how the item responses Y_{ij} can be modeled using latent factors, loadings, and residuals, which are characterized by their mean, covariance structure, and shrinkage parameters in a hierarchical form. The Bayesian hierarchical structure can be estimated using MCMC, which is a simulation-based algorithm that iteratively resamples from the probability distributions based on a stochastic process of Markov chains (Gill, 2002). Under mild regularity conditions, the Markov chains will converge to a stationary posterior distribution after a sufficiently large number of iterations called the burn-in period. Two types of MCMC are used: the Gibbs sampler (Casella & George, 1992), which is easier to implement but requires the explicit full conditional distributions, and the Metropolis-Hastings algorithm (Chib & Greenberg, 1995), which relies only on the joint posterior distribution. Except for the factorial correlation matrix $\boldsymbol{\Phi}$, which is estimated using the Metropolis-Hastings algorithm as described above, all model parameters are estimated using the Gibbs sampler or block Gibbs sampler.

The procedure for sampling the parameters of interest from their full conditional distributions has six steps:

Step 1: Draw $\boldsymbol{\Omega}$ from $p(\boldsymbol{\Omega} \mid \mathbf{Y}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi})$;

Step 2: Draw $\boldsymbol{\Lambda}_j^{''}$ from $p(\boldsymbol{\Lambda}_j^{''} \mid \mathbf{Y}, \boldsymbol{\Lambda}_{-j}, \boldsymbol{\Omega}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \delta_{lj})$ for $j = 1$ to J;

Step 3: Draw $\boldsymbol{\Lambda}_j'$ from $p(\boldsymbol{\Lambda}_j' \mid \mathbf{Y}, \boldsymbol{\Lambda}_{-j}, \boldsymbol{\Omega}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Psi})$ for $j = 1$ to J;

Step 4: Draw $\boldsymbol{\mu}$ from $p(\boldsymbol{\mu} \mid \mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi})$, if needed;

Step 5: Draw $\boldsymbol{\Phi}$ from $p(\boldsymbol{\Phi} \mid \mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\Psi})$;

Step 6: Draw $\boldsymbol{\Sigma}$ from $p(\boldsymbol{\Sigma} \mid \mathbf{Y}, \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\Lambda}, \boldsymbol{\Phi}, \delta_s)$ for the C-step, or $p(\boldsymbol{\Sigma} \mid \mathbf{Y}, \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\Lambda}, \boldsymbol{\Phi})$ for the E-step, and compute $\boldsymbol{\Psi} = \boldsymbol{\Sigma}^{-1}$.

Multiple chains with different initial values can be run to monitor the convergence of the algorithm. After the burn-in period, the

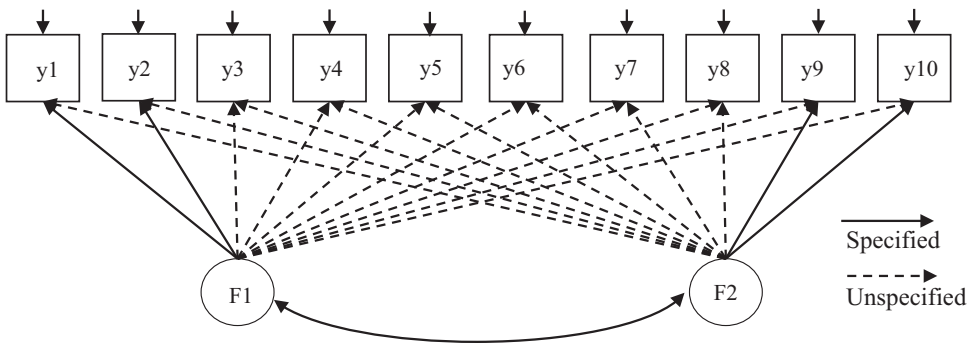

Figure 2. An E-step example in PCFA with two factors and 10 items.

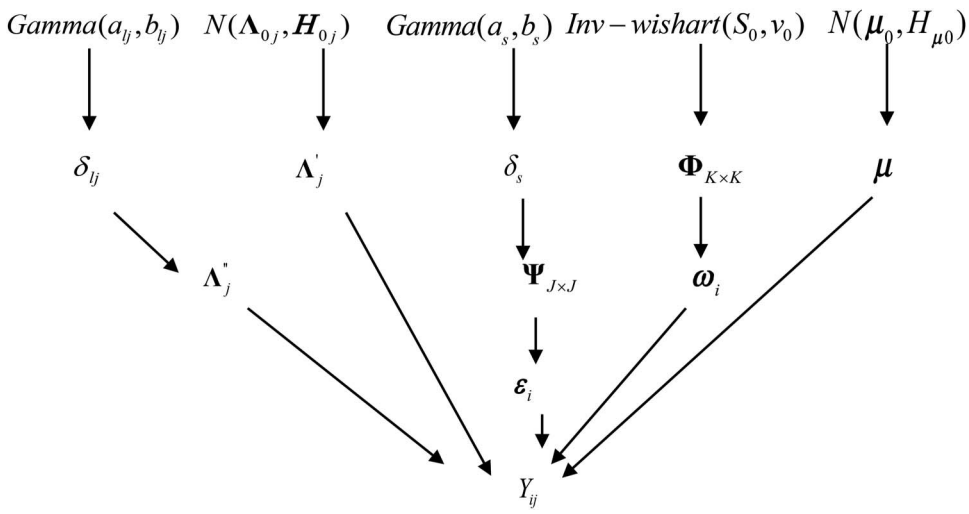

Figure 3. A directed acyclic graph for the C-step.

convergence for the parameters of interest can be determined using the estimated potential scale reduction (EPSR) value (Gelman, 1996). The EPSR value compares the ratio of the weighted average of the within-chain variance and between-chain variance to the within-chain variance. The chains are said to converge to the stationary distribution if this ratio is less than 1.1 (Gelman, Carlin, Stern, & Rubin, 2004). In this research, we found that all chains had reached stationary status within 8,000 iterations and then set the burn-in as 10,000 iterations for all subsequent studies. After the burn-in phase, parameter estimates are inferenced based on 10,000 and 20,000 draws for the simulation and real-data analyses, respectively. The uncertainty of estimates is characterized with the concept of highest posterior density (HPD) intervals, or more specifically the $100(1 - \alpha)\%$ HPD interval (Box & Tiao, 1973). Both the EPSR value and HPD interval can be obtained with the R package coda (Plummer, Best, Cowles, & Vines, 2006). All programming was conducted on the R platform (R Development Core Team, 2010). Finally, two sets of prior values were evaluated for sensitivity analysis in a preliminary study, as shown in Table 1, and the difference turned out to be trivial. We adopted the first set of values for subsequent analyses, which was less informative.

Empirical Studies

Study 1: Performance in a Relatively Simple Case

Study 1 evaluated the performance of the E-step and C-step in a relatively simple case with two factors and 10 items, namely, $J = 10$ and $K = 2$, through a simulation study. The factorial correlation was $\phi_{12} = \phi_{21} = .3$; the loading matrix was

$$\mathbf{\Lambda}^T = \begin{bmatrix} \lambda_{11} & \lambda_{21} & \lambda_{31} & \lambda_{41} & \lambda_{51} & \lambda_{61} & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 \\ \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_{52} & \lambda_{62} & \lambda_{72} & \lambda_{82} & \lambda_{92} & \lambda_{10,2} \end{bmatrix},$$

with $\lambda_{11} = \lambda_{21} = \lambda_{31} = \lambda_{41} = \lambda_{72} = \lambda_{82} = \lambda_{92} = \lambda_{10,2} = .7$, $\lambda_{51} = \lambda_{61} = \lambda_{52} = \lambda_{62} = .5$, and all other loadings $\lambda_0 = 0$. For

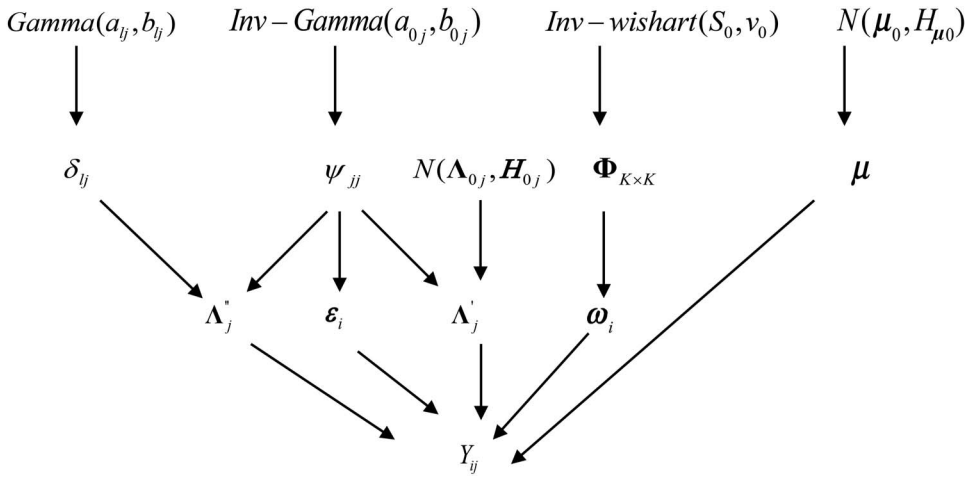

Figure 4. A directed acyclic graph for the E-step.

Table 1
Two Sets of Prior Values

Set	a_{lj}	b_{lj}	a_s	b_s	a_0	b_0	$\mathbf{\Lambda}_{0j}$	$\boldsymbol{H}_{0j}$	$\boldsymbol{S}_0$	$\boldsymbol{v}_0$
1	1	.01	1	.01	1	.01	0	$4\mathbf{I}$	$\mathbf{I} + .1_{\text{od}}$	$K + 2$
2	1	.1	1	.1	1	.1	0	$\mathbf{I}$	$\mathbf{I} + .5_{\text{od}}$	$K + 8$

Note. $\mathbf{I}$ = identity matrix; $\mathbf{I} + .1_{\text{od}}$ = diagonal elements as 1 and off-diagonal elements as .1; $\mathbf{I} + .5_{\text{od}}$ = diagonal elements as 1 and off-diagonal elements as .5; $\boldsymbol{\mu}_0$ and $\mathbf{H}_{\boldsymbol{\mu}0}$ not set as the intercept is not of interest in this research.

the residual structure in $\mathbf{\Psi}$, the diagonal elements were $\psi_{jj} = .51$ for $j = 1$ to 4 and 7 to 10 and $\psi_{jj} = .35$ for $j = 5$ and 6, and the nonzero off-diagonal elements were $\psi_{12} = \psi_{21} = \psi_{74} = \psi_{47} = .3$ with all other off-diagonal elements $\psi_0 = 0$. Two levels of sample sizes were simulated, with $N = 250$ and 500, and the number of replications was 200 for each sample size.

For the E-step recovery, a PCFA model with two specified loadings per factor was fitted, that is, λ_{11} and λ_{21} for Factor F1 and λ_{92} and $\lambda_{10,2}$ for Factor F2 (see Figure 2). To clarify, a structure matrix $\mathbf{R}$ with specified loading denoted by 1 and unspecified loading denoted by -1 can be defined corresponding to the factor loading matrix $\mathbf{\Lambda}$, as follows:

$$\mathbf{R}^T = \begin{bmatrix} 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix},$$

which implies four specified loadings and 16 Lasso loadings (8 nonzero and 8 zero, i.e., λ_0). For the C-step recovery, the PCFA model in Figure 1 with one loading per item was fitted. The equivalent structure matrix is as follows:

$$\mathbf{R}^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix},$$

which implies 10 specified loadings and 10 Lasso loadings (8 zero and 2 nonzero). Performance assessment involved computing the bias of the parameter estimates (BIAS), the mean of the standard error estimates (*SE*), the root mean square error (RMSE) between the estimates and the true values, and the percentage of estimates significantly different from zero at $\alpha = .05$ (SIG%) based on the HPD interval.

Simulation results for the E-step can be found in Table 2. The loading recovery in terms of BIAS and RMSE could be poor when the items were associated with off-diagonal residual elements, but was acceptable otherwise. Poor loading estimates could not be improved with a larger sample size, which suggests a systemic error due to ignoring the nonzero off-diagonal elements. Additional simulation studies revealed that the biases of estimates were also related to the level of residual correlations, the sparsity of the residual covariance matrix, and the magnitude of the cross-loadings (see Appendix D for details). Recovery of factorial correlations was also poor with large *SE* and lower power (i.e., SIG%). However, the Type I error rates for misidentifying zero loadings (i.e., λ_0) as significant were almost zero, and the power for detecting significant loadings was close to one. This result suggests that all significant loading estimates will be correctly identified in the E-step and accordingly can be used as specified loadings to satisfy the sufficient constraint for the C-step.

Additional simulation studies were conducted with the same data sets by fitting the model in the E-step with one specified loading per factor (that is, λ_{11} for F1 and $\lambda_{10,2}$ for F2) or no

Table 2
E-Step Results in Study 1

		$N = 250$				$N = 500$			
Par	True	BIAS	RMSE	*SE*	SIG%	BIAS	RMSE	*SE*	SIG%
λ_{11}	0.7	0.210	0.219	0.122	0.962	0.230	0.240	0.105	0.985
λ_{21}	0.7	0.220	0.228	0.123	0.962	0.231	0.240	0.105	0.985
λ_{31}	0.7	−0.111	0.125	0.096	0.971	−0.107	0.118	0.077	0.990
λ_{41}	0.7	−0.130	0.142	0.098	0.976	−0.131	0.141	0.079	0.990
λ_{51}	0.5	−0.054	0.079	0.133	0.986	−0.059	0.081	0.118	0.990
λ_{61}	0.5	−0.063	0.085	0.133	0.971	−0.060	0.081	0.117	0.980
λ_{52}	0.5	−0.002	0.058	0.133	0.981	0.002	0.058	0.118	0.965
λ_{62}	0.5	0.002	0.059	0.133	0.976	−0.002	0.057	0.118	0.965
λ_{72}	0.7	−0.030	0.068	0.124	0.929	−0.023	0.077	0.099	0.955
λ_{82}	0.7	−0.030	0.071	0.122	0.929	−0.025	0.078	0.098	0.955
λ_{92}	0.7	−0.030	0.075	0.123	0.929	−0.025	0.078	0.098	0.955
$\lambda_{10,2}$	0.7	−0.035	0.072	0.122	0.929	−0.032	0.082	0.097	0.955
λ_0	0	−0.006	0.102	0.168	0.002	−0.009	0.110	0.151	0.004
ϕ_{12}	0.3	0.039	0.072	0.219	0.176	0.063	0.091	0.204	0.475
ψ_{55}	0.35	0.051	0.068	0.044	1.000	0.037	0.050	0.030	1.000
ψ_{66}	0.35	0.050	0.070	0.044	1.000	0.042	0.053	0.030	1.000
ψ_{jj}	0.51	−0.019	0.129	0.055	1.000	−0.028	0.121	0.038	1.000
δ_l	—	4.148	0.571	2.231	1.000	4.153	0.354	2.199	1.000

Note. ψ_{jj} averaged across elements from $j = 1$ to 4 and 7 to 10; λ_0 averaged across all zero loading estimates; for δ_l, BIAS = mean; RMSE = *SD*.

specified loading (that is, all loadings were unspecified). The results presented in Table 3 indicate that the Type I error rates decreased and the power increased when the number of specified loadings per factor increased from zero to two. The case of one specified loading per factor was acceptable in general and can serve as the minimum constraint, while two or more specified loadings per factor is preferred for more stable performance in the E-step.

Table 4 summarizes the simulation results for the C-step. The loading estimates were acceptable in general, with those for the first factor relatively worse likely due to the within-factor residual covariance ψ_{12}. The recovery for nonzero Lasso loadings (i.e., λ_{61}, λ_{52}) was also slightly worse. Recovery of factorial correlation was better than that in the E-step, but the power for the smaller sample size was still low. The Type I error rates for both zero Lasso loadings and zero residual terms were zero, suggesting the conservative nature of the Bayesian Lasso. The power for ψ_{12} was also close to zero, making it difficult to identify nonzero residual covariance within one factor. The reason was likely because the related λ_{11} and λ_{21} estimates were positively biased. Additional simulation studies revealed that power was also related to the position of the nonzero elements (that is, within-factor or between-factor) and the magnitude of residual correlation (see Appendix D for details). Finally, all estimates improved with a larger sample size.

The estimates of the shrinkage parameters (Tables 2 and 4) were about the same across different sample sizes, suggesting that the penalty was largely independent of the sample size. The increase of δ_l from the E-step to the C-step might hint at interference between the L_1 penalties on the loadings and off-diagonal residual elements. To better investigate the impact of the Lasso technique in parameter estimation, one simulated data set was randomly chosen, and the loadings $\lambda_{10,1}$, λ_{12}, λ_{61}, and λ_{52} were treated as specified and unspecified, respectively. The corresponding posterior distributions (see Figure 5) suggest that, compared with the specified loadings, the unspecified loadings always pushed toward zero with the Lasso prior, and cases with smaller values were more remarkable.

Study 2: Performance in a More Complex Case

Study 2 evaluated the performance of the E-step and C-step in a more complex case with three factors and 18 items, namely, $J = 18$ and $K = 3$, through a simulation study. The factorial correlations were set as $\phi_{kk'} = .3$ for k and $K' = 1$ to 3 and $k \neq k'$. The loading matrix was

$$\mathbf{\Lambda}^T = \begin{bmatrix} \lambda_{11} & \lambda_{21} & \lambda_{31} & \lambda_{41} & \lambda_{51} & \lambda_{61} & \lambda_{71} & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_{18,1} \\ \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_{62} & \lambda_{72} & \lambda_{82} & \lambda_{92} & \lambda_{10,2} & \lambda_{11,2} & \lambda_{12,2} & \lambda_{13,2} & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 \\ \lambda_{13} & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_{12,3} & \lambda_{13,3} & \lambda_{14,3} & \lambda_{15,3} & \lambda_{16,3} & \lambda_{17,3} & \lambda_{18,3} \end{bmatrix}.$$

The following loadings were set as .7: λ_{j1} for $j = 2$ to 5, λ_{j2} for $j = 8$ to 11, and λ_{j3} for $j = 14$ to 17; the following loadings were set as .5: λ_{11}, λ_{61}, λ_{71}, $\lambda_{18,1}$, λ_{62}, λ_{72}, $\lambda_{12,2}$, $\lambda_{13,2}$, λ_{13}, $\lambda_{12,3}$, $\lambda_{13,3}$, and $\lambda_{18,3}$; and all other loadings were set as $\lambda_0 = 0$. For the residual structure in $\mathbf{\Psi}$, the diagonal elements were $\psi_{jj} = .51$ for $j = 2$ to 5, 8 to 11, and 14 to 17 and $\psi_{jj} = .35$ for $j = 1$, 6, 7, 12, 13, and 18. The nonzero off-diagonal elements for the lower triangle were $\psi_{43} = \psi_{16,15} = \psi_{10,9} = \psi_{85} = \psi_{14,11} = \psi_{17,2} = .3$ with a symmetric upper triangle. The first three terms were within-factor whereas the last three were between-factor. All other off-diagonal elements were set as $\psi_0 = 0$. This was a balanced design, as each factor was measured by four items with cross-loadings, two items with within-factor residual covariance, and two items with between-factor residual covariance.

The structure matrix $\mathbf{R}^T$ for the E-step and C-step were, respectively,

Table 3
Type I Error Rate and Power of Bayesian Lasso in the E-Step

		$N = 250$			$N = 500$		
Par	True	SIG%(0)	SIG%(1)	SIG%(2)	SIG%(0)	SIG%(1)	SIG%(2)
λ_{11}	0.7	0.270	0.890	0.962	0.330	0.950	0.985
λ_{21}	0.7	0.270	0.890	0.962	0.330	0.950	0.985
λ_{31}	0.7	0.370	0.900	0.971	0.350	0.960	0.990
λ_{41}	0.7	0.450	0.910	0.976	0.460	0.970	0.990
λ_{51}	0.5	0.690	0.690	0.986	0.770	0.750	0.990
λ_{61}	0.5	0.640	0.630	0.971	0.760	0.800	0.980
λ_{52}	0.5	0.770	1.000	0.981	0.750	0.990	0.965
λ_{62}	0.5	0.750	0.990	0.976	0.750	0.990	0.965
λ_{72}	0.7	0.290	0.890	0.929	0.380	0.960	0.955
λ_{82}	0.7	0.290	0.890	0.929	0.370	0.960	0.955
λ_{92}	0.7	0.290	0.890	0.929	0.370	0.960	0.955
$\lambda_{10,2}$	0.7	0.300	0.890	0.929	0.370	0.960	0.955
λ_0	0	0.270	0.014	0.002	0.401	0.017	0.004

Note. SIG%(0), SIG%(1), and SIG%(2) represent the percentage of estimates significantly different from zero at $\alpha = .05$ with no specified loading, one specified loading, and two specified loadings per factor, respectively.

Table 4
C-step Results in Study 1

		$N = 250$				$N = 500$			
Par	True	BIAS	RMSE	*SE*	SIG%	BIAS	RMSE	*SE*	SIG%
λ_{11}	0.7	0.081	0.092	0.137	1.000	0.078	0.087	0.123	1.000
λ_{21}	0.7	0.084	0.096	0.137	1.000	0.078	0.086	0.123	1.000
λ_{31}	0.7	−0.067	0.081	0.131	1.000	−0.058	0.068	0.115	1.000
λ_{41}	0.7	−0.087	0.096	0.130	0.995	−0.082	0.089	0.116	1.000
λ_{51}	0.5	−0.039	0.064	0.133	0.995	−0.043	0.059	0.116	0.995
λ_{61}	0.5	−0.128	0.139	0.139	0.780	−0.105	0.114	0.125	0.875
λ_{52}	0.5	−0.082	0.102	0.151	0.775	−0.061	0.075	0.131	0.910
λ_{62}	0.5	0.013	0.050	0.134	0.985	0.001	0.036	0.117	0.995
λ_{72}	0.7	−0.054	0.070	0.137	0.990	−0.046	0.056	0.117	1.000
λ_{82}	0.7	−0.029	0.053	0.134	0.990	−0.015	0.035	0.112	1.000
λ_{92}	0.7	−0.025	0.052	0.134	0.990	−0.016	0.036	0.112	1.000
$\lambda_{10,2}$	0.7	−0.027	0.054	0.135	1.000	−0.017	0.036	0.112	1.000
λ_0	0	−0.008	0.043	0.111	0.000	−0.008	0.037	0.104	0.000
ϕ_{12}	0.3	0.069	0.100	0.191	0.560	0.086	0.101	0.172	0.850
ψ_{21}	0.3	−0.109	0.118	0.146	0.005	−0.102	0.110	0.138	0.040
ψ_{74}	0.3	−0.077	0.086	0.098	0.755	−0.070	0.076	0.089	0.915
ψ_{55}	0.35	0.106	0.116	0.114	1.000	0.083	0.091	0.095	1.000
ψ_{66}	0.35	0.097	0.109	0.111	1.000	0.082	0.090	0.093	1.000
ψ_{jj}	0.51	0.025	0.091	0.133	1.000	0.014	0.075	0.118	1.000
ψ_0	0	0.023	0.043	0.082	0.000	0.019	0.036	0.071	0.000
δ_l	—	7.584	0.488	2.653	1.000	7.543	0.355	2.528	1.000
δ_s	—	2.188	0.125	0.354	1.000	2.276	0.100	0.360	1.000

Note. ψ_{jj} averaged across elements from $j = 1$ to 4 and 7 to 10; λ_0 averaged across all zero loading estimates; for δ_l and δ_s, BIAS = mean; RMSE = *SD*.

$$\begin{bmatrix} 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

This implies that there were six specified loadings and 48 Lasso loadings (30 zero and 18 nonzero) for the E-step, and 18 specified loadings and 36 Lasso loadings (30 zero and 6 nonzero) for the C-step. The sample size was $N = 250$, and other conditions were similar to the above.

Table 5 summarizes the simulation results for the E-step and C-step. The performance of both steps in Study 2 was clearly better than for the simpler Study 1 case above. For the E-step, loading estimates involving residual covariance were worse, but all nonsignificant and significant loadings could be successfully identified. Not surprisingly, the C-step estimates were better than the E-step counterpart. In the C-step, the nonzero Lasso loadings (e.g., λ_{71}, $\lambda_{18,1}$) performed similarly well as specified loadings. Loading estimates with within-factor residual covariance (e.g., λ_{31}, λ_{41}) were still positively biased, but within a reasonable range. The extremely low Type I error rates for both zero Lasso loadings and zero residual terms were desirable. Recovery of factorial correlations and most residual terms was also satisfactory. Of a little concern was the power to identify within-factor residual covariance. In general, the results suggest that the PCFA approach has more satisfactory performance in this more complex case.

Additional simulation studies were conducted across different situations, including more heterogeneous factor loadings, different patterns of the loading and residual covariance matrix, and a different number of specified loadings. In brief, the results were generally similar to or consistent with the findings in these two simulation studies, and details can be found in Appendix D.

Study 3: Questionnaire of Mathematics Learning

During scale development, it is common to have some items with unclear factor loading, and this issue is well known within the scope of the PCFA approach. Table 6 lists potential items for a questionnaire addressing anxiety, confidence (i.e., self-efficacy), and utility (valuing) in mathematics learning. A similar background questionnaire can be found in large-scale assessments such as the Trends in International Mathematics and Science Study (TIMSS; International Association for the Evaluation of Educational Achievement, 2011). For content analysis, three mathematics educators were asked to rate if the items measured specific factors clearly, clearly not, or unclearly with 1, −1, or 0, respectively. The educators were encouraged to be more conservative, namely, to give a ranking of 1 or −1 only

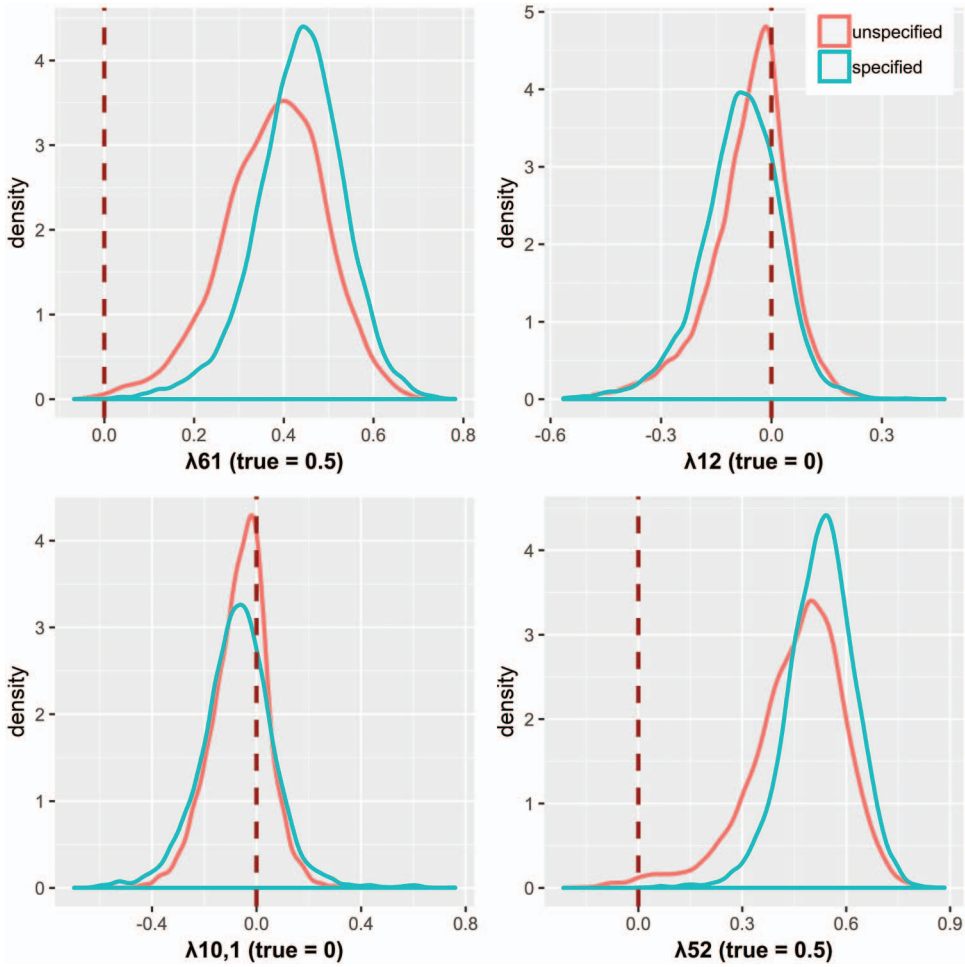

Figure 5. Posterior distributions of specified and unspecified loadings. See the online article for the color version of this figure.

when they were very confident and to use 0 as the default value. The mean ratings in Table 6 reflect the degree of uncertainty for item-factor fitness from a substantive perspective, which cannot be easily accommodated under the traditional EFA or CFA approach. Under the PCFA approach, one can simply designate those elements with 1 as specified loadings and all others as Lasso loadings.

Responses from 218 Chinese fifth grade students were collected and analyzed using the E-step and C-step procedures in sequence, with estimation results given in Table 7. Significant Lasso loadings identified in the E-step were specified to satisfy the constraint necessary for the C-step. Note that for Item 14, two loadings were found significant with close values in the E-step and accordingly were both specified in the C-step. No significant residual covariance can be identified in the C-step, which can explain the similarity of estimation results between the two steps. One can also triangulate the results with the BLCCFA model by fixing all significant Lasso loadings in the C-step as specified and nonsignificant Lasso loadings as zero. Not surprisingly, the difference was trivial (results not shown to save space, but available upon request).

Study 4: Humor Styles Questionnaire

When a more confirmatory approach is adopted to develop a psychological scale, the C-step can be used directly. The Humor Styles Questionnaire (HSQ; Martin, Puhlik-Doris, Larsen, Gray, & Weir, 2003) with four factors and 32 items can be found in Appendix E. As each item supposedly loads on only one factor, the development process of the HSQ was oriented toward a confirmatory approach. Although the major loading of each item was confirmed, researchers were concerned about cross-loadings for some items, as the related behaviors tended to be multidimensional (Heintz, 2017).

Data for HSQ from 1,070 respondents were publicly available at https://openpsychometrics.org/_rawdata/. Dissatisfactory fitting as shown in Table 8 was obtained using standard CFA, with no cross-loading or residual covariance in Mplus (L. K. Muthén & Muthén, 1998–2010) with the robust maximum likelihood estimation. Moreover, the PMM approach was challenging since many cross-loadings and residual covariance terms were suggested based on the modification index. In contrast, the C-step with major and Lasso loadings and residual covariance structure

Table 5
Simulation Results for Study 2

		E-step				C-step			
Par	True	BIAS	RMSE	*SE*	SIG%	BIAS	RMSE	*SE*	SIG%
λ_{11}	0.5	−0.014	0.064	0.093	1.000	−0.010	0.051	0.096	1.000
λ_{21}	0.7	−0.056	0.083	0.083	1.000	−0.035	0.053	0.104	1.000
λ_{31}	0.7	0.162	0.169	0.085	1.000	0.082	0.093	0.106	1.000
λ_{41}	0.7	0.163	0.168	0.085	1.000	0.084	0.093	0.106	1.000
λ_{51}	0.7	−0.069	0.090	0.082	1.000	−0.034	0.054	0.104	1.000
λ_{61}	0.5	−0.039	0.068	0.096	1.000	−0.007	0.049	0.096	1.000
λ_{71}	0.5	−0.041	0.076	0.095	1.000	−0.056	0.081	0.099	1.000
$\lambda_{18,1}$	0.5	−0.018	0.063	0.093	1.000	−0.052	0.076	0.099	0.995
λ_{62}	0.5	−0.033	0.067	0.094	1.000	−0.061	0.084	0.100	1.000
λ_{72}	0.5	−0.039	0.073	0.094	1.000	−0.017	0.055	0.097	1.000
λ_{82}	0.7	−0.057	0.082	0.083	1.000	−0.042	0.062	0.107	1.000
λ_{92}	0.7	0.179	0.184	0.087	1.000	0.080	0.090	0.108	1.000
$\lambda_{10,2}$	0.7	0.174	0.180	0.086	1.000	0.087	0.096	0.108	1.000
$\lambda_{11,2}$	0.7	−0.066	0.085	0.083	1.000	−0.038	0.055	0.105	1.000
$\lambda_{12,2}$	0.5	−0.039	0.067	0.094	1.000	−0.012	0.049	0.098	1.000
$\lambda_{13,2}$	0.5	−0.036	0.071	0.094	1.000	−0.057	0.080	0.101	0.990
λ_{13}	0.5	−0.048	0.081	0.096	1.000	−0.054	0.078	0.101	1.000
$\lambda_{12,3}$	0.5	−0.046	0.076	0.093	1.000	−0.057	0.078	0.100	0.995
$\lambda_{13,3}$	0.5	−0.046	0.075	0.094	1.000	−0.012	0.054	0.097	1.000
$\lambda_{14,3}$	0.7	−0.089	0.109	0.082	1.000	−0.042	0.061	0.106	1.000
$\lambda_{15,3}$	0.7	0.190	0.195	0.087	1.000	0.081	0.091	0.108	1.000
$\lambda_{16,3}$	0.7	0.191	0.196	0.087	1.000	0.082	0.092	0.108	1.000
$\lambda_{17,3}$	0.7	−0.084	0.105	0.083	1.000	−0.041	0.060	0.105	1.000
$\lambda_{18,3}$	0.5	−0.053	0.079	0.096	1.000	−0.014	0.053	0.096	1.000
λ_0	0	−0.013	0.070	0.107	0.002	−0.007	0.038	0.086	0.000
$\phi_{kk'}$	0.3	0.027	0.068	0.156	0.612	0.031	0.074	0.135	0.805
ψ_{jj}	0.35	0.064	0.079	0.043	1.000	0.071	0.084	0.073	1.000
ψ_{ii}	0.51	−0.065	0.149	0.052	1.000	−0.009	0.081	0.105	1.000
ψ_w	0.3	—	—	—	—	−0.090	0.101	0.106	0.702
ψ_b	0.3	—	—	—	—	−0.062	0.071	0.070	1.000
ψ_0	0	—	—	—	—	0.005	0.028	0.053	0.000
δ_l	—	6.059	0.558	1.912	1.000	7.789	0.302	1.468	1.000
δ_s	—	—	—	—	—	2.615	0.130	0.248	1.000

Note. For $\phi_{kk'}$, k and $K' = 1$ to 3 and $k \neq k'$; for ψ_{jj}, $j = 2$ to 4, 8 to 12, and 14 to 17; for ψ_{ii}, $i = 1$, 6, 7, 11, 12, and 18; ψ_w averaged across ψ_{43}, $\psi_{16,15}$, and $\psi_{10,9}$ (within factor); ψ_b averaged across ψ_{85}, $\psi_{14,11}$, and $\psi_{17,2}$ (between factors); for δ_l and δ_s, BIAS = mean; RMSE = *SD*.

can be directly applied. The estimation resulted in seven significant cross-loadings and 15 significant residual covariance terms, as shown in Table 9. It appeared that both cross-loadings and residual covariance contributed to model fitness. Note that most significant cross-loadings were around .1 and only six significant residual terms were above .1. This result suggests that many estimates were significant merely due to the sample size.

A BLCCFA model with all significant loadings in the C-step as specified was used for triangulation. It resulted in the same sig-

Table 6
Questionnaire of Mathematics Learning and the Mean Rating of Item-Factor Fitness

#	Content	F1	F2	F3
1	Mathematics makes me nervous	1.00	0.00	−1.00
2	Mathematics exams are easy for me	−0.33	1.00	−1.00
3	I worry about mathematics exams	1.00	0.00	−1.00
4	I am not bothered by mathematics courses	0.33	0.67	−0.67
5	I find mathematics confusing	0.67	0.33	−0.33
6	My brain is empty when solving mathematics problem	0.67	0.33	−0.33
7	Solving difficult mathematics problem is one of my strengths	−0.33	1.00	−1.00
8	I usually do well in mathematics	−0.67	1.00	−1.00
9	I learn things quickly in mathematics	−0.67	0.67	−1.00
10	Mathematics helps me in my daily life	−0.67	−1.00	1.00
11	I think mathematics will help me in my future life	−0.67	0.00	0.67
12	I need to do well in mathematics to get into the university of my choice	−1.00	−0.33	1.00
13	I need to do well in mathematics to get the job I want	−1.00	−0.33	1.00
14	I am willing to take more mathematics courses	−1.00	0.33	0.33

Note. All items employed a 5-point Likert scale, with 1 indicating *agree a lot* and 5 indicating *disagree a lot*; F1 = anxiety; F2 = confidence; F3 = utility.

Table 7
Parameter Estimates for the Questionnaire of Mathematics Learning

	E-step				C-step			
Item	F1	F2	F3	RV	F1	F2	F3	RV
1	0.708*	0.056	−0.044	0.522	0.660*	0.024	−0.041	0.587
2	−0.157	0.573*	−0.057	0.623	−0.116	0.560*	−0.019	0.657
3	0.653*	−0.078	0.131	0.603	0.584*	−0.026	0.016	0.676
4	−0.058	0.269*	0.099	0.856	−0.027	0.366*	0.040	0.850
5	0.652*	0.005	−0.044	0.568	0.627*	−0.016	−0.025	0.615
6	0.607*	−0.054	−0.064	0.562	0.637*	−0.057	−0.032	0.573
7	0.007	0.746*	−0.070	0.519	−0.008	0.671*	−0.010	0.586
8	0.028	0.704*	0.034	0.501	0.013	0.650*	0.063	0.565
9	0.088	0.485*	0.149	0.681	0.051	0.574*	0.051	0.681
10	−0.113	0.064	0.427*	0.710	−0.066	0.016	0.534*	0.692
11	−0.047	0.085	0.560*	0.584	−0.006	0.042	0.653*	0.569
12	−0.066	−0.042	0.593*	0.649	−0.060	−0.008	0.551*	0.690
13	0.076	−0.006	0.783*	0.440	0.052	0.042	0.678*	0.564
14	−0.042	0.319*	0.306*	0.650	−0.014	0.330*	0.339*	0.654

Note. F1 = anxiety; F2 = confidence; F3 = utility; RV = residual variance. Underscored are specified loadings.
* Significant loadings.

nificant loadings and 88 significant residual covariance terms, 19 of which were above .1 (see Table 9). The loading estimates were generally close, whereas there was some difference between the residual structures, suggesting the impact of the Lasso loadings.

Three standard CFA models were fitted with the same significant loadings identified in the C-step but with different residual structures: residual covariance above .1 in the C-step (M1), all significant residual covariance in the C-step (M2), and residual covariance above .1 in BLCCFA (M3). As shown in Table 8, all three models were acceptable and the difference was small. Moreover, although M2 was the best model, M1 was not a bad choice considering the small number of residual terms involved. The results suggest that the C-step can offer some flexibility in choosing the final model when the sample size is large.

Discussion

A different amount of substantive knowledge can be available during scale development, making the EFA and CFA two ends of a continuum. The partially confirmatory approach with Bayesian Lasso is flexible in addressing a wide range of the substantive continuum that would be challenging otherwise. Specifically, a C-step can be adopted when the constraint of one specified loading per item is satisfied, and an additional E-step is recommended before applying the C-step when only a few loadings per factor can be specified. Note that the C-step reduces to the Bayesian covariance Lasso CFA when there is no Lasso loading, while the E-step is similar to the BSEM without residual covariance. The PCFA approach with Bayesian Lasso can be derived using the Bayesian hierarchical formulation with different Lasso and regular priors by combining the Bayesian regression and covariance Lasso. Moreover, this approach is implemented using MCMC estimation with both a block Gibbs sampler and an adapted Metropolis-Hastings algorithm.

Simulation studies showed that both steps can perform as expected, and the performance was better for the complex case with three factors than for the simple case with two factors. This finding is consistent with the nature of the L_1 norm penalty that can better select variables among many candidates and is meaningful for scale development with many factors, which is not uncommon in social and behavioral science. Study 3 suggested that a conservative approach can be adopted to specify the item-factor connections substantively with the help of both the E-step and C-step of the PCFA approach, allowing the data to give more information. This would be especially valuable when consensus or agreement among content experts is difficult to achieve. The two steps are also especially beneficial when the sample size is relatively small (e.g., ~200). Practitioners can apply an intermediate stage after qualitative content analysis but before collecting data on a large scale.

Table 8
CFA Model Fitness for the Humor Styles Questionnaire

Model	RMSEA	90%CI	CFI	TLI	SRMR	BIC	*df*	#RCT
M0	0.054	[0.052, 0.057]	0.844	0.831	0.065	96697	458	0
M1	0.040	[0.037, 0.043]	0.917	0.907	0.046	95981	445	6
M2	0.037	[0.035, 0.040]	0.929	0.920	0.044	95898	436	15
M3	0.038	[0.035, 0.041]	0.928	0.918	0.043	95923	432	19

Note. RMSEA = root mean square error; CFA = confirmatory factor analysis; M0 = no cross-loading or residual covariance; M1–M3 = all significant loadings identified in the C-step; M1 = residual covariance above .1 in the C-step; M2 = all significant residual covariance in the C-step; M3 = residual covariance above .1 in BLCCFA; #RCT = number of residual covariance term.

Table 9
Significant Loadings and Residual Estimates for the Humor Styles Questionnaire

	C-step				BLCCFA				RCV		
Item	F1	F2	F3	F4	F1	F2	F3	F4	Par	C-step	BLC
1	0.649				0.575				$\psi_{1,13}$		0.116
2		0.612				0.615			$\psi_{13,14}$		0.116
3			0.531				0.637		$\psi_{15,16}$	0.094	0.126
4				0.631				0.593	$\psi_{1,17}$		0.122
5	0.615				0.616				$\psi_{10,18}$	0.146	0.114
6	0.203	0.378			0.242	0.327			$\psi_{11,19}$	0.112	0.109
7	−0.142		0.580		−0.110		0.539		$\psi_{8,20}$	0.153	0.165
8				0.714				0.698	$\psi_{11,21}$		−0.130
9	0.452				0.461				$\psi_{17,21}$		0.142
10		0.726				0.732			$\psi_{11,23}$	−0.079	−0.073
11			0.540				0.483		$\psi_{4,24}$	0.096	0.123
12				0.635				0.610	$\psi_{15,24}$	−0.072	−0.084
13	0.614				0.530				$\psi_{1,25}$		0.124
14		0.674				0.669			$\psi_{5,25}$	−0.085	
15			0.647				0.579		$\psi_{13,25}$	0.147	0.263
16				0.567				0.592	$\psi_{17,25}$		0.130
17	0.697				0.624				$\psi_{21,25}$		0.142
18		0.723				0.690			$\psi_{23,25}$	0.086	0.116
19	0.178		0.452		0.344		0.435		$\psi_{27,28}$	0.092	
20				0.710				0.679	$\psi_{5,29}$	0.110	0.158
21	0.651				0.583				$\psi_{22,29}$	0.069	0.093
22		0.422				0.449			$\psi_{24,29}$		−0.113
23			0.489				0.496		$\psi_{25,29}$		0.110
24				0.483				0.423	$\psi_{6,30}$	0.218	0.235
25	0.634				0.484				$\psi_{22,31}$	0.071	0.097
26		0.666				0.686					
27			0.551				0.549				
28		0.140	0.160	0.249		0.190	0.271	0.221			
29	0.561		0.125	−0.163	0.472		0.128	−0.187			
30		0.447				0.397					
31			0.637				0.574				
32				0.659				0.688			

Note. BLCCFA = Bayesian covariance Lasso for CFA; F1 = affiliative humor; F2 = self-enhancing humor; F3 = aggressive humor; F3 = self-defeating humor; RCV = residual covariance; BLC = BLCCFA; Underscored are cross-loadings or residual covariance above .1; only residual covariance significant at the C-step or significant and above .1 in BLCCFA is presented.

Study 4 demonstrated the advantage of the C-step to identify significant cross-loadings and residual covariance when encountering many factors with a large number of items, a task that would be challenging with the traditional PMM approach. Although unnecessary because the constraint of one specified loading per item was well satisfied with the available literature, we tried the E-step with only two major loadings per factor and found that all other major loadings were successfully recovered. Accordingly, a combination of the E-step and C-step can offer more freedom during scale development. The C-step can also offer some flexibility in choosing a succinct model when the sample size is large. Finally, the Bayesian covariance Lasso CFA without any Lasso loading can be used for triangulation, as shown in both studies.

Although the PCFA with Bayesian Lasso approach appears to be promising, some issues can be addressed in future research to make the framework more comprehensive and versatile.

First, the PCFA approach assumes that substantive knowledge about the loading structure is partially available but knowledge about the factors (i.e., number of factors and their operational definitions) is largely available. During scale development, however, it is likely that there is a different amount of knowledge about the factors (e.g., a range of factor numbers or blurred definitions). It would be valuable but also challenging to extend the PCFA to accommodate these scenarios.

Second, it would be useful to equip the proposed framework with powerful means of model comparison to allow comparison of different substantive knowledge. One possible but also challenging means is to rely on Bayes factor for comparison.

Third, categorical data are widely available in social and behavioral research, as shown in Studies 3 and 4. These data are treated as continuous in the existing approach, which is suboptimal and can result in substantial bias without favorite conditions (e.g., normality). Accordingly, it would be valuable to extend the approach to address dichotomous and polytomous responses, possibly under the umbrella of item response theory.

Finally, Lu, Chow, and Loken (2016) proposed a Bayesian CFA with the spike-and-slab prior (SSP) and found that their approach outperformed the ridge regression prior and yielded more parsimonious loading structures. However, there is no research to compare the performance of Bayesian Lasso in estimating the loadings with that of the SSP under the CFA context. Future studies can further investigate if the SSP can be incorporated into the PCFA framework, and if it is possible, explain how that differs from the existing approach.

References

Bentler, P. M. (2007). On tests and indices for evaluating structural models. *Personality and Individual Differences, 42,* 825–829. http://dx.doi.org/10.1016/j.paid.2006.09.024

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. Reading, MA: Addison Wesley.

Casella, G., & George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician, 46,* 167–174. http://dx.doi.org/10.1080/00031305.1992.10475878

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician, 49,* 327–335. http://dx.doi.org/10.1080/00031305.1995.10476177

Gelman, A. (1996). Inference and monitoring convergence. In W. R. Gilks, S. Richardson, & D. J. Spiegelharter (Eds.), *Markov chain Monte Carlo in practice* (pp. 131–144). London, UK: Chapman & Hall.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). London, UK: Chapman & Hall.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6,* 721–741. http://dx.doi.org/10.1109/TPAMI.1984.4767596

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (Eds.). (1996). *Markov chain Monte Carlo in practice*. London, UK: Chapman & Hall.

Gill, J. (2002). *Bayesian methods: A social and behavioral sciences approach*. Boca Raton, FL: Chapman & Hall/CRC. http://dx.doi.org/10.1201/9781420057478

Hans, C. (2009). Bayesian lasso regression. *Biometrika, 96,* 835–845. http://dx.doi.org/10.1093/biomet/asp047

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their application. *Biometrika, 57,* 97–109. http://dx.doi.org/10.1093/biomet/57.1.97

Heintz, S. (2017). Putting a spotlight on daily humor behaviors: Dimensionality and relationships with personality, subjective well-being, and humor styles. *Personality and Individual Differences, 104,* 407–412. http://dx.doi.org/10.1016/j.paid.2016.08.042

Hsiang, T. C. (1975). A Bayesian view on ridge regression. *The Statistician, 24,* 267–268. http://dx.doi.org/10.2307/2987923

International Association for the Evaluation of Educational Achievement. (2011). *TIMSS 2011 student questionnaire (Grade 8)*. Boston, MA: Author.

Jennrich, R. I., & Sampson, P. F. (1966). Rotation for simple loadings. *Psychometrika, 31,* 313–323. http://dx.doi.org/10.1007/BF02289465

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34,* 183–202. http://dx.doi.org/10.1007/BF02289343

Kaplan, D. (1990). Evaluating and modifying covariance structure models: A review and recommendation. *Multivariate Behavioral Research, 25,* 137–155. http://dx.doi.org/10.1207/s15327906mbr2502_1

Kaplan, D. (2009). *Structural equation modeling: Foundations and extensions* (2nd ed.). Thousand Oaks, CA: Sage. http://dx.doi.org/10.4135/9781452226576

Khondker, Z. S., Zhu, H., Chu, H., Lin, W., & Ibrahim, J. G. (2013). The Bayesian covariance lasso. *Statistics and Its Interface, 6,* 243–259. http://dx.doi.org/10.4310/SII.2013.v6.n2.a8

Lee, S.-Y. (2007). *Structural equation modeling: A Bayesian approach*. Hoboken, NJ: Wiley. http://dx.doi.org/10.1002/9780470024737

Liu, X. (2008). Parameter expansion for sampling a correlation matrix: An efficient GPX-RPMH algorithm. *Journal of Statistical Computation and Simulation, 78,* 1065–1076. http://dx.doi.org/10.1080/00949650701519635

Liu, X., & Daniels, M. J. (2006). A new efficient algorithm for sampling a correlation matrix based on parameter expansion and re-parameterization. *Journal of Computational and Graphical Statistics, 15,* 897–914. http://dx.doi.org/10.1198/106186006X160681

Lu, Z. H., Chow, S. M., & Loken, E. (2016). Bayesian factor analysis as a variable-selection problem: Alternative priors and consequences. *Multivariate Behavioral Research, 51,* 519–539. http://dx.doi.org/10.1080/00273171.2016.1168279

MacCallum, R. C. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin, 100,* 107–120. http://dx.doi.org/10.1037/0033-2909.100.1.107

Martin, R. A., Puhlik-Doris, P., Larsen, G., Gray, J., & Weir, K. (2003). Individual differences in uses of humor and their relation to psychological well-being: Development of the Humor Styles Questionnaire. *Journal of Research in Personality, 37,* 48–75. http://dx.doi.org/10.1016/S0092-6566(02)00534-2

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equations of state calculations by fast computing machine. *The Journal of Chemical Physics, 21,* 1087–1091. http://dx.doi.org/10.1063/1.1699114

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods, 17,* 313–335. http://dx.doi.org/10.1037/a0026802

Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Author.

O'Hara, R. B., & Sillanpää, M. J. (2009). A review of Bayesian variable selection methods: What, how and which. *Bayesian Analysis, 4,* 85–117. http://dx.doi.org/10.1214/09-BA403

Pan, J., Ip, E. H., & Dubé, L. (2017). An alternative to post hoc model modification in confirmatory factor analysis: The Bayesian lasso. *Psychological Methods, 22,* 687–704. http://dx.doi.org/10.1037/met0000112

Park, T., & Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association, 103,* 681–686. http://dx.doi.org/10.1198/016214508000000337

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News, 6,* 7–11.

R Development Core Team. (2010). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Sörbom, D. (1989). Model modification. *Psychometrika, 54,* 371–384. http://dx.doi.org/10.1007/BF02294623

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25,* 173–180. http://dx.doi.org/10.1207/s15327906mbr2502_4

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B. Methodological, 58,* 267–288. http://dx.doi.org/10.1111/j.2517-6161.1996.tb02080.x

Wang, H. (2012). Bayesian graphical lasso models and efficient posterior computation. *Bayesian Analysis, 7,* 867–886. http://dx.doi.org/10.1214/12-BA729

(*Appendices follow*)

Appendix A

Full Conditional Distribution for $\mathbf{\Lambda}_j''$ in the C-Step

Equation 1 can be rewritten as $y_{ij} = \boldsymbol{\mu}_j + \mathbf{\Lambda}_j^T\boldsymbol{\omega}_i + \boldsymbol{\varepsilon}_{ij}$, where $\mathbf{\Lambda}_j^T = (\lambda_{j1}, \lambda_{j2}, \ldots \lambda_{jK})$ is the jth row of $\mathbf{\Lambda}$. After rearrange the latent factors and loading vector for item j, the factor vector is $\boldsymbol{\omega}_i = \begin{pmatrix}\boldsymbol{\omega}_i' \\ \boldsymbol{\omega}_i''\end{pmatrix}$ where the $K' \times 1$ vector $\boldsymbol{\omega}_i'$ and $K'' \times 1$ vector $\boldsymbol{\omega}_i''$ are associated with the zero or specified loadings and unspecified loadings, respectively. Similarly, denote the rearranged factor matrix as $\mathbf{\Omega} = \begin{pmatrix}\mathbf{\Omega}' \\ \mathbf{\Omega}''\end{pmatrix} = \begin{pmatrix}\boldsymbol{\omega}_1' \ldots \boldsymbol{\omega}_i' \ldots \boldsymbol{\omega}_N' \\ \boldsymbol{\omega}_1'' \ldots \boldsymbol{\omega}_i'' \ldots \boldsymbol{\omega}_N''\end{pmatrix}$. The conditional distribution for $\mathbf{\Lambda}_i''$ can be expressed as:

$$
\begin{aligned}
&p(\mathbf{\Lambda}_j''|\mathbf{Y}, \mathbf{\Lambda}_{-j}, \mathbf{\Omega}, \boldsymbol{\mu}, \mathbf{\Phi}, \mathbf{\Psi}, \boldsymbol{\delta}_{lj}) \propto p(\mathbf{\Lambda}_j''|\mathbf{y}_j, \mathbf{y}_{-j}, \mathbf{\Lambda}_{-j}, \mathbf{\Omega}, \mathbf{\Psi}, \boldsymbol{\delta}_l) \\
&\propto p(\mathbf{y}_j|\mathbf{y}_{-j}, \mathbf{\Lambda}_j'', \mathbf{\Lambda}_{-j}, \mathbf{\Omega}, \mathbf{\Psi})p(\mathbf{\Lambda}_j'') \\
&= \prod_{i=1}^{N} p(y_{ij}|y_{i(-j)}, \mathbf{\Lambda}_j'', \mathbf{\Lambda}_{-j}, \boldsymbol{\omega}_i, \mathbf{\Psi})p(\mathbf{\Lambda}_j'') \\
&\propto \prod_{i=1}^{N} \exp\Big\{\frac{1}{2}\big[y_{ij} - \mu_j - (\mathbf{\Lambda}_j'')^T\boldsymbol{\omega}_i'' - \psi_j^T\mathbf{\Psi}_{-jj}^{-1}(\mathbf{y}_{i(-j)} - \boldsymbol{\mu}_{-j} \\
&\quad - \mathbf{\Lambda}_{-j}^T\boldsymbol{\omega}_i)\big]^T \times (\psi_{jj} - \psi_j^T\mathbf{\Psi}_{-jj}^{-1}\psi_j)^{-1}\big[y_{ij} - \mu_j - (\mathbf{\Lambda}_j'')^T\boldsymbol{\omega}_i'' \\
&\quad - \psi_j^T\mathbf{\Psi}_{-jj}^{-1}(\mathbf{y}_{i(-j)} - \boldsymbol{\mu}_{-j} - \mathbf{\Lambda}_{-j}^T\boldsymbol{\omega}_i)\big]\Big\} \times p(\mathbf{\Lambda}_j'').
\end{aligned}
$$

where ψ_{jj} is the jth diagonal element of $\mathbf{\Psi}$, $\psi_j = (\psi_{j1}, \ldots, \psi_{j,j-1}, \psi_{j,j+1}, \ldots, \psi_{jJ})^T$ is the vector of all off-diagonal elements of the jth column, and $\mathbf{\Psi}_{-jj}$ is the $(J-1) \times (J-1)$ matrix resulting from deleting the jth row and jth column from $\mathbf{\Psi}$.

Let $\mathbf{\Psi}_j^* = \psi_{jj} - \psi_j^T\mathbf{\Psi}_{-jj}^{-1}\psi_j$, $\mathbf{y}_{ij}^* = y_{ij} - \mu_j - \psi_j^T\mathbf{\Psi}_{-jj}^{-1}(\mathbf{y}_{i(-j)} - \boldsymbol{\mu}_{-j} - \mathbf{\Lambda}_{-j}^T\boldsymbol{\omega}_i)$, and $\mathbf{Y}_j^* = (\mathbf{y}_{1j}^*, \mathbf{y}_{2j}^*, \ldots \mathbf{y}_{Nj}^*)^T$. With Equation 5 as the prior for $\mathbf{\Lambda}_j''$, the right hand side becomes:

$$
\begin{aligned}
&\prod_{i=1}^{N} \exp\Big\{-\frac{1}{2}(\mathbf{y}_{ij}^* - (\mathbf{\Lambda}_j'')^T\boldsymbol{\omega}_i'')^T\mathbf{\Psi}_j^{*-1}(\mathbf{y}_{ij}^* - (\mathbf{\Lambda}_j'')^T\boldsymbol{\omega}_i'')\Big\} \times p(\mathbf{\Lambda}_j'') \\
&= \exp\Big\{-\frac{1}{2}\mathbf{\Psi}_j^{*-1}(\mathbf{Y}_j^{*T}\mathbf{Y}_j^* - 2(\mathbf{\Lambda}_j'')^T\mathbf{\Omega}''\mathbf{Y}_j^* + (\mathbf{\Lambda}_j'')^T(\mathbf{\Omega}''(\mathbf{\Omega}'')^T)\mathbf{\Lambda}_j'')\Big\} \times p(\mathbf{\Lambda}_j'').
\end{aligned}
$$

$\Rightarrow p(\mathbf{\Lambda}_j''|\mathbf{Y}, \mathbf{\Lambda}_{-j}, \mathbf{\Omega}, \boldsymbol{\mu}, \mathbf{\Phi}, \mathbf{\Psi}) \propto N(m, var)$, where $m = (\mathbf{\Psi}_j^{*-1}(\mathbf{\Omega}''(\mathbf{\Omega}'')^T) + \mathbf{D}_{\tau_j}^{-1})^{-1}(\mathbf{\Psi}_j^{*-1}\mathbf{\Omega}''\mathbf{Y}_j^*)$ and $var = (\mathbf{\Psi}_j^{*-1}(\mathbf{\Omega}''(\mathbf{\Omega}'')^T) + \mathbf{D}_{\tau_j}^{-1})^{-1}$.

The conditional distribution for τ_{jk} can be expressed as:

$$
\begin{aligned}
p(\tau_{jk}^2|\mathbf{Y}, \mathbf{\Lambda}, \mathbf{\Omega}, \boldsymbol{\mu}, \mathbf{\Phi}, \mathbf{\Psi}) &\propto p(\lambda_{jk}|\tau_{jk}^2)p(\tau_{jk}^2) \\
&\propto (\tau_{jk}^2)^{-\frac{1}{2}}\exp\Big\{-\frac{1}{2\tau_{jk}^2}\lambda_{jk}^2\Big\}\exp\Big\{-\frac{\delta_l^2}{2}\tau_{jk}^2\Big\}.
\end{aligned}
$$

$$
\Rightarrow p\Big(\frac{1}{\tau_{jk}^2}|\mathbf{Y}, \mathbf{\Lambda}, \mathbf{\Omega}, \boldsymbol{\mu}, \mathbf{\Phi}, \mathbf{\Psi}\Big) \propto \textit{Inv-Gaussian}\Big(\sqrt{\frac{\delta_l^2}{\lambda_{jk}^2}}, \delta_l^2\Big).
$$

The conditional distribution for δ_l can be expressed as:

$$
\begin{aligned}
p(\delta_l^2|\mathbf{Y}, \mathbf{\Lambda}, \mathbf{\Omega}, \boldsymbol{\mu}, \mathbf{\Phi}, \mathbf{\Psi}) &\propto p(\tau_{jk}^2|\delta_l^2)p(\delta_l^2) \\
&\propto \frac{\delta_l^2}{2}\exp\Big\{-\frac{\delta_l^2}{2}\tau_{jk}^2\Big\} \times (\delta_l^2)^{a_l-1}\exp\{-b_l(\delta_l^2)\} \\
&\propto (\delta_l^2)^{(a_l+1)-1}\exp\Big\{-\Big(\frac{\tau_{jk}^2}{2} + b_l\Big)(\delta_l^2)\Big\} \\
&\propto \textit{Gamma}\Big(a_l + 1, \frac{\tau_{jk}^2}{2} + b_l\Big).
\end{aligned}
$$

(*Appendices continue*)

Appendix B

Full Conditional Distribution for $\Lambda_j^{''}$ in the E-Step

Different from the C-step, $\boldsymbol{\Psi} = diag(\psi_{jj})$ is modeled as a diagonal matrix. Let $\mathbf{Y}_j^* = (\mathbf{y}_{1j}, \mathbf{y}_{2j}, \ldots \mathbf{y}_{Nj})^T$. With notations similar above and Equation 6 as the prior, the conditional distribution for $\Lambda_i^{''}$ can be expressed as:

$$p(\Lambda_j^{''} \mid \mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \delta_{lj}) \propto p(\mathbf{y}_j \mid \Lambda_j^{''}, \boldsymbol{\Omega}, \boldsymbol{\Psi}) p(\Lambda_j^{''} \mid \psi_{jj})$$

$$\propto \prod_{i=1}^{N} \exp\Big\{-\frac{1}{2}(y_{ij} - (\boldsymbol{\Lambda}_j^{''})^T \boldsymbol{\omega}_i^{''})^T \psi_{jj}^{-1} (y_{ij} - (\boldsymbol{\Lambda}_j^{''T} \boldsymbol{\omega}_i^{''}) \times p(\Lambda_j^{''} \mid \psi_{jj})$$

$\Rightarrow p(\Lambda_j^{''} \mid \mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \delta_l) \propto N(m, var)$, where $m = (\boldsymbol{\Omega}^{''}(\boldsymbol{\Omega}^{''})^T + \mathbf{D}_{\tau_j}^{-1})^{-1}(\Omega^{''}\mathbf{Y}_j^*)$ and $var = \psi_{jj}(\boldsymbol{\Omega}^{''}(\boldsymbol{\Omega}^{''})^T + \mathbf{D}_{\tau_j}^{-1})^{-1}$.

The conditional distribution for ψ_j can be expressed as:

$$p(\psi_{jj}^{-1} \mid \mathbf{Y}, \Lambda_j^{''}, \boldsymbol{\Omega}, \boldsymbol{\mu}, \boldsymbol{\Phi}) \propto p(\mathbf{y}_j \mid \mathbf{y}_{-j}, \Lambda_j^{''}, \boldsymbol{\Omega}, \boldsymbol{\Psi}) p(\Lambda_j^{''} \mid \psi_{jj}) p(\psi_{jj}^{-1}),$$

which can be simplified as

$$\propto Gamma\Big(a_{0j} + \frac{N}{2} + \frac{K}{2} - 1, b_{0j} + \frac{1}{2}\Big[\sum_{i=1}^{N}(y_{ij} - (\boldsymbol{\Lambda}_j^{''})^T \boldsymbol{\omega}_i^{''})^2 + (\boldsymbol{\Lambda}_j^{''})^T \mathbf{D}_{\tau_j}^{-1} \boldsymbol{\Lambda}_j^{''}\Big]\Big).$$

Similar to the C-step, the conditional distribution for τ_{jk} and δ_j can be expressed, respectively, as:

$$p(\frac{1}{\tau_{jk}^2} \mid \mathbf{Y}, \boldsymbol{\Lambda}, \boldsymbol{\Omega}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Psi}) \propto Inv\text{-}Gaussian\Big(\sqrt{\frac{\delta_l^2}{\lambda_{jk}^2}\psi_{jj}}, \delta_l^2\Big),$$

and

$$p(\delta_l^2 \mid \mathbf{Y}, \boldsymbol{\Lambda}, \boldsymbol{\Omega}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Psi}) \propto Gamma(a_l + 1, \frac{\tau_{jk}^2}{2} + b_l).$$

Appendix C

Estimation of the Residual Covariance Matrix in the C-Step

The conditional distribution $p(\boldsymbol{\Sigma} \mid \mathbf{Y}, \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\tau}, \delta_s)$ can be decomposed as follows:

$$p(\boldsymbol{\Sigma} \mid \mathbf{Y}, \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\tau}, \delta_s) \propto p(\mathbf{Y} \mid \boldsymbol{\Sigma}, \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\Lambda}, \boldsymbol{\Phi}) p(\boldsymbol{\tau}, \delta_s)$$

$$\propto |\boldsymbol{\Sigma}|^{N2} \exp\Big[-\Big(-\frac{1}{2}\mathbf{S}\boldsymbol{\Sigma}\Big)\Big] \prod_{i<j} \exp\Big(-\frac{\sigma_{ij}^2}{2\tau_{ij}}\Big) \times \prod_{j=1}^{J} \exp\Big(-\frac{\delta_s \sigma_{jj}}{2}\Big) \mathrm{I}(\boldsymbol{\Sigma} > 0),$$

where $\boldsymbol{\tau} = (\tau_{ij})_{i<j}$ is the vector of the latent scale parameters, and $\mathbf{S} = \sum_{i=1}^{N}(\mathbf{y}_i - \boldsymbol{\mu} - \boldsymbol{\Lambda}\boldsymbol{\omega}_i)(\mathbf{y}_i - \boldsymbol{\mu} - \boldsymbol{\Lambda}\boldsymbol{\omega}_i)$.

For $k = 1, 2, \ldots, J$, and without loss of generality, partition and rearrange the columns of $\boldsymbol{\Sigma}$ and $\mathbf{S}$ as follows:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{-kk} & \boldsymbol{\sigma}_k \\ \boldsymbol{\sigma}_k^T & \sigma_{kk} \end{pmatrix}, \mathbf{S} = \begin{pmatrix} \mathbf{S}_{-kk} & \boldsymbol{s}_k \\ \boldsymbol{s}_k^T & s_{kk} \end{pmatrix}.$$

Where σ_{kk} is the k-th diagonal element of $\boldsymbol{\Sigma}$, $\boldsymbol{\sigma}_k = (\sigma_{k1}, \ldots, \sigma_{k,k-1}, \sigma_{k,k+1}, \ldots, \sigma_{kJ})^T$ is the vector of all off-diagonal elements of the k-th column, and Σ_{-kk} is the $(p-1) \times (p-1)$ matrix resulting from deleting the k-th row and k-th column from $\boldsymbol{\Sigma}$. Similar, s_{kk} is the k-th diagonal element of $\mathbf{S}$, $\boldsymbol{s}_k$ is the vector of all-diagonal elements of the k-th column of $\mathbf{S}$, and $\mathbf{S}_{-kk}$ is the matrix with the k-th row and k-th column of $\mathbf{S}$ deleted. Therefore we have:

$$p(\boldsymbol{\sigma}_k, \sigma_{kk} \mid \Sigma_{-kk}, \mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\tau}, \delta_s) \propto (\sigma_{kk} - \boldsymbol{\sigma}_k^T \boldsymbol{\Sigma}_{-kk}^{-1} \boldsymbol{\sigma}_k)^{\frac{N}{2}} \times \exp\Big\{-\frac{1}{2}[\boldsymbol{\sigma}_k^T \mathbf{M}_\tau^T \boldsymbol{\sigma}_k + 2\boldsymbol{s}_k^T \boldsymbol{\sigma}_k + (s_{kk} + \delta_s)\sigma_{kk}]\Big\}$$

where $\mathbf{M}_\tau$ is the diagonal matrix with diagonal elements $\tau_{k1}, \ldots, \tau_{k,k-1}, \tau_{k,k+1}, \ldots, \tau_{kJ}$.

Let $\boldsymbol{\beta} = \boldsymbol{\sigma}_k$ and $\boldsymbol{\gamma} = \sigma_{kk} - \boldsymbol{\sigma}_k^T \boldsymbol{\Sigma}_{-kk}^{-1} \boldsymbol{\sigma}_k$. It can be shown that:

$$p(\boldsymbol{\beta} \mid \boldsymbol{\Sigma}_{-kk}, \mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\tau}, \delta_s)$$

$$\propto N(-[(s_{kk} + \delta_s)\boldsymbol{\Sigma}_{-kk}^{-1} + \mathbf{M}_\tau^{-1}]^{-1}\boldsymbol{s}_k, [(s_{kk} + \delta_s)\boldsymbol{\Sigma}_{-kk}^{-1} + \mathbf{M}_\tau^{-1}]^{-1}),$$

$$p(\boldsymbol{\gamma} \mid \boldsymbol{\Sigma}_{-kk}, \mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\tau}, \delta_s) \propto Gamma\Big(\frac{N}{2} + 1, \frac{s_{kk} + \delta_s}{2}\Big).$$

After simulating observations from the above conditional distributions, we can obtain $\boldsymbol{\sigma}_k = \boldsymbol{\beta}$, $\boldsymbol{\sigma}_k^T = \boldsymbol{\beta}^T$ and $\sigma_{kk} = \boldsymbol{\gamma} + \boldsymbol{\sigma}_k^T \Sigma_{-kk}^{-1} \boldsymbol{\sigma}_k$, then the k-th column and row of $\boldsymbol{\Sigma}$ are updated at a time. At the end, $\boldsymbol{\psi} = \boldsymbol{\Sigma}^{-1}$ is computed.

The conditional distribution for $\boldsymbol{\tau} = (\tau_{ij})_{i<j}$ can be expressed as:

$$p(\boldsymbol{\tau} \mid \mathbf{Y}, \boldsymbol{\Lambda}, \boldsymbol{\Omega}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}, \delta_s) \propto p(\boldsymbol{\Sigma} \mid \boldsymbol{\tau}, \delta_s) p(\boldsymbol{\tau} \mid \delta_s)$$

$$\propto \prod_{i<j} \tau_{ij}^{-\frac{1}{2}} \exp\Big\{-\frac{\sigma_{ij}^2 + \tau_{ij}^2 \delta_s^2}{2\tau_{ij}}\Big\}.$$

It can be shown that for $i < j$,

$$p\Big(\frac{1}{\tau_{ij}} \mid \mathbf{Y}, \boldsymbol{\Lambda}, \boldsymbol{\Omega}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}, \delta_s\Big) \propto Inv - Gaussian\Big(\sqrt{\frac{\delta_s^2}{\sigma_{ij}^2}}, \delta_s^2\Big).$$

The conditional distribution for δ_s can be expressed as:

$$p(\delta_s \mid \mathbf{Y}, \boldsymbol{\Lambda}, \boldsymbol{\Omega}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}, \boldsymbol{\tau}) \propto p(\boldsymbol{\tau} \mid \delta_s) p(\delta_s)$$

$$\propto Gamma\Big(a_s + \frac{J(J+1)}{2}, b_s + \frac{1}{2}\sum_{i=1}^{J}\sum_{j=1}^{J}|\sigma_{ij}|\Big) \quad .$$

(*Appendices continue*)

Appendix D

Additional Simulation Studies

Additional simulation studies were conducted under more heterogeneous factor loadings, as shown in Table D1. The model settings were the same as in Study 2, except that the data were generated by a different factor loading matrix with more heterogenous loading values. In general, the results indicated that the performance under the condition with more heterogeneous loading was similar to that presented in Study 2.

We found that if at least one specified loading per item was imposed, parameter recovery was satisfactory in the C-step. To better understand the consequences of imposing this identifiability constraint, additional simulation studies were conducted. The settings were the same as those in Study 1 except that the data were generated with diagonal $\boldsymbol{\Psi}$ (i.e., the residuals in the true model were assumed to be independent). The C-step was fitted with three specified loadings per factor (that is, λ_{11}, λ_{21} and λ_{31} are specified for F1 and λ_{82}, λ_{92}, $\lambda_{10,2}$ are specified for F2). As shown in Table D2, the recoveries of most parameters were unacceptable. Accordingly, it is difficult to relax the constraint. Moreover, the constraint is easy to understand and to follow since there is usually a target factor for each item in scale development.

The consequence of correlated residuals is the existence of local dependence, which needs to be accounted for during modeling. Treating the correlated residuals as independent can distort the loading estimates, especially those associated with the correlated residuals, as shown in the E-step of Study 1 and 2. Additional simulation studies were conducted to further understand the consequence of correlated residuals. The settings were the same as those of Study 1 and 2 except that the data were generated with diagonal $\boldsymbol{\Psi}$. The results of the E-step (Table D3 and D4) indicated that the estimates of all loading parameters were satisfactory. Accordingly, the biases of loading estimates in the E-step of Study 1 and 2 did come from the correlated residuals.

To investigate whether the model performs depending on the number of unspecified Lasso loadings or on the structure of the

Table D1
Simulation Studies With More Heterogeneous Loading Values

		E-step				C-step			
Par	True	BIAS	RMSE	*SE*	SIG%	BIAS	RMSE	*SE*	SIG%
λ_{11}	0.8	0.034	0.062	0.096	1.000	−0.016	0.048	0.071	1.000
λ_{21}	0.8	0.024	0.066	0.091	1.000	0.001	0.054	0.076	1.000
λ_{31}	0.8	0.071	0.094	0.092	1.000	0.036	0.067	0.076	1.000
λ_{41}	0.3	0.124	0.156	0.087	1.000	0.071	0.107	0.109	0.960
λ_{51}	0.5	0.008	0.077	0.089	1.000	0.022	0.069	0.090	1.000
λ_{61}	0.6	0.006	0.055	0.092	1.000	0.018	0.056	0.080	1.000
λ_{71}	0.3	−0.017	0.063	0.108	0.870	−0.060	0.084	0.087	0.890
$\lambda_{18,1}$	0.3	−0.004	0.046	0.132	0.620	−0.046	0.061	0.090	0.930
λ_{62}	0.3	−0.018	0.058	0.099	0.950	−0.068	0.087	0.089	0.900
λ_{72}	0.6	−0.019	0.058	0.091	0.990	0.017	0.056	0.089	1.000
λ_{82}	0.8	−0.002	0.069	0.097	0.990	−0.019	0.050	0.085	1.000
λ_{92}	0.3	0.191	0.211	0.091	0.990	0.107	0.129	0.130	0.980
$\lambda_{10,2}$	0.8	0.070	0.101	0.098	0.990	0.036	0.067	0.086	1.000
$\lambda_{11,2}$	0.5	−0.022	0.083	0.087	0.990	0.007	0.066	0.097	1.000
$\lambda_{12,2}$	0.6	−0.027	0.067	0.097	1.000	0.027	0.064	0.093	1.000
$\lambda_{13,2}$	0.3	−0.067	0.084	0.135	0.130	−0.058	0.071	0.090	0.900
λ_{13}	0.3	−0.086	0.100	0.137	0.070	−0.022	0.053	0.083	0.960
$\lambda_{12,3}$	0.3	0.017	0.057	0.116	0.950	−0.051	0.075	0.097	0.810
$\lambda_{13,3}$	0.8	0.019	0.062	0.117	0.990	0.011	0.055	0.079	1.000
$\lambda_{14,3}$	0.3	−0.024	0.092	0.088	0.870	0.004	0.070	0.099	0.940
$\lambda_{15,3}$	0.5	0.060	0.099	0.103	0.990	0.046	0.083	0.099	1.000
$\lambda_{16,3}$	0.6	0.051	0.094	0.107	0.990	0.042	0.084	0.093	1.000
$\lambda_{17,3}$	0.5	−0.032	0.084	0.097	0.990	0.018	0.063	0.097	1.000
$\lambda_{18,3}$	0.8	−0.024	0.065	0.120	0.990	0.003	0.051	0.080	1.000
λ_0	0	−0.024	0.072	0.110	0.012	−0.013	0.034	0.073	0.000
$\phi_{kk'}$	0.3	0.067	0.094	0.168	0.670	0.069	0.095	0.118	0.973
ψ_{jj}	0.53	−0.010	0.066	0.054	1.000	0.014	0.064	0.077	1.000
ψ_w	0.3	—	—	—	—	−0.056	0.074	0.082	0.903
ψ_b	0.3	—	—	—	—	−0.038	0.058	0.063	1.000
ψ_0	0	—	—	—	—	−0.001	0.028	0.046	0.003

(*Appendices continue*)

Table D2
C-step Results With Three Specified Loadings per Factor

		$N = 250$				$N = 500$			
Par	True	BIAS	RMSE	*SE*	SIG%	BIAS	RMSE	*SE*	SIG%
λ_{11}	0.7	−0.064	0.106	0.259	0.575	−0.052	0.121	0.239	0.650
λ_{21}	0.7	−0.064	0.107	0.260	0.600	−0.050	0.119	0.239	0.680
λ_{31}	0.7	−0.312	0.324	0.257	0.100	−0.287	0.310	0.243	0.290
λ_{41}	0.7	−0.429	0.438	0.227	0.008	−0.389	0.407	0.220	0.110
λ_{51}	0.5	−0.300	0.309	0.184	0.000	−0.275	0.290	0.176	0.080
λ_{61}	0.5	−0.330	0.336	0.183	0.000	−0.305	0.318	0.177	0.030
λ_{52}	0.5	−0.307	0.319	0.220	0.000	−0.268	0.300	0.207	0.120
λ_{62}	0.5	−0.245	0.263	0.264	0.000	−0.211	0.248	0.240	0.130
λ_{72}	0.7	−0.360	0.376	0.305	0.000	−0.329	0.365	0.283	0.100
λ_{82}	0.7	−0.376	0.387	0.306	0.008	−0.340	0.372	0.280	0.120
λ_{92}	0.7	−0.497	0.508	0.253	0.000	−0.449	0.480	0.240	0.080
$\lambda_{10,2}$	0.7	−0.496	0.507	0.254	0.000	−0.447	0.479	0.239	0.070
λ_0	0	0.009	0.046	0.157	0.000	0.007	0.051	0.155	0.000
ϕ_{12}	0.3	−0.119	0.161	0.411	0.000	−0.073	0.162	0.386	0.080
ψ_{55}	0.35	−0.023	0.092	0.161	0.283	−0.044	0.137	0.144	0.300
ψ_{66}	0.35	−0.025	0.084	0.165	0.275	−0.051	0.137	0.146	0.310
ψ_{jj}	0.51	−0.221	0.322	0.140	0.299	−0.237	0.338	0.126	0.351
ψ_0	0	0.326	0.338	0.165	0.316	0.306	0.328	0.147	0.364

Note. ψ_{ij} averaged across elements from $j = 1$ to 4 and 7 to 10; λ_0 averaged across all zero loading estimates; ψ_0 averaged across all zero residual covariance estimates.

loading matrix, additional simulation studies were conducted. The model settings were the same as Study 1 (10 items, two factors) except that the data were generated with diagonal $\boldsymbol{\Psi}$ (i.e., the residuals in the true model were assumed to be independent). In the E-step, the model was fitted with one specified loading per factor (that is, λ_{11} for Factor 1 and $\lambda_{10,2}$ for Factor 2) or no specified loadings (that is, all loadings were unspecified). The results, found in Tables D5 and D6, can be compared with those in Table D3. The estimates in the case of no specified loading (Table D6) were poor. In contrast, the case of one specified loading per factor (Table D5) was acceptable, but slightly worse than the case of two specified loadings per factor (Table D3). This is consistent with the findings in Study 1 (with correlated residuals).

Table D3
E-Step Results With True Diagonal ψ and Two Specified Loadings per Factor (10 Items, Two Factors)

		$N = 250$			$N = 500$		
Par	True	BIAS	RMSE	*SE*	BIAS	RMSE	*SE*
λ_{11}	0.7	0.021	0.078	0.134	0.004	0.068	0.118
λ_{21}	0.7	0.010	0.070	0.132	0.006	0.066	0.119
λ_{31}	0.7	−0.001	0.071	0.130	0.001	0.068	0.118
λ_{41}	0.7	0.003	0.073	0.131	−0.013	0.065	0.114
λ_{51}	0.5	−0.019	0.065	0.145	−0.017	0.051	0.130
λ_{61}	0.5	−0.025	0.071	0.144	−0.021	0.053	0.132
λ_{52}	0.5	−0.018	0.063	0.144	−0.025	0.053	0.130
λ_{62}	0.5	−0.019	0.060	0.143	−0.012	0.051	0.130
λ_{72}	0.7	0.004	0.058	0.132	0.005	0.066	0.118
λ_{82}	0.7	0.000	0.071	0.131	0.003	0.071	0.117
λ_{92}	0.7	0.015	0.069	0.133	0.009	0.067	0.118
$\lambda_{10,2}$	0.7	0.005	0.063	0.130	0.003	0.067	0.118
λ_0	0	−0.033	0.069	0.177	−0.027	0.066	0.168
ϕ_{12}	0.3	0.075	0.094	0.219	0.075	0.087	0.211
ψ_{55}	0.35	0.001	0.040	0.041	0.006	0.029	0.029
ψ_{66}	0.35	0.012	0.038	0.042	0.006	0.028	0.029
ψ_{jj}	0.51	0.007	0.058	0.059	0.006	0.042	0.041

Note. λ_0 averaged across all zero loading estimates.

(*Appendices continue*)

Table D4
E-Step Results With True Diagonal ψ and Two Specified Loadings per Factor (18 Items, Three Factors)

		$N = 250$			$N = 500$		
Par	True	BIAS	RMSE	*SE*	BIAS	RMSE	*SE*
λ_{11}	0.5	0.023	0.066	0.105	0.018	0.045	0.087
λ_{21}	0.7	0.023	0.065	0.094	0.024	0.062	0.077
λ_{31}	0.7	0.014	0.068	0.093	0.027	0.061	0.077
λ_{41}	0.7	0.007	0.057	0.092	0.013	0.052	0.076
λ_{51}	0.7	0.020	0.060	0.093	0.015	0.051	0.077
λ_{61}	0.5	−0.001	0.057	0.108	−0.009	0.045	0.092
λ_{71}	0.5	−0.005	0.056	0.107	−0.005	0.048	0.092
$\lambda_{18,1}$	0.5	0.000	0.062	0.105	0.013	0.042	0.086
λ_{62}	0.5	−0.007	0.055	0.109	−0.009	0.043	0.093
λ_{72}	0.5	−0.012	0.068	0.108	−0.010	0.043	0.093
λ_{82}	0.7	0.033	0.074	0.099	0.021	0.052	0.080
λ_{92}	0.7	0.014	0.065	0.097	0.016	0.049	0.080
$\lambda_{10,2}$	0.7	0.016	0.069	0.097	0.006	0.045	0.079
$\lambda_{11,2}$	0.7	0.009	0.057	0.097	0.011	0.048	0.079
$\lambda_{12,2}$	0.5	−0.020	0.061	0.108	−0.005	0.041	0.091
$\lambda_{13,2}$	0.5	−0.009	0.056	0.108	−0.003	0.047	0.091
λ_{13}	0.5	−0.016	0.064	0.110	−0.019	0.044	0.092
$\lambda_{12,3}$	0.5	−0.009	0.064	0.106	−0.012	0.047	0.090
$\lambda_{13,3}$	0.5	−0.010	0.061	0.106	−0.011	0.048	0.091
$\lambda_{14,3}$	0.7	0.006	0.063	0.095	0.003	0.062	0.080
$\lambda_{15,3}$	0.7	0.018	0.067	0.096	0.013	0.059	0.080
$\lambda_{16,3}$	0.7	0.022	0.068	0.097	0.014	0.068	0.080
$\lambda_{17,3}$	0.7	−0.003	0.059	0.094	0.004	0.054	0.080
$\lambda_{18,3}$	0.5	−0.016	0.063	0.108	−0.024	0.049	0.092
λ_0	0	−0.019	0.052	0.114	−0.018	0.045	0.101

Note. λ_0 averaged across all zero loading estimates.

Additional simulation studies were also conducted for a more complex loading structure in Study 2 (18 items, three factors). The model settings were the same as those of Study 2 except that the data were generated with diagonal $\boldsymbol{\Psi}$. In the E-step, the model was fitted with one specified loading per factor (that is, λ_{11} for Factor 1, λ_{82} for Factor 2, and $\lambda_{15,3}$ for Factor 3) or no specified loadings (that is, all loadings were unspecified). The results can be found in Table D7 and D8 and can be compared with those in Table D4.

Table D5
E-Step Results With One Specified Loading per Factor (10 Items, Two Factors)

		$N = 250$			$N = 500$		
Par	True	BIAS	RMSE	*SE*	BIAS	RMSE	*SE*
λ_{11}	0.7	−0.036	0.147	0.193	−0.023	0.114	0.149
λ_{21}	0.7	−0.057	0.143	0.187	−0.025	0.106	0.150
λ_{31}	0.7	−0.056	0.148	0.187	−0.022	0.115	0.148
λ_{41}	0.7	−0.054	0.141	0.190	−0.038	0.111	0.146
λ_{51}	0.5	−0.011	0.059	0.159	−0.017	0.053	0.145
λ_{61}	0.5	−0.018	0.069	0.159	−0.021	0.053	0.147
λ_{52}	0.5	−0.023	0.063	0.160	−0.025	0.053	0.146
λ_{62}	0.5	−0.023	0.066	0.158	−0.012	0.051	0.146
λ_{72}	0.7	−0.056	0.150	0.188	−0.017	0.099	0.145
λ_{82}	0.7	−0.061	0.155	0.187	−0.020	0.100	0.145
λ_{92}	0.7	−0.061	0.160	0.190	−0.020	0.096	0.144
$\lambda_{10,2}$	0.7	−0.058	0.151	0.188	−0.020	0.102	0.144
λ_0	0	0.029	0.142	0.226	−0.002	0.102	0.197

(*Appendices continue*)

Table D6
E-Step Results With All Loadings Being Unspecified (10 Items, Two Factors)

		$N = 250$			$N = 500$		
Par	True	BIAS	RMSE	*SE*	BIAS	RMSE	*SE*
λ_{11}	0.7	−0.385	0.474	0.289	−0.361	0.464	0.269
λ_{21}	0.7	−0.378	0.462	0.280	−0.352	0.460	0.269
λ_{31}	0.7	−0.380	0.464	0.281	−0.356	0.462	0.273
λ_{41}	0.7	−0.381	0.466	0.286	−0.356	0.454	0.264
λ_{51}	0.5	−0.017	0.063	0.173	−0.034	0.069	0.166
λ_{61}	0.5	−0.022	0.065	0.173	−0.030	0.061	0.168
λ_{52}	0.5	−0.018	0.063	0.173	−0.012	0.058	0.163
λ_{62}	0.5	−0.021	0.059	0.173	−0.007	0.057	0.165
λ_{72}	0.7	−0.384	0.470	0.290	−0.334	0.432	0.265
λ_{82}	0.7	−0.380	0.467	0.285	−0.335	0.433	0.263
λ_{92}	0.7	−0.384	0.469	0.290	−0.334	0.436	0.262
$\lambda_{10,2}$	0.7	−0.391	0.473	0.283	−0.337	0.433	0.260
λ_0	0	0.355	0.445	0.283	0.317	0.425	0.267

Note. λ_0 averaged across all zero loading estimates.

Similar to the above, the estimates in the case of no specified loading (Table D8) were poor. In contrast, the case of one specified loading per factor (Table D7) was acceptable, but slightly worse than the case of two specified loadings per factor (Table D4) in general.

In sum, the case of one specified loading per factor was acceptable in general and can serve as the minimum constraint, while two or more specified loadings per factor is preferred for more stable performance in the E-step.

The same simulated data sets were also used to investigate the same issue with the C-step. Table D2 above revealed that the C-step performed poorly up to three specified loadings per factor. In contrast, Table D9 confirmed that the constraint of one specified

Table D7
E-Step Results With One Specified Loading per Factor (18 Items, Three Factors)

		$N = 250$			$N = 500$		
Par	True	BIAS	RMSE	*SE*	BIAS	RMSE	*SE*
λ_{11}	0.5	0.015	0.086	0.133	0.012	0.077	0.102
λ_{21}	0.7	−0.072	0.219	0.130	−0.027	0.158	0.095
λ_{31}	0.7	−0.071	0.223	0.131	−0.018	0.159	0.097
λ_{41}	0.7	−0.081	0.224	0.132	−0.028	0.154	0.094
λ_{51}	0.7	−0.068	0.218	0.135	−0.030	0.162	0.097
λ_{61}	0.5	−0.049	0.153	0.127	−0.031	0.104	0.105
λ_{71}	0.5	−0.052	0.153	0.127	−0.028	0.111	0.105
$\lambda_{18,1}$	0.5	−0.009	0.087	0.131	0.008	0.071	0.100
λ_{62}	0.5	−0.019	0.082	0.125	−0.009	0.044	0.102
λ_{72}	0.5	−0.024	0.096	0.125	−0.011	0.048	0.103
λ_{82}	0.7	−0.018	0.180	0.126	0.002	0.120	0.092
λ_{92}	0.7	−0.047	0.175	0.122	−0.010	0.117	0.090
$\lambda_{10,2}$	0.7	−0.032	0.183	0.124	−0.015	0.122	0.091
$\lambda_{11,2}$	0.7	−0.040	0.176	0.123	−0.008	0.119	0.090
$\lambda_{12,2}$	0.5	−0.040	0.117	0.126	−0.017	0.082	0.106
$\lambda_{13,2}$	0.5	−0.031	0.115	0.126	−0.014	0.086	0.107
λ_{13}	0.5	−0.047	0.118	0.128	−0.029	0.074	0.103
$\lambda_{12,3}$	0.5	−0.057	0.151	0.128	−0.035	0.118	0.104
$\lambda_{13,3}$	0.5	−0.054	0.147	0.128	−0.034	0.113	0.105
$\lambda_{14,3}$	0.7	−0.080	0.230	0.134	−0.030	0.159	0.093
$\lambda_{15,3}$	0.7	−0.068	0.234	0.132	−0.028	0.173	0.095
$\lambda_{16,3}$	0.7	−0.081	0.236	0.132	−0.028	0.173	0.093
$\lambda_{17,3}$	0.7	−0.088	0.226	0.130	−0.033	0.171	0.093
$\lambda_{18,3}$	0.5	−0.044	0.115	0.125	−0.032	0.072	0.102
λ_0	0	0.020	0.142	0.138	0.000	0.101	0.115

(*Appendices continue*)

Table D8
E-Step Results With All Loadings Being Unspecified (18 Items, Three Factors)

		$N = 250$			$N = 500$		
Par	True	BIAS	RMSE	*SE*	BIAS	RMSE	*SE*
λ_{11}	0.5	−0.173	0.275	0.159	−0.169	0.293	0.116
λ_{21}	0.7	−0.452	0.557	0.165	−0.467	0.578	0.131
λ_{31}	0.7	−0.446	0.555	0.168	−0.462	0.575	0.132
λ_{41}	0.7	−0.461	0.565	0.167	−0.470	0.578	0.132
λ_{51}	0.7	−0.453	0.561	0.170	−0.474	0.579	0.131
λ_{61}	0.5	−0.159	0.276	0.161	−0.191	0.304	0.121
λ_{71}	0.5	−0.179	0.290	0.159	−0.191	0.306	0.121
$\lambda_{18,1}$	0.5	−0.172	0.273	0.158	−0.167	0.288	0.115
λ_{62}	0.5	−0.191	0.295	0.160	−0.178	0.301	0.123
λ_{72}	0.5	−0.179	0.287	0.158	−0.175	0.301	0.123
λ_{82}	0.7	−0.451	0.559	0.178	−0.484	0.589	0.131
λ_{92}	0.7	−0.449	0.543	0.176	−0.476	0.579	0.130
$\lambda_{10,2}$	0.7	−0.441	0.553	0.177	−0.481	0.584	0.129
$\lambda_{11,2}$	0.7	−0.439	0.546	0.176	−0.479	0.581	0.131
$\lambda_{12,2}$	0.5	−0.164	0.270	0.153	−0.178	0.297	0.121
$\lambda_{13,2}$	0.5	−0.161	0.268	0.152	−0.177	0.297	0.123
λ_{13}	0.5	−0.164	0.277	0.158	−0.195	0.313	0.118
$\lambda_{12,3}$	0.5	−0.196	0.300	0.154	−0.174	0.291	0.121
$\lambda_{13,3}$	0.5	−0.188	0.295	0.155	−0.172	0.294	0.122
$\lambda_{14,3}$	0.7	−0.463	0.567	0.172	−0.490	0.586	0.134
$\lambda_{15,3}$	0.7	−0.475	0.568	0.175	−0.487	0.586	0.133
$\lambda_{16,3}$	0.7	−0.452	0.552	0.177	−0.491	0.588	0.134
$\lambda_{17,3}$	0.7	−0.467	0.561	0.175	−0.488	0.588	0.132
$\lambda_{18,3}$	0.5	−0.179	0.288	0.156	−0.201	0.320	0.116
λ_0	0	0.230	0.374	0.174	0.245	0.405	0.127

Note. λ_0 averaged across all zero loading estimates.

loading per item was still appropriate even when the residuals in the true model were assumed to be independent.

Additional simulation studies were conducted to compare the performance of the E-step with the BSEM approach using the ridge regression priors proposed by Muthén and Asparouhov (2012) in estimating the loadings. The BSEM approach relies on normal priors concentrated around zero with small variance. Mplus was used to analyze the same simulated data sets with diagonal $\boldsymbol{\Psi}$. The results of loading estimates are presented in Tables D10 and D11. Compared with Tables D3 and D4, the E-step with the

Table D9
C-Step Results With One Specified Loading per Item (10 Items, Two Factors)

		$N = 250$			$N = 500$		
Par	True	BIAS	RMSE	*SE*	BIAS	RMSE	*SE*
λ_{11}	0.7	−0.015	0.066	0.119	−0.020	0.047	0.096
λ_{21}	0.7	−0.021	0.063	0.117	−0.018	0.046	0.095
λ_{31}	0.7	−0.022	0.065	0.118	−0.019	0.052	0.096
λ_{41}	0.7	−0.018	0.067	0.117	−0.028	0.046	0.095
λ_{51}	0.5	0.016	0.059	0.120	0.004	0.040	0.101
λ_{61}	0.5	−0.089	0.110	0.137	−0.067	0.079	0.116
λ_{52}	0.5	−0.083	0.102	0.137	−0.071	0.080	0.116
λ_{62}	0.5	0.017	0.059	0.121	0.010	0.040	0.102
λ_{72}	0.7	−0.022	0.055	0.118	−0.014	0.048	0.098
λ_{82}	0.7	−0.021	0.064	0.119	−0.016	0.048	0.097
λ_{92}	0.7	−0.020	0.062	0.119	−0.016	0.044	0.097
$\lambda_{10,2}$	0.7	−0.029	0.063	0.116	−0.020	0.050	0.096
λ_0	0	−0.012	0.034	0.105	−0.013	0.027	0.094

Note. λ_0 averaged across all zero loading estimates.

(*Appendices continue*)

Table D10
Loading Estimates for BSEM With the Ridge Regression Priors (10 Items, Two Factors)

		$N = 250$				$N = 500$			
Par	True	BIAS	RMSE	*SE*	SIG%	BIAS	RMSE	*SE*	SIG%
λ_{11}	0.7	−0.199	0.220	0.111	1.000	0.016	0.055	0.057	1.000
λ_{21}	0.7	−0.193	0.211	0.112	1.000	0.019	0.052	0.057	1.000
λ_{31}	0.7	−0.459	0.466	0.076	0.990	−0.119	0.130	0.047	1.000
λ_{41}	0.7	−0.456	0.464	0.076	1.000	−0.129	0.137	0.046	1.000
λ_{51}	0.5	−0.156	0.158	0.073	1.000	−0.1	0.105	0.050	1.000
λ_{61}	0.5	−0.161	0.162	0.072	1.000	−0.104	0.107	0.050	1.000
λ_{52}	0.5	−0.158	0.160	0.073	1.000	−0.113	0.116	0.049	1.000
λ_{62}	0.5	−0.160	0.162	0.073	1.000	−0.103	0.107	0.050	1.000
λ_{72}	0.7	−0.460	0.468	0.076	1.000	−0.116	0.126	0.047	1.000
λ_{82}	0.7	−0.460	0.467	0.076	1.000	−0.12	0.130	0.046	1.000
λ_{92}	0.7	−0.192	0.211	0.113	1.000	0.02	0.052	0.057	1.000
$\lambda_{10,2}$	0.7	−0.201	0.215	0.111	1.000	0.013	0.052	0.057	1.000
λ_0	0	0.090	0.125	0.087	0.446	−0.087	0.100	0.060	0.334

Bayesian Lasso was better in terms of BIAS and RMSE. More importantly, the Type I error rates for the BSEM approach were high, making it likely to misidentify truly zero loadings as significant.

To better understand whether the bias in the E-step is a function of the residual correlations, additional simulation studies were conducted. The settings were similar to those of Study 1, except that the data were generated with two additional level of residual correlations (0.3 and 0.7). Comparing the results (Tables D12 and D13) with the case of no correlated residual (Table D3), it seems that the bias was a function of the residual correlations to some extent, and the bias became larger when the residual correlation increased.

Table D11
Loading Estimates for BSEM With the Ridge Regression Priors (18 Items, Three Factors)

		$N = 250$				$N = 500$			
Par	True	BIAS	RMSE	*SE*	SIG%	BIAS	RMSE	*SE*	SIG%
λ_{11}	0.5	0.156	0.187	0.120	1.000	−0.008	0.044	0.061	1.000
λ_{21}	0.7	−0.133	0.188	0.117	1.000	0.006	0.048	0.056	1.000
λ_{31}	0.7	−0.412	0.439	0.076	0.729	−0.105	0.111	0.046	1.000
λ_{41}	0.7	−0.415	0.440	0.076	0.708	−0.117	0.122	0.046	1.000
λ_{51}	0.7	−0.406	0.434	0.076	0.708	−0.118	0.123	0.046	1.000
λ_{61}	0.5	−0.260	0.270	0.075	0.958	−0.110	0.114	0.050	1.000
λ_{71}	0.5	−0.266	0.275	0.075	0.917	−0.106	0.112	0.051	1.000
$\lambda_{18,1}$	0.5	−0.181	0.200	0.076	1.000	−0.065	0.073	0.054	1.000
λ_{62}	0.5	−0.202	0.217	0.067	1.000	−0.070	0.076	0.050	1.000
λ_{72}	0.5	−0.206	0.224	0.067	0.979	−0.073	0.079	0.051	1.000
λ_{82}	0.7	−0.056	0.156	0.103	1.000	0.045	0.063	0.058	1.000
λ_{92}	0.7	−0.064	0.145	0.102	1.000	0.038	0.061	0.058	1.000
$\lambda_{10,2}$	0.7	−0.314	0.356	0.072	0.708	−0.098	0.103	0.047	1.000
$\lambda_{11,2}$	0.7	−0.269	0.291	0.057	1.000	−0.122	0.124	0.038	1.000
$\lambda_{12,2}$	0.5	−0.069	0.131	0.057	1.000	0.078	0.082	0.038	1.000
$\lambda_{13,2}$	0.5	−0.160	0.187	0.067	1.000	−0.010	0.040	0.047	1.000
λ_{13}	0.5	−0.393	0.415	0.096	0.292	−0.137	0.140	0.055	1.000
$\lambda_{12,3}$	0.5	−0.312	0.323	0.073	0.604	−0.182	0.184	0.048	1.000
$\lambda_{13,3}$	0.5	−0.292	0.305	0.075	0.667	−0.158	0.161	0.048	1.000
$\lambda_{14,3}$	0.7	−0.432	0.456	0.074	0.667	−0.148	0.153	0.045	1.000
$\lambda_{15,3}$	0.7	−0.184	0.215	0.116	1.000	−0.041	0.057	0.053	1.000
$\lambda_{16,3}$	0.7	−0.174	0.207	0.116	1.000	−0.041	0.071	0.053	1.000
$\lambda_{17,3}$	0.7	−0.434	0.456	0.074	0.667	−0.144	0.148	0.044	1.000
$\lambda_{18,3}$	0.5	−0.197	0.211	0.076	1.000	−0.131	0.135	0.053	1.000
λ_0	0	0.026	0.105	0.079	0.181	−0.053	0.073	0.057	0.188

Note. λ_0 averaged across all zero loading estimates.

(*Appendices continue*)

Table D12
E-Step Results With Two Specified Loadings per Factor (Residual Correlation = .3)

		$N = 250$			$N = 500$		
Par	True	BIAS	RMSE	*SE*	BIAS	RMSE	*SE*
λ_{11}	0.7	0.102	0.123	0.122	0.110	0.121	0.107
λ_{21}	0.7	0.113	0.137	0.123	0.107	0.116	0.107
λ_{31}	0.7	−0.035	0.069	0.110	−0.028	0.062	0.093
λ_{41}	0.7	−0.066	0.095	0.109	−0.054	0.073	0.091
λ_{51}	0.5	−0.043	0.072	0.139	−0.050	0.068	0.125
λ_{61}	0.5	−0.053	0.081	0.140	−0.052	0.069	0.124
λ_{52}	0.5	−0.007	0.059	0.135	0.003	0.047	0.116
λ_{62}	0.5	0.002	0.059	0.133	0.001	0.046	0.116
λ_{72}	0.7	−0.005	0.068	0.112	0.011	0.047	0.091
λ_{82}	0.7	0.005	0.066	0.114	0.017	0.049	0.091
λ_{92}	0.7	0.026	0.079	0.116	0.026	0.056	0.093
$\lambda_{10,2}$	0.7	0.028	0.071	0.117	0.026	0.060	0.093
λ_0	0	−0.039	0.078	0.163	−0.039	0.070	0.150
ϕ_{12}	0.3	0.083	0.101	0.209	0.090	0.100	0.197
ψ_{55}	0.35	0.017	0.045	0.042	0.016	0.034	0.029
ψ_{66}	0.35	0.014	0.044	0.042	0.016	0.034	0.029
ψ_{jj}	0.51	−0.007	0.080	0.057	−0.012	0.065	0.040

To better understand whether the bias in the E-step is a function of the sparsity of the residual covariance matrix, additional simulation studies were conducted. The settings were similar to those of Study 1 except that the data were generated with four residual correlations each with a magnitude of 0.3. Comparing the results (Table D14) with the case of two residual correlations each with the same magnitude (Table D12), it seems the difference was not significant. More work is needed to systematically investigate this issue.

To better understand whether the bias in the E-step diminishes when the cross-loadings are close to zero, additional simulation studies were conducted. The settings were similar to those of Study 1 except that the data were generated with residual correlations of 0.3 and no cross-loadings. Comparing the results in Table D15 with those in Table D12 above, the bias seemed to diminish a little bit but still existed when the cross-loadings were closed to zero.

To better understand how the Bayesian Lasso works when the aim is to estimate the residual covariance matrix, additional simulation studies were conducted. The settings were similar as those of Study 1 and 2 except that the loading structure was assumed to be known in the C-step (i.e., all the nonzero load-

Table D13
E-Step Results With Two Specified Loadings per Factor (Residual Correlation = .7)

		$N = 250$			$N = 500$		
Par	True	BIAS	RMSE	*SE*	BIAS	RMSE	*SE*
λ_{11}	0.7	0.249	0.255	0.114	0.244	0.252	0.108
λ_{21}	0.7	0.247	0.253	0.113	0.245	0.254	0.108
λ_{31}	0.7	−0.196	0.207	0.091	−0.193	0.201	0.077
λ_{41}	0.7	−0.233	0.248	0.095	−0.224	0.231	0.083
λ_{51}	0.5	−0.195	0.204	0.131	−0.193	0.198	0.124
λ_{61}	0.5	−0.187	0.195	0.130	−0.185	0.192	0.124
λ_{52}	0.5	0.100	0.112	0.098	0.107	0.116	0.087
λ_{62}	0.5	0.098	0.114	0.099	0.105	0.114	0.088
λ_{72}	0.7	0.019	0.070	0.089	0.033	0.069	0.079
λ_{82}	0.7	−0.008	0.062	0.088	0.000	0.057	0.078
λ_{92}	0.7	0.002	0.061	0.089	−0.002	0.058	0.078
$\lambda_{10,2}$	0.7	0.014	0.069	0.090	0.006	0.056	0.079
λ_0	0	−0.003	0.113	0.136	−0.001	0.117	0.131
ϕ_{12}	0.3	0.058	0.081	0.187	0.062	0.079	0.184
ψ_{55}	0.35	0.056	0.070	0.045	0.043	0.054	0.031
ψ_{66}	0.35	0.052	0.069	0.045	0.045	0.054	0.031
ψ_{jj}	0.51	−0.036	0.161	0.053	−0.042	0.149	0.037

Note. ψ_{jj} averaged across elements from $j = 1$ to 4 and 7 to 10; λ_0 averaged across all zero loading estimates.

(*Appendices continue*)

Table D14
E-Step With Bayesian Lasso in Estimating the Loadings (Four Residual Correlations Each With a Magnitude of 0.3)

		$N = 250$			$N = 500$		
Par	True	BIAS	RMSE	*SE*	BIAS	RMSE	*SE*
λ_{11}	0.7	0.095	0.121	0.125	0.094	0.109	0.109
λ_{21}	0.7	0.090	0.115	0.125	0.083	0.102	0.109
λ_{31}	0.7	−0.024	0.079	0.115	−0.020	0.053	0.098
λ_{41}	0.7	−0.051	0.077	0.113	−0.048	0.074	0.096
λ_{51}	0.5	−0.066	0.091	0.137	−0.075	0.090	0.125
λ_{61}	0.5	−0.055	0.080	0.135	−0.071	0.084	0.125
λ_{52}	0.5	−0.003	0.061	0.129	0.008	0.050	0.114
λ_{62}	0.5	−0.015	0.068	0.129	0.008	0.048	0.114
λ_{72}	0.7	0.005	0.066	0.114	0.015	0.057	0.099
λ_{82}	0.7	0.035	0.082	0.120	0.048	0.073	0.103
λ_{92}	0.7	0.002	0.072	0.114	−0.006	0.054	0.096
$\lambda_{10,2}$	0.7	−0.010	0.071	0.113	−0.008	0.060	0.096
λ_0	0	−0.028	0.069	0.163	−0.030	0.063	0.151
ϕ_{12}	0.3	0.107	0.121	0.205	0.104	0.114	0.195
ψ_{55}	0.35	0.031	0.053	0.043	0.028	0.041	0.030
ψ_{66}	0.35	0.034	0.056	0.043	0.033	0.045	0.030
ψ_{jj}	0.51	−0.008	0.076	0.058	−0.010	0.065	0.041

Note. ψ_{jj} averaged across elements from $j = 1$ to 4 and 7 to 10; λ_0 averaged across all zero loading.

ings including the cross-loadings were assumed to be specified and all the zero loadings were fixed to zero). The results in Table D16 showed that, by assuming the loading structure as known, the power of detecting the within-factor nonzero residual covariance was still low, but higher than that with an unknown structure in Table 4 (0.020 vs. 0.005 in sample size 250; 0.200 vs. 0.040 in sample size 500). The power of detecting the between-factor nonzero residual covariance was satisfactory and higher than that in Table 4 (0.830 vs. 0.755 in sample size 250; 0.980 vs. 0.915 in sample size 500). It is interesting to observe that, similar to Table 4, the estimates of λ_{11} and λ_{21} were positively biased and the estimate of ψ_{21} was negatively biased. Moreover, the estimate of the between-factor residual covariance ψ_{74} was also more accurate and stable than that of the within-factor residual covariance ψ_{21}. Accordingly, the cases of known and unknown loading structures in the C-step exhibited similar patterns, and the power of detecting the nonzero off-diagonal elements of the covariance matrix was not only related to the Bayesian Lasso estimation, but also related to the position of the nonzero elements (that is, within factor or

Table D15
E-Step With Bayesian Lasso Priors in Estimating the Loadings (No Cross-Loading)

		$N = 250$			$N = 500$		
Par	True	BIAS	RMSE	*SE*	BIAS	RMSE	*SE*
λ_{11}	0.7	0.092	0.110	0.076	0.097	0.106	0.055
λ_{21}	0.7	0.091	0.107	0.076	0.098	0.108	0.055
λ_{31}	0.7	−0.068	0.094	0.075	−0.045	0.067	0.054
λ_{41}	0.7	−0.067	0.094	0.077	−0.065	0.078	0.055
λ_{51}	0.5	−0.042	0.089	0.075	−0.041	0.058	0.053
λ_{62}	0.5	−0.023	0.057	0.075	−0.004	0.045	0.054
λ_{72}	0.7	−0.007	0.070	0.074	−0.005	0.052	0.054
λ_{82}	0.7	−0.019	0.066	0.074	−0.003	0.050	0.053
λ_{92}	0.7	−0.003	0.067	0.074	0.008	0.041	0.054
$\lambda_{10,2}$	0.7	0.013	0.063	0.075	0.003	0.043	0.054
λ_0	0	0.011	0.058	0.104	0.008	0.042	0.090
ϕ_{12}	0.3	−0.023	0.073	0.167	−0.022	0.059	0.152
ψ_{55}	0.75	0.017	0.076	0.074	0.029	0.066	0.053
ψ_{66}	0.75	0.010	0.076	0.074	0.010	0.051	0.052
ψ_{jj}	0.51	−0.011	0.085	0.060	−0.014	0.073	0.042

Note. ψ_{jj} averaged across elements from $j = 1$ to 4 and 7 to 10; λ_0 averaged across all zero loading estimates.

(*Appendices continue*)

Table D16
C-Step With Known Loading Structure (10 Items, Two Factors)

		$N = 250$				$N = 500$			
Par	True	BIAS	RMSE	*SE*	SIG%	BIAS	RMSE	*SE*	SIG%
λ_{11}	0.7	0.096	0.120	0.097	1.000	0.086	0.096	0.084	1.000
λ_{21}	0.7	0.098	0.121	0.098	1.000	0.089	0.100	0.084	1.000
λ_{31}	0.7	−0.042	0.072	0.097	1.000	−0.029	0.055	0.080	1.000
λ_{41}	0.7	−0.085	0.102	0.083	1.000	−0.075	0.086	0.067	1.000
λ_{51}	0.5	−0.026	0.065	0.088	1.000	−0.036	0.055	0.069	1.000
λ_{61}	0.5	−0.035	0.064	0.087	1.000	−0.033	0.051	0.068	1.000
λ_{52}	0.5	−0.010	0.048	0.087	1.000	−0.008	0.035	0.068	1.000
λ_{62}	0.5	−0.006	0.056	0.087	1.000	−0.011	0.039	0.068	1.000
λ_{72}	0.7	−0.072	0.092	0.083	1.000	−0.055	0.069	0.065	1.000
λ_{82}	0.7	−0.002	0.063	0.090	1.000	0.005	0.041	0.071	1.000
λ_{92}	0.7	−0.003	0.064	0.090	1.000	−0.003	0.042	0.071	1.000
$\lambda_{10,2}$	0.7	0.010	0.063	0.090	1.000	0.005	0.042	0.071	1.000
ϕ_{12}	0.3	−0.029	0.084	0.101	0.830	−0.027	0.064	0.075	0.980
ψ_{21}	0.3	−0.136	0.144	0.117	0.020	−0.126	0.131	0.110	0.200
ψ_{74}	0.3	0.007	0.049	0.069	1.000	0.007	0.037	0.052	1.000
ψ_{55}	0.35	0.085	0.096	0.093	1.000	0.074	0.080	0.081	1.000
ψ_{66}	0.35	0.079	0.089	0.092	1.000	0.070	0.077	0.079	1.000
ψ_{jj}	0.51	0.011	0.097	0.109	1.000	0.003	0.078	0.094	1.000
ψ_0	0	0.022	0.041	0.060	0.003	0.019	0.035	0.050	0.003

Note. ψ_{jj} averaged across elements from $j = 1$ to 4 and 7 to 10; ψ_0 averaged across all zero residual covariance estimates.

between factor). However, more work might be needed to systematically investigate this issue.

To study how the value of the correlation of the residuals of variables that load to the same factor influence the estimation of the true covariance structure using the Bayesian Lasso, additional simulation studies were conducted. The settings were similar to those of Study 1 except that the data were generated by $\mathbf{\Psi}$ with nonzero within-factor residual covariance ψ_{12} and all other off-diagonal elements equal to 0. Two levels of residual correlations (0.3 and 0.7) were considered, which led to $\psi_{12} = \psi_{21} = 0.152$ and $\psi_{12} = \psi_{21} = 0.357$, respectively. For ease of comparison, we only focused on the estimates of the elements in the residual covariance matrix. The results presented in Tables D17 and D18 indicate that the power of detecting the nonzero ψ_{12} was low whether the residual correlation was low (0.3) or high (0.7) and could be improved with a larger sample size. However, the accuracy and stability of estimates, in terms of BIAS and SE, became poorer when the correlation increased.

To study how the value of the correlation of the residuals of variables that load to different factors influences the estimation of the true covariance structure using the Bayesian Lasso, additional simulation studies were conducted. The settings were similar to those of Study 1 except that the data were generated by $\mathbf{\Psi}$ with nonzero between-factor residual covariance ψ_{74} and with all other off-diagonal elements equal to zero. Two levels of residual correlations (0.3 and 0.7) were considered, which led to $\psi_{74} = \psi_{47} = 0.152$ and $\psi_{74} = \psi_{47} = 0.357$, respectively. The results presented in Tables D19 and D20 indicate that the power of detecting the nonzero ψ_{74} was low when the residual correlation was low (0.3), but became satisfactory (>0.800) when the residual correlation was high (0.7). Moreover, the power improved with a larger sample size. Similar to the above, the accuracy and stability of estimates, in terms of BIAS and SE, became poorer when the correlation increased. However, more work might be needed to systematically investigate this issue.

Table D17
C-Step With the Residuals of Variables That Load to the Same Factor (Residual Correlation = .3)

		$N = 250$				$N = 500$			
Par	True	BIAS	RMSE	*SE*	SIG%	BIAS	RMSE	*SE*	SIG%
ψ_{21}	0.153	−0.023	0.038	0.105	0.000	−0.029	0.041	0.094	0.010
ψ_{55}	0.35	0.055	0.066	0.088	1.000	0.052	0.060	0.071	1.000
ψ_{66}	0.35	0.058	0.068	0.087	1.000	0.044	0.051	0.069	1.000
ψ_{jj}	0.51	0.042	0.072	0.114	1.000	0.030	0.056	0.096	1.000
ψ_0	0	0.015	0.035	0.067	0.000	0.012	0.029	0.056	0.000

(*Appendices continue*)

Table D18
C-Step With the Residuals of Variables That Load to the Same Factor (Residual Correlation = .7)

		$N = 250$				$N = 500$			
Par	True	BIAS	RMSE	*SE*	SIG%	BIAS	RMSE	*SE*	SIG%
ψ_{21}	0.357	−0.155	0.162	0.165	0.010	−0.133	0.141	0.161	0.070
ψ_{55}	0.35	0.102	0.112	0.120	1.000	0.083	0.091	0.100	1.000
ψ_{66}	0.35	0.092	0.103	0.114	1.000	0.076	0.082	0.097	1.000
ψ_{jj}	0.51	0.025	0.109	0.138	1.000	0.020	0.093	0.124	1.000
ψ_0	0	0.026	0.045	0.082	0.000	0.021	0.039	0.072	0.000

Note. ψ_{jj} averaged across elements from $j = 1$ to 4 and 7 to 10; λ_0 averaged across all zero loading estimates; ψ_0 averaged across all zero residual covariance estimates.

Table D19
C-Step With the Residuals of Variables That Load to Different Factors (Residual Correlation = .3)

		$N = 250$				$N = 500$			
Par	True	BIAS	RMSE	*SE*	SIG%	BIAS	RMSE	*SE*	SIG%
ψ_{74}	0.153	−0.070	0.077	0.070	0.020	−0.060	0.066	0.061	0.110
ψ_{55}	0.35	0.053	0.064	0.086	1.000	0.048	0.055	0.070	1.000
ψ_{66}	0.35	0.056	0.069	0.087	1.000	0.043	0.051	0.069	1.000
ψ_{jj}	0.51	0.049	0.070	0.110	1.000	0.036	0.051	0.093	1.000
ψ_0	0	0.015	0.037	0.067	0.000	0.012	0.030	0.057	0.000

Table D20
C-Step With the Residuals of Variables That Load to Different Factors (Residual Correlation = .7)

		$N = 250$				$N = 500$			
Par	True	BIAS	RMSE	*SE*	SIG%	BIAS	RMSE	*SE*	SIG%
ψ_{74}	0.357	−0.102	0.111	0.103	0.890	−0.088	0.092	0.091	0.980
ψ_{55}	0.35	0.078	0.089	0.100	1.000	0.062	0.069	0.078	1.000
ψ_{66}	0.35	0.081	0.090	0.099	1.000	0.061	0.067	0.077	1.000
ψ_{jj}	0.51	0.041	0.066	0.120	1.000	0.026	0.046	0.100	1.000
ψ_0	0	0.019	0.040	0.076	0.000	0.014	0.031	0.062	0.000

Note. ψ_{jj} averaged across elements from $j = 1$ to 4 and 7 to 10; λ_0 averaged across all zero loading estimates; ψ_0 averaged across all zero residual covariance estimates.

(*Appendices continue*)

Appendix E

Humor Styles Questionnaire

Item	Content	Factor
1	I usually don't laugh or joke around much with other people	F1
2	If I am feeling depressed, I can usually cheer myself up with humor	F2
3	If someone makes a mistake, I will often tease them about it	F3
4	I let people laugh at me or make fun at my expense more than I should	F4
5	I don't have to work very hard at making other people laugh—I seem to be a naturally humorous person	F1
6	Even when I'm by myself, I'm often amused by the absurdities of life	F2
7	People are never offended or hurt by my sense of humor	F3
8	I will often get carried away in putting myself down if it makes my family or friends laugh	F4
9	I rarely make other people laugh by telling funny stories about myself	F1
10	If I am feeling upset or unhappy I usually try to think of something funny about the situation to make myself feel better	F2
11	When telling jokes or saying funny things, I am usually not very concerned about how other people are taking it	F3
12	I often try to make people like or accept me more by saying something funny about my own weaknesses, blunders, or faults	F4
13	I laugh and joke a lot with my closest friends	F1
14	My humorous outlook on life keeps me from getting overly upset or depressed about things	F2
15	I do not like it when people use humor as a way of criticizing or putting someone down	F3
16	I don't often say funny things to put myself down	F4
17	I usually don't like to tell jokes or amuse people	F1
18	If I'm by myself and I'm feeling unhappy, I make an effort to think of something funny to cheer myself up	F2
19	Sometimes I think of something that is so funny that I can't stop myself from saying it, even if it is not appropriate for the situation	F3
20	I often go overboard in putting myself down when I am making jokes or trying to be funny	F4
21	I enjoy making people laugh	F1
22	If I am feeling sad or upset, I usually lose my sense of humor	F2
23	I never participate in laughing at others even if all my friends are doing it	F3
24	When I am with friends or family, I often seem to be the one that other people make fun of or joke about	F4
25	I don't often joke around with my friends	F1
26	It is my experience that thinking about some amusing aspect of a situation is often a very effective way of coping with problems	F2
27	If I don't like someone, I often use humor or teasing to put them down	F3
28	If I am having problems or feeling unhappy, I often cover it up by joking around, so that even my closest friends don't know how I really feel	F4
29	I usually can't think of witty things to say when I'm with other people	F1
30	I don't need to be with other people to feel amused—I can usually find things to laugh about even when I'm by myself	F2
31	Even if something is really funny to me, I will not laugh or joke about it if someone will be offended	F3
32	Letting others laugh at me is my way of keeping my friends and family in good spirits	F4

Note. F1 = affiliative humor; F2 = self-enhancing humor; F3 = aggressive humor; F3 = self-defeating humor.

Received April 9, 2019
Revision received April 17, 2020
Accepted April 24, 2020 ■