Published in final edited form as:

Multivariate Behav Res. 2016 ; 51(4): 519–539. doi:10.1080/00273171.2016.1168279.

Bayesian Factor Analysis as a Variable Selection Problem: Alternative Priors and Consequences

Zhao-Hua Lu, Sy-Miin Chow, and Eric Loken

Department of Human Development and Family Studies, Pennsylvania State University,
University Park, Pennsylvania

Abstract

Factor analysis is a popular statistical technique for multivariate data analysis. Developments in the structural equation modeling framework have enabled the use of hybrid confirmatory/exploratory approaches in which factor loading structures can be explored relatively flexibly within a confirmatory factor analysis (CFA) framework. Recently, a Bayesian structural equation modeling (BSEM) approach (Muthén & Asparouhov, 2012) has been proposed as a way to explore the presence of cross-loadings in CFA models. We show that the issue of determining factor loading patterns may be formulated as a Bayesian variable selection problem in which Muthén and Asparouhov's approach can be regarded as a BSEM approach with ridge regression prior (BSEM-RP). We propose another Bayesian approach, denoted herein as the Bayesian structural equation modeling with spike and slab prior (BSEM-SSP), which serves as a one-stage alternative to the BSEM-RP. We review the theoretical advantages and disadvantages of both approaches and compare their empirical performance relative to two modification indices-based approaches and exploratory factor analysis with target rotation. A teacher stress scale data set (Byrne, 2012; Pettegrew & Wolf, 1982) is used to demonstrate our approach.

Keywords

Factor analysis; Bayesian structural equation modeling; variable selection; Markov chain Monte Carlo algorithms

Factor analysis is one popular statistical technique for multivariate data analysis. It is a dimension reduction procedure which extracts information from multivariate observed indicators to lower-dimensional latent factors through

$$y_i = \mu + \Lambda \omega_i + \varepsilon_i, \quad (1)$$

Correspondence concerning this article can be addressed to Zhao-Hua Lu, 413A Biobehavioral Health Building, the Pennsylvania State University, University Park, PA 16802, or by to zhaohua.lu@gmail.com.
Zhao-Hua Lu, Department of Human Development and Family Studies, Pennsylvania State University, University Park, Pennsylvania;
Sy-Miin Chow, Department of Human Development and Family Studies, Pennsylvania State University, University Park, Pennsylvania;
Eric Loken, Department of Human Development and Family Studies, Pennsylvania State University, University Park, Pennsylvania.

where $i = 1, \dots, N$, N is the sample size, $\mathbf{y}_i$ is a $p \times 1$ vector of observed indicators, $\boldsymbol{\mu}$ is a $p \times 1$ vector of intercepts, $\boldsymbol{\Lambda}$ is a $p \times q$ loading matrix, $\boldsymbol{\omega}_i$ is a $q \times 1$ vector of latent factors, and $\boldsymbol{\varepsilon}_i$ is a $p \times 1$ vector of measurement errors. It is assumed that $\boldsymbol{\omega}_i \sim N_q(\mathbf{0}, \boldsymbol{\Phi})$, $\boldsymbol{\varepsilon}_i \sim N_p(\mathbf{0}, \boldsymbol{\Psi})$, and $\boldsymbol{\omega}_i$ and $\boldsymbol{\varepsilon}_i$ are independent. $\boldsymbol{\Lambda}$ reflects the relations between observed indicators and latent factors. It is well-known that the model in eq. (1) is not identified without further constraints. Constraints can be applied to elements in $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$. In this paper, we adopt the common assumptions that $\boldsymbol{\Phi}$ is a positive definite matrix, and $\boldsymbol{\Psi}$ is a diagonal matrix.

Two main variants of factor analysis are exploratory (EFA, Jennrich & Sampson, 1966) and confirmatory (CFA, Jöreskog, 1969) factor analysis, each adopting different assumptions and approaches to estimating $\boldsymbol{\Lambda}$. EFA and CFA have certain advantages and disadvantages. EFA first generates preliminary estimates by optimizing a target function with respect to the parameters using a minimally identified version of model in eq. (1). Then, rotation is applied to the preliminary estimation to yield a parsimonious $\boldsymbol{\Lambda}$ that optimizes a specific simplicity function. With fewer constraints, EFA is more flexible. In comparison, CFA starts with stronger, theoretically informed assumptions, with substantially more elements in $\boldsymbol{\Lambda}$ fixed to zero than in EFA. With fewer free parameters to estimate, the standard errors may be smaller in CFA, leading to more accurate estimates and more powerful hypothesis tests. However, prior knowledge may not be available, and specifying oversimplified factor loading structures may lead to biased estimates of the parameters in $\boldsymbol{\Phi}$ (Asparouhov & Muthén, 2009). In addition, the factors in CFA are typically allowed to be correlated (via $\boldsymbol{\Phi}$), whereas the factors in EFA are initially assumed to be orthogonal before rotation (Browne, 2001).

One possible “hybrid” approach combining the parsimony of the CFA while retaining some flexibility of the EFA is to utilize modification indices (MI, Sörbom, 1989) in the context of a CFA to identify additional elements in $\boldsymbol{\Lambda}$ that may be freed to improve model fit. However, analogous to the challenge faced in stepwise regression analysis, factor loadings may not be orthogonal to one another, and researchers often have to perform multiple iterations of model fitting and MI procedures to free up additional parameters in a stepwise manner to maximize power. One general difficulty resides in how to appropriately quantify the variation from repeated application of the MI procedure to the same data, and from sampling variability across independent data sets. No sampling distribution is available – or has been derived (Hu & Bentler, 1995) – for the purpose of quantifying the uncertainties associated with MIs. Moreover, model complexity is not explicitly accounted for in the use of MIs. In addition, as with other test statistics (see Lee, 2007, for a discussion), model inference based on large sample sizes tends to reject models with satisfactory goodness of fit.

We argue that an alternative way to combine the strengths of EFA and CFA is to structure factor analysis as a variable selection problem. Structuring the issue of estimating factor loadings as a Bayesian variable selection problem opens up myriad new possibilities to adopt different prior choices and procedures to resolve issues associated with the MIs. Doing so drastically reduces the computational costs involved in determining factor loading patterns using Bayesian model evaluation criteria such as the Bayesian Information Criterion (Schwarz, 1978), Deviance Information Criterion (Spiegelhalter, Best, Carlin, & Van Der Linde, 2002) and Bayes factor (Kass & Raftery, 1995), which traditionally requires calculating these criteria for 2^P candidate models corresponding to P candidate cross-

loadings. Already within this variable selection framework – though not presented this way by the authors – is the Bayesian structural equation modeling (BSEM) approach by Muthén and Asparouhov (2012). In this framework, the number of factors has to be chosen, and each latent factor is linked to at least one indicator for identification of the scale of the factor as in CFA. However, as distinct from a standard CFA with informative main loading structures, it allows relatively flexible cross-loading structures that are similar to the EFA.¹ We show in this paper that Muthén and Asparouhov's approach may be conceptualized as a special case of Bayesian variable selection in which one possible informative prior specification based on the ridge prior is used to shrink all the cross-loading elements in Λ toward zero. We denote their approach herein as the BSEM approach with ridge prior (BSEM-RP).

A table summarizing the acronyms used is given in Table 1. The remainder of the article is organized as follows. First, we briefly review variable selection approaches in the frequentist and Bayesian frameworks, along with popular priors for Bayesian variable selection other than the ridge prior, including the Bayesian lasso (Park & Casella, 2008) and the spike and slab prior (O'Hara & Sillanpää, 2009). Then we show that model exploration in CFA can be conceived and formulated naturally as a variable selection problem. Our proposed BSEM approach with the spike and slab prior (denoted herein as BSEM-SSP) for selecting and estimating the structure of Λ is presented. Moreover, simulation studies comparing the BSEM-RP and BSEM-SSP, as well as other frequentist methods, are included. We then show that the BSEM-SSP achieves better estimation of the loading structure by shrinking elements with weak signals (i.e., small effect sizes) closer to zero while exerting less shrinkage to elements with moderate and strong signals. Our proposed approach combines model exploration (i.e., to determine the structure of Λ) as well as final parameter estimation within a single step, as distinct from the two-step procedure adopted by Muthén and Asparouhov (2012) in their BSEM-RP. Also, we apply the Bayesian methods to the problem of estimating cross-loadings in real data. Finally, a discussion is given to conclude the paper.

Variable Selection

Variable selection has been a major topic in the statistical literature on regression. Consider the regression model

$$y_i = \mathbf{x}_i \boldsymbol{\beta} + \varepsilon_i, \quad (2)$$

where y_i is a dependent variable, ε_i is an error term, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ are a $p \times 1$ vector of independent variables and coefficients, respectively. It is usually assumed that $\varepsilon_i \sim N(0, \sigma^2)$. To ease presentation we assume that no intercept term exists in the model. Classic variable selection procedures have been developed, such as best subset selection, sequential searches including forward selection, backward elimination and

¹We note that substantively motivated cross-loadings may also be assumed in CFA models. However, to ease prior selection, we follow the convention adopted by Muthén and Asparouhov (2012) and use the term *main loadings* to denote elements in the loading matrix that are supported by substantive knowledge; all other elements are referred to as cross-loadings. As such, all loadings freed up in a CFA, provided that they are substantively motivated, are referred to as main loadings in the current context even though some associated indicators may be linked to multiple latent factors.

stepwise method (Chatterjee & Hadi, 2012). Test statistics such as the likelihood ratio test, score test, and Wald test assess the statistical significance of the elements in β . Model comparison criteria such as the Akaike Information Criterion (Akaike, 1974) and Bayesian Information Criterion (Schwarz, 1978) have also been used. Both test statistics and model comparison criteria are usually based on the maximum likelihood estimator (MLE). When ε_i are i.i.d. and follow $N(0, \sigma^2)$,

$$\hat{\beta}_{MLE} = \arg \min_{\beta} \left\{ \sum_{i=1}^N (y_i - \mathbf{x}_i \beta)^2 \right\}, \quad (3)$$

which is also the ordinary least squares estimator $\hat{\beta}_{OLS}$. The notation “arg min” denotes the operation to find values of the parameters in β for which the function enclosed in the braces is minimized.

Classic variable selection procedures have several disadvantages. The best subset selection requires the comparison of 2^p models, which is not feasible even for moderate p . Sequential search algorithms circumvent this problem, but still face several issues besides those mentioned for MIs. First, regressors are either retained or dropped by a binary decision from the model. Model modification may capitalize on chance characteristics in the data (MacCallum, Roznowski, & Necowitz, 1992), and significant variables in one study may not be significant in other studies, leading to extremely variable parameter estimates and discrepant conclusions across different data sets. Later steps in sequential testing assume the selected loadings are the deterministic null model, ignoring the randomness of sequential testing. As pointed out in the literature of model comparison and selection (see for example, Chatfield, 1995; Draper, 1995; Hoeting, Madigan, Raftery, & Volinsky, 1999, among others), failing to account for model selection uncertainty may lead to underestimated standard errors and overfitting. Also, Altman and Andersen (1989) pointed out that stepwise procedures may yield inflated regression estimates and biased standard errors.

Second, without a proper correction procedure, the type I error rate may be far from the specified significance level. Even though correction procedures exist, the actual performance of these procedures in CFAs with complex factorial structures is not well understood and is, in fact, one of the aspects evaluated in our simulation study. Finally, the parameter and standard error estimates may be severely inflated by collinearity problems, and may lead to falsely identified significant variables (Harrell, 2001). Other methods for variable selection that do not suffer as severely from these disadvantages have been actively developed, one approach of which comprises various regularization and shrinkage methods.

Variable Selection through Regularization

Regularization methods for variable selection use carefully designed penalty functions to *regulate* the loss functions/log-likelihood functions to achieve sparse coefficient estimates and thus variable selection (Tibshirani, 1996; Zou, 2006; Zou & Hastie, 2005). It is possible to obtain Bayesian analogues of well-known regularization methods by specifying the penalty functions as prior distributions in Bayesian models (Leng, Tran, & Nott, 2014; Q. Li

& Lin, 2010; Park & Casella, 2008). These Bayesian regularization (Polson & Scott, 2011) methods differ from traditional Bayesian methods in that the prior distributions not only specify the distributions of unknown parameter values, but also affect the model structure by helping to exclude unimportant parameters more effectively. To provide background on the generality of the Bayesian variable selection approach, we first review the “historical origin” of regularization approaches in regression.

Regularization in the regression framework—The goal of variable selection is to control the complexity of the model characterized by β . Unlike testing methods which quantify the complexity by the number of non-zero elements in β , regularization methods (Bickel et al., 2006) incorporate the norms of β into the loss/log-likelihood functions, and estimate β with

$$\hat{\beta}_R = \arg \min_{\beta} \left\{ \sum_{i=1}^N (y_i - \mathbf{x}_i \beta)^2 + \text{norm}(\beta) / \tau^2 \right\}, \quad (4)$$

where $\text{norm}(\beta)$ is a function of β that measures the length of β as a vector and quantifies the model complexity of eq. (2). Commonly used norms include: (1) the L_0 -norm $\|\beta\|_0$, namely, the number of nonzero elements in β ; (2) the L_1 -norm $\|\beta\|_1 = \sum_{j=1}^p |\beta_j|$; and (3) the L_2 -norm $\|\beta\|_2 = \sqrt{\sum_{j=1}^p \beta_j^2}$. The $\text{norm}(\beta) / \tau^2$ is often referred to as the penalty. The magnitude of the vector $\hat{\beta}_R$ as a whole is less than, or equal to $\hat{\beta}_{MLE}$. Smaller τ^2 ($\tau \geq 0$) weighs the norm more and thus penalizes more heavily, representing a preference for less complex over more complex regression models. τ^2 is called the inverse tuning parameter, and is usually determined by predictive criteria such as cross-validation.

Two well-known special cases of eq. (4) are the ridge² (Hoerl & Kennard, 1970) and lasso³ (Tibshirani, 1996) estimators which use L_2 -norm and L_1 -norm respectively as the penalty function in eq. (4) (for an introduction see Hastie et al., 2009). The ridge estimator shrinks the magnitude of all elements of β towards zero.⁴ The L_2 penalty function reflects the subjective belief that estimates of β with large magnitudes are less preferable. It was designed to improve the stability of the MLE estimator in eq. (3) when the covariates in $\mathbf{X}$ are strongly correlated, and $(\mathbf{X}^T \mathbf{X})^{-1}$ is sensitive to random errors in the data. The ridge estimator may be solved analytically by taking the first derivative of the function after argmin in eq. (4).

²The ridge estimator was designed to solve the unstable estimation caused by the flat “ridge-like” hyper-surface of

$-\sum_{i=1}^N (y_i - \mathbf{x}_i \beta)^2$, which may be caused by multicollinearity among the elements in $\mathbf{x}_i$.

³Lasso is the acronym for *least absolute shrinkage and selection operator*, which uses absolute values as shrinkage functions and the consequence that it selects important variables through setting the others to zero.

⁴Let $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ and $\mathbf{y} = (y_1, \dots, y_n)^T$. Compared to ordinary least squares $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$, which is also the MLE under i.i.d. normal error assumption, $\hat{\beta}_{Ridge} = (\mathbf{X}^T \mathbf{X} + \mathbf{I} / \tau^2)^{-1} \mathbf{X}^T \mathbf{y}$, where $\mathbf{I}$ is an identity matrix. Let $\hat{\mathbf{x}}_j$ be the j th column of $\mathbf{X}$. When the covariates are orthogonal and $\mathbf{X}^T \mathbf{X} = \mathbf{I}$, the j th elements in $\hat{\beta}_{MLE}$ is $\hat{\mathbf{x}}_j$ and that in $\hat{\beta}_{Ridge}$ is given by $\hat{\beta}_{MLE} / (1 + \tau^2)$, which shrink the estimates to zero proportional to the magnitude of $\hat{\beta}_{MLE}$.

The lasso estimator, in contrast, was motivated by variable selection problems. As illustrated in Hastie et al. (2009), the lasso estimator sets the elements of the MLE estimator that are smaller than $1/\tau^2$ to exactly zero, and shrinks the other elements to zero by $1/\tau^2$. Utilizing the L_1 penalty function reveals subjective belief that the coefficients of unimportant covariates may be set directly to zero, as opposed to being shrunk to zero as in ridge regression. Covariates are selected and estimated jointly rather than sequentially. The lasso estimator may be viewed as a constrained optimization problem and may be solved by the least angle regression algorithm (Efron, Hastie, Johnstone, & Tibshirani, 2004) or the coordinate descent algorithm (Friedman, Hastie, & Tibshirani, 2010).

In sum, regularization in the frequentist regression framework optimizes a penalized form of the objective function. The choice of the penalty function reflects the subjective beliefs of the researchers. Bayesian approaches, which are based on posterior distributions formed by likelihoods and prior distributions, are another natural way to formulate these problems. We consider Bayesian analogues of the regularization methods described above.

Bayesian Regularization Methods—Bayesian analysis is a broad topic, so we do not attempt to provide a detailed introduction. The reader may refer to Muthén and Asparouhov (2012) for an excellent introduction to Bayesian analysis and important practical considerations. The key point with respect to the Bayesian regularization methods is that they accomplish the same goals as the frequentist approaches through assigning appropriate prior distributions derived from subjective knowledge about variable selection, which also shapes the penalty functions in the frequentist approaches.

For $\mathbf{y} = (y_1, \dots, y_n)^T$, Bayesian analysis is usually based on the posterior distribution of the parameters, $p(\boldsymbol{\beta}|\mathbf{y})$, which depends on the likelihood, $p(\mathbf{y}|\boldsymbol{\beta})$, and the prior distribution, $p(\boldsymbol{\beta})$, through the Bayes' theorem $p(\boldsymbol{\beta}|\mathbf{y}) = p(\mathbf{y}|\boldsymbol{\beta})p(\boldsymbol{\beta})/p(\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\beta})p(\boldsymbol{\beta})$, where $p(\mathbf{y})$ is the marginal distribution of $\mathbf{y}$. The mean and mode of $p(\boldsymbol{\beta}|\mathbf{y})$ are commonly used as the estimators of $\boldsymbol{\beta}$.

For the regression model in eq. (2), two popular prior choices for $\boldsymbol{\beta}$ are the Gaussian distribution and the Laplace distributions, expressed respectively as:

$$p(\boldsymbol{\beta}|\sigma^2) \propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^p \beta_j^2 \right\}, \quad (5)$$

$$p(\boldsymbol{\beta}|\sigma^2) \propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^p |\beta_j| \right\}, \quad (6)$$

where σ^2 is a variance parameter with prior distribution $p(\sigma^2)$. Following Bayes' theorem, the posterior mode of $p(\boldsymbol{\beta}|\mathbf{y})$ under the prior choices in eqs. (5) and (6) can be shown to be equivalent to the frequentist ridge and lasso estimator of $\boldsymbol{\beta}$, respectively, given the same σ^2 .

This forms the basis of the Bayesian lasso (Park & Casella, 2008), in which the prior in eq. (6) is imposed in place of the penalty function in the frequentist lasso to perform Bayesian variable selection (Park & Casella, 2008). Estimators of σ^2 may be computed from the mean or mode of the posterior distribution of σ^2 , which may be calculated with numerical and Monte Carlo algorithms.

Bayesian regularization approaches have several advantages compared to their frequentist counterparts. Sampling from the posterior distribution provides more information for estimation and inference, including valid standard errors which are difficult to derive for the frequentist lasso (Kyung, Gill, Ghosh, & Casella, 2010). In addition, uncertainties in model selection can be characterized by the posterior probabilities of different models, and the probabilities of inclusion or exclusion of single or multiple parameters can also be quantified. Powerful Bayesian computational techniques have further enabled the estimation of a wide range of models: penalty functions that often lead to optimization difficulties in the frequentist framework may be handled with relative ease in the Bayesian framework, often yielding better estimation results. We discuss each these advantages in more detail in the context of variable selection in Bayesian CFA.

Regularization of Cross-Loadings in Bayesian CFA via Prior Distributions

We compare two prior distributions for selecting the loading structure in factor analysis. One is the BSEM-RP method proposed by Muthén and Asparouhov (2012); the other is a novel approach, BSEM-SSP, based on the spike and slab prior.

BSEM-RP for Factor Analysis

Muthén and Asparouhov (2012), in discussing some powerful features afforded by the Bayesian analytic framework, proposed a prior parameter specifications where exact zeros in BSEMs are replaced with approximate zeros based on informative, small-variance priors. In the special case of CFA, the proposed approach operates as follows. Let λ_{jk} be the element in the j th row and k th column of Λ . Independent Gaussian prior distributions $N(0, \sigma_{jk}^2)$ are assigned to λ_{jk} as

$$p(\Lambda | \Sigma_0) \propto \exp \left\{ \sum_{j=1}^p -\frac{1}{2\sigma_{jk}^2} \lambda_{jk}^2 \right\} \quad (7)$$

where σ_{jk}^2 is a hyperparameter assumed to be known from prior knowledge and Σ_0 is a matrix containing all σ_{jk}^2 , and the proportionality symbol indicates that the left side of eq. (7) equals its right side times a constant for all values of Λ and a given Σ_0 . The λ_{jk} s are further divided into two groups. The first group consists of *main-loadings* generally implemented in standard CFA as supported by substantive knowledge. For these main loadings, σ_{jk}^2 is assumed to be large enough to allow λ_{jk} to take on values that deviate substantially from zero.⁵ The second group consists of the remaining elements of λ_{jk} s that are typically set to

zero in CFA (i.e., the *cross-loadings*) as “hard” constraints. Muthén and Asparouhov (2012) replaced the “hard” constraints with “soft” constraints characterized by prior distributions of λ_{jk} with very small variance σ_{jk}^2 (e. g. , $\sigma_{jk}^2=0.01$). In the original formulation of the BSEM-RP, no cross-loadings are fixed to zero for identification and the minimum number q^2 constraints needed for identification in factor analysis are not imposed. Instead, the cross-loadings are assigned priors concentrated around zero with a small σ_{jk}^2 to reflect the prior belief that these λ_{jk} s have large prior probability near 0. This informative prior structure concentrates the posterior distributions for λ_{jk} around zero, unless there is strong evidence from the likelihood indicating otherwise.

Although Muthén and Asparouhov (2012) did not link their proposed prior to the Bayesian variable selection framework, comparison to eq. (5) shows that their prior for the cross-loadings is a special case of the Bayesian ridge regression prior (RP). The only minor distinction is that the prior variance in the Bayesian RP in eq. (5) is also estimated as a parameter (i.e., with its own prior distribution), as opposed to being fixed at known constants as in Muthén and Asparouhov (2012). As such, we refer to Muthén and Asparouhov’s approach herein as the BSEM-RP.

The BSEM-RP is implemented by first assigning prior distributions to ϕ , $\Psi = \text{diag}(\psi_1, \dots, \psi_p)$, μ_j and the main-loadings in Λ . The following (proper) conjugate prior distributions are used,

$$\begin{aligned} p(\lambda_{jk}|\psi_j) &= N(\lambda_{0jk}, \psi_j \sigma_{\lambda_{0jk}}^2), & p(\mu_j|\psi_j) &= N(\mu_{0j}, \psi_j \sigma_{\mu_{0j}}^2), \\ p(\psi_j) &= IG(\alpha_{1j}, \alpha_{2j}), & p(\Phi) &= IW(\rho_0, \Phi_0), \end{aligned} \quad (8)$$

where *IG* and *IW* stand for inverse-gamma and inverse-Wishart distributions, respectively.

The terms $\rho_0 > q - 1$, $\alpha_{1j} > 0$, $\alpha_{2j} > 0$, λ_{0jk} , $\sigma_{\lambda_{0jk}}^2 > 0$, μ_{0j} , $\sigma_{\mu_{0j}}^2 > 0$, and positive definite matrix Φ_0 are hyperparameters whose values are based on prior information. We also considered the default prior setting of BSEM for ϕ , Ψ in Mplus (Muthén & Asparouhov, 2012),

$$p(\psi_j) \propto 1, \quad p(\Phi) \propto 1, \quad (9)$$

and we refer to this specific prior choice as the noninformative prior.⁶

Using Gibbs sampling, the empirical distribution of the N_1 Markov chain Monte Carlo (MCMC) samples after burn-in, denoted as $\{\Lambda^{(t)}, \mathbf{R}^{(t)}, \Phi^{(t)}, \Psi^{(t)} / N_0 < t \leq N_0 + N_1\}$,

⁵Where substantive knowledge about the mean and variance of the main-loadings is available, informative prior distributions may also be used.

⁶Note that these prior distributions are improper prior distributions (thus integration of this density function does not entail 1 or even a finite number). They are usually used to convey the subjective belief that there is little information in the prior distribution.

approximates the posterior distribution on which Bayesian estimates and inference are based. For instance, sample means and percentiles are used to approximate posterior means (used as Bayesian point estimates) and credible intervals (for deducing whether the point estimates are substantially different from zero). Many quantities which involve integration, i.e., expectations, with respect to the posterior distributions may also be estimated with MCMC samples. In the context of the BSEM-RP, Muthén and Asparouhov (2012) proposed using the 95% percent credible interval to determine if the cross-loadings are substantially different from zero.⁷ It is important to emphasize that even though the BSEM-RP does not involve sequential model comparison as in the MI-based approaches for determining factor loading structures, the original formulation of the BSEM-RP as advocated by Muthén and Asparouhov (2012) is still a two-step procedure. Specifically, the authors suggest that after the loading structure has been determined through inspection of the credible intervals, a revised CFA model should be refitted for final parameter estimates and inference.

BSEM-SSP for Factor Analysis

Several issues inherent to BSEM-RP motivated us to study an alternative method of determining the structure of the loading matrix. As an alternative method of selecting λ_{jk} , we propose using the spike and slab prior (SSP, Ishwaran & Rao, 2005; Mitchell & Beauchamp, 1988),

$$p(\lambda_{jk}|r_{jk}) = (1-r_{jk})\delta_0 + r_{jk}N(0, c_{jk}^2), \text{ with } p(r_{jk}) = \text{Bernoulli}(p_{jk}), \quad (10)$$

for all cross-loadings in $\mathbf{\Lambda}$ except for the cross-loadings that are fixed to zero to satisfy the minimum constraints of the model.⁸ Here, r_{jk} is a binary latent variable commonly used in the formulation of mixture models, which indicates whether the loading in the j th row and k th column of $\mathbf{\Lambda}$ should be set to zero or freed to deviate from zero; δ_0 is a point mass function at 0, and c_{jk}^2 (where c_{jk}^2 is usually substantially greater than 0) and p_{jk} are hyperparameters reflecting the user's prior knowledge or subjective belief.

This prior is a mixture of two components: (1) the “*spike*” – representing the prior belief that λ_{jk} should be fixed at zero and (2) the “*slab*” – the wider normal distribution representing the belief that λ_{jk} should be estimated from data because it may deviate substantially from 0. For variable selection purposes, the promising subsets of predictors can be identified as those with higher posterior inclusion probabilities, namely, $p(r_{jk} = 1|y)$, which weighs both the data and the prior to quantify the probability for λ_{jk} to be non-zero given the data. Note that $p(r_{jk} | y)$ is often different from $p(r_{jk})$ because information from the likelihood is incorporated.

⁷To avoid confusion with the frequentist significance, we use the word “substantial” to describe a loading that is large enough to represent a nonignorable relation between a pair of indicator and factor in Bayesian approaches.

⁸In BSEM-SSP, the spike and slab prior is not used to identify the factor model. The spike and slab prior is only used to determine all the remaining cross-loadings after minimum identification conditions are set.

Various mixture distributions have been used in Bayesian regularization (O'Hara & Sillanpää, 2009). To aid comparisons between the BSEM-RP and BSEM-SSP, it is helpful to note that the SSP is the limiting case of the hierarchical normal mixture prior (George & McCulloch, 1993),

$$p(\lambda_{jk}|r_{jk}) = (1-r_{jk})N(0, \sigma_{jk}^2) + r_{jk}N(0, c_{jk}^2), \text{ with } p(r_{jk}) = \text{Bernoulli}(p_{jk}), \quad (11)$$

as σ_{jk}^2 tends to zero. That is, when σ_{jk}^2 is small, the first component would center narrowly around zero, as opposed to being specified as a point mass (i.e., a spike) at zero.

Interestingly, the BSEM-RP approach can be expressed as a special case of eq. (11) by

setting $r_{jk} = 0$ and $\sigma_{jk}^2 = 0.01$. A similar prior as eq. (11) was used recently by Mavridis and Ntzoufras (2014) as a Bayesian item selection method for EFA. Even though the mixture framework shown in eq. (11) may appear at first to provide a more general framework for comparing the BSEM-SSP and BSEM-RP, the comparison is computationally cumbersome with limited practical advantages. In fact, the SSP has been shown to converge⁹ faster (O'Hara & Sillanpää, 2009) and is more parsimonious because with the point mass component, λ_{jk} may be shrunk exactly to zero as in the lasso. We focus on comparing the BSEM-RP and the BSEM-SSP, first at the theoretical level, followed by a simulation study to address unresolved questions.

Posterior Estimation and Inference for the BSEM-SSP

Let $\mathbf{R} = (r_{jk})$ be a $p \times q$ matrix, where the r_{jk} , as noted earlier, are latent binary variables that can take on the value of either 0 or 1. For BSEMs, our primary interest is in evaluating the properties of the posterior distribution $p(\boldsymbol{\theta}, \mathbf{R}|\mathbf{Y})$. To do so we first assign the same conjugate priors to other parameters as in the BSEM-RP (see eq. (8)). In the BSEM-SSP, we can obtain point estimates of the cross-loadings that take into consideration uncertainty inherent in variable selection. Let $\mathbf{M}$ be a hypothetical factor analytic model (see eq. (1)) of interest and $\hat{\mathbf{R}} = (\hat{r}_{jk})$ represents the corresponding cross-loading structure. The three estimates available for this purpose in the BSEM-SSP framework are $E(\lambda_{jk}|\mathbf{Y}, \mathbf{R} = \hat{\mathbf{R}})$, $E(\lambda_{jk}|\mathbf{Y}, r_{jk} = 1)$ and $E(\lambda_{jk}|\mathbf{Y})$, which we call the SSP confirmatory estimator, SSP conditional estimator and SSP marginal estimator, respectively. Estimating of the loading matrix in a CFA model, e.g., the final CFA model chosen, should be equivalent to the confirmatory estimator given the appropriate $\hat{\mathbf{R}}$. This expectation can be estimated by the mean of the subset of MCMC samples for which $\mathbf{R}^{(t)} = \hat{\mathbf{R}}$. However, the proportion of such MCMC samples may be small because BSEM-SSP explores many possible models and the MCMC samples are distributed among these models according to the posterior probabilities, $p(\mathbf{R}|\mathbf{Y})$. The mean of the small subset of MCMC samples may not approximate the expectation $E(\lambda_{jk}|\mathbf{Y}, \mathbf{R} = \hat{\mathbf{R}})$ well enough. The SSP conditional estimator may be a good alternative estimator because more MCMC samples are typically available. As demonstrated in the simulation study, the SSP conditional estimator provides similar results to CFA models where the loading structure

⁹In the Bayesian framework, convergence is defined as convergence of the empirical distribution of the Monte Carlo samples of parameters to the target posterior distributions.

characterized by $\hat{\mathbf{R}}$ are assumed to be known. Hence, if one is interested in obtaining results from a “final” CFA based variable selection procedure, the estimates are already available from the BSEM-SSP through the SSP conditional estimator and through refitting the CFA model after model selection is complete. The SSP confirmatory and conditional estimators are recommended if the researcher is confident with the model and is willing to bypass the model selection uncertainty. In comparison, the SSP marginal estimator is the posterior mean as averaged over two components of the mixture distribution as $E(\lambda_{jk} | \mathbf{Y}, r_{jk} = 1)P(r_{jk} = 1 | \mathbf{Y}) + E(\lambda_{jk} | \mathbf{Y}, r_{jk} = 0)P(r_{jk} = 0 | \mathbf{Y})$ and it incorporates model selection uncertainty. This estimator may thus be preferred in situations where the researcher does not regard the final model as “definitive” but rather, wants to weigh the cross-loading estimates by their variable selection uncertainties.

Even though the BSEM-SSP may be regarded as a one-step approach for model selection, parameter estimation, and inference, alternative model comparison measures may also be computed using MCMC samples. To select a final model from a collection of candidate models, one approach is to select a final model in which only cross-loadings with $p(r_{jk} = 1 | \mathbf{Y}) > 0.5$ are retained, resulting in the median probability model (Berger & Barbieri, 2004). This specific model is the optimal predictive model in regressions with independent and identically distributed Gaussian errors. Model comparison may also be performed by evaluating the posterior probabilities of all candidate models. Let M be a hypothetical factor analytic model of interest and $\hat{\mathbf{R}} = (\hat{r}_{jk})$ represents the corresponding cross-loading structure. The posterior probability of model M , $p(\mathbf{R} = \hat{\mathbf{R}} | \mathbf{Y})$ can be computed using MCMC samples from fitting the BSEM-SSP, from which the Bayes factor can then be computed. Details on computing the posterior model probability and the Bayes factor are summarized in the supplemental materials.

Theoretical Advantages and Limitations of the BSEM-SSP and BSEM-RP

Successful variable selection in the hybrid BSEM-CFA framework is contingent on correctly classifying zero- and non-zero loadings. In this sense, the SSP prior in eq. (10) has several appealing characteristics. Compared to the BSEM-RP and EFA, BSEM-SSP facilitates the classification of zero cross-loadings by shrinking them closer to zero, leading to more parsimonious models and interpretations. For non-zero loadings, the SSP provides higher power for detection by learning the posterior distribution of $r_{ij} = 1$ and exerting less effect of shrinkage on the moderate or large cross-loadings. In comparison, the BSEM-RP uses strong prior distributions with small σ_{jk}^2 s, corresponding to a small τ^2 in eq. (4) and consequently, a large penalty on large λ_{jk} . If most cross-loadings are small, this specific prior yields a parsimonious cross-loading structure. However, if the cross-loadings are moderate or large, this prior may over-shrink toward zero. In other words, as in ridge regression, the prior specification used for the cross-loadings in BSEM-RP shrinks all parameters proportionally to zero, instead of performing targeted variable selection as lasso does.

Second, BSEM-SSP is a one-step approach for model selection, parameter estimation, and inference, which makes the data analytic process more convenient. BSEM-RP replaces sequential model modification procedures such as MI in a one-step model selection process. As with the BSEM-SSP, the BSEM-RP circumvents having to compute quantities that are

hard to calculate in sequential testing (e.g., type I error in sequential model modification). However, the model selection process is then followed up in stage 2 with a Bayesian CFA in which cross-loadings whose credible intervals included zero in stage 1 are set to zero, while the other cross-loadings are estimated with large σ_{jk}^2 . Unlike the two-stage BSEM-RP, BSEM-SSP does not require model refits based on CFA models.

Third, the less informative prior distributions of the cross-loadings in the BSEM-SSP exert less influence on the posterior distribution. Our simulation results showed that the BSEM-SSP is not very sensitive to the prior distributions specified in eqs. (8), (9), and (10). Reasonably different hyperparameters of these prior distributions lead to similar posterior distributions and estimates. In comparison, Muthén and Asparouhov (2012) and Asparouhov, Muthén, and Morin (2015) suggested running BSEM-RP multiple times by using several prior variances in the range of [0.001, 0.1] for sensitivity checking, and selecting the appropriate prior variances through subjective selection with two criteria.

Finally, as mentioned previously, the BSEM-SSP enables fast calculation of Bayes factors (Pang & Gill, 2009) not available for Bayesian lasso and BSEM-RP. Moreover, the cross-loading selection uncertainties are quantified in the BSEM-SSP approach by various posterior inclusion probabilities that offer continuous inclusion/exclusion measurements, and can be incorporated into parameter estimation. This is not applicable in BSEM-RP where inferences based on the estimates from stage 2 are performed without consideration of the model selection uncertainty in the first stage. We compare the estimates of cross-loadings and the covariance of the latent factors in a simulation study.

Simulation

Simulation Design

To compare the performance of the BSEM-SSP, BSEM-RP and MI approaches with regard to cross-loading selection, parameter estimation, and inference, we conducted a simulation study with two factor loading structures. The structures of the loading matrices, Λ , were defined to be

$$\begin{pmatrix} 1^* & 0 & 0 \\ 1 & x & 0 \\ 1 & 0 & 0 \\ 0 & 1^* & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1^* \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1^* & 0 & x \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ x & 1^* & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & x & 1^* \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad (12)$$

where the elements marked with an asterisk were fixed for identification purpose; other elements with the value 1 were main-loadings, and elements marked with 'x' were non-zero cross-loadings. The elements with asterisk were fixed to identify the scale of the latent

factors. The first factor loading structure was characterized by only one non-zero cross-loading and is referred to herein as the “*Single Cross-Loading Condition*.” The second factor loading structure, with multiple non-zero cross loadings, are referred to herein as the “*Multiple Cross-Loading Condition*.” For each loading structure, three sample sizes (100, 200, and 300) and three magnitudes of cross-loadings ($x = 0.1, 0.2, 0.3$) were considered, resulting in 2 (loading structures) $\times 9$ (sample sizes by cross-loading magnitude variations) = 18 conditions. For each condition, we generated 100 datasets. For each replication, a total of 4000 MCMC samples were discarded as burn-in samples and 4000 additional samples were acquired for posterior estimation and inference. The other parameters were set to $\mu_j = 0$, $\psi_j = .3$ for $j=1, \dots, 9$, and $\Phi = \begin{bmatrix} 1 & .5 & .5; .5 & 1 & .5; .5 & .5 & 1 \end{bmatrix}$.

Methods for Comparison—The goal of the simulation study was to compare cross-loading selection and estimates of Λ and Φ from the BSEM-SSP, BSEM-RP, and selected frequentist methods. In all cases, we used the frequentist and Bayesian CFA in which the true loading structures eq. (12) were assumed to be known and only the main-loadings and the cross-loadings marked as “x” were freed as the gold standards.

As mentioned previously, we suggested applying BSEM-SSP in an identified model with q^2 identification constraints, including q constraints of the main-loadings to identify the scale of the factors, and $q(q-1)$ constraints on the remaining cross-loadings. In comparison, the BSEM-RP used the informative RP but no additional identification constraints. However, in these simulation conditions, we found that the BSEM-SSP worked well without the $q(q-1)$ constraints on the cross-loadings. For fair comparison, we freed up all cross-loadings in the BSEM-RP and BSEM-SSP. For the BSEM-SSP, the spike and slab prior in eq. (10) was used for all the cross-loadings (including all the zero and ‘x’ elements in eq. (12)) with $c_{jk}^2 = 10^4$ and $p_{jk} = 0.5$. For the BSEM-RP, the normal distribution, $N(0, 0.01)$ was used as the prior for all the cross-loadings. For both the BSEM-RP and BSEM-SSP, we use the prior in eq. (8) for the main-loadings λ_{ij} and the intercept μ_j with $\lambda_{0jk} = \mu_{0j} = 0$, and $\sigma_{\lambda_{0jk}}^2 = \sigma_{\mu_{0j}}^2 = 10^{10}$, and use the noninformative prior in eq. (9) for Φ and Ψ .

Three frequentist-based methods were selected as comparisons to the Bayesian approaches. First, we considered a forward-backward stepwise model modification (FBS-MI) procedure. A similar method can be found in Kano (2002). Specifically, starting from a model without cross-loadings, the most significant cross-loading with the p -value of modification index smaller than α (a prespecified threshold) was added to the model. The new model was refitted, and additional cross-loadings were added sequentially (one at a time) until no cross-loading with p -value smaller than α existed. Then, the Wald test statistics of all cross-loadings were checked, and the one with the largest p -value that was greater than α was excluded from the model. The new model was refitted and a final model was reached until the Wald test p -values of all cross-loadings were smaller than α . We used 21 levels of α in the range of 0.223 and 0.002, equivalent to $\exp(-1.5)$, $\exp(-1.75)$, ..., $\exp(-6.5)$, and ran the whole FBS-MI separately at each target significance level. This stepwise procedure was implemented in R (R Core Team, 2013) with the MplusAutomation package (Hallquist & Wiley, 2014).

We also considered a forward selection procedure with a Scheffé-type test (FS, Hancock, 1999), which was developed to better control the type I error across the set of all possible post hoc model modifications. In each step, a chi-square test with degree of freedom equal to the degree of freedom of the model was used. Similarly, we used 25 levels of α in the range of 0.779 and 0.002, equivalent to $\exp(-0.25)$, $\exp(-0.5)$, ..., $\exp(-6.25)$, and ran the whole FS separately at each significance level.

The third frequentist method considered was EFA with target rotation (Browne, 2001). The factor loading matrices are rotated to match a pre-specified target matrix. We set the target matrix to contain only the main-loadings. Confidence intervals were used to decide the significance of the cross-loadings with and without the Bonferroni correction.

Sensitivity Analyses—We performed additional analyses to investigate the sensitivity of the BSEM-SSP and the BSEM-RP to (1) prior specification, (2) choices of σ_{jk}^2 in the BSEM-RP, and (3) variations in the factor structure. To test sensitivity to prior specification, we compared results from our earlier simulation to those from a set of “informative priors”, obtained by setting the hyperparameters in eq. (8) to $a_{1j}=11$, $a_{2j}=3$, $\rho_0=7$, Φ_0 = the true value of Φ , $\lambda_{0jk}=0$, and $\sigma_{\lambda_{0jk}}^2=1$. The informative prior distributions were specified to have means that were equal to the true values used in the simulation. Additionally, we also considered the prior settings with means equal to two times and one half of the true values. We calculated the estimated potential scale reduction (EPSR, Gelman, 1996) values of each parameter using the samples of the three MCMC chains, each of which is based on one of the three prior specifications, to assess possible divergence in posterior estimates with changes in prior specifications. EPSR values of all parameters that are smaller than 1.2 after certain iterations indicate that the MCMC chains converge to similar regions at the corresponding iterations.

As pointed out by Muthén and Asparouhov (2012) and also noted earlier, BSEM-RP is an approach using informative, small-variance prior for the cross-loadings to identify models. Consequently, BSEM-RP has been shown to be sensitive to the choice of the variance σ_{jk}^2 (Muthén & Asparouhov, 2012). To replicate this earlier finding, we first investigated the effects of σ_{jk}^2 on cross-loading selection with three values of σ_{jk}^2 , specifically, 0.001, 0.01 and 0.1. In addition, instead of setting σ_{jk}^2 to be a constant, we also tested an alternative scenario wherein σ_{jk}^2 was allowed to be estimated as a parameter with its own inverse-gamma prior, a widely used conjugate prior distribution in Bayesian analysis (Lee, 2007). This approach is referred to herein as the BSEM-RP-VP, where VP stands for variance prior. This option is currently unavailable in Mplus; thus we implemented the corresponding model in R.

To examine the performance and sensitivity of the two approaches to increased complexity in the factor structure, we extended the “*Multiple Cross-Loading*” structure shown in eq. (12) to yield a condition with 8 factors identified by 24 measurement variables, which is denoted herein as the “*Complex Multiple Cross-Loading Condition*.” The true structure of Λ is as follows. Each factor loads on 3 distinct measurement variables like eq. (12) with the

main-loadings equal to 1. The non-zero cross-loadings are $\lambda_{22,1}$ and $\lambda_{3(k-1)+1,k+1}$, where $k = 1, \dots, 7$. We evaluated the performance of the BSEM-SSP and BSEM-RP using the same measures as described earlier. The main-loading fixed for identification is $\lambda_{3(k-1)+2,k}$, where $k = 1, \dots, 8$.

One caveat to note is as follows. The original BSEM-RP formulation presented by Muthén and Asparouhov (2012), and considered in our simulation, requires the use of small values of σ_{jk}^2 because these values serve both as “informative priors” and as identification constraints. That is, in the absence of additional constraints, the CFA models considered in our simulation above are under-identified (i.e., only q loadings are set to 1.0 to identify the scales of the latent factors, compared to the minimum q^2 constraints needed to identify the CFA). To test the sensitivity of the BSEM-RP under the BSEM-RP-VP approach, we fixed $q(q-1)$ additional cross-loadings to zero to yield an identified model for all the approaches considered (BSEM-RP, BSEM-RP-VP and BSEM-SSP). Specifically, all the cross-loadings in the $(3(k-1) + 2)$ th rows are fixed at zero for $k = 1, \dots, 8$.

To explore sensitivity of the BSEM-SSP and BSEM-RP to changes in the identification constraints in the factor loading structure, we used the “*Multiple Cross-Loading Condition*”, and changed the locations of the $q^2 = 9$ constraints to yield an identified model. We tried three different cases. In each case, the main-loadings and cross-loadings in three rows were set to 1 and 0, respectively: (1) in the 1st, 4th, and 7th rows of Λ ; (2) in the 2nd, 5th, and 8th rows of Λ ; and (3) in the 3rd, 6th and 9th rows of Λ . We then calculated the EPSR values from these three models to evaluate possible divergence in posterior estimates with these changes.

Simulation Results

Cross-loadings Selection—The receiver operating characteristic (ROC) curves summarizing the performance of all the approaches considered are shown in Figure 1. The ROC curves show the trade-off of true positives (TPs) and false positives (FPs) as the thresholds of positive incidents vary. TPs correspond to the elements in eq. (12) marked as ‘x’, which represent the cross-loadings with non-zero true values that were correctly identified as non-zero, while FPs correspond to the elements in eq. (12) marked as 0, namely, cross-loadings with zero true values that were incorrectly identified as non-zero. Each cross-loading was identified as “statistically different” from zero (i.e., a “positive incident”) for the five approaches as follows: (i) for the BSEM-SSP, when $p(r_{jk} = 1|\mathbf{Y})$ was greater than a threshold (0.5 as default); (ii) for the BSEM-RP, when the absolute value of the ratio between a point estimate and its standard error was greater than a threshold. The default threshold was set to 1.96 for tests without Bonferroni correction and 2.99 for tests with Bonferroni correction; (iii) for the FBS-MI, when a particular cross-loading was retained in the final model ($\alpha = 0.05$ as default); (iv) for EFA with target rotation, the criteria used were identical to BSEM-RP; (v) for the forward Scheffé-type test (FS), the criteria were identical to FBS-MI. The TP rates were computed as the ratios of the sum of TPs to the sum of the number of cross-loadings with non-zero true values over all replications, while the FP rates were computed as the ratios of the sum of FPs to the sum of the number of cross-loadings with zero true values over all replications.¹⁰ The points on the ROC curves

represent the TP and FP rates obtained at different thresholds for the same simulated datasets, and the ROC curves essentially depict the collection of points (or TP and FP rates) obtained across the set of prespecified thresholds. Specifically, for the non-iterative methods, i.e., the BSEM-SSP, BSEM-RP, and EFA with target rotation, we basically used the empirically observed values of the quantities used to determine the “statistical difference” of the cross-loadings (e.g., $p(r_{jk} = 1|Y)$) as thresholds to calculate the TPs and FPs and to draw the ROC curves. For FBS-MI and FS, the 21 and 25 levels of alpha values described previously were used, respectively.

The corresponding TP/FP rates under the default thresholds are indicated by distinct symbols in Figure 1. For BSEM-RP and EFA, we plotted the TP/FP rates at two thresholds, that is, 0.05 with and without Bonferroni correction, which are shown with unfilled and solid symbols, respectively. In some settings, the FP rates at the cut-off are too large, and the symbols are out of the range of display. A more accurate classifier produces more true positives given the same level of false positive, or equivalently, fewer false positives given the same level of true positive, which correspond to ROC curves closer to the upper left corner. The lower left corner of the ROC curves, which show the TPs/FPs corresponding to the default thresholds were of particular interest to us. In addition, the numbers of replications in which the correct loading structures in eq. (12) were accurately recovered are shown in Table 2.

Results from the two factor loading conditions (i.e., single vs. multiple cross-loading condition) were largely comparable. When the effect size was 0.1 (see Figures 1(a), 1(c), 1(e) and 1(g)), BSEM-SSP, BSEM-RP, FBS-MI, EFA, and Hancock’s FS method performed similarly. FBS-MI and EFA performed slightly better in the multiple cross-loading condition when the non-zero cross-loadings were 0.1. In conditions with larger sample sizes and larger magnitudes of nonzero cross-loadings (see Figures 1(b), 1(d), 1(f) and 1(h)), BSEM-SSP and FBS-MI yielded better ROC curves than the BSEM-RP and EFA. In these conditions, the prior $N(0, 0.01)$ in the BSEM-RP assigned small prior probability for loadings that deviated from 0, thus leading to reduced power for the BSEM-RP to detect TPs due to the discrepancy between the prior and the data.

Overall, the performance of the BSEM-SSP was comparable to that of the FBS-MI in the conditions considered. The FBS-MI is not always superior to BSEM-SSP and vice versa because neither BSEM-SSP nor FBS-MI gives a larger false positive rate at all levels. At the default threshold, the FBS-MI was associated with better TP rates, especially in conditions with small effects and sample sizes; in contrast, the BSEM-SSP produces lower FP rates. The default threshold of the FBS-MI appeared to be too liberal and produced inflated FPs that were not compensated to the same degree by the increase in TPs, as indicated by the relatively flat dot-dashed lines to the left of the triangles. Without appropriate multiple comparison correction, the type I errors of the sequential procedure appeared inflated,

¹⁰FP rate is conceptually related to the type I error rate, which is the proportion of rejection of the null hypothesis given that the null hypothesis is true. However, type I error rates are set by the researcher based on model assumptions before hypothesis testing. FP rates appear in much wider context related to binary classification, which are calculated after multiple cases are classified, and requires the knowledge of the true state of each case (true or false).

leading to too many falsely identified cross-loadings and reduced power for detecting the true loading structure in conditions with larger sample sizes (see Table 2).

The balance between true and false discovery rates for forward Scheffé-type test (FS) approach was also less than satisfactory. In contrast to the FBS-MI, the FS approach was found to have a better control over the false positive rates compared to the FBS-MI, but at the expense of low detection power. In particular, the rate of correct detection of truly non-zero cross-loadings at the default threshold (0.05) was much lower than those of the other methods. The ROC curves of the FS approach are only visible as a dot in each of the Figures 1(a), 1(c), 1(e), and 1(g) because the TPs and FPs are zero under all 25 levels of alpha. The ROC curves of the FS approach in Figures 1(b), 1(d), 1(f), and 1(h) are vertical lines at $FP = 0$, and overlap with the ROC curves of the other methods around $FP = 0$. These ROC curves demonstrate that the FS approach may be too conservative, with low false detection rates of zero cross-loadings and the low detection power of non-zero cross-loadings.

EFA with target rotation performed less satisfactorily with larger sample sizes and larger magnitudes of nonzero cross-loadings. EFA begins from a model having a complex loading matrix with many non-zero loadings, but no shrinkage and variable selection is incorporated during the parameter estimation and inference. The power of detecting non-zero cross-loading is thus reduced in the EFA. In comparison, parsimonious models with few spurious non-zero parameters are generally associated with better detection power and higher true discovery rates. Model parsimony may be achieved by model modification starting from the simplest model, for example, FBS-MI, or using shrinkage and variable selection, for example, the BSEM-SSP, starting from the most complex model.

While we did not show results beyond a sample sizes of 300, we can comment briefly on some general trends. Specifically, larger sample sizes and effect sizes would further increase the true positive rates and decrease the false positive rates of all methods. Under these circumstances, the ROC curves were pushed to the upper left corner, leaving very little discernible differences among them. The FPs of BSEM-SSP under the default threshold remains to be lower than those of BSEM-RP, FBS-MI, and EFA.

Comparison of Parameter Estimates—Similar results were obtained for both the “*Single Cross-Loading*” and “*Multiple Cross-Loading*” conditions with regard to estimates of the zero and non-zero cross-loadings in Λ (see Figure 2 for the “*Single Cross-Loading*” condition). For cross-loadings that were truly zero (see the first row in Figure 2), BSEM-SSP provided more efficient estimates that are closer to zero compared to BSEM-RP and FBS-MI due to the use of the spike and slab prior.

The BSEM-SSP provides more information concerning the non-zero cross-loadings under different model uncertainty scenarios compared to the FBS-MI. As shown in the second to the fourth rows in Figure 2, when the non-zero cross-loadings were small, estimates of FBS-MI were similar to the “shrinkage estimates” obtained using the marginal estimator computed using MCMC samples from the BSEM-SSP (see eq. (S12) in the supplemental materials), based on posterior means as averaged over the spike and slab components. As the magnitudes of the non-zero cross-loadings increased, the estimates from the FBS-MI

became similar to the conditional posterior mean estimates from the BSEM-SSP (see eq. (S11) in the supplemental materials).

In addition, the posterior means of the mixture component corresponding to $r = 1$ in the BSEM-SSP yielded cross-loading estimates that were comparable to the gold standards obtained from fitting the Bayesian CFA with correctly specified factor loading structure. Hence, unlike BSEM-RP, a refit is not required after cross-loading selection, leading to a one-step selection, estimation and inference procedure. Also, the marginal and conditional estimators of BSEM-SSP are more informative concerning the non-zero cross-loadings than the posterior mean or median of BSEM-RP.

We also compared the parameter estimates of Φ . We plotted the root mean square errors (RMSEs) of the main diagonal element ϕ_{11} and off diagonal elements ϕ_{12} and ϕ_{13} of Φ under the “*Single Cross-Loading*” and “*Multiple Cross-Loading*” conditions with noninformative prior in Figure 3. For the “*Single Cross-Loading*” condition, the performance of the BSEM-SSP is comparable to that of frequentist and Bayesian CFA with known factor loading structure. In comparison, the RMSEs given by the BSEM-RP grew with the true magnitude of the cross-loading, λ_{22} . This was because the prior $N(0, 0.01)$ imposed too strong of a prior – but in the wrong direction – when the cross-loadings were relatively large. The unexplained proportion λ_{22} was “absorbed into λ_{21} ” and led to estimation biases in λ_{21} , the first factor score, and related elements in Φ . These biases, which were not noted by Muthén and Asparouhov (2012), may be circumvented in part by their proposal to use a two-step approach in which the structure of the loading matrix is first determined using an informative but conservative prior in eq. (7), followed by a Bayesian CFA with noninformative prior for parameter estimation. However, even if such a two-stage procedure is used, the loading structure determined in the first step may be affected by the strong prior, and may contain false positive and negative elements as shown in Figure 1. In comparison, BSEM-SSP uses more flexible prior distributions for the cross-loadings and combines variable selection and parameter estimation in one step.

As the number of non-zero cross-loadings increased, so did the complexity of the variable selection problem for the “*Multiple Cross-Loading*” condition. In this context, some further strengths and limitations of the various approaches can be noted. BSEM-RP had better performance due to the correct informative prior distributions when the non-zero cross-loadings are small. This was validated by their smaller RMSEs near 0.1. Similarly, with increased magnitudes of the non-zero cross-loadings, the RMSEs from the BSEM-RP also increased due to increased discrepancies between the informative prior and the data; in contrast, the RMSEs from the BSEM-SSP were found to decrease as more information became available for the proposed algorithm to “learn” the correct loading structure. Generally, larger sample sizes (300) and effect sizes (0.2, 0.3) are required for BSEM-SSP to perform better than the BSEM-RP. In these scenarios, the BSEM-SSP outperformed the BSEM-RP and its RMSEs were close to, but slightly higher than, those from the frequentist and Bayesian CFA, reflecting in part the model selection uncertainty associated with the BSEM-SSP. The differences in RMSEs vanished, however, as stronger information became available. The RMSEs of ϕ_{22} , ϕ_{33} , and ϕ_{23} were found to follow a similar trend.

Results from Sensitivity Analyses—Inspection of the plots of EPSR values indicated that the two Bayesian variable selection approaches considered were not very sensitive to these specific variations in prior specification in eq. (8), and the parameter estimates were only slightly different between the three “informative prior” conditions. The means of the informative prior distributions used in Table 2 were equal to the true parameter values. Correct informative prior distributions in eq. (8) are helpful to improve the estimation efficiency and thus the detection power. If incorrect informative priors or other uninformative priors were used, the true positive rates and true negative rates would likely be reduced, unless the effect size and sample size considered were large enough. Using the results based on this informative prior distribution, we drew the ROC curves of BSEM-SSP and BSEM-RP similar to those in Figure 1, which lay closer to the upper left corner and represented higher true positive rates (not shown to conserve space). We emphasize that informative prior distributions should only be used if strong prior information exists, and sensitivity analysis should always be conducted whenever possible.

In evaluating the EPSR plots generated by varying the locations of the identification constraints, we also found that results from the two identification conditions converged to similar posterior distributions given sufficient MCMC iterations, indicating that the BSEM-SSP is not very sensitive to the choice of identification conditions if the cross-loadings fixed at zero are unimportant, namely, small and close to zero. In contrast, if important cross-loadings were set to zero for identification, the effects of these important cross-loadings may be expressed through other free cross-loadings, thus yielding biases. Similar sensitivity results were observed for the BSEM-RP. We suggest first applying the EFA to get a sense of viable unimportant cross-loadings to be fixed.

ROC curves from the BSEM-RP with different prior variance of the cross-loadings in the “*Complex Multiple Cross-loading Condition*” were shown Figure 4. Three observations can be noted. First, consistent with previous results in Muthén and Asparouhov (2012), the BSEM-RP was quite sensitive to the prior variance of the cross-loading. Prior variances 0.001 and 0.01 yielded better results for smaller and larger cross-loading size, respectively. Second, even though the BSEM-RP-VP, which assigned the additional prior distribution on the variance, provided more robust results compared to the other BSEM-RP results, the BSEM-RP-VP only outperformed the BSEM-RP with fixed variance at a sample size of 300, suggesting that larger sampler sizes were required for the BSEM-RP-VP to learn the variance adaptively. Third, the BSEM-SSP outperformed all the versions of the BSEM-RP and BSEM-RP-VP under conditions with larger sample and effect sizes.

Empirical Example

We illustrate the use of BSEM-SSP by building on a previously published structural equation modeling example in Byrne (2012, page 153) using a sample of 1430 teachers. The general model is shown in Figure 5. Six latent constructs, including *Role Ambiguity* (denoted in the path diagram as ROLEA), *Role Conflict* (ROLEC), *Work Overload* (WORK), *Decision Making* (DECI), *Superior Support* (SSUP), and *Peer Support* (PSUP), were each indicated by two parcels based on questions from the Teacher Stress Scale (TSS, Pettegrew & Wolf, 1982). Information concerning the specific items that were assigned to

each parcel was no longer available (B. M. Byrne, personal communication, Feb 15, 2016), but the parcels were constructed by assigning the questions to parcels according to content to equalize the measurement weighting across parcels (Byrne, 1994).

We reanalyzed the dataset using the BSEM-SSP approach and compared the results with those from a EFA and a CFA replicating the six-factor structure proposed by Byrne (1994). Both the EFA and CFA were conducted using Mplus with no correction for multiple comparison. In the CFA analysis, some redundancies were found among some factors: for instance, the estimated correlation between work overload and role conflict was 0.951, and the estimated correlation between decision making and superior support was 0.961. Some evidence for factor redundancies was also revealed in the EFA analysis. The estimated loading matrix from EFA with 6 factors using the default Geomin rotation in Mplus revealed that all loadings related to the 6th factor were insignificant.

Given some of the large inter-factor correlations and the insignificant factor loadings in EFA, we decided to reduce the number of factors. We encountered convergence problems in EFA with 5 factors. The 5th factor was only associated with one indicator and the factor was also strongly correlated with some other factors. We thus adopted the revised confirmatory structure shown in the right panel of Figure 5 with four factors. The revised model yielded the following fit statistics: RMSEA in [0.053, 0.071] (root mean square error of approximation, Steiger, 1990), SRMR = 0.014 (standardized root mean square residual, Jöreskog & Sörbom, 1996), TLI = 0.959 (Tucker-Lewis index, Tucker & Lewis, 1973), CFI = 0.985 (comparative fit index, Bentler, 1990), which indicated reasonable fit of the four factor model. Standardized estimates of Λ and Φ are shown in Table 3.

Building on the four-factor structure from the EFA, we used BSEM-SSP to further explore the cross-loading structure in this model. For identification purposes, we fixed q^2 elements in the loading matrix by setting q main-loadings to 1 and $q(q - 1)$ cross-loadings to zero, selected in descending order from the EFA estimates whose magnitudes were either close to zero, or were statistically insignificant.

To aid comparisons, the estimates of BSEM-SSP were standardized after estimation to make the variances of the factors equal to 1. Standardized estimates of Λ and Φ are shown in the right panel of Table 3, where the underlined elements were fixed for identification during model fitting. The posterior inclusion probabilities of the significant elements in EFA are shown in superscripts. For comparison, we also analyzed the data with BSEM-RP and FBS-MI. The standardized estimates of Λ and Φ of BSEM-RP and FBS-MI are shown in Table 3. The elements with asterisks were elements for which the associated 95% credible or confidence intervals did not include zero.

Compared to the BSEM-RP, the BSEM-SSP tended to identify cross-loadings that were greater than 0.1 (based on EFA) with less shrinkage, and shrank the others to be closer to zero. For example, the loading of the factor Role Ambiguity on the first parcel measuring *Role Conflict*, while estimated not to be different from zero in BSEM-RP, was identified in our approach with a posterior inclusion probability of 0.8 and was significant in EFA (0.13). Given the same threshold (0.05) as BSEM-RP, the FBS-MI identified more significant cross-

loadings than the substantial elements identified by BSEM-RP as well as the BSEM-SSP. One possible reason might be that both Bayesian variable selection approaches assigned shrinkage to the cross-loading estimates while the FBS-MI did not. With fewer free parameters, the FBS-MI identified more significant cross-loadings than the EFA, the magnitudes of which were generally small. Without correct multiple comparison correction, the significance of these cross-loadings remains in question.

Consistent with results from the simulation study, the estimated factor correlation matrix from the BSEM-SSP was closer to those from the methods without shrinkage, e.g., EFA estimates, than the BSEM-RP estimates. This suggests that the strong priors imposed in the BSEM-RP might have overshrunk some truly non-zero cross-loadings and forced them to exert their associations with the latent factors through other main or cross-loadings, as was found in our simulation study. The prior distributions for the loading matrix in the BSEM-SSP were less informative in contrast and exerted less influence on the estimation of the other parameters in the factor model than the BSEM-RP.

The latent constructs *Role Ambiguity* and *Role Conflict* were positively correlated as shown in Table 4. Additionally, the first parcel for *Role Conflict* may share other sources of covariation with *Role Ambiguity* that were not captured by the inter-factor correlation. EFA, FBS-MI and BSEM-SSP captured these connections with the cross-loading λ_{31} , the cross-loading of *Role Ambiguity* on the aforementioned parcel for *Role Conflict*. In comparison, BSEM-RP provided a smaller estimate of λ_{31} , and explained these connections as a higher correlation between *Role Ambiguity* and *Role Conflict/Work Overload*. Consequently, the latent variables in BSEM-SSP might represent more distinct latent constructs (i.e., with lower factor inter-correlations) compared to those in BSEM-RP.

The first parcel for *Decision Making* appeared to show complex cross-loading patterns that differed among the approaches. While all methods found substantial cross-loadings of *Role Ambiguity* and *Peer Support* on the first parcel for *Decision Making*, they yielded slightly different conclusions concerning λ_{72} , namely, the loading of *Role Conflict/Work Overload* on the first parcel for *Decision Making*. Specifically, the estimated values of λ_{72} from BSEM-RP, EFA and FBS-MI were small. In comparison, the BSEM-SSP simply shrank λ_{72} to zero and “assigned” the remaining variance to λ_{71} . These results help highlight the strengths of the BSEM-SSP in yielding more parsimonious and discernible factor structure (e.g., by setting the small λ_{72} to zero and also lowering the correlations among factors) compared to the BSEM-RP, EFA as well as the FBS-MI.

Discussion

We proposed a one-step Bayesian variable selection approach for model selection in confirmatory factor analysis using the spike and slab prior. The work was motivated by an earlier approach, BSEM-RP, proposed by Muthén and Asparouhov (2012) as an illustration of how different prior distributions might be used in common psychometric models, including hybrid EFA/CFA with flexible cross-loading structures. Extending Muthén and Asparouhov (2012)’s work, we compared BSEM-SSP with BSEM-RP, as well as three frequentist approaches to determining the loading structure. In doing so, we also

implemented and evaluated new extensions to the original BSEM-RP formulation, including modeling the prior variance, σ_{jk}^2 , as a parameter with its own prior in the BSEM-RP-VP approach, and adding the minimum constraints needed to identify a CFA to improve the numerical stability of the BSEM-RP approaches. Overall, the BSEM-SSP outperformed the other approaches considered when the sample size and effect size were relatively large, and yielded more parsimonious loading structures (by shrinking weaker cross-loadings to zero) in situations involving smaller sample and/or effect sizes.

As shown in the simulation study, FBS-MI detects more important cross-loading elements given all possible false positive levels than the other frequentist methods. In comparison, the BSEM-SSP is comparable to the FBS-MI when the sample sizes or the magnitudes of the cross-loadings are not small. However, the FBS-MI with the default p -value threshold is characterized by heightened false positive rates – a problem that actually aggravates as the sample size grows. In contrast, the BSEM-SSP with the default PIP threshold controls the false positive rates at very low levels, and the rates of detecting the correct loading structure increase with increased sample size and magnitudes of the cross-loading. Furthermore, the BSEM-SSP provides quantifiable measures of the uncertainties associated with cross-loading inclusion and model choice, and these uncertainties are already reflected in the parameter estimates and corresponding inferential results, thereby leading to more reproducible results.

EFA with target rotation performed less satisfactorily in identifying the important cross-loadings with larger sample sizes and larger magnitudes of nonzero cross-loadings. One possible way to improve the EFA with target rotation is to use it iteratively.¹¹ Specifically, if the estimated loading structure is different from the target structure, the structure containing all the main-loadings and the significant cross-loadings may be used as the new target matrix for another target rotation. This iterative procedure stops when the estimated loading matrix of EFA after rotation agrees with the target matrix. This approach is an extension to the standard target rotation approach considered in the present study, and its performance in selecting important cross-loadings compared to the BSEM-RP and BSEM-SSP warrants further investigation.

Even though many other frequentist model fit indices exist for model selection purposes (see Hu & Bentler, 1998, 1999, for reviews), we only considered three frequentist approaches, namely, the FBS-MI, FS, and EFA because MIs were shown to be effective in detecting model misspecification (Saris, Satorra, & van der Veld, 2009). Two other reasons add to the appeal of using variable selection approaches to select cross-loading structures. First, variable selection approaches can be used to broadly and efficiently select important cross-loadings to be freed up, as opposed to only comparing a number of candidate models. Second, most of these model comparison criteria are descriptive measures without known analytical distributions. Thus, it is difficult to quantify the uncertainty in variable selection.

As in the frequentist framework, many other model selection criteria also exist for latent variable models in the Bayesian framework, including but not limited to Bayesian

¹¹We thank the suggestion from a reviewer

information criterion (Schwarz, 1978), deviance information criterion (Spiegelhalter et al., 2002), and the L_V measure (Ibrahim, Chen, & Sinha, 2001; Y.-X. Li, Kano, Pan, & Song, 2012). A thorough comparison of the Bayesian variable selection approaches considered in this article to these alternative model comparison criteria is warranted in future studies.

Several other extensions of the current study may be pursued in the future. First, we only developed BSEM-SSP with continuous data. However, binary, categorical, and nominal data are frequently encountered in psychological, behavioral and medical studies. Extending the BSEM-SSP to handle latent variable models with discrete data is important. Second, Ψ was assumed to be a diagonal matrix in the examples considered in our study. Off-diagonal elements may be freed to provide a more realistic representation of empirical data. Third, the spike and slab prior is only used in factor analysis models. In SEMs, selection of the structural model and SEMs with nonlinear and interaction effects of latent variables are also important issues (Cham, West, Ma, & Aiken, 2012; Kenny & Judd, 1984). In Muthén and Asparouhov (2012), the ridge prior was also used to select the correlation elements in Ψ , and also to identify substantial nonlinear interaction effects between latent variables in nonlinear SEMs. Given the potential of the spike and slab prior shows in the current study, it will be interesting to investigate its performance in these situations, particularly in comparison to other commonly adopted approaches utilizing e.g., FBS-MI and model comparison criteria (Cai, Song, & Lee, 2008; Y.-X. Li et al., 2012; Song & Lee, 2006). Finally, the current simulation study is not intended to be exhaustive. There are other variables that may affect the results, including the number of indicators and factors, and the size of measurement error variances. Moreover, in our setting the true model is among the candidate models. It would be interesting to investigate the model selection performance under model misspecification.

As noted previously, the prior distributions in BSEM-SSP, BSEM-RP and Bayesian lasso can be formulated as mixture distributions. Even in the case of the BSEM-RP-VP, imposing a gamma prior on the variance of the normal distribution has the effect of dictating that the cross-loadings come from a mixture of distributions, the weights of which can be described by a gamma prior. This hierarchical normal mixture setting is widely used in Bayesian variable selection and shrinkage estimates (Gelman, 2006). Much research has been devoted to developing more efficient mixture distributions (Leng et al., 2014; Q. Li & Lin, 2010; Park & Casella, 2008), including using different prior distributions of the variance to represent different mixture weights. In comparison, the BSEM-SSP uses a discrete prior distribution for the mixture weight. All these methods are useful approaches of Bayesian variable selection, and are worthy of further studies in latent variable models.

Mixture approaches such as the BSEM-SSP are known to yield “by-products” that can be used to compute the Bayes factors (BF, Kass & Raftery, 1995). However, the SSP and BF are different in several aspects. Specifically, to calculate the BFs, each candidate model needs to be estimated separately through the MCMC algorithms, which may be time consuming. In comparison, the SSP applies the MCMC algorithm to the model which subsumes the candidate models as special cases. The marginal probabilities and the BFs of the most likely candidate models can be estimated all at once. The less likely models are eliminated automatically due to the MCMC algorithm.¹² The mixture prior formulation also

offers some unique advantages besides serving as a tool for computing the BF (for a review see Kamary, Mengersen, Robert, & Rousseau, 2014). First, the BF is based on a limited number of selected candidate models. In comparison, SSP considers all possible models at once. Second, even though the prior probabilities of the candidate models are important to the performance of BF, they are rarely studied and hard to elicit from prior knowledge. The SSP offers a more intuitive and flexible way to specify the prior probabilities through the prior distributions of the parameters. Third, BF cannot be used with non-informative prior distribution. Fourth, compared to the BF, which is typically used only in pairwise comparison in a categorical way (e.g., yielding conclusion such as Model 1 is preferred over Model 2) the posterior mixture probability offers additional information about the degree of support for a given model based on the data.

The two-step procedure of BSEM-RP suggested by Muthén and Asparouhov (2012) is characterized by a few additional problems which we did not resolve in the present study. For instance, the second step of the BSEM-RP is based on the same dataset used in the first step, which leads to two problems. First, the reuse of the same data may lead to overfitting in the second step. Second, as pointed out previously, the variation of the cross-loading selection in the first step of the BSEM-RP is not accounted for in the second step (estimation step), which leads to underestimation of the standard errors. One possible approach to improve the over-fitting problem is to use cross-validation, which splits the data and uses different data subsets in the two steps. Doing so may help identify redundant loadings in the second step, thus improving the robustness of the BSEM-RP. However, cross-validation does not resolve the second problem because splitting the data set into two portions still does not provide a way to quantify the uncertainty in Stage 1 of the BSEM-RP. In addition, the accuracy of model selection and parameter estimation is reduced because of the reduced sample sizes for estimation purposes. Thus, further studies are needed to examine the pros and cons of using cross-validation procedures to improve the robustness of the BSEM-RP and all Bayesian variable selection approaches in general.

Our simulation and empirical examples indicated that the BSEM-RP and BSEM-SSP yielded slightly different factor solutions. So, which of these solutions should one adopt? Our view is that the different prior choices represent different beliefs about the data at hand and no single prior distribution should be considered superior in all situations. Substantive researchers are advised to pick prior distributions that fit their prior beliefs concerning the data analytic problem at hand. Overall, framing the estimation of “hybrid” CFA models in the context of variable selection provides useful insight into the conceptual and computation issues involved. We see many potential future directions for this approach in factor analysis, along with extension to more general latent variable models.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

¹²We thank a reviewer for pointing out that mixture priors (like those in this paper) have been used to calculate marginal probabilities and BFs, and “trans-dimensional Markov Chains” (Green, 1995; Sisson, 2005) are used for technical computing. In comparison, our approach does not require “trans-dimensional Markov Chains”, which is computationally challenging.

Acknowledgments

Funding for this study was provided by NSF grant BCS-0826844, and NIH grant R01GM105004.

References

- Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. 1974; 19:716–723. DOI: 10.1109/TAC.1974.1100705
- Altman DG, Andersen PK. Bootstrap investigation of the stability of a cox regression model. *Statistics in Medicine*. 1989; 8:771–783. DOI: 10.1002/sim.4780080702 [PubMed: 2672226]
- Asparouhov T, Muthén BO. Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*. 2009; 16:397–438. DOI: 10.1080/10705510903008204
- Asparouhov T, Muthén BO, Morin AJS. Bayesian structural equation modeling with cross-loadings and residual covariances comments on stromeyer et al. *Journal of Management*. 2015; 0149206315591075. doi: 10.1177/0149206315591075
- Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin*. 1990; 107:238–246. DOI: 10.1037/0033-2909.107.2.238 [PubMed: 2320703]
- Berger JO, Barbieri MM. Optimal predictive model selection. *The Annals of Statistics*. 2004; 32:870–897. DOI: 10.1214/009053604000000238
- Bickel PJ, Li B, Tsybakov AB, Geer SAvd, Yu B, Valdés T, ... Vaart Avd. *Regularization in statistics*. Test. 2006; 15:271–344. DOI: 10.1007/BF02607055
- Browne MW. An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*. 2001; 36:111–150. DOI: 10.1207/S15327906MBR3601_05
- Byrne BM. Burnout: testing for the validity, replication, and invariance of causal structure across elementary, intermediate, and secondary teachers. *American Educational Research Journal*. 1994; 31:645–673. DOI: 10.3102/00028312031003645
- Byrne, BM. *Structural equation modeling with Mplus: basic concepts, applications, and programming*. New York, NY: Routledge, Taylor & Francis Group, LLC; 2012.
- Cai JH, Song XY, Lee SY. Bayesian analysis of nonlinear structural equation models with mixed continuous, ordered and unordered categorical, and nonignorable missing data. *Statistics and Its Interface*. 2008; 1:99–114. DOI: 10.4310/SII.2008.v1.n1.a9
- Cham H, West SG, Ma Y, Aiken LS. Estimating latent variable interactions with nonnormal observed data: a comparison of four approaches. *Multivariate Behavioral Research*. 2012; 47:840–876. DOI: 10.1080/00273171.2012.732901 [PubMed: 23457417]
- Chatfield C. Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*. 1995; 158:419–466. DOI: 10.2307/2983440
- Chatterjee, S.; Hadi, AS. *Regression analysis by example*. 5. Hoboken, NJ: John Wiley & Sons, Inc; 2012.
- Draper D. Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1995; 57(1):45–97. 01394. Retrieved August 27, 2015, from <http://www.jstor.org.ezaccess.libraries.psu.edu/stable/2346087>.
- Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. *The Annals of Statistics*. 2004; 32:407–499. DOI: 10.1214/009053604000000067
- Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*. 2010; 33:1–22. DOI: 10.18637/jss.v033.i01 [PubMed: 20808728]
- Gelman, A. Inference and monitoring convergence. In: Gilks, WR.; Richardson, S.; Spiegelhalter, DJ., editors. *Markov chain Monte Carlo in practice*. Boca Raton, FL: CRC; 1996. p. 131–143.
- Gelman A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*. 2006; 1:515–534. DOI: 10.1214/06-BA117A
- George EI, McCulloch RE. Variable selection via Gibbs sampling. *Journal of the American Statistical Association*. 1993; 88:881–889. DOI: 10.1080/01621459.1993.10476353

- Green PJ. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*. 1995; 82:711–732. DOI: 10.1093/biomet/82.4.711
- Hallquist, M.; Wiley, J. MplusAutomation: automating Mplus model estimation and interpretation. R package version 0.6-2. 2014. Retrieved from <http://CRAN.R-project.org/package=MplusAutomation>
- Hancock GR. A sequential Scheffé-type respecification procedure for controlling type I error in exploratory structural equation model modification. *Structural Equation Modeling: A Multidisciplinary Journal*. 1999; 6:158–168. DOI: 10.1080/10705519909540126
- Harrell, FE. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York, NY: Springer-Verlag New York, Inc; 2001.
- Hastie, T.; Tibshirani, R.; Friedman, J.; Hastie, T.; Friedman, J.; Tibshirani, R. The elements of statistical learning. 2. New York, NY: Springer Science+Business Media, LCC; 2009.
- Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*. 1970; 12:55–67. DOI: 10.1080/00401706.1970.10488634
- Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian model averaging: a tutorial. *Statistical Science*. 1999; 14:382–417. DOI: 10.1214/ss/1009212519
- Hu, L-T.; Bentler, PM. Structural equation modeling: concepts, issues, and applications. Thousand Oaks, CA, USA: Sage Publications, Inc; 1995. Evaluating model fit; p. 76-99.
- Hu LT, Bentler PM. Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. *Psychological Methods*. 1998; 3:424–453. DOI: 10.1037/1082-989X.3.4.424
- Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*. 1999; 6:1–55. DOI: 10.1080/10705519909540118
- Ibrahim JG, Chen M-H, Sinha D. Criterion-based methods for Bayesian model assessment. *Statistica Sinica*. 2001; 11(2):419–444.
- Ishwaran H, Rao JS. Spike and slab variable selection: frequentist and Bayesian strategies. *The Annals of Statistics*. 2005; 33:730–773. DOI: 10.1214/009053604000001147
- Jennrich RI, Sampson PF. Rotation for simple loadings. *Psychometrika*. 1966; 31:313–323. DOI: 10.1007/BF02289465 [PubMed: 5221128]
- Jöreskog KG. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*. 1969; 34:183–202. DOI: 10.1007/BF02289343
- Jöreskog, KG.; Sörbom, D. LISREL 8: user's reference guide. Lincolnwood, IL: Scientific Software International, Inc; 1996.
- Kamary, K.; Mengersen, K.; Robert, CP.; Rousseau, J. Testing hypotheses via a mixture estimation model. 2014. *arXiv:1412.2044 [stat]*. 00004. arXiv: 1412.2044. Retrieved August 22, 2015, from <http://arxiv.org/abs/1412.2044>
- Kano Y. Variable selection for structural models. *Journal of Statistical Planning and Inference*. 2002; 108:173–187. C.R. Rao 80th Birthday Felicitation Volume, Part II. DOI: 10.1016/S0378-3758(02)00277-X
- Kass RE, Raftery AE. Bayes factors. *Journal of the American Statistical Association*. 1995; 90:773–795. DOI: 10.1080/01621459.1995.10476572
- Kenny DA, Judd CM. Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*. 1984; 96:201–210. DOI: 10.1037/0033-2909.96.1.201
- Kyung M, Gill J, Ghosh M, Casella G. Penalized regression, standard errors, and Bayesian lassos. *Bayesian Analysis*. 2010; 5:369–411. DOI: 10.1214/10-BA607
- Lee, S-Y. Structural equation modeling: a Bayesian approach. Chichester, West Sussex, England: John Wiley & Sons; 2007.
- Leng C, Tran MN, Nott D. Bayesian adaptive lasso. *Annals of the Institute of Statistical Mathematics*. 2014; 66:221–244. DOI: 10.1007/s10463-013-0429-6
- Li Q, Lin N. The Bayesian elastic net. *Bayesian Analysis*. 2010; 5:151–170. 00097. DOI: 10.1214/10-BA506

- Li YX, Kano Y, Pan JH, Song XY. A criterion-based model comparison statistic for structural equation models with heterogeneous data. *Journal of Multivariate Analysis*. 2012; 112:92–107. DOI: 10.1016/j.jmva.2012.05.010
- MacCallum RC, Roznowski M, Necowitz LB. Model modifications in covariance structure analysis: the problem of capitalization on chance. *Psychological Bulletin*. 1992; 111:490–504. DOI: 10.1037/0033-2909.111.3.490 [PubMed: 16250105]
- Mavridis D, Ntzoufras I. Stochastic search item selection for factor analytic models. *British Journal of Mathematical and Statistical Psychology*. 2014; 67:284–303. DOI: 10.1111/bmsp.12019 [PubMed: 23837882]
- Mitchell TJ, Beauchamp JJ. Bayesian variable selection in linear regression. *Journal of the American Statistical Association*. 1988; 83:1023–1032. DOI: 10.1080/01621459.1988.10478694
- Muthén BO, Asparouhov T. Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychological Methods*. 2012; 17:313–335. DOI: 10.1037/a0026802 [PubMed: 22962886]
- O'Hara RB, Sillanpää MJ. A review of Bayesian variable selection methods: what, how and which. *Bayesian Analysis*. 2009; 4:85–117. DOI: 10.1214/09-BA403
- Pang, X.; Gill, J. Spike and slab prior distributions for simultaneous Bayesian hypothesis testing, model selection, and prediction, of nonlinear outcomes. Washington University in St. Louis; St. Louis, MO: 2009. Unpublished manuscript Retrieved from <http://polmeth.wustl.edu/node/303>
- Park T, Casella G. The Bayesian lasso. *Journal of the American Statistical Association*. 2008; 103:681–686. DOI: 10.1198/016214508000000337
- Pettegrew LS, Wolf GE. Validating measures of teacher stress. *American Educational Research Journal*. 1982; 19:373–396. DOI: 10.3102/00028312019003373
- Polson, NG.; Scott, JG. Shrink globally, act locally: sparse Bayesian regularization and prediction. In: Bernardo, JM.; Bayarri, MJ.; Berger, JO.; Dawid, AP.; Heckerman, D.; Smith, AFM.; West, M., editors. *Bayesian statistics*. Vol. 9. Oxford, UK: Oxford University Press; 2011. p. 501-538.
- R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013. Retrieved from <http://www.R-project.org/>
- Saris WE, Satorra A, van der Veld WM. Testing structural equation models or detection of misspecifications? *Structural Equation Modeling: A Multidisciplinary Journal*. 2009; 16:561–582. DOI: 10.1080/10705510903203433
- Schwarz G. Estimating the dimension of a model. *The Annals of Statistics*. 1978; 6:461–464. DOI: 10.1214/aos/1176344136
- Sisson SA. Transdimensional Markov chains: a decade of progress and future perspectives. *Journal of the American Statistical Association*. 2005; 100:1077–1089. DOI: 10.1198/016214505000000664
- Song XY, Lee SY. A maximum likelihood approach for multisample nonlinear structural equation models with missing continuous and dichotomous data. *Structural Equation Modeling: A Multidisciplinary Journal*. 2006; 13:325–351. DOI: 10.1207/s15328007sem1303_1
- Sörbom D. Model modification. *Psychometrika*. 1989; 54:371–384. DOI: 10.1007/BF02294623
- Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2002; 64:583–639. DOI: 10.1111/1467-9868.00353
- Steiger JH. Structural model evaluation and modification: an interval estimation approach. *Multivariate Behavioral Research*. 1990; 25:173–180. DOI: 10.1207/s15327906mbr2502_4 [PubMed: 26794479]
- Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1996; 58(1):267–288. Retrieved July 31, 2014, from <http://www.jstor.org/stable/2346178>.
- Tucker LR, Lewis C. A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*. 1973; 38:1–10. DOI: 10.1007/BF02291170
- Zou H. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*. 2006; 101:1418–1429. DOI: 10.1198/016214506000000735

Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2005; 67:301–320. 03906. DOI: 10.1111/j.1467-9868.2005.00503.x

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.

ROC curves for the “*Single Cross-Loading Condition*” (top) and “*Multiple Cross-Loading Condition*” (bottom) with the noninformative prior setting. The caption of each figure states the sample size and the value of the non-zero cross-loadings. BSEM-SSP: Bayesian structural equation modeling with the spike and slab prior in eq. (10); FBS-MI: forward-backward stepwise model modification procedure; BSEM-RP: Bayesian structural equation modeling with the ridge prior in eq. (7); EFA: exploratory factor analysis with target rotation; Hancock: forward selection procedure with a Scheffé-type test. The square, circle, triangle, diamond and asterisk on the ROC curves represent the true and false positive rates obtained under the default thresholds of the methods. The unfilled and solid symbols correspond to the thresholds with and without Bonferroni correction.

Figure 2.

Boxplots of estimates of nonzero cross-loadings based on the 100 samples generated in the “*Single Cross-Loading Condition*” with the noninformative prior setting. BCFA: Bayesian CFA with known factor loading structure; CFA: frequentist CFA with known factor loading structure; FBS: MLE of the final model from the forward-backward stepwise model modification procedure; RP: posterior mean of BSEM-RP; SSPr1: the conditional estimator of BSEM-SSP; SSP: the marginal estimator of BSEM-SSP; EFA: MLE estimates of EFA with target rotation. The figures in the first, second, and third columns are based on samples with sample size 100, 200, and 300, respectively. The figures from the first to the fourth rows correspond to the samples with cross-loadings 0.0, 0.1, 0.2, and 0.3, respectively.

Figure 3.

Plots of root mean square error (RMSE) of λ_1 , λ_2 , and λ_3 , which are shown in column 1, 2, and 3, respectively. BSEM-SSP: Bayesian structural equation modeling with the spike and slab prior in eq. (10); BSEM-RP: Bayesian structural equation modeling with the ridge prior in eq. (7); CFA: frequentist CFA with known factor loading structure; BCFA: Bayesian CFA with known factor loading structure. The first two rows are based on “*Single Cross-Loading Condition*” with the noninformative prior setting, while the last two rows are based on “*Multiple Cross-Loading Condition*” with the noninformative prior setting. The caption of each figure states the sample size and the element in Φ .

Figure 4.

ROC curves for the “*Complex Multiple Cross-loading Condition*” with 24 measurement variables and 8 factors. BSEM-SSP: BSEM with the spike and slab prior in eq. (10); BSEM-RP: BSEM with the ridge prior in eq. (7) and fixed prior variance; BSEM-RP-VP: BSEM with the ridge prior in eq. (7) and random prior variance with gamma prior distribution. From top to bottom are cases with sample size 100, 200, and 300. From left to right are cases with non-zero cross-loadings 0.1, 0.2, and 0.3. The symbols on the ROC curves represent the true and false positive rates obtained under the default thresholds of the methods. The unfilled and solid symbols correspond to the thresholds with and without Bonferroni correction.

Figure 5.

Path diagrams of the factor models. ROLEA1–2: parcels measuring *Role Ambiguity*; ROLEC1–2: parcels measuring *Role Conflict*; WORK1–2: parcels measuring *Work Overload*; DECI1–2: parcels measuring *Decision Making*; SSUP1–2: parcels measuring *Superior Support*; PSUP1–2: parcels measuring *Peer Support*. The left panel shows the CFA model illustrated in Byrne (2012). The right panel is a revised 4-factor model based on results from an EFA.

Table 1

A table summarizing the acronyms used in the paper.

Acronyms	Meaning
BF	Bayes factor
BSEM	Bayesian structural equation modeling
BSEM-RP	BSEM with ridge regression prior (Muthén & Asparouhov, 2012)
BSEM-RP-VP	BSEM-RP using variance with prior
BSEM-SSP	BSEM with spike and slab prior
CFA	confirmatory factor analysis
CFI	comparative fit index
EFA	exploratory factor analysis
EPSR	estimated potential scale reduction (Gelman, 1996)
FPS	false positives
FS	forward selection with a Scheffé-type test (Hancock, 1999)
MCMC	Markov chain Monte Carlo
MI	modification indices
MLE	maximum likelihood estimator
ROC	receiver operating characteristic
RMSEA	root mean square error of approximation
SRMR	standardized root mean square residual
TLI	Tucker-Lewis index
TPs	true positives

Table 2

Comparisons of the number of replications where the structure of the loading matrix was estimated correctly for BSEM-SSP, FBS-MI and BSEM-RP. The approach with the largest number of true positives is highlighted.

Scenarios	Prior	BSEM-SSP			FBS-MI			BSEM-RP			
		0.1	0.2	0.3	0.1	0.2	0.3	0.1	0.2	0.3	
1	1	100	1	5	34	9	32	47	0	1	6
		200	2	26	75	16	48	54	0	9	48
		300	4	55	95	17	34	38	0	30	74
	2	100	1	11	50	9	32	47	0	1	7
		200	4	30	78	16	48	54	0	11	48
		300	4	59	94	17	34	38	0	32	77
2	1	100	0	0	5	0	11	41	0	0	0
		200	0	5	54	3	34	52	0	1	18
		300	0	18	77	6	39	45	0	2	19
	2	100	0	0	22	0	11	41	0	0	0
		200	0	12	69	3	34	52	0	2	28
		300	0	29	87	6	39	45	0	4	28

Note. BSEM-SSP: Bayesian structural equation modeling with the ridge prior in eq. (7). The approach with the largest number of true positives is highlighted. Scenarios 1 and 2 correspond to the single and multiple cross-loading conditions. Prior settings 1 and 2 represent the non-informative and informative prior distributions, the means of which are equal to the true parameter values.

Table 3

The estimated loading matrices obtained via the EFA, FBS-MI, BSEM-RP and BSEM-SSP.

	Ambi	RCWO	DMSS	Peer	Ambi	RCWO	DMSS	Peer
Estimated loading matrix								
EFA					FBS-MI			
ROLEA1	0.650	0.026	0.009	-0.058	0.690	0.000	0.000	0.000
ROLEA2	0.827	0.003	-0.020	0.021	0.876	0.000	0.000	0.087*
ROLEC1	0.130*	0.516	-0.073*	-0.040	0.188*	0.486	-0.079*	0.000
ROLEC2	0.075	0.728	-0.026	0.022	0.119*	0.705	0.000	0.000
WORK1	-0.079	0.821	0.008	-0.008	0.000	0.761	0.000	0.000
WORK2	0.013	0.651	0.038	0.004	0.000	0.644	0.000	0.000
DECI1	-0.254*	-0.126*	0.305	0.194*	-0.274*	-0.109*	0.284	0.191*
DECI2	0.001	-0.007	0.787	0.037	0.000	0.000	0.807	0.000
SSUP1	-0.112*	0.011	0.816	-0.030	-0.074*	0.000	0.811	0.000
SSUP2	0.046*	-0.006	0.964	0.004	0.070*	0.000	0.985	0.000
PSUP1	-0.049	-0.016	-0.029	0.774	0.000	0.000	-0.145*	0.954
PSUP2	0.025	0.017	0.033	0.915	0.000	0.000	0.000	0.821
BSEM-RP					BSEM-SSP			
ROLEA1	0.735	-0.025	0.017	-0.010	0.696	0.000	0.000	0.000
ROLEA2	0.842	-0.006	-0.012	0.053	0.809	0.000	0.000	0.000
ROLEC1	0.078	0.575	-0.043	-0.033	0.160 ^(.80)	0.545	-0.012 ^(.11)	-0.002
ROLEC2	0.029	0.778	-0.009	0.035	0.003	0.806	0.000	0.000
WORK1	-0.100	0.843	0.014	0.007	-0.003	0.743	0.000	0.000
WORK2	-0.012	0.671	0.030	0.022	0.000	0.637	0.000	0.000
DECI1	-0.212*	-0.112*	0.346	0.167*	-0.355 ^(1.0)	-0.005 ^(.05)	0.303	0.190 ^(1.0)
DECI2	0.038	0.010	0.831	0.013	0.000	0.000	0.811	0.000
SSUP1	-0.059*	0.020	0.861	-0.050	-0.047 ^(.42)	0.000	0.833	0.000
SSUP2	0.090	0.015	1.012	-0.019	0.073 ^(.59)	0.000	0.990	0.000
PSUP1	-0.017	-0.007	-0.025	0.801	0.000	0.000	-0.007	0.825

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

	Ambi	RCWO	DMSS	Peer	Ambi	RCWO	DMSS	Peer
PSUP2	0.045	0.027	0.024	0.931	0.000	0.000	0.000	0.882

Note. ROLEA1–2: parcels measuring *Role Ambiguity*; ROLEC1–2: parcels measuring *Role Conflict*; WORK1–2: parcels measuring *Work Overload*; DECI1–2: parcels measuring *Decision Making*; SSUP1–2: parcels measuring *Superior Support*; PSUP1–2: parcels measuring *Peer Support*; Ambi: latent factor *Role Ambiguity*; RCWO: latent factor *Role Conflict and Work Overload*; DMSS: latent factor *Decision Making and Superior Support*; Peer: latent factor *Peer Support*; EFA: exploratory factor analysis with the default Geomin rotation; FBS-MI: forward-backward stepwise model modification procedure; BSEM-RP: Bayesian structural equation modeling with the ridge prior in eq. (7); BSEM-SSP: Bayesian structural equation modeling with the spike and slab prior in eq. (10). The main-loadings are bolded. The elements with asterisks were elements for which the associated 95% credible or confidence intervals did not include zero.

Table 4

The estimated standardized covariance matrices obtained via the EFA, FBS-MI, BSEM-RP and BSEM-SSP.

	Ambi	RCWO	DMSS	Peer	Ambi	RCWO	DMSS	Peer
Estimated standardized covariance matrix								
EFA								
RCWO	0.693				0.668			
DMSS	-0.535	-0.492			-0.584	-0.485		
Peer	-0.467	-0.379	0.469		-0.564	-0.393	0.546	
BSEM-RP								
RCWO	0.744				0.727			
DMSS	-0.601	-0.556			-0.582	-0.519		
Peer	-0.536	-0.434	0.518		-0.495	-0.388	0.492	
BSEM-SSP								
RCWO	0.744				0.727			
DMSS	-0.601	-0.556			-0.582	-0.519		
Peer	-0.536	-0.434	0.518		-0.495	-0.388	0.492	

Note. Ambi: latent factor *Role Ambiguity*; RCWO: latent factor *Role Conflict and Work Overload*; DMSS: latent factor *Decision Making and Superior Support*; Peer: latent factor *Peer Support*. EFA: exploratory factor analysis with the default Geomin rotation; FBS-MI: forward-backward stepwise model modification procedure; BSEM-RP: Bayesian structural equation modeling with the ridge prior in eq. (7); BSEM-SSP: Bayesian structural equation modeling with the spike and slab prior in eq. (10). The main-loadings are bolded. The elements marked with an asterisk are significant or substantial cross-loadings.