

An Alternative to Post Hoc Model Modification in Confirmatory Factor Analysis: The Bayesian Lasso

Junhao Pan
Sun Yat-sen University

Edward Haksing Ip
Wake Forest University

Laurette Dubé
McGill University

Abstract

As a commonly used tool for operationalizing measurement models, confirmatory factor analysis (CFA) requires strong assumptions that can lead to a poor fit of the model to real data. The post hoc modification model approach attempts to improve CFA fit through the use of modification indexes for identifying significant correlated residual error terms. We analyzed a 28-item emotion measure collected for $n = 175$ participants. The post hoc modification approach indicated that 90 item-pair errors were significantly correlated, which demonstrated the challenge in using a modification index, as the error terms must be individually modified as a sequence. Additionally, the post hoc modification approach cannot guarantee a positive definite covariance matrix for the error terms. We propose a method that enables the entire inverse residual covariance matrix to be modeled as a sparse positive definite matrix that contains only a few off-diagonal elements bounded away from zero. This method circumvents the problem of having to handle correlated residual terms sequentially. By assigning a Lasso prior to the inverse covariance matrix, this Bayesian method achieves model parsimony as well as an identifiable model. Both simulated and real data sets were analyzed to evaluate the validity, robustness, and practical usefulness of the proposed procedure.

Translational Abstract

Confirmatory factor analysis (CFA) is a commonly used tool by psychologists to determine the underlying dimensions of a survey. The “Big Five,” which were 5 key traits that were thought to characterize one’s personality, were identified using CFA. CFA requires strong model assumptions, which sometimes are not realistic because the restrictions could lead to poor fit between real data and the model. Existing literature tends to use ad hoc methods to fix the problem. Instead of using a patching-up approach, this study describes a scientifically justifiable way to relax model assumptions. Based on sound statistical principles, the proposed method adopts analytic techniques from the machine learning field to solve the potentially complex interrelatedness problem between survey items. By using both simulation experiments and a real data set regarding emotion measurement, it is shown that the method can be used to help researchers implement CFA in more flexible ways.

Keywords: confirmatory factor analysis, post hoc model modification, Lasso prior, Bayesian method

Supplemental materials: <http://dx.doi.org/10.1037/met0000112.supp>

Confirmatory factor analysis (CFA) is a commonly used technique for studying theory-driven hypotheses regarding observed and unobserved variables. CFA has been extensively applied to research areas in psychology and the social sciences including personality assessment (Briggs & Cheek, 1986; Marsh & Hocevar, 1985), organizational studies (Stone-Romero, Weaver, & Glenar, 1995), and quality of life research (Hays, Anderson, & Revicki,

1993). Using CFA, the number of factors is assumed to be known a priori and is thus treated as a fixed number in the estimation procedure. Unlike exploratory factor analysis (EFA), CFA requires the observed variables to have zero loadings on all factors except the appropriate ones, which are a priori specified in the model. The increasing use of CFA is partly due to its relative ease to set up, its robustness to mild deviations from the underlying assumptions

Junhao Pan, Department of Psychology, Sun Yat-sen University; Edward Haksing Ip, School of Medicine, Wake Forest University; Laurette Dubé, Desautels Faculty of Management, McGill University.

This work was supported by the National Natural Science Foundation of China (NSFC 11301555, PI: Junhao Pan) and the China Scholarship Council, and the National Science Foundation (SES-1424875, PI: Ed-

ward Haksing Ip) and the Wake Forest Clinical and Translational Science Institute through a grant from the National Institute of Health (UL1TR001420).

Correspondence concerning this article should be addressed to Edward Haksing Ip, School of Medicine, Wake Forest University, Medical Center Boulevard, WC23, NC27157, Winston Salem, NC. E-mail: eip@wakehealth.edu

(e.g., Curran, West, & Finch, 1996; Flora & Curran, 2004; Hu, Bentler, & Kano, 1992), and its important role in structural equation modeling (SEM; see Brown, 2006, p. xi; Thompson, 2004, p. 110), which has undergone rapid growth as a tool for studying causal relationships (Hershberger, 2003). For a recent review of CFA, see DiStefano and Hess (2005).

Researchers using CFA for various purposes often encounter a dilemma. Although in many ways theory-based CFA is more compelling than its counterpart of more exploratory work, sometimes the theory being tested simply does not fit the data well. Using the Big Five Personality data, Muthén and Asparouhov (2012) demonstrated that even in well-established theory, CFA may not fit well and possible correlated residuals may exist among items. When such a scenario occurs, it is tempting for researchers to revert to EFA, compare the results derived from the two different approaches, and make changes for a separate round of CFA exercise. For example, items that were found to cross-load on more than one factor in EFA are considered for deletion in the subsequent analysis. Recent work in Bayesian factor analysis proposed using different priors such as ridge regression prior or spike and slab prior to shrink cross-loadings in EFA toward zero (Muthén & Asparouhov, 2012; Lu, Chow, & Loken, 2016). Another possibility is to use modification indexes (MIs) for identifying components in the model that could be tweaked for the purpose of improving overall goodness-of-fit. Typically, a constrained parameter is modified to become a free parameter and if the corresponding MI suggests that the improvement of model fit is significant, the model with the freed parameter will be used. This approach, sometimes known as post hoc model modification (PMM, Kaplan, 1990; Sörbom, 1989), has been popularized by software programs such as LISREL (Jöreskog & Sörbom, 1984) and Mplus (Muthén & Muthén, 1998–2013), both of which allow users to examine MIs and make changes to a number of parameters. Although critics of the method views MI as either a form of data snooping or a devious departure from the theory-driven paradigm (Steiger, 1990), there are several advantages to the modification approach from a practical point of view. First, CFA is a restrictive tool because hypotheses are rarely available that allow the specification of a simple factor pattern that agrees with real data. Second, the standard for acceptance for CFA models, such as criterion of percentage of variance accounted for, is higher than the standard of acceptance for EFA models (Bentler & Bonett, 1980). Such a high standard cannot always be achieved in practice. Lastly, when used appropriately, the PMM approach may pinpoint pockets of misspecification in the model, which could lead to modification and a more robust model (MacCallum, 1995; Sörbom, 1989).

One problem in using the PMM methodology is that the related statistical hypothesis testing can only be sequentially applied to nested models. Suppose Model A containing a fixed parameter $\psi_1 = 0$ is nested within Model B, which contains the freed parameter of ψ_1 . A chi-squared statistic of one degree of freedom is used to evaluate whether the improvement due to the change of freeing the ψ_1 parameter is statistically significant. If another parameter ψ_2 needs to be tested, it must be tested after the testing of parameter ψ_1 . Although the sequential testing procedure in PMM using MI can be used to modify a CFA model, the post hoc procedure cannot simultaneously identify locations where modifications are needed. In other words, PMM may lead to a suboptimal solution. MacCallum (1986) examined specification searches and

found methods such as MI tend to produce incorrect results. A second problem with PMM is that the procedure may not be practical when the number of items is large. For a CFA on 30 items, even when a small percentage, say 5%, of covariances deviate from the diagonal matrix assumption, there would be 22 flagged covariances. This is complicated by the possibility of false discovery; how PMM would control for spurious significant findings is far from clear. Additionally, there is no guarantee that the modified covariance matrix is positive definite, a critical requirement for the model based on the estimated parameters to form a proper joint density. We will return to these issues in more depth in the Discussion section.

In this article we adopt the view that CFA is a modeling effort in which (a) the researcher has a theory-based factor structure in mind for the measures in the traditional CFA sense, and (b) the residual covariance structure in the error terms across all variables is not necessarily diagonal, but only a few of the off-diagonal elements are bounded away from zero. In particular, these covariance terms are treated as nuisance parameters—deviations from the theory-driven factor model that need to be accounted for but otherwise of little substantive interest. In other words, we do not adjust a CFA model post hoc but rather start with a comprehensive CFA model that allows for correlated residual errors between measures after accounting for the latent factors. Using a full correlated-error model has its own problems that the covariance matrix for the error terms can be quite large. Using the aforementioned example of CFA for 30 measures, the 30×30 covariance matrix contains 435 distinct off-diagonal entries. We propose the use of a highly sparse representation of the covariance matrix. The technical implementation of such an idea is through the Bayesian Lasso (least absolute shrinkage and selection operator) approach, in which the entire residual covariance matrix for all observed measures are modeled as a sparse structure that contains only a few covariance entries bounded away from zero.

Consider a researcher who uses SEM for a substantive analysis that requires CFA on a 20-item response data set. Suppose that although the three-factor model is an interpretable model and consistent with the theory, the four-factor model offers better fit in terms of goodness-of-fit indexes such as CFI and RMSEA. Additionally, further examining the three-factor model result reveals that 5%, or a 10-residual correlation, still lingers in the covariance matrix. Suppose the researcher prefers the three-factor solution but has concerns using the diagonal three-factor structure because the CFA result could be biased, and perhaps more important, that the bias could percolate to the other SEM components. In such a scenario, what can the researcher do? S/he can dismiss the error covariances and defend the solution using theory, or S/he can use PMM whose limitations we have discussed. An alternative is to use the Bayesian Lasso. The greatest practical benefits that we see for the Bayesian Lasso in this situation include (a) the ability to retain a more parsimonious factor structure that could be more consistent with the theory by treating off-diagonal errors as a nuisance factor, and (b) the use of a defensible statistical method for handling the limited and yet non-negligible number of error covariances. In contrast to PMM, the Lasso approach avoids the form of p value hacking inherent in PMM. In other words, truly zero covariance that shows up as significant due to sampling error is likely to be shrunk to near zero, and truly nonzero covariance will likely be detected and adjusted for in the CFA analysis. As a

result, the researcher can specify a theory-based confirmatory factor-structure and adjust the model using a data-driven process instead of tediously modifying one parameter at a time.

The third practical benefit of the Bayesian Lasso method is that it can prevent or at least reduce the propagation of bias in CFA to SEM estimates. In a study using simulated data that we provide later, we demonstrate this effect using a simple SEM example. In addition to the above-mentioned practical benefits, the Bayesian Lasso also enjoys many technical benefits, compared with other methods including PMM. First, it is a formal statistical model that maintains the positive definiteness property of the covariance matrix. Second, unlike a CFA model with a fully specified error covariance under a frequentist framework, the Bayesian Lasso model is identifiable. Third, as we shall see, the Bayesian Lasso can easily incorporate an additional step of imputation in handling missing values. This could be very useful for data sets that contain partially missing values—for example, when participants miss a response on some items but not all. Fourth, the Bayesian Lasso model generally fits the data better. By limiting the denseness of the covariance in terms of deviation from a diagonal form, the Bayesian Lasso is able to achieve an optimal tradeoff between overall fit and model complexity. Finally, the Bayesian Lasso can be used as a diagnostic tool for affirming or refuting theory-based factor configurations. As we shall see from the simulation experiments, specific patterns in the residual covariance matrix often signal certain forms of misspecification. An excessive amount of covariance error terms would suggest that the CFA is not fitting well and that the researcher should investigate—this is an advantage that a global goodness-of-fit index would not offer.

The Bayesian Lasso factor model proposed in this article can be viewed as a specific implementation of Bayesian structural equation modeling (SEM) as discussed in Muthén and Asparouhov (2012). The authors argued that in many SEM applications, unnecessary restrictions to models are applied to represent hypotheses derived from substantive theories. As hypotheses are reflected in parameters fixed at zero, they proposed to replace parameters of exact zeros with approximate zeros via a Bayesian analysis using small variance priors. For the Bayesian Lasso factor model, limiting the numbers of nonzero (or not-so-close-to-zero) parameters is especially important as we focus on the residual covariance matrix and the number of parameters that grow quickly with the dimension of the problem. Without the Lasso component, a Bayesian approach for handling residual covariance would be to free the covariance parameters and assign prior distributions to the covariance entries. However, such a generic Bayesian approach would not necessarily lead to a sparse solution. As previously pointed out, Muthén and Asparouhov (2012) demonstrated the use of a ridge-regression prior for exploring the presence of cross-loadings in CFA as an example of the Bayesian SEM approach. The Bayesian Lasso in this article is a special implementation of the Bayesian SEM for CFA covariance structures, not for cross-loading. The current approach is also analogous to some local dependent models in item response theory (IRT; e.g., Ip et al., 2004). Similar to local dependent item response models, the Bayesian Lasso approach treats the set of nonzero entries in the error covariance matrix as a means of achieving a better goodness-of-fit, but is otherwise only of peripheral interest.

The remainder of the article is organized as follows: First, we provide an overview of the Bayesian and Lasso approaches. We then describe the CFA model with Bayesian covariance Lasso prior and outline the estimation procedure. Next we describe several simulation studies for evaluating the proposed procedure. An analysis of a real data set collected from a study on emotion is then presented. Finally, we provide a discussion.

The Bayesian Covariance Lasso Prior CFA Model

Bayesian Analysis

The Bayesian approach is well recognized in the statistics and psychology literature as an attractive way of analyzing a wide variety of models (see, e.g., Congdon, 2005, 2007, 2014, and the references therein). The most attractive features of the Bayesian approach in statistical modeling and data analysis are as follows (Gelman et al., 2004; Congdon, 2007). First, it allows the use of genuine prior information in addition to the information available in observed data to produce results. Second, sampling-based Bayesian methods rely less on large-sample asymptotic theory, thereby producing reliable results even with relatively small sample sizes. Third, it provides better statistics for goodness-of-fit and model comparison as well as other useful statistics; for example, point estimates and interval estimates on parameters and probability values on hypotheses. Lastly, compared with maximum likelihood (ML) based methods, the Bayesian approach is conceptually simple and can be implemented in complex models that are more difficult to fit using classical methods.

A typical Bayesian analysis can be outlined in the following steps:

1. Define a model M with unknown parameter θ .
2. Assign prior distribution of θ under model M , $p(\theta|M)$, which reflects the prior information from some sources, such as the knowledge of experts, analyses of similar data or past data, and so on.
3. Construct the likelihood function $p(y|\theta, M)$ based on the observed data y and the model M defined in Step 1.
4. Determine the posterior distribution of θ by incorporating the sample information in $p(y|\theta, M)$ and prior information in $p(\theta|M)$ through $p(\theta|y, M) \propto p(y|\theta, M) \times p(\theta|M)$, where $p(\theta|y, M)$ is the posterior of θ .
5. Calculate the quantities of interest for θ (e.g., point estimates and interval estimates) based on the posterior distribution of θ in Step 4.

Lasso Model

Regularization is a general approach that requires additional model information for improving a solution that is inherently unstable or too complex for meaningful interpretation. In frequentist framework, the Lasso approach implements regularization by adding a penalty term to the usual likelihood so that the model would move toward a solution that contains fewer parameters. In general, frequentist Lasso tends to produce some coefficients that

are exactly zero, which not only helps to increase predictive accuracy, but also enjoys desirable properties including improved interpretability and numerical stability. Suppose the set of parameters in a model is denoted by $\boldsymbol{\theta}$ and θ_j denotes individual parameters, $j = 1, \dots, p$. Specifically a Lasso approach uses the following objective function:

$$PL(\boldsymbol{\theta}) = \log(p(\mathbf{y}|\boldsymbol{\theta}, M)) + \lambda \sum_{j=1}^n |\theta_j| = LL(\boldsymbol{\theta}) + \lambda \sum_{j=1}^n |\theta_j|, \quad (1)$$

where $PL(\boldsymbol{\theta})$ and $LL(\boldsymbol{\theta})$ are respectively the penalized and the usual log-likelihoods based on model M , and $\lambda \geq 0$ is a tuning parameter that determines the amounts of shrinkage (Tibshirani, 1996). A larger λ value tends to increase the penalty for more complex models that have more parameters. Because each tuning parameter value corresponds to a fitted model, selecting λ can be viewed as a model selection problem. Tibshirani (1996) proposed the quadratic programming algorithm to find the Lasso solutions and described methods such as cross-validation to select the tuning parameter λ in Eq. 1.

In Bayesian framework, the key quantity is the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y}, M) \propto p(\mathbf{y}|\boldsymbol{\theta}, M) \times p(\boldsymbol{\theta}|M)$, where $p(\boldsymbol{\theta}|M)$ is the prior distribution. Compared with a frequentist approach in which inference is based on the likelihood term $p(\mathbf{y}|\boldsymbol{\theta}, M)$, the Bayesian posterior contains the additional term for the prior $p(\boldsymbol{\theta}|M)$. This is an important observation as it is the connection to a regularization approach such as Lasso. The log posterior in a Bayesian approach takes the general form

$$\log(p(\mathbf{y}|\boldsymbol{\theta}, M)) + \log(p(\boldsymbol{\theta}|M)) = LL(\boldsymbol{\theta}) + \text{LPrior}(\boldsymbol{\theta}),$$

where $\text{LPrior}(\boldsymbol{\theta})$ represents the log prior (possibly multivariate). Thus, if the appropriate form of prior distribution is chosen, the log prior distribution in Bayesian analysis tends to play the role of the penalty function in Lasso. For example, Tibshirani (1996) suggested that Lasso estimates can be interpreted under the Bayesian framework when the θ_j 's are assigned independent and identical double-exponential priors, $\frac{\lambda}{2} \exp(-\lambda|\theta_j|)$. The probability density functions of double exponential with some different values of λ were presented in Figure 1. Like the normal distribution with zero mean, this distribution is unimodal and symmetrical. However, it has a sharper peak than the normal distribution. The plots revealed that the larger the value of λ , the more the probability density function concentrated around zero.

Algorithms for implementing Lasso include Efron, Hastie, Johnstone, and Tibshirani (2004), Friedman, Hastie, Hofling, and Tibshirani (2007), and Wu and Lange (2008). However, none of these algorithms provide a valid measure of standard error, which can be problematic for the frequentist Lasso (Kyung et al., 2010). Park and Casella (2008) proposed Gibbs sampling for the Lasso with the double-exponential priors in the hierarchical model, which was the first explicit treatment of a Bayesian Lasso regression (see also Hans, 2009). Compared with the frequentist Lasso, the Bayesian Lasso provides a valid measure of standard error (Kyung et al., 2010). Moreover, the sampling-based Bayesian framework also provides a flexible way of estimating the shrinkage parameter λ along with other parameters simultaneously, which lessens the heavily computational burden in selecting the optimal values of λ via a commonly used cross-validation procedure.

In recent years, a tremendous amount of work has been done to apply Lasso and L_1 -like penalties to the traditional models (Tib-

shirani, 2011). One example is the graphical Lasso (Yuan & Lin, 2007; Friedman et al., 2007) for fitting a sparse Gaussian graph that applied the L_1 -penalty—that is, the sum of $|\theta_j|$, to the inverse covariance matrix for graph-edge selection. The modeling of the inverse covariance matrix (or the precision matrix), instead of the covariance matrix, was an innovation proposed by Dempster (1972). For a joint multivariate normal distribution, zero entry of the precision matrix at the (i, j) position corresponds to conditional independence of the corresponding variables i and j given all other variables. Compared with zero entry of the covariance matrix at the (i, j) position which only implies noncorrelation between variables i and j , the conditional independence assumption is both stronger and more interpretable (Edwards, 2000). Indeed, a sparse inverse covariance matrix implies a sparse graphical model or path diagram of which conditional independence relationships between variables can be directly read off. The graphical Lasso is also known as covariance Lasso (Khondker, Zhu, Chu, Lin, & Ibrahim, 2013). Because our proposed model was inspired by the Bayesian approach for covariance Lasso in Khondker, Zhu, Chu, Lin, and Ibrahim (2013) and Wang (2012) hereafter we refer the proposed model as Bayesian covariance Lasso CFA.

Bayesian Covariance Lasso for Confirmatory Factor Analysis

Suppose $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ are independent random observations and each $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{ip})^T$ satisfies the following factor analysis model:

$$\mathbf{y}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda} \boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i, \quad i = 1, 2, \dots, n, \quad (2)$$

where $\boldsymbol{\mu}$ is $p \times 1$ vector of intercepts, and $\boldsymbol{\Lambda}$ is a $p \times q$ factor-loading matrix that reflects the relation of observed variables in \mathbf{y}_i with the $q \times 1$ latent factors in $\boldsymbol{\omega}_i$. Moreover, it is

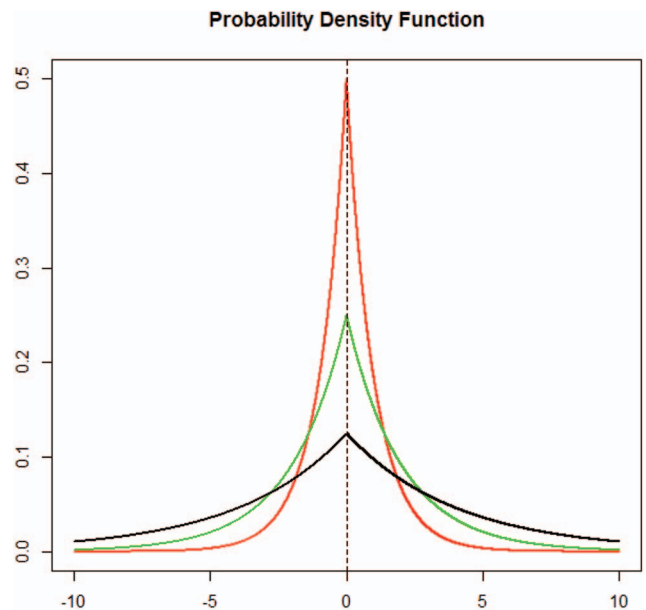


Figure 1. Probability density function of double exponential with three different values of λ : 1.0 (solid), 0.5 (dot-dash) and 0.25 (long-dash). See the online article for the color version of this figure.

assumed that ω_i follows $N[0, \Phi]$, and ϵ_i is a $p \times 1$ random vector of measurement errors, or residuals, which is independent of ω_i and follows $N[0, \Psi]$. Following a standard modeling approach for CFA, the number of latent factors and the structure of Λ are specified a priori. In contrast to the common diagonality assumption about Ψ in factor analysis, in our model, the variance-covariance matrix Ψ is not necessarily diagonal.

The proposed Bayesian covariance Lasso for CFA assigns prior distributions for the inverse of the covariance matrix instead of directly for the covariance matrix. Let $\Sigma = \Psi^{-1} = (\sigma_{ij})_{p \times p}$. Together with a positive definiteness constraint on the covariance matrix, independent exponential priors and the double exponential priors are respectively assigned for the diagonal and the off-diagonal elements of Σ (Wang, 2012). The double exponential density has the form $\frac{\lambda}{2} \exp(-\lambda |\sigma_{ij}|)$, $i < j$, whereas the exponential density function has the form $\frac{\lambda}{2} \exp(-\frac{\lambda}{2} \sigma_{ii})$. For a full Bayesian Lasso approach, instead of using cross-validation methods, an appropriate hyperprior is assigned to λ . Specifically, $\lambda \sim \text{Gamma}(\alpha_{\lambda 0}, \beta_{\lambda 0})$, where $\alpha_{\lambda 0}$ and $\beta_{\lambda 0}$ are hyperparameters whose values are preassigned. This Bayesian method of modeling λ has been applied in the context of Bayesian Lasso regression models (Hans, 2009; Park & Casella, 2008). The common choices for these hyperparameters are $\alpha_{\lambda 0} = 1$ and $\beta_{\lambda 0}$ to be small.

For the structural parameters involved in μ , Λ , and Φ , the following conjugate prior distributions are assigned (Lee, 2007). For $k = 1, \dots, p$,

$$\mu \sim N(\mu_0, H_{\mu 0}), \Lambda_k \sim N(\Lambda_{0k}, H_{0k}), \Phi^{-1} \sim \text{Wishart}(R_0, \rho_0), \quad (3)$$

where Λ_k^T is the k th row of Λ ; μ_0 , Λ_{0k} , ρ_0 , and positive definite matrices $H_{\mu 0}$, H_{0k} , and R_0 are hyperparameters whose values are assumed to be given to represent the available prior knowledge that may be obtained from the subjective knowledge of field experts and/or analysis of past or closely related data.

Gibbs Sampler and Positive Definiteness of Covariance Matrix

The Gibbs sampler (Geman & Geman, 1984), which is a form of Markov chain Monte Carlo (MCMC; Gilks, Richardson, & Spiegelhalter, 1996), combined with the idea of data augmentation (Tanner & Wong, 1987), iteratively simulates observations from the full conditional distributions such that eventually the posterior distribution of the parameters of interest and latent quantities can be empirically represented by a collection of simulated values. For example, in a simple model of two parameters θ_1 , θ_2 , the Gibbs sampler would iteratively draw values from the full conditional distributions $p(\theta_1 | \theta_2, y)$ and $p(\theta_2 | \theta_1, y)$, where y represents the data. If the full conditional distributions are standard distributions, simulating observations from them is rather straightforward and fast. For nonstandard conditional distributions, the Metropolis-Hastings (MH) algorithm can be used (see Hastings, 1970; Metropolis et al., 1953 for more details). Suppose the parameters of interest from the k th iteration in the Gibbs sampler is denoted by $\theta^{(k)} = (\theta_1^{(k)}, \theta_2^{(k)}, \dots, \theta_m^{(k)})$. It has been shown that under mild regularity conditions, the joint distribution of $(\theta_1^{(k)}, \theta_2^{(k)}, \dots, \theta_m^{(k)})$ converges to the desired posterior distribution $[0|y, M]$, after a

sufficiently large number of iterations (called a burn-in phase), say K . After discarding the observations obtained at the burn-in phase (or say, after convergence), statistical inference of θ can be carried out on the basis of the simulated samples $\{\theta^{(K+j)} : j = 1, 2, \dots, J\}$. For example, Bayesian point estimates and standard error estimates of θ can be produced by using the corresponding posterior means and their posterior covariance matrix of the simulated samples.

The specification of the Bayesian covariance Lasso CFA model allows a block Gibbs sampling scheme to be implemented for statistical inference. An outline of the implementation of the block Gibbs sampler is given in Appendix A. The block Gibbs sampler offers a convenient way to handle missing values. Under the missing-at-random (MAR) assumption (Little & Rubin, 1987), missing values were imputed. In other words, the conditional distributions of a missing variable uses data from other variables to inform the imputation. The block Gibbs sampler eventually provides an estimate of Σ that can be inverted to obtain $\Psi = \Sigma^{-1}$. Detailed model specification and full conditional distributions are described within the R codes provided in the online supplementary materials Section 1.

Under the proposed block Gibbs sampler, it can be proved that covariance matrix estimate is positive definite. The mathematical proof of positive definiteness of Σ is included in Appendix B. The positive definiteness of Σ implies that Ψ is also positive definite. The result is an important technical innovation because the positive definiteness of the covariance matrix is a necessary condition for a valid CFA solution.

Convergence Monitoring, Goodness-of-Fit, Uncertainty Estimates, and Identifiability Checking

The convergence of the MCMC algorithm is monitored by the estimated potential scale reduction (EPSR) value (Gelman, 1996) of each individual parameter of interest. The EPSR value can be calculated from several parallel sequences of simulated observations generated independently via different starting values. For more details of computation of the EPSR value, please refer to Gelman (1996). The whole simulation procedure is said to be converged if all the EPSR values are less than 1.2. Once the algorithm is judged to converge, a large number of draws can then be sampled from the joint posterior distribution to produce Bayesian point estimates and standard errors estimates of the unknown parameters and latent variables. The EPSR value can be obtained by using an R package called coda (Plummer et al., 2006).

The goodness-of-fit of the posited model was tested via the posterior predictive (PP) p value (Gelman, Meng, & Stern, 1996; Scheines, Hojink, & Boomsma, 1999) as introduced in Lee (2007) for SEMs and related models. As discussed in Muthén and Asparouhov (2012), the PP p value is akin to a model fit index. Here we used the PP p value as a complementary statistic for assessing the goodness-of-fit of a single model, given that the Bayesian Lasso was already used as a model selection tool. The implementation of the PP procedure in Mplus can be found in Asparouhov and Muthén (2010) and a brief overview can be found in Muthén and Asparouhov (2012). A model is considered plausible if the PP p value estimate is not far from 0.5.

Analogous to the measure of confidence interval in a frequentist's approach, the Bayesian approach uses the concept of highest posterior density (HPD) intervals to characterize the uncertainty associated with an estimated parameter. The $100(1 - \alpha)\%$ HPD interval (Box & Tiao, 1992; Chen & Shao, 1999) for an unknown parameter θ is defined as

$$H(\delta) = \{\theta : p(\theta | y, M) \geq \delta\},$$

where δ is the largest constant such that $\int_{H(\delta)} p(\theta | y, M) d\theta = 1 - \alpha$. The HPD interval can also be obtained by using the R package coda, which was used for the analysis in this article.

A key issue using a full covariance matrix for errors in CFA is model identifiability, for example, the model is not identified if the parameters are redundant. Under a frequentist framework, freeing all the residual covariance parameters violates the counting rule (Kaplan, 2009)—that is, the number of free parameters is less than or equal to $p(p + 3)/2$, where p is the dimension of y_i —that generally leads to an unidentified model. Using the data setting in the simulation Study 1 as an example, when $p = 10$, the number of free parameters is 76, which is greater than $10 \times (10 + 3)/2 = 65$, meaning that the model is unidentified. In a practical implementation such as using Mplus, when all residual covariances are specified to be freely estimated under the ML framework, an error message appears, stating that the degree of freedom for this model is negative and that the model is not identified will be returned. Under the Bayesian framework, model identification can be achieved by the provision of prior information in restricting the range of individual elements in the covariance matrix. The Bayesian Lasso method does not, however, exactly assign zero values to off-diagonal elements in the covariance matrix; it shrinks the more weakly related parameters close to zero faster (Park & Casella, 2008), and as a result, avoids the nonidentification problem. Kaufman and Press (1973) proved that the Bayesian CFA model with an inverse-Wishart prior for the full Ψ matrix is identified. For the proposed CFA model with a covariance Lasso prior, identifiability can be empirically evaluated by examining the convergence of an MCMC sequence of the parameters. One common way to do this is to generate parallel MCMC sequences with different starting values and check that the sequences mix well via trace plots. An alternative is to check EPSR values (< 1.2). Both methods were used to check identifiability in our subsequent data analysis.

Empirical Studies

Study 1: Parameter Recovery When the Model is Correctly Specified

The main purpose of this section is to evaluate the parameter recovery of the proposed CFA model with a Bayesian covariance Lasso prior when the model is correctly specified. A data set was simulated based on the model specified in Eq. 2 with the number of observed variables set at $p = 10$ and the number of factors set at $q = 2$. It was assumed that Eq. 2 contained the following structure:

$$\Lambda^T = \begin{pmatrix} 1.0 & \lambda_{21} & \lambda_{31} & \lambda_{41} & \lambda_{51} & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & \lambda_{72} & \lambda_{82} & \lambda_{92} & \lambda_{10,2} \end{pmatrix},$$

$$\Phi = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix},$$

where the 1's and 0's are fixed for identification purpose, and

$\Psi =$

$$\begin{pmatrix} \psi_{11} & 0.0 & 0.0 & 0.0 & 0.0 & \psi_{16} & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & \psi_{22} & 0.0 & 0.0 & 0.0 & 0.0 & \psi_{27} & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & \psi_{33} & 0.0 & \psi_{35} & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & \psi_{44} & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & \psi_{53} & 0.0 & \psi_{55} & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ \psi_{61} & 0.0 & 0.0 & 0.0 & 0.0 & \psi_{66} & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & \psi_{72} & 0.0 & 0.0 & 0.0 & 0.0 & \psi_{77} & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & \psi_{88} & 0.0 & \psi_{8,10} \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & \psi_{99} & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & \psi_{10,8} & 0.0 & \psi_{10,10} \end{pmatrix}.$$

All elements in Ψ were considered free and accordingly estimated. The model structure was presented in Figure 2(a). The true values of the structural parameters were set as follows: $\mu = (\mu_1, \mu_2, \dots, \mu_{10})^T = (0.5, 0.5, \dots, 0.5)^T$, $\lambda_{21} = \lambda_{72} = \lambda_{82} = 0.8$, $\lambda_{31} = \lambda_{41} = \lambda_{92} = 0.5$, $\lambda_{51} = \lambda_{10,2} = 0.3$, $\phi_{11} = \phi_{22} = 1.0$, $\phi_{21} = 0.3$, and $\psi_{kk} = 0.36$, for $k = 1, 2, 6, 7, 8$; $\psi_{kk} = 0.50$, for $k = 3, 4, 5, 9, 10$; $\psi_{16} = \psi_{61} = \psi_{27} = \psi_{72} = \psi_{35} = \psi_{53} = \psi_{8,10} = \psi_{10,8} = 0.3$. Two levels of sample sizes were used, $N = 200$ and $N = 500$, and for each sample size 100 replications were generated.

Additionally, a sensitivity analysis regarding inputs in the prior distributions was conducted by perturbing the prior input as follows:

Perturbed Input I: The elements in μ_0 and Λ_{0k} were taken as 0.0, and H_{μ_0} and H_{0k} were taken as diagonal matrices with all diagonal elements equal to 4.0 in appropriate order. $\rho_0 = 6$, $R_0 = 6I_2$, where I_2 is an identity matrix, $\alpha_{\lambda_0} = 1$, and $\beta_{\lambda_0} = 0.01$.

Perturbed Input II: The elements in μ_0 and Λ_{0k} were taken as true values, and H_{μ_0} and H_{0k} were taken as identity matrices in appropriate order. $\rho_0 = 3$, $R_0 = 6I_2$, where I_2 is an identity matrix, $\alpha_{\lambda_0} = 1$, and $\beta_{\lambda_0} = 0.005$.

Several test runs were initially conducted to determine the approximate number of MCMC iterations that would be required for achieving convergence. It was observed that the algorithm converged in less than 5,000 iterations, as indicated by the EPSR values being less than 1.2. In the actual run, we used the more conservative number of iterations of 10,000 for the burn-in. For each replication, we used 10,000 draws subsequent to the burn-in period for deriving required statistics. Based on 100 replications, the bias of the estimates (BIAS), the mean of the standard error estimates (SE), and the root mean squares (RMS) error between the estimates and the true values were computed. Table 1 summarizes the results of the simulation study with $N = 200$. Both BIAS and RMS values were quite small, indicating that the Bayesian estimates of the unknown parameters are generally accurate. Additionally, it can be seen that for the given level of sample size, the Bayesian estimates perform well under different prior inputs, suggesting that the analysis results are not sensitive to how prior inputs are specified. The result for $N = 500$ is similar to that for $N = 200$.

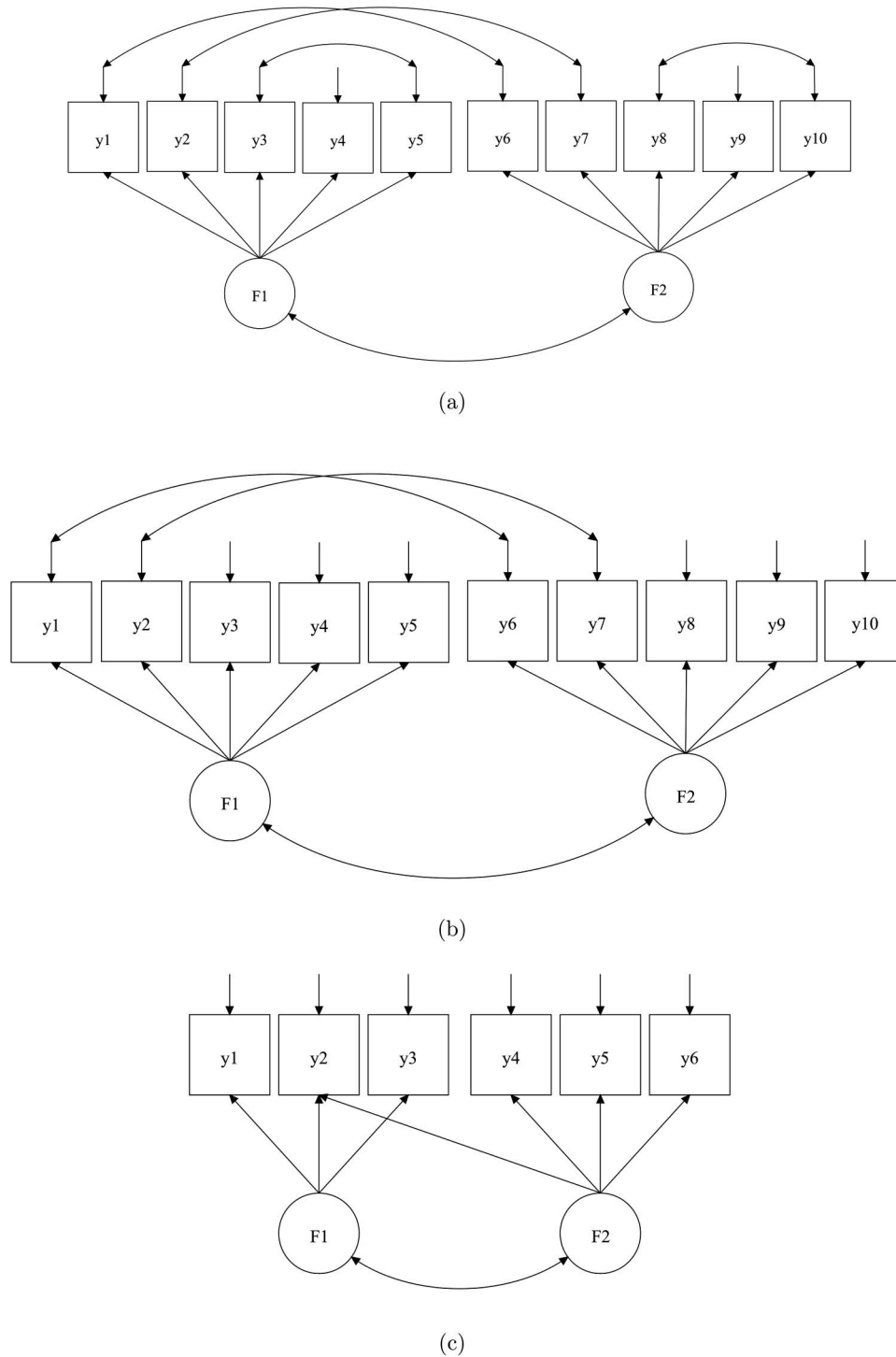


Figure 2. Model Structure in Simulation Studies 1–3.

Briefly, for most unknown parameters, the BIAS, SE, and RMS reduced when sample size was increased to $N = 500$. To save space, we do not present the full result for $N = 500$.

We also compared the performance of the Bayesian CFA that used the proposed Lasso prior to a standard Wishart prior distribution using the same simulated data sets. The simulation

experiment, which is not fully reported here due to space limitations, suggests that the proposed Lasso prior distribution performs better than the Wishart prior in terms of BIAS and RMS. For example, in $N = 200$ and 500 , compared with the Wishart prior, the L_1 loss for the Bayesian Lasso method was respectively reduced by 67% and 75%. The better performance

Table 1
Summary Statistics for Bayesian Estimates Under Perturbation
Input I and II Priors ($N = 200$, Covariance Lasso Prior)

Par	True	Perturbed input I			Perturbed input II		
		BIAS	SE	RMS	BIAS	SE	RMS
μ_1	.5	-.008	.083	.086	-.017	.083	.086
μ_2	.5	-.011	.071	.074	-.019	.071	.074
μ_3	.5	-.006	.062	.054	-.011	.062	.054
μ_4	.5	.001	.062	.074	-.004	.062	.073
μ_5	.5	-.004	.055	.054	-.008	.055	.054
μ_6	.5	-.011	.083	.082	-.021	.081	.083
μ_7	.5	-.014	.071	.068	-.022	.071	.069
μ_8	.5	.001	.072	.068	-.008	.071	.068
μ_9	.5	.002	.062	.063	-.003	.062	.063
μ_{10}	.5	.003	.055	.058	.001	.055	.058
λ_{21}	.8	.006	.045	.034	.007	.045	.034
λ_{31}	.5	.059	.135	.094	.063	.136	.094
λ_{41}	.5	-.003	.094	.067	-.001	.095	.065
λ_{51}	.3	.063	.124	.095	.067	.125	.098
λ_{72}	.8	.010	.043	.031	.010	.042	.031
λ_{82}	.8	.036	.124	.093	.039	.123	.095
λ_{92}	.5	-.006	.084	.053	.005	.084	.054
$\lambda_{10,2}$.3	.082	.127	.118	.084	.127	.120
ψ_{11}	.36	.035	.143	.101	.039	.145	.102
ψ_{22}	.36	.017	.103	.072	.018	.103	.073
ψ_{33}	.50	-.028	.106	.062	-.030	.107	.062
ψ_{44}	.50	.011	.082	.056	.010	.082	.056
ψ_{55}	.50	-.031	.079	.062	-.033	.080	.063
ψ_{66}	.36	.016	.132	.098	.018	.132	.096
ψ_{77}	.36	.008	.095	.064	.008	.095	.062
ψ_{88}	.36	-.016	.134	.060	-.019	.113	.061
ψ_{99}	.50	.018	.076	.058	.017	.076	.057
$\psi_{10,10}$.50	-.042	.083	.074	-.044	.084	.075
ψ_{16}	.3	-.051	.102	.096	-.049	.102	.095
ψ_{27}	.3	-.034	.075	.065	-.034	.075	.064
ψ_{35}	.3	-.045	.085	.065	-.047	.086	.066
$\psi_{8,10}$.3	-.065	.100	.086	-.067	.101	.087
ϕ_{11}	1.0	.002	.168	.106	-.001	.170	.107
ϕ_{21}	.3	.039	.115	.091	.038	.114	.092
ϕ_{22}	1.0	.009	.164	.120	.007	.163	.119

Note. BIAS = bias of the estimates; RMS = root mean squares; SE = standard error estimates.

of the Lasso prior may be due to (a) the priors were applied to correctly specified models and the Lasso prior expects the covariance structure to be sparse, which is indeed the case; and (b) the flexibility of the Lasso prior as compared to the Wishart prior, which tends to concentrate probability mass at the expected value.

Study 2: The Power of Detecting Significant Residual Covariances and the Potential Problem of Capitalization on Chance

Simulation Study 2 evaluates the performance of the proposed model in detecting genuine nonzero residual covariances and investigates the extent of capitalization on the chance the proposed model may suffer from as pointed out by MacCallum (1986) and thus both power and Type I error rates will be examined. A data set was simulated based on the model structure specified with the number of observed variables set at $p = 10$ and the number of factors set at $q = 2$. Two residual correlations ψ_{16} and ψ_{27} were set to nonzero values. The set-up

is visualized in Figure 2b. Two levels of factor loadings (0.5, 0.8), two levels of factor correlation (0.3, 0.7), and three levels of residual correlation (0.0, 0.3 and 0.7) were specified, resulting in a total of $2 \times 2 \times 3 = 12$ different conditions. A sample size of 250 was used for all the conditions and 100 replications were generated for each condition. For each condition, power for detecting individual nonzero residual correlation ψ_{16} and ψ_{27} as well as the average Type I error rate over zero residual correlations was calculated based on the 100 replications. The results are summarized in the top half of Table 2. It can be seen that in general the power of detecting ψ_{16} and ψ_{27} is satisfactory except for ψ_{27} under Conditions 5 and 7, of which residual correlations are at the lower level of 0.3. By comparing these results to the patterns of other conditions of which residual correlation is low, we suspected that lower power is due to the larger ratio between factor loading (0.8) and residual correlation (0.3). In order to test this hypothesis, four additional settings (13–18) were considered for different levels of factor loading (0.3, 0.5, 0.8) and one level of residual correlation (0.2). The results are reported in the bottom half of Table 2. It seems plausible that a low ratio of residual correlation and factor loading would result in low power.

For Type I error rates—that is, capitalization on chance when residual correlation is actually zero—the results in Table 2 reveal that the Bayesian covariance Lasso CFA procedure can maintain detection of spurious nonzeros at approximately the nominal rate of 5%. When both the correlation between the factors and the residual correlation are high (Conditions 4 and 20), the Type I error rate tends to elevate slightly above the nominal rate.

Study 3: Diagnostic Tool for Misspecification With Cross-Loading

Because the CFA is a strong model that places strict assumptions on the pattern of factor loading, it is prone to misspecification. For example, an item may cross-load on more than one factor. The performance of the Bayesian Lasso under such a situation would be of interest because residual correlations may manifest and sound an alarm. Thus, a study of the performance of the Bayesian covariance Lasso under a misspecified model is important for an understanding of the relationship between the various model components as well as potential use as a diagnostic tool.

We conducted simulation experiments for two scenarios of misspecification with cross-loading. A data set was simulated based on the model specified in Eq. 2 with the number of observed variables set at $p = 6$ and the number of factors set at $q = 2$. It was assumed that Λ contained the following overlapping structure:

$$\Lambda^T = \begin{pmatrix} 1.0 & 0.8 & 0.8 & 0.0 & 0.0 & 0.0 \\ 0.0 & \lambda_{22} & 0.0 & 1.0 & 0.8 & 0.8 \end{pmatrix},$$

where the 1's and 0's are fixed for identification purpose, and Ψ is assumed to be a diagonal matrix. Item 2 is cross-loaded on both Factors 1 and 2 with respective loadings of 0.8 and λ_{22} . The model structure is visualized in Figure 2c. Four settings were considered: $\{\lambda_{22} = 0.3, N = 250\}$, $\{\lambda_{22} = 0.3, N = 500\}$, $\{\lambda_{22} = 0.8, N = 250\}$, and $\{\lambda_{22} = 0.8, N = 500\}$. For each setting, 100 replications were generated. The data set was analyzed under the proposed

Table 2

Power of Detecting the Significant Residual Covariances and the Average of Type I Error Rates of Falsely Detecting the Nonsignificant Residual Covariances

	Loading	Factor Corr.	Residual Corr.	Power1	Power2	Ave.Type I
1	.5	.3	.3	.99	.90	1.76%
2	.5	.3	.7	1.00	1.00	3.27%
3	.5	.7	.3	1.00	.91	2.08%
4	.5	.7	.7	1.00	1.00	7.88%
5	.8	.3	.3	.74	.47	1.00%
6	.8	.3	.7	1.00	1.00	1.29%
7	.8	.7	.3	.76	.56	.00%
8	.8	.7	.7	1.00	1.00	1.00%
9	.5	.3	.0	—	—	1.53%
10	.5	.7	.0	—	—	2.12%
11	.8	.3	.0	—	—	1.00%
12	.8	.7	.0	—	—	1.00%
13	.3	.3	.2	.86	.78	1.91%
14	.3	.7	.2	.95	.91	5.19%
15	.5	.3	.2	.77	.49	1.73%
16	.5	.7	.2	.87	.52	1.78%
17	.8	.3	.2	.22	.15	1.00%
18	.8	.7	.2	.21	.19	1.00%
19	.3	.3	.7	1.00	1.00	3.12%
20	.3	.7	.7	1.00	1.00	7.75%

Note. Power1 = The power of detecting the significance of ψ_{16} . Power2 = The power of detecting the significance of ψ_{27} . Ave.Type I = The average of Type I error rates for falsely detecting the nonsignificant residual covariances.

CFA model with the Bayesian covariance Lasso prior and the following nonoverlapping factor structure of Λ :

$$\Lambda^T = \begin{pmatrix} 1.0 & \lambda_{21} & \lambda_{31} & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & \lambda_{52} & \lambda_{62} \end{pmatrix}. \quad (4)$$

Due to space limitations, we only report the full results for the high cross-loading cases of which $\lambda_{22} = 0.8$ with $N = 500$, and highlight

findings for $N = 250$, as well as for the low cross-loading cases of which $\lambda_{22} = 0.3$. Table 3 reports the patterns of significant residual covariances detected in the 100 replications. The most frequently detected significant residual covariances were $\{\psi_{42}, \psi_{52}, \psi_{62}\}$ (across Factor 1 and 2), and $\{\psi_{54}, \psi_{64}, \psi_{65}\}$ (within Factor 2). It can be seen that when the cross-loading for Item 2 is substantial ($\lambda_{22} = 0.8$) and the sample size is $N = 500$ (250), the total numbers of significant

Table 3

Patterns of Significant Residual Covariances and the Total Number of Specific Patterns During 100 Replications ($\lambda_{22} = 0.8$, $N = 500$)

	Pattern															Total1
	ψ_{21}	ψ_{31}	ψ_{32}	ψ_{41}	ψ_{42}	ψ_{43}	ψ_{51}	ψ_{52}	ψ_{53}	ψ_{54}	ψ_{61}	ψ_{62}	ψ_{63}	ψ_{64}	ψ_{65}	
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	0	0	0	0	1	0	0	1	0	1	0	1	0	0	1	1
	0	0	0	0	0	0	0	1	0	1	0	1	0	1	1	2
	0	0	0	0	1	0	0	1	0	1	0	1	0	1	1	72
	0	0	0	0	1	0	1	1	0	1	0	1	0	1	1	1
	1	0	0	0	1	0	1	1	0	1	0	1	0	1	1	1
	0	0	0	0	1	0	0	1	0	1	1	1	0	1	1	2
	1	0	0	0	1	0	1	1	0	1	1	1	0	1	1	2
	1	0	0	1	1	0	1	1	0	1	1	1	0	1	1	3
	1	0	0	1	1	0	1	1	1	1	1	1	0	1	1	1
	1	0	0	0	1	0	1	1	0	1	0	1	1	1	1	1
	0	0	0	0	1	0	0	1	1	1	0	1	1	1	1	1
	0	0	1	0	1	1	0	1	1	1	0	1	1	1	1	1
	1	0	1	1	1	0	0	1	0	1	1	1	1	1	1	1
	1	0	0	1	1	0	1	1	0	1	1	1	1	1	1	1
	1	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1
	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1
	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	7
Total2	19	0	10	14	97	9	19	99	12	99	19	99	14	98	99	

Note. Each row corresponds to a pattern where “1” represents significance detected by 95% highest posterior density (HPD). Total1 = The total number of specific pattern during 100 replications; Total2 = The total number of the significance of ψ during 100 replications.

residual covariances for ψ_{42} , ψ_{52} , ψ_{62} , ψ_{54} , ψ_{64} , and ψ_{65} are then respectively 97 (94), 99 (91), 99 (93), 99 (96), 98 (96), and 99 (97), all out of 100 (see Table 3). The patterns reveal that when the Bayesian Lasso model was misspecified by treating λ_{22} as zero, significant residual covariances are present between Item 2 and the items that are loaded on Factor 2 as well as the residual covariances among items that are on Factor 2. Interestingly, the cross-loading (in this case Item 2) appears to induce correlations to all of the item pairs within the factor to which the item belongs. When the cross-loading for Item 2 is small ($\lambda_{22} = 0.3$), the pattern of significant residual covariance is not as obvious. For example, the total number of significant residual covariances for the above parameters for $N = 250$ are 4, 4, 7, 5, 5, and 3, and for $N = 500$, 18, 17, 17, 21, 24, and 15. The Bayesian Lasso procedure is not powered to detect small residuals at these sample sizes. In summary, we observe that for both low and high cross-loading, when the sample size becomes larger, the trend of significant residual covariances becomes more apparent. When an item that belongs to a primary factor is cross-loaded on a different factor, it tends to “infect” all of the items within its primary factor. We conducted a separate experiment to examine more complex cross-loading. The patterns of residual covariances were consistent with what we observed but became more complex when the cross-loading patterns varied. The full results of the additional simulation experiments are not reported here.

Study 4: Analysis of Emotion With Bayesian Covariance Lasso CFA

The Bayesian covariance Lasso CFA, was applied to a data set collected from a study on emotions and diet in women (Lu, Huet, & Dubé, 2011). Data were collected from a sample of $n = 186$ Caucasian adult nonobese women in a large North American city. Participants, who were recruited through a local advertisement, all signed an informed consent form before engaging in the study. Each participant received a small incentive for participating. The study protocol was approved by the human subjects ethics committee of McGill University, Canada. Twenty-eight items were selected to cover the entire spectrum of positive and negative affects. The selected items were adapted from the Consumption Emotions Set (CES; Richins, 1997), which includes 47 emotion descriptors (items) representing the broader range of emotions consumers most frequently experience in consumption situations. To keep the questionnaire concise, Lu, Huet, and Dubé (2011) selected 28 emotion items that were found most relevant to food consumption from the literature on eating behaviors (Canetti, Bachar, & Berry, 2002) and excluded those less eating-relevant emotion items. Participants were asked to indicate the degree to which they experienced each of the emotions at the present moment by placing a mark on a 15-cm visual analogue scale with 0 cm indicating *don't feel the emotion at all* and 15 cm indicating *feel it intensely*. The scales were then transformed into numerical scales ranging from 0 to 150 mm with 1-mm gradations. Each participant reported her emotional states every 2 hr, 6 times per day over 10 observation days. In the current study, the response data from the third time point of observation were used to demonstrate the proposed method.

For the data used in the current study, out of 186 participants, $n = 11$ of cases were deemed to have high proportions (60% or more) of missing values were not included in the analysis, resulting in a sample size of $n = 175$. Six participants had 3.6% and one participant had 25% of missing values in their responses. The missing values were assumed to be MAR and the Gibbs sampler scheme imputed values for the missing data in our implementation (see section on Gibbs sampler). All raw data were treated as continuous and standardized.

To conduct the proposed Bayesian covariance Lasso CFA, we first needed to determine the factor structure of the items. Arguably, the CES contains the most extensive list of emotion descriptors of emotions. However, the number of identified emotion factors (hereafter called emotions) in either valence—positive emotion (PE) or negative emotion (NE)—are substantial in the CES. Table 4 contains the list of identified emotions and emotion descriptors from CES. Although prior work on eating-related emotion has identified four to six emotions, only a few used a two-factor model of positive and negative emotional valence (e.g., Han & Back, 2008). Macht (1999) reported that four emotions—joy, anger, fear, sadness—could change the characteristics of eating. In a study of eating in everyday life, Macht and Simons (2000) found six emotions that covered different emotional states relaxation, joy, anger, sadness, and tension. In the original analytic work in Lu, Huet, and Dubé (2011), a total of 28 items (emotion descriptors) and five emotions were represented by the following descriptors and identified in a CFA happy, peaceful, angry, ashamed, and worried. A comparison of the five identified factors to the CES emotions is provided in Table 4.

Partly because the focus of Lu, Huet, and Dubé (2011) was to test how location (home vs. away-from-home) moderated the relationship between emotion and food choice, some low-arousal positive emotions such as peaceful and calm that were often experienced at home were included while other positive emotions not directly related to this context such as surprised and amazed were not included. On the other hand, negative emotions such as guilty and ashamed have been known to be associated with overeating and junk food consumption. The ashamed factor in the five-factor solution (see Table 4) reflects this salient feature of the study. While the original CFA in Lu, Huet, and Dubé (2011) appears to have some limitations (e.g., the use of an averaged data set), as pointed out by a reviewer, we followed the five-factor model for our illustration of the Bayesian covariance Lasso CFA after examining the emotion structure and seeing its agreement with the CES. In the next section we provide alternative factor analyses of the same data set and compare the results to the five-factor CFA.

In summary, the Bayesian covariance Lasso CFA that we considered is defined in Eq. 2 with $p = 28$, $q = 5$. The PE factors were labeled happy/joyful (ξ_1 ; 12 items), peaceful (ξ_2 ; four items), and the NE factors were labeled Angry/depressed (ξ_3 ; six items), ashamed (ξ_4 ; three items), and worried/tense (ξ_5 ; three items). A nonoverlapping factor structure of Λ with fixed zero entries at the appropriate positions and the covariance matrix among the factors Φ can be expressed as follows:

Table 4
Identified Emotion Factors and Emotion Descriptors

Emotions in the consumption emotions set	Emotions in Lu, Huet, and Dubé (2011) article	Emotions in six-factor model
Positive emotions		
Joy: happy; joyful; pleased	Happy/joyful: happy; joyful; elated; amused;	Joyful: joyful; elated; amused; loving;
Love: loving; sentimental; warm hearted	loving; sentimental; warm-hearted;	sentimental; warm-hearted; optimistic;
Optimism: optimistic; encouraged; hopeful	optimistic; encouraged; hopeful; fulfilled;	encouraged; hopeful
Contentment: contented; fulfilled	accomplished	Contented: contented; fulfilled; accomplished
Peacefulness: calm; peaceful	Peaceful: calm; peaceful; serene; contented	Peaceful: calm; peaceful; serene
Romantic love: sexy; romantic; passionate		
Excitement: excited; thrilled; enthusiastic		
Surprise: surprised; amazed; astonished		
Negative emotions		
Sadness: depressed; sad; miserable	Angry: depressed; sad; miserable; frustrated;	Angry: depressed; sad; miserable; angry;
Anger: frustrated; angry; irritated	angry; annoyed	happy
Worry: nervous; worried; tense	Worried: nervous; worried; tense	Worried: frustrated; annoyed; nervous;
Shame: embarrassed; ashamed; humiliated	Ashamed: embarrassed; ashamed; guilty	worried; tense
Envy: envious; jealous		Ashamed: embarrassed; ashamed; guilty
Discontent: unfulfilled; discontented		
Loneliness: lonely; homesick		
Fear: scared; afraid; panicky		
Others: guilty; proud; eager; relieved		

$$\Lambda = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \lambda_{21} & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \lambda_{12,1} & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & \lambda_{14,2} & 0 & 0 & 0 \\ 0 & \lambda_{15,2} & 0 & 0 & 0 \\ 0 & \lambda_{16,2} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \lambda_{18,3} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \lambda_{22,3} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \lambda_{24,4} & 0 \\ 0 & 0 & 0 & \lambda_{25,4} & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \lambda_{27,5} \\ 0 & 0 & 0 & 0 & \lambda_{28,5} \end{pmatrix},$$

$$\Phi = \begin{pmatrix} \phi_{11} & \phi_{12} & \phi_{13} & \phi_{14} & \phi_{15} \\ \phi_{21} & \phi_{22} & \phi_{23} & \phi_{24} & \phi_{25} \\ \phi_{31} & \phi_{32} & \phi_{33} & \phi_{34} & \phi_{35} \\ \phi_{41} & \phi_{42} & \phi_{43} & \phi_{44} & \phi_{45} \\ \phi_{51} & \phi_{52} & \phi_{53} & \phi_{54} & \phi_{55} \end{pmatrix}.$$

Following common practice in CFA, each factor contained a loading value of one for one of the measurement variables for the identification of the scale of the latent factor.

The following hyperparameter values were used in the analysis: μ_0 and Λ_{0k} fixed at vector $\mathbf{0}$, $\mathbf{H}_{\mu 0}$, \mathbf{H}_{0k} fixed at four times the identity matrices with appropriate dimensions, and $\rho_0 = 12$, $\mathbf{R}_0 = 6\mathbf{I}$, $\alpha_{\lambda 0} = 1$, and $\beta_{\lambda 0} = 0.01$.

To assess the convergence behavior of the Gibbs sampler in the analysis, the EPSR values of the parameter sequences using different starting values were calculated, which suggested that

the Gibbs sampler converged in fewer than 3,000 iterations. In our actual implementation, the Bayesian analysis results were obtained using 20,000 simulated draws from the posterior distributions subsequent to a burn-in period of 5,000 iterations. The program is written in R (R Core Team, 2013). Conducting the real data analysis using a PC with Intel Core i7-2600@3.40 GHz CUP and 8G RAM took about 45 min. Sample codes for fitting the Bayesian covariance Lasso CFA to the emotion data are included in online supplementary materials Section 2.

For the proposed Bayesian Lasso model, the PP p value was 0.517, providing evidence that the model fitted reasonably well to the sample data. The Bayesian estimates of the unknown parameters in Λ , Φ , the diagonal elements in Ψ , and their corresponding 95% highest posterior density (HPD) intervals are presented in Table 5. It can be seen that all factor loading estimates are statistically significant and substantial in magnitude, which suggest strong associations between the latent factors and their corresponding indicators. As expected, from the estimates of ϕ_{ij} 's, the PE factors are positively correlated, the NE factors are positively correlated, and the PE factors and NE factors are negatively correlated.

Altogether, 29, or 7.6% significant residual covariance estimates ψ_{ij} ($i < j$) resulted out of 378 ($= C(28, 2)$) in Ψ . Table 6 reports the 95% HPD intervals of significant residual covariances between pairs of observed variables. The results show that after conditioning on the PE and/or NE factors, dependence between measurements—some within the same factor and some between factors—remains. The strongest covariances are observed for the within-factor item pair miserable/sad (0.30), and angry/miserable (0.25) (for NE), and fulfilled/accomplished (0.20), and joyful/happy (0.16) (for PE). Some cross-factor negative correlations, for example happy/sad (-0.27), and happy/depressed (-0.2), are also noted.

Table 5

Bayesian Estimates (Est) of the Unknown Parameters in Λ , Φ , the Diagonal Elements in Ψ , and Their Corresponding 95% HPD Intervals in Real Data

Item	Est of λ	HPD of λ	Est of ψ_{kk}	HPD of ψ_{kk}
1. Joyful	1.000*	—	.497	(.349, .641)
2. Happy	.780	(.583, .963)	.670	(.487, .859)
3. Elated	.840	(.623, 1.055)	.648	(.497, .819)
4. Amused	.735	(.523, .942)	.728	(.547, .907)
5. Optimistic	1.067	(.848, 1.303)	.423	(.289, .561)
6. Hopeful	1.056	(.822, 1.309)	.436	(.304, .582)
7. Encouraged	1.038	(.803, 1.270)	.457	(.324, .594)
8. Fulfilled	.972	(.758, 1.186)	.518	(.371, .678)
9. Accomplished	.967	(.755, 1.190)	.526	(.384, .681)
10. Warm-hearted	1.102	(.876, 1.337)	.410	(.289, .548)
11. Loving	1.020	(.792, 1.255)	.490	(.356, .638)
12. Sentimental	.758	(.540, .995)	.722	(.554, .908)
13. Serene	1.000*	—	.338	(.227, .469)
14. Calm	.925	(.756, 1.101)	.441	(.307, .590)
15. Peaceful	.998	(.844, 1.168)	.336	(.211, .473)
16. Contented	.774	(.579, .989)	.610	(.433, .797)
17. Frustrated	1.000*	—	.338	(.210, .465)
18. Annoyed	.984	(.830, 1.131)	.373	(.240, .511)
19. Angry	.718	(.533, .917)	.631	(.471, .810)
20. Miserable	.635	(.424, .877)	.684	(.491, .886)
21. Sad	.573	(.374, .779)	.734	(.540, .935)
22. Depressed	.763	(.566, .975)	.583	(.405, .764)
23. Ashamed	1.000*	—	.462	(.270, .659)
24. Guilty	.840	(.551, 1.135)	.595	(.384, .818)
25. Embarrassed	.694	(.413, .973)	.707	(.492, .940)
26. Nervous	1.000*	—	.277	(.188, .374)
27. Tense	1.001	(.860, 1.137)	.241	(.157, .339)
28. Worried	.951	(.818, 1.082)	.302	(.202, .407)

Par	Est	HPD	Par	Est	HPD
ϕ_{11}	.547	(.367, .742)	ϕ_{41}	-.068	(-.173, .042)
ϕ_{22}	.688	(.478, .897)	ϕ_{42}	-.145	(-.275, -.017)
ϕ_{33}	.719	(.522, .933)	ϕ_{43}	.330	(.193, .470)
ϕ_{44}	.610	(.388, .851)	ϕ_{51}	-.135	(-.252, -.022)
ϕ_{55}	.782	(.585, 1.013)	ϕ_{52}	-.250	(-.382, -.118)
ϕ_{21}	.403	(.265, .538)	ϕ_{53}	.622	(.459, .796)
ϕ_{31}	-.143	(-.252, -.036)	ϕ_{54}	.379	(.232, .536)
ϕ_{32}	-.264	(-.398, -.136)			

Note. HPD = highest posterior density. The items with asterisks are fixed to identify the scale of the latent factor. PE-general = Items 1–12; PE-peacefulness = Items 13–16; NE-general = Items 17–22; NE-shame = Items 23–25; NE-worry = Items 26–28.

Comparison With Other Factor Analytic Models

To gain a deeper understanding of the behavior of the Bayesian covariance Lasso model, the following CFA models were used in the comparison: the five-factor CFA model with the Bayesian covariance Lasso prior (M_1), the five-factor CFA model with diagonal Ψ (M_2), and a CFA model with diagonal Ψ (M_3), whose number of factors would be determined by an EFA. For the model (M_3), the EFA (with oblique rotation), based on factors with eigenvalues >1 , revealed a six-factor structure (see Table 4). The six-factor structure agreed reasonably well with the emotions identified in the CES. Two observations were made. First, compared with the five-factor model, fulfilled and accomplished from the factor happy/joyful, together with contented from the factor peaceful, were separated out as a distinct factor. Second, the item happy, to our surprise, was loaded in the angry factor with high negative loading. Referring to Table 6, most of the significant residual

covariances found in M_1 were among these items. In other words, if one is willing to let the data select the number of factors (via EFA), additional factors that group observed measurements with high error covariance would be revealed. The total explained variance of the six-factor structure was 70.40%, compared with 66.49% in the five-factor structure (M_2). Hereafter, M_3 refers to the six-factor CFA model with diagonal Ψ .

We applied multiple criteria to ascertain the quality of the three comparison models. First we examined traditional goodness-of-fit indexes. Several model fit indices, which are commonly reported in SEM software such as Mplus (Muthén & Muthén, 1998–2013), are presented in Table 7. The results indicate that neither five-factor diagonal model M_2 nor the six-factor diagonal M_3 fit the data very well, although the six-factor model is slightly better than the five-factor model. Model M_1 , which contains 29 significant residual covariances as identified by the Bayesian covariance

Table 6
The Significant Residual Covariance Estimates (Est) and Their Corresponding 95% HPD Intervals in Real Data

Residual covariance	Est	HPD
Joyful with happy	.157	(.030, .296)
Happy with amused	.140	(.004, .270)
Happy with sentimental	-.123	(-.234, -.008)
Happy with contented	.130	(.006, .260)
Happy with miserable	-.162	(-.280, -.043)
Happy with sad	-.266	(-.401, -.132)
Happy with depressed	-.200	(-.323, -.073)
Elated with amused	.160	(.038, .298)
Optimistic with hopeful	.121	(.009, .232)
Optimistic with guilty	-.111	(-.217, -.015)
Hopeful with encouraged	.112	(.010, .226)
Fulfilled with accomplished	.198	(.078, .330)
Fulfilled with contented	.163	(.048, .284)
Fulfilled with sad	-.108	(-.222, -.002)
Accomplished with loving	-.093	(-.186, -.009)
Accomplished with contented	.128	(.020, .241)
Loving with frustrated	.086	(.001, .169)
Sentimental with sad	.176	(.050, .298)
Contented with sad	-.159	(-.287, -.045)
Annoyed with depressed	-.098	(-.193, -.009)
Angry with miserable	.245	(.087, .392)
Angry with sad	.139	(.008, .284)
Angry with embarrassed	.222	(.093, .359)
Miserable with sad	.304	(.140, .476)
Miserable with depressed	.180	(.034, .337)
Miserable with embarrassed	.224	(.090, .365)
Sad with depressed	.226	(.070, .383)
Sad with worried	.112	(.007, .237)
Guilty with embarrassed	.227	(.047, .409)

Note. HPD = highest posterior density.

Lasso, fitted the data better in terms of all of the conventional fit indices (see Table 7). For example, the RMSEA of M_3 is only approximately half of that of either the five- and six-factor diagonal models, and the CFI of M_3 is higher than the cutoff of 0.9 while the other two models were below.

We further examined the impact of relaxing the diagonality requirement of the covariance on model goodness-of-fit in terms of differences in observed and expected (fitted) correlations between pairs of measurements. Various similar diagnostics have been proposed for latent class models; see for example, Garrett and Zeger (2000), and Reboussin, Ip, and Wolfson (2008). Unlike a latent class analysis that focuses on binary indicators, the CFA model contains

continuous indicators. We defined an analogous version of Reboussin, Ip, and Wolfson's (2008) log-odds-ratio check as follows:

$$z_{ij} = \frac{r_{ij}^{obs} - r_{ij}^{est}}{\sigma_{ij}^{est}}, \quad (5)$$

where r_{ij}^{obs} is the correlation coefficient between Item i and Item j calculated from the observed data, and r_{ij}^{est} and σ_{ij}^{est} are, respectively, the Bayesian estimates of the correlation coefficient between Item i and Item j and the corresponding standard error estimates calculated from the simulated samples of the parameters. The idea behind the use of the z -value is that if the model is correctly specified, the observed and expected correlations between pairs of measurements in a covariance structure analysis should be close. In other words, a z -value plot can be used as a visual tool for assessing what might have gone wrong at the item-pair level. As suggested by Reboussin, Ip, and Wolfson (2008), the relative magnitude of the z -values were examined and threshold levels of ± 1.5 were used for determining a significant deviation. A model that produces many large z -values would suggest that the model could be misspecified.

For these three comparison models, Figure 3 displays the z -values for different pairs of emotion items among the 28 emotion items. Under model M_2 , there is strong evidence that dependence exists between many pairs of items. Increasing the number of factors to six (M_3) apparently reduces the number of large z -values. Additionally, the proposed model M_1 results in smaller z -values and no z -value exceeds the threshold of 1.5 in either direction. In other words, M_1 (five factors with residual correlation) fits the data well and the fit is comparable, if not superior to, M_3 (six factors with no residual correlation).

A conundrum that a researcher using factor analytic method often faces is that, while more factors can help to achieve a better fit, some factors may not be of substantive interest. As a highly restrictive model, the CFA requires that the number of factors and to which factor an item belongs are predetermined, and that no correlation between item pairs is allowed given the factor structure. As demonstrated by the emotion example of 28 items, result from data analysis may not perfectly agree well with a theory-guided factor structure. While increasing the number of factors (from five to six in the emotion example) could improve goodness-of-fit, the interpretation of the model becomes less clear as one of the factor contains both PE (happy) and NE (angry) emotions. The previous analysis and comparison are not meant to be support one theory over another; the purpose is to demonstrate that the relax-

Table 7
Goodness-of-Fit Statistics of Different Models Fitted to the Emotion Data by Mplus

Comparison models	AIC	BIC	RMSEA	CFI	TLI	SRMR
5-factor with diagonal Ψ	11504.630	11802.120	.103	.797	.775	.074
6-factor with diagonal Ψ	11440.847	11754.161	.098	.819	.796	.090
5-factor with 29 residual covariances	10907.638	11292.615	.054	.925	.909	.065

Note. For both TLI and CFI, .9 represents adequate fit and .95 good fit. RMSEA < .05 indicates good fit, between .05 and .08 acceptable fit. SRMR < .08 considered adequate. For AIC and BIC, smaller value indicates better fit. AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; RMSEA = Root Mean Square Error of Approximation; CFI = Comparative Fit Index; TLI = Tucker Lewis Index; SRMR = Standardized Root Mean Square Residual.

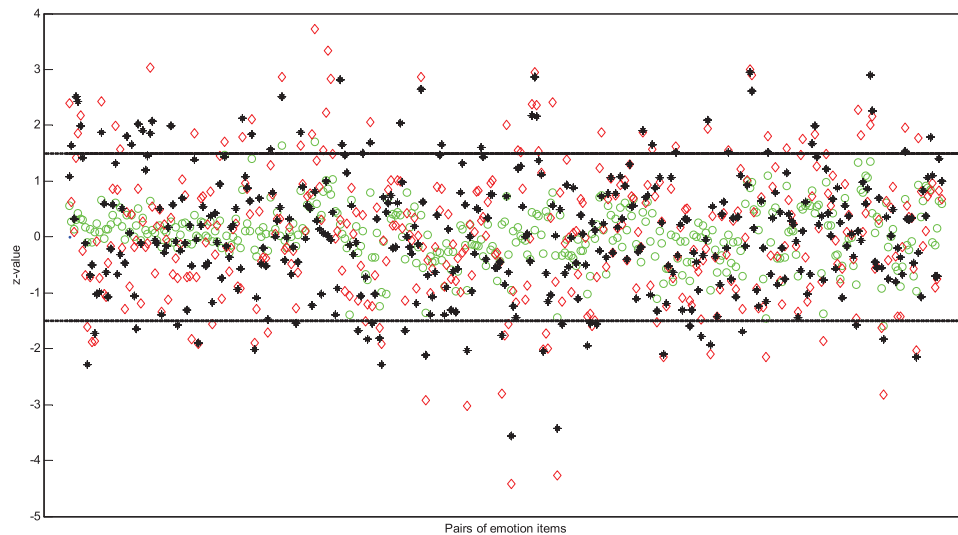


Figure 3. z-values for the five-factor confirmatory factor analysis (CFA) model with a covariance Lasso prior (o), the five-factor CFA model with diagonal Ψ (\diamond), and the six-factor CFA model with diagonal Ψ (*) fitted to the emotion data. The z-value is calculated for each pair of the 28 emotion items. See the online article for the color version of this figure.

ation of the diagonal assumption of the covariance matrix by replacing the covariance with a sparse structure can be used as an approximation to the theory. The solution keeps the factor structure intact and moves the solution empirically closer to real data. The empirical example also shows how the Bayesian covariance Lasso CFA can be used as a diagnostic tool. First, if a large number of significant residual covariance terms are present relative to the size of the matrix, or large individual covariances exist, it would then suggest that the entire CFA approach needs to be revised, or that additional steps need to be taken, such as removing items. Although we are not aware of any existing guideline for how large is “large,” without controlling for multiple comparisons, we would surmise that $>10\%$ of significant covariances or absolute values of correlation >0.5 should raise serious concern. Further studies are needed to provide guidelines for determining when a CFA needs to be refuted. The emotion example here shows weak-to-moderate residual covariances between some item pairs that signifies a certain level of item redundancy within the same factor. However the covariances do not appear to suggest a high level of clustering within and across factors. We suspect that a residual dependence between some item pairs arises partly because participants were unable to clearly distinguish certain descriptors of emotions that only have subtle differences in meaning—for example, miserable and sad, and optimistic and hopeful.

Study 5: Bias Reduction in Structural Equation Models

To illustrate the bias propagation problem for SEM and how the Bayesian covariance Lasso method can reduce bias in SEM parameter estimates, we simulated response data using a simple SEM example. The path diagram of the model is presented in Figure 4. The SEM involves 10 observed variables that are related with three latent factors $\omega_i = (\eta_i, \xi_{i1}, \xi_{i2})$, where η_i is endogenous factor, and ξ_{i1} and ξ_{i2} are exogenous factors that follow a bivariate normal distribution with

mean zero and variance-covariance Φ . The specification of Λ and Φ , and the values of their parameters are given by:

$$\Lambda^T = \begin{pmatrix} 1.0 & 0.8 & 0.5 & 0.3 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.5 & 0.8 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.8 & 0.5 \end{pmatrix},$$

$$\Phi = \begin{pmatrix} 1.0 & 0.3 \\ 0.3 & 1.0 \end{pmatrix},$$

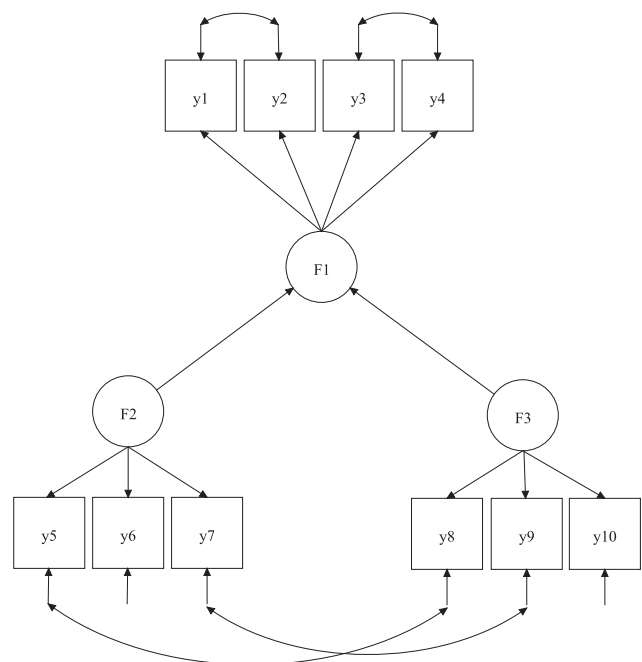


Figure 4. Model structure in simulation Study 5.

Table 8
Summary Statistics for Bayesian Estimates Under SEM With the Proposed Lasso Prior to Ψ and SEM With the Diagonal Ψ

	Par	True	N = 200			N = 500		
			BIAS	SE	RMS	BIAS	SE	RMS
SEM with Lasso prior to Ψ	γ_1	.6	.033	.085	.087	.028	.065	.063
	γ_2	-.6	-.030	.086	.079	-.035	.066	.068
SEM with diagonal Ψ	γ_1	.6	.149	.089	.173	.122	.056	.136
	γ_2	-.6	-.144	.090	.167	-.133	.056	.143

Note. BIAS = bias of the estimates; RMS = root mean squares; SE = standard error estimates; SEM = structural equation modeling.

where the 1's and 0's in Λ are fixed for identification purpose.

$\mu = (\mu_1, \mu_2, \dots, \mu_{10})^T = (0.5, 0.5, \dots, 0.5)^T$ and

$\Psi =$

$$\begin{pmatrix} 0.36 & 0.30 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.30 & 0.36 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.36 & 0.30 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.30 & 0.36 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.36 & 0.0 & 0.0 & 0.30 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.36 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.36 & 0.0 & 0.30 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.30 & 0.0 & 0.0 & 0.36 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.30 & 0.0 & 0.36 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.36 \end{pmatrix}.$$

All elements in Ψ were considered free parameters and estimated accordingly. The structural component of the generative model is given by the regression equation:

$$\eta_i = \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + \delta_i,$$

where $\gamma_1 = 0.6$, $\gamma_2 = -0.6$, and δ_i follows a normal distribution with mean 0 and variance 0.36. Two levels of sample sizes were used, $N = 200$ and $N = 500$, and for each sample size 100 replications were generated. The generated data was analyzed by two models: SEM with the proposed Lasso prior to Ψ and SEM with the diagonal Ψ . Based on the 100 replications, BIAS, SE, and RMS values were computed. Table 8 shows the results for the parameters of interest γ_1 and γ_2 .

Compared with results using diagonal covariance, the Bayesian Lasso method applied to Ψ results in smaller values in BIAS and RMS in the SEM regression parameters, suggesting that the Bayesian estimates of the γ 's are more accurate than those estimated by SEM with the diagonal Ψ . The simulation result here is not meant to fully capture the patterns of bias in SEM due to a violation of the diagonal error covariance assumption, but rather to illustrate the potential effects of the propagation of biases to other parts of the SEM and the reduction in propagation bias when the residual covariance is properly modeled.

Discussion

Whether using EFA or CFA, the goal of factor analysis is to explain the covariances between multiple observed variables by means of a small and known number of underlying latent variables, or factors (Bollen, 1989). From a purist point of view, the ideal factor structure for CFA would be one in which each measurement

is loaded on only one factor, and the factors completely explain away the covariation in the measurements—that is, the factors are independent and no correlated residual errors exist in the measurements. This kind of “ideal” CFA may not exist in reality, and if it does, a price has to be paid in terms of sacrifice of validity. The items written for attaining an ideal factor structure would likely be highly restrictive and lead to a construct that is too narrow to meaningfully function in practice—a phenomenon referred to by Ip (2010) as the validity-versus-dimensionality dilemma (the number of factors signifies the dimension of the model).

The dilemma can be examined by going back to a fundamental question: How do factor analytic models capture covariation among measured variables? In the current factor analysis literature, there is, in fact, more than one mechanism that can be utilized to explain measurement covariation. Three mechanisms have been identified: (C_1) latent factors, (C_2) correlation between factors, and (C_3) correlated residual error terms. For the ease of discussion, we call these covariation mechanisms. The mechanism C_2 is routinely implemented in CFA, and its presence actually implies that observed variables that belong to different factors are conditionally correlated given the factors (see also Fabrigar et al., 1999 for a discussion of oblique versus orthogonal factors). Thus a purist would not have included C_2 in the CFA. In this sense, there is nothing magical about C_3 , although its inclusion in the CFA often creates more controversy than its counterpart C_2 . From a purist view, C_1 is the only signal and the other mechanisms should not happen—the noise component should be entirely captured by the variance of the error component. A more realistic view is that C_1 is a signal as per the researcher's specification (based on theory), but noise does appear in various forms, including C_2 and C_3 .

Modification Index and Implementation

Examining the MI procedure in Mplus as an example can illuminate how the covariation mechanisms, especially the way in which C_3 is handled in practice. Under the maximum likelihood framework, Mplus first fits a CFA with diagonal residual covariance matrix Ψ to the data and identifies a set of candidates for modification by applying a threshold to the MIs of all estimated pairwise residual covariances. In the analysis of the emotion data, Mplus produced 90 residual covariances of which the MI values were greater than the prescribed threshold value of 3.84. The number of residual covariances identified by PMM is substantially higher than that identified by the Bayesian Lasso procedure. If the PMM approach is to be fully implemented, one would begin with

the largest MI of the 90 residual covariances, add the corresponding covariance term to the CFA, rerun the model, and examine the chi-squared test statistic. Rerunning the model is necessary because changing in a single parameter in a model could affect other parts of the solution (MacCallum, Roznowski, & Necowitz, 1992). The procedure thus has to be repeated many times, leading to a long series of modifications to the initial CFA model. While it is generally suggested that modifications should be used sparingly (MacCallum, 1995), no standard or guideline exists for choosing the number of modifications—only that the number probably should not be as high as 90 as our real data example suggested. Adding to the complexity of finding a rule for stopping model modification is the fact that adding off-diagonal entries to a positive definite covariance matrix may make it nonpositive definite, a technical and highly challenging issue (Pourahmadi, 2011) in covariance estimation that cannot be satisfactorily solved by directly manipulating individual off-diagonal elements. Adding to the list of procedure woes is the incremental test statistic for MI, which is not exactly distributed as χ^2 with one degree of freedom when the statistic is based on a post hoc model modification (Bentler, 2007). It is possible that the deviation is not substantial. However, it is not at all clear when the chi-squared distribution is a good approximation. As pointed out by Lu, Chow, and Loken (2016), no sampling distribution is currently available for the purpose of quantifying the uncertainties associated with MIs, and model complexity is not explicitly accounted for in the use of MIs.

Post Hoc Modification Versus the Bayesian Lasso

The practical implementation of PMM described above makes it clear that researchers can, or are at least tempted to, selectively report incremental MI statistics that support their desired model. Researchers, of course, can be encouraged to provide justification of added parameters based on theory (MacCallum, 1995). However, as asked by Steiger (1990) in a rhetorical question regarding the PMM approach: “What percentage of researchers would ever find themselves unable to think up a theoretical justification for freeing a parameter?” If one assumes that the answer is near zero, as Steiger did, it means that different researchers could report different CFA models of different fit indices even when they apply the same method to the same data, which is an unflattering situation that is not beneficial to the scientific enterprise of psychology (Jackson, Gillaspay, & Purc-Stephenson, 2009).

The residual covariance matrix reveals important information about the quality of the CFA. Unfortunately, the matrix is often underreported in the literature. From our own experience with secondary data analysis, we suspect many studies using CFA contained substantial numbers of residual covariances that were either not carefully examined and ignored, or not reported. Unlike other goodness-of-fit indexes, the failure to report on residual covariance is often overlooked in the review process of CFA studies. McDonald and Ho (2002) encouraged SEM researchers to publish a correlation matrix and discrepancies as well as goodness of fit indexes “so that readers can exercise independent and critical judgment.” We believe that the same can be said about residual covariance.

From the covariation mechanism perspective, PMM represents a half-hearted effort that attempts to “fix” the lack of fit problem using a questionable covariance matrix modification technique, a questionable test statistic, and questionable tactics to achieve model parsimony. There are often “rules” and cautionary remarks about making minor tweaks using MIs, which are more or less reflections of the ad hoc nature of the procedure and the ambiguity regarding what is considered a minor adjustment and sufficiently good fit. Indeed, the issue of false discovery of spurious residual covariance in PMM is technically challenging, partly because of the iterative nature of PMM. It was suggested that the temptation of tweaking more MIs of even better fit should be resisted because (a) overuse of modifications violates the purpose of CFA of precise theoretical propositions and the goal of parsimony, and (b) MIs are likely to include model improvement suggestions that are specific to the sample being examined (Bowen, 2014; Bowen & Guo, 2011). These are good points, but these concerns are precisely the reason modern statistics exist. Specifically, modern statistical techniques which appropriately penalizes overfitting to the sample data, have been extensively developed to address such problems. The Bayesian covariance Lasso CFA offers such a solution that has many practical and technical benefits and should be considered to be included into a quantitative psychologist’s toolbox.

References

- Asparouhov, T., & Muthén, B. (2010). *Bayesian analysis using Mplus Technical Implementation (Technical appendix)*. Los Angeles, CA: Muthén & Muthén.
- Bentler, P. M. (2007). On tests and indices for evaluating structural models. *Personality and Individual Differences*, 42, 825–829.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588–606.
- Bollen, K. A. (1989). *Structural equation models*. New York, NY: Wiley, Ltd.
- Bowen, N. K. (2014). *Requesting and using modification indices in Mplus*. Retrieved from <http://ssw.unc.edu/sswsig/sites/default/files/Modification%20Indices%20in>
- Bowen, N. K., & Guo, S. (2011). *Structural equation modeling*. New York, NY: Oxford University Press.
- Box, G. E. P., & Tiao, G. C. (1992). *Bayesian inference in statistical analysis*. New York, NY: Wiley.
- Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality*, 54, 106–148.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.
- Canetti, L., Bachar, E., & Berry, E. M. (2002). Food and emotion. *Behavioural Processes*, 60, 157–164.
- Chen, M. H., & Shao, Q. M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics*, 8, 69–92.
- Congdon, P. (2005). *Bayesian models for categorical data*. New York, NY: Wiley.
- Congdon, P. (2007). *Bayesian statistical modelling*. New York, NY: Wiley.
- Congdon, P. (2014). *Applied Bayesian modelling*. New York, NY: Wiley.
- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1, 16–29.
- Dempster, A. P. (1972). Covariance selection. *Biometrics*, 28, 157–175.

- DiStefano, C., & Hess, B. (2005). Using confirmatory factor analysis for construct validation: An empirical review. *Journal of Psychoeducational Assessment*, 23, 225–241.
- Edwards, D. (2000). Introduction to graphical modelling (2nd ed.) New York, NY: Springer.
- Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32, 407–499.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4, 272–299.
- Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, 9, 468–491.
- Friedman, J., Hastie, T., Hofling, H., & Tibshirani, R. (2007). Pathwise coordinate optimization. *The Annals of Applied Statistics*, 1, 302–332.
- Garrett, E. S., & Zeger, S. L. (2000). Latent class model diagnosis. *Biometrics*, 56, 1055–1067.
- Gelman, A. (1996). Inference and monitoring convergence. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in practice* (pp. 131–144), London, UK: Chapman & Hall.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis*. London, UK: Chapman & Hall/CRC.
- Gelman, A., Meng, X. L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6, 733–759.
- Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.
- Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice*. London, UK: Chapman & Hall.
- Han, H., & Back, K. (2008). Relationships among image congruence, consumption emotions, and customer loyalty in the lodging industry. *Journal of Hospitality & Tourism Research*, 32, 467–490.
- Hans, C. (2009). Bayesian lasso regression. *Biometrika*, 96, 835–845.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their application. *Biometrika*, 57, 97–109.
- Hays, R. D., Anderson, R., & Revicki, D. (1993). Psychometric considerations in evaluating health-related quality of life measures. *Quality of Life Research*, 2, 441–449.
- Hershberger, S. L. (2003). The growth of structural equation modeling: 1994–2001. *Structural Equation Modeling*, 10, 35–46.
- Hu, L. T., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, 112, 351–361.
- Ip, E. H. (2010). Empirically indistinguishable multidimensional IRT and locally dependent unidimensional item response models. *British Journal of Mathematical and Statistical Psychology*, 63, 395–415.
- Ip, E. H., Wang, Y., de Boeck, P., & Meulders, M. (2004). Locally dependent latent trait model for polytomous responses with application to inventory of hostility. *Psychometrika*, 62, 191–216.
- Jackson, D. L., Gillaspay, J. A., Jr., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, 14, 6–23.
- Jöreskog, K. G., & Sörbom, D. (1984). *LISREL VI user's guide*. Mooresville, IN: Scientific Software.
- Kaplan, D. (1990). Evaluating and modifying covariance structure models: A review and recommendation. *Multivariate Behavioral Research*, 25, 137–155.
- Kaplan, D. (2009). *Structural equation modeling: Foundations and extensions* (2nd ed.). Thousand Oaks, CA: Sage.
- Kaufman, G. M., & Press, S. J. (1973). *Bayesian factor analysis*. Sloan School of Management, University Chicago.
- Khondker, Z. S., Zhu, H. T., Chu, H. T., Lin, W. L., & Ibrahim, J. G. (2013). The Bayesian covariance lasso. *Statistics and Its Interface*, 6, 243–259.
- Kyung, M., Gill, J., Ghosh, M., & Casella, G. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian Analysis*, 5, 369–411.
- Lee, S. Y. (2007). *Structural equation modelling: A Bayesian approach*. New York, NY: Wiley.
- Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York, NY: Wiley.
- Lu, J., Huet, C., & Dubé, L. (2011). Emotional reinforcement as a protective factor for healthy eating in home settings. *The American Journal of Clinical Nutrition*, 94, 254–261.
- Lu, Z., Chow, S., & Loken, E. (2016). Bayesian factor analysis as a variable selection problem: Alternative priors and consequences. *Multivariate Behavioral Research*, 51, 519–539.
- MacCallum, R. C. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin*, 100, 107–120.
- MacCallum, R. C. (1995). Model specification: Procedures, strategies, and related issues. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications*. Thousand Oaks, CA: Sage.
- MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, 111, 490–504.
- Macht, M. (1999). Characteristics of eating in anger, fear, sadness and joy. *Appetite*, 33, 129–139.
- Macht, M., & Simons, G. (2000). Emotions and eating in everyday life. *Appetite*, 35, 65–71.
- Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher order factor models and their invariance across groups. *Psychological Bulletin*, 97, 562–582.
- McDonald, R. P., & Ho, M. (2002). Principles and practices in reporting structural equation analyses. *Psychological Methods*, 7, 64–82.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equations of state calculations by fast computing machine. *Journal of Chemical Physics*, 21, 1087–1091.
- Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, 17, 313–335.
- Muthén, L. K., & Muthén, B. O. (1998–2013). *Mplus user's guide*. 7th edition. Los Angeles, CA: Author.
- Park, T., & Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, 103, 681–686.
- Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R news*, 6, 7–11.
- Pourahamadi, M. (2011). Covariance estimation: The GLM and regularization perspectives. *Statistical Science*, 26, 369–387.
- R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Reboussin, B. A., Ip, E. H., & Wolfson, M. (2008). Locally dependent latent class models with covariates: An application to under-age drinking in the USA. *Journal of the Royal Statistical Society*, 171, 877–897.
- Richins, M. L. (1997). Measuring emotions in the consumption experience. *Journal of Consumer Research*, 24, 127–146.
- Scheines, R., Hojnik, H., & Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika*, 64, 37–52.
- Sörbom, D. (1989). Model modification. *Psychometrika*, 54, 371–384.
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173–180.
- Stone-Romero, E. F., Weaver, A. E., & Glenar, J. L. (1995). Trends in research design and data analytic strategies in organizational research. *Journal of Management*, 21, 141–157.
- Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association*, 82, 528–550.

- Thompson, B. (2004). *Exploratory and confirmatory factor analysis*, Washington, DC: American Psychological Association.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B*, 58, 267–288.
- Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: A retrospective. *Journal of the Royal Statistical Society*, 73, 273–282.

- Wang, H. (2012). Bayesian graphical lasso models and efficient posterior computation. *Bayesian Analysis*, 7, 867–886.
- Wu, T. T., & Lange, K. (2008). Coordinate descent algorithms for lasso penalized regression. *The Annals of Applied Statistics*, 2, 224–244.
- Yuan, M., & Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika*, 94, 19–35.

Appendix A

Gibbs Sampler Procedure for Bayesian Covariance Lasso CFA

The Gibbs sampler used for the Bayesian covariance Lasso CFA model is better described as a block Gibbs sampler because a group of parameters can be simultaneously simulated from a multivariate conditional distribution. Let $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$ be the observed data and $\mathbf{\Omega} = (\boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_n)$ be the matrices of latent factors. The details of the block Gibbs sampler which is implemented by iteratively simulating observations from the full conditional distributions are provided as follows:

1. Generate $\mathbf{\Omega}$ from $p(\mathbf{\Omega} | \mathbf{Y}, \boldsymbol{\mu}, \mathbf{\Lambda}, \Phi, \Psi)$;
2. Generate $\mathbf{\Lambda}$ from $p(\mathbf{\Lambda} | \mathbf{Y}, \mathbf{\Omega}, \boldsymbol{\mu}, \Phi, \Psi)$;
3. Generate Φ from $p(\Phi | \mathbf{Y}, \mathbf{\Omega}, \boldsymbol{\mu}, \mathbf{\Lambda}, \Psi)$;

4. Generate $\boldsymbol{\mu}$ from $p(\boldsymbol{\mu} | \mathbf{Y}, \mathbf{\Omega}, \Phi, \mathbf{\Lambda}, \Psi)$;
5. Generate Σ from $p(\Sigma | \mathbf{Y}, \mathbf{\Omega}, \boldsymbol{\mu}, \mathbf{\Lambda}, \Phi, \tau, \lambda)$ and compute $\Psi = \Sigma^{-1}$.

Except for $p(\Sigma | \mathbf{Y}, \mathbf{\Omega}, \boldsymbol{\mu}, \mathbf{\Lambda}, \Phi, \tau, \lambda)$, the other conditional distributions are standard densities including the normal and inverse-Wishart distributions. Simulating observations from these standard distributions is generally straightforward and efficient. The details for both the standard and non-standard conditional distributions are given in online supplementary materials Section 1.

Appendix B

Theorem and Proof of Positive Definiteness of the Inverse Covariance Matrix

Theorem

Suppose we start with a positive definite current value of Σ and sample from the Gibbs procedure described in Appendix A. It is guaranteed that the positive definite constraint on Σ is maintained at all subsequent iterations.

Proof

Following the iterative block Gibbs sampler described in Appendix A and the specific conditional distributions in the online supplementary materials Section 1, assume that at iteration t the value of $\Sigma^{(t)}$ is positive definite. Therefore, all of the leading principal minors of $\Sigma^{(t)}$ are positive and so are all its diagonal elements.

Define

$$\Sigma = \begin{pmatrix} \Sigma_{-kk} & \sigma_k \\ \sigma_k^T & \sigma_{kk} \end{pmatrix}. \quad (\text{A1})$$

where σ_{kk} is the k th diagonal element of Σ , $\sigma_k =$

$(\sigma_{k1}, \dots, \sigma_{k,k-1}, \sigma_{k,k+1}, \dots, \sigma_{kp})^T$ is the vector of all off-diagonal elements of the k th column, and Σ_{-kk} is the $(p-1) \times (p-1)$ matrix resulting from deleting the k th row and k th column from Σ . After updating the k th column $(\sigma_k^T, \sigma_{kk})^T$, we only need to prove that $\sigma_{kk}^{(t+1)}$ is positive and $|\Sigma^{(t+1)}|$ is positive.

Because $\Sigma_{-kk}^{(t)}$ is positive definite, its inverse $(\Sigma_{-kk}^{(t)})^{-1}$ is also positive definite, so there exists a unique lower triangular matrix \mathbf{L} with real and strictly positive diagonal elements, such that $(\Sigma_{-kk}^{(t)})^{-1} = \mathbf{L}\mathbf{L}^T$. For γ simulated from the prior distribution $\text{Gamma}(\alpha_{\lambda 0}, \beta_{\lambda 0})$, the element $\sigma_{kk}^{(t+1)} = \gamma + (\sigma_k^{(t+1)})^T (\Sigma_{-kk}^{(t)})^{-1} \sigma_k^{(t+1)} = \gamma + (\sigma_k^{(t+1)})^T (\mathbf{L}\mathbf{L}^T) \sigma_k^{(t+1)} = \gamma + (\mathbf{L}^T \sigma_k^{(t+1)})^T (\mathbf{L}^T \sigma_k^{(t+1)}) > 0$ since $(\mathbf{L}^T \sigma_k^{(t+1)})^T (\mathbf{L}^T \sigma_k^{(t+1)}) \geq 0$.

To complete the proof, observe that $|\Sigma_{-kk}^{(t)}| > 0$ and $\gamma > 0$, $|\Sigma^{(t+1)}| = |\Sigma_{-kk}^{(t)}| \times |\sigma_{kk}^{(t+1)} - (\sigma_k^{(t+1)})^T (\Sigma_{-kk}^{(t)})^{-1} \sigma_k^{(t+1)}| = \gamma |\Sigma_{-kk}^{(t)}| > 0$. \square

Received November 20, 2015

Revision received August 9, 2016

Accepted August 10, 2016 ■