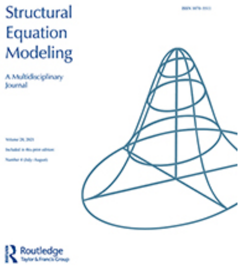

Structural Equation Modeling: A Multidisciplinary Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hsem20

A Bayesian Regularized Approach to Exploratory Factor Analysis in One Step

Jinsong Chen

STRUCTURAL EQUATION MODELING: A MULTIDISCIPLINARY JOURNAL
2021, VOL. 28, NO. 4, 518–528
https://doi.org/10.1080/10705511.2020.1854763

A Bayesian Regularized Approach to Exploratory Factor Analysis in One Step

Jinsong Chen

The University of Hong Kong

ABSTRACT

This research proposes a one-step Bayesian regularized approach to exploratory factor analysis (EFA) with an unknown number of factors. The proposed Bayesian regularized exploratory factor analysis (BREFA) model builds on the idea of bi-level Bayesian sparse group selection and can produce exact zero estimates at both the factor and loading levels. It can distinguish true factors from spurious factors and provide estimations of model and tuning parameters simultaneously. In addition to achieving model simplicity at both the factor and item levels, the approach provides interval estimates that can be used for significance testing, making it capable of addressing both uncorrelated and correlated factors. The Bayesian hierarchical formulation is implemented using Markov chain Monte Carlo estimation with the multivariate spike and slab priors and posterior median estimator. Based on simulated and real data analysis, BREFA demonstrates clear advantages or flexibility compared with traditional and Bayesian EFA, in terms of factor extraction, parameter estimation, and model interpretation.

KEYWORDS

Factor analysis; bayesian regularization; factor extraction; bilevel selection; spike and slab prior

Statistical regularizations is increasingly adopted in factor analysis (FA) or latent variable modeling for model sparsity (e.g., Feng et al., 2017; Jacobucci et al., 2016; Lu et al., 2016; Trendafilov et al., 2017), which can greatly benefit the development of psychometric scales or models. Compared with traditional methods, regularization provides more flexibility and/or viable alternatives to exploratory factor analysis (EFA) across different scenarios. When the number of factors is known, factor rotation is conducted conditional on an initial solution to obtain model simplicity under traditional EFA (Browne, 2001; Mulaik, 2010). There are a variety of rotation methods that differ in their ability to recover true parameters across different population patterns (Asparouhov & Muthén, 2009; Sass & Schmitt, 2010), making the selection of methods challenging and subjective in practice. In contrast, the performance of regularized EFA is comparable to that of the optimal rotational methods in different practical situations (Scharf & Nestler, 2019). In addition, the Bayesian regularized approach can offer interval estimates and related significance testing, which is challenging using rotational methods.

When the number of factors is unclear, an additional step is necessary to first extract the factors under traditional EFA settings. Adding to the complexity, there are a variety of methods for factor extraction, each with its own pros and cons under different situations, and a combination of them is suggested (Auerswald & Moshagen, 2019). Bayesian EFA can address factor extraction and parameter estimation in one step (Conti et al., 2014). But it's limited to the dedicated factor model where cross-loadings are not currently allowed. Regularized EFA under both the frequentist and Bayesian approaches can provide a one-step solution, but conditional on an important assumption of orthogonal or uncorrelated factors (Frühwirth-Schnatter & Lopes, 2018; Papastamoulis, 2020; Trendafilov et al., 2017). A one-step regularized approach for an oblique solution or correlated factors, which is more common in practice, is not yet available. Moreover, although the traditional multi-step approach is theoretically suboptimal owing to the possible loss of information between steps, it is empirically unclear under what conditions and/or to what extent a one-step approach can perform better.

Building on the recent development of statistical regularization, especially the Bayesian approach, this research proposes a one-step Bayesian regularized exploratory factor analysis (BREFA) model for EFA with an unknown number of factors. Compared with its frequentist counterpart (Huang et al., 2017; Jacobucci et al., 2016), Bayesian regularization with latent variable modeling (e.g., Feng et al., 2017; Lu et al., 2016; Muthén & Asparouhov, 2012) enjoys the benefits of straightforward estimation of the tuning parameters, interval estimates, and scalability to complex models that are analytically intractable. For instance, a one-step approach to EFA with unknown number of possibly correlated factors is analytically challenging under frequentist regularization. BREFA can distinguish true factors from spurious factors and provide simultaneous estimations of factor correlations, item parameters, and tuning parameters. In addition to achieving model sparsity at both the factor and item levels, the approach provides interval estimates that can be used for significance testing, making it capable of addressing both uncorrelated and correlated factors.

The BREFA extends the idea of Bayesian sparse group selection (Xu & Ghosh, 2015) to the FA context. With a similar mechanism of bilevel selection, factor selection and loading selection can coexist for EFA with an unknown number of factors. The multivariate spike and slab priors with the posterior median estimator (PME) were used at the factor and loading levels, which can produce exact zero estimates at both

CONTACT Jinsong Chen jinsong.chen@live.com Faculty of Education, The University of Hong Kong, Hong Kong.

the group and individual levels. The full Bayesian hierarchical formulation can be implemented with Markov chain Monte Carlo (MCMC; Gilks et al., 1996) estimation, which was a combination of the Gibbs sampling (Casella & George, 1992; Geman & Geman, 1984), empirical Bayes Gibbs sampling (Casella, 2001; Park & Casella, 2008), and Metropolis-Hastings algorithm (Hastings, 1970; Metropolis et al., 1953). Although currently limited to EFA, BREFA could be extended to other FA settings in future study by incorporating substantive knowledge at the loading level.

Through simulation studies, the current research further evaluated the performance of the BREFA model, with comparisons to the multi-step traditional EFA and one-step Bayesian EFA (BEFA) models across different conditions. For traditional EFA, a combination of parallel analysis (PA; Horn, 1965) and the sequential χ^2 model test (SMT), a recommended traditional method (Auerswald & Moshagen, 2019), was used to determine the number of factors. Next, Geomin rotation (Yates, 1987) is adopted for factor rotation and parameter estimations based on an initial model with the maximum likelihood (ML) estimation. Here, the BEFA for dedicated factor model was evaluated as an alternative one-step approach for comparison. The simulation results demonstrated the strengths and weaknesses of the different models and methods. Real data sets were analyzed to evaluate and compare the practical usefulness of such models and methods in practice.

Bayesian regularized factor analysis

Bayesian sparse group selection: Lasso vs. SSP

Factor and loading selection in BREFA can be started with the least absolute shrinkage and selection operator (Lasso), which is equivalent to the L_1-norm regularization that can shrink the unspecified parameters toward zero to obtain a sparse model (Tibshirani, 1996). The Lasso estimator for variable selection and estimation can be extended to cover group data at both the group and within-group levels. In a linear regression model with G groups, each with a coefficient vector $\boldsymbol{\beta}_g$ of length m_g, one has:

$$\mathbf{y}_{n\times 1}=\sum\nolimits_{g=1}^{G}\mathbf{X}_g\boldsymbol{\beta}_g+\boldsymbol{\varepsilon}, \tag{1}$$

where the error distribution is $\boldsymbol{\varepsilon}_{n\times 1} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}_n)$, $\mathbf{I}_n$ is the n-dimensional identity matrix, and $\mathbf{X}_g$ is an $n \times m_g$ covariate matrix corresponding to the group $\boldsymbol{\beta}_g$, $g = 1, 2, \ldots, G$. Based on the concept of the group-level Lasso (Yuan & Lin, 2006), a sparse group Lasso estimator to shrink the coefficient vector toward zero at both the group and within-group levels was presented by solving (Simon et al., 2013):

$$\min_{\boldsymbol{\beta}}\left(\left\|\mathbf{y}-\sum\nolimits_{g=1}^{G}\mathbf{X}_g\boldsymbol{\beta}_g\right\|_2^2+\gamma_1\|\boldsymbol{\beta}\|_1+\gamma_2\sum\nolimits_{g=1}^{G}\left\|\boldsymbol{\beta}_g\right\|_2\right), \tag{2}$$

where $\boldsymbol{\beta}$ was the coefficient vector containing all group vector $\boldsymbol{\beta}_g$, and γ_1 and γ_2 are the shrinkage parameters at the group and individual levels, respectively. Under the Bayesian approach, Equation 2 can be reframed as the Bayesian sparse group Lasso with the prior of the form (Xu & Ghosh, 2015):

$$p(\boldsymbol{\beta}) \propto \exp\left(-\gamma_1\|\boldsymbol{\beta}\|_1-\gamma_2\sum\nolimits_{g=1}^{G}\left\|\boldsymbol{\beta}_g\right\|_2\right), \tag{3}$$

which can be expressed as a scale mixture of normals with an exponential mixing density. To produce exact zero estimates as the frequentist Lasso does, the Bayesian Lasso estimator requires modification to incorporate the spike and slab prior (SSP). SSP is a multivariate zero inflated mixture prior mixed with point mass priors for the spike part and Lasso-type priors (with heavy tail) for the slab part (Yuan & Lin, 2005). Combining the power of the point mass mixture prior and the double exponential distribution in variable selection and estimation, the SSP with the posterior median estimator (PME) can offer an empirical Bayes estimation and is closely related to the Lasso estimator.

Although both the Bayesian group Lasso and SSP estimators share the desirable property of group-wise and within-group sparsity, only the latter can produce exact zero estimates. Moreover, the PME in the spike and slab-type models is consistent in model selection and has an optimal asymptotic estimation rate, while the group Lasso must sacrifice the estimation rate to achieve selection consistency (Xu & Ghosh, 2015). The selected model based on PME also has a far lower false-positive rate than the model chosen through the Lasso methods. Finally, whilst it is natural to extend the Bayesian sparse group approach to factor selection in FA, the ability to produce an exact zero estimate at the group level is important for eliminating the interference of unneeded factors. Otherwise, the system can become unstable owing to resampling behaviors associated with the spurious factors.

Bayesian regularized EFA with SSP

Consider a psychological scale with J items, K factors, and n respondents. Denote the data matrix as $\mathbf{Y} = (\mathbf{Y}_1, \cdots, \mathbf{Y}_J) = (y_{ij})_{n\times J}$ with vector $\mathbf{y}_i = (y_{i1},\cdots, y_{iJ})^{\mathrm{T}}$, where respondent $i = 1, \cdots, n$ and item $j = 1, \cdots, J$. Associated with the response data are the latent factors in the K-dimensional Euclidean space, $\mathbf{H} = (\mathbf{H}_1, \cdots, \mathbf{H}_K) = (\eta_{ik})_{n\times K}$ with vector $\boldsymbol{\eta}_i = (\eta_{i1}, \cdots, \eta_{iK})^{\mathrm{T}}$ and $k = 1, \cdots, K$. A general FA model satisfies the following equation:

$$\mathbf{y}_i = \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i, \tag{4}$$

where model parameters include the loading matrix $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_K) = (\lambda_{jk})_{J\times K}$ and the error term $\boldsymbol{\varepsilon}_i \sim N_J(\mathbf{0}, \boldsymbol{\Psi})$ with diagonal covariance matrix $\boldsymbol{\Psi} = \mathrm{diag}(\psi_{jj})_{J\times J}$. In addition, the factors are assumed to follow multivariate normal distributions, $\boldsymbol{\eta}_i \sim N_K(\mathbf{0}, \boldsymbol{\Phi})$, with $\boldsymbol{\Phi} = (\phi_{kk'})_{K\times K}$ as a correlation matrix for scale determinacy. Alternatively, one can estimate a covariance matrix while fixing one item per factor to determine the factor scale. Similar to classical FA, it is assumed that the data have been standardized and centered so that the intercept is dropped from the equation. However, an intercept vector can be easily included if needed.

Elements in $\boldsymbol{\Lambda}$ can be fixed as a specific value (usually zero), specified as free to estimate (specified loading; determined by expert and/or substantive knowledge), or unspecified (i.e., to be selected or determined by data). In this section, all loadings are considered to be unspecified, which is equivalent to full exploratory FA. Non-zero loading estimates based on the PME will be considered as selected. Note that there can be more factors than needed for selection. Factor k will be considered as selected if its eigenvalue based on the PME of the loading vector $\mathbf{\Lambda}_k$ is larger than zero, as $\|\hat{\mathbf{\Lambda}}_k\|_2^2 > 0$, implying that at least one loading estimate in $\mathbf{\Lambda}_k$ has been selected.

Accordingly, there are two types of sparsity that are of concern in the loading matrix, namely at the factor level (i.e., the loading vector) and for individual loading. Similar to the Bayesian sparse group approach with spike and slab prior, one can re-parameterize the loading vector to tackle the two concerns separately, as:

$$\mathbf{\Lambda}_k = \mathbf{V}_k^{1/2}\mathbf{b}_k, \tag{5}$$

where $\mathbf{b}_k = (b_{1k}, \ldots, b_{Jk})^{\mathrm{T}}$ controls the loading vector, $\mathbf{V}_k^{1/2} = \mathrm{diag}\{\tau_{1k}, \cdots, \tau_{Jk}\}$, and $\tau_{jk} \geq 0$ controls the magnitude of individual loading. To select the loading vector at the factor level, the following multivariate spike and slab prior is assumed for $\mathbf{b}_k$:

$$\mathbf{b}_k \sim (1 - \pi_0)N_J(\mathbf{0}, \mathbf{I}_J) + \pi_0\delta_0(\mathbf{b}_k), \tag{6}$$

where $\delta_0(\mathbf{b}_k)$ denotes a point mass at $\mathbf{0}$ for $\mathbf{b}_k$. This means that $\mathbf{b}_k$, when non-zero, has a $\mathbf{0}$ mean multivariate normal distribution with the identity matrix as its covariance matrix. Note that $\tau_{jk} = 0$, λ_{jk} is essentially dropped from the loading vector even when $b_{jk} \neq 0$. Therefore, the following spike and slab prior can be assumed for τ_{jk} to select individual loading within each loading vector:

$$\tau_{jk} \sim (1 - \pi_1)N^+(0, s_k^2) + \pi_1\delta_0(\tau_{jk}), \tag{7}$$

where $N^+(0, s_k^2)$ denotes a normal distribution $N(0, s_k^2)$ truncated below 0. Note that different from the Bayesian sparse group approach with an orthogonal design, one shrinkage parameter is needed on individual factor k as $1/s_k^2$, with $s_k^2 \sim \mathrm{Inv} - \mathrm{Gamma}(1, r_k)$. Conjugate beta hyper-priors can be assigned on hyperparameters π_0 and π_1, as:

$$\pi_0 \sim \mathrm{Beta}(a_{01}, a_{02}), \pi_1 \sim \mathrm{Beta}(a_{11}, a_{12}), \tag{8}$$

where one can obtain uniform hyper-priors with $a_{01} = a_{02} = a_{11} = a_{12} = 1$.

The j-th diagonal term of $\mathbf{\Psi}$, ψ_{jj}, can be assigned with an inverse gamma prior:

$$\psi_{jj} \sim \mathrm{Inv} - \mathrm{Gamma}(d_{j1}, d_{j2}), \tag{9}$$

and the common choices for hyperparameters are $d_{j1} = 1$ and d_{j2} to be small (e.g., .01). To obtain $\mathbf{\Phi}$, one can first directly sample from a covariance matrix with conjugate prior $\mathbf{\Phi}* \sim \mathrm{Inv} - \mathrm{Wishart}(\mathbf{S}_0^{-1}, q_0)$, where $q_0 = K + 2$ and the diagonal and off-diagonal elements of $\mathbf{S}_0$ can be set as 1s and 0.1s, respectively. The sampled $\mathbf{\Phi}*$ is then transformed into $\mathbf{\Phi}$ with a Metropolis-Hastings acceptance probability (Liu, 2008; Liu & Daniels, 2006). More details are provided in the section below. For uncorrelated factors, one can simply fix $\mathbf{\Phi}$ as $\mathbf{I}_K$.

For model identification, Lu et al. (2016) suggested that K minimum constraints can be used on the loading matrix to identify the Bayesian EFA with the spike and slab priors, instead of the K^2 minimum constraints needed for the regular EFA. This means that no further constraint is needed on the loading matrix in our case, as a correlation matrix is assumed between factors, which is equivalent to imposing K constraints on the loading matrix with a covariance matrix. This minimum constraint of identification suggests that all loadings can be unspecified, which was further tested through simulation studies.

Posterior analysis and MCMC estimation

With the above model specification, the joint posterior of $\mathbf{b} = (\mathbf{b}_1, \cdots, \mathbf{b}_K)^{\mathrm{T}}$, $\boldsymbol{\tau} = (\tau_{jk})_{J\times K}$, π_0, π_1, $\mathbf{\Psi}$, $\mathbf{H}$, and $\mathbf{\Phi}$ conditional on the data $\mathbf{Y}$ is:

$$p(\mathbf{b}, \tau, \pi_0, \pi_1, \mathbf{\Psi}, \mathbf{H}, \mathbf{\Phi}|\mathbf{Y}) \propto |\mathbf{\Psi}|^{-n/2} \exp\left\{-\frac{1}{2}\mathrm{Tr}\left[\left(\mathbf{Y} - \sum_{k=1}^{K}\mathbf{H}_k(\mathbf{V}_k^{1/2}\mathbf{b}_k)^{\mathrm{T}}\right)\mathbf{\Psi}^{-1}\right.\right.$$
$$\left.\left.\left(\mathbf{Y} - \sum_{k=1}^{K}\mathbf{H}_k(\mathbf{V}_k^{1/2}\mathbf{b}_k)^{\mathrm{T}}\right)^{\mathrm{T}}\right]\right\} \times \prod_{k=1}^{K}\left[(1 - \pi_0)(2\pi)^{-J/2}\exp\{-\mathbf{b}_k^{\mathrm{T}}\mathbf{b}_k/2\}\right.$$
$$\left.\mathrm{I}(\mathbf{b}_k \neq \mathbf{0}) + \pi_0\delta_0(\mathbf{b}_k)\right] \times \prod_{k=1}^{K}\prod_{j=1}^{J}\left[2(1 - \pi_1)(2\pi s_k^2)^{-1/2}\exp\left\{-\frac{\tau_{jk}^2}{2s_k^2}\right\}\right.$$
$$\left.\mathrm{I}(\tau_{kj} > 0) + \pi_1\delta_0(\tau_{jk})\right] \times \pi_0^{a_{01}-1}(1 - \pi_0)^{a_{02}-1} \times \pi_1^{a_{11}-1}(1 - \pi_1)^{a_{12}-1}$$
$$\times \prod_{k=1}^{K}\left[r_k(s_k^2)^{-2}\exp\left\{-\frac{r_k}{s_k^2}\right\}\right] \times \prod_{j=1}^{J}\left[(\psi_{jj}^2)^{-d_{j1}-1}\exp\left\{-\frac{d_{j2}}{\psi_{jj}^2}\right\}\right]$$
$$\times |\mathbf{\Phi}|^{-n/2}\exp\left\{-\frac{1}{2}\mathrm{Tr}\left[\mathbf{\Phi}^{-1}(\mathbf{H}^{\mathrm{T}}\mathbf{H})\right]\right\}$$
$$\times |\mathbf{\Phi}*|^{-(q_0+K+1)/2}\exp\left\{-\frac{1}{2}\mathrm{Tr}\left[\mathbf{S}_0^{-1}\mathbf{\Phi}*^{-1}\right]\right\},$$

where I(A) is an indicator function that takes the value 1 if A is true and 0 otherwise.

Although the joint posterior distribution has a complicated form difficult to handle, most resulting full conditionals are standard distributions that can be directly sampled with a Gibbs sampler or block Gibbs sampler, as shown in Appendix A. The procedure for sampling the parameters of interest from their full conditional distributions can be implemented in the following steps:

(1) Draw $\boldsymbol{\eta}_i$ from $p(\boldsymbol{\eta}_i|\text{rest})$ for $i = 1$ to N.
(2) Draw $\mathbf{b}_k$ from $p(\mathbf{b}_k|\text{rest})$ for $k = 1$ to K.
(3) For $\mathbf{V}_k^{1/2} = \mathrm{diag}\{\tau_{1k}, \cdots, \tau_{Jk}\}$, draw τ_{jk} from $p(\tau_{jk}|\text{ rest})$ for $j = 1$ to J, $k = 1$ to K.
(4) Draw π_0 and π_1 from $p(\pi_0|\text{rest})$ and $p(\pi_1|\text{rest})$, respectively.
(5) Obtain $\mathbf{\Phi}$ with procedures specified in Appendix A.
(6) Draw ψ_{jj} from $p(\psi_{jj}|\text{rest})$.
(7) Draw s_k^2 from $p(s_k^2|\text{rest})$.

Multiple chains with different initial values can be run to monitor the convergence of the algorithm. After the burn-in period, the convergence for the parameters of interest can be determined using the estimated potential scale reduction (EPSR) value (Gelman, 1996). The EPSR value compares the ratio of the weighted average of the within-chain variance and between-chain variance to the within-chain variance. The chains are said to converge to the stationary distribution if this ratio is less than 1.1 (Gelman et al., 2004). The standard errors of estimates can be characterized using the concept of highest posterior density (HPD) intervals, or more specifically the 100(1 − α)% HPD interval (Box & Tiao, 1973).

Simulation studies

Simulation studies were conducted to evaluate the performance of the proposed model and how it compared with that of traditional methods across different settings. After taking into consideration the data conditions in previous research (e.g., Auerswald & Moshagen, 2019; Scharf & Nestler, 2019) and real-life examples, the following conditions were simulated: two sample sizes N = 200 and 1,000; four true factors with eight items per factor, namely $K_t = 4$ and $J = 32$; factorial correlations $\phi_{kk'} = .5$ for k and $k' = 1$ to 4 and $k \neq k'$; and eight primary loadings per factor, λ_1 to λ_8, set as .6. Different conditions of cross-loadings and local dependence were manipulated as shown in the following two studies. All other loadings were set as $\lambda_0 = 0$. For each condition cell, 200 datasets were simulated and analyzed.

For the proposed model, two sets of prior values were evaluated for sensitivity analysis with the two sample sizes in a preliminary study, as shown in Table 1. Although Bayesian structural equation models may perform differently with different sets of priors, especially when the sample size is small (Van Erp et al., 2017), the difference in all of our cases turned out to be trivial. The first set of values was adopted for subsequent analyses, which was less informative. All Markov chains were found to reach stationary status within 5,000 and 10,000 iterations for N = 200 and 1,000, respectively. The author set the burn-in at 10,000 and 20,000 iterations, and parameters were estimated based on 10,000 and 20,000 draws after the burn-in phase for the two scenarios, respectively. The number of possible factors for extraction was set as K = 8. Larger numbers of factors (e.g., K = 10) were tried, with trivial difference except for a much longer running time.

Additionally, the following models or methods were considered for comparisons.

For factor extraction, SMT, PA, a combination of both (SMT-PA), and BEFA were compared. For parameter recovery, the Geomin rotation after the initial solution based on ML-EFA estimation was compared. Although BEFA is a dedicated factor model and theoretically inappropriate for cross-loadings, it could still be meaningful if the number of factors or primary loadings can be correctly recovered.

For performance assessment, the bias of the parameter estimates (BIAS), the root mean squares error (RMSE) between the estimates and true values, the mean of the standard error estimates (SE), and the proportion of estimates significantly different from zero at α = .05 (SIG) based on the HPD interval were computed. Note that SIG gives the Type I error and statistical power of the estimates. However, interval estimates and related SIG are not available in traditional methods. To compare the different methods, we reported the proportion of absolute estimates larger than .1 ($SIG_{.1}$) for reference.

All programming was conducted on the R platform (R Development Core Team, 2020). Program codes to implement BREFA are available from OSF at https://osf.io/a8hvc/?view_only=b1d453539fa7420691f932a61075f162. Both the EPSR value and HPD interval can be obtained using the *coda* package (Plummer et al., 2006). PA and SMT were implemented following Auerswald and Moshagen (2019)'s procedures.[1] Geomin was implemented using the *GPArotation* package (Bernaards & Jennrich, 2005) based on an initial ML-EFA with the *psych* package (Revelle, 2020). BEFA was implemented using the *BayesFM* package (Piatek, 2020).

Study 1: Model performance and comparisons under local independence

In this study, local independence was assumed by setting all off-diagonal elements in $\boldsymbol{\Psi}$ as 0. Two conditions of cross-loadings were simulated: 1) two cross-loadings per factor (i.e., λ_{c1}, λ_{c2}), and 2) four cross-loadings per factor (i.e., λ_{c1}, λ_{c2}, λ_{c3}, λ_{c4}). The cross-loadings were placed on the first succeeding factor and the last two or four items associated with the set of eight primary loadings. The value of cross-loadings was fixed at .2.

As shown in Table 2, both BREFA and SMT can correctly extract the number of factors most of the time, although SMT was slightly worse. PA was problematic when the sample size was small and the number of cross-loadings was large; therefore, the SMT-PA combination was also problematic in this case. BEFA performed the worst among all of the methods, and surprisingly, was much worse with larger samples. In line with this, BEFA's parameter estimates were barely useful and were excluded from further analysis.

When factor extraction was incorrect, it was difficult to match parameter estimates with the true structure, which was not surprising. In most cases, BREFA produced a spurious factor, which was uncorrelated with true factors and measured only by a few cross-loadings. In contrast, the traditional method was more irregular for both loading and factor recovery. To ensure a fair comparison, only data sets with correct factor extraction under each method or model were evaluated for parameter recovery. To maximize the benefit for traditional methods, data sets based on SMT (with higher accuracy of factor extraction) were adopted.

As shown in Table 3, the parameter estimates from BREFA for N = 200 were largely satisfactory across all cases, except for

Table 1. Two sets of prior values.

Set	a_{01}	a_{02}	a_{11}	a_{12}	d_{j1}	d_{j2}	λ_{0j}	H_{0j}	S_0	q_0
1	1	1	1	1	1	.01	0	4**I**	**I**+.1$_{od}$	$K+2$
2	1.5	1.5	1.5	1.5	1	.1	0	**I**	**I**+.5$_{od}$	$K+8$

I+.1$_{od}$: identity matrix with off-diagonal elements as.1; **I**+.5$_{od}$: off-diagonal elements as.5.

Table 2. Accuracy of different factor extraction methods in study 1.

N	CPF	BREFA	BEFA	SMT	PA	SMT-PA
200	2	.995	.580	.953	.930	.890
	4	.945	.550	.903	.610	.557
1000	2	1.000	.000	.943	1.000	.943
	4	1.000	.000	.953	1.000	.953

CPF: cross-loadings per factor.

[1]The syntax is available at https://osf.io/gqma2/?view_only?d03efba1fd0f4c849a87db82e6705668.

Table 3. Parameter recovery for BREFA in study 1, N = 200.

	CPF = 2					4				
Par	BIAS	RMSE	SE	SIG	$SIG_{.1}$	BIAS	RMSE	SE	SIG	$SIG_{.1}$
λ_1	−.082	.103	.043	.997	1.000	−.093	.114	.045	.984	1.000
λ_2	−.082	.104	.043	.997	1.000	−.093	.113	.046	.983	1.000
λ_3	−.084	.105	.043	.997	1.000	−.096	.118	.045	.984	1.000
λ_4	−.086	.106	.043	.997	1.000	−.093	.115	.045	.983	1.000
λ_5	−.089	.109	.043	.997	1.000	−.032	.084	.051	.987	1.000
λ_6	−.081	.102	.043	.997	1.000	−.036	.086	.051	.987	1.000
λ_7	−.040	.080	.046	.997	1.000	−.033	.085	.052	.985	1.000
λ_8	−.037	.080	.045	.999	1.000	−.031	.084	.052	.985	1.000
λ_{c1}	−.022	.093	.057	**.510**	.763	−.051	.112	.064	**.339**	.636
λ_{c2}	−.019	.095	.057	**.504**	.762	−.050	.113	.064	**.336**	.627
λ_{c3}	-	-	-	-	-	−.056	.117	.064	**.313**	.610
λ_{c4}	-	-	-	-	-	−.054	.115	.064	**.323**	.618
λ_0	.023	.043	.031	.009	.054	.023	.045	.030	.010	.057
$\phi_{kk'}$	−.094	.127	.098	.985	1.000	−.064	.122	.109	.961	.984

For $\phi_{kk'}$, k and k' = 1 to 4 and $k \neq k'$; CPF: cross-loadings per factor; BIAS = bias of the parameter estimates; SE: mean of the standard error estimates; RMSE: root mean square error between the estimates and true values; SIG = proportion of estimates significantly different from zero at α =.05; $SIG_{.1}$ = proportion of absolute estimates larger than.1; values of concern are highlighted.

the low SIG on the cross-loadings, which suggested low statistical power due to the small magnitude of the cross-loadings. However, the power based on $SIG_{.1}$ was better. All of the BREFA estimates were slightly better, with a smaller number of cross-loadings; and were substantially better when the sample size increased from 200 to 1,000 (Table 4). In comparison, there was one salient difference in the Geomin estimates when N = 200 (Table 5): Based on $SIG_{.1}$, the power for cross-loadings was higher but the Type I error for zero loadings was also larger. Stated differently, there was a compromise between the power for cross-loadings and the Type I error for zero loadings between the two methods, and BREFA offered a simpler model under the same subjective cutoff criterion. Other than that, or when N = 1,000, the two methods were quite close, except that SIG based on interval estimates were not available from Geomin.

Table 4. Parameter recovery for BREFA in study 1, N = 1000.

	CPF = 2					4				
Par	BIAS	RMSE	SE	SIG	$SIG_{.1}$	BIAS	RMSE	SE	SIG	$SIG_{.1}$
λ_1	−.012	.028	.018	1.000	1.000	−.022	.035	.019	1.000	1.000
λ_2	−.012	.028	.018	1.000	1.000	−.023	.036	.019	1.000	1.000
λ_3	−.013	.028	.018	1.000	1.000	−.024	.036	.019	1.000	1.000
λ_4	−.011	.026	.018	1.000	1.000	−.023	.036	.019	1.000	1.000
λ_5	−.010	.026	.018	1.000	1.000	−.009	.029	.019	1.000	1.000
λ_6	−.012	.028	.018	1.000	1.000	−.010	.032	.019	1.000	1.000
λ_7	−.006	.028	.017	1.000	1.000	−.010	.033	.019	1.000	1.000
λ_8	−.004	.031	.017	1.000	1.000	−.010	.031	.018	1.000	1.000
λ_{c1}	.001	.032	.018	.998	1.000	.001	.033	.023	.998	1.000
λ_{c2}	−.001	.036	.018	.985	.988	.003	.037	.023	.995	.995
λ_{c3}	-	-	-	-	-	.003	.039	.022	.998	.998
λ_{c4}	-	-	-	-	-	.006	.034	.022	.995	.995
λ_0	.004	.009	.010	.001	.001	.007	.015	.012	.004	.003
$\phi_{kk'}$	−.015	.038	.034	1.000	1.000	−.028	.048	.041	1.000	1.000

For $\phi_{kk'}$, k and k' = 1 to 4 and $k \neq k'$; CPF: cross-loadings per factor; BIAS = bias of the parameter estimates; SE: mean of the standard error estimates; RMSE: root mean square error between the estimates and true values; SIG = proportion of estimates significantly different from zero at α =.05; $SIG_{.1}$ = proportion of absolute estimates larger than.1.

Study 2: Model performance and comparisons under local dependence

There were eight terms of local dependence, namely eight non-zero off-diagonal elements for the upper triangle in $\boldsymbol{\Psi}$ with a symmetric lower triangle. Four of them were within-factor (one per factor), while the other four were between-factor (equally distributed among factors). Meanwhile, two effects of local dependence were simulated: .2 and .3. All local dependence terms were on items without cross-loadings. There were two cross-loadings per factor (i.e., λ_{c1}, λ_{c2}), which was set similarly to above, with the loading value fixed at .2.

As shown in Table 6, both BREFA and PA can extract the number of factors correctly most of the times, but both SMT and BEFA were problematic in all cases.

Accordingly, the SMT-PA combination was also problematic. Like the above, BEFA's parameter estimates were barely useful and excluded from further analysis, and only data sets with correct factor extraction under each method or model were evaluated for parameter recovery. To maximize the benefit for traditional methods, data sets based on PA (with higher accuracy of factor extraction) were adopted.

As shown in Table 7, parameter estimates from BREFA for N = 200 were largely satisfactory across all cases, except for the low SIG on the cross-loadings, which suggested low statistical power due to the small magnitude of the cross-loadings. However, the power based on $SIG_{.1}$ was better. Overall, the BREFA estimates tended to be slightly better with smaller local dependence effects, and were substantially better when the sample size increased from 200 to 1,000 (Table 8).

Relative to BREFA, a similar trend can be found in the Geomin estimates when N = 200 (Table 9): Based on $SIG_{.1}$, the power for cross-loadings was higher but the Type I error for zero loadings was also larger. Similarly, one can either opt for a better Type I error for zero loading with BREFA or a better power to detect cross-loadings with Geomin when the sample size is small. This also implies that BREFA offered a simpler model based on the same criteria. Moreover, the Type I error for BREFA was even smaller based on SIG. Other than that, or when N = 1,000, the two methods were quite close, except that interval estimates and related SIG were not available from Geomin.

Real-life example

The Humor Styles Questionnaire (HSQ; Martin et al., 2003) was used for illustration. There were four factors and 32 items, each targeting at specific factor, as shown in Appendix B. Although the primary loading of each item was confirmed, cross-loadings were of concern for some items because the related behaviors tended to be multidimensional (Heintz, 2017). Public data for HSQ are available online at https://openpsychometrics.org/_rawdata/, and complete responses from 443 females were analyzed in this example. Only BREFA and PA extracted the correct number of factors, whereas both SMT and BEFA seriously over-extracted the number of factors.

Table 5. Parameter recovery for Geomin in study 1.

	N = 200						1000					
	CPF = 2			4			CPF = 2			4		
Par	BIAS	RMSE	$SIG_{.1}$	BIAS	RMSE	$SIG_{.1}$	BIAS	RMSE	$SIG_{.1}$	BIAS	RMSE	$SIG_{.1}$
λ_1	−.025	.082	1.000	−.019	.083	1.000	−.012	.036	1.000	.001	.036	1.000
λ_2	−.028	.082	1.000	−.015	.082	1.000	−.011	.036	1.000	−.002	.035	1.000
λ_3	−.029	.081	1.000	−.019	.082	1.000	−.011	.036	1.000	.001	.036	1.000
λ_4	−.030	.084	1.000	−.018	.082	1.000	−.010	.036	1.000	.002	.036	1.000
λ_5	−.031	.082	1.000	−.005	.073	1.000	−.009	.035	1.000	.004	.033	1.000
λ_6	−.025	.079	1.000	−.010	.075	1.000	−.011	.037	1.000	.005	.035	1.000
λ_7	−.020	.073	1.000	−.006	.076	1.000	−.006	.031	1.000	.005	.034	1.000
λ_8	−.017	.075	1.000	−.009	.076	1.000	−.005	.032	1.000	.005	.033	1.000
λ_{c1}	.004	.080	.908	−.026	.089	.795	−.006	.037	.996	−.033	.050	.965
λ_{c2}	.003	.083	.894	−.027	.088	.793	−.007	.038	.992	−.032	.052	.957
λ_{c3}	-	-	-	−.031	.091	.775	-	-	-	−.033	.052	.945
λ_{c4}	-	-	-	−.027	.090	.803	-	-	-	−.031	.050	.961
λ_0	.024	.079	**.149**	.020	.080	**.145**	.009	.036	.005	.006	.040	.008
$\phi_{kk'}$	−.100	.117	1.000	−.060	.090	1.000	−.036	.047	1.000	−.003	.037	1.000

For $\phi_{kk'}$, k and k' = 1 to 4 and $k \neq k'$; CPF: cross-loadings per factor; BIAS = bias of the parameter estimates; RMSE: root mean squares error between the estimates and the true values; $SIG_{.1}$ = proportion of absolute estimates larger than.1; values of concern are highlighted.

Table 6. Accuracy of different factor extraction methods in study 2.

N	LD	BREFA	BEFA	SMT	PA	SMT-PA
200	.2	.995	.035	.023	.960	.023
	.3	.985	.000	.000	.970	.000
1000	.2	.990	.000	.000	1.000	.000
	.3	.980	.000	.000	1.000	.000

LD: local dependence effect.

Table 7. Parameter recovery for BREFA in study 2, N = 200.

	LD =.2					0.3				
Par	BIAS	RMSE	SE	SIG	$SIG_{.1}$	BIAS	RMSE	SE	SIG	$SIG_{.1}$
λ_1	−.096	.116	.043	.997	1.000	−.117	.139	.045	.989	1.000
λ_2	−.095	.114	.043	.997	1.000	−.117	.139	.045	.990	1.000
λ_3	−.103	.125	.044	.997	1.000	−.125	.151	.046	.989	1.000
λ_4	−.101	.122	.043	.997	1.000	−.124	.149	.046	.985	1.000
λ_5	−.020	.062	.042	.997	1.000	.023	.075	.046	.989	1.000
λ_6	−.022	.065	.042	.997	1.000	.025	.074	.046	.987	1.000
λ_7	−.054	.087	.045	.997	1.000	−.078	.111	.047	.992	1.000
λ_8	−.053	.088	.045	.997	1.000	−.076	.110	.047	.992	1.000
λ_{c1}	−.017	.092	.055	**.546**	.780	−.006	.095	.054	**.595**	.799
λ_{c2}	−.017	.094	.055	**.548**	.776	−.007	.096	.054	**.584**	.786
λ_0	.027	.049	.031	.018	.073	.034	.059	.033	.035	.109
$\phi_{kk'}$	−.120	.150	.099	.966	1.000	−.154	.180	.101	.925	1.000

For $\phi_{kk'}$, k and k' = 1 to 4 and $k \neq k'$; LD: local dependence effect; BIAS = bias of the parameter estimates; SE: mean of the standard error estimates; RMSE: root mean squares error between the estimates and true values; SIG = proportion of estimates significantly different from zero at α =.05; $SIG_{.1}$ = proportion of absolute estimates larger than.1; values of concern are highlighted.

Loading estimates for both BREFA and Geomin can be found in Table 10. Note that we assumed the correct number of factors was extracted before using Geomin, whereas such an assumption was not required in BREFA. As shown, the loading estimates were generally close between the two methods and all primary loadings were successfully identified in both cases. However, BREFA provided fewer cross-loadings than Geomin based on a .1 cutoff (15 vs. 21). Significant cross-loadings for BREFA were even fewer (11). There were some differences in the estimates of factorial correlations between the two methods (Table 11). However, it is difficult to tell whether there was any statistical difference, because interval estimates are not available for Geomin. However, in the case of

Table 8. Parameter recovery for BREFA in study 2, N = 1000.

	LD =.2					0.3				
Par	BIAS	RMSE	SE	SIG	$SIG_{.1}$	BIAS	RMSE	SE	SIG	$SIG_{.1}$
λ_1	−.024	.039	.022	1.000	1.000	−.069	.090	.024	.999	1.000
λ_2	−.027	.043	.022	1.000	1.000	−.070	.092	.024	.999	1.000
λ_3	−.035	.054	.023	1.000	1.000	−.082	.105	.024	.999	1.000
λ_4	−.041	.059	.024	.998	1.000	−.085	.107	.024	.999	1.000
λ_5	.041	.048	.023	.998	1.000	.085	.092	.024	.995	1.000
λ_6	.039	.047	.023	.998	1.000	.085	.092	.024	.996	1.000
λ_7	−.024	.041	.020	1.000	1.000	−.061	.084	.023	.999	1.000
λ_8	−.020	.038	.020	1.000	1.000	−.058	.083	.023	.999	1.000
λ_{c1}	.011	.041	.022	.975	.993	.019	.052	.023	.985	.992
λ_{c2}	.010	.039	.022	.985	.995	.018	.051	.023	.985	.986
λ_0	.009	.018	.015	.016	.019	.023	.038	.019	.073	.075
$\phi_{kk'}$	−.044	.061	.039	.998	1.000	−.123	.146	.048	.974	1.000

For $\phi_{kk'}$, k and k' = 1 to 4 and $k \neq k'$; LD: local dependence effect; BIAS = bias of the parameter estimates; SE: mean of the standard error estimates; RMSE: root mean square error between the estimates and true values; SIG = proportion of estimates significantly different from zero at α =.05; $SIG_{.1}$ = proportion of absolute estimates larger than.1.

BREFA, it is clear that an orthogonal solution for uncorrelated factors is not appropriate with two significant correlations.

Discussion

This research proposes a one-step Bayesian regularization approach for EFA with an unknown number of factors. With a mechanism similar to bi-level Bayesian sparse group selection, BREFA can distinguish true factors from spurious factors and provide estimations of model and tuning parameters simultaneously. In addition to achieving model simplicity at both the factor and item levels, the approach also provides interval estimates that can be used for significance testing, making it capable of addressing both uncorrelated and correlated factors. The simulation studies showed that BREFA can extract factors correctly in most cases across different settings, with satisfactory parameter recovery except for relatively low power to detect cross-loadings when the sample size was small. When factor extraction was incorrect, BREFA produced a spurious factor, uncorrelated with other factors and measured only by cross-loadings.

Table 9. Parameter recovery for Geomin in study 2.

	N = 200						1000					
	LD =.2			0.3			LD =.2			0.3		
Par	BIAS	RMSE	$SIG_{.1}$	BIAS	RMSE	$SIG_{.1}$	BIAS	RMSE	$SIG_{.1}$	BIAS	RMSE	$SIG_{.1}$
λ_1	−.049	.094	1.000	−.077	.116	1.000	−.032	.048	1.000	−.061	.073	1.000
λ_2	−.050	.094	1.000	−.078	.118	1.000	−.033	.048	1.000	−.061	.073	1.000
λ_3	−.057	.103	1.000	−.081	.128	.999	−.040	.055	1.000	−.066	.079	1.000
λ_4	−.055	.101	1.000	−.080	.126	1.000	−.042	.057	1.000	−.067	.081	1.000
λ_5	.058	.097	1.000	.120	.148	1.000	.077	.084	1.000	.144	.149	1.000
λ_6	.057	.098	1.000	.121	.148	1.000	.076	.084	1.000	.144	.149	1.000
λ_7	−.036	.082	1.000	−.062	.104	1.000	−.029	.043	1.000	−.054	.065	1.000
λ_8	−.037	.082	1.000	−.062	.104	1.000	−.026	.042	1.000	−.052	.064	1.000
λ_{c1}	.002	.080	.895	.008	.086	.890	−.003	.037	.994	.004	.039	.996
λ_{c2}	.001	.083	.888	.006	.088	.878	−.004	.037	.995	.003	.039	.996
λ_0	.025	.083	**.167**	.028	.090	**.186**	.011	.041	.023	.014	.049	.050
$\phi_{kk'}$	−.108	.123	1.000	−.119	.134	.999	−.046	.054	1.000	−.060	.066	1.000

For $\phi_{kk'}$, k and k' = 1 to 4 and $k \neq k'$; LD: local dependence effect; BIAS = bias of the parameter estimates; RMSE: root mean squares error between the estimates and true values; $SIG_{.1}$ = proportion of absolute estimates larger than.1; values of concern are highlighted.

Table 10. Loading estimates for the humor styles questionnaire.

	BREFA				Geomin			
Item	F1	F2	F3	F4	F1	F2	F3	F4
1	−0.626				−0.657			
2	0.103	0.550				0.579		
3	0.176		0.446	0.172	0.132		0.450	0.171
4				0.632				0.655
5	0.636				0.615			
6	0.245	0.374			0.220	0.410		
7			−0.505		0.168		−0.551	
8				0.679				0.703
9	−0.497				−0.501			
10		0.732		0.134		0.778		0.131
11			0.567				0.602	
12	0.158			0.588	0.122		−0.111	0.623
13	0.655				0.659			
14	0.236	0.594			0.208	0.627		
15			−0.564				−0.575	
16			−0.119	−0.538			−0.108	−0.564
17	−0.719				−0.740			
18		0.713				0.775		
19	0.261		0.564		0.201		0.580	
20			0.161	0.673			0.143	0.702
21	0.668				0.693			
22		−0.418				−0.439	−0.114	
23			−0.414				−0.432	
24				0.470	−0.178			0.507
25	−0.633				−0.642			
26	0.150	0.641			0.111	0.673		
27			0.526				0.550	
28	0.254		0.198	0.263	0.193	0.115	0.201	0.269
29	−0.441				−0.489		−0.126	0.164
30	0.103	0.384			0.103	0.413		−0.103
31			−0.662				−0.696	
32	0.136			0.690				0.714
EV	3.440	2.593	2.410	2.800	3.498	2.976	2.648	3.096

EV denotes eigenvalue; only absolute estimates >.1 are presented; for BREFA, loading estimates insignificant at α =.05 are underscored.

Table 11. Factorial correlation for the humor styles questionnaire.

	BREFA			Geomin		
	F1	F2	F3	F1	F2	F3
F2	**0.350**			0.401		
F3	**0.134**	0.009		0.206	0.077	
F4	0.101	0.071	0.063	0.195	0.135	0.148

Correlation estimates significant at α =.05 are highlighted.

For traditional methods of factor extraction, the highly recommended SMT-PA combinations only worked when there was no local dependence and the sample size was large. Instead, PA was suggested except in cases of a large number of cross-loadings and small sample size. Bayesian EFA failed to extract factors correctly in most conditions. For parameter estimation, Geomin was largely satisfactory aside from two shortcomings, namely 1) a relatively large Type I error for cross-loadings, which implied a less sparse model, and 2) the unavailability of interval estimates, making it difficult to determine whether the factors were significantly correlated or not.

In summary, BREFA demonstrated clear advantages over traditional and Bayesian EFA under simulated conditions, which were further demonstrated by real-life data. Moreover, BREFA offers more flexibility for model simplicity and interpretation, using either subjective cutoff criteria or statistical significance. Although limited to EFA at the moment, BREFA has the potential to incorporate substantive knowledge with a prespecified loading pattern and accordingly to cover a wide range of the entire FA family. In addition to developing a partially exploratory approach to accommodate different amounts of knowledge, future studies can extend the existing regularized approach to include a structural component, likely comparable to exploratory structural equation modeling (Asparouhov & Muthén, 2009). At the moment, a two-step approach can be adopted, using BREFA as the first step for a measurement model and a regular structural equation model with latent variables as the second step. Regularization for multi-group settings can also be addressed similarly to its frequentist counterpart (Huang, 2018).

The existing simulation studies are limited, and more comprehensive investigation is needed to understand the strengths and weaknesses of the proposed models across a wider range of conditions (e.g., nonnormality, more complex structure). The finding that BREFA can produce spurious factors uncorrelated with true factors and measured only by a few cross-loadings may also deserve further investigation, especially under local dependence. It suggests that the true factors can be identified and that most parameters can be recovered, even when factor extraction is incorrect. Moreover, the extent to which the spurious factor may be connected to different magnitudes or patterns of local dependence could be investigated. Finally, in addition to comparisons with multi-step traditional methods, one could consider a multi-step regularized approach with the orthogonal EFA version (e.g., Frühwirth-Schnatter & Lopes, 2018; Papastamoulis, 2020; Trendafilov et al., 2017) for an unknown number of factors and an oblique FA version (e.g., Huang et al., 2017; Jacobucci et al., 2016; Lu et al., 2016) for parameter estimates, or a mix of regularized and traditional methods.

ORCID

Jinsong Chen http://orcid.org/0000-0002-0157-5469

References

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, *16*, 397–438. https://doi.org/10.1080/10705510903008204

Auerswald, M., & Moshagen, M. (2019). How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychological Methods*, *24*, 468–491. https://doi.org/10.1037/met0000200

Bernaards, C. A., & Jennrich, R. I. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement*, *65*, 676–696. https://doi.org/10.1177/0013164404272507

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

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, *36*, 111–150. https://doi.org/10.1207/S15327906MBR3601_05

Casella, G. (2001). Empirical Bayes Gibbs sampling. *Biostatistics*, *2*, 485–500. https://doi.org/10.1093/biostatistics/2.4.485

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

Conti, G., Frühwirth-Schnatter, S., Heckman, J. J., & Piatek, R. (2014). Bayesian exploratory factor analysis. *Journal of Econometrics*, *183*, 31–57. https://doi.org/10.1016/j.jeconom.2014.06.008

Feng, X.-N., Wu, H.-T., & Song, X.-Y. (2017). Bayesian regularized multivariate generalized latent variable models. *Structural Equation Modeling*, *24*, 341–358. https://doi.org/10.1080/10705511.2016.1257353

Frühwirth-Schnatter, S., & Lopes, H. F. (2018). *Sparse Bayesian factor analysis when the number of factors is unknown*. arXiv.org. https://arxiv.org/abs/1804.04231

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

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

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

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

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

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

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179–185. https://doi.org/10.1007/BF02289447

Huang, P.-H. (2018). A penalized likelihood method for multi-group structural equation modelling. *British Journal of Mathematical and Statistical Psychology*, *71*, 499–522. https://doi.org/10.1111/bmsp.12130

Huang, P.-H., Chen, H., & Weng, L.-J. (2017). A penalized likelihood method for structural equation modeling. *Psychometrika*, *82*, 329–354. https://doi.org/10.1007/s11336-017-9566-9

Jacobucci, R., Grimm, K. J., & McArdle, J. J. (2016). Regularized structural equation modeling. *Structural Equation Modeling*, *23*, 555–566. https://doi.org/10.1080/10705511.2016.1154793

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

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

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

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

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

Mulaik, S. A. (2010). *Foundations of factor analysis* (2nd ed.). Chapman & Hall/CRC.

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

Papastamoulis, P. (2020). Clustering multivariate data using factor analytic Bayesian mixtures with an unknown number of components. *Statistics and Computing*, *30*, 485–506. https://doi.org/10.1007/s11222-019-09891-z

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

Piatek, R. (2020). *BayesFM: Bayesian inference for factor modeling (R package version 0.1.4)*. CRAN. https://CRAN.R-project.org/package=BayesFM

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, *6*, 7–11. https://cran.r-project.org/doc/Rnews/Rnews_2006-1.pdf#page=7

R Development Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org

Revelle, W. (2020). *psych: Procedures for psychological, psychometric, and personality research (R package version 2.0.9)*. CRAN. https://CRAN.R-project.org/package=psych

Sass, D. A., & Schmitt, T. A. (2010). A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research*, *45*, 73–103. https://doi.org/10.1080/00273170903504810

Scharf, F., & Nestler, S. (2019). Should regularization replace simple structure rotation in exploratory factor analysis? *Structural Equation Modeling*, *26*, 576–590. https://doi.org/10.1080/10705511.2018.1558060

Simon, N., Friedman, J., Hastie, T., & Tibshirani, R. (2013). A sparse-group Lasso. *Journal of Computational and Graphical Statistics*, *22*, 231–245. https://doi.org/10.1080/10618600.2012.681250

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

Trendafilov, N. T., Fontanella, S., & Adachi, K. (2017). Sparse exploratory factor analysis. *Psychometrika*, *82*, 778–794. https://doi.org/10.1007/s11336-017-9575-8

Van Erp, S., Mulder, J., & Oberski, D. L. (2017). Prior sensitivity analysis in default Bayesian structural equation modeling. *Psychological Methods*, *23*, 363–388. https://doi.org/10.1037/met0000162

Xu, X., & Ghosh, M. (2015). Bayesian variable selection and estimation for group Lasso. *Bayesian Analysis*, *10*, 909–936. https://doi.org/10.1214/14-BA929

Yates, A. (1987). *Multivariate exploratory data analysis: A perspective on exploratory factor analysis*. State Univ. of New York Press.

Yuan, M., & Lin, Y. (2005). Efficient empirical Bayes variable selection and estimation in linear models. *Journal of the American Statistical Association*, *100*, 1215–1225. https://doi.org/10.1198/016214505000000367

Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *68*, 49–67. https://doi.org/10.1111/j.1467-9868.2005.00532.x

Appendix A. Posterior analysis and MCMC estimation

- The posterior distribution of $\mathbf{b}_k$ conditional on everything else is still a multivariate spike and slab distribution:

$$\mathbf{b}_k|\text{rest}\sim(1-l_k)N_J(\boldsymbol{\mu}_k,\Sigma_k)+l_k\delta_0(\mathbf{b}_k), \tag{A1}$$

where $\boldsymbol{\mu}_k=\boldsymbol{\Psi}^{-1}\Sigma_k\mathbf{V}_k^{1/2}\mathbf{Y}^{\mathrm{T}}\mathbf{H}_k$, and $\boldsymbol{\Sigma}_k=\left(\mathrm{I}_J+\boldsymbol{\Psi}^{-1}\mathbf{V}_k^{1/2}(\mathbf{H}_k^{\mathrm{T}}\mathbf{H}_k)\mathbf{V}_k^{1/2}\right)^{-1}$. l_k is the posterior probability of $\mathbf{b}_k=\mathbf{0}$ given the remaining parameters, as:

$$l_k=p(\mathbf{b}_k=\mathbf{0}|\text{rest})=\frac{\pi_0}{\pi_0+(1-\pi_0)|\Sigma_k|^{1/2}\exp\{\mathrm{Tr}[(\boldsymbol{\mu}_k)^{\mathrm{T}}\Sigma_k^{-1}\boldsymbol{\mu}_k]/2\}}. \tag{A2}$$

- The conditional posterior of τ_{jk} is also a spike and slab distribution:

$$\tau_{jk}|\text{rest}\sim(1-m_{jk})N^+(u_{jk},v_{jk}^2)+m_{jk}\delta_0(\tau_{jk}), \tag{A3}$$

where $u_{jk}=\psi_{jj}^{-1}v_{jk}^2(\mathbf{Y}_j-\sum_{k'\neq k}\lambda_{jk'}\mathbf{H}_{k'})\mathbf{H}_k^{\mathrm{T}}$, $v_{jk}^2=\left(\frac{1}{s_k^2}+\psi_{jj}^{-1}\mathbf{H}_k^{\mathrm{T}}\mathbf{H}_kb_{jk}^2\right)^{-1}$, and

$$m_{jk}=p(\tau_{jk}=\mathbf{0}|\text{rest})=\frac{\pi_1}{\pi_1+2(1-\pi_1)(s_k^2)^{-1/2}(v_{jk}^2)^{1/2}\exp\left(\frac{u_{jk}^2}{2v_{jk}^2}\right)\Phi(\frac{u_{jk}}{v_{jk}})}, \tag{A4}$$

with Φ as the standard cumulative normal distribution function.

- The conditional posteriors of π_0 and π_1 continue to be beta distributions:

$$\begin{aligned}\pi_0|\text{rest}&\sim\text{Beta}(\#(\mathbf{b}_k=\mathbf{0})+a_{01},\#(\mathbf{b}_k\neq\mathbf{0})+a_{02})\\ \pi_1|\text{rest}&\sim\text{Beta}\big(\#(\tau_{kj}=0)+a_{11},\#(\tau_{kj}\neq0)+a_{12}\big).\end{aligned} \tag{A5}$$

Alternatively, one can fix π_0 and/or π_1 as.5, and the differences are usually trivial.

- The conditional posterior of s_k^2 is still an inverse gamma distribution:

$$s_k^2|\text{rest}\sim\text{Inv}-\text{Gamma}\left(1+\frac{1}{2}\#(\tau_{jk}=0),r_k+\frac{1}{2}\sum_{j=1}^{J}\tau_{jk}^2\right). \tag{A6}$$

- The conditional posterior of ψ_{jj} is also an inverse gamma distribution:

$$\psi_{jj}|\text{rest}\sim\text{Inv}-\text{Gamma}\left(d_{j1}+\frac{n-1}{2},d_{j2}+\frac{1}{2}(\mathbf{Y}_j-\sum\nolimits_{k=1}^{K}\mathbf{H}_k\boldsymbol{\Lambda}_k^{\mathrm{T}})\right). \tag{A7}$$

- The conditional posterior of $\mathbf{H}$ is still a multivariate normal distribution:

$$\boldsymbol{\omega}_i|\text{rest}\sim N_K\left((\boldsymbol{\Phi}^{-1}+\boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\mathbf{y}_i,(\boldsymbol{\Phi}^{-1}+\boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda})^{-1}\right). \tag{A8}$$

- To obtain $\boldsymbol{\Phi}$, one can first sample from the conditional posterior of $\boldsymbol{\Phi}*$ as Inv $-$ *Wishart*$(\mathbf{S}_n^{-1},q_n)$, where $q_n=q_0+n$ and $\mathbf{S}_n=\mathbf{S}_0+\mathbf{H}^{\mathrm{T}}\mathbf{H}$. Then one can obtain a provisional $\boldsymbol{\Phi}$ as $\boldsymbol{\Phi}=\mathbf{D}^{-1}\boldsymbol{\Phi}*\mathbf{D}^{-1}$, where $\mathbf{D}=\mathrm{diag}(\phi_{11}^*,\cdots,\phi_{KK}^*)^{1/2}$. The provisional $\boldsymbol{\Phi}$ is accepted with a probability of $\min\left(\left(\frac{|\boldsymbol{\Phi}^{(t)}|}{|\boldsymbol{\Phi}^{(t-1)}|}\right)^{(K+1)/2},1\right)$, where t is the current draw and (t– 1) is the previous draw.
- An empirical Bayes Gibbs sampling or Monte Carlo expectation-maximization algorithm (Casella, 2001; Park & Casella, 2008) can be used to estimate r_k. For the t-th expectation-maximization update,

$$r_k^{(t)}=\frac{1}{E_{r_k^{(t-1)}}\left(\frac{1}{s_k^2}|\mathbf{Y}\right)}, \tag{A9}$$

where the posterior expectation of $1/s_k^2$ is estimated from the Gibbs samples based on $r_k^{(t-1)}$.

Appendix B. Humor styles questionnaire

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

Note. F1 = Affiliative humor; F2 = Self-enhancing humor; F3 = Aggressive humor; F3 = Self-defeating humor.