

Efficient Bayesian Model Averaging in Factor Analysis

David B. Dunson

Biostatistics Branch, MD A3-03

National Institute of Environmental Health Sciences

P.O. Box 12233, RTP, NC 27709, USA

dunson1@niehs.nih.gov

Summary. Although factor analytic models have proven useful for covariance structure modeling and dimensionality reduction in a wide variety of applications, a challenging problem is uncertainty in the number of latent factors. This article proposes an efficient Bayesian approach for model selection and averaging in hierarchical models having one or more factor analytic components. In particular, the approach relies on a method for embedding each of the smaller models within the largest possible model. Bayesian computation can proceed within the largest model, while moving between sub-models based on posterior model probabilities. The approach represents a type of parameter expansion, as one always samples within an encompassing model, incorporating extra parameters and latent variables when a smaller model is true. This results in a highly efficient stochastic search factor selection algorithm (SSFS) for identifying good factor models and performing model-averaged inferences. The approach is illustrated using simulated examples and a toxicology application.

Key words: Covariance structure model; Latent variables; Model selection; Mixture model; Number of factors; Shrinkage; Variance components.

1 Introduction

With an increasing need for flexible models for characterizing multivariate data in a wide variety of application areas, interest in factor analytic and latent structure models has increased greatly in recent years outside of traditional social science application areas. For example, latent factor regression models have been used as a dimensionality reduction tool for modeling of sparse covariance structures in genomic applications (West, 2003; Carvalho et al., 2006). In addition, structural equation models and other generalizations of factor analysis are widely useful in epidemiologic studies involving complex health outcomes and exposures (Sanchez et al, 2005). Improvements in Bayesian computation permit the routine implementation of latent factor models via Markov chain Monte Carlo (MCMC) algorithms, and a very broad class of models can be fitted easily using the freely-available software package WinBUGS. The literature on methods for fitting and inferences in latent factor models is vast (for recent books, see Loehlin, 2004, and Thompson, 2004).

In using a factor analytic model for inferences on a covariance structure or for prediction, it is appealing to formally account for uncertainty in selecting the number of factors. There has been some focus in the frequentist and Bayesian literature on the problem of selection of the number of factors. Press and Shigemasu (1999) propose to choose the number of factors having the highest posterior probability, noting that such an approach improves upon the commonly-used AIC (Akaike, 1987) and BIC (Schwarz, 1978) criteria. For hierarchical models, such as latent factor models, the BIC justification as an approximation to the Bayes factor breaks down (Berger, Ghosh and Mukhopadhyay, 2003), and one may need a different penalty for model complexity (Zhang and Kocka, 2004).

If posterior probabilities corresponding to models having different numbers of latent factors can be estimated, one can implement Bayesian model averaging (BMA), which is expected to have better predictive performance compared with analyses based on a single

selected model (Hoeting et al., 1999). However, major challenges include (1) how to choose priors for the factor loadings in the list of models corresponding to different numbers of factors; and (2) how to efficiently and accurately estimate posterior model probabilities. Polasek (1997) considered approaches for estimating posterior probabilities based on separate MCMC analyses of models differing only in the number of factors. In implementing such an approach, it is necessary to estimate the marginal likelihood for each model. Although an estimate of the marginal likelihood is not automatically available from the MCMC output, a number of black box algorithms have been proposed (Chib, 1995; DiCiccio et al., 1997; Gelfand and Dey, 1994; Meng and Wong, 1996).

Lopes and West (2004) proposed a reversible jump MCMC (RJMCMC) algorithm (Green, 1995) to move between models with different numbers of factors, and conducted a thorough comparison with estimators for approximating marginal likelihoods from separate MCMC analyses under each model. In simulation studies, they found that a number of the methods perform poorly relative to RJMCMC and bridge sampling (Meng and Wong, 1996) in terms of proportions of simulations in which the true number of factors is assigned highest posterior probability. A computational challenge in implementing RJMCMC for factor model selection is the difficulty of choosing efficient proposal distributions. Lopes and West (2004) address this problem by constructing proposals using the results of a preliminary MCMC run under each model. Such an approach is highly computationally demanding, becoming infeasible as the sample size and potential number of factors increases. Motivated by this problem, Carvalho et al. (2006) proposed an evolutionary search algorithm, which provides a useful approach for searching for good factor models in high-dimensions.

The approach proposed in this article is quite different, being based on a method for moving between models with different numbers of factors by collapsing. We initially develop a parameter-expanded Gibbs sampler (Liu and Wu, 1999; Gelman et al., 2004) for the model with the largest possible number of factors. This approach effectively generalizes

the approach of Gelman (2005) to the multivariate case. We then move between models stochastically based on posterior model probabilities as the Gibbs sampler proceeds. By treating the larger models as parameter-expanded versions of smaller models, we obtain a highly efficient stochastic search factor selection (SSFS) algorithm. Stochastic search variable selection (SSVS) algorithms (George and McCulloch, 1993; Geweke, 1996) have been widely used in variable selection problems, but are not straightforward to extend to the factor selection problem.

Section 2 defines the model and develops a parameter-expansion approach for efficient posterior computation when the number of factors is known. Section 3 extends this approach to allow unknown number of factors. Section 4 presents the results of a simulation study. Section 5 considers generalizations to more complex factor models. Section 6 contains an application to data from a toxicology study, and Section 7 discusses the results.

2 Bayesian Factor Models

We initially focus on the case in which the number of factors is known to be k , and the model has the following structure:

$$\mathbf{y}_i = \mathbf{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}), \quad (1)$$

where $\mathbf{\Lambda}$ is a $p \times k$ factor loading matrix, $\boldsymbol{\eta}_i = (\eta_{i1}, \dots, \eta_{ik})' \sim N_k(\mathbf{0}, \mathbf{I}_k)$ is a vector of standard normal latent factors, and $\boldsymbol{\epsilon}_i$ is a residual with diagonal covariance matrix $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$. The introduction of the latent factors, $\boldsymbol{\eta}_i$, induces dependency, as the marginal distribution of \mathbf{y}_i is $N_p(\mathbf{0}, \boldsymbol{\Omega})$, with $\boldsymbol{\Omega} = \mathbf{\Lambda}\mathbf{\Lambda}' + \boldsymbol{\Sigma}$. In practice, the number of factors is small relative to the number of outcomes ($k \ll p$). Very small k corresponds to a sparse model for $\boldsymbol{\Omega}$ containing many fewer than $p(p+1)/2$ parameters. For this reason, factor models provide a convenient and flexible framework for modeling of a covariance matrix, particularly in applications with moderate to large p .

For simplicity in exposition, we leave the intercept out of expression (1), though the methods generalize directly to the case in which the intercept is unknown and even depends on predictors. Note that the factor model (1) is subject to a non-identifiability problem. As is well known, one can obtain an identical $\mathbf{\Omega}$ by multiplying $\mathbf{\Lambda}$ by an orthonormal matrix \mathbf{P} defined so that $\mathbf{PP}' = \mathbf{I}_k$. Following a common convention to ensure identifiability, we assume that $\mathbf{\Lambda}$ has a full-rank lower triangular structure. The number of free parameters in $\mathbf{\Lambda}, \mathbf{\Sigma}$ is then $q = p(k + 1) - k(k - 1)/2$, and k must be chosen so that $q \leq p(p + 1)/2$. We let m denote the number of factors in the largest identifiable model.

The lower triangular structure of $\mathbf{\Lambda}$ implies that the model is not invariant to changes in the ordering of the elements of \mathbf{y}_i . However, as argued by Lopes and West (2004), the ordering does not impact inferences or predictions when k is known. In particular, by multiplying \mathbf{y}_i by a rotation matrix \mathbf{A} , one obtains a reordered outcome vector, with modified factor loadings matrix $\mathbf{A}\mathbf{\Lambda}$. By further multiplying by an appropriately-chosen orthonormal matrix \mathbf{P}' , one can always obtain a lower-triangular factor loadings matrix $\mathbf{A}\mathbf{\Lambda}\mathbf{P}'$ for the reordered outcomes, obtaining a model with the same structure as (1). That said, the ordering can impact selection of the number of factors, which makes it all the more important to allow for uncertainty in this selection through Bayesian model averaging, as will be described in Section 3.

To complete a Bayesian specification of model (1), the typical choice specifies truncated normal priors for the diagonal elements of $\mathbf{\Lambda}$, normal priors for the lower triangular elements, and inverse-gamma priors for $\sigma_1^2, \dots, \sigma_p^2$. These choices are convenient, because they represent conditionally-conjugate forms that lead to straightforward posterior computation by a Gibbs sampler (Arminger, 1998; Rowe, 1998; Song and Lee, 2001).

Unfortunately, this choice may lead to a poorly-behaved sampler having substantial slow-mixing problems for a number of reasons. First, as noted by Gelfand et al. (1995), centering parameterizations often lead to substantially better mixing than uncentered parameteriza-

tions. Because the variance of the latent variable is fixed at one, expression (1) is uncentered. One can solve this problem by replacing (1) with the equivalent model

$$\mathbf{y}_i = \mathbf{\Lambda}^* \boldsymbol{\eta}_i^* + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\eta}_i^* \sim N_k(\mathbf{0}, \boldsymbol{\Psi}), \quad \boldsymbol{\epsilon}_i \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}), \quad (2)$$

where $\mathbf{\Lambda}^*$ is $p \times k$ lower triangular with ones on the diagonal. We then have $\mathbf{\Lambda} = \mathbf{\Lambda}^* \boldsymbol{\Psi}^{1/2}$ and $\boldsymbol{\eta}_i = \boldsymbol{\Psi}^{-1/2} \boldsymbol{\eta}_i^*$. Conditionally-conjugate inverse-gamma priors can then be assigned to ψ_1, \dots, ψ_k .

Unfortunately, inverse-gamma densities tend to be a poor choice of prior for variance components in hierarchical models, particularly when one has limited prior information about the variance parameters (Gelman, 2005). In particular, if one chooses $\pi(\psi_h^{-1}) = \mathcal{G}(\epsilon, \epsilon)$, with ϵ a small positive constant, there can be sensitivity of the posterior to the specific ϵ chosen, an improper posterior is obtained in the limit as $\epsilon \rightarrow 0$, and mixing of the Gibbs sampler is often horrendous. To solve this problem in the setting of simple variance component models, Gelman (2005) recommends a *half-t* prior on the standard deviation $\psi_h^{1/2}$. Conveniently, such a prior can be induced through a parameter-expansion approach (Gelman et al., 2004).

We generalize the Gelman prior to the multivariate factor model case by replacing expression (2) with the parameter-expanded model:

$$\mathbf{y}_i = \mathbf{\Lambda}^* \boldsymbol{\Gamma} \boldsymbol{\zeta}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\zeta}_i \sim N_k(\mathbf{0}, \boldsymbol{\Phi}), \quad \boldsymbol{\epsilon}_i \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}), \quad (3)$$

where $\boldsymbol{\Gamma} = \text{diag}(\gamma_1, \dots, \gamma_k)$ and $\boldsymbol{\Phi} = (\phi_1, \dots, \phi_k)$. Note that this model has an extra k parameters, which are not identified. We can collapse to obtain expression (2) by letting $\eta_{ih}^* = |\gamma_h| \zeta_{ih}$ and $\psi_h = \gamma_h^2 \phi_h$, for $h = 1, \dots, k$. Then, assigning γ_h a $N(0, 1)$ prior and choosing an inverse-gamma prior for ϕ_h , we induce a half- t prior on the latent variable standard deviation $\psi_h^{1/2}$, parameterized in terms of a scale τ and degrees of freedom ν . As a reasonable default choice, we recommend letting $\nu = 1$ and $\tau = n$, which results in a half-Cauchy prior (used throughout the remainder of the paper).

A Bayesian specification is completed with normal priors for the free elements of $\mathbf{\Lambda}^*$ and inverse-gamma priors for $\sigma_1^2, \dots, \sigma_p^2$. One can then run an efficient blocked, parameter-expanded and centered Gibbs sampling algorithm for posterior computation. The conditional distributions are described in the appendix. We have founded this algorithm to be highly efficient, in terms of rates of convergence and mixing, in a wide variety of simulated and real data examples.

3 Unknown Number of Factors

3.1 Identifiability issues when k is chosen incorrectly

Section 2 considers prior specification and efficient computation when the number of factors k is known. To set up the framework for allowing unknown k , we first discuss identifiability issues that arise when the true value of k is less than the specified value of k . In particular, consider the example in which $p = 5$ and the true model has $k = 1$ but one fits the model with $k = 2$. Then, letting $\mathbf{\Omega}$ and $\tilde{\mathbf{\Omega}}$ denote the covariance matrices under the $k = 1$ and $k = 2$ models, respectively, we have

$$\mathbf{\Omega} = \begin{pmatrix} \lambda_{11}^2 + \sigma_1^2 & \lambda_{11}\lambda_{21} & \lambda_{11}\lambda_{31} & \lambda_{11}\lambda_{41} & \lambda_{11}\lambda_{51} \\ \lambda_{11}\lambda_{21} & \lambda_{21}^2 + \sigma_2^2 & \lambda_{21}\lambda_{31} & \lambda_{21}\lambda_{41} & \lambda_{21}\lambda_{51} \\ \lambda_{11}\lambda_{31} & \lambda_{21}\lambda_{31} & \lambda_{31}^2 + \sigma_3^2 & \lambda_{31}\lambda_{41} & \lambda_{31}\lambda_{51} \\ \lambda_{11}\lambda_{41} & \lambda_{21}\lambda_{41} & \lambda_{31}\lambda_{41} & \lambda_{41}^2 + \sigma_4^2 & \lambda_{41}\lambda_{51} \\ \lambda_{11}\lambda_{51} & \lambda_{21}\lambda_{51} & \lambda_{31}\lambda_{51} & \lambda_{41}\lambda_{51} & \lambda_{51}^2 + \sigma_5^2 \end{pmatrix}$$

$$\tilde{\mathbf{\Omega}} = \begin{pmatrix} \tilde{\lambda}_{11}^2 + \tilde{\sigma}_1^2 & \tilde{\lambda}_{11}\tilde{\lambda}_{21} & \tilde{\lambda}_{11}\tilde{\lambda}_{31} & \tilde{\lambda}_{11}\tilde{\lambda}_{41} & \tilde{\lambda}_{11}\tilde{\lambda}_{51} \\ \tilde{\lambda}_{11}\tilde{\lambda}_{21} & \tilde{\lambda}_{21}^2 + \tilde{\lambda}_{22}^2 + \tilde{\sigma}_2^2 & \tilde{\lambda}_{21}\tilde{\lambda}_{31} + \tilde{\lambda}_{22}\tilde{\lambda}_{32} & \tilde{\lambda}_{21}\tilde{\lambda}_{41} + \tilde{\lambda}_{22}\tilde{\lambda}_{42} & \tilde{\lambda}_{21}\tilde{\lambda}_{51} + \tilde{\lambda}_{22}\tilde{\lambda}_{52} \\ \tilde{\lambda}_{11}\tilde{\lambda}_{31} & \tilde{\lambda}_{21}\tilde{\lambda}_{31} + \tilde{\lambda}_{22}\tilde{\lambda}_{32} & \tilde{\lambda}_{31}^2 + \tilde{\lambda}_{32}^2 + \tilde{\sigma}_3^2 & \tilde{\lambda}_{31}\tilde{\lambda}_{41} + \tilde{\lambda}_{32}\tilde{\lambda}_{42} & \tilde{\lambda}_{31}\tilde{\lambda}_{51} + \tilde{\lambda}_{32}\tilde{\lambda}_{52} \\ \tilde{\lambda}_{11}\tilde{\lambda}_{41} & \tilde{\lambda}_{21}\tilde{\lambda}_{41} + \tilde{\lambda}_{22}\tilde{\lambda}_{42} & \tilde{\lambda}_{31}\tilde{\lambda}_{41} + \tilde{\lambda}_{32}\tilde{\lambda}_{42} & \tilde{\lambda}_{41}^2 + \tilde{\lambda}_{42}^2 + \tilde{\sigma}_4^2 & \tilde{\lambda}_{41}\tilde{\lambda}_{51} + \tilde{\lambda}_{42}\tilde{\lambda}_{52} \\ \tilde{\lambda}_{11}\tilde{\lambda}_{51} & \tilde{\lambda}_{21}\tilde{\lambda}_{51} + \tilde{\lambda}_{22}\tilde{\lambda}_{52} & \tilde{\lambda}_{31}\tilde{\lambda}_{51} + \tilde{\lambda}_{32}\tilde{\lambda}_{52} & \tilde{\lambda}_{41}\tilde{\lambda}_{51} + \tilde{\lambda}_{42}\tilde{\lambda}_{52} & \tilde{\lambda}_{51}^2 + \tilde{\lambda}_{52}^2 + \tilde{\sigma}_5^2 \end{pmatrix}$$

An observation is that, when the true model has $k = 1$ but we fit the $k = 2$ model, an identifiability problem occurs in that many different values of $\tilde{\mathbf{\Lambda}}$ are consistent with the same $\mathbf{\Omega}$. Although an MLE can still be obtained under the $k = 2$ model, this weak identifiability problem leads to a high variance estimator. In addition, the Gibbs sampler for the $k = 2$

model is poorly behaved, with extreme slow mixing in the chains for the elements of $\tilde{\Lambda}$, particularly if a diffuse prior is chosen. As noted by Lopes and West (2004), this behavior is so extreme that one can select the number of factors by running chains for different k and choosing the largest k such that a well behaved chain is obtained.

Our viewpoint is that, when the $k = 1$ model is true, the $k = 2$ model has too many parameters and can be considered as effectively parameter-expanded. Thus, one can potentially run a Gibbs sampler under the $k = 2$ model but then collapse the samples back to the $k = 1$ model using the transformation:

$$\lambda_{11} = \tilde{\lambda}_{11}, \quad \lambda_{j1} = \mathcal{S}(\tilde{\lambda}_{j1})(\tilde{\lambda}_{j1}^2 + \tilde{\lambda}_{j2}^2)^{1/2}, \quad \sigma_j^2 = \tilde{\sigma}_j^2, \quad j = 2, \dots, 5,$$

where $S(z) = (-1)^{1(z < 0)}$ preserves the sign. Such an approach will provide approximate samples from the posterior distribution of $\boldsymbol{\lambda}, \boldsymbol{\sigma}$ under the $k = 1$ model. The samples will not have exactly the same target distribution as obtained running the Gibbs sampler directly for the $k = 1$ model, because collapsing from the $k = 2$ model induces a different prior. However, if a diffuse prior is chosen, say by placing a half-Cauchy prior on the latent variable standard deviations, then the approaches should produce very similar results.

To illustrate this, we implemented a simulation example in which the true model had $k = 1$, with $\boldsymbol{\lambda} = (0.975, 0.949, 0.922, 0.894, 0.837)'$ and $\text{diag}(\boldsymbol{\Sigma}) = (0.05, 0.1, 0.15, 0.2, 0.3)$, and we let $n = 100$. Two different Gibbs samplers were then implemented separately for 20,000 iterations: (1) the parameter-expanded Gibbs sampler of Section 2 choosing $k = 1$; and (2) the parameter-expanded Gibbs sampler of Section 2 choosing $k = 2$ but then collapsing to the $k = 1$ model. The first 100 iterations of the chains are plotted in Figure 1 along with the true values. There are no clear differences in rates of convergence and mixing or in the target distribution. Discarding a burn-in of 1,000 iterations, posterior summaries are as follows:

Loading	True Value	Approach 1			Approach2		
		Mean	Median	95% CI	Mean	Median	95% CI
λ_1	0.975	0.941	0.938	[0.802,1.103]	0.942	0.938	[0.803,1.103]
λ_2	0.949	0.941	0.937	[0.802,1.100]	0.941	0.937	[0.801,1.103]
λ_3	0.922	0.927	0.923	[0.787,1.088]	0.927	0.923	[0.787,1.090]
λ_4	0.894	0.908	0.904	[0.763,1.070]	0.907	0.907	[0.764,1.075]
λ_5	0.837	0.852	0.849	[0.701,1.023]	0.850	0.846	[0.698,1.019]

Hence, the results are indistinguishable. We have found this effective equivalence to extend to more complex examples and to hold across many simulated data sets.

3.2 Model uncertainty

Based on these results, one can envision an approach in which posterior computation is implemented under the $k = m$ model, with m denoting an upper bound on the number of factors (based on identifiability considerations or user input). Then, to obtain approximate draws from the posterior distribution for any model having $k \leq m$ factors, one can simply collapse the samples using:

$$\lambda_{jh}^{(k)} = \lambda_{jh}^{(m)}, \quad h = 1, \dots, k-1, \quad \lambda_{jk}^{(k)} = \mathcal{S}(\lambda_{jk}^{(m)}) \left(\sum_{h=k}^m \lambda_{jh}^{(m)2} \right)^{1/2}, \quad (4)$$

where the (k) superscript denotes the parameter for the model with k factors.

To allow unknown k , we choose a multinomial prior distribution, with $\Pr(k = h) = \kappa_h$, for $h = 1, \dots, m$. We then complete a Bayesian specification through priors on the coefficients within each of the models in the list $k \in \{1, \dots, m\}$. This is accomplished by choosing a prior for the coefficients in the m factor model having the form described in Section 2, which then automatically induces priors on the coefficients in all smaller models through the collapsing approach.

Bayesian selection of the number of factors relies on posterior model probabilities:

$$\Pr(k = h | \mathbf{y}) = \frac{\kappa_h \pi(\mathbf{y} | k = h)}{\sum_{l=1}^m \kappa_l \pi(\mathbf{y} | k = l)}, \quad (5)$$

where the marginal likelihood under model k , $\pi(\mathbf{y} | k = h)$, is obtained by integrating the likelihood $\prod_i N_p(\mathbf{y}_i; \mathbf{0}, \mathbf{\Lambda}^{(k)} \mathbf{\Lambda}^{(k)'} + \mathbf{\Sigma})$ across the prior for the factor loadings $\mathbf{\Lambda}^{(k)}$ and resid-

ual variances Σ . As is well known, the marginal likelihood and resulting posterior model probabilities can be sensitive to the choice of the prior, particularly for those parameters that vary across the models in the list. Hence, we expect the posterior probabilities to be sensitive to the priors $\pi(\mathbf{\Lambda}^{(k)})$, for $k = 1, \dots, m$.

In general, this sensitivity is problematic in that one typically does not know how the priors should be changed to best express uncertainty in the coefficients as one moves across the model in the list. One natural way to express prior information about the parameters is in terms of the marginal covariance matrix $\mathbf{\Omega}$, which is defined for model k as $\mathbf{\Omega}^{(k)} = \mathbf{\Lambda}^{(k)} \mathbf{\Lambda}^{(k)'} + \Sigma$. Our approach induces a prior on $\mathbf{\Omega}$ for each model in the list by starting with a prior for $\mathbf{\Omega}^{(m)}$ specified through priors for $\mathbf{\Lambda}^{(m)}$ and Σ . An appealing property of the specification is that the resulting prior for the diagonal elements of $\mathbf{\Omega}^{(k)}$ is invariant to $k \in \{1, \dots, m\}$. This is due to the fact that the j th diagonal element of $\mathbf{\Omega}^{(k)}$ equals:

$$\omega_{jj}^{(k)} = \left\{ \sum_{h=1}^{\min(j,m)} \lambda_{jh}^{(m)2} \right\} + \sigma_j^2 = \omega_{jj},$$

which is constant for $k = 1, \dots, m$. The specification also tends to minimize changes in the off-diagonal elements of $\mathbf{\Omega}$ across the different models. In particular, for model k the (j, l) element ($j < l$) of $\mathbf{\Omega}^{(k)}$ equals

$$\omega_{jl}^{(k)} = \begin{cases} \sum_{h=1}^j \lambda_{jh}^{(m)} \lambda_{lh}^{(m)} & \text{for } j < k \\ \left\{ \sum_{h=1}^{k-1} \lambda_{jh}^{(m)} \lambda_{lh}^{(m)} \right\} + \mathcal{S}(\lambda_{jk}^{(m)}) \mathcal{S}(\lambda_{lk}^{(m)}) \left(\sum_{h=k}^m \lambda_{jh}^{(m)2} \right)^{1/2} \left(\sum_{h=k}^m \lambda_{lh}^{(m)2} \right)^{1/2} & \text{for } j \geq k \end{cases}$$

Due to this construction, we expect that sensitivity of the posterior model probabilities to the choice of hyperparameters in the prior for $\mathbf{\Lambda}^{(m)}$, Σ should be minimal. This is evaluated through simulation experiments in Section 4.

We still need to consider the problem of how to estimate $\Pr(k = h | \mathbf{y})$ and obtain model-averaged estimates of the posterior distribution of $\mathbf{\Omega}$. Our approach relies on expressing the posterior model probability (5) as

$$\int \Pr(k = h | \mathbf{y}, \mathbf{\Lambda}^{(m)}, \Sigma) \pi(d\mathbf{\Lambda}^{(m)}, d\Sigma | \mathbf{y}) = \int \frac{\kappa_h \pi(\mathbf{y} | k = h, \mathbf{\Lambda}^{(m)}, \Sigma)}{\sum_{l=1}^m \kappa_l \pi(\mathbf{y} | k = l, \mathbf{\Lambda}^{(m)}, \Sigma)} \pi(d\mathbf{\Lambda}^{(m)}, d\Sigma | \mathbf{y}),$$

where $\pi(\mathbf{\Lambda}^{(m)}, \mathbf{\Sigma} | \mathbf{y})$ is the parameter-expanded posterior distribution, and

$$\pi(\mathbf{y} | k = h, \mathbf{\Lambda}^{(m)}, \mathbf{\Sigma}) = \prod_i N_p(\mathbf{y}_i; \mathbf{0}, \mathbf{\Lambda}^{(k)} \mathbf{\Lambda}^{(k)'} + \mathbf{\Sigma}),$$

where $\mathbf{\Lambda}^{(k)}$ is calculated from $\mathbf{\Lambda}^{(m)}$ using (4). As draws from $\pi(\mathbf{\Lambda}^{(m)}, \mathbf{\Sigma} | \mathbf{y})$ can be obtained easily using the parameter-expanded Gibbs sampler, it is trivial to obtain a Monte Carlo estimate of $\Pr(k = h | \mathbf{y})$ by calculating $\pi(\mathbf{y} | k = h, \mathbf{\Lambda}^{(m)}, \mathbf{\Sigma})$, for $h = 1, \dots, m$, at each iteration, normalizing across the models after weighting by the prior probabilities, and then averaging across the Gibbs iterates. By instead sampling from the multinomial conditional posterior distribution of k given $\mathbf{\Lambda}^{(m)}, \mathbf{\Sigma}$ at each iteration and calculating $\mathbf{\Omega}$ under the chosen model, we can obtain draws from the model-averaged posterior for the covariance matrix $\mathbf{\Omega}$.

Note that this approach is clearly much more computationally efficient than the RJMCMC approach of Lopes and West (2004), which requires running a separate MCMC algorithm for $k = 1, \dots, m$ in order to construct efficient proposals in the final RJMCMC run. Instead we can rely on a single run under $k = m$, and then conduct a simple post-processing procedure. In addition, the approach automatically defines a prior in a manner, which minimizes differences in the marginal covariance across models with different numbers of factors. As we illustrate in the next section, this results in a procedure with excellent performance.

4 Simulation Study

4.1 One factor case

We first compared the results of our approach to the simulation results reported in Lopes and West (2004) for their RJMCMC approach and a number of other methods for approximating posterior model probabilities. In their first simulation case, they let $p = 7, n = 100$, with the true model having $k = 1$ and

$$\mathbf{\Lambda} = (0.995, 0.975, 0.949, 0.922, 0.894, 0.866, 0.837)'$$

$$\text{diag}(\mathbf{\Sigma}) = (0.01, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30). \quad (6)$$

As shown in Table 1 of Lopes and West (2004), the RJMCMC assigned the true model ($k = 1$) highest posterior probability in 1000/1000 simulations, while several of the approximation approaches had poor performance.

We repeated this simulation exercise for 100 simulated data sets analyzed using our proposed approach. To specify the prior, we chose half-Cauchy priors for the latent variable standard deviations, $\psi_h^{1/2}$, $\mathcal{G}(1, 1)$ priors for the residual precisions, σ_j^{-2} , and $N(0, 1)$ priors for the lower triangular elements of $\mathbf{\Lambda}^*$. In addition, we let $\log \kappa_h \propto -hn$ in order to favor sparser factor models. For each simulated data set, we ran the Gibbs sampler for 5,000 iterations, discarding the first 1,000 iterations as a burn-in. Based on examination of trace plots across the simulations, apparent convergence occurred very rapidly and mixing was good. Interestingly, the mixing tended to be good only after collapsing to the chosen model. The extra parameters in the models having too many factors mixed poorly, which is as expected and consistent with the observations of Lopes and West (2004). This does not present a problem to our approach, as one expects the extra parameters to have poor mixing in a parameter-expanded Gibbs sampler. The good performance is seen only after transforming back to the original parameterization.

Although we expected good performance given that many of the procedures considered by Lopes and West (2004) assigned highest posterior probability to the true model in 100% of the simulations, the approach did amazingly well. In fact, the estimated posterior probability assigned to the true model ($k = 1$) was 1.0 in all 100 of the simulated data sets. This does not represent poor mixing between models, as that is not an issue with our approach. To assess estimation bias, we plot the posterior mean factor loading matrix in each of the simulations versus the true value in Figure 2. The estimates are distributed about the true value, suggesting minimal bias. The only exception is for λ_1 , which is slightly under-estimated,

most likely because the true value of $\sigma_1^{-2} = 100$ is assigned low value under the $\mathcal{G}(1, 1)$ prior, leading to slight underestimation of σ_1^{-2} .

4.2 Three factor case

We also compared our results to the three factor case considered by Lopes and West (2004). This case is more challenging in that $p = 9$, the sample size is only $n = 50$ and $k = 3$. The true parameter values were

$$\Lambda' = \begin{pmatrix} 0.99 & 0.00 & 0.00 & 0.99 & 0.99 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.95 & 0.00 & 0.00 & 0.00 & 0.95 & 0.95 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.90 & 0.00 & 0.00 & 0.00 & 0.00 & 0.90 & 0.90 \end{pmatrix}$$

$$\text{diag}(\Sigma) = (0.02, 0.19, 0.36, 0.02, 0.02, 0.19, 0.19, 0.36, 0.36).$$

The procedure of Lopes and West (2004) identified the correct model 993/1000 times.

We implemented our approach exactly as in the first simulation case, and assigned highest posterior probability to the correct $k = 3$ model in 100/100 of the simulations. In addition, as illustrated in Table 1, the posterior probability assigned to the true model was close to one in each of the simulations. Given that there is only 50 samples, this result is quite surprisingly. Figure 3 plots the posterior mean factor loading matrix in each of the simulations versus the true value. Again, the estimates are distributed about the true value.

Given the performance, a concern is that the cases considered by Lopes and West (2004) represented scenarios in which model selection is particularly easy. For example, the $k = 3$ case suggests a tight clustering of the outcomes into three groups $\{y_{i1}, y_{i4}, y_{i5}\}$, $\{y_{i2}, y_{i6}, y_{i7}\}$ and $\{y_{i3}, y_{i8}, y_{i9}\}$, and the loadings are all non-negative.

Therefore, we repeated the simulation for a case in which there were $p = 26$ outcomes, $n = 500$, $k = 3$ factors, and some of the loadings were negative. We set $m = 10$, and otherwise ran the simulation exactly as in the other cases. Summary statistics of the probabilities assigned in each choice of k are shown in Table 2. The $k = 3$ model was correctly chosen in 99/100 simulations. In addition, the posterior means of the factor loadings under the selected

model are shown for each simulation in Figure 4. The bias was low and the posterior means were close to the true values on average.

5 Generalizations

5.1 Mixed categorical and continuous

Although we have focused on the normal, linear factor model (1), the method generalizes trivially to many other cases. For example, it is often of interest to model the joint distribution of mixed categorical and continuous variables using a factor model. For example, suppose that $\mathbf{y}_i = (\mathbf{y}'_{i1}, \mathbf{y}'_{i2})'$, with \mathbf{y}_{i1} a $p_1 \times 1$ vector of continuous variables and \mathbf{y}_{i2} a $p_2 \times 1$ vector of ordered categorical variables having L_j levels, for $j = 1, \dots, p$. Then, we can consider the following generalization of (1):

$$\begin{aligned} y_{ij} &= h_j(y_{ij}^*; \boldsymbol{\tau}_j), \quad \text{for } j = 1, \dots, p \\ \mathbf{y}_i^* &= \boldsymbol{\alpha} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\eta}_i \sim N_k(\mathbf{0}, \mathbf{I}_k), \quad \boldsymbol{\epsilon}_i \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}), \end{aligned} \quad (7)$$

where $h_j(\cdot)$ is the identity link, for $j = 1, \dots, p_1$, and is a threshold link:

$$h_j(z; \boldsymbol{\tau}_j) = \sum_{c=1}^{L_j} c \mathbf{1}(\tau_{j,c-1} \leq z \leq \tau_{jc}), \quad \text{for } j = p_1 + 1, \dots, p,$$

with $p = p_1 + p_2$. Here, y_{ij}^* is an underlying normal variable, so that probit-type models are assigned to the categorical items. For identifiability, the $p_1 + 1, \dots, p$ diagonal elements of $\boldsymbol{\Sigma}$ are set equal to one.

The approach described in Sections 2 and 3 can be implemented with the modifications:

1. In the Gibbs sampler, the underlying y_{ij}^* for the categorical items are imputed by sampling from their truncated normal full conditional distributions (following Albert and Chib, 1993). The y_{ij} 's are then replaced with y_{ij}^* 's in the other updating steps.

2. Choosing a normal prior for $\boldsymbol{\alpha}$, these intercepts can be updated from their multivariate normal full conditional posterior in implementing the Gibbs sampler.
3. The threshold parameters $\boldsymbol{\tau}$ need to be assigned a prior and updated, which can proceed via a Metropolis-Hastings step.

5.2 Selecting loadings structure

Although we have considered selection of the number of factors k , there is an additional model selection question, which is often of interest in factor analysis, pertaining to selection of the lower triangular elements of $\mathbf{\Lambda}$ that can be set to zero. In the setting of extremely-high dimensional data, Carvalho et al (2006) addressed this approach by using a variable selection-type mixture prior for the lower triangular elements, with the mixture consisting of a point mass at zero and a normal component. Similar priors have been widely used in variable selection applications (George and McCulloch, 1993; Geweke, 1996). Such priors could be applied directly to the lower triangular elements of $\tilde{\mathbf{\Lambda}}$ in place of the normal priors considered above. As the mixture priors are conjugate, one could still implement the Gibbs sampling algorithm after appropriately modifying the conditional posterior distributions of the free elements of $\tilde{\mathbf{\Lambda}}$ in a straightforward manner.

6 Application

We illustrate the approach through application to organ weight data from a U.S. National Toxicology Program (NTP) 13 week study of Anthraquinone in female Fischer rats. The goal of the NTP 13 week studies is to assess the short term toxicological effects of test agents on a variety of outcomes, including animal and organ body weights. Studies are routinely conducted with 60 animals randomized to six dose groups, including a vehicle control. In the Anthraquinone study, doses included 0, 1875, 3750, 7500, 15000 and 30000 ppm.

At the end of the study, animals are sacrificed and a necropsy is conducted, with overall body weight obtained along with weights for the heart, liver, lungs, kidneys (combined) and thymus. Although body and organ weights are clearly correlated, a challenge in the analysis of these data is the dimensionality of the covariance matrix. In particular, even assuming a constant covariance across dose groups, it is still necessary to estimate $p(p + 1)/2 = 21$ covariance parameters using data from only $n = 60$ animals. Hence, routine analyses rely on univariate approaches applied separately to body weight and the different organ weights.

An alternative is to use a factor model to reduce dimensionality, but it is not clear whether it is appropriate to assume a single factor underlying the different weights or if additional factors need to be introduced. To address this question using the Anthraquinone data, we repeated the approach described in Sections 2 and 3 in the same manner as implemented in the simulation examples, though we ran the algorithm for 15,000 iterations. The maximum possible number of factors was $m = 3$. Body weights were normalized within each dose group prior to analysis for purposes of studying the correlation structure.

The probability assigned to the one factor model was 0.999, suggesting that one factor is sufficient. The estimated factor loadings are as follows:

Weight	Parameter	Mean	95% CI
body	λ_1	0.800	[0.555,1.081]
heart	λ_2	0.377	[0.064,0.739]
liver	λ_3	0.590	[0.301,0.922]
lungs	λ_4	0.480	[0.107,0.882]
kidneys	λ_5	0.724	[0.724,0.458]
thymus	λ_6	0.503	[0.202,0.850]

Given the standardization, these estimates suggest that body weight and kidney weight are the best surrogates for the latent factor, while most of the variability in heart weight is idiosyncratic. Supporting this, we have $\Pr(\lambda_1 > \lambda_2 | \text{data}) = 0.99$ and $\Pr(\lambda_5 > \lambda_2 | \text{data}) = 0.97$.

These results suggest that a reasonable strategy for analysis of body and organ weight data from short-term toxicology studies is to assume a one factor model, with the factor

potentially dependent on dose of exposure.

7 Discussion

This article has proposed an efficient and easy to implement approach for allowing for uncertainty in the number of factors in Bayesian models having one or more factor analytic components. Being essentially no more difficult to implement than a Gibbs sampler for a single factor model, this approach should be practical to implement routinely. There are a number of settings in which the method may be particularly useful.

First, in analyzing high-dimensional, or even moderate-dimensional, multivariate data, one is faced with the problem of estimating a large number of covariance parameters. The factor model provides a convenient dimensionality-reduction technique, and one can solve the problem of uncertainty in the number of factors using the approach proposed in this article, with a prior that favors sparse models. This would represent an alternative to previous methods for Bayesian shrinkage estimation of covariance matrices (Barnard et al., 2000; Daniels and Kass, 2001). By using a factor model to describe the correlation in underlying normal variables, one can easily apply this approach to multivariate categorical or mixed categorical and continuous data.

Another setting in which a similar dimensionality reduction technique is needed is in studying the effects of many, correlated predictors. For example, in the Agricultural Health Study, investigators are interested in the health effects of farm exposures, including different pesticides, insecticides, and herbicides. There are many classes of pesticides and different chemicals within each class, with the doses of exposure correlated within individuals. Thus, one faces problems in putting all these exposures into a regression model. An alternative is to assume that there are a few latent factors underlying the exposures, and one can then use the methods proposed in this article to allow for uncertainty in the number of factors in

performing model-averaged inferences on the pesticide effects.

Because it is straightforward to incorporate multiple factor analytic components within one model, one can potentially use the method for both predictors and outcomes. For example, this would be relevant to studying pesticide effects on neurological health in the Agricultural Health Study using a structural equation model.

Appendix: Steps in Parameter-Expanded Gibbs Sampler

Letting $\tilde{\Lambda} = \Lambda^* \Gamma$, model (3) can be re-expressed as

$$y_{ij} = \mathbf{z}'_{ij} \tilde{\lambda}_j + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma_j^2),$$

where $\mathbf{z}_{ij} = (\zeta_{i1}, \dots, \zeta_{im_j})'$, $\tilde{\lambda}_j = (\tilde{\Lambda}_{j1}, \dots, \tilde{\Lambda}_{jm_j})'$ denotes the free elements of row j of $\tilde{\Lambda}$, and $m_j = \min(j, m)$ is the number of free elements. Then, letting $\pi(\tilde{\lambda}_j) = N_{m_j}(\tilde{\lambda}_{0k}, \Sigma_{\tilde{\lambda}_j})$ denote the prior for $\tilde{\lambda}_j$, the posterior distribution of $\tilde{\lambda}_j$ is

$$\pi(\tilde{\lambda}_j | \zeta, \Psi, \Sigma, \mathbf{y}) = N_{m_j} \left(\left(\Sigma_{\tilde{\lambda}_j} + \sigma_j^{-2} \mathbf{Z}'_j \mathbf{Z}_j \right)^{-1} \left(\Sigma_{\tilde{\lambda}_j}^{-1} \tilde{\lambda}_{0k} + \sigma_j^{-1} \mathbf{Z}'_j \mathbf{Y}_j \right), \left(\Sigma_{\tilde{\lambda}_j} + \sigma_j^{-2} \mathbf{Z}'_j \mathbf{Z}_j \right)^{-1} \right),$$

where $\mathbf{Z}_j = (\mathbf{z}_{1j}, \dots, \mathbf{z}_{nj})'$ and $\mathbf{Y}_j = (y_{1j}, \dots, y_{nj})'$.

With the model in this hierarchical, normal linear structure it is straightforward to derive the remaining conditional posterior distributions used for Gibbs sampling:

$$\begin{aligned} \pi(\zeta_i | \tilde{\Lambda}, \Sigma, \Psi, \mathbf{y}) &= N_m \left((\Psi^{-1} + \tilde{\Lambda}' \Sigma^{-1} \tilde{\Lambda})^{-1} \tilde{\Lambda}' \Sigma^{-1} \mathbf{y}_i, (\Psi^{-1} + \tilde{\Lambda}' \Sigma^{-1} \tilde{\Lambda})^{-1} \right), \\ \pi(\phi_h^{-1} | \zeta, \tilde{\Lambda}, \Sigma, \mathbf{y}) &= \mathcal{G} \left(a_h + \frac{n}{2}, b_h + \frac{1}{2} \sum_{i=1}^n \zeta_{ih}^2 \right), \\ \pi(\sigma_j^{-2} | \zeta, \tilde{\Lambda}, \Psi, \mathbf{y}) &= \mathcal{G} \left(c_j + \frac{n}{2}, b_j + \frac{1}{2} \sum_{i=1}^n (y_{ij} - \mathbf{z}'_{ij} \tilde{\lambda}_j)^2 \right), \end{aligned}$$

where $\mathcal{G}(a_h, b_h)$ is the prior for ϕ_h^{-1} , for $h = 1, \dots, m$, and $\mathcal{G}(c_j, d_j)$ is the prior for σ_j^{-2} , for $j = 1, \dots, p$.

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika* **52**, 317-332.

- Albert, J.H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**, 669-679.
- Arminger, G. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika* **63**, 271-300.
- Barnard, J., McCulloch, R. and Meng, X.L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with applications to shrinkage. *Statistica Sinica* **10**, 1281-1311.
- Berger, J.O., Ghosh, J.K. and Mukhopadhyay, N. (2003). Approximation and consistency of Bayes factors as model dimension grows. *Journal of Statistical Planning and Inference* **112**, 241-258.
- Carvalho, C., Lucas, J., Wang, Q., Nevins, J. and West, M. (2005). High-dimensional sparse factor models & latent factor regression. *Discussion Paper 2005-15*, Institute of Statistics and Decision Sciences, Duke University.
- Chib, S. (1995). Marginal likelihoods from the Gibbs output. *Journal of the American Statistical Association* **90**, 1313-1321.
- Daniels, M.J. and Kass, R.E. (2001). Shrinkage estimators for covariance matrices. *Biometrics* **57**, 1173-1184.
- DiCiccio, T.J., Kass, R., Raftery, A. and Wasserman, L. (1997). Computing Bayes factors by combining simulations and asymptotic approximations. *Journal of the American Statistical Association* **92**, 903-915.
- Gelfand, A.E. and Dey, D.K. (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society* **B**, 501-514.

- Gelfand, A.E., Sahu, S.K. and Carlin, B.P. (1995). Efficient parameterisations for normal linear mixed models. *Biometrika* **82**, 479-488.
- Gelman, A., Huang, Z., van Dyk, D. and Boscardin, W.J. (2004). Transformed and parameter-expanded Gibbs samplers for multilevel and generalized linear models. Technical report, Department of Statistics, Columbia University.
- George, E.I. and McCulloch, R.E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association* **88**, 881-889.
- Geweke, J. (1996). Variable selection and model comparison in regression. in J.O. Berger, J.M. Bernardo, A.P. Dawid, and A.F.M. Smith (eds.), *Bayesian Statistics 5*. Oxford: Oxford University Press.
- Green, P.J. (1995). Reversible jump Markov chain Monte Carlo and Bayesian model determination. *Biometrika* **82**, 711-732.
- Hoeting, J., Madigan, D., Raftery, A. and Volinsky, C. (1999). Bayesian model averaging. *Statistical Science* **14**, 382-401.
- Liu, J. and Wu, Y.N. (1999). Parameter expansion for data augmentation. *Journal of the American Statistical Association* **94**, 1264-1274.
- Loehlin, J.C. (2004). Latent Variable Models: An Introduction to Factor, Path and Structural Equation Analysis. *Lawrence Erlbaum Assoc Inc*.
- Lopes, H.F. and West, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica* **14**, 41-67.
- Meng, X.L. and Wong, W.H. (1996). Simulating ratios of normalising constants via a simple identity. *Statistica Sinica* **11**, 552-586.

- Press, S.J. and Shigemasu, K. (1999). A note on choosing the number of factors. *Communications in Statistics - Theory and Methods* **28**, 1653-1670.
- Rowe, D.B. (1998). Correlated Bayesian factor analysis. Ph.D. Thesis, Department of Statistics, University of California, Riverside, CA.
- Sanchez, B.N., Budtz-Jorgensen, E., Ryan, L.M. and Hu, H. (2005). Structural equation models: A review with applications to environmental epidemiology. *Journal of the American Statistical Association* **100**, 1442-1455.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**, 461-464.
- Song, X.Y. and Lee, S.Y. (2001). Bayesian estimation and test for factor analysis model with continuous and polytomous data in several populations. *British Journal of Mathematical & Statistical Psychology* **54**, 237-263.
- Thompson, B. (2004). Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications. APA Books.
- West, M. (2003). Bayesian factor regression models in the “large p, small n” paradigm. *Bayesian Statistics* **7**, J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith and M. West (eds). Oxford University Press.
- Zhang, N.L. and Kocka, T. (2004). Effective dimensions of hierarchical latent class models. *Journal of Artificial Intelligence Research* **21**, 1-17.

Table 1*Summary statistics of posterior probabilities for each number of factors (k)**across 100 simulated data sets in case 2*

k	mean	median	95% interval
1	0.000	0.000	[0.000, 0.000]
2	0.009	0.000	[0.000, 0.089]
3	0.990	0.999	[0.911, 1.000]
4	0.001	0.000	[0.000, 0.001]
5	0.000	0.000	[0.000, 0.000]

Table 2*Summary statistics of posterior probabilities for each number of factors (k)**across 100 simulated data sets in case 3*

k	mean	median	95% interval
1	0.014	0.003	[0.000, 0.102]
2	0.123	0.081	[0.005, 0.606]
3	0.864	0.907	[0.391, 0.994]
4	.000	0.000	[0.000, 0.000]
5	.000	0.000	[0.000, 0.000]
6	.000	0.000	[0.000, 0.000]
7	.000	0.000	[0.000, 0.000]
8	.000	0.000	[0.000, 0.000]
9	.000	0.000	[0.000, 0.000]
10	.000	0.000	[0.000, 0.000]

Figure Captions:

1. First 100 iterations of the parameter expand Gibbs sampler for a simulated data case with $p = 5$ and $k = 1$ using algorithms in which one samples (1) under the $k = 1$ model (dashed lines); or (2) under the $k = 2$ model and then collapses (solid line).
2. Posterior means of the factor loadings in each of the 100 simulated data sets under case 1. Solid line shows true values, while dashed line shows averaging across the simulations.
3. Posterior means of the factors loadings in each of the 100 simulated data sets under case 2. Solid lines show true values, while dashed lines show averages across simulations.
4. Posterior means of the factors loadings in each of the 100 simulated data sets under case 3. Solid lines show true values, while dashed lines show averages across simulations.







