Fitting Structural Equation Models via Variational Approximations

Khue-Dung Dang

School of Mathematics and Statistics
University of Melbourne
kd.dang@unimelb.edu.au

Coauthor: Luca Maestrini, The Australian National University

Structural equation models

- Structural equation models (SEMs) are commonly used in social and behavioral sciences to study the structural relationship between manifest (observed) variables, such as test scores or answers from a questionnaire, and latent constructs (unobservable) variables

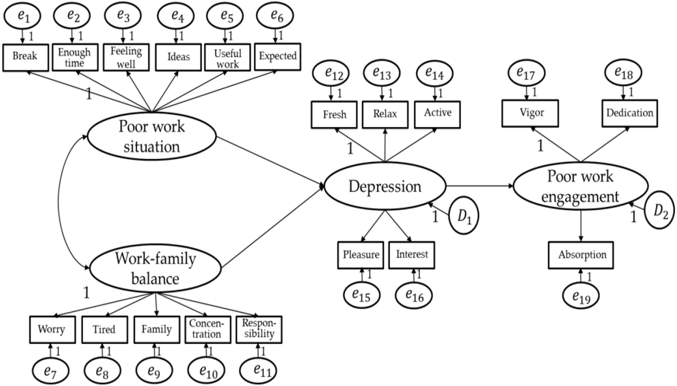


Figure 1: Path diagram of a SEM (figure from Shin and Jeong (2021))

Estimation of SEMs

- Maximum likelihood estimation and weighted least squares estimation of SEMs have already been implemented in software such as Mplus, OpenMX and the package *lavaan* in R
  - Classical methods for SEMs typically analyze the sample covariance matrix (Lee, 2007)
  - May encounter computational and theoretical problems when the sample size is small or the data is non-normal
- Bayesian estimation of SEMs (Ansari and Jedidi, 2000; Ansari et al., 2000; Dunson, 2000) has gotten more attention recently
  - Works better with small data sets
  - Facilitates more flexible model structures

Bayesian estimation of SEMs

- In most research on Bayesian SEMs, Markov chain Monte Carlo (MCMC) is used to fit the model
- MCMC for SEMs typically suffers from slow convergence and long running time, compared to the frequentist approach
- Variational inference methods for models presenting similarities with SEMs have also been developed (Ghahramani and Beal, 1999; Ghahramani and Beal, 2000; Zhao and Philip, 2009; Khan et al., 2010; Klami, 2015), but a fully Bayesian framework that considers models with correlated factors which may also be dependent on covariates is not available

Bayesian estimation of SEMs

- Tiwari(2016) proposed a fast method to fit a special case of latent growth curve models by variational approximations
- The current literature on factor analysis and SEMs does not include studies concerning the accuracy of variational inference, nor does it provide solutions to overcome low accuracy issues
- In this paper, we introduce a mean field Variational Bayes (MFVB) approach to fit elemental SEMs and a strategy to improve the accuracy of MFVB by nonparametric bootstrap

Confirmatory factor analysis (CFA)

- We first consider a confirmatory factor analysis (CFA) model with one latent factor, which is one of the basic SEMs
- Let $\boldsymbol{y}_i$ be the observed outcome of individual $i, \quad i = 1, \ldots, n$

$$
\begin{gathered}
\boldsymbol{y}_i \,|\, \boldsymbol{\nu}, \boldsymbol{\lambda}, \eta_i, \boldsymbol{\psi} \overset{\text{ind.}}{\sim} N\big(\boldsymbol{\nu} + \boldsymbol{\lambda}\eta_i, \operatorname{diag}(\boldsymbol{\psi})\big), \;\; \eta_i \,|\, \sigma^2 \overset{\text{ind.}}{\sim} N(0, \sigma^2), \\
\lambda_j \,|\, \psi_j \overset{\text{ind.}}{\sim} N(\mu_\lambda, \sigma_\lambda^2 \psi_j), \;\; j = 2, \ldots, m. \\
\nu_j \overset{\text{ind.}}{\sim} N(0, \sigma_\nu^2), \;\; \psi_j \overset{\text{ind.}}{\sim} \text{Inverse-}\chi^2(\kappa_\psi, \delta_\psi), \;\; j = 1, \ldots, m, \\
\sigma^2 \sim \text{Inverse-}\chi^2(\kappa_{\sigma^2}, \delta_{\sigma^2})
\end{gathered} \tag{1}
$$

Mean field Variational Bayes (MFVB)

- Bayesian inference for statistical models requires the computation of the posterior distribution of the parameter vector $\boldsymbol{\theta}$

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

- The idea of MFVB is to find an approximating density q for which the Kullback-Leibler divergence between q and the posterior density function $p(\boldsymbol{\theta}|\boldsymbol{y})$ is minimized
- If the approximating density q is factorized according to a partition $(\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_K})$ of $\boldsymbol{\theta}$ such that $q(\boldsymbol{\theta}) = \prod_{k=1}^{K} q(\boldsymbol{\theta_k})$, then the optimal approximating densities satisfy

$$q^*(\boldsymbol{\theta_k}) \propto \exp[E_{q(\boldsymbol{\theta}\backslash\boldsymbol{\theta_k})}\{\log p(\boldsymbol{\theta_k}|\boldsymbol{y}, \boldsymbol{\theta}\backslash\boldsymbol{\theta_k})\}], \quad k = 1, \ldots, K \qquad (2)$$

Fitting SEM with mean field Variational Bayes

- For our problem, we restrict the approximating density to the following product of densities:

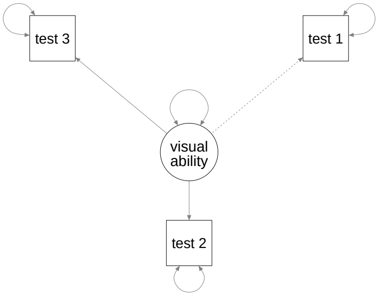
$$q(\boldsymbol{\nu}, \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\psi}, \sigma^2) = q(\boldsymbol{\nu})q(\boldsymbol{\lambda})q(\boldsymbol{\psi}) \prod_{i=1}^{n} q(\eta_i)q(\sigma^2) \tag{3}$$

- For this choice of priors for $\boldsymbol{\nu}, \boldsymbol{\lambda}, \boldsymbol{\eta}, \psi, \sigma^2$, we can easily derive the full conditional density functions for ν_j, λ_j, ψ_j for $j = 1, \ldots, m$; σ^2 and η_i for $i = 1, \ldots, n$
- Applying (2) to the conditional density functions, we can derive the MFVB approximations for $\boldsymbol{\nu}, \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\psi}, \sigma^2$

Holzinger & Swineford (1939) data

- The classic Holzinger and Swineford (1939) data set (Holzinger and Swineford, 1939) consists of mental ability test scores of 301 students from Pasteur and Grant-White Schools

- For illustration purpose, we only consider the first 3 outcomes: visual perception test, cubes test and lozenges test. These outcomes are hypothesized to be associated with the participant's visual ability



Holzinger & Swineford (1939) data

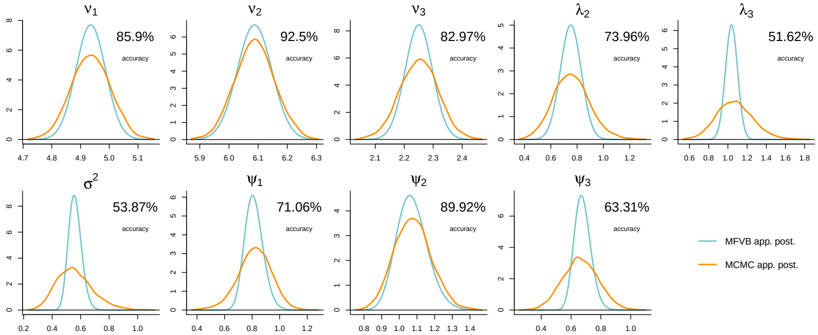


Figure 2: Approximate marginal posterior densities of the parameters of model (1) applied to the Holzinger & Swineford (1939) data. The curves are obtained via MFVB (light blue) and MCMC (orange). MFVB converged after 98 iterations and took 0.1 seconds while MCMC took 307 seconds

Improving performance by bootstrap

- Usage of bootstrap in conjuction with variational inference has been discussed in the literature (Chen et al., 2018) but has not been widely used
- Here we study the use of nonparametric bootstrap to enhance the accuracy of MFVB when it is used for fitting SEMs
- In each bootstrap iteration, we sample with replacement from the original dataset, then recompute the variational estimates of parameters
- The empirical distribution of these bootstrapped estimates or other quantities related to them are used to derive uncertainty measures

Improving performance by bootstrap

Holzinger & Swineford (1939) data

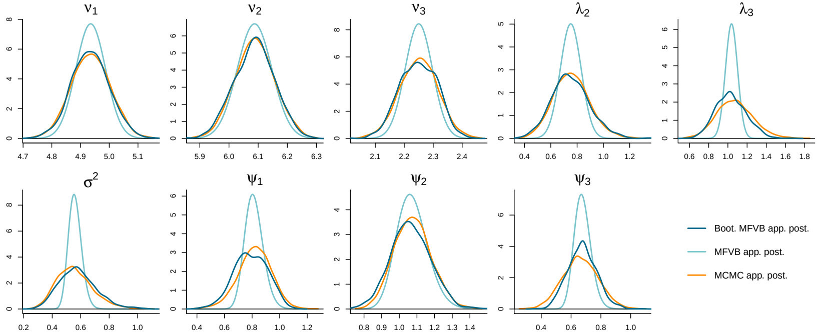


Figure 3: Approximate marginal posterior densities of the parameters of model (1) applied to the Holzinger & Swineford (1939) data. The curves are obtained via kernel density estimation applied to the MFVB point estimates from 1,000 bootstrap samples (navy blue), simple MFVB (light blue) and MCMC (orange). MFVB with 1,000 bootstrap iterations takes 43 seconds

Bootstrap credible intervals

To obtain a **percentile bootstrap credible interval** for a generic SEM parameter θ

- Find the variational inference estimate $\hat{\theta}_{\text{VI}}$ of θ using the original dataset
- For $b = 1, \ldots, B$:
  - Sample with replacement n vectors from $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ to form the bth bootstrap dataset
  - Find the variational inference estimate $\hat{\theta}^{(b)}_{\text{B-VI}}$ of θ using the bth bootstrap dataset
  - Calculate $\delta^{(b)}_{\text{B-VI}} = \hat{\theta}^{(b)}_{\text{B-VI}} - \hat{\theta}_{\text{VI}}$
- Given a credible level α, compute the $\alpha/2$ and $1-\alpha/2$ quantiles of the empirical distribution of $\delta^{(1)}_{\text{B-VI}}, \ldots, \delta^{(B)}_{\text{B-VI}}$, $q^{per}_{\alpha/2}$ and $q^{per}_{1-\alpha/2}$
- The credible interval is given by:

$$\left[\hat{\theta}_{\text{VI}} + q^{per}_{\alpha/2}, \;\; \hat{\theta}_{\text{VI}} + q^{per}_{1-\alpha/2}\right]$$

Simulated data study

We do a simulation exercise to examine the effect of data resampling on the coverage performances of MFVB credible intervals for the simple SEM parameters of interest

- We simulate 1,000 data sets using the MCMC point estimates of the parameters of model (1) fitted using the Holzinger & Swineford data
- For each of the 1,000 simulated datasets we ran MCMC and simple MFVB, and implemented the bootstrap procedure with $B =$ 100, 500 or 1,000
- We construct the 95% percentile credible intervals for each parameter of interest
- For each parameter, the percentage of coverage is calculated as the proportion of simulations where the true parameter falls inside a 95% credible interval

Simulated data study

Results

We report the computational times of simple MFVB and MCMC from the simulation study. In this case MFVB would still provide accurate inference at reduced time even when running 1,000 bootstrap iterations in a non-parallelized system

	1st quartile	median	3rd quartile
MFVB	0.038	0.048	0.060
MCMC	205.3	240.5	287.4

Table 1: Computational times in seconds of MFVB (no bootstrap) and MCMC from the simulation study

Simulated data study

Results

	ν_1	ν_2	ν_3	λ_2	λ_3	σ^2	ψ_1	ψ_2	ψ_3
MCMC	**0.941**	**0.946**	**0.939**	**0.955**	**0.944**	**0.949**	**0.944**	**0.968**	**0.952**
MFVB	**0.851**	**0.914**	**0.839**	**0.776**	**0.543**	**0.552**	**0.735**	**0.930**	**0.679**
MFVB with jackknife	0.940	0.943	0.939	0.951	0.890	0.949	0.943	0.970	0.935
MFVB with per. boot. (100)	0.917	0.926	0.922	0.936	0.894	0.913	0.895	0.950	0.922
MFVB with per. boot. (500)	0.938	0.945	0.940	0.955	0.904	0.936	0.918	0.959	0.937
MFVB with per. boot. (1,000)	**0.941**	**0.947**	**0.940**	**0.957**	**0.905**	**0.938**	**0.925**	**0.958**	**0.940**

Table 2: Performance comparisons for the simulation study. For each fitting method and parameter, the average and sample standard deviation (in brackets) of percentages of coverage over the 1000 simulations are displayed

A model with multiple factors

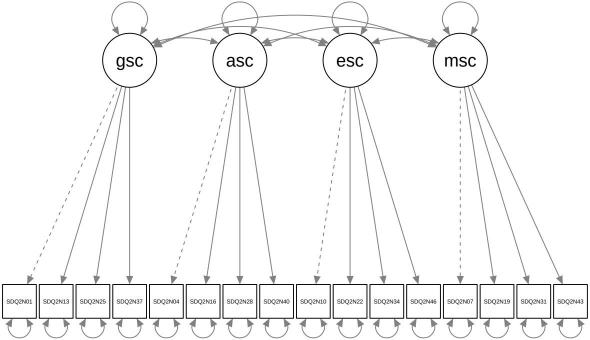
- We have considered the case with one latent factor
- In practice, most of the time researchers are interested in constructs with multiple latent factors
- Our framework can be easily extended to the case with $p > 1$ latent factors

$$\begin{gathered}
\boldsymbol{y}_i \,|\, \boldsymbol{\nu}, \boldsymbol{\Lambda}, \boldsymbol{\eta}_i, \boldsymbol{\psi} \overset{\text{ind.}}{\sim} N\big(\boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i, \mathrm{diag}(\boldsymbol{\psi})\big), \;\; \boldsymbol{\eta}_i \,|\, \boldsymbol{\Sigma} \overset{\text{ind.}}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}), \\
\boldsymbol{\Lambda}_j^T \,|\, \psi_j \overset{\text{ind.}}{\sim} N(\boldsymbol{\mu}_{\boldsymbol{\Lambda}}, \psi_j \boldsymbol{\Sigma}_{\boldsymbol{\Lambda}}), \;\; \nu_j \overset{\text{ind.}}{\sim} N(0, \sigma_\nu^2), \;\; \psi_j \overset{\text{ind.}}{\sim} \text{Inverse-}\chi^2(\kappa_\psi, \delta_\psi) \\
\boldsymbol{\Sigma} \sim \text{Inverse G-Wishart}(G_{\text{full}}, \xi_{\boldsymbol{\Sigma}}, \boldsymbol{\Lambda}_{\boldsymbol{\Sigma}}),
\end{gathered} \tag{4}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Here $\boldsymbol{\eta}_i$ is a vector of length p

Application to self-concept data

- The dataset was collected from 265 early adolescents in grade 7 and consists of 16 observed variables from four subscales of the Self Description Questionnaire II (Byrne, 2016; Marsh, 1992)
- A CFA was used to test the hypothesis that self-concept (SC) is a multidimensional construct composed of $p = 4$ intercorrelated factors: general SC (GSC), academic SC (ASC), English SC (ESC), and mathematics SC (MSC)



Application to self-concept data

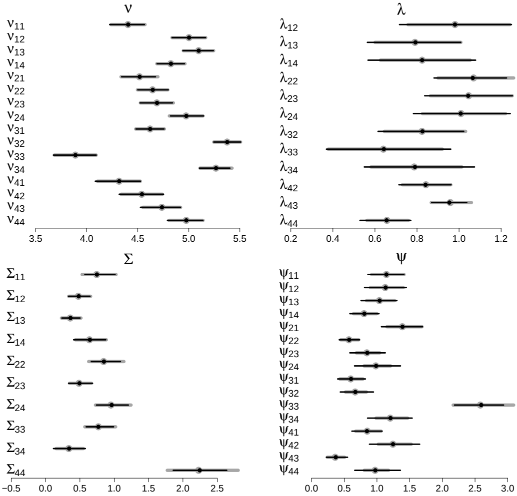


Figure 4: Visualization of 95% credible intervals for the parameters of model (6) applied to the self-concept data. The black lines correspond to percentile bootstrap credible intervals obtained via MFVB and 1,000 bootstrap iterations. These are compared to the MCMC benchmark provided by thicker gray lines

Summary

- We developed fast MFVB algorithms for fitting Bayesian SEMs
- We studied the use of bootstrap and showed improved variational inference for the model parameters of interest
- We are working on MFVB algorithms for more challenging situations
- Our paper has been published recently
  Dang, K. D., & Maestrini, L. (2022). Fitting structural equation models via variational approximations. *Structural Equation Modeling: A Multidisciplinary Journal,* 1-15

Thank you !

References I

Ansari, A. and Jedidi, K. (2000).
Bayesian factor analysis for multilevel binary observations.
Psychometrika, 65(4):475–496.

Ansari, A., Jedidi, K., and Jagpal, S. (2000).
A hierarchical Bayesian methodology for treating heterogeneity in structural equation models.
Marketing Science, 19(4):328–347.

Byrne, B. M. (2016).
Structural equation modeling with AMOS: basic concepts, applications, and programming (multivariate applications series).
New York: Taylor & Francis Group, 3rd edition.

Chen, Y.-C., Wang, Y. S., Erosheva, E. A., et al. (2018).
On the use of bootstrap with variational inference: Theory, interpretation, and a two-sample test example.
Annals of Applied Statistics, 12(2):846–876.

Dunson, D. B. (2000).
Bayesian latent variable models for clustered mixed outcomes.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 62(2):355–366.

Ghahramani, Z. and Beal, M. (1999).
Variational inference for Bayesian mixtures of factor analysers.
Advances in neural information processing systems, 12.

Ghahramani, Z. and Beal, M. J. (2000).
Graphical models and variational methods.
In *Advanced Mean Field Methods-Theory and Practice*. Citeseer.

Holzinger, K. J. and Swineford, F. (1939).
A study in factor analysis: The stability of a bi-factor solution.
Supplementary educational monographs.

References II

Khan, M. E. E., Bouchard, G., Murphy, K. P., and Marlin, B. M. (2010).
Variational bounds for mixed-data factor analysis.
Advances in Neural Information Processing Systems, 23.

Klami, A. (2015).
Polya-gamma augmentations for factor models.
In *Asian Conference on Machine Learning*, pages 112–128. PMLR.

Lee, S.-Y. (2007).
Structural equation modeling: A Bayesian approach.
John Wiley & Sons.

Marsh, H. W. (1992).
Self Description Questionnaire (SDQ) II: A theoretical and empirical basis for the measurement of multiple dimensions of adolescent self-concept: An interim test manual and a research monograph.
New South Wales, Australia: University of Western Sydney, Faculty of Education, pages 53–63.

Zhao, J.-h. and Philip, L. (2009).
A note on variational bayesian factor analysis.
Neural Networks, 22(7):988–997.