Contributed Discussion

Ihnwhi Heo[*], Julius M. Pfadt[†], and Eric-Jan Wagenmakers[†]

Identifying the number of factors has remained a fundamental challenge in factor analysis research. We commend and congratulate Frühwirth-Schnatter et al. (2024) for their inspiring work in addressing this methodological issue. Their approach to achieving sparsity in overfitting exploratory factor analysis models is supported by two computational strategies. First, the authors introduce an unordered generalized lower triangular (UGLT) representation of the factor loading matrix to address rotational invariance. Second, they develop a customized reversible jump Markov chain Monte Carlo (MCMC) algorithm for efficient posterior inference of the number of factors and other model parameters. Their contributions are substantial and have been well summarized in invited discussions. The thorough theoretical exposition, supplemented by empirical examples, positions this work as a stepping stone for further exploration.

In this contributed discussion, we point out avenues for expanding upon the ideas of Frühwirth-Schnatter et al. (2024). The methods proposed by the authors implicitly assume the homogeneity of the population of interest. Put differently, the estimated factor structure is assumed to apply uniformly across the entire population. However, this assumption often needs to be relaxed due to the inherent heterogeneity in the population. One approach to account for population heterogeneity and consider differences between qualitatively distinct subpopulations—termed latent classes—is the mixture modeling framework (Frühwirth-Schnatter, 2010; McLachlan and Peel, 2000). With this consideration in mind, we suggest that an interesting direction is the incorporation of the mixture modeling framework to induce sparsity across different latent classes.

Within the mixture modeling framework, one important consideration is that the number of latent classes is typically predetermined using statistical criteria such as information criteria. Afterward, the factor structure for each latent class is estimated conditional on the estimated group membership. The next step will be to examine whether estimating separate factor structures for each latent class based on the UGLT representation is computationally feasible. Here, the label switching problem must be addressed. Common solutions include adding order constraints to parameters during estimation or post-processing the chains using some permutation techniques to reorder the MCMC output. Developing a scalable MCMC algorithm is essential in this context.

Another direction of research regards the estimation of the number of latent classes without resorting solely to information criteria. Instead of predetermining the dimensionality of factors based on information criteria, the dimension can be solely determined based on the data. A related idea can arise by borrowing the principles of Bayesian nonparametrics (Gershman and Blei, 2012; Ghahramani, 2013). Relatedly, Grushanina and

[*]Department of Psychological Sciences, University of California, Merced, ihnwhi.heo@gmail.com
[†]Department of Psychology, University of Amsterdam, julius.pfadt@gmail.com; ej.wagenmakers@gmail.com

Frühwirth-Schnatter (2023) developed an automatic inference process to assign a nonparametric prior for estimating the dimension of factors. The development of MCMC algorithms that integrate the data-driven estimation of factor dimensionality and induce sparsity warrants further research.

An important parameter in implementing mixture models is the mixing proportion. In the Bayesian framework, Dirichlet distributions are commonly used as prior distributions for this parameter, with hyperparameters representing the prior proportion of individuals in each class. Implementing this prior alongside the spike and slab priors for simultaneously imposing row and column sparsity while establishing factor structures could be an important contribution to future research.

We believe that expanding the ideas of Frühwirth-Schnatter et al. (2024) into the factor mixture modeling framework can reveal important and nuanced details about population heterogeneity, particularly with regard to the factor dimensions within each latent class. Looking ahead, the integration of their approach, along with its mixture extensions, into open-source statistical software such as JASP (JASP Team, 2024) could augment the accessibility and dissemination of these methods. We are confident that the work of Frühwirth-Schnatter et al. (2024) represents a substantial contribution that will inspire further advancements in both methodological and applied research. We once again commend the authors for their outstanding work.

References

Frühwirth-Schnatter, S. (2010). *Finite Mixture and Markov Switching Models*. Springer. MR2265601. 295

Frühwirth-Schnatter, S., Hosszejni, D., and Lopes, H. F. (2024). "Sparse Bayesian factor analysis when the number of factors is unknown." *Bayesian Analysis*, 1–31. 295, 296

Gershman, S. J. and Blei, D. M. (2012). "A tutorial on Bayesian nonparametric models." *Journal of Mathematical Psychology*, 56(1): 1–12. MR2903470. doi: https://doi.org/10.1016/j.jmp.2011.08.004. 295

Ghahramani, Z. (2013). "Bayesian non-parametrics and the probabilistic approach to modelling." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1984): 20110553. MR3005667. doi: https://doi.org/10.1098/rsta.2011.0553. 295

Grushanina, M. and Frühwirth-Schnatter, S. (2023). "Dynamic mixture of finite mixtures of factor analysers with automatic inference on the number of clusters and factors." *arXiv preprint arXiv:2307.07045*. MR2265601. 295

JASP Team (2024). "JASP (Version 0.19.0) [Computer software]." https://jasp-stats.org/. 296

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. John Wiley & Sons. MR1789474. doi: https://doi.org/10.1002/0471721182. 295