A discussion on: "Sparse Bayesian factor analysis when the number of factors is unknown" by Frühwirth-Schnatter, S., Hosszejni D., and Lopes, H. F.

DOI:10.1214/24-BA1423

(Article begins on next page)

Bayesian Analysis (2024) **TBA**, Number TBA, pp. 1

Contributed Discussion on "Sparse Bayesian Factor Analysis When the Number of Factors Is Unknown" by Sylvia Frühwirth-Schnatter, Darjus Hosszejni and Hedibert Freitas Lopes

Marta Catalano[*], Beatrice Franzolini[†], Matteo Giordano[‡] and Giovanni Rebaudo[‡]

We congratulate the authors for their contribution to the field of Bayesian factor analysis, which develops an appealing methodology for sparse recovery. We believe that their approach may become a useful tool in applications for statisticians and practitioners, thanks to its ability to simultaneously perform inference on identifiable sparse factor loadings and achieve data-driven model selection in terms of the number of factors. This works also stimulates a wealth of interesting follow up questions, both on the specifics of the model at hand and also drawing from the broader contemporary Bayesian literature. In this discussion we expand upon some of these questions, with specific reference to multiple testing, dependence among slab probabilities, continuous alternatives to spike-and-slab priors, and the learning of the number of factors.

Multiplicity and multiple testing. The sparse Bayesian factor model with UGLT structures estimates all parameters simultaneously. Nonetheless, the estimation process can be viewed as addressing two inferential challenges: first, a multiple testing problem, where each entry of the factor loading matrix is tested to determine whether it is active or not; and second, the estimation of the values of the active entries. In the context of multiple testing, the Bayesian framework naturally accommodates two desirable yet distinct types of penalties. The first is an Ockham's razor penalty, typically arising from the use of marginal likelihoods, which penalizes models with more active loadings in factor analysis, thus *promoting sparsity*. The second is a multiplicity penalty, thanks to which the posterior inclusion probabilities for each factor loading decrease as the dimensions m or H increase, leading to a framework of *adaptive sparsity*. Multiplicity penalties relate to the frequentist challenge of multiple testing and generally provide better control of noisy signals and false discoveries by adopting a more cautious approach (Scott and Berger, 2010). It would be valuable to assess the presence and potential influence of the multiplicity penalty in the Bayesian model with UGLT stuctures. Specifically, while the prior on τ_j seems to induce a multiplicity penalty in H (i.e., the number of columns), it would be worth exploring whether a similar penalty also applies to m (i.e., the number of rows) and whether such penalties are preserved after the post-processing procedure. Additionally, it would be interesting to investigate how these penalties affect the model's recovery performance, for instance, by examining changes in the ROC curve.

[*]Luiss University, Italy, mcatalano@luiss.it
[†]Bocconi University, Italy, beatrice.franzolini@unibocconi.it
[‡]University of Torino, Italy, matteo.giordano@unito.it, giovanni.rebaudo@unito.it

Individual components in the slab probability. The prior for the factor loadings $(\beta_{i,j})_{i,j}$ in (3.1) requires a multivariate spike and slab prior. The dependence between the marginals $\beta_{i,j}$ is induced by column-specific slab probabilities τ_j and a multivariate hierarchical slab distribution with row-specific dependence, which shrinks each component towards zero. However, to achieve additional shrinkage for individual factor loadings, in (3.12) the authors propose to add an individual shrinking factor in the multivariate hierarchical slab distribution. Could a similar result be achieved by allowing for individual components in the slab probabilities? This would bring to the need for multivariate versions of the exchangeable or cumulative shrinkage process, which could possibly converge to a multivariate version of the Indian buffet process, e.g., the hierarchical Indian buffet process (Thibaux and Jordan, 2007; James et al., 2024).

A role for other Bayesian approaches to sparsity and model selection in Bayesian factor analysis? Implementing posterior-based inference with spike-and-slab priors is a notoriously challenging task, due to the underlying combinatorial problem of exploring the space containing all possible sparsity patterns. One of the key contributions of the present paper is the construction of an ad-hoc MCMC sampler (cf. Algorithm 1), which cleverly alternates between the two formulations, Exploratory and Confirmatory, of the Factor Analysis model. On the other hand, the construction of Bayesian models for sparse structures and variable selection is an issue of general interest in the broader Bayesian literature, where some continuous (and easier to work with) alternatives to the spike-and-slab prior have been shown, either theoretically or empirically, to be potential effective approaches. For example, Laplace priors are known to possess desirable sparsity-promoting properties at the level of the maximum-a-posterior estimators (Agapiou et al., 2018) and here could be deployed either directly onto the factor loadings β_{ij}, or column-wise for the matrix β_H. Excitingly, hierarchical Gaussian priors have recently been shown to possess variable selection properties when endowed with a horseshoe hyper-prior on the length-scale without the need for the additional spike-and-slab structure (Castillo and Randrianarisoa, 2024). Some further possibilities include the horseshoe priors themselves (and extensions thereof), Dirichlet–Laplace and the so-called R2-D2 priors, see Hirsh et al. (2022). These developments suggest the questions as to whether such sparse (or approximately sparse) continuous priors models could also be employed with success in Bayesian factor analysis.

Learning the number of factors. Learning the number of factors simultaneously with estimating the factors is a challenging problem, which the authors address effectively in their proposal. Specifically, by imposing a UCLT structure, they facilitate the joint identification of the unknown number of factors r and the underlying factor model parameters Λ and Σ_0 from the overfitting BFA model. Inference on the number of factors is validated through a simulation study, where the authors empirically demonstrate the model's ability to recover the true number of factors, r_{true}. These findings give hope that the model can consistently estimate the true number of factors under a well-specified data-generating process. Establishing such consistency for the discrete parameter r_{true} under an identifiable and well-specified model might be achievable by leveraging Doob's Theorem (Doob, 1949). In particular, demonstrating consistency for the true number of factors $r_{\text{true}}(< H)$ in a subset of the possible values of r of prior probability one would imply consistency across all possible values of r, given that the prior distribution assigns positive probability to each possible value. This strategy has been successfully employed, for instance, in proving the consistency of the number of mixture components in Bayesian finite mixture models with a prior on the number of components (Nobile, 1994; Miller, 2023).

References

Agapiou, S., Burger, M., Dashti, M., and Helin, T. (2018). "Sparsity-promoting and edge-preserving maximum a posteriori estimators in non-parametric Bayesian inverse problems." *Inverse Probl.*, 34: 1–37. 2

Castillo, I. and Randrianarisoa, T. (2024). "Deep horseshoe Gaussian processes." *Preprint at arXiv: 2403.01737*. 2

Doob, J. L. (1949). "Application of the theory of martingales." In *Le calcul des probabilites et ses applications*, volume 13, 23–27. 2

Hirsh, S. M., Barajas-Solano, D. A., and Kutz, J. N. (2022). "Sparsifying priors for Bayesian uncertainty quantification in model discovery." *R. Soc. Open Sci.*, 9: 1–20. 2

James, L. F., Lee, J., and Pandey, A. (2024). "Bayesian analysis of generalized hierarchical Indian buffet processes for within and across group sharing of latent features." *Preprint at arXiv: 2304.05244*. 2

Miller, J. W. (2023). "Consistency of mixture models with a prior on the number of components." *Depend. Model.*, 11: 1–9. 3

Nobile, A. (1994). "Bayesian Analysis of Finite Mixture Distributions." Ph.D. thesis, Carnegie Mellon Univ. 3

Scott, J. G. and Berger, J. O. (2010). "Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem." *Ann. Stat.*, 38: 2587–2619. 1

Thibaux, R. and Jordan, M. I. (2007). "Hierarchical Beta Processes and the Indian Buffet Process." In Meila, M. and Shen, X. (eds.), *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*, volume 2 of *Proceedings of Machine Learning Research*, 564–571. San Juan, Puerto Rico: PMLR. 2