# Sparse Bayesian Factor Analysis when the Number of Factors is Unknown

Simran Bilkhu

McMaster University

# Introduction

- Goal: Estimating a factor analysis model:

$$f_t \sim MVN_r\left(0, I_r\right), \quad y_t = \Lambda f_t + \epsilon_t, \quad \epsilon_t \sim MVN_m\left(0, \Sigma_0\right)$$

- Traditional methods struggle with identifying the correct number of factors
- The proposed approach starts with: Over-fitting model with $k \geq r$ factors

$$y_t = \beta_k f_t^k + \epsilon_t$$

# What types of factors are we including if we overfit?

Two types of excess factors:

- Factors that capture already existing relationships
- Spurious factors that add no meaningful information

It's easier to detect spurious factors, since they will only have one factor loading that is non-zero.

# Intoducing restrictions

### Definition (Sparsity Matrix)

If $\delta_k$ is a binary matrix with $\delta_{i,j} = 1 \iff \beta_{k_{i,j}} \neq 0$, then $\delta_k$ is a sparsity matrix for $\beta_k$

### Definition (UGLT Structure)

Let $\delta_k$ be a binary matrix with $r$ columns. If all pivots lie in different rows, $\delta_k$ is a UGLT structure.

### Definition (3579 Counting Rule)

A binary matrix $\delta_k$ satisfies the 3579 counting rule if each sub-matrix of $\delta_k$ consisting of $q$ columns has $2q + 1$ 1's in it.

# UGLT Structure Solution

Consider a model where we fit $k = r + 1$ factors, then, if $\beta_k$ as a UGLT structure, it turns out that we can easily identify the original factor matrix and the "spurious" loading:

$$\beta_{r+1} = \left( \quad \Lambda \quad \middle| \quad \Xi \quad \right) \quad \Xi = \begin{pmatrix} 0 \\ \vdots \\ \Xi_{lsp} \\ \vdots \\ 0 \end{pmatrix}$$

# UGLT Structure Solution

$$\Sigma_r = Diag\left(\sigma_1^2, \ldots, \sigma_{lsp}^2 - \Xi_{lsp}^2, \ldots, \sigma_m^2\right)$$

Where:

- $\Xi$ has only one non-zero entry $\Xi_{lsp}$
- $\Sigma_{r+1}$ is modified to remove the impact of this spurious factor

The above generalizes for a general $k \geq r$. So, this allows us to:

- Eliminate spurious columns
- Recover $\Lambda$ up to a signed permutation
- Recover the original covariance matrix

# Prior Specifications

**Imposing Sparsity through Spike and Slab Priors**:

$$\beta_{ij}|\tau_j \sim (1-\tau_j)\,\Delta_0 + \tau_j P_{\text{slab}}\,(\beta_{ij})$$

**Where**:

$$\tau_j|k \sim Beta\,(a_k, b_k)$$

**Imposing a UGLT Structure**: Let $\mathbf{l}_r, \mathbf{l}_{r_{sp}}$ be the pivots of the active and spurious columns in $\beta_k$, we define a prior on the pivots as follows: Let $\mathbf{l} \subseteq \{1, 2, \ldots, m\}$ $L\,(\mathbf{l}) = \{i \in \{1, 2, \ldots, m\} : i \notin \mathbf{l}\}$, and set:

$$p(l_j|\mathbf{l}_{r,-j}) = \frac{1}{m-r+1}$$

Where $\mathbf{l_{r,-j}} = \{l_1, l_2, \ldots, l_r\} - \{l_j\}$

# Impose a UGLT Structure

To impose the UGLT structure into the sparsity matrix:

$$p(\delta_{ij} = 1 | l_j, \tau_j) = \begin{cases} 0, & i < l_j \\ 1, & i = l_j \\ \tau_j, & i = l_j + 1, \ldots, m \end{cases}$$

The rest of the priors can be introduced through a normal mixture, or any other reasonable choice of prior.

# Conclusion

Model estimation is achieved using a reversible jump MCMC algorithm - next set of slides. Key Contributions:

- Novel approach to factor selection
- UGLT structure for handling excess factors
- Ability to recover true factor structure (up to a signed permutation) through the overfitting model

# References


- Frühwirth-Schnatter, Sylvia, Darjus Hosszejni, and Hedibert Freitas Lopes. "Sparse Bayesian factor analysis when the number of factors is unknown." Bayesian Analysis 1.1 (2024): 1-44.