Mathematical Details on the Bayesian model underlying the Bayesian Multitarget Latent Factors package

Bruno Ursino

February 19, 2024

Abstract

This document presents the mathematical foundation of the Bayesian model implemented in the Bayesian Multitarget Latent Factors Python package for the analysis of multivariate functional datasets. We explore the theoretical background, model specification, and practical implementation details, focusing on the Bayesian approach's advantages for data analysis.

1 Introduction

Introduce the Python package and the need for a Bayesian approach in data analysis. Discuss the goals of the document and the structure.

2 Theoretical Background

Discuss Bayesian statistics fundamentals, including Bayes' theorem, prior distributions, likelihood functions, and posterior distributions. Explain the philosophical underpinnings of Bayesian inference.

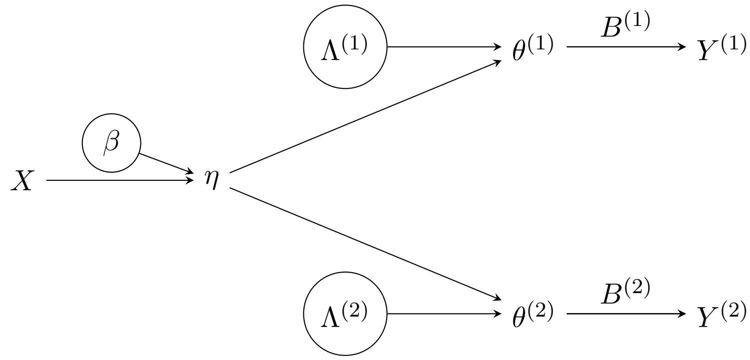

Figure 1: Schematic representation of the model showing the progression from covariates X through latent variables η and their components $\Lambda^{(1)}$ and $\Lambda^{(2)}$, leading to the functional targets $Y^{(1)}$ and $Y^{(2)}$.

3 Model Specification

3.1 Overview of the Model

Introduction. This section introduces a Bayesian framework based on latent factors, extending and generalizing the regression methodology presented by Montagna et al. [1] to the case of multiple functional outcomes. Our approach finds parallels in the work of Noh and Choi [2], who applied distinct latent factors for each functional target. In contrast, our approach involves latent variables linked to individuals and shared across the latent factors. Adopted from the physical context of our investigation, this methodology enhances the simplicity and interpretability of the posterior distributions.

Model parameters and Data. The model is characterized by a comprehensive parameter set Θ, encompassing:

- $\boldsymbol{\beta}$, regression coefficients linking the covariates to the latent variables $\boldsymbol{\eta}$ $(k \times r)$,
- $\boldsymbol{\Lambda}^{(1)}$ and $\boldsymbol{\Lambda}^{(2)}$, components of latent factors for each functional target $(p_i \times k)$ for $i = 1, 2$,
- $\boldsymbol{\tau}_\lambda^{(1)}$ and $\boldsymbol{\tau}_\lambda^{(2)}$, precision parameters for the latent factors (vectors of length k),
- $\boldsymbol{\tau}_\theta^{(1)}$ and $\boldsymbol{\tau}_\theta^{(2)}$, precision parameters for the basis function coefficients (vectors of length p_i for $i = 1, 2$),
- $\psi^{(1)}$ and $\psi^{(2)}$, standard deviations of the observation noise around the functional targets.

Basis matrices $\boldsymbol{B}^{(1)}$ and $\boldsymbol{B}^{(2)}$, along with coefficients $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$, form the basis function representations for the functional targets. The design matrix $\boldsymbol{X}$ includes the covariates, while $\boldsymbol{Y}^{(1)}$ and $\boldsymbol{Y}^{(2)}$ represent the target data matrices.

3.2 Complete model specification

Likelihood:

$$
\begin{aligned}
Y^{(1)}_{h_1,s}\Big|\Theta,\theta^{(1)},\theta^{(2)},\eta &\overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\left(B^{(1)}\theta^{(1)}\right)_{h_1,s}, {\psi^{(1)}}^2\right)\\
Y^{(2)}_{h_2,s}\Big|\Theta,\theta^{(1)},\theta^{(2)},\eta &\overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\left(B^{(2)}\theta^{(2)}\right)_{h_2,s}, {\psi^{(2)}}^2\right)\\
\theta^{(1)}_{j_1,s}\Big|\Theta,\eta &\overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\left(\Lambda^{(1)}\eta\right)_{j_1,s}, \quad 1/\tau^{(1)}_{\theta\, j_1}\right)\\
\theta^{(2)}_{j_2,s}\Big|\Theta,\eta &\overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\left(\Lambda^{(2)}\eta\right)_{j_2,s}, \quad 1/\tau^{(2)}_{\theta\, j_2}\right)\\
\eta_{\ell,s}|\Theta &\overset{\text{i.i.d.}}{\sim} \mathcal{N}\left((\beta X)_{\ell,s}, 1\right)
\end{aligned}
\tag{1}
$$

for:

$$
\begin{aligned}
h_1 &= 1,\dots,L_1, & h_2 &= 1,\dots,L_2,\\
j_1 &= 1,\dots,p_1, & j_2 &= 1,\dots,p_2,\\
\ell &= 1,\dots,k, & s &= 1,\dots,N
\end{aligned}
$$

Prior:

$$
\begin{aligned}
\beta_{\ell,i} &\overset{\text{i.i.d.}}{\sim} \mathbf{t}_\chi(0,1)\\
\Lambda^{(1)}_{j_1,\ell}\Big|\tau^{(1)}_{\lambda\,\ell} &\overset{\text{i.i.d.}}{\sim} \mathbf{t}_\nu\left(0, 1/\tau^{(1)}_{\lambda\,\ell}\right)\\
\boldsymbol{\tau}^{(1)}_\lambda &\sim \text{MGPS}(\alpha_1,\alpha_2)\\
\Lambda^{(2)}_{j_2,\ell}\Big|\tau^{(2)}_{\lambda\,\ell} &\overset{\text{i.i.d.}}{\sim} \mathbf{t}_\nu\left(0, 1/\tau^{(2)}_{\lambda\,\ell}\right)\\
\boldsymbol{\tau}^{(2)}_\lambda &\sim \text{MGPS}(\alpha_1,\alpha_2)\\
\tau^{(1)}_{\theta\, j_1} &\overset{\text{i.i.d.}}{\sim} \Gamma(\alpha_\sigma,\beta_\sigma)\\
\tau^{(2)}_{\theta\, j_2} &\overset{\text{i.i.d.}}{\sim} \Gamma(\alpha_\sigma,\beta_\sigma)\\
{\psi^{(1)}}^2 &\overset{\text{i.i.d.}}{\sim} \text{Inv-}\Gamma(\alpha_\psi,\beta_\psi)\\
{\psi^{(2)}}^2 &\overset{\text{i.i.d.}}{\sim} \text{Inv-}\Gamma(\alpha_\psi,\beta_\psi)
\end{aligned}
\tag{2}
$$

for:

$$
\begin{aligned}
j_1 &= 1,\dots,p_1, & j_2 &= 1,\dots,p_2,\\
\ell &= 1,\dots,k, & i &= 1,\dots,r
\end{aligned}
$$

Where the Multiplicative Gamma Process Shrinkage (MGPS) prior is defined as:

$$
\boldsymbol{\tau} \sim \text{MGPS}(\alpha_1,\alpha_2) \quad \Longleftrightarrow \quad
\begin{aligned}
&\tau_\ell = \prod_{\gamma=1}^{\ell}\delta_\gamma, && \ell = 1,\dots,k\\
&\delta_1 \sim \Gamma(\alpha_1, 1)\\
&\delta_\ell \sim \Gamma(\alpha_2, 1), && \ell = 2,\dots,k
\end{aligned}
\tag{3}
$$

3.3 Likelihood Function

Standard Formulation. The likelihood function quantifies the probability of observing the data given the model parameters. In the context of our latent factor model, the likelihood of the functional data $\boldsymbol{Y}_s \left[= \left(\boldsymbol{Y}_s^{(1)}, \boldsymbol{Y}_s^{(2)}\right)\right]$, is formulated under the assumption of independence between samples, conditioned on the parameters Θ.

The schematic description in figure 1 can be expressed as:

$$
\begin{aligned}
\boldsymbol{\eta}_s &= \boldsymbol{\beta} X_{\cdot,s} &&+ \boldsymbol{\varepsilon}_{\eta,s} \\
\boldsymbol{\theta}_s^{(1)} &= \boldsymbol{\Lambda}^{(1)} \boldsymbol{\eta}_s &&+ \boldsymbol{\varepsilon}_{\theta_1,s} \\
\boldsymbol{\theta}_s^{(2)} &= \boldsymbol{\Lambda}^{(2)} \boldsymbol{\eta}_s &&+ \boldsymbol{\varepsilon}_{\theta_2,s} \\
\boldsymbol{Y}_s^{(1)} &= \boldsymbol{B}^{(1)} \boldsymbol{\theta}^{(1)} &&+ \boldsymbol{\varepsilon}_{Y_1,s} \\
\boldsymbol{Y}_s^{(2)} &= \boldsymbol{B}^{(2)} \boldsymbol{\theta}^{(2)} &&+ \boldsymbol{\varepsilon}_{Y_2,s}
\end{aligned}
$$

In particular, the description in (1) is equivalent to:

$$
\boldsymbol{Y}_s = \begin{bmatrix} \boldsymbol{Y}_s^{(1)} \\ \boldsymbol{Y}_s^{(2)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{B}^{(1)} \boldsymbol{\Lambda}^{(1)} \boldsymbol{\beta} X_{\cdot,s} \\ \boldsymbol{B}^{(2)} \boldsymbol{\Lambda}^{(2)} \boldsymbol{\beta} X_{\cdot,s} \end{bmatrix} + \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{B}^{(1)} & \boldsymbol{0} & \boldsymbol{B}^{(1)} \boldsymbol{\Lambda}^{(1)} \\ \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{B}^{(2)} & \boldsymbol{B}^{(2)} \boldsymbol{\Lambda}^{(2)} \end{bmatrix} \boldsymbol{\mathcal{E}}_s \tag{4}
$$

where

$$
\boldsymbol{\mathcal{E}}_s = (\boldsymbol{\varepsilon}_{Y_1,s}, \boldsymbol{\varepsilon}_{Y_2,s}, \boldsymbol{\varepsilon}_{\theta_1,s}, \boldsymbol{\varepsilon}_{\theta_2,s}, \boldsymbol{\varepsilon}_{\eta,s},)
$$

is normally distributed with a diagonal covariance structure:

$$
\boldsymbol{\mathcal{E}}_s | \, \Theta \sim \mathcal{N}(0, \boldsymbol{\Sigma}) \tag{5}
$$

$$
\boldsymbol{\Sigma} = \left[\begin{array}{c|c|ccc|ccc|c}
{\psi^{(1)}}^2 \boldsymbol{I} & \boldsymbol{0} & & \boldsymbol{0} & & & \boldsymbol{0} & & \boldsymbol{0} \\
\hline
\boldsymbol{0} & {\psi^{(2)}}^2 \boldsymbol{I} & & \boldsymbol{0} & & & \boldsymbol{0} & & \boldsymbol{0} \\
\hline
 & & 1/\tau_{\theta\,1}^{(1)} & \cdots & 0 & & & & \\
\boldsymbol{0} & \boldsymbol{0} & \vdots & \ddots & \vdots & & \boldsymbol{0} & & \boldsymbol{0} \\
 & & 0 & \cdots & 1/\tau_{\theta\,p_1}^{(1)} & & & & \\
\hline
 & & & & & 1/\tau_{\theta\,1}^{(2)} & \cdots & 0 & \\
\boldsymbol{0} & \boldsymbol{0} & & \boldsymbol{0} & & \vdots & \ddots & \vdots & \boldsymbol{0} \\
 & & & & & 0 & \cdots & 1/\tau_{\theta\,p_2}^{(2)} & \\
\hline
\boldsymbol{0} & \boldsymbol{0} & & \boldsymbol{0} & & & \boldsymbol{0} & & \boldsymbol{I}
\end{array}\right] = \begin{bmatrix}
{\psi^{(1)}}^2 \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\
\boldsymbol{0} & {\psi^{(1)}}^2 \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\
\boldsymbol{0} & \boldsymbol{0} & \Delta_\theta^{(1)} & \boldsymbol{0} & \boldsymbol{0} \\
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \Delta_\theta^{(2)} & \boldsymbol{0} \\
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}
\end{bmatrix}
$$

Marginal Likelihood. Recognizing that (4) shows that $\boldsymbol{Y}_s$ is an affine transformation of a multivariate normal distribution (5), we derive an alternative formulation of the likelihood function, which marginalizes over the latent variables $\boldsymbol{\eta}$ and basis function coefficients $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}$:

$$
\boldsymbol{Y}_s | \, \Theta \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\begin{bmatrix} \boldsymbol{B}^{(1)} \boldsymbol{\Lambda}^{(1)} \boldsymbol{\beta} X_{\cdot,s} \\ \boldsymbol{B}^{(2)} \boldsymbol{\Lambda}^{(2)} \boldsymbol{\beta} X_{\cdot,s} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right), \qquad s = 1, \ldots, N \tag{6}
$$

where

$$
\boldsymbol{\Sigma}_{ii} = {\psi^{(i)}}^2 \boldsymbol{I} + \boldsymbol{B}^{(i)} \left[\Delta_\theta^{(i)} + \boldsymbol{\Lambda}^{(i)} {\boldsymbol{\Lambda}^{(i)}}^T \right] {\boldsymbol{B}^{(i)}}^T \tag{7}
$$

$$
\boldsymbol{\Sigma}_{ij} = \boldsymbol{B}^{(i)} \boldsymbol{\Lambda}^{(i)} {\boldsymbol{\Lambda}^{(j)}}^T {\boldsymbol{B}^{(j)}}^T \tag{8}
$$

This yields a direct relationship between the observed data and the covariates, which is encoded in the mean of the resulting multivariate normal distribution, at the same gives light to the relationship between the two functional targets $\boldsymbol{Y}^{(1)}$ and $\boldsymbol{Y}^{(2)}$ which is encoded in the covariance matrix $\Sigma_{12} = \Sigma_{21}^T$.

Likelihood and latent variables. Another related useful expression can be derived by marginalizing only the basis function coefficients, in which case:

$$\left[\begin{array}{c} \boldsymbol{Y}_s^{(1)} \\ \boldsymbol{Y}_s^{(2)} \\ \boldsymbol{\eta}_s \end{array}\right] \Bigg| \Theta \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\left[\begin{array}{c} \boldsymbol{B}^{(1)}\boldsymbol{\Lambda}^{(1)}\boldsymbol{\beta}X_{\cdot,s} \\ \boldsymbol{B}^{(2)}\boldsymbol{\Lambda}^{(2)}\boldsymbol{\beta}X_{\cdot,s} \\ \boldsymbol{\beta}X_{\cdot,s} \end{array}\right], \left[\begin{array}{ccc} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \boldsymbol{B}^{(1)}\boldsymbol{\Lambda}^{(1)} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \boldsymbol{B}^{(2)}\boldsymbol{\Lambda}^{(2)} \\ \boldsymbol{\Lambda}^{(1)^T}\boldsymbol{B}^{(1)^T} & \boldsymbol{\Lambda}^{(2)^T}\boldsymbol{B}^{(2)^T} & \boldsymbol{I} \end{array}\right]\right), \quad s = 1, \ldots, N \tag{9}$$

3.4 Prior Distributions

General Considerations. In Bayesian analysis, prior distributions are integral for incorporating existing knowledge or conjectures regarding parameters into the analytical framework. Our selection of priors for the latent factor model is pivotal for managing model complexity and ensuring computational tractability.

Prior Selection for Regression Coefficients $\boldsymbol{\beta}$. The regression coefficients $\boldsymbol{\beta}$ link the covariates to the latent variables. We employ a standard Student's t-distribution with χ degrees of freedom, allowing some coefficients to significantly diverge from the mean, when supported by the data.

Prior Selection for Latent Factors $\boldsymbol{\Lambda}^{(i)}$ and Precision Parameters $\boldsymbol{\tau}_\lambda^{(i)}$. The latent factors $\boldsymbol{\Lambda}^{(1)}$ and $\boldsymbol{\Lambda}^{(2)}$ are crucial for capturing the underlying structure of the functional targets. For these elements, we adopt a centered Student's t-distribution prior,

$$\Lambda_{j,\ell}^{(i)} \sim \mathbf{t}_\nu\left(0, 1/\tau_{\lambda\,\ell}^{(i)}\right),$$

while the precision parameters $\boldsymbol{\tau}_\lambda^{(i)}$ follow a Multiplicative Gamma Process Shrinkage prior (3):

$$\boldsymbol{\tau}_\lambda^{(i)} \sim MGPS\left(\alpha_1, \alpha_2\right)$$

where α_1 and α_2 are hyperparameters that modulate the regularization degree.

$$\mathbb{V}\text{ar}\left[\Lambda_{j,\,\ell}^{(i)}\middle|\,\boldsymbol{\tau}_\lambda^{(i)}\right] = \frac{\nu}{\nu-2}\left(\tau_{\lambda\,\ell}^{(i)}\right)^{-1}$$

which leads to an interest on the moments of the reciprocal of the precision for informed hyperparameter selection. The expected value and variance of the reciprocal of τ_λ are presented below with proofs deferred to the appendix. If $\alpha_1 > 2$ and $\alpha_2 > 2$:

$$\begin{aligned} \mathbb{E}\left[\left(\tau_{\lambda\,\ell}^{(i)}\right)^{-1}\right] &= \frac{1}{\alpha_1-1}\left(\frac{1}{\alpha_2-1}\right)^{\ell-1}, \qquad \ell = 1, \ldots, k \\ \mathbb{V}\text{ar}\left[\left(\tau_{\lambda\,\ell}^{(i)}\right)^{-1}\right] &= \frac{1}{(\alpha_1-1)(\alpha_1-2)}\left[\frac{1}{(\alpha_2-1)(\alpha_2-2)}\right]^{\ell-1} \cdot \left[1 - \frac{\alpha_1-2}{\alpha_1-1}\left(\frac{\alpha_2-2}{\alpha_2-1}\right)^{\ell-1}\right] \end{aligned} \tag{10}$$

If $1 < \alpha_i < 2$, the expectation remains valid but the variance becomes undefined.

The choice of $\alpha_2 > 2$ ensures diminishing effect sizes for higher indexed latent factors. Additionally, practical experience indicates difficulties in sampling using Hamiltonian Monte Carlo from distributions lacking variance, advocating for $\alpha_i > 2$. Still, excessively large α_2 values, implying a limited number of latent factors with significant effect sizes, are not recommended.

Since α_1 influences only the 'initial condition' $\tau_{\lambda\,1}^{(i)}$, other useful formulas are:

$$\begin{aligned} \mathbb{E}\left[\left(\frac{\tau_{\lambda\,\ell}^{(i)}}{\tau_{\lambda\,1}^{(i)}}\right)^{-1}\right] &= \left(\frac{1}{\alpha_2-1}\right)^{\ell-1}, \qquad \ell = 1,\ldots,k \\ \mathbb{V}\mathrm{ar}\left[\left(\frac{\tau_{\lambda\,\ell}^{(i)}}{\tau_{\lambda\,1}^{(i)}}\right)^{-1}\right] &= \left[\frac{1}{(\alpha_2-1)(\alpha_2-2)}\right]^{\ell-1}\cdot\left[1-\left(\frac{\alpha_2-2}{\alpha_2-1}\right)^{\ell-1}\right] \end{aligned} \tag{11}$$

Prior Selection for Precision Parameters of Basis Function Coefficients $\boldsymbol{\tau}_\theta^{(i)}$. A conventional choice for the precision of a normally distributed quantity, is the Gamma prior. Individual $\boldsymbol{\tau}_{\theta\,j}^{(i)}$ for each basis function coefficient allows heteroscedasticity in the functional representation of the targets, allowing non vanishing likelihood for residual functions that the latent factors failed to capture.

$$\tau_{\theta\,j}^{(i)} \sim \Gamma\left(\alpha_\sigma, \beta_\sigma\right)$$

Prior Selection for Observation Noise $\psi^{(i)}$. The noise component is assumed to be homoscedastic, delegating the management of heteroscedasticity to the basis function coefficients' variance:

$$\left(\psi^{(i)}\right)^2 \sim \text{Inv-}\Gamma\left(\alpha_\psi, \beta_\psi\right)$$

4 Moments of the Prior Predictive and Prior Elicitation

4.1 Moments of the Prior Predictive

The proofs of all the results in this subsection are postponed in appendix B.

Otherwise explicitly stated, the presented results in this section hold under the assumption of $\alpha_2, \nu, \chi > 2$ and $\alpha_1, \alpha_\sigma, \alpha_\psi > 1$.

Expected Value. As all quantities that enter in the likelihood model are centered, the expected value of the prior predictive is 0:

$$\mathbb{E}[\boldsymbol{Y}_s] = \boldsymbol{0}$$

Total Variance. We define the total variance Σ_T^k as the trace of the covariance matrix $\mathbb{C}\mathrm{ov}[\boldsymbol{Y}_s]$, where the number of latent factors k is explicitly indicated, while Σ_T^∞ indicates the limit as $k \to \infty$ which is shown to exist under the assumptions of this section.

The total variance can be expressed as:

$$\Sigma_{T,s}^k = \sigma_{\mathcal{WN}}^2 + \mathbf{Tr}\left(\boldsymbol{B}\boldsymbol{B}^T\right)\left[\sigma_{\mathcal{I}}^2 + \rho_\Lambda^{(k)}\sigma_\Lambda^2\left(1 + \frac{\chi}{\chi-2}\|\boldsymbol{X}_{\cdot,s}\|_2^2\right)\right]$$

We also introduce the Total Variance explained by the latent factors:

$$\Sigma_{T\Lambda,s}^k = \rho_\Lambda^{(k)}\mathbf{Tr}\left(\boldsymbol{B}\boldsymbol{B}^T\right)\left[\sigma_\Lambda^2\left(1 + \frac{\chi}{\chi-2}\|\boldsymbol{X}_{\cdot,s}\|_2^2\right)\right]$$

where we introduced the following quantities:

- the Prior Observation Variance $\sigma_{\mathcal{WN}}^2 = \frac{(L_1+L_2)\beta_\psi}{\alpha_\psi - 1}$,
- the Trace of the Hat Matrix $\mathbf{Tr}\left(\boldsymbol{B}\boldsymbol{B}^T\right)$,
- the Idiosyncratic Functional Variance $\sigma_{\mathcal{I}}^2 = \frac{\beta_\sigma}{\alpha_\sigma - 1}$,
- $\rho_\Lambda^{(k)} = \Sigma_{T\Lambda}^k / \Sigma_{T\Lambda}^\infty = 1 - \left(\frac{1}{\alpha_2 - 1}\right)^k$ the portion of latent factor's variance, explained by the first k latent factors,
- the variance of the latent factors $\sigma_\Lambda^2 = \frac{\nu}{\nu-2}\frac{1}{\alpha_1-1}\frac{\alpha_2-1}{\alpha_2-2}$,

To complete the analysis we assume the covariates are whitened:

$$\boldsymbol{X}_{\cdot,s} \sim \mathcal{N}(0, V_X\boldsymbol{I})$$

$$\Sigma_T^k := \lim_{N\to\infty}\frac{1}{N}\sum_{s=1}^N \Sigma_{T,s}^k = \sigma_{\mathcal{WN}}^2 + \mathbf{Tr}\left(\boldsymbol{B}\boldsymbol{B}^T\right)\left[\sigma_{\mathcal{I}}^2 + \rho_\Lambda^{(k)}\sigma_\Lambda^2\left(1 + \frac{\chi}{\chi-2}rV_X\right)\right]$$

$$\Sigma_{T\Lambda}^k := \lim_{N\to\infty}\frac{1}{N}\sum_{s=1}^N \Sigma_{T\Lambda,s}^k = \rho_\Lambda^{(k)}\mathbf{Tr}\left(\boldsymbol{B}\boldsymbol{B}^T\right)\left[\sigma_\Lambda^2\left(1 + \frac{\chi}{\chi-2}rV_X\right)\right]$$

5 Exploration of the Posterior distribution

6 Implementation

6.1 Python Package Overview

Introduce the Python package, discussing its structure and main functionalities.

6.2 Bayesian Model Implementation

Detail how the Bayesian model is implemented in the package, including any specific libraries or frameworks used (e.g., PyMC3, Stan).

```
# Python code example for Bayesian model implementation
import pymc3 as pm
print("PyMC3 version:", pm.__version__)

# Bayesian model code snippet
with pm.Model() as model:
    # Model specification here
```

Listing 1: Bayesian Model Implementation Example

7 Results

Present results obtained using the package, possibly including data analysis examples, model diagnostics, and interpretation of the posterior distributions.

8 Conclusion

Summarize the key findings and implications of using the Bayesian approach in the Python package for data analysis. Discuss future directions and improvements.

References

[1] Silvia Montagna, Surya Tokdar, Brian Neelon, and David Dunson. Bayesian latent factor regression for functional and longitudinal data. *Biometrics*, 68:1064–1073, 12 2012.

[2] Heesang Noh, Taeryon Choi, Jinsu Park, and Yeonseung Chung. Bayesian latent factor regression for multivariate functional data with variable selection. *Journal of the Korean Statistical Society*, 49(3):901–923, 2020.

A Prior Statistical Moments of the Latent Factors

Proposition A.1 (Expected Latent Factor)**.**

$$\mathbb{E}\left[\Lambda_{j\,\ell}^{(i)}\right] = 0$$

0

Proof.

$$\mathbb{E}\left[\Lambda_{j\,\ell}^{(i)}\right] = \mathbb{E}\left[\mathbb{E}\left[\Lambda_{j\,\ell}^{(i)}\middle|\,\tau_{\lambda\,\ell}^{(i)}\right]\right] = \mathbb{E}\left[0\right] = 0$$

□

Proposition A.2 (Expected Reciprocal of the Precision of the Latent Factors τ_λ)**.** *Let $\alpha_1 > 1$ and $\alpha_2 > 1$, then:*

$$\mathbb{E}\left[\left(\tau_{\lambda\,\ell}^{(i)}\right)^{-1}\right] = \frac{1}{\alpha_1 - 1}\left(\frac{1}{\alpha_2 - 1}\right)^{\ell-1}$$

Proof.

$$\mathbb{E}\left[\left(\tau_{\lambda\,\ell}^{(i)}\right)^{-1}\right] = \mathbb{E}\left[\prod_{\gamma=1}^{\ell}\left(\delta_{\gamma}^{(i)}\right)^{-1}\right] = \prod_{\gamma=1}^{\ell}\mathbb{E}\left[\left(\delta_{\gamma}^{(i)}\right)^{-1}\right]$$

If $\ell = 1$,

$$\mathbb{E}\left[\left(\tau_{\lambda\,1}^{(i)}\right)^{-1}\right] = \frac{1}{\alpha_1 - 1} = \frac{1}{\alpha_1 - 1}\left(\frac{1}{\alpha_2 - 1}\right)^{0} = \frac{1}{\alpha_1 - 1}\left(\frac{1}{\alpha_2 - 1}\right)^{1-1}$$

using the well known moments of the Inverse-Gamma distribution.

If $\ell \geq 2$:

$$\mathbb{E}\left[\left(\delta_{1}^{(i)}\right)^{-1}\right]\prod_{\gamma=2}^{\ell}\mathbb{E}\left[\left(\delta_{\gamma}^{(i)}\right)^{-1}\right] = \frac{1}{\alpha_1 - 1}\prod_{\gamma=1}^{\ell}\frac{1}{\alpha_2 - 1} = \frac{1}{\alpha_1 - 1}\left(\frac{1}{\alpha_2 - 1}\right)^{\ell-1}$$

□

Remark A.2.1. *Following the same proof, let $\alpha_2 > 1$, then:*

$$\mathbb{E}\left[\left(\frac{\tau_{\lambda\,\ell}^{(i)}}{\tau_{\lambda\,1}^{(i)}}\right)^{-1}\right] = \left(\frac{1}{\alpha_2 - 1}\right)^{\ell-1}$$

Proposition A.3 (Variance of the Latent Factor)**.** *Let $\alpha_1 > 1$, $\alpha_2 > 1$ and $\nu > 2$,*

$$\mathbb{V}ar\left[\Lambda_{j\,\ell}^{(i)}\right] = \frac{\nu}{\nu - 2}\frac{1}{\alpha_1 - 1}\left(\frac{1}{\alpha_2 - 1}\right)^{\ell-1}$$

Proof.

$$\mathbb{V}\mathrm{ar}\left[\Lambda_{j\,\ell}^{(i)}\right] = \mathbb{V}\mathrm{ar}\left[\mathbb{E}\left[\Lambda_{j\,\ell}^{(i)}\middle|\,\tau_{\lambda\,\ell}^{(i)}\right]\right] + \mathbb{E}\left[\mathbb{V}\mathrm{ar}\left[\Lambda_{j\,\ell}^{(i)}\middle|\,\tau_{\lambda\,\ell}^{(i)}\right]\right] = \mathbb{E}\left[\mathbb{V}\mathrm{ar}\left[\Lambda_{j\,\ell}^{(i)}\middle|\,\tau_{\lambda\,\ell}^{(i)}\right]\right]$$

$$\mathbb{E}\left[\mathbb{V}\mathrm{ar}\left[\Lambda_{j\,\ell}^{(i)}\middle|\,\tau_{\lambda\,\ell}^{(i)}\right]\right] = \mathbb{E}\left[\frac{\nu}{\nu - 2}\left(\tau_{\lambda\,\ell}^{(i)}\right)^{-1}\right] = \frac{\nu}{\nu - 2}\mathbb{E}\left[\left(\tau_{\lambda\,\ell}^{(i)}\right)^{-1}\right]$$

□

Proposition A.4 (Variance of the Reciprocal of the Precision of the Latent Factors τ_λ)**.** *Let $\alpha_1 > 2$ and $\alpha_2 > 2$, then:*

$$\mathbb{V}ar\left[\left(\tau_{\lambda\,\ell}^{(i)}\right)^{-1}\right] = \frac{1}{(\alpha_1-1)(\alpha_1-2)}\left[\frac{1}{(\alpha_2-1)(\alpha_2-2)}\right]^{\ell-1}\cdot\left[1-\frac{\alpha_1-2}{\alpha_1-1}\left(\frac{\alpha_2-2}{\alpha_2-1}\right)^{\ell-1}\right]$$

Proof. If $\ell = 1$ then:

$$\mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,1}^{(i)}\right)^{-1}\right] = \mathbb{V}\mathrm{ar}\left[\left(\delta_{\lambda\,1}^{(i)}\right)^{-1}\right] = \frac{1}{(\alpha_1-1)^2(\alpha_1-2)}$$

If $\ell \geq 2$ then:

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,\ell}^{(i)}\right)^{-1}\right] &= \mathbb{V}\mathrm{ar}\left[\mathbb{E}\left[\left(\tau_{\lambda\,\ell-1}^{(i)}\right)^{-1}\left(\delta_{\lambda\,\ell}^{(i)}\right)^{-1}\middle|\,\delta_{\lambda\,\ell}^{(i)}\right]\right]+\\ &+\mathbb{E}\left[\mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,\ell-1}^{(i)}\right)^{-1}\left(\delta_{\lambda\,\ell}^{(i)}\right)^{-1}\middle|\,\delta_{\lambda\,\ell}^{(i)}\right]\right]\end{aligned}$$

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left[\mathbb{E}\left[\left(\tau_{\lambda\,\ell-1}^{(i)}\right)^{-1}\left(\delta_{\lambda\,\ell}^{(i)}\right)^{-1}\middle|\,\delta_{\lambda\,\ell}^{(i)}\right]\right] &= \mathbb{V}\mathrm{ar}\left[\left(\delta_{\lambda\,\ell}^{(i)}\right)^{-1}\mathbb{E}\left[\left(\tau_{\lambda\,\ell-1}^{(i)}\right)^{-1}\middle|\,\delta_{\lambda\,\ell}^{(i)}\right]\right] =\\ &= \mathbb{V}\mathrm{ar}\left[\left(\delta_{\lambda\,\ell}^{(i)}\right)^{-1}\mathbb{E}\left[\left(\tau_{\lambda\,\ell-1}^{(i)}\right)^{-1}\right]\right] =\\ &= \mathbb{V}\mathrm{ar}\left[\left(\delta_{\lambda\,\ell}^{(i)}\right)^{-1}\frac{1}{\alpha_1-1}\left(\frac{1}{\alpha_2-1}\right)^{\ell-2}\right] =\\ &= \left[\frac{1}{\alpha_1-1}\left(\frac{1}{\alpha_2-1}\right)^{\ell-2}\right]^2\mathbb{V}\mathrm{ar}\left[\delta_{\lambda\,\ell}^{(i)}\right] =\\ &= \left[\frac{1}{\alpha_1-1}\left(\frac{1}{\alpha_2-1}\right)^{\ell-2}\right]^2\frac{1}{(\alpha_2-1)^2(\alpha_2-2)}\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[\mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,\ell-1}^{(i)}\right)^{-1}\left(\delta_{\lambda\,\ell}^{(i)}\right)^{-1}\middle|\,\delta_{\lambda\,\ell}^{(i)}\right]\right] &= \mathbb{E}\left[\left(\delta_{\lambda\,\ell}^{(i)}\right)^{-2}\mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,\ell-1}^{(i)}\right)^{-1}\middle|\,\delta_{\lambda\,\ell}^{(i)}\right]\right] =\\ &= \mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,\ell-1}^{(i)}\right)^{-1}\right]\mathbb{E}\left[\left(\delta_{\lambda\,\ell}^{(i)}\right)^{-2}\right] \overset{(A.4.1)}{=}\\ &= \frac{1}{(\alpha_2-1)(\alpha_2-2)}\mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,\ell-1}^{(i)}\right)^{-1}\right]\end{aligned}$$

Where we used:

$$\begin{aligned}\mathbb{E}\left[\left(\delta_{\lambda\,\ell}^{(i)}\right)^{-2}\right] &= \mathbb{E}^2\left[\left(\delta_{\lambda\,\ell}^{(i)}\right)^{-1}\right]+\mathbb{V}\mathrm{ar}\left[\left(\delta_{\lambda\,\ell}^{(i)}\right)^{-1}\right] =\\ &= \left(\frac{1}{\alpha_2-1}\right)^2+\frac{1}{(\alpha_2-1)^2(\alpha_2-2)} = \frac{1}{(\alpha_2-1)(\alpha_2-2)}\end{aligned} \tag{A.4.1}$$

This proves the following recurrence equation:

$$\mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,\ell}^{(i)}\right)^{-1}\right] = \frac{1}{(\alpha_2-1)(\alpha_2-2)}\mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,\ell-1}^{(i)}\right)^{-1}\right]+\frac{1}{\alpha_2-2}\left(\frac{\alpha_2-1}{\alpha_1-1}\right)^2\left(\frac{1}{\alpha_2-1}\right)^{2\ell}$$

$$\mathbb{V}\mathrm{ar}\left[\left(\tau_{\lambda\,1}^{(i)}\right)^{-1}\right] = \frac{1}{(\alpha_1-1)^2(\alpha_1-2)}$$

Lemma A.4.1. *Let $0 < A < 1$ and $0 < B < 1$. A recurrence equation of the type:*

$$F_\ell = \frac{B^2}{1-B} F_{\ell-1} + \frac{A^2}{1-B} B^{2\ell-1}$$
$$F_1 = \frac{A^3}{1-A}$$

has solution:

$$F_\ell = \frac{A^2}{1-A} \left(\frac{B^2}{1-B} \right)^{\ell-1} \left\{ A + (1-A) \left[1 - (1-B)^{\ell-1} \right] \right\}$$

Now, to prove the proposition, we recognize the relationship between the recurrence equation of the lemma and the one for $\mathbb{V}\mathrm{ar}\left[\left(\tau^{(i)}_{\lambda\,\ell} \right)^{-1} \right]$, due to the substitution:

$$A := \frac{1}{\alpha_1 - 1}, \qquad B := \frac{1}{\alpha_2 - 1}$$

which induces:

$$\frac{1}{\alpha_1 - 2} = \frac{A}{1-A}, \qquad \frac{1}{\alpha_2 - 2} = \frac{B}{1-B}$$

Direct substitution in the solution of Lemma A.4.1 yields:

$$\mathbb{V}\mathrm{ar}\left[\left(\tau^{(i)}_{\lambda\,\ell} \right)^{-1} \right] = \frac{1}{(\alpha_1 - 1)(\alpha_1 - 2)} \left[\frac{1}{(\alpha_2 - 1)(\alpha_2 - 2)} \right]^{\ell-1} \cdot \left[1 - \frac{\alpha_1 - 2}{\alpha_1 - 1} \left(\frac{\alpha_2 - 2}{\alpha_2 - 1} \right)^{\ell-1} \right]$$

$\square$

Remark A.4.1. *By observing that*

$$\left(\frac{\tau^{(i)}_{\lambda\,\ell}}{\tau^{(i)}_{\lambda\,1}} \right)^{-1} = \prod_{\gamma=2}^{\ell} \delta^{(i)}_\gamma$$

one can infer the formula for $\mathbb{V}ar\left[\left(\frac{\tau^{(i)}_{\lambda\,\ell}}{\tau^{(i)}_{\lambda\,1}} \right)^{-1} \right]$ *using the same result of the previous proposition, with the substitution $\alpha_1 = \alpha_2$ and correcting for the value of ℓ. It follows:*
Let $\alpha_2 > 2$,

$$\mathbb{V}ar\left[\left(\frac{\tau^{(i)}_{\lambda\,\ell}}{\tau^{(i)}_{\lambda\,1}} \right)^{-1} \right] = \left[\frac{1}{(\alpha_2 - 1)(\alpha_2 - 2)} \right]^{\ell-1} \cdot \left[1 - \left(\frac{\alpha_2 - 2}{\alpha_2 - 1} \right)^{\ell-1} \right]$$

Remark A.4.2.

$$\alpha_1, \alpha_2 > 2 \quad \longrightarrow \quad \frac{\mathbb{V}ar\left[\left(\tau^{(i)}_{\lambda\,\ell}\right)^{-1}\right]}{\mathbb{E}^2\left[\left(\tau^{(i)}_{\lambda\,\ell}\right)^{-1}\right]} = \frac{\alpha_1 - 1}{\alpha_1 - 2}\left(\frac{\alpha_2 - 1}{\alpha_2 - 2}\right)^{\ell-1} - 1$$

$$\alpha_2 > 2 \quad \longrightarrow \quad \frac{\mathbb{V}ar\left[\left(\frac{\tau^{(i)}_{\lambda\,\ell}}{\tau^{(i)}_{\lambda\,1}}\right)^{-1}\right]}{\mathbb{E}^2\left[\left(\frac{\tau^{(i)}_{\lambda\,\ell}}{\tau^{(i)}_{\lambda\,1}}\right)^{-1}\right]} = \left(\frac{\alpha_2 - 1}{\alpha_2 - 2}\right)^{\ell-1} - 1$$

Proposition A.5. *If* $(i_1, j_1, \ell_1) \neq (i_2, j_2, \ell_2)$*:*

$$\mathbb{C}ov\left[\Lambda^{(i_1)}_{j_1\,\ell_1}, \Lambda^{(i_2)}_{j_2\,\ell_2}\right] = 0$$

Proof. This is a consequence of 1) the independence of the entries of the latent factors conditional on $\boldsymbol{\tau}_\lambda$ and 2) the latent factors are centered conditional on $\boldsymbol{\tau}_\lambda$. □

Remark A.5.1. *The covariance is* 0 *due to the centered nature of the latent factors, but an internal structure is still present and is captured by the forth mixed moments* $\mathbb{E}\left[\left(\Lambda^{(i_1)}_{j_1\,\ell_1}\Lambda^{(i_2)}_{j_2\,\ell_2}\right)^2\right]$*. This allows to forget the sign of the deviation from the mean, and focus only on the co-deviation in absolute terms:*

If $(i_1, \ell_1) = (i_2, \ell_2)$*, but* $j_1 \neq j_2$*:*

$$\mathbb{E}\left[\left(\Lambda^{(i)}_{j_1\,\ell}\Lambda^{(i)}_{j_2\,\ell}\right)^2\right] = \mathbb{E}\left[\left(\tau^{(i)}_{\lambda\,\ell}\right)^{-2}\right] = \mathbb{V}ar\left[\left(\tau^{(i)}_{\lambda\,\ell}\right)^{-1}\right] + \mathbb{E}^2\left[\left(\tau^{(i)}_{\lambda\,\ell}\right)^{-1}\right] =$$
$$= \frac{1}{\alpha_1 - 1}\frac{1}{\alpha_1 - 2}\left[\frac{1}{(\alpha_2 - 1)(\alpha_2 - 2)}\right]^{\ell-1} \tag{A.5.1.1}$$

If $i_1 = i_2$ *but* $\ell_1 \neq \ell_2$*, define* $\ell_m := \min(\ell_1, \ell_2)$*,* $\ell_M := \max(\ell_1, \ell_2)$*:*

$$\mathbb{E}\left[\left(\Lambda^{(i)}_{j_1\,\ell_1}\Lambda^{(i)}_{j_2\,\ell_2}\right)^2\right] = \mathbb{E}\left[\left(\tau^{(i)}_{\lambda\,\ell_1}\right)^{-1}\left(\tau^{(i)}_{\lambda\,\ell_2}\right)^{-1}\right] = \left\{\mathbb{V}ar\left[\left(\tau^{(i)}_{\lambda\,\ell_m}\right)^{-1}\right] + \mathbb{E}^2\left[\left(\tau^{(i)}_{\lambda\,\ell_m}\right)^{-1}\right]\right\}\left(\frac{1}{\alpha_2 - 1}\right)^{|\ell_1 - \ell_2|} =$$
$$= \frac{1}{\alpha_1 - 1}\frac{1}{\alpha_1 - 2}\left(\frac{1}{\alpha_2 - 1}\right)^{\ell_M - 1}\left(\frac{1}{\alpha_2 - 2}\right)^{\ell_m - 1} \tag{A.5.1.2}$$

Finally, if $i_1 \neq i_2$*:*

$$\mathbb{E}\left[\left(\Lambda^{(1)}_{j_1\,\ell_1}\Lambda^{(2)}_{j_2\,\ell_2}\right)^2\right] = \mathbb{E}\left[\left(\tau^{(1)}_{\lambda\,\ell_1}\right)^{-1}\left(\tau^{(2)}_{\lambda\,\ell_2}\right)^{-1}\right] = \mathbb{E}\left[\left(\tau^{(1)}_{\lambda\,\ell_1}\right)^{-1}\right]\mathbb{E}\left[\left(\tau^{(2)}_{\lambda\,\ell_2}\right)^{-1}\right] = \left(\frac{1}{\alpha_1 - 1}\right)^2\left(\frac{1}{\alpha_2 - 1}\right)^{\ell_1 + \ell_2 - 2} \tag{A.5.1.3}$$

B Prior Predictive Moments

As previously discussed:

Proposition B.1 (Prior Predictive Distribution)**.**

$$\boldsymbol{Y}_s | \Theta \overset{i.i.d.}{\sim} \mathcal{N}\left(\begin{bmatrix} \boldsymbol{B}^{(1)}\boldsymbol{\Lambda}^{(1)}\boldsymbol{\beta} X_{\cdot,s} \\ \boldsymbol{B}^{(2)}\boldsymbol{\Lambda}^{(2)}\boldsymbol{\beta} X_{\cdot,s} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right), \qquad s = 1, \ldots, N \tag{6}$$

where

$$\boldsymbol{\Sigma}_{ii} = {\psi^{(i)}}^2 \boldsymbol{I} + \boldsymbol{B}^{(i)} \left[\Delta_\theta^{(i)} + \boldsymbol{\Lambda}^{(i)} {\boldsymbol{\Lambda}^{(i)}}^T \right] {\boldsymbol{B}^{(i)}}^T \tag{7}$$

$$\boldsymbol{\Sigma}_{ij} = \boldsymbol{B}^{(i)} \boldsymbol{\Lambda}^{(i)} {\boldsymbol{\Lambda}^{(j)}}^T {\boldsymbol{B}^{(j)}}^T \tag{8}$$

Proposition B.2 (Expectation of the Prior Predictive)**.**

$$\mathbb{E}\left[\boldsymbol{Y}_s^{(\iota)} \right] = \boldsymbol{0}$$

Proof.

$$\mathbb{E}\left[\boldsymbol{Y}_s^{(\iota)} \right] = \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{Y}_s^{(\iota)} \middle| \Theta \right] \right] = \mathbb{E}\left[\boldsymbol{B}^{(\iota)} \boldsymbol{\Lambda}^{(\iota)} \boldsymbol{\beta} X_{\cdot,s} \right] = \sum_{j=1}^{p_\iota} \sum_{\ell=1}^{k} \sum_{i=1}^{r} \boldsymbol{B}_{\cdot,j}{}^{(\iota)} \mathbb{E}\left[\Lambda_{j\ell}^{(\iota)} \right] \mathbb{E}\left[\beta_{\ell i} \right] X_{i,s} = \boldsymbol{0}$$

□

Proposition B.3 (Variance of the Prior Predictive)**.**

Assuming $\nu, \chi > 2$ and $\alpha_\sigma, \alpha_\psi, \alpha_1, \alpha_2 > 1$:

If $\alpha_2 \neq 2$:

$$\mathbb{V}ar\left[\boldsymbol{Y}_{h,s}^{(\iota)} \right] = \frac{\beta_\psi}{\alpha_\psi - 1} + \left(\sum_{j=1}^{p_\iota} {B_{h,j}^{(\iota)}}^2 \right) \cdot \left\{ \frac{\beta_\sigma}{\alpha_\sigma - 1} + \right.$$
$$\left. + \frac{\nu}{\nu - 2} \frac{\alpha_2 - 1}{\alpha_1 - 1} \frac{1}{\alpha_2 - 2} \left[1 - \left(\frac{1}{\alpha_2 - 1} \right)^k \right] \left[1 + \frac{\chi}{\chi - 2} \left(\sum_{i=1}^{r} X_{i,s}^2 \right) \right] \right\}$$

If $\alpha_2 = 2$:

$$\mathbb{V}ar\left[\boldsymbol{Y}_{h,s}^{(\iota)} \right] = \frac{\beta_\psi}{\alpha_\psi - 1} + \left(\sum_{j=1}^{p_\iota} {B_{h,j}^{(\iota)}}^2 \right) \cdot \left\{ \frac{\beta_\sigma}{\alpha_\sigma - 1} + \frac{\nu}{\nu - 2} \frac{k}{\alpha_1 - 1} \left[1 + \frac{\chi}{\chi - 2} \left(\sum_{i=1}^{r} X_{i,s}^2 \right) \right] \right\}$$

Proof.

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left[\boldsymbol{Y}_{h,s}^{(\iota)} \right] &= \mathbb{E}\left[\mathbb{V}\mathrm{ar}\left[\boldsymbol{Y}_{h,s}^{(\iota)} \middle| \Theta \right] \right] + \mathbb{V}\mathrm{ar}\left[\mathbb{E}\left[\boldsymbol{Y}_{h,s}^{(\iota)} \middle| \Theta \right] \right] \\
\mathbb{E}\left[\mathbb{V}\mathrm{ar}\left[\boldsymbol{Y}_{h,s}^{(\iota)} \middle| \Theta \right] \right] &= \mathbb{E}\left[{\psi^{(\iota)}}^2 + \boldsymbol{B}_{h,\cdot}^{(\iota)} \left[\Delta_\theta^{(\iota)} + \boldsymbol{\Lambda}^{(\iota)} {\boldsymbol{\Lambda}^{(\iota)}}^T \right] {\boldsymbol{B}_{h,\cdot}^{(\iota)}}^T \right] = \\
&= \frac{\beta_\psi}{\alpha_\psi - 1} + \boldsymbol{B}_{h,\cdot}^{(\iota)} \left\{ \mathbb{E}\left[\Delta_\theta^{(\iota)} \right] + \mathbb{E}\left[\boldsymbol{\Lambda}^{(\iota)} {\boldsymbol{\Lambda}^{(\iota)}}^T \right] \right\} {\boldsymbol{B}_{h,\cdot}^{(\iota)}}^T \\
\mathbb{V}\mathrm{ar}\left[\mathbb{E}\left[\boldsymbol{Y}_{h,s}^{(\iota)} \middle| \Theta \right] \right] &= \mathbb{V}\mathrm{ar}\left[\boldsymbol{B}_{h,\cdot}^{(\iota)} \boldsymbol{\Lambda}^{(\iota)} \boldsymbol{\beta} X_{\cdot,s} \right] = \mathbb{E}\left[\left(\boldsymbol{B}_{h,\cdot}^{(\iota)} \boldsymbol{\Lambda}^{(\iota)} \boldsymbol{\beta} X_{\cdot,s} \right)^2 \right]
\end{aligned}$$

Lemma B.3.1.

$$\mathbb{E}\left[\Delta_\theta^{(\iota)}\right] = \frac{\beta_\sigma}{\alpha_\sigma - 1}\boldsymbol{I}$$

Lemma B.3.2.

$$\mathbb{E}\left[\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\Lambda}^{(\iota)T}\right] = \frac{\nu}{\nu-2}\frac{\alpha_2-1}{\alpha_1-1}\frac{1}{\alpha_2-2}\left[1-\left(\frac{1}{\alpha_2-1}\right)^k\right]\boldsymbol{I}$$

Indeed:

$$\mathbb{E}\left[\left(\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\Lambda}^{(\iota)T}\right)_{j_1 j_2}\right] = \sum_{\ell=1}^{k}\mathbb{E}\left[\Lambda_{j_1\ell}^{(\iota)}\Lambda_{j_2\ell}^{(\iota)}\right] = \begin{cases}\sum_{\ell=1}^{k}\frac{\nu}{\nu-2}\frac{1}{\alpha_1-1}\left(\frac{1}{\alpha_2-1}\right)^{\ell-1} & \text{if } j := j_1 = j_2,\\ 0 & \text{if } j_1 \neq j_2.\end{cases}$$

$$\sum_{\ell=1}^{k}\left(\frac{1}{\alpha_2-1}\right)^{\ell-1} = \frac{1-\left(\frac{1}{\alpha_2-1}\right)^k}{1-\frac{1}{\alpha_2-1}} = \frac{\alpha_2-1}{\alpha_2-2}\left[1-\left(\frac{1}{\alpha_2-1}\right)^k\right]$$

(With the understanding that in the case $\alpha_2 = 2$ we get k)

Lemma B.3.3.

$$\mathbb{E}\left[\left(\boldsymbol{B}_{h,\cdot}^{(\iota)}\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta}X_{\cdot,s}\right)^2\right] = \frac{\chi}{\chi-2}\frac{\nu}{\nu-2}\frac{\alpha_2-1}{\alpha_1-1}\frac{1}{\alpha_2-2}\left[1-\left(\frac{1}{\alpha_2-1}\right)^k\right]\sum_{j=1}^{p_\iota}\sum_{i=1}^{r}{B_{h,j}^{(\iota)}}^2 X_{i,s}^2$$

Indeed:

$$\begin{aligned}\mathbb{E}\left[\left(\boldsymbol{B}_{h,\cdot}^{(\iota)}\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta}X_{\cdot,s}\right)^2\right] &= \mathbb{E}\left[\sum_{j_1=1}^{p_\iota}\sum_{j_2=1}^{p_\iota}B_{h,j_1}^{(\iota)}B_{h,j_2}^{(\iota)}\left(\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta}\boldsymbol{X}\right)_{j_1,s}\left(\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta}\boldsymbol{X}\right)_{j_2,s}\right] =\\ &= \sum_{j_1=1}^{p_\iota}\sum_{j_2=1}^{p_\iota}B_{h,j_1}^{(\iota)}B_{h,j_2}^{(\iota)}\mathbb{E}\left[\left(\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta}\boldsymbol{X}\right)_{j_1,s}\left(\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta}\boldsymbol{X}\right)_{j_2,s}\right] =\\ &= \sum_{j_1=1}^{p_\iota}\sum_{j_2=1}^{p_\iota}\sum_{i_1=1}^{r}\sum_{i_2=1}^{r}B_{h,j_1}^{(\iota)}B_{h,j_2}^{(\iota)}X_{i_1,s}X_{i_2,x}\mathbb{E}\left[\left(\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta}\right)_{j_1,i_1}\left(\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta}\right)_{j_2,i_2}\right]\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[\left(\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta}\right)_{j_1,i_1}\left(\boldsymbol{\Lambda}^{(\iota)}\boldsymbol{\beta}\right)_{j_2,i_2}\right] &= \sum_{\ell_1=1}^{k}\sum_{\ell_2=1}^{k}\mathbb{E}\left[\Lambda_{j_1,\ell_1}^{(\iota)}\Lambda_{j_2,\ell_2}^{(\iota)}\beta_{\ell_1,i_1}\beta_{\ell_2,i_2}\right] =\\ &= \sum_{\ell_1=1}^{k}\sum_{\ell_2=1}^{k}\mathbb{E}\left[\Lambda_{j_1,\ell_1}^{(\iota)}\Lambda_{j_2,\ell_2}^{(\iota)}\right]\mathbb{E}\left[\beta_{\ell_1,i_1}\beta_{\ell_2,i_2}\right] =\\ &= \frac{\chi}{\chi-2}\sum_{\ell_1=1}^{k}\sum_{\ell_2=1}^{k}\mathbb{E}\left[\Lambda_{j_1,\ell_1}^{(\iota)}\Lambda_{j_2,\ell_2}^{(\iota)}\right]\delta^K(\ell_1,\ell_2)\,\delta^K(i_1,i_2) =\\ &= \frac{\chi}{\chi-2}\sum_{\ell=1}^{k}\mathbb{E}\left[\Lambda_{j_1,\ell}^{(\iota)}\Lambda_{j_2,\ell}^{(\iota)}\right]\delta^K(i_1,i_2) =\\ &= \frac{\chi}{\chi-2}\frac{\nu}{\nu-2}\frac{\alpha_2-1}{\alpha_1-1}\frac{1}{\alpha_2-2}\left[1-\left(\frac{1}{\alpha_2-1}\right)^k\right]\delta^K(i_1,i_2)\,\delta^K(j_1,j_2)\end{aligned}$$

□

Proposition B.4 (Limiting behavior of the Variance of the Prior Predictive)**.**

Let $\nu, \chi, \alpha_2 > 2$ *and* $\alpha_\sigma, \alpha_\psi, \alpha_1 > 1$*:*

$$\lim_{k\to+\infty} \mathbb{V}ar\left[\boldsymbol{Y}^{(\iota)}_{h,s}\right] = \frac{\beta_\psi}{\alpha_\psi - 1} + \left(\sum_{j=1}^{p_\iota} {B^{(\iota)}_{h,j}}^2\right) \cdot \left\{ \frac{\beta_\sigma}{\alpha_\sigma - 1} + \frac{\nu}{\nu - 2} \frac{1}{\alpha_1 - 1} \frac{\alpha_2 - 1}{\alpha_2 - 2} \left[1 + \frac{\chi}{\chi - 2} \left(\sum_{i=1}^{r} X^2_{i,s}\right)\right] \right\}$$

Proof. A direct consequence of the fact that for $\alpha_2 > 2$,

$$\lim_{k\to+\infty} \left(\frac{1}{\alpha_2 - 1}\right)^k = 0$$

$\square$

Proposition B.5 (Total Variance of the Prior Predictive)**.**

Let $\nu, \chi, \alpha_2 > 2$ *and* $\alpha_\sigma, \alpha_\psi, \alpha_1 > 1$*:*

$$\Sigma^k_T(\boldsymbol{Y}_s) := \left(\sum_{h_1=1}^{L_1} \mathbb{V}ar[\boldsymbol{Y}_{h_1,s}]\right) + \left(\sum_{h_2=1}^{L_2} \mathbb{V}ar[\boldsymbol{Y}_{h_2,s}]\right) = (L_1 + L_2) \frac{\beta_\psi}{\alpha_\psi - 1} + \mathbf{Tr}\left(\boldsymbol{B}\boldsymbol{B}^T\right) \times$$
$$\times \left\{ \frac{\beta_\sigma}{\alpha_\sigma - 1} + \frac{\nu}{\nu - 2} \frac{\alpha_2 - 1}{(\alpha_1 - 1)(\alpha_2 - 2)} \left[1 - \left(\frac{1}{\alpha_2 - 1}\right)^k\right] \left(1 + \frac{\chi}{\chi - 2} \|\boldsymbol{X}_{\cdot,s}\|^2_2\right) \right\}$$

$$\Sigma^\infty_T(\boldsymbol{Y}_s) = (L_1 + L_2) \frac{\beta_\psi}{\alpha_\psi - 1} + \mathbf{Tr}\left(\boldsymbol{B}\boldsymbol{B}^T\right) \left\{ \frac{\beta_\sigma}{\alpha_\sigma - 1} + \frac{\nu}{\nu - 2} \frac{\alpha_2 - 1}{(\alpha_1 - 1)(\alpha_2 - 2)} \left(1 + \frac{\chi}{\chi - 2} \|\boldsymbol{X}_{\cdot,s}\|^2_2\right) \right\}$$

where

$$\boldsymbol{B} := \begin{bmatrix} \boldsymbol{B}^{(1)} \\ \boldsymbol{B}^{(2)} \end{bmatrix}$$

C Third Appendix