A Bayesian Approach for Large Asset Allocation

Mihnea S. Andrei[1] & John S. J. Hsu[2]

[1] Department of Mathematical Sciences, Worcester Polytechnic Institute, USA

[2] Department of Statistics and Applied Probability, University of California, Santa Barbara, USA

Correspondence: Mihnea S. Andrei, Department of Mathematical Sciences, Worcester Polytechnic Institute, USA

Abstract

The Black-Litterman model combines investor's personal views with historical data and gives optimal portfolio weights. In (Andrei & Hsu, 2020), they reviewed the original Black-Litterman model and modified it in order to fit it into a Bayesian framework, when a certain number of assets is considered. They used the idea by (Leonard & Hsu, 1992) for a multivariate normal prior on the logarithm of the covariance matrix. When implemented and applied to a large number of assets such as all the $S\&P500$ companies, they ran into memory allocation and running time issues. In this paper, we reduce the dimensions by considering Bayesian factor models, which solve the asset allocation problems for a large number of assets. In addition, we will conduct sensitivity analysis for the confidence levels that the investors have to input.

Keywords: Black-Litterman, covariance matrix, investor's views prior, logarithmic covariance prior, portfolio allocation, statistics

1. Background Introduction

1.1 Black-Litterman Asset Allocation Model

The Black-Litterman asset allocation model developed by Black and Litterman in 1992 has been widely used for decades and it is presented in more detail in the paper by (He & Litterman, 2002). Suppose that m assets in the market are considered. The returns of those assets $\mathbf{r} = (r_1, r_2, \ldots, r_m)^T$ follow a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. That is

$$\mathbf{r} \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{1}$$

Black and Litterman proposed the following CAPM (Capital Asset Pricing Model) prior for the mean of the return:

$$\boldsymbol{\mu} \sim N_m(\boldsymbol{\pi}, \tau\boldsymbol{\Sigma}) \tag{2}$$

where $\boldsymbol{\pi}$ is the equilibrium risk parameter and τ indicates the uncertainty of the CAPM prior. In addition to the CAPM prior, they also take investor's views into consideration. Suppose that the investor has k views. His or her views can be expressed in the equation:

$$\mathbf{P}\boldsymbol{\mu} \sim N_k(\mathbf{q}_0, \boldsymbol{\Omega}) \tag{3}$$

where the matrix $\mathbf{P}$ is a $k \times m$ matrix, $\mathbf{q}_0$ is a $k \times 1$ vector, and $\boldsymbol{\Omega}$ is a $k \times k$ matrix, usually diagonal. Each row in $\mathbf{P}$ and $\mathbf{q}_0$ represents a personal view.

2. A Bayesian Approach for Large Asset Allocation, When Historical Data is Unavailable

2.1 Introduction

Black and Litterman derived the corresponding optimal portfolio weights and these weights can be computed by inputting all involved parameters into their probability model. Along the same lines, (Andrei & Hsu, 2020) considered a Bayesian statistical model instead for asset allocation for a certain fixed number of assets. They assumed that the n returns $\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_n$ are independent, where each $\mathbf{r}_i$ is a vector containing the returns of m assets, which are normally distributed as in equation (1) for $i \in \{1, 2, ..., n\}$. The priors for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are assumed to be independent, where the prior on $\boldsymbol{\mu}$ can be expressed as the investor's views in equation (3). Following (Leonard & Hsu, 1992), they considered a multivariate normal prior for the logarithm of the covariance matrix $\mathbf{A} = \log(\boldsymbol{\Sigma})$. The corresponding covariance matrix involved in the multivariate normal distribution is of size $\frac{1}{2}m(m+1) \times \frac{1}{2}m(m+1)$. Therefore, if one considers the whole $S\&P500$, the size of this random matrix in terms of memory would be of around 106GB. Because of this issue, we decided to incorporate factors in order to reduce the dimension. Hence, as we will see in the following sections, after applying factor models, we will incorporate a prior on the covariance matrix of the common factors instead of incorporating a prior directly on the covariance matrix of the returns. The dimension of the covariance matrix of the common factors is $q \times q$, where q

represents the number of common factors, which is much smaller than $m \times m$, the dimension of the covariance matrix of the returns, where m is the number of assets.

Our work consists of combining the two Bayesian versions for the Black-Litterman model from (Andrei & Hsu, 2020) with the work of (Lee et al., 2007) and the work of (Lee & Shi, 2000).

2.2 Assumptions

We consider a factor model on n returns with m assets each and k investor's views:

$$\mathbf{r}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f}_i + \mathbf{e}_i$$

Where we introduced the following notation:

- $\boldsymbol{\mu}$ is a vector of expected returns for each of the m assets

- $\boldsymbol{\Lambda}$ is the loading matrix from the factor model

- $\mathbf{f}_i$ are common factors

- $\mathbf{e}_i$ are unobserved errors which satisfy the following assumption:

$$\mathbf{e}_i | \boldsymbol{\Psi} \stackrel{iid.}{\sim} N_m(0, \boldsymbol{\Psi}) \text{ for all } i \in \{1, 2, ..., n\} \text{ and } \boldsymbol{\Psi} = diag(\Psi_1, ..., \Psi_m)$$

(1) Hence, from the above factor model, we obtain that the conditional distribution of the returns $\mathbf{r}_i$, given the factor $\mathbf{f}_i$ is:

$$\mathbf{r}_i | \boldsymbol{\mu}, \mathbf{f}_i, \boldsymbol{\Lambda}, \boldsymbol{\Psi} \sim N_m(\boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f}_i, \boldsymbol{\Psi}) \text{ for all } i \in \{1, 2, ..., n\}$$
$$\text{where } \mathbf{f}_i | \boldsymbol{\Phi} \stackrel{iid.}{\sim} N_q(0, \boldsymbol{\Phi}) \text{ for all } i \in \{1, 2, ..., n\}$$

(2) Next, we consider a two-stage prior for $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$:

$$\boldsymbol{\Lambda}_j | \Psi_j \stackrel{indep.}{\sim} N_q(\boldsymbol{\Lambda}_{0j}, \Psi_j\mathbf{H}_j)$$
$$\Psi_j \stackrel{indep.}{\sim} IG(\alpha_j, \beta_j) \text{ for all } j \in \{1, 2, ..., m\}$$

Here, $\boldsymbol{\Lambda}_j^T$ is the jth row in $\boldsymbol{\Lambda}$.

(3) Following the Black-Litterman approach in equation (3), we include a prior on the mean of the returns, which is projected through the investor's views:

$$\mathbf{P}\boldsymbol{\mu} \sim N_k(\mathbf{q}_0, \boldsymbol{\Omega})$$

(4) Following (Leonard & Hsu, 1992), we consider a multivariate normal prior for the logarithm of the covariance $\mathbf{A} = \log(\boldsymbol{\Phi}) = \left(a_{ij}\right)_{i,j=\{1,2,...,q\}}$. Let $\boldsymbol{\alpha} = \left[\alpha_1, ..., \alpha_q, \alpha_{q+1}, ..., \alpha_d\right] = Vec^*(\mathbf{A}) = \left[a_{11}, a_{22}, ..., a_{qq}|a_{12}, a_{23}, ..., a_{q-1q}|...|a_{1q}\right]$, where $d = \frac{1}{2}q(q + 1)$. Suppose that the variance related components $\alpha_1, ..., \alpha_q$ and covariance related components $\alpha_{q+1}, ..., \alpha_d$ are independent. We further assume that the variance related components follow independent normal distributions, each with mean θ_1 and variance σ_1^2 and the covariance related components follow independent normal distributions, each with mean θ_2 and variance σ_2^2. That is,

$$\boldsymbol{\alpha}|\boldsymbol{\theta}, \boldsymbol{\Delta} \sim N_d(\mathbf{J}\boldsymbol{\theta}, \boldsymbol{\Delta})$$
$$\text{where } \mathbf{J} = \begin{bmatrix} \mathbf{1}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{d-q} \end{bmatrix}, \boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \text{ and } \boldsymbol{\Delta} = \begin{bmatrix} \sigma_1^2\mathbf{I}_q & \mathbf{0} \\ \mathbf{0} & \sigma_2^2\mathbf{I}_{d-q} \end{bmatrix},$$

where $\mathbf{1}_q$ is a q dimensional vector of ones and $\mathbf{I}_q$ is a $q \times q$ identity matrix. Further, suppose that only vague information about $\theta_1, \theta_2, \sigma_1^2, \sigma_2^2$ is available. That is $\pi(\theta_1, \theta_2) \propto 1$ and $\pi(\sigma_1^2, \sigma_2^2) \propto 1$.

Therefore, all the model assumptions are:

$$\mathbf{r}_i|\boldsymbol{\mu}, \mathbf{f}_i, \boldsymbol{\Lambda}, \boldsymbol{\Psi} \sim N_m(\boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f}_i, \boldsymbol{\Psi}), \text{for all } i \in \{1, 2, ..., n\}$$

$$\mathbf{f}_i|\boldsymbol{\Phi} \overset{iid.}{\sim} N_q(\mathbf{0}, \boldsymbol{\Phi}), \text{ for all } i \in \{1, 2, ..., n\}$$

$$\mathbf{P}\boldsymbol{\mu} \sim N_k(\mathbf{q}_0, \boldsymbol{\Omega})$$

$$\boldsymbol{\alpha}|\boldsymbol{\theta}, \boldsymbol{\Delta} = Vec^*(log(\boldsymbol{\Phi}))|\boldsymbol{\theta}, \boldsymbol{\Delta} \sim N(\mathbf{J}\boldsymbol{\theta}, \boldsymbol{\Delta})$$

$$\pi(\theta_1, \theta_2) \propto 1$$

$$\pi(\sigma_1^2, \sigma_2^2) \propto 1$$

$$\boldsymbol{\Lambda}_j|\Psi_j \overset{iid.}{\sim} N_q(\boldsymbol{\Lambda}_{0j}, \Psi_j\mathbf{H}_j)$$

$$\Psi_j \overset{iid.}{\sim} IG(\alpha_j, \beta_j), \text{ for all } j \in \{1, 2, ..., m\}$$

(4)

2.3 Posterior Distributions

(a) Approximate conditional posterior for α given $\boldsymbol{\mu}, \sigma_1^2, \sigma_2^2, \mathbf{r}_1, ..., \mathbf{r}_n$.

Following (Andrei & Hsu, 2020), with the common factors playing the role of the returns, we can approximate the conditional posterior distribution of α given σ_1^2 and σ_2^2 with a normal distribution that has as mean the vector α^* and covariance matrix $(\mathbf{Q} + \mathbf{G})^{-1}$. Hence, we obtain that:

$$\alpha|\cdot \approx\sim N(\alpha^*, (\mathbf{Q} + \mathbf{G})^{-1}), \text{ where } \alpha^* = (\mathbf{Q} + \mathbf{G})^{-1}\mathbf{Q}\lambda,$$

$$\mathbf{G} = \left(\mathbf{I}_d - \mathbf{J}(\mathbf{J}^T\boldsymbol{\Delta}^{-1}\mathbf{J})^{-1}\mathbf{J}^T\boldsymbol{\Delta}^{-1}\right)^T \boldsymbol{\Delta}^{-1} \left(\mathbf{I}_d - \mathbf{J}(\mathbf{J}^T\boldsymbol{\Delta}^{-1}\mathbf{J})^{-1}\mathbf{J}^T\boldsymbol{\Delta}^{-1}\right)$$

(5)

and the matrix $\mathbf{Q}$ is computed in the following way. If we let $\mathbf{e}_i, d_i$ be the ith normalized eigenvector with its corresponding eigenvalue of $\mathbf{S} = \frac{1}{n}(\mathbf{r}_i - \boldsymbol{\mu})(\mathbf{r}_i - \boldsymbol{\mu})^T$, respectively, then $\mathbf{h}_{ij}$ is obtained by looking at the equation $Vec^*(log(\boldsymbol{\Phi}))^T\mathbf{h}_{ij} = \mathbf{e}_i^T log(\boldsymbol{\Phi})\mathbf{e}_j$ and identifying the coefficients of the entries in the $log(\boldsymbol{\Phi})$ matrix. With those $\mathbf{h}_{ij}$, we can finally compute $\mathbf{Q}$:

$$\mathbf{Q} = \frac{n}{2}\sum_{i=1}^{q}\mathbf{h}_{ii}\mathbf{h}_{ii}^T + n\sum_{i<j}^{q}\xi_{ij}\mathbf{h}_{ij}\mathbf{h}_{ij}^T, \text{ where}$$

$$\xi_{ij} = \frac{(d_i - d_j)^2}{d_id_j(log(d_i) - log(d_j))^2}$$

(6)

(b) Conditional posteriors for σ_1^2 and σ_2^2 given α

Following (Andrei & Hsu, 2020), with the number of stocks m replaced by the number of common factors q and $d = \frac{1}{2}q(q + 1)$:

$$\sigma_1^2|\cdot \sim IG\left(\frac{q-3}{2}, \frac{1}{2}\sum_{i=1}^{q}(\alpha_i - \overline{\alpha}_v)^2\right), \text{ where } \overline{\alpha}_v = \frac{1}{q}\sum_{i=1}^{q}\alpha_i$$

$$\sigma_2^2|\cdot \sim IG\left(\frac{d-q-3}{2}, \frac{1}{2}\sum_{i=q+1}^{d}(\alpha_i - \overline{\alpha}_c)^2\right), \text{ where } \overline{\alpha}_c = \frac{1}{d-q}\sum_{i=q+1}^{d}\alpha_i$$

(c) Conditional distribution for $\mathbf{f}_i$ given $\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\Phi}, \mathbf{r}_1, ..., \mathbf{r}_n$.

Let us find now the updated density for $\mathbf{f}_i$:

$$\pi(\mathbf{f}_i|\cdot) \propto \exp\left\{-\frac{1}{2}\left((\boldsymbol{\Lambda}\mathbf{f}_i - (\mathbf{r}_i - \boldsymbol{\mu}))^T \boldsymbol{\Psi}^{-1}(\boldsymbol{\Lambda}\mathbf{f}_i - (\mathbf{r}_i - \boldsymbol{\mu})) + \mathbf{f}_i^T\boldsymbol{\Phi}^{-1}\mathbf{f}_i\right)\right\}$$

Let us focus on the term in the exponential. For simplicity of formulas, let us drop the $\exp\{\cdot\}$ and the $-\frac{1}{2}$ factor:

$$\mathbf{f}_i^T\boldsymbol{\Lambda}^T\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}\mathbf{f}_i - 2\mathbf{f}_i^T\boldsymbol{\Lambda}^T\boldsymbol{\Psi}^{-1}(\mathbf{r}_i - \boldsymbol{\mu}) + (\mathbf{r}_i - \boldsymbol{\mu})^T\boldsymbol{\Psi}^{-1}(\mathbf{r}_i - \boldsymbol{\mu}) + \mathbf{f}_i^T\boldsymbol{\Phi}^{-1}\mathbf{f}_i$$

$$= \mathbf{f}_i^T(\boldsymbol{\Lambda}^T\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda} + \boldsymbol{\Phi}^{-1})\mathbf{f}_i - 2(\mathbf{r}_i - \boldsymbol{\mu})^T\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}\mathbf{f}_i + (\mathbf{r}_i - \boldsymbol{\mu})^T\boldsymbol{\Psi}^{-1}(\mathbf{r}_i - \boldsymbol{\mu})$$

We will repeatedly make use of the following **Lemma**:

Lemma 1. *Let $\mathbf{M}$ be a symmetric and invertible matrix, then the following identity holds:*

$$\mathbf{x}^T \mathbf{M} \mathbf{x} - 2\mathbf{b}^T \mathbf{x} = (\mathbf{x} - \mathbf{M}^{-1} \mathbf{b})^T \mathbf{M} (\mathbf{x} - \mathbf{M}^{-1} \mathbf{b}) - \mathbf{b}^T \mathbf{M}^{-1} \mathbf{b}$$

To update the density for $\mathbf{f}_i$, we first apply **Lemma 1** for $\mathbf{x} = \mathbf{f}_i$, $\mathbf{M} = \mathbf{\Lambda}^T \mathbf{\Psi}^{-1} \mathbf{\Lambda} + \mathbf{\Phi}^{-1}$, $\mathbf{b}^T = (\mathbf{r}_i - \boldsymbol{\mu})^T \mathbf{\Psi}^{-1} \mathbf{\Lambda}$ and we obtain that the term in the exponential for the posterior of $\mathbf{f}_i$ multiplied by $-\frac{1}{2}$ is:

$$\left[\mathbf{f}_i - \left(\mathbf{\Lambda}^T \mathbf{\Psi}^{-1} \mathbf{\Lambda} + \mathbf{\Phi}^{-1} \right)^{-1} \mathbf{\Lambda}^T \mathbf{\Psi}^{-1} (\mathbf{r}_i - \boldsymbol{\mu}) \right]^T \left(\mathbf{\Lambda}^T \mathbf{\Psi}^{-1} \mathbf{\Lambda} + \mathbf{\Phi}^{-1} \right)$$

$$\times \left[\mathbf{f}_i - \left(\mathbf{\Lambda}^T \mathbf{\Psi}^{-1} \mathbf{\Lambda} + \mathbf{\Phi}^{-1} \right)^{-1} \mathbf{\Lambda}^T \mathbf{\Psi}^{-1} (\mathbf{r}_i - \boldsymbol{\mu}) \right]$$

$$- (\mathbf{r}_i - \boldsymbol{\mu})^T \mathbf{\Psi}^{-1} \mathbf{\Lambda} \left(\mathbf{\Lambda}^T \mathbf{\Psi}^{-1} \mathbf{\Lambda} + \mathbf{\Phi}^{-1} \right)^{-1} \mathbf{\Lambda}^T \mathbf{\Psi}^{-1} (\mathbf{r}_i - \boldsymbol{\mu})$$

Here, only the first term depends on $\mathbf{f}_i$ and we actually observe that it is the kernel of a normal distribution. Therefore, we obtain that:

$$\mathbf{f}_i |\cdot \overset{indep.}{\sim} N_q \left(\left(\mathbf{\Lambda}^T \mathbf{\Psi}^{-1} \mathbf{\Lambda} + \mathbf{\Phi}^{-1} \right)^{-1} \mathbf{\Lambda}^T \mathbf{\Psi}^{-1} (\mathbf{r}_i - \boldsymbol{\mu}), \left(\mathbf{\Lambda}^T \mathbf{\Psi}^{-1} \mathbf{\Lambda} + \mathbf{\Phi}^{-1} \right)^{-1} \right)$$

(d) Conditional posterior for $\boldsymbol{\mu}$, given $\mathbf{\Lambda}, \mathbf{f}_i, \mathbf{\Phi}$

Next, we are ready to find the posterior for $\boldsymbol{\mu}$:

$$\pi(\boldsymbol{\mu}|\mathbf{\Lambda}, \mathbf{f}_i, \mathbf{\Phi}) \propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^{n} (\mathbf{r}_i - \mathbf{\Lambda} \mathbf{f}_i - \boldsymbol{\mu})^T \mathbf{\Psi}^{-1} (\mathbf{r}_i - \mathbf{\Lambda} \mathbf{f}_i - \boldsymbol{\mu}) \right\}$$

$$\times \exp \left\{ -\frac{1}{2} (\mathbf{P} \boldsymbol{\mu} - \mathbf{q}_0)^T \mathbf{\Omega}^{-1} (\mathbf{P} \boldsymbol{\mu} - \mathbf{q}_0) \right\}$$

Let $\mathbf{r}_i^* = \mathbf{r}_i - \mathbf{\Lambda} \mathbf{f}_i$, for all $i \in \{1, 2, ..., n\}$ and $\bar{\mathbf{r}}^* = \frac{1}{n} \sum_{i=1}^{n} \mathbf{r}_i^* = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{r}_i - \mathbf{\Lambda} \mathbf{f}_i)$. If we focus only on the first exponential, we can apply the typical trick of subtracting and adding $\bar{\mathbf{r}}^*$ and we obtain that the term in the first exponential is equal to:

$$\sum_{i=1}^{n} (\mathbf{r}_i^* - \boldsymbol{\mu})^T \mathbf{\Psi}^{-1} (\mathbf{r}_i^* - \boldsymbol{\mu}) = \sum_{i=1}^{n} (\mathbf{r}_i^* - \bar{\mathbf{r}}^*)^T \mathbf{\Psi}^{-1} (\mathbf{r}_i^* - \bar{\mathbf{r}}^*) + n (\bar{\mathbf{r}}^* - \boldsymbol{\mu})^T \mathbf{\Psi}^{-1} (\bar{\mathbf{r}}^* - \boldsymbol{\mu})$$

Therefore, by using **Lemma 1**, we obtain that the posterior for $\boldsymbol{\mu}$ is:

$$\pi(\boldsymbol{\mu}|\cdot) \propto \exp \left\{ -\frac{1}{2} n (\bar{\mathbf{r}}^* - \boldsymbol{\mu})^T \mathbf{\Psi}^{-1} (\bar{\mathbf{r}}^* - \boldsymbol{\mu}) \right\} \exp \left\{ -\frac{1}{2} (\mathbf{P} \boldsymbol{\mu} - \mathbf{q}_0)^T \mathbf{\Omega}^{-1} (\mathbf{P} \boldsymbol{\mu} - \mathbf{q}_0) \right\}$$

$$\propto \exp \left\{ -\frac{1}{2} \left(\boldsymbol{\mu}^T \left(n \mathbf{\Psi}^{-1} + \mathbf{P}^T \mathbf{\Omega}^{-1} \mathbf{P} \right) \boldsymbol{\mu} - 2 \left(\bar{\mathbf{r}}^{*T} n \mathbf{\Psi}^{-1} + \mathbf{q}_0^T \mathbf{\Omega}^{-1} \mathbf{P} \right) \boldsymbol{\mu} \right) \right\}$$

$$\propto \exp \left\{ -\frac{1}{2} \left(\boldsymbol{\mu} - \left(n \mathbf{\Psi}^{-1} + \mathbf{P}^T \mathbf{\Omega}^{-1} \mathbf{P} \right)^{-1} \left(n \mathbf{\Psi}^{-1} \bar{\mathbf{r}}^* + \mathbf{P}^T \mathbf{\Omega}^{-1} \mathbf{q}_0 \right) \right)^T \left(n \mathbf{\Psi}^{-1} + \mathbf{P}^T \mathbf{\Omega}^{-1} \mathbf{P} \right) \right.$$

$$\left. \times \left(\boldsymbol{\mu} - \left(n \mathbf{\Psi}^{-1} + \mathbf{P}^T \mathbf{\Omega}^{-1} \mathbf{P} \right)^{-1} \left(n \mathbf{\Psi}^{-1} \bar{\mathbf{r}}^* + \mathbf{P}^T \mathbf{\Omega}^{-1} \mathbf{q}_0 \right) \right) \right\}$$

The conditional distribution of $\boldsymbol{\mu}$ given $\mathbf{\Lambda}, \mathbf{f}_i, \mathbf{\Phi}, \mathbf{r}$ is a multivariate normal. That is

$$\boldsymbol{\mu}|\cdot \sim N_m \left(\left(n \mathbf{\Psi}^{-1} + \mathbf{P}^T \mathbf{\Omega}^{-1} \mathbf{P} \right)^{-1} \left(n \mathbf{\Psi}^{-1} \bar{\mathbf{r}}^* + \mathbf{P}^T \mathbf{\Omega}^{-1} \mathbf{q}_0 \right), \left(n \mathbf{\Psi}^{-1} + \mathbf{P}^T \mathbf{\Omega}^{-1} \mathbf{P} \right)^{-1} \right),$$

$$\text{where } \bar{\mathbf{r}}^* = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{r}_i - \mathbf{\Lambda} \mathbf{f}_i)$$

(e) Posterior for $\mathbf{\Lambda}, \mathbf{\Psi}$

By looking at the assumptions in equations (4) and by collecting the terms depending on $\mathbf{\Lambda}$ and $\mathbf{\Psi}$, we obtain:

$$\pi(\boldsymbol{\Lambda}, \boldsymbol{\Psi}|\tilde{\boldsymbol{\theta}}) \propto \prod_{j=1}^{m} \left(\Psi_j^{-\frac{n}{2}-\alpha_j-1} |\boldsymbol{\Psi}_j \mathbf{H}_j|^{-\frac{1}{2}} \right) \exp\left\{ -\sum_{j=1}^{m} \frac{\beta_j}{\Psi_j} \right\}$$

$$\times \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} (\mathbf{r}_i - \boldsymbol{\mu} - \boldsymbol{\Lambda}\mathbf{f}_i)^T \boldsymbol{\Psi}^{-1} (\mathbf{r}_i - \boldsymbol{\mu} - \boldsymbol{\Lambda}\mathbf{f}_i) \right\}$$

$$\times \exp\left\{ -\frac{1}{2} \sum_{j=1}^{m} \left(\boldsymbol{\Lambda}_j - \boldsymbol{\Lambda}_{0j} \right)^T \frac{1}{\Psi_j} \mathbf{H}_j^{-1} \left(\boldsymbol{\Lambda}_j - \boldsymbol{\Lambda}_{0j} \right) \right\} \tag{7}$$

Let us first focus our attention on the last two exponentials. We notice that one sum is over columns (the one over i), while the other sum is over the rows (the one over j). However, we can write the sum over i as a sum over j in the following way:

$$\sum_{i=1}^{n} (\mathbf{r}_i - \boldsymbol{\mu} - \boldsymbol{\Lambda}\mathbf{f}_i)^T \boldsymbol{\Psi}^{-1} (\mathbf{r}_i - \boldsymbol{\mu} - \boldsymbol{\Lambda}\mathbf{f}_i) = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\mathbf{r}_{ji} - \mu_j - \mathbf{f}_i^T \boldsymbol{\Lambda}_j \right)^2 \frac{1}{\Psi_j}$$

$$= \sum_{j=1}^{m} \left(\mathbf{r}_{j\cdot}^T - \mu_j \mathbf{1} - \mathbf{F}^T \boldsymbol{\Lambda}_j \right)^T \frac{1}{\Psi_j} \left(\mathbf{r}_{j\cdot}^T - \mu_j \mathbf{1} - \mathbf{F}^T \boldsymbol{\Lambda}_j \right)$$

Here, we introduced the following notation:

- $\mathbf{r}_{j\cdot}$ = the jth row in the matrix of returns $\mathbf{R} = \begin{bmatrix} \mathbf{r}_1 & \dots & \mathbf{r}_n \end{bmatrix}$

- $\mathbf{F} = \begin{bmatrix} \mathbf{f}_1 & \dots & \mathbf{f}_n \end{bmatrix}$ is the matrix in which we have as columns the common factors.

- μ_j is the jth entry in the vector of means $\boldsymbol{\mu}$.

- $\boldsymbol{\Lambda}_j$ is the jth row in the matrix $\boldsymbol{\Lambda}$.

Since we managed to change the summation so that it is with respect to the rows, we can now combine the last two exponentials from the joint posterior density presented above:

$$\pi\left(\boldsymbol{\Lambda}_j, \Psi_j|\cdot \right) \propto \left(\Psi_j^{-\frac{n}{2}-\alpha_j-1} |\boldsymbol{\Psi}_j \mathbf{H}_j|^{-\frac{1}{2}} \right) \exp\left\{ -\frac{\beta_j}{\Psi_j} \right\}$$

$$\times \exp\left\{ -\frac{1}{2} \left(\mathbf{r}_{j\cdot}^T - \mu_j \mathbf{1} - \mathbf{F}^T \boldsymbol{\Lambda}_j \right)^T \frac{1}{\Psi_j} \left(\mathbf{r}_{j\cdot}^T - \mu_j \mathbf{1} - \mathbf{F}^T \boldsymbol{\Lambda}_j \right) \right\}$$

$$\times \exp\left\{ -\frac{1}{2} \left(\boldsymbol{\Lambda}_j - \boldsymbol{\Lambda}_{0j} \right)^T \frac{1}{\Psi_j} \mathbf{H}_j^{-1} \left(\boldsymbol{\Lambda}_j - \boldsymbol{\Lambda}_{0j} \right) \right\} \tag{8}$$

(e1) Posterior for $\boldsymbol{\Lambda}$

Only the last two exponentials in equation (8) depend on $\boldsymbol{\Lambda}$ and for simplicity we omit the $-\frac{1}{2}$ factor:

$$\left(\mathbf{r}_{j\cdot}^T - \mu_j \mathbf{1} - \mathbf{F}^T \boldsymbol{\Lambda}_j \right)^T \frac{1}{\Psi_j} \left(\mathbf{r}_{j\cdot}^T - \mu_j \mathbf{1} - \mathbf{F}^T \boldsymbol{\Lambda}_j \right) + \left(\boldsymbol{\Lambda}_j - \boldsymbol{\Lambda}_{0j} \right)^T \frac{1}{\Psi_j} \mathbf{H}_j^{-1} \left(\boldsymbol{\Lambda}_j - \boldsymbol{\Lambda}_{0j} \right)$$

$$= \boldsymbol{\Lambda}_j^T \mathbf{F} \frac{1}{\Psi_j} \mathbf{F}^T \boldsymbol{\Lambda}_j + \boldsymbol{\Lambda}_j^T \frac{1}{\Psi_j} \mathbf{H}_j^{-1} \boldsymbol{\Lambda}_j - 2\boldsymbol{\Lambda}_j^T \mathbf{F} \frac{1}{\Psi_j} \left(\mathbf{r}_{j\cdot}^T - \mu_j \mathbf{1} \right) - 2\boldsymbol{\Lambda}_j^T \frac{1}{\Psi_j} \mathbf{H}_j^{-1} \boldsymbol{\Lambda}_{0j}$$

$$+ \left(\mathbf{r}_{j\cdot}^T - \mu_j \mathbf{1} \right)^T \frac{1}{\Psi_j} \left(\mathbf{r}_{j\cdot}^T - \mu_j \mathbf{1} \right) + \boldsymbol{\Lambda}_{0j}^T \frac{1}{\Psi_j} \mathbf{H}_j^{-1} \boldsymbol{\Lambda}_{0j}$$

$$= \boldsymbol{\Lambda}_j^T \frac{1}{\Psi_j} \left(\mathbf{F}\mathbf{F}^T + \mathbf{H}_j^{-1} \right) \boldsymbol{\Lambda}_j - 2\boldsymbol{\Lambda}_j^T \frac{1}{\Psi_j} \left(\mathbf{F} \left(\mathbf{r}_{j\cdot}^T - \mu_j \mathbf{1} \right) + \mathbf{H}_j^{-1} \boldsymbol{\Lambda}_{0j} \right) + \left(\mathbf{r}_{j\cdot}^T - \mu_j \mathbf{1} \right)^T \frac{1}{\Psi_j} \left(\mathbf{r}_{j\cdot}^T - \mu_j \mathbf{1} \right)$$

$$+ \boldsymbol{\Lambda}_{0j}^T \frac{1}{\Psi_j} \mathbf{H}_j^{-1} \boldsymbol{\Lambda}_{0j}$$

Since the only terms that depend on $\mathbf{\Lambda}_j$ are the first two, for now we can focus only on them and they will give us the posterior. We can apply again **Lemma 1** for $\mathbf{x} = \mathbf{\Lambda}_j$, $\mathbf{M} = \frac{1}{\Psi_j}(\mathbf{FF}^T + \mathbf{H}_j^{-1})$, $\mathbf{b} = \frac{1}{\Psi_j}\left(\mathbf{F}(\mathbf{r}_{j\cdot}^T - \mu_j\mathbf{1}) + \mathbf{H}_j^{-1}\mathbf{\Lambda}_{0j}\right)$ and we obtain:

$$
\left(\mathbf{\Lambda}_j - \Psi_j\left(\mathbf{FF}^T + \mathbf{H}_j^{-1}\right)^{-1}\frac{1}{\Psi_j}\left(\mathbf{F}\left(\mathbf{r}_{j\cdot}^T - \mu_j\mathbf{1}\right) + \mathbf{H}_j^{-1}\mathbf{\Lambda}_{0j}\right)\right)^T \frac{1}{\Psi_j}\left(\mathbf{FF}^T + \mathbf{H}_j^{-1}\right)
$$

$$
\times \left(\mathbf{\Lambda}_j - \Psi_j\left(\mathbf{FF}^T + \mathbf{H}_j^{-1}\right)^{-1}\frac{1}{\Psi_j}\left(\mathbf{F}\left(\mathbf{r}_{j\cdot}^T - \mu_j\mathbf{1}\right) + \mathbf{H}_j^{-1}\mathbf{\Lambda}_{0j}\right)\right)
$$

$$
-\frac{1}{\Psi_j}\left(\mathbf{F}\left(\mathbf{r}_{j\cdot}^T - \mu_j\mathbf{1}\right) + \mathbf{H}_j^{-1}\mathbf{\Lambda}_{0j}\right)^T \Psi_j\left(\mathbf{FF}^T + \mathbf{H}_j^{-1}\right)^{-1}\frac{1}{\Psi_j}\left(\mathbf{F}\left(\mathbf{r}_{j\cdot}^T - \mu_j\mathbf{1}\right) + \mathbf{H}_j^{-1}\mathbf{\Lambda}_{0j}\right)
$$

$$
+\left(\mathbf{r}_{j\cdot}^T - \mu_j\mathbf{1}\right)^T\frac{1}{\Psi_j}\left(\mathbf{r}_{j\cdot}^T - \mu_j\mathbf{1}\right) + \mathbf{\Lambda}_{0j}^T\frac{1}{\Psi_j}\mathbf{H}_j^{-1}\mathbf{\Lambda}_{0j} \tag{9}
$$

Therefore, the posterior for $\mathbf{\Lambda}_j$ is proportional to:

$$
\pi\left(\mathbf{\Lambda}_j|\mathbf{\Psi}_j, \widetilde{\boldsymbol{\theta}}\right) \propto \exp\left\{-\frac{1}{2}\left(\mathbf{\Lambda}_j - \Psi_j\left(\mathbf{FF}^T + \mathbf{H}_j^{-1}\right)^{-1}\frac{1}{\Psi_j}\left(\mathbf{F}\left(\mathbf{r}_{j\cdot}^T - \mu_j\mathbf{1}\right) + \mathbf{H}_j^{-1}\mathbf{\Lambda}_{0j}\right)\right)^T\right.
$$

$$
\left.\times\frac{1}{\Psi_j}\left(\mathbf{FF}^T + \mathbf{H}_j^{-1}\right)\left(\mathbf{\Lambda}_j - \Psi_j\left(\mathbf{FF}^T + \mathbf{H}_j^{-1}\right)^{-1}\frac{1}{\Psi_j}\left(\mathbf{F}\left(\mathbf{r}_{j\cdot}^T - \mu_j\mathbf{1}\right) + \mathbf{H}_j^{-1}\mathbf{\Lambda}_{0j}\right)\right)\right\}
$$

That is, given $\mathbf{\Psi}_j$, the conditional posterior for $\mathbf{\Lambda}_j$ is a normal with mean vector $\overline{\boldsymbol{\mu}}_j$ and covariance matrix $\Psi_j\overline{\mathbf{\Omega}}_j$, where:

$$
\overline{\mathbf{\Omega}}_j = \left(\mathbf{FF}^T + \mathbf{H}_j^{-1}\right)^{-1}
$$
$$
\overline{\boldsymbol{\mu}}_j = \overline{\mathbf{\Omega}}_j\left[\mathbf{F}\left(\mathbf{r}_{j\cdot}^T - \mu_j\mathbf{1}\right) + \mathbf{H}_j^{-1}\mathbf{\Lambda}_{0j}\right] \tag{10}
$$

(e2) Posterior for $\mathbf{\Psi}$

If we go back to equation (8) and collect the terms that depend on $\mathbf{\Psi}$, we obtain that the posterior is:

$$
\pi(\Psi_j|\cdot) \propto \Psi_j^{-\frac{n}{2}-\frac{q}{2}-\alpha_j-1}\exp\left\{-\frac{1}{\Psi_j}\left(\beta_j + \frac{1}{2}\left(\mathbf{r}_{j\cdot}^T - \mu_j\mathbf{1} - \mathbf{F}^T\mathbf{\Lambda}_j\right)^T\left(\mathbf{r}_{j\cdot}^T - \mu_j\mathbf{1} - \mathbf{F}^T\mathbf{\Lambda}_j\right)\right)\right\}\cdot
$$

$$
\cdot\exp\left\{-\frac{1}{2}\frac{1}{\Psi_j}\left(\mathbf{\Lambda}_j - \mathbf{\Lambda}_{0j}\right)^T\mathbf{H}_j^{-1}\left(\mathbf{\Lambda}_j - \mathbf{\Lambda}_{0j}\right)\right\}
$$

Therefore, we obtain that the posterior for Ψ_j is:

$$
\Psi_j|\cdot \overset{indep.}{\sim} IG\left(\alpha_{\Psi_j}, \beta_{\Psi_j}\right)
$$
$$
\text{where } \alpha_{\Psi_j} = \frac{n}{2} + \frac{q}{2} + \alpha_j
$$
$$
\text{and } \beta_{\Psi_j} = \beta_j + \frac{1}{2}\left(\left(\mathbf{r}_{j\cdot}^T - \mu_j\mathbf{1} - \mathbf{F}^T\mathbf{\Lambda}_j\right)^T\left(\mathbf{r}_{j\cdot}^T - \mu_j\mathbf{1} - \mathbf{F}^T\mathbf{\Lambda}_j\right) + \left(\mathbf{\Lambda}_j - \mathbf{\Lambda}_{0j}\right)^T\mathbf{H}_j^{-1}\left(\mathbf{\Lambda}_j - \mathbf{\Lambda}_{0j}\right)\right)
$$

2.4 Implementation

Now that we have derived our posteriors, we are ready to implement them using a Gibbs Sampler. We will use a Metropolis-Hastings algorithm for sampling α, for which we need both the exact posterior distribution and the approximation obtained with the Volterra integral equation. It is an approach introduced by (Leonard & Hsu, 1992) and (Albert et al., 2000) and used in (Andrei & Hsu, 2020). Given σ_1^2, σ_2^2 and μ, the posterior for α can be approximated by a normal density with mean vector α^* and covariance matrix $(\mathbf{Q} + \mathbf{G})^{-1}$, where α^*, $\mathbf{G}$ and $\mathbf{Q}$ are defined in (5) and (6), respectively. The exact conditional posterior given σ_1^2, σ_2^2 and μ is

$$
\pi(\alpha|\cdot) \propto \exp\left\{-\frac{n}{2}Tr\left(\mathbf{A} + \mathbf{S}e^{-\mathbf{A}}\right) - \frac{1}{2}\alpha^T\mathbf{G}\alpha\right\}
$$

$$
\text{where } \mathbf{S} = \frac{1}{n}\left(\mathbf{r}_i - \mu\right)\left(\mathbf{r}_i - \mu\right)^T
$$

The Metropolis-Hastings step at i^{th} iteration would be that we would simulate a candidate value from the approximate posterior distribution: $\widetilde{\alpha} \approx \sim N(\alpha^*, (\mathbf{Q} + \mathbf{G})^{-1})$ and we would accept it with probability $min(\rho, 1)$, where

$$\rho = \frac{\pi(\widetilde{\alpha}|\cdot)}{\pi(\alpha^{(t)}|\cdot)} \cdot \frac{\pi^*\left(\alpha^{(t)}|\cdot\right)}{\pi^*(\widetilde{\alpha}|\cdot)}$$

It is useful at this point to remember that there is a connection between π^* and π, since there is one between $\mathbf{A}$ and α, namely:

$$\alpha = Vec^*(\log(\mathbf{\Phi})), \mathbf{A} = \log(\mathbf{\Phi})$$

Albeit the Gibbs Sampler converges to the same distribution no matter the starting points, we should try to initialize it with good estimates. Also, we have to make sure that we specify the hyper-parameters with values that would make sense in the real world:

- n = number of returns in the historical dataset= number of returns from 1/2/2014 to 12/29/2017.

- $\hat{\mathbf{F}}_{init} = \left[\hat{\mathbf{f}}_1\hat{\mathbf{f}}_2...\hat{\mathbf{f}}_n\right]$, where $\hat{\mathbf{f}}_i$ for $i \in \{1, 2, ..., n\}$ are the common factors obtained by fitting a factor model on the historical dataset with an optimal number of factors of $q = 18$ determined from a scree plot of eigenvalues.

- We also have that $\mathbf{f}_i|\mathbf{\Phi} \overset{indep.}{\sim} N_q(0, \mathbf{\Phi})$. In order to specify $\hat{\mathbf{\Phi}}_{init}$, we take the covariance of the above found common factors: $\hat{\mathbf{\Phi}}_{init} = Cov(\hat{\mathbf{f}}_i)$.

- We also have the following assumption.

$$\alpha|\theta, \Delta = Vec^*(log(\mathbf{\Phi}))|\theta, \Delta \sim N_d\left(\begin{bmatrix}\theta_1\mathbf{1}_q \\ \theta_2\mathbf{1}_{d-q}\end{bmatrix}, \begin{bmatrix}\sigma_1^2\mathbf{I}_q & \mathbf{0} \\ \mathbf{0} & \sigma_2^2\mathbf{I}_{d-q}\end{bmatrix}\right)$$

  - In order to initialize σ_1^2, we have to take the variance of the first q entries in $Vec^*(\log(\hat{\mathbf{\Phi}}_{init}))$.

  - In order to initialize σ_2^2, we have to take the variance of the last $d - q$ entries in $Vec^*(\log(\hat{\mathbf{\Phi}}_{init}))$.

- We remember that $\mathbf{\Lambda}_j|\Psi_j \overset{indep.}{\sim} N_q(\mathbf{\Lambda}_{0j}, \Psi_j\mathbf{H}_j)$. Since, in general, we do not have any prior information on the factor weights, we specify the hyper-parameters to be:

  - $\hat{\mathbf{\Lambda}}_{0jinit} = 0$

  - We initialize the variance $\Psi_j\mathbf{H}_j$ with a big value: $\hat{\Psi}_{jinit} = 1$ and $\hat{\mathbf{H}}_{jinit} = 10^{10}\mathbf{I}_q$.

- Also, we remember that $\Psi_j \overset{indep.}{\sim} IG(\alpha_j, \beta_j)$, for all $j \in \{1, 2, ..., m\}$. Similarly to the previous point made, in the real world, we do not have any prior information on Ψ_j and this should be reflected in our choice of α_j and β_j. If we let $\alpha_j \to 0$ and $\beta_j \to 0$ in the pdf of the $IG(\alpha_j, \beta_j)$, we notice that we obtain an uninformative prior. Therefore, we initialize $\hat{\alpha}_{jinit} = \hat{\beta}_{jinit} = 10^{-10}$.

Using the Metropolis Hastings step that was just discussed, we arrive at the following Gibbs Sampler:

Algorithm 1 Gibbs Sampler $\log(\mathbf{\Phi})$

1:   $\alpha^{(t+1)} = \begin{cases} \widetilde{\alpha} \sim N\left(\left(\mathbf{Q}^{(t)} + \mathbf{G}^{(t)}\right)^{-1} \mathbf{Q}^{(t)} \lambda^{(t)}, \left(\mathbf{Q}^{(t)} + \mathbf{G}^{(t)}\right)^{-1}\right) \text{w.p. } min(\rho, 1) \\ \alpha^{(t)} \text{otherwise} \end{cases}$

2:   Since $\alpha = Vec^*\left(\log(\mathbf{\Phi})\right) \Rightarrow \begin{cases} \text{compute } \mathbf{\Phi}^{(t+1)} = \exp\left\{Vec^{*-1}\left(\alpha^{(t+1)}\right)\right\} \\ \text{keep } \mathbf{\Phi}^{(t)} \end{cases}$

3:   $\begin{cases} \sigma_1^{2(t+1)} \sim IG\left(\frac{q-3}{2}, \frac{1}{2}\sum_{i=1}^{q}\left(\alpha_i^{(t+1)} - \overline{\alpha_v}^{(t+1)}\right)^2\right) \\ \sigma_2^{2(t+1)} \sim IG\left(\frac{d-q-3}{2}, \frac{1}{2}\sum_{i=q+1}^{d}\left(\alpha_i^{(t+1)} - \overline{\alpha_c}^{(t+1)}\right)^2\right) \end{cases} \Rightarrow$

$\Rightarrow \mathbf{\Delta}^{(t+1)} = \begin{bmatrix} \sigma_1^{2(t+1)}\mathbf{I}_q & \mathbf{0} \\ \mathbf{0} & \sigma_2^{2(t+1)}\mathbf{I}_{d-q} \end{bmatrix}$

4:   Let $\Sigma_{\mu} = \left(n\mathbf{\Psi}^{(t)-1} + \mathbf{P}^T\mathbf{\Omega}^{-1}\mathbf{P}\right)^{-1} \Rightarrow \mu^{(t+1)} \sim N\left(\Sigma_{\mu}\left(n\mathbf{\Psi}^{(t)-1}\overline{\mathbf{r}}^{*(t)} + \mathbf{P}^T\mathbf{\Omega}^{-1}\mathbf{q}_0\right), \Sigma_{\mu}\right)$, where $\overline{\mathbf{r}}^{*(t)} = \frac{\sum_{i=1}^{n}\mathbf{r}_i - \mathbf{\Lambda}^{(t)}\mathbf{f}_i^{(t)}}{n}$.

5:   Let

$$\Sigma_{\mathbf{f}} = \left(\mathbf{\Lambda}^{(t)T}\mathbf{\Psi}^{(t)-1}\mathbf{\Lambda}^{(t)} + \mathbf{\Phi}^{(t+1)-1}\right)^{-1} \Rightarrow \mathbf{f}_i^{(t+1)} \sim N\left(\Sigma_{\mathbf{f}}\mathbf{\Lambda}^{(t)T}\mathbf{\Psi}^{(t)-1}\left(\mathbf{r}_i - \mu^{(t+1)}\right), \Sigma_{\mathbf{f}}\right)$$

6:   $\mathbf{\Psi}_j|\cdot \overset{indep.}{\sim} IG\left(\alpha_{\Psi_j}, \beta_{\Psi_j}\right)$, where $\alpha_{\Psi_j} = \frac{n}{2} + \frac{q}{2} + \alpha_j$

$$\text{and } \beta_{\Psi_j} = \beta_j + \frac{1}{2}\left(\left(\mathbf{r}_{j\cdot}^T - \mu_j\mathbf{1} - \mathbf{F}^T\mathbf{\Lambda}_j\right)^T\left(\mathbf{r}_{j\cdot}^T - \mu_j\mathbf{1} - \mathbf{F}^T\mathbf{\Lambda}_j\right) + \left(\mathbf{\Lambda}_j - \mathbf{\Lambda}_{0j}\right)^T \mathbf{H}_j^{-1}\left(\mathbf{\Lambda}_j - \mathbf{\Lambda}_{0j}\right)\right)$$

7:   $\mathbf{\Lambda}_j^{(t+1)} \overset{indep.}{\sim} N\left(\overline{\mu}_j^{(t+1)}, \mathbf{\Psi}_j^{(t+1)}\overline{\mathbf{\Omega}}_j^{(t+1)}\right)$

8:   Compute $\mathbf{S}_{\mathbf{f}}^{(t+1)} = \frac{\sum_{i=1}^{n}\mathbf{f}_i\mathbf{f}_i^T}{n}$, $\lambda^{(t+1)} = Vec^*\left(\log\left(\mathbf{S}_{\mathbf{f}}^{(t+1)}\right)\right)$, $d_j^{(t+1)}$ and $\mathbf{e}_j^{(t+1)}$ the eigenvalue and normalized eigenvector of $\mathbf{S}_{\mathbf{f}}^{(t+1)}$ respectively.

9:   Compute $\mathbf{f}_{ij}^{(t+1)}$ by identifying the coefficients of the entries of the $log(\mathbf{\Phi})$ matrix from the equation $Vec^*\left(\log\left(\mathbf{\Phi}^{(t+1)}\right)\right)^T\mathbf{f}_{ij}^{(t+1)} = \mathbf{e}_i^{(t+1)T}log\left(\mathbf{\Phi}^{(t+1)}\right)\mathbf{e}_j^{(t+1)}$

10:   Compute $\xi_{ij}^{(t+1)} = \frac{(d_i^{(t+1)} - d_j^{(t+1)})^2}{d_i^{(t+1)}d_j^{(t+1)}\left(\log\left(d_i^{(t+1)}\right) - log\left(d_j^{(t+1)}\right)\right)^2}$

11:   Compute $\mathbf{Q}^{(t+1)} = \frac{n}{2}\sum_{i=1}^{q}\mathbf{f}_{ii}^{(t+1)}\mathbf{f}_{ii}^{(t+1)T} + T\sum_{i<j}^{q}\xi_{ij}^{(t+1)}\mathbf{f}_{ij}^{(t+1)}\mathbf{f}_{ij}^{(t+1)T}$

12:   Compute

$$\mathbf{G}^{(t+1)} = \left(\mathbf{I}_d - \mathbf{J}\left(\mathbf{J}^T\mathbf{\Delta}^{(t+1)-1}\mathbf{J}\right)^{-1}\mathbf{J}^T\mathbf{\Delta}^{(t+1)-1}\right)^T \mathbf{\Delta}^{(t+1)-1}$$

$$\times\left(\mathbf{I}_d - J\left(\mathbf{J}^T\mathbf{\Delta}^{(t+1)-1}\mathbf{J}\right)^{-1}\mathbf{J}^T\mathbf{\Delta}^{(t+1)-1}\right)$$

2.5 Results-Personal Views on 4 Stocks

We would like to conduct sensitivity analysis for the confidence in the investor views (the diagonal elements of the $\mathbf{\Omega}$ matrix denoted by ω_i). We would expect the model to behave in a similar manner as before:

- The more confident the investor is in the inputted views, the closer the model should follow them
- The less confident the investor is in the inputted views, the closer the model should follow history

We choose to have views for the following 4 stocks: AAPL, FB, GOOG, MSFT and we will consider views on industry sectors in the next section. We will use the daily returns from 1/2/2014 to 12/29/2017. We create the following investor inputs (again the columns are in order AAPL, FB, GOOG, MSFT and the rows represent the views). Please notice that the matrix $\mathbf{P}$ in our implementation has a lot more columns (one for each stock actively traded in $S\&P500$), but the vast majority of the entries are 0:

$$\mathbf{q}_0 = \begin{bmatrix} 0.02 \\ 0.05 \end{bmatrix}, \mathbf{P} = \begin{array}{c c c c c} & \text{AAPL} & \text{FB} & \text{GOOG} & \text{MSFT} \\ \text{view1} & -1 & 1 & 0 & 0 \\ \text{view2} & 0 & 0 & 1 & -1 \end{array}$$

Albeit the memory allocation problem which originally motivated the incorporation of the Bayesian factor model was solved, the version presented in this paper is still computationally expensive since we have to sample from many distributions. Therefore, an exhaustive search was ran in parallel on multiple cores (each core running the Gibbs Sampler for 1 pair (ω_1, ω_2), the grid was split into 16 evenly split ranges, each one running 6 simulations). The burn period is 10^3 and the iteration period is 10^4.

In the following plot, 2 of the axis are represented by the two confidence levels (ω_1 and ω_2) and the third one is represented by the l^2 distance $\|\mathbf{P}\mu_{\text{post}} - \mathbf{q}_0\|_2$. As mentioned previously, this distance should go to 0 as ω_1 and ω_2 go to 0, which can easily be observed in Figure 1.
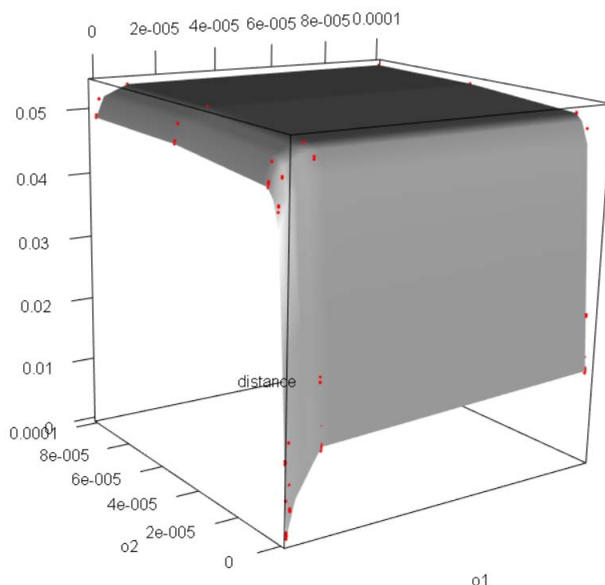
Figure 1. Distance when considering only 4 stocks

Furthermore, similar to the versions introduced previously, as ω_1 and ω_2 increase, the distance converges to the same number. Since ω_1 and ω_2 are standard deviations, a high standard deviation represents a lack of confidence in the personal views inputted. Therefore, intuitively, the model should only take into consideration the history. This is precisely how the model behaves. If we just consider the historical returns, the unbiased estimator for μ is the sample mean of the returns, $\bar{\mathbf{r}}$. The distance $\|\mathbf{P}\bar{\mathbf{r}} - \mathbf{q}_0\|_2 = 0.05386381$, which is the level at which the curve in Figure 1 flattens.

One could use this model to hold a portfolio with an initial starting capital of \$100,000 over a testing data set consisting of the daily returns during the month of January 2018. Please keep in mind that, in order to obtain portfolio weights, we estimate from the Gibbs Sampler $\Sigma_{post} = \Lambda_{post}\Phi_{post}\Lambda_{post}{}^T + \Psi_{post}$ and we try to maximize the portfolio returns, while minimizing the portfolio risk. Hence, we would like to find $\max_{\mathbf{w}} \mathbf{w}^T\mu_{post} - \frac{\lambda}{2}\mathbf{w}^T\Sigma_{post}\mathbf{w}$, where λ is the investor's risk aversion coefficient. In his paper, **?** suggests that $\lambda = 2.5$ is a reasonable choice for equities. By making the derivative with respect to $\mathbf{w}$ equal to 0, and by solving the resulting equation for $\mathbf{w}$, we obtain: $\mathbf{w}^* = \frac{1}{2.5}\Sigma_{post}^{-1}\mu_{post}$. The profits without considering any fees on a testing data set consisting of the returns over the month of January 2018 for all the previously mentioned combinations of confidence levels (ω_1 and ω_2) averaged \$40,075.87 with a standard deviation of \$14,373.88

2.6 Results-Personal Views on Industry Sectors

In this section, we will present similar results to the ones presented in the previous section. We will have the exact same training and testing data sets as before. The only change is in the personal views inputted in our model. However, this time we would like to enter personal views about different industry sectors. The stocks in the $S\&P500$ are divided into 11 industry sectors. These industry sectors are described in Appendix 2.6.

In order to have good personal views and not just random guesses, as we have done so far, we will use the weighting recommendations provided by CFRA [1], an independent fundamental and forensic investment research firm. Each stock within the same sector receives equal weight that sum up to 1, with a positive weight for the ones outperforming and a negative weight for the ones under-performing. We will have the following 4 personal views [2]:

(1) *Information technology* outperforms *utilities* by 0.2% with confidence level ω_1.

(2) *Energy* outperforms *industrials* by 0.1% with confidence level ω_2.

(3) *Real Estate* outperforms *consumer staples* by 0.2% with confidence level ω_2.

(4) *Consumer discretionary* outperforms *financials* by 0.3% with confidence level ω_2.

---

[1]CFRA: https://eresearch.fidelity.com/eresearch/markets_sectors/sectors/si_weighting_recommendations.jhtml?tab=sirecommendationsFidelity Investments link. The companies in $S\&P500$ were divided into 11 industry sectors based on economic characteristics (please see **2.6. Appendix**)

[2]For details on which companies are in each industry sector, please see https://www.barchart.com/stocks/indices/sp-sector/industrialsbarchart.com

Hence, we have that $\mathbf{q}_0 = (0.002, 0.001, 0.002, 0.003)^T$.

We picked 2 distinct confidence levels for the 4 views simply because we wanted to have another 3D plot with 2 of the axis represented by ω_1 and ω_2 and the third axis represented by the distance $\|\mathbf{P}\mu_{post} - \mathbf{q}_0\|_2$. Just as before, the exhaustive search was ran in parallel on multiple cores. The same burning and iteration periods were used also.
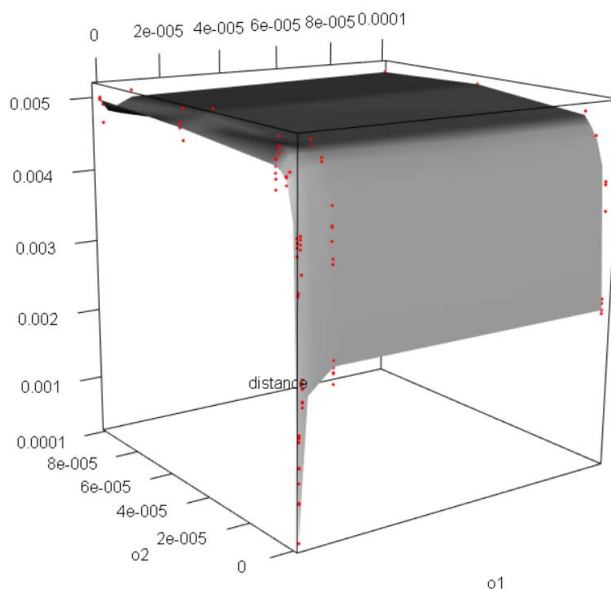

Figure 2. Distance when considering industry sectors

Again, the model behaves exactly as our intuition suggests. As ω_1 and ω_2 go to 0, the distance $\|\mathbf{P}\mu_{post} - \mathbf{q}_0\|_2$ converges to 0. Moreover, for bigger values of ω_1 and ω_2 (small confidence in views) the distance converges to $0.004999748 = \|\mathbf{P}\overline{\mathbf{r}} - \mathbf{q}_0\|_2$. This confirms our intuition that the less confident the investor is in his or her views, the more the model takes into consideration the history.

Moving on to presenting the profits, we used the same starting capital of \$100,000, the same testing data set over the month of January 2018 and the same methodology for computing the portfolio weights. The mean of the profits over all the simulated pairs (ω_1, ω_2) was \$37,576.68 with a standard deviation of \$5,857.198.

Acknowledgement

References

Albert, J. H., Leonard, T., & Hsu, J. S. J. (2000, Jun). Bayesian methods: An analysis for statisticians and interdisciplinary researchers. *Journal of the American Statistical Association, 95*(450), 679. https://doi.org/10.2307/2669420

Andrei, M. S., & Hsu, J. S. J. (2020, May). A bayesian approach for asset allocation. *International Journal of Statistics and Probability, 9*(4), 1. https://doi.org/10.5539/ijsp.v9n4p1

Bellman, R. (1997). Introduction to matrix analysis, second edition. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611971170

He, G., & Litterman, R. (2002). The intuition behind blacklitterman model portfolios. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.334304

Ledoit, O., & Wolf, M. (2004, feb). A wellconditioned estimator for largedimensional covariance matrices. *Journal of Multivariate Analysis, 88*(2), 365-411. https://doi.org/10.1016/S0047-259X(03)00096-4

Lee, S. Y., Poon, W. Y., & Song, X. Y. (2007, Jun). Bayesian analysis of the factor model with finance applications. *Quantitative Finance, 7*(3), 343-356. https://doi.org/10.1080/14697680601009838

Lee, S. Y., & Shi, J. Q. (2000, Dec). Joint bayesian analysis of factor scores and structural parameters in the factor analysis

model. *Annals of the Institute of Statistical Mathematics, 52*(4), 722-736. https://doi.org/10.1023/A:1017529427433

Leonard, T., & Hsu, J. S. J. (1992, Dec). Bayesian inference for a covariance matrix. *The Annals of Statistics, 20*(4), 1669-1696. https://doi.org/10.1214/aos/1176348885

Appendix

*S&P*500 **Industry Sectors**

The stocks in the *S&P*500 are divided into broad groupings based on economic characteristics. Currently there are 11 industry sectors[3]

- Communication Services: from telephone access to high-speed internet, this sector of the economy keeps us all connected.

- Consumer Discretionary: businesses that have demand that rises and falls based on general economic conditions such as washers and dryers, sporting goods, new cars, and diamond engagement rings

- Consumer Staples: businesses that sell the necessities of life, ranging from bleach and laundry detergent to toothpaste and packaged food.

- Energy: businesses that source, drill, extract, and refine the raw commodities we need to keep the country going, such as oil and gas.

- Financials: banks, insurance companies, real estate investment trusts, credit card issuers, and a host of other money-centric enterprises that keep the debits and credits of the economy flowing.

- Health Care: drug companies, medical supply companies, and other scientific-based operations that are concerned with improving and healing human life.

- Industrials: from railroads and airlines to military weapons and industrial conglomerates.

- Information Technology: hardware, software, computer equipment, and IT services operations.

- Materials sector manufacturers, logs, and mines everything from precious metals, paper, and chemicals to shipping containers, wood pulp, and industrial ore.

- Real Estate: all Real Estate Investment Trusts (REITs) with the exception of Mortgage REITs, which is housed under the financial sector. The sector also includes companies that manage and develop properties.

- Utilities sector is home to the firms that make our lights work when we flip the switch, let our stoves erupt in flame when we want to cook food, make water come out of the tap when we are thirsty, and more.

Copyrights

---

[3] According to https://www.thebalance.com/what-are-the-sectors-and-industries-of-the-sandp-500-3957507thebalance.com