A Bayesian Approach for Large Asset Allocation

Mihnea S. Andrei¹ & John S. J. Hsu²

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

Received: October 10, 2020 Accepted: November 6, 2020 Online Published: December 2, 2020
doi:10.5539/ijsp.v10n1p58 URL: https://doi.org/10.5539/ijsp.v10n1p58

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 \(r = (r_1, r_2, \ldots, r_m)^T\) follow a multivariate normal distribution with mean \(\mu\) and covariance matrix \(\Sigma\). That is

\[
\begin{align*}
\mathbf{r} & \sim N_m(\mu, \Sigma) \\
\mu & \sim N_m(\pi, \tau\Sigma)
\end{align*}
\]

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

\[
\begin{align*}
\mathbf{P}\mu & \sim N_q(q_0, \Omega)
\end{align*}
\]

where \(\pi\) is the equilibrium risk parameter and \(\tau\) 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:

\[
\begin{align*}
\mathbf{P}\mu & \sim N_k(q_0, \Omega)
\end{align*}
\]

where the matrix \(\mathbf{P}\) is a \(k \times m\) matrix, \(q_0\) is a \(k \times 1\) vector, and \(\Omega\) is a \(k \times k\) matrix, usually diagonal. Each row in \(\mathbf{P}\) and \(q_0\) 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 \(r_1, r_2, \ldots, r_n\) are independent, where each \(r_i\) is a vector containing the returns of \(m\) assets, which are normally distributed as in equation (1) for \(i \in \{1, 2, \ldots, n\}\). The priors for \(\mu\) and \(\Sigma\) are assumed to be independent, where the prior on \(\mu\) 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 \(\mathbf{A} = \log(\Sigma)\). 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:

\[
\mathbf{r}_i = \mathbf{\mu} + \mathbf{A}_i + \mathbf{e}_i
\]

Where we introduced the following notation:

- \( \mathbf{\mu} \) is a vector of expected returns for each of the \( m \) assets
- \( \mathbf{A} \) 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:

\[
\mathbf{e}_i | \Psi \sim \mathcal{N}_m(0, \Psi) \text{ for all } i \in \{1, 2, ..., n\} \text{ and } \Psi = \text{diag}(\Psi_1, ..., \Psi_m)
\]

(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:

\[
\mathbf{r}_i | \mathbf{\mu}, \mathbf{f}_i, \mathbf{A}, \Psi \sim \mathcal{N}_m(\mathbf{\mu} + \mathbf{A}_i, \Psi) \text{ for all } i \in \{1, 2, ..., n\}
\]

where \( \mathbf{f}_i | \Phi \sim \mathcal{N}_q(0, \Phi) \) for all \( i \in \{1, 2, ..., n\} \)

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

\[
\mathbf{\Lambda}_j | \Psi_j \sim \mathcal{N}_q(\mathbf{0}_j, \Psi_j \mathbf{H}_j)
\]

\[
\Psi_j \sim \mathcal{IG}(\alpha_j, \beta_j) \text{ for all } j \in \{1, 2, ..., m\}
\]

Here, \( \mathbf{\Lambda}_j^T \) is the \( j \)-th row in \( \mathbf{\Lambda} \).

(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:

\[
\mathbf{P}_i \mathbf{\mu} \sim \mathcal{N}_q(\mathbf{0}, \Omega)
\]

(4) Following (Leonard & Hsu, 1992), we consider a multivariate normal prior for the logarithm of the covariance \( \mathbf{\Lambda} = \log(\Phi) = \{a_{ij}\}_{i,j=1,2,...,q} \). Let \( \mathbf{a} = [a_1, ..., a_q, a_q+1, ..., a_d] = \text{Vec}^T(\mathbf{A}) = [a_{11}, a_{22}, ..., a_{qq}, a_{12}, a_{23}, ..., a_{q-1,q}, ..., a_{1d}, ..., a_{d-1,d}] \), 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 \( \theta_1 \) and variance \( \sigma_1^2 \) and the covariance related components follow independent normal distributions, each with mean \( \theta_2 \) and variance \( \sigma_2^2 \). That is,

\[
\alpha | \theta, \mathbf{\Lambda} \sim \mathcal{N}_d(\mathbf{J} \theta, \mathbf{\Lambda})
\]

where \( \mathbf{J} = \begin{bmatrix} \mathbf{1}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{d-q} \end{bmatrix}, \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \) and \( \mathbf{\Lambda} = \begin{bmatrix} \sigma_1^2 \mathbf{I}_q & \mathbf{0} \\ \mathbf{0} & \sigma_2^2 \mathbf{I}_{d-q} \end{bmatrix} \).

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:

\[ \begin{align*}
    r_i | \mu, f_i, \Lambda, \Psi & \sim N_m(\mu + \Lambda f_i, \Psi), \text{ for all } i \in \{1, 2, ..., n\} \\
    f_i | \Phi & \sim N_q(0, \Phi), \text{ for all } i \in \{1, 2, ..., n\} \\
    P \mu & \sim N_k(q_0, \Omega) \\
    \alpha | \theta, \Lambda & = \text{Vec}^*(\log(\Phi)) | \theta, \Lambda \sim N(J \theta, \Lambda) \\
    \pi(\theta_1, \theta_2) & \propto 1 \\
    \pi(\sigma_\alpha^2, \sigma_\beta^2) & \propto 1 \\
    \Lambda_j | \Psi_j & \sim N_q(\Lambda_{0j}, \Psi_j H_j) \\
    \Psi_j & \sim IG(\alpha_j, \beta_j), \text{ for all } j \in \{1, 2, ..., m\}
\end{align*} \]

2.3 Posterior Distributions

(a) Approximate conditional posterior for \( \alpha \) given \( \mu, \sigma_\alpha^2, \sigma_\beta^2, r_1, ..., r_n \).

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

\[ \alpha^*_j = N(\alpha^*, (Q + G)^{-1}), \text{ where } \alpha^* = (Q + G)^{-1}Q d, \]

\[ G = (I_d - J(J^T \Lambda^{-1} J)^{-1}J^T \Lambda^{-1}) (I_d - J(J^T \Lambda^{-1} J)^{-1}J^T \Lambda^{-1})^T \]

and the matrix \( Q \) is computed in the following way. If we let \( e_i, d_i \) be the \( i \)th normalized eigenvector with its corresponding eigenvalue of \( S = \frac{1}{n} (r_i - \mu)(r_i - \mu)^T \), respectively, then \( h_j \) is obtained by looking at the equation \( \text{Vec}^*(\log(\Phi))^T h_{ij} = e_i^T \log(\Phi)e_j \) and identifying the coefficients of the entries in the log(\( \Phi \)) matrix. With those \( h_{ij} \), we can finally compute \( Q \):

\[ Q = \frac{n}{2} \sum_{i=1}^{q} h_i h_i^T + n \sum_{i,j} \xi_{ij} h_i h_j^T, \text{ where } \]

\[ \xi_{ij} = \frac{(d_i - d_j)^2}{d_j (\log(d_i) - \log(d_j))^2} \]

(b) Conditional posteriors for \( \sigma_\alpha^2 \) and \( \sigma_\beta^2 \) given \( \alpha \)

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

\[ \sigma_\alpha^2 | \alpha - \bar{\alpha}, \sigma_\alpha^2 \sim IG \left( \frac{q - 3}{2}, \frac{1}{2} \sum_{i=1}^{q} (\alpha_i - \bar{\alpha})^2 \right), \text{ where } \bar{\alpha} = \frac{1}{q} \sum_{i=1}^{q} \alpha_i \]

\[ \sigma_\beta^2 | \alpha - \bar{\alpha}, \sigma_\beta^2 \sim IG \left( \frac{d - q - 3}{2}, \frac{1}{2} \sum_{i=q+1}^{d} (\alpha_i - \bar{\alpha})^2 \right), \text{ where } \bar{\alpha} = \frac{1}{d-q} \sum_{i=q+1}^{d} \alpha_i \]

(c) Conditional distribution for \( f_i \) given \( \mu, \Lambda, \Psi, \Phi, r_1, ..., r_n \).

Let us find now the updated density for \( f_i \):

\[ \pi(f_i | \cdot) \propto \exp \left\{ -\frac{1}{2} \left( (\Lambda f_i - (r_i - \mu))^T \Psi^{-1} (\Lambda f_i - (r_i - \mu)) + f_i^T \Phi^{-1} f_i \right) \right\} \]

Let us focus on the term in the exponential. For simplicity of formulas, let us drop the exp [·] and the \(-\frac{1}{2}\) factor:

\[ f_i^T \Lambda^T \Psi^{-1} \Lambda f_i - 2 f_i^T \Lambda^T \Psi^{-1} (r_i - \mu) + (r_i - \mu)^T \Psi^{-1} (r_i - \mu) + f_i^T \Phi^{-1} f_i \]

\[ = f_i^T (\Lambda^T \Psi^{-1} \Lambda + \Phi^{-1}) f_i - 2 (r_i - \mu)^T \Psi^{-1} \Lambda f_i + (r_i - \mu)^T \Psi^{-1} (r_i - \mu) \]

We will repeatedly make use of the following Lemma:
Lemma 1. Let $M$ be a symmetric and invertible matrix, then the following identity holds:

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

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

$$\left[ f_i - \left( \Lambda^T \Psi^{-1} A + \Phi^{-1} \right)^{-1} \Lambda^T \Psi^{-1} \left( r_i - \mu \right) \right]^T \left( \Lambda^T \Psi^{-1} A + \Phi^{-1} \right) \times \left[ f_i - \left( \Lambda^T \Psi^{-1} A + \Phi^{-1} \right)^{-1} \Lambda^T \Psi^{-1} \left( r_i - \mu \right) \right]$$

$$- \left( r_i - \mu \right)^T \Psi^{-1} A \left( \Lambda^T \Psi^{-1} A + \Phi^{-1} \right)^{-1} \Lambda^T \Psi^{-1} \left( r_i - \mu \right)$$

Here, only the first term depends on $f_i$ and we actually observe that it is the kernel of a normal distribution. Therefore, we obtain that:

$$f_i \sim_{	ext{ind.}} N_0 \left( \left( \Lambda^T \Psi^{-1} A + \Phi^{-1} \right)^{-1} \Lambda^T \Psi^{-1} \left( r_i - \mu \right), \left( \Lambda^T \Psi^{-1} A + \Phi^{-1} \right)^{-1} \right)$$

(d) Conditional posterior for $\mu$, given $\Lambda, f, \Phi$

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

$$\pi (\mu | \Lambda, f, \Phi) \propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^{n} (r_i - \Lambda f_i - \mu)^T \Psi^{-1} \left( r_i - \Lambda f_i - \mu \right) \right\} \times \exp \left\{ -\frac{1}{2} (P \mu - q_0)^T \Omega^{-1} (P \mu - q_0) \right\}$$

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

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

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

$$\pi (\mu | \cdot) \propto \exp \left\{ -\frac{1}{2} \left( \bar{r}^* - \mu \right)^T \Psi^{-1} \left( \bar{r}^* - \mu \right) \right\} \exp \left\{ -\frac{1}{2} (P \mu - q_0)^T \Omega^{-1} (P \mu - q_0) \right\}$$

$$\propto \exp \left\{ -\frac{1}{2} \left( \mu^T \left( n \Psi^{-1} + P^T \Omega^{-1} P \right) \mu - 2 \left( \bar{r}^*^T n \Psi^{-1} + P^T \bar{r}^* \Omega^{-1} q_0 \right) \right) \right\}$$

$$\propto \exp \left\{ -\frac{1}{2} \left( \mu - \left( n \Psi^{-1} + P^T \Omega^{-1} P \right)^{-1} \left( n \Psi^{-1} \bar{r}^* + P^T \Omega^{-1} q_0 \right) \right)^T \left( n \Psi^{-1} + P^T \Omega^{-1} P \right) \mu \right\}$$

$$\times \left\{ \mu - \left( n \Psi^{-1} + P^T \Omega^{-1} P \right)^{-1} \left( n \Psi^{-1} \bar{r}^* + P^T \Omega^{-1} q_0 \right) \right\}$$

The conditional distribution of $\mu$ given $\Lambda, f, \Phi, r$ is a multivariate normal. That is

$$\mu | \cdot \sim N_m \left( \left( n \Psi^{-1} + P^T \Omega^{-1} P \right)^{-1} \left( n \Psi^{-1} \bar{r}^* + P^T \Omega^{-1} q_0 \right), \left( n \Psi^{-1} + P^T \Omega^{-1} P \right)^{-1} \right)$$

where $\bar{r}^* = \frac{1}{n} \sum_{i=1}^{n} (r_i - \Lambda f_i)$

(e) Posterior for $\Lambda, \Psi$

By looking at the assumptions in equations (4) and by collecting the terms depending on $\Lambda$ and $\Psi$, we obtain:
Since we managed to change the summation so that it is with respect to the rows, we can now combine 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:

\[
\sum_{i=1}^{n} (r_i - \mu - \Lambda f)^T \Psi^{-1} (r_i - \mu - \Lambda f) = \sum_{j=1}^{m} \sum_{i=1}^{m} (r_{ji} - \mu_j - f_j^T \Lambda^j)^2 \frac{1}{\Psi_j} \\
= \sum_{j=1}^{m} (r_j^T - \mu_j 1 - F^T \Lambda^j)^T \frac{1}{\Psi_j} (r_j^T - \mu_j 1 - F^T \Lambda^j)
\]

Here, we introduced the following notation:

- \( r_j \) is the \( j^{th} \) row in the matrix of returns \( \mathbf{R} = [r_1 \ldots r_n] \)
- \( F = [f_1 \ldots f_n] \) is the matrix in which we have as columns the common factors.
- \( \mu_j \) is the \( j^{th} \) entry in the vector of means \( \mu \).
- \( \Lambda^j \) is the \( j^{th} \) row in the matrix \( \Lambda \).

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:

\[
\pi(\Lambda, \Psi | \hat{\theta}) \propto \left| \Psi_j^{-\frac{1}{2}} \right| \left| \Psi_j \right| \left| \Psi_j \right|^{-\frac{1}{2}} \exp \left\{ -\beta_j \Psi_j \right\} \\
\times \exp \left\{ -\frac{1}{2} (r_{ij}^T - \mu_j 1 - F^T \Lambda^j)^T \frac{1}{\Psi_j} (r_{ij}^T - \mu_j 1 - F^T \Lambda^j) \right\} \\
\times \exp \left\{ -\frac{1}{2} (\Lambda_j - \Lambda_0)^T \frac{1}{\Psi_j} \mathbf{H}_j^{-1} (\Lambda_j - \Lambda_0) \right\}
\]

(e1) Posterior for \( \Lambda \)

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

\[
(r_{ij}^T - \mu_j 1 - F^T \Lambda^j)^T \frac{1}{\Psi_j} (r_{ij}^T - \mu_j 1 - F^T \Lambda^j) + (\Lambda_j - \Lambda_0)^T \frac{1}{\Psi_j} \mathbf{H}_j^{-1} (\Lambda_j - \Lambda_0) \\
= \Lambda_j^T F \frac{1}{\Psi_j} F^T \Lambda_j + \Lambda_j^T \frac{1}{\Psi_j} \mathbf{H}_j^{-1} \Lambda_j - 2 \Lambda_j^T \frac{1}{\Psi_j} (r_{ij}^T - \mu_j 1) - 2 \Lambda_j^T \frac{1}{\Psi_j} \mathbf{H}_j^{-1} \Lambda_0j \\
\quad + (r_{ij}^T - \mu_j 1)^T \frac{1}{\Psi_j} (r_{ij}^T - \mu_j 1) + \Lambda_0j^T \frac{1}{\Psi_j} \mathbf{H}_j^{-1} \Lambda_0j \\
= \Lambda_j^T \frac{1}{\Psi_j} (FF^T + \mathbf{H}_j^{-1}) \Lambda_j - 2 \Lambda_j^T \frac{1}{\Psi_j} (F (r_{ij}^T - \mu_j 1) + \mathbf{H}_j^{-1} \Lambda_0j) + (r_{ij}^T - \mu_j 1)^T \frac{1}{\Psi_j} (r_{ij}^T - \mu_j 1) \\
\quad + \Lambda_0j^T \frac{1}{\Psi_j} \mathbf{H}_j^{-1} \Lambda_0j
\]
Since the only terms that depend on $\Lambda_j$ 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 $x = \Lambda_j, m = \frac{1}{\Psi_j}(F\Sigma F^T + H_j^{-1}), b = \frac{1}{\Psi_j}(F(r_j^T - \mu, 1) + H_j^{-1}\Lambda_0)$ and we obtain:

$$
\left(\Lambda_j - \Psi_j (F\Sigma F^T + H_j^{-1})^{-1} F\Sigma r_j - \mu, 1 + H_j^{-1}\Lambda_0\right)^T \left(\Lambda_j - \Psi_j (F\Sigma F^T + H_j^{-1})^{-1} F\Sigma r_j - \mu, 1 + H_j^{-1}\Lambda_0\right)
\times
\left(\Lambda_j - \Psi_j (F\Sigma F^T + H_j^{-1})^{-1} F\Sigma r_j - \mu, 1 + H_j^{-1}\Lambda_0\right)^T
\frac{1}{\Psi_j} (F\Sigma F^T + H_j^{-1})
\times
\left(\Lambda_j - \Psi_j (F\Sigma F^T + H_j^{-1})^{-1} F\Sigma r_j - \mu, 1 + H_j^{-1}\Lambda_0\right)
\frac{1}{\Psi_j} (F\Sigma F^T + H_j^{-1})
+ \left(r_j^T - \mu, 1\right)^T \frac{1}{\Psi_j} (r_j^T - \mu, 1) + \Lambda_0^T \frac{1}{\Psi_j} H_j^{-1} \Lambda_0
$$

Therefore, the posterior for $\Lambda_j$ is proportional to:

$$
\pi (\Lambda_j | \Psi_j, \theta) \propto \exp \left\{ -\frac{1}{2} \left(\Lambda_j - \Psi_j \left(\begin{array}{c} F\Sigma F^T + H_j^{-1} \end{array}\right)^{-1} F\Sigma r_j - \mu, 1 + H_j^{-1}\Lambda_0\right) \right\}^T \left(\Lambda_j - \Psi_j \left(\begin{array}{c} F\Sigma F^T + H_j^{-1} \end{array}\right)^{-1} F\Sigma r_j - \mu, 1 + H_j^{-1}\Lambda_0\right)
\times
\frac{1}{\Psi_j} (F\Sigma F^T + H_j^{-1})
\times \frac{1}{\Psi_j} (F\Sigma F^T + H_j^{-1})
\left(\Lambda_j - \Psi_j \left(\begin{array}{c} F\Sigma F^T + H_j^{-1} \end{array}\right)^{-1} F\Sigma r_j - \mu, 1 + H_j^{-1}\Lambda_0\right)
\} \}
$$

That is, given $\Psi_j$, the conditional posterior for $\Lambda_j$ is a normal with mean vector $\bar{\mu}_j$ and covariance matrix $\Psi_j \bar{\Omega}_j$, where:

$$
\bar{\Omega}_j = \left(\begin{array}{c} F\Sigma F^T + H_j^{-1} \end{array}\right)^{-1}
\bar{\mu}_j = \bar{\Omega}_j \left(\begin{array}{c} F\Sigma r_j - \mu, 1 + H_j^{-1}\Lambda_0 \end{array}\right)
$$

(e2) Posterior for $\Psi$

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

$$
\pi (\Psi_j | \Lambda_j^i) \propto \Psi_j^{\frac{n}{2} - \frac{1}{2} - \alpha_j} \exp \left\{ -\frac{1}{2} \left(\begin{array}{c} \beta_j + \frac{1}{2} \left(\begin{array}{c} r_j^T - \mu, 1 - F\Sigma \Lambda_j \end{array}\right) \right) \left(\begin{array}{c} r_j^T - \mu, 1 - F\Sigma \Lambda_j \end{array}\right) \right\} \cdot \exp \left\{ -\frac{1}{2} \left(\begin{array}{c} \left(\begin{array}{c} \Lambda_j - \Lambda_0 \end{array}\right) \right) \right\} \}.
$$

Therefore, we obtain that the posterior for $\Psi_j$ is:

$$
\Psi_j \overset{\text{indep}}{\sim} IG \left(\alpha_{\Psi_j} + \beta_{\Psi_j} \right)
$$

where $\alpha_{\Psi_j} = \frac{n}{2} + \frac{q}{2} + \alpha_j$

and $\beta_{\Psi_j} = \beta_j + \frac{1}{2} \left(\begin{array}{c} \left(\begin{array}{c} r_j^T - \mu, 1 - F\Sigma \Lambda_j \end{array}\right) \right) \left(\begin{array}{c} \left(\begin{array}{c} r_j^T - \mu, 1 - F\Sigma \Lambda_j \end{array}\right) \right) + \left(\begin{array}{c} \left(\begin{array}{c} \Lambda_j - \Lambda_0 \end{array}\right) \right) \right) \left(\begin{array}{c} \left(\begin{array}{c} \Lambda_j - \Lambda_0 \end{array}\right) \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 $\alpha$, 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 $\sigma_1^2, \sigma_2^2$ and $\mu$, the posterior for $\alpha$ can be approximated by a normal density with mean vector $\alpha^*$ and covariance matrix $(Q + G)^{-1}$, where $\alpha^*, G$ and $Q$ are defined in (5) and (6), respectively. The exact conditional posterior given $\sigma_1^2, \sigma_2^2$ and $\mu$ is

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

where $S = \frac{1}{n} \left(\begin{array}{c} r_i - \mu \end{array}\right) \left(\begin{array}{c} r_i - \mu \end{array}\right)^T$
The Metropolis-Hastings step at \(i\)th iteration would be that we would simulate a candidate value from the approximate posterior distribution: \(\tilde{\alpha} \approx \sim N(\alpha^*, (Q + \tilde{G})^{-1})\) and we would accept it with probability \(\min(\rho, 1)\), where

\[
\rho = \frac{\pi(\tilde{\alpha} \mid \cdot)}{\pi(\alpha^* \mid \cdot)} \cdot \frac{\pi^*(\alpha^* \mid \cdot)}{\pi^*(\tilde{\alpha} \mid \cdot)}
\]

It is useful at this point to remember that there is a connection between \(\pi^*\) and \(\pi\), since there is one between \(A\) and \(\alpha\), namely:

\[
\alpha = \text{Vec}^*(\log(\Phi))\quad A = \log(\Phi)
\]

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.
- \(\hat{F}_{init} = \left[\hat{f}_1 \hat{f}_2 \ldots \hat{f}_n\right]\), where \(\hat{f}_i\) for \(i \in \{1, 2, \ldots, n\}\) 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 \(f_i^j \phi \sim N_q(0, \Phi)\). In order to specify \(\Phi_{init}\), we take the covariance of the above found common factors: \(\Phi_{init} = \text{Cov}(\hat{f})\).

- We also have the following assumption.

\[
\alpha \mid \theta, \Lambda = \text{Vec}^*(\log(\Phi)) \mid \theta, \Lambda \sim N_d\left(\left[\begin{array}{c}
\theta_1 1_q \\
\theta_2 1_{d-q}
\end{array}\right], \left[\begin{array}{cc}
\sigma_1^2 1_q & 0 \\
0 & \sigma_2^2 1_{d-q}
\end{array}\right]\right)
\]

- In order to initialize \(\sigma_1^2\), we have to take the variance of the first \(q\) entries in \(\text{Vec}^*(\log(\hat{\Phi}_{init}))\).

- In order to initialize \(\sigma_2^2\), we have to take the variance of the last \(d - q\) entries in \(\text{Vec}^*(\log(\hat{\Phi}_{init}))\).

- We remember that \(\Lambda_j \mid \Psi_j \sim N_q(\Lambda_{0j}, \Psi_j H_j)\). Since, in general, we do not have any prior information on the factor weights, we specify the hyper-parameters to be:

\[
\hat{\Lambda}_{0j init} = 0
\]

We initialize the variance \(\Psi_j H_j\) with a big value: \(\hat{\Psi}_{j init} = 1\) and \(\hat{H}_{j init} = 10^{10} I_q\).

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

Using the Metropolis Hastings step that was just discussed, we arrive at the following Gibbs Sampler:
Algorithm 1: Gibbs Sampler log(Φ)

1: \[ a \sim N\left((Q(i) + G(i))^{-1} Q(i)F(i), (Q(i) + G(i))^{-1}\right) \text{ w.p. } \min(p, 1) \]

2: Since \( \alpha \sim \text{Vec}^\ast \left(\log(\Phi)\right) \Rightarrow \text{compute } \Phi^{(i+1)} = \exp \left( V_{cc}^{-1} \left(a^{(i+1)}\right) \right) \]

3: \[ \begin{align*}
    \omega_{1}^{(i+1)} & \sim IG \left( \frac{1}{2} + \frac{1}{2} \sum_{j} \left( \alpha_{j}^{(i+1)} - \mu_{j}\right) \right) \\
    \omega_{2}^{(i+1)} & \sim IG \left( \frac{1}{2} + \frac{1}{2} \sum_{j} \left( \alpha_{j}^{(i+1)} - \mu_{j}\right) \right) \\
    \Rightarrow & \quad \Lambda^{(i+1)} = \begin{pmatrix}
    \omega_{1}^{(i+1)}I_{p} & 0 \\
    0 & \omega_{2}^{(i+1)}I_{p}\end{pmatrix}
\end{align*} \]

4: Let \( \Sigma_{\phi} = \left(n^{(i)} + 1 + \Sigma \Omega^{-1}F \right)^{-1} \Rightarrow \mu^{(i+1)} = \left( \Sigma_{\phi} \left( n^{(i)} + 1 + \mu^{(i+1)} \right) \right)^{-1} \]

5: Let \( \Sigma = \left( A^{(i+1)} \Psi \Omega^{-1} + \Phi^{(i+1)} \right)^{-1} \Rightarrow \mu^{(i+1)} = \left( \Sigma_{\phi} \left( n^{(i)} + 1 + \mu^{(i+1)} \right) \right)^{-1} \]

2.5 Results: Personalized Views on 4 Stocks

We would like to conduct sensitivity analysis for the confidence in the investor views (the diagonal elements of the \( \Omega \) matrix denoted by \( \omega_{l} \)). 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 \( P \) 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:

\[
\begin{bmatrix}
0.02 \\
0.05
\end{bmatrix}, \quad \begin{bmatrix}
\text{AAPL} & \text{FB} & \text{GOOG} & \text{MSFT}
\end{bmatrix} = \begin{bmatrix}
-1 & 1 & 0 & 0 \\
0 & 0 & 1 & -1
\end{bmatrix}
\]

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 run in parallel on multiple cores (each core running the Gibbs Sampler for 1 pair (\( \omega_{1}, \omega_{2} \)), 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 (\( \omega_{1} \) and \( \omega_{2} \)) and the third one is represented by the \( F \) distance \( ||P_{\text{post}} - q_{0}||_2 \). As mentioned previously, this distance should go to 0 as \( \omega_{1} \) and \( \omega_{2} \) go to 0, which can easily be observed in Figure 1.
Furthermore, similar to the versions introduced previously, as $\omega_1$ and $\omega_2$ increase, the distance converges to the same number. Since $\omega_1$ and $\omega_2$ 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 $\mu$ is the sample mean of the returns, $\bar{r}$. The distance $||PF - q_0||_2 = 0.05386381$, 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 $\Sigma_{post} = A_{post} \Phi_{post} A_{post}^T + \Psi_{post}$ and we try to maximize the portfolio returns, while minimizing the portfolio risk. Hence, we would like to find $\max_w w^T \mu_{post} - \frac{1}{2} w^T \Sigma_{post} w$, where $\lambda$ is the investor’s risk aversion coefficient. In his paper, \(^1\) suggests that $\lambda = 2.5$ is a reasonable choice for equities. By making the derivative with respect to $w$ equal to 0, and by solving the resulting equation for $w$, we obtain: $w^* = \frac{1}{2\lambda} \Sigma_{post}^{-1} \mu_{post}$. 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 ($\omega_1$ and $\omega_2$) 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 $\omega_1$.
2. **Energy** outperforms **industrials** by 0.1% with confidence level $\omega_2$.
3. **Real Estate** outperforms **consumer staples** by 0.2% with confidence level $\omega_2$.
4. **Consumer discretionary** outperforms **financials** by 0.3% with confidence level $\omega_2$.

\(^1\)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 $q_0 = (0.002, 0.001, 0.002, 0.003)^T$.

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 $\omega_1$ and $\omega_2$ and the third axis represented by the distance $||P\mu_{post} - q_0||$. 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](image)

Again, the model behaves exactly as our intuition suggests. As $\omega_1$ and $\omega_2$ go to 0, the distance $||P\mu_{post} - q_0||$ converges to 0. Moreover, for bigger values of $\omega_1$ and $\omega_2$ (small confidence in views) the distance converges to 0.004999748 = $||P\mu - q_0||$. 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 $100,000, the same testing data set over the month of January 2018 and the same methodology for computing the portfolio weights. The mean of the profits over all the simulated pairs ($\omega_1, \omega_2$) was $37,576.68 with a standard deviation of $5,857.198.

Acknowledgement

Use was made of computational facilities purchased with funds from the National Science Foundation (CNS-1725797) and administered by the Center for Scientific Computing (CSC). The CSC is supported by the California NanoSystems Institute and the Materials Research Science and Engineering Center (MRSEC; NSF DMR 1720256) at UC Santa Barbara.

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
Appendix

S&P500 Industry Sectors

The stocks in the S&P500 are divided into broad groupings based on economic characteristics. Currently there are 11 industry sectors³

- 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

Copyright for this article is retained by the author(s), with first publication rights granted to the journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).

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