

A Bayesian Approach for Asset Allocation

Mihnea S. Andrei^{1*} & 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: April 10, 2020 Accepted: May 11, 2020 Online Published: May 15, 2020

doi:10.5539/ijsp.v9n4p1 URL: <https://doi.org/10.5539/ijsp.v9n4p1>

Abstract

The Black-Litterman model combines investors' personal views with historical data and gives optimal portfolio weights. In this paper we will introduce the original Black-Litterman model (**Section 1**), we will modify the model such that it fits in a Bayesian framework by considering the investors' personal views to be a direct prior on the means of the returns and by including a typical Inverse Wishart prior on the covariance matrix of the returns (**Section 2**). We will also consider an idea of Leonard & Hsu [1992] for a prior on the logarithm of the covariance matrix (**Section 3**). Sensitivity analysis for the level of confidence that investors have in their own personal views was performed and performance of the models was assessed on a test data set consisting of returns over the month of January 2018.

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

1. Black-Litterman Asset Allocation Model

The Black-Litterman asset allocation model, developed by Fischer Black and Robert Litterman in the early 90's while working for Goldman Sachs, 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 $\mathbf{r} = (r_1, r_2, \dots, r_m)^T$ follow a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. That is

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

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

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

where $\boldsymbol{\pi}$ is the equilibrium risk parameter, which can be expressed as

$$\boldsymbol{\pi} = \delta\boldsymbol{\Sigma}\mathbf{w}_{eq}$$

where the parameter δ is the investor's risk aversion parameter, \mathbf{w}_{eq} is the vector of m equilibrium weights, and the parameter τ in equation (2) 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 following equation:

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

where the matrix \mathbf{P} is a $k \times m$ matrix, \mathbf{q} 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} represents a personal view. To illustrate all terms in equation (3), we consider an example with four assets: Apple Inc (AAPL), Amazon.com Inc (AMZN), Alphabet Inc Class C (GOOG), and Microsoft Corporation (MSFT). Let $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3, \mu_4)^T$ represent the mean returns for AAPL, AMZN, GOOG, and MSFT, respectively. Suppose that the investor believes that AAPL will outperform AMZN by 2% and that GOOG will have returns that amount to 5%. In that case,

$$\mathbf{P} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{q} = \begin{bmatrix} 0.02 \\ 0.05 \end{bmatrix}$$

* 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.

The covariance matrix Ω is usually a diagonal matrix. The diagonal elements represent the uncertainty of each view. A small value reflects a high confidence in the view and vice versa.

He & Litterman [2002] reported that, by combining the prior information in equations (2) and (3), the mean return follows a normal distribution

$$\mu \sim N(\bar{\mu}, \bar{M}^{-1}) \tag{4}$$

where

$$\bar{\mu} = [(\tau\Sigma)^{-1} + \mathbf{P}^T\Omega^{-1}\mathbf{P}]^{-1}[(\tau\Sigma)^{-1}\pi + \mathbf{P}^T\Omega^{-1}\mathbf{q}]$$

and

$$\bar{M}^{-1} = [(\tau\Sigma)^{-1} + \mathbf{P}^T\Omega^{-1}\mathbf{P}]^{-1}$$

The combined prior in equation (4) compromises the proposed CAPM prior in (2) and the investor's views in (3). The combined mean $\bar{\mu}$ is a weighted average of the CAPM prior mean π and the mean \mathbf{q} of the investor's views, with the weights, $(\tau\Sigma)^{-1}$ and $\mathbf{P}^T\Omega\mathbf{P}$, respectively. The combined prior mean $\bar{\mu}$ is closer to the CAPM prior mean π when the uncertainty parameter τ is small, or equivalently, when we are more certain about the CAPM prior.

The unconditional distribution of \mathbf{r} is therefore

$$\mathbf{r} \sim N(\bar{\mu}, \bar{\Sigma}) \tag{5}$$

which is obtained by combining equations (1) and (4) and then integrating the vector μ out, where $\bar{\Sigma} = \Sigma + \bar{M}^{-1}$. Based on the unconditional mean $\bar{\mu}$ and covariance matrix $\bar{\Sigma}$ in equation (5), the optimal portfolio can be determined by using the standard mean-variance optimization method to maximize $\mathbf{w}^T\bar{\mu} - \frac{\delta}{2}\mathbf{w}^T\bar{\Sigma}\mathbf{w}$ with respect to the weights \mathbf{w} for an investor with the risk aversion parameter δ . He & Litterman [2002] reported that the optimal portfolio weights \mathbf{w}^* can be expressed as

$$\mathbf{w}^* = \frac{1}{1 + \tau}(\mathbf{w}_{eq} + \mathbf{P}^T\mathbf{w}) \tag{6}$$

where

$$\mathbf{w} = \frac{\tau}{\delta}\Omega^{-1}\mathbf{q} - \frac{1}{1 + \tau}\mathbf{A}^{-1}\mathbf{P}\Sigma\mathbf{w}_{eq} - \frac{1}{\delta(1 + \tau)}\mathbf{A}^{-1}\mathbf{P}\Sigma\mathbf{P}^T\Omega^{-1}\mathbf{q}$$

and

$$\mathbf{A} = \frac{1}{\tau}\Omega + \frac{1}{1 + \tau}\mathbf{P}\Sigma\mathbf{P}^T.$$

Black-Litterman suggested to replace the covariance matrix Σ by a matrix estimated from historical data, then treated Σ as a known covariance matrix in their model. The optimal portfolio weights \mathbf{w}^* can be obtained by plugging in all known parameters, the CAPM prior mean π , the uncertainty parameter τ , personal views parameters \mathbf{P} , \mathbf{q} , Ω , and the covariance matrix Σ . The model they proposed was a probability model. The optimal portfolio weights were easily obtained by plugging in all parameters. No data were collected, only the covariance matrix was obtained using historical data. In this paper, we will propose instead a statistical approach, indeed, a complete Bayesian statistical approach, which takes into consideration investor's views. We will focus our attention on 2 cases:

- (1) when historical data is available
- (2) when historical data is not available

2. A Bayesian Approach, When Historical Data Is Available

2.1 Prior and Posterior Distributions

Let $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n$ be n independent returns, where each \mathbf{r}_i represents the returns of m assets and follows the distribution specified in equation (1), that is,

$$\mathbf{r}_i \sim N_m(\mu, \Sigma), \text{ for } i = 1, 2, \dots, n \tag{7}$$

We consider commonly used priors for μ and Σ :

$$\mu \sim N_m(\pi, \Delta) \tag{8}$$

and Σ follows an Inverse Wishart distribution with ν degrees of freedom and a location parameter Σ_0 , that is,

$$\Sigma \sim W^{-1}(\nu, \Sigma_0) \tag{9}$$

Smaller values of the degrees of freedom parameter ν imply an increasingly more diffuse distribution and larger values for the degrees of freedom parameter ν yield a more highly concentrated distribution about the location matrix parameter Σ_0 . Please note, the prior for μ in equation (8) is similar to equation (2) used for the Black-Litterman approach, but with

the more general covariance matrix Λ replacing the more restricted $\tau\Sigma$. Suppose that historical data is available, then we can determine the prior parameters in equations (8) and (9) using historical data.

We further suppose that investor’s views specified in equation (3) are also available:

$$\mathbf{P}\boldsymbol{\mu} \sim N_k(\mathbf{q}, \boldsymbol{\Omega}) \tag{10}$$

and we would like to include that as part of our prior information. It is obvious that the two proposed priors on $\boldsymbol{\mu}$ from equations (8) and (10) would suggest inconsistent information in practice. We would like to retain the information as much as possible, whether the information is from investor’s views or historical data. However, we believe that in practice investor’s views are more valuable than the more objective prior information based on historical data because this gives investors more power if they also use other models in order to create their inputs for the views.

Let us take a look at the matrix \mathbf{P} in detail. Suppose that the investor has k views. Those k views are expressed in equation (10). Each row in \mathbf{P} represents a view about the m assets. Those views can be classified as relative views (the rows that sum up to 0) and absolute views (only one 1 in a row). The k views should be linearly independent. If they are not all linearly independent, then some views would be either redundant or inconsistent. As a result, at most $k = m$ linearly independent views can be expressed. In the case when $k = m$, the matrix \mathbf{P} in equation (10) is invertible and only investor’s views will be used and the prior information in equation (8) will be automatically ignored. In the case when the number of views k is less than the number of assets m , we will use all information based on investor’s views, and additional information based on historical data as much as possible. We now consider an augmented matrix \mathbf{P}^* based on the matrix \mathbf{P} in equation (10), such that the augmented matrix \mathbf{P}^* is invertible.

In here, we will present a method in which we can add rows to \mathbf{P} such that the resulting square matrix \mathbf{P}^* is invertible. The main idea is based on the way in which one would row reduce a matrix to the echelon form. It is well known that a matrix is invertible if and only if its row reduced echelon form is the identity matrix. This gives us the idea of taking our matrix \mathbf{P} and adding rows to it in order to make it invertible:

- For each column in \mathbf{P} that has only 0’s, we have to create a new row that will have only one 1 in the respective column and 0’s in all the others.
- If a row has more than 1 nonzero entry, for each such entry except the entries in the pivot columns, we have to create a row in which we have a 1.

For example, if we consider the illustrative matrix \mathbf{P} used in section 1, the above procedure gives us:

$$\mathbf{P} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} = \mathbf{P}^* = \begin{bmatrix} \mathbf{P} \\ \mathbf{P}_2 \end{bmatrix}$$

The augmented matrix \mathbf{P}^* consists of two parts, the original personal views matrix \mathbf{P} and a newly created $(m - k) \times m$ matrix \mathbf{P}_2 to make the augmented matrix \mathbf{P}^* invertible.

We now transform the data through \mathbf{P}^* : $\mathbf{r}_i^* = \mathbf{P}^* \mathbf{r}_i$ for $i = 1, 2, \dots, n$. Then $\mathbf{r}_1^*, \mathbf{r}_2^*, \dots, \mathbf{r}_n^*$ are independent and

$$\mathbf{r}_i^* \sim N_m(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*), \text{ for } i = 1, 2, \dots, n \tag{11}$$

where $\boldsymbol{\mu}^* = \mathbf{P}^* \boldsymbol{\mu}$ and $\boldsymbol{\Sigma}^* = \mathbf{P}^* \boldsymbol{\Sigma} \mathbf{P}^{*T}$. We further consider the priors in equations (8) and (9), but on $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$,

$$\boldsymbol{\mu}^* \sim N_m(\boldsymbol{\pi}^*, \boldsymbol{\Delta}^*) \tag{12}$$

and

$$\boldsymbol{\Sigma}^* \sim W^{-1}(\nu^*, \boldsymbol{\Sigma}_0^*) \tag{13}$$

Suppose that historical data is available, we can objectively specify those prior parameters in equations (12) and (13) in practice, as follows: we calculate the sample covariance matrix using the whole historical data and that covariance matrix will be used as the location parameter matrix $\boldsymbol{\Sigma}_0^*$ and the number of historical returns will be used as the degrees of freedom parameter ν^* . We then split the historical data into many groups of size n . The sample mean vectors are calculated for each group. Then the average of those sample mean vectors will be used as $\boldsymbol{\pi}^*$ and the sample covariance matrix based on those sample mean vectors will be used as $\boldsymbol{\Delta}^*$. Note that, in equation (12),

$$\boldsymbol{\mu}^* = \mathbf{P}^* \boldsymbol{\mu} = \begin{bmatrix} \mathbf{P} \\ \mathbf{P}_2 \end{bmatrix} \boldsymbol{\mu} = \begin{bmatrix} \mathbf{P}\boldsymbol{\mu} \\ \mathbf{P}_2\boldsymbol{\mu} \end{bmatrix} \sim N(\boldsymbol{\pi}^*, \boldsymbol{\Delta}^*)$$

where

$$\boldsymbol{\pi}^* = \begin{bmatrix} \pi_1^* \\ \pi_2^* \end{bmatrix} \text{ and } \boldsymbol{\Delta}^* = \begin{bmatrix} \Delta_{11}^* & \Delta_{12}^* \\ \Delta_{21}^* & \Delta_{22}^* \end{bmatrix}$$

are partitioned accordingly, $\boldsymbol{\pi}_1^*$ is a $k \times 1$ vector and $\boldsymbol{\Delta}_{11}^*$ is a $k \times k$ matrix. We now impose the investor's views specified in (10)

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

by replacing $\boldsymbol{\pi}_1^*$ and $\boldsymbol{\Delta}_{11}^*$ by \mathbf{q} and $\boldsymbol{\Omega}$, respectively. Therefore,

$$\boldsymbol{\mu}^* \sim N_m(\boldsymbol{\pi}^*, \boldsymbol{\Delta}^*), \tag{14}$$

where

$$\boldsymbol{\pi}^* = \begin{bmatrix} \mathbf{q} \\ \boldsymbol{\pi}_2^* \end{bmatrix} \text{ and } \boldsymbol{\Delta}^* = \begin{bmatrix} \boldsymbol{\Omega} & \Delta_{12}^* \\ \Delta_{21}^* & \Delta_{22}^* \end{bmatrix} \tag{15}$$

We now successfully combined two sources of prior information, \mathbf{q} and $\boldsymbol{\Omega}$ according to investor's views, and $\boldsymbol{\pi}_2^*$, Δ_{12}^* , Δ_{21}^* and Δ_{22}^* , objectively specified according to historical data.

The sampling distribution of $\mathbf{r}_1^*, \mathbf{r}_2^*, \dots, \mathbf{r}_n^*$ in equation (11) gives the following density:

$$L(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^* | \mathbf{r}_1^*, \mathbf{r}_2^*, \dots, \mathbf{r}_n^*) \propto |\boldsymbol{\Sigma}^*|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (\mathbf{r}_i^* - \boldsymbol{\mu}^*)^T \boldsymbol{\Sigma}^{*-1} (\mathbf{r}_i^* - \boldsymbol{\mu}^*) \right\} \tag{16}$$

Also, we had independent prior distributions on $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ in equations (12) and (13), respectively:

$$\pi(\boldsymbol{\mu}^* | \boldsymbol{\pi}^*, \boldsymbol{\Delta}^*) \propto |\boldsymbol{\Delta}^*|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\boldsymbol{\mu}^* - \boldsymbol{\pi}^*)^T \boldsymbol{\Delta}^{*-1} (\boldsymbol{\mu}^* - \boldsymbol{\pi}^*) \right\} \tag{17}$$

and

$$\pi(\boldsymbol{\Sigma}^* | \nu, \boldsymbol{\Sigma}_0) \propto |\boldsymbol{\Sigma}^*|^{-\frac{1}{2}(\nu+m+1)} \exp \left\{ -\frac{1}{2} Tr(\boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}^{*-1}) \right\} \tag{18}$$

The joint posterior density of $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ given $\mathbf{r}_1^*, \mathbf{r}_2^*, \dots, \mathbf{r}_n^*$ is proportional to the product of equations (16), (17), and (18) and can be represented as

$$\begin{aligned} \pi(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^* | \mathbf{r}_1^*, \mathbf{r}_2^*, \dots, \mathbf{r}_n^*) &\propto |\boldsymbol{\Sigma}^*|^{-\frac{1}{2}(\nu+n+m+1)} \exp \left\{ -\frac{1}{2} Tr(\boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}^{*-1}) \right\} \\ &\times \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (\mathbf{r}_i^* - \boldsymbol{\mu}^*)^T \boldsymbol{\Sigma}^{*-1} (\mathbf{r}_i^* - \boldsymbol{\mu}^*) \right\} \\ &\times |\boldsymbol{\Delta}^*|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\boldsymbol{\mu}^* - \boldsymbol{\pi}^*)^T \boldsymbol{\Delta}^{*-1} (\boldsymbol{\mu}^* - \boldsymbol{\pi}^*) \right\} \end{aligned} \tag{19}$$

We rearrange the second exponent, then combine the two quadratic functions of $\boldsymbol{\mu}^*$ in the last two exponents in equation (19). The joint density becomes

$$\begin{aligned} \pi(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^* | \mathbf{r}_1^*, \mathbf{r}_2^*, \dots, \mathbf{r}_n^*) &\propto \exp \left\{ -\frac{1}{2} (\boldsymbol{\mu}^* - \bar{\boldsymbol{\mu}}^*)^T (n\boldsymbol{\Sigma}^{*-1} + \boldsymbol{\Delta}^{*-1}) (\boldsymbol{\mu}^* - \bar{\boldsymbol{\mu}}^*) \right\} \\ &\times \exp \left\{ -\frac{1}{2} (\bar{\mathbf{r}}^* - \boldsymbol{\pi}^*)^T \mathbf{H} (\bar{\mathbf{r}}^* - \boldsymbol{\pi}^*) + (n-1)\mathbf{S} \right\} \\ &\times |\boldsymbol{\Sigma}^*|^{-\frac{1}{2}(\nu+n+m+1)} \exp \left\{ -\frac{1}{2} Tr(\boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}^{*-1}) \right\} \end{aligned} \tag{20}$$

Where

$$\begin{aligned} \mathbf{S} &= \frac{1}{n-1} \sum_{i=1}^n (\mathbf{r}_i^* - \bar{\mathbf{r}}^*)^T \boldsymbol{\Sigma}^{*-1} (\mathbf{r}_i^* - \bar{\mathbf{r}}^*), \quad \bar{\mathbf{r}}^* = \frac{1}{n} \sum_{i=1}^n \mathbf{r}_i^*, \\ \bar{\boldsymbol{\mu}}^* &= (n\boldsymbol{\Sigma}^{*-1} + \boldsymbol{\Delta}^{*-1})^{-1} (n\boldsymbol{\Sigma}^{*-1} \bar{\mathbf{r}}^* + \boldsymbol{\Delta}^{*-1} \boldsymbol{\pi}^*) \end{aligned} \tag{21}$$

and

$$\mathbf{H} = \left(\frac{1}{n} \boldsymbol{\Sigma}^* + \boldsymbol{\Delta}^* \right)^{-1}$$

We can implement a Gibbs Sampler/Markov Chain Monte Carlo (MCMC) procedure to compute the posterior means for $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$. To facilitate the procedure, all conditional posterior distributions, $\boldsymbol{\mu}^*$ given $\boldsymbol{\Sigma}^*$ and $\boldsymbol{\Sigma}^*$ given $\boldsymbol{\mu}^*$ are needed. Following equation (20), the conditional posterior distribution of $\boldsymbol{\mu}^*$, given $\boldsymbol{\Sigma}^*$ can be represented as

$$\pi(\boldsymbol{\mu}^* | \mathbf{r}_1^*, \mathbf{r}_2^*, \dots, \mathbf{r}_n^*, \boldsymbol{\Sigma}^*) \propto \exp \left\{ -\frac{1}{2} (\boldsymbol{\mu}^* - \bar{\boldsymbol{\mu}}^*)^T (n\boldsymbol{\Sigma}^{*-1} + \boldsymbol{\Delta}^{*-1}) (\boldsymbol{\mu}^* - \bar{\boldsymbol{\mu}}^*) \right\} \quad (22)$$

That is, conditional on $\boldsymbol{\Sigma}^*$, the vector $\boldsymbol{\mu}^*$ follows a normal distribution with posterior mean $\bar{\boldsymbol{\mu}}^*$ specified in equation (21) and posterior variance $(n\boldsymbol{\Sigma}^{*-1} + \boldsymbol{\Delta}^{*-1})^{-1}$. Following equation (19), the conditional posterior distribution of $\boldsymbol{\Sigma}^*$ given $\boldsymbol{\mu}^*$ is

$$\pi(\boldsymbol{\Sigma}^* | \mathbf{r}_1^*, \mathbf{r}_2^*, \dots, \mathbf{r}_n^*, \boldsymbol{\mu}^*) \propto |\boldsymbol{\Sigma}^*|^{-\frac{1}{2}(\nu+n+m+1)} \times \exp \left\{ -\frac{1}{2} Tr \left[\left(\boldsymbol{\Sigma}_0 + \sum_{i=1}^n (\mathbf{r}_i^* - \boldsymbol{\mu}^*)(\mathbf{r}_i^* - \boldsymbol{\mu}^*)^T \right) \boldsymbol{\Sigma}^{*-1} \right] \right\} \quad (23)$$

That is, conditional on $\boldsymbol{\mu}^*$, the matrix $\boldsymbol{\Sigma}^*$ follows an Inverse Wishart with degree of freedom parameter, $\nu^* = \nu + n$ and location matrix parameter $\boldsymbol{\Sigma}_0^* = \boldsymbol{\Sigma}_0 + \sum_{i=1}^n (\mathbf{r}_i^* - \boldsymbol{\mu}^*)(\mathbf{r}_i^* - \boldsymbol{\mu}^*)^T$. The posterior means can be calculated according to Algorithm 1.

Algorithm 1 Gibbs Sampler

1: $\boldsymbol{\Sigma}^{*(t+1)} | \mathbf{r}_1^*, \mathbf{r}_2^*, \dots, \mathbf{r}_n^*, \boldsymbol{\mu}^{*(t)} \sim W^{-1}(\nu + n, \boldsymbol{\Sigma}_0^*)$, where

$$\boldsymbol{\Sigma}_0^* = \boldsymbol{\Sigma}_0 + \sum_{i=1}^n (\mathbf{r}_i^* - \boldsymbol{\mu}^{*(t)})(\mathbf{r}_i^* - \boldsymbol{\mu}^{*(t)})^T$$

2: $\boldsymbol{\mu}^{*(t+1)} | \mathbf{r}_1^*, \dots, \mathbf{r}_n^*, \boldsymbol{\Sigma}^{*(t+1)} \sim N_m(\bar{\boldsymbol{\mu}}^{*(t+1)}, \bar{\boldsymbol{\Sigma}}^{*(t+1)})$, where

$$\begin{aligned} \bar{\boldsymbol{\mu}}^{*(t+1)} &= \left(n\boldsymbol{\Sigma}^{*(t+1)-1} + \boldsymbol{\Delta}^{*-1} \right)^{-1} \left(n\boldsymbol{\Sigma}^{*(t+1)-1} \bar{\mathbf{r}}^* + \boldsymbol{\Delta}^{*-1} \boldsymbol{\pi}^* \right) \\ \bar{\boldsymbol{\Sigma}}^{*(t+1)} &= \left(n\boldsymbol{\Sigma}^{*(t+1)-1} + \boldsymbol{\Delta}^{*-1} \right)^{-1} \end{aligned}$$

Once the posterior means $E(\boldsymbol{\mu}^* | \mathbf{r}_1^*, \mathbf{r}_2^*, \dots, \mathbf{r}_n^*)$ and $E(\boldsymbol{\Sigma}^* | \mathbf{r}_1^*, \mathbf{r}_2^*, \dots, \mathbf{r}_n^*)$ for $\boldsymbol{\mu}^* = \mathbf{P}^* \boldsymbol{\mu}$ and $\boldsymbol{\Sigma}^* = \mathbf{P}^* \boldsymbol{\Sigma} \mathbf{P}^{*T}$ are calculated, we can transform back to the means for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$

$$\tilde{\boldsymbol{\mu}} = E(\boldsymbol{\mu} | \mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n) = \mathbf{P}^{*-1} E(\boldsymbol{\mu}^* | \mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n)$$

and

$$\tilde{\boldsymbol{\Sigma}} = E(\boldsymbol{\Sigma} | \mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n) = (\mathbf{P}^{*-1}) E(\boldsymbol{\Sigma}^* | \mathbf{r}_1^*, \mathbf{r}_2^*, \dots, \mathbf{r}_n^*) (\mathbf{P}^{*-1})^T$$

The optimal portfolio can then be determined by using the standard mean-variance optimization method: maximize $\mathbf{w}^T \tilde{\boldsymbol{\mu}} - \frac{\delta}{2} \mathbf{w}^T \tilde{\boldsymbol{\Sigma}} \mathbf{w}$ with respect to the weights \mathbf{w} for an investor with the risk aversion parameter δ . The resulting optimal portfolio weights \mathbf{w}^* can be expressed as

$$\mathbf{w}^* = \frac{1}{\delta} \tilde{\boldsymbol{\Sigma}}^{-1} \tilde{\boldsymbol{\mu}}$$

2.2 Implementation

The simulations in this section were done a couple of years ago, which is the reason why the data ends in 2017, as we will soon see. At that time, downloading closing prices was easily done using the *quantmod* package in *R*. For illustration purposes, 4 stocks were considered: Apple(AAPL), Amazon(AMZN), Alphabet Inc Class C.(GOOG) and Microsoft(MSFT). Closing prices for those 4 stocks from January 2nd 2015 to May 1st 2017 were considered and the returns were computed. This data is split into two parts: one representing the *current data* (the last n returns $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n$, here $n = 21$) and the rest representing *historical data* used to determine the prior parameters in the model. We chose $n = 21$ because we are thinking of modeling the returns that happen within a period of approximately a month and 21

is the average number of trading days in a month. For this example, suppose that the investor believes that AAPL will outperform AMZN by 2% and that GOOG will outperform MSFT by 5%. The investor's views are presented as

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

where

$$\mathbf{P} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \mathbf{q} = \begin{bmatrix} 0.02 \\ 0.05 \end{bmatrix}, \boldsymbol{\Omega} = \begin{bmatrix} 0.0001 & 0 \\ 0 & 0.0001 \end{bmatrix}$$

where the columns in \mathbf{P} represent the four stocks AAPL, AMZN, GOOG, MSFT, respectively. The following augmented matrix \mathbf{P}^* was created according to the procedure suggested in section 2.1.

$$\mathbf{P}^* = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

We split the historical data into groups, where each group contains 21 returns. The mean vectors were calculated for each group. Then the average and the covariance matrix of those mean vectors were used to determine the entries in the vector $\boldsymbol{\pi}^*$ and covariance matrix $\boldsymbol{\Delta}^*$, in addition to the already supplied \mathbf{q} and $\boldsymbol{\Omega}$ according to investor's views, in equation (15).

$$\boldsymbol{\pi}^* = \begin{bmatrix} 0.02 \\ 0.05 \\ 0.0011579 \\ 0.0007917 \end{bmatrix}$$

$$\boldsymbol{\Delta}^* = \begin{bmatrix} \omega_1 & 0 & -1.072 \times 10^{-5} & -2.665 \times 10^{-7} \\ 0 & \omega_2 & 1.980 \times 10^{-6} & -5.312 \times 10^{-6} \\ -1.072 \times 10^{-5} & 1.980 \times 10^{-6} & 1.487 \times 10^{-5} & 3.732 \times 10^{-6} \\ -2.665 \times 10^{-7} & -5.312 \times 10^{-6} & 3.732 \times 10^{-6} & 9.331 \times 10^{-6} \end{bmatrix}$$

Where $\omega_1 = \omega_2 = 0.0001$. Our prior parameters are all fully supplied and Algorithm 1 was implemented.

A burning period of 10^3 was chosen and the number of iterations for the Gibbs Sampler is 10^4 . After the Gibbs Sampler is completed, one would only have to take the mean of the simulated $\boldsymbol{\mu}^{*(t)}$, call it $\tilde{\boldsymbol{\mu}}^*$, and the average of the simulated $\boldsymbol{\Sigma}^{(t)}$, call it $\tilde{\boldsymbol{\Sigma}}^*$. The specific values obtained from the Gibbs Sampler for those examples are:

$$\tilde{\boldsymbol{\mu}}^* = \begin{bmatrix} 0.0001361794 \\ 0.0017267513 \\ 0.0011532351 \\ 0.0009999096 \end{bmatrix}$$

$$\tilde{\boldsymbol{\Sigma}}^* = \begin{bmatrix} 8.867 \times 10^{-5} & -1.423 \times 10^{-5} & -7.09 \times 10^{-5} & -2.234 \times 10^{-6} \\ -1.423 \times 10^{-5} & 9.588 \times 10^{-5} & 1.517 \times 10^{-5} & -1.385 \times 10^{-5} \\ -7.09 \times 10^{-5} & 1.517 \times 10^{-5} & 9.562 \times 10^{-5} & 2.4 \times 10^{-5} \\ -2.234 \times 10^{-6} & -1.385 \times 10^{-5} & 2.400 \times 10^{-5} & 4.488 \times 10^{-5} \end{bmatrix}$$

However, one has to remember that those were transformed using \mathbf{P}^* , hence now we would have to transform them back into the original space: $\tilde{\boldsymbol{\mu}} = \mathbf{P}^{*-1}\tilde{\boldsymbol{\mu}}^*$, $\tilde{\boldsymbol{\Sigma}} = \mathbf{P}^{*-1}\tilde{\boldsymbol{\Sigma}}^*\mathbf{P}^{*-T}$. As shown before, the weights are computed according to $\mathbf{w}^* = \frac{1}{\delta}\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{\mu}}$, where $\delta = 2.5$, as chosen in the original Black-Litterman model. Also there has been extensive research when it comes to choosing δ (please see Janacek [2004]). For trading stocks, a risk aversion coefficient between 2 and 3 is reasonable. More specifically, the weights in this example are:

$$\mathbf{w}^* = \begin{bmatrix} 0.46667044 \\ -0.03372112 \\ 0.46770841 \\ -0.03190002 \end{bmatrix}$$

The weights are normalized so that the sum of the absolute values of their entries is 1. Also, please note that a negative weight is possible and it corresponds in finance to short selling. Short selling 1 share of Google is done by borrowing 1 share of Google from the market manager, selling it on the market instantly, but with the promise of buying it back and giving it back to the market manager at some future date.

2.3 Sensitivity Analysis

In the analysis presented in this section, we will use the same data-set and investor personal views inputs (\mathbf{P} and \mathbf{q}). In practice, it is of great interest to see how sensitive the models are to changes in the confidence levels that the investor inputs ($\mathbf{\Omega}$). Our intuition about the world says that:

- 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

Our intuition of the world should be reflected in the model assumptions and also in the results. The following remark shows that they are indeed:

Remark 1. Since $\mathbf{P}\boldsymbol{\mu} \sim N(\mathbf{q}, \mathbf{\Omega})$, we have that $\lim_{\mathbf{\Omega} \rightarrow \mathbf{0}_2} \mathbf{P}\boldsymbol{\mu} = \mathbf{q}$ a.s.

Therefore, as the diagonal entries of $\mathbf{\Omega}$ get smaller and smaller ($diag(\mathbf{\Omega})$ contains standard deviations squared), we expect to get closer and closer to \mathbf{q} .

This remark suggests also the way in which we will conduct sensitivity analysis. For both models an exhaustive method was implemented that would compute for each pair of diagonal entries in $\mathbf{\Omega}$ a posterior mean $\hat{\boldsymbol{\mu}}^*$. This is transformed back to the original space of returns: $\hat{\boldsymbol{\mu}} = \mathbf{P}^{*-1}\hat{\boldsymbol{\mu}}^*$. Once this is obtained, the distance $|\mathbf{P}\hat{\boldsymbol{\mu}} - \mathbf{q}|$ can be calculated for both models. The following graphs have as 2 of the axes the 2 diagonal entries in $\mathbf{\Omega}$ and the third axis represents the Euclidean distance $\|\mathbf{P}(\mathbf{P}^{*-1}\hat{\boldsymbol{\mu}}^*) - \mathbf{q}\|_2$:

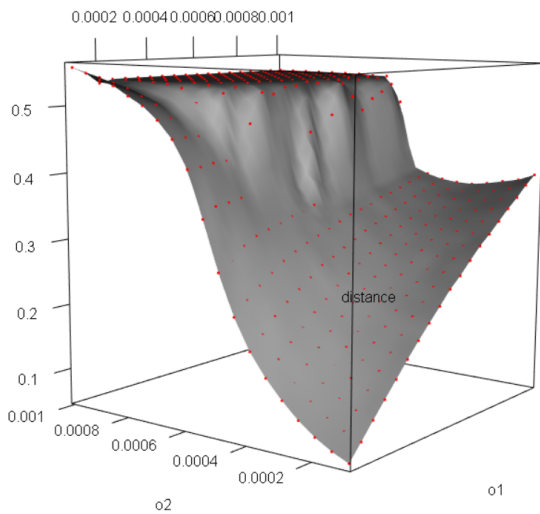


Figure 1. Results of $\mathbf{\Omega}$ for the extension model

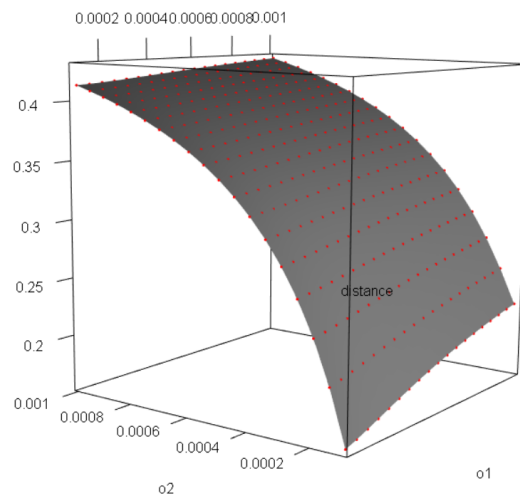


Figure 2. Results of $\mathbf{\Omega}$ for original model

Both Figure 1 and Figure 2 confirm our intuition on how the models should work in the real world. As ω_1 (confidence in the first view labeled as $o1$ in the figure) and ω_2 (confidence in the second view labeled as $o2$ in the figure) decrease, the distance gets closer and closer to 0. As $o1$ and $o2$ increase, the distance seems to plateau to a certain value (the history).

3. A Bayesian Approach, When Historical Data is Unavailable

3.1 Introduction

A very interesting idea for a different prior on the covariance matrix is presented by Leonard & Hsu [1992] and by Albert et al. [2000]. Again, suppose that $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n$ are the n independent returns, where each \mathbf{r}_i follows a normal distribution specified in equation (7).

$$\mathbf{r}_i \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \text{ for } i = 1, 2, \dots, n$$

The likelihood function for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n) = (2\pi)^{-\frac{nm}{2}} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (\mathbf{r}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{r}_i - \boldsymbol{\mu}) \right\}$$

Let $\mathbf{A} = \log(\mathbf{\Sigma}) = (a_{ij})_{i,j=\{1,2,\dots,m\}}$ and $\mathbf{S} = \frac{1}{n} \sum_{i=1}^n (\mathbf{r}_i - \boldsymbol{\mu})(\mathbf{r}_i - \boldsymbol{\mu})^T$, the likelihood of $\boldsymbol{\mu}$ and \mathbf{A} can be represented as

$$L(\boldsymbol{\mu}, \mathbf{A} | \mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n) = (2\pi)^{-\frac{nm}{2}} \exp \left\{ -\frac{n}{2} Tr(\mathbf{A} + \mathbf{S}e^{-\mathbf{A}}) \right\} \tag{24}$$

Here we define the operator $Vec^*(.)$ that stacks in a vector the entries of a matrix parallel to the main diagonal. For example,

$$\boldsymbol{\alpha} = Vec^*(\mathbf{A}) = [a_{11} \ a_{22} \ \dots \ a_{mm} \ a_{12} \ a_{23} \ \dots \ a_{n-1n} \ \dots \ a_{1n}]^T$$

Leonard & Hsu [1992] approximated the likelihood of $\boldsymbol{\mu}$ and \mathbf{A} in equation (24) using Bellman’s iterative solution (please see Bellman [1997]) to the linear Volterra integral equation, and reported the corresponding approximated likelihood function for $\boldsymbol{\alpha} = Vec^*(\mathbf{A}) = Vec^*(\log(\mathbf{\Sigma}))$ and $\boldsymbol{\mu}$:

$$L^*(\boldsymbol{\mu}, \boldsymbol{\alpha} | \mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n) = (2\pi e)^{-\frac{nm}{2}} |\mathbf{S}|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} (\boldsymbol{\alpha} - \boldsymbol{\lambda})^T \mathbf{Q} (\boldsymbol{\alpha} - \boldsymbol{\lambda}) \right\}. \tag{25}$$

Here, $\boldsymbol{\lambda} = Vec^*(\log(\mathbf{S}))$ and the $(d \times d)$ symmetric almost surely positive definite matrix \mathbf{Q} is the likelihood information matrix of $\boldsymbol{\alpha}$ and is a function of the eigenvalues and normalized eigenvectors of \mathbf{S} , where $d = \frac{1}{2}m(m + 1)$. In particular,

$$\mathbf{Q} = \frac{n}{2} \sum_{i=1}^m \mathbf{f}_i \mathbf{f}_i^T + n \sum_{i=1}^m \sum_{j=1}^m \xi_{ij} \mathbf{f}_i \mathbf{f}_j^T \tag{26}$$

where $\xi_{ij} = \frac{(d_i - d_j)^2}{d_i d_j [\log(d_i) - \log(d_j)]^2}$, d_i and d_j are the i^{th} and j^{th} eigenvalues of \mathbf{S} for $i, j = \{1, \dots, m\}$ and the vectors \mathbf{f}_{ij} satisfy the condition $\boldsymbol{\alpha}^T \mathbf{f}_{ij} = \mathbf{e}_i^T \mathbf{A} \mathbf{e}_j$, \mathbf{e}_i and \mathbf{e}_j are the normalized eigenvectors of \mathbf{S} that correspond to the i^{th} eigenvalue d_i and the j^{th} eigenvalue d_j .

We notice that the approximate likelihood function (25) has a multivariate Normal form with respect to $\boldsymbol{\alpha}$. Specifically, the approximate likelihood function for $\boldsymbol{\alpha}$ is a $d = \frac{1}{2}m(m + 1)$ dimensional multivariate Normal distribution with mean vector equal to $\boldsymbol{\lambda}$ and covariance matrix equal to \mathbf{Q}^{-1} . This functional form of the approximate likelihood function in equation (25) will be the driving mechanism in the Bayesian analysis for $\boldsymbol{\alpha}$.

3.2 The Model

In the case when the historical information is not available and we do not have substantial prior information about the covariance $\mathbf{\Sigma}$, we consider a vague prior for

$$\boldsymbol{\alpha} = Vec^*(\log(\mathbf{\Sigma})) = (\alpha_1, \alpha_2, \dots, \alpha_m, \alpha_{m+1}, \dots, \alpha_d)^T, \text{ where } d = \frac{1}{2}m(m + 1)$$

The prior is specified in two stages:

- In the first stage, given $\theta_1, \sigma_1^2, \theta_2$, and σ_2^2 , the diagonal elements of $\mathbf{A} = \log(\mathbf{\Sigma})$, $\alpha_1, \alpha_2, \dots, \alpha_m$ which correspond to the variance components of the covariance matrix $\mathbf{\Sigma}$ and each component α_i follows an independent normal distribution with a common mean θ_1 and a common variance σ_1^2 . The off-diagonal elements of $\mathbf{A} = \log(\mathbf{\Sigma})$, $\alpha_{m+1}, \alpha_{m+2}, \dots, \alpha_d$, which correspond to the covariance components of the covariance matrix $\mathbf{\Sigma}$, each follow an independent normal distribution with a common mean θ_2 and a common variance σ_2^2 and those diagonal elements and off-diagonal elements are independent to each other.
- In the second stage, we assume independent diffuse priors for θ_1, θ_2 and for σ_1^2, σ_2^2 , respectively. In addition to our priors for $\boldsymbol{\alpha}$, we include our investors views for consideration.

We summarize our model when the historical information is not available as follows:

$$\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n | \boldsymbol{\mu}, \mathbf{\Sigma} \stackrel{iid.}{\sim} N(\boldsymbol{\mu}, \mathbf{\Sigma}) \tag{27}$$

$$\mathbf{P}\boldsymbol{\mu} \sim N(\mathbf{q}, \mathbf{\Omega}) \tag{28}$$

$$\boldsymbol{\alpha} | \boldsymbol{\theta}, \Delta \sim N(\mathbf{J}\boldsymbol{\theta}, \Delta) \tag{29}$$

$$\pi(\boldsymbol{\theta}) \propto 1$$

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

where

$$\mathbf{J} = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix}, \quad \Delta = \begin{bmatrix} \sigma_1^2 \mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & \sigma_2^2 \mathbf{I}_{d-m} \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$$

The parameters θ_1 and θ_2 can be integrated out from the joint density of $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ given σ_1^2 and σ_2^2 :

$$\begin{aligned} \pi(\boldsymbol{\alpha} | \sigma_1^2, \sigma_2^2) &\propto \int_{\boldsymbol{\theta}} \pi(\boldsymbol{\alpha} | \boldsymbol{\theta}, \Delta) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &\propto \int_{\boldsymbol{\theta}} |\Delta|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\boldsymbol{\alpha} - \mathbf{J}\boldsymbol{\theta})^T \Delta^{-1} (\boldsymbol{\alpha} - \mathbf{J}\boldsymbol{\theta}) \right\} d\boldsymbol{\theta} \\ &= |\Delta|^{-\frac{1}{2}} |\mathbf{J}^T \Delta^{-1} \mathbf{J}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \boldsymbol{\alpha}^T \mathbf{G} \boldsymbol{\alpha} \right\}, \end{aligned} \tag{30}$$

where

$$\mathbf{G} = (\mathbf{I}_d - \mathbf{J}(\mathbf{J}^T \Delta^{-1} \mathbf{J})^{-1} \mathbf{J}^T \Delta^{-1})^T \Delta^{-1} (\mathbf{I}_d - \mathbf{J}(\mathbf{J}^T \Delta^{-1} \mathbf{J})^{-1} \mathbf{J}^T \Delta^{-1}) \tag{31}$$

Combining the approximated likelihood in equation (25), the prior for the covariance matrix in equation (30) and the investors views in equation (28), we obtain that the approximate joint distribution is:

$$\begin{aligned} \pi^*(\boldsymbol{\alpha}, \boldsymbol{\mu}, \sigma_1^2, \sigma_2^2, \mathbf{r}_1, \dots, \mathbf{r}_n) &\propto |\Delta|^{-\frac{1}{2}} |\mathbf{J}^T \Delta^{-1} \mathbf{J}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \boldsymbol{\alpha}^T \mathbf{G} \boldsymbol{\alpha} \right\} \\ &\quad \times |\mathbf{S}|^{-\frac{m}{2}} \exp \left\{ -\frac{1}{2} (\boldsymbol{\alpha} - \boldsymbol{\lambda})^T \mathbf{Q} (\boldsymbol{\alpha} - \boldsymbol{\lambda}) \right\} \\ &\quad \times |\boldsymbol{\Omega}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{P}\boldsymbol{\mu} - \mathbf{q})^T \boldsymbol{\Omega}^{-1} (\mathbf{P}\boldsymbol{\mu} - \mathbf{q}) \right\} \end{aligned} \tag{32}$$

The approximated conditional posterior density of $\boldsymbol{\alpha}$ given σ_1^2, σ_2^2 and $\boldsymbol{\mu}$ is:

$$\pi^*(\boldsymbol{\alpha} | \mathbf{r}_1, \dots, \mathbf{r}_n, \sigma_1^2, \sigma_2^2, \boldsymbol{\mu}) \propto \exp \left\{ -\frac{1}{2} (\boldsymbol{\alpha}^T \mathbf{G} \boldsymbol{\alpha} + (\boldsymbol{\alpha} - \boldsymbol{\lambda})^T \mathbf{Q} (\boldsymbol{\alpha} - \boldsymbol{\lambda})) \right\}$$

By completing the square, we can combine the two quadratics in $\boldsymbol{\alpha}$ from the exponential and we obtain the posterior:

$$\pi^*(\boldsymbol{\alpha} | \mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n, \sigma_1^2, \sigma_2^2, \boldsymbol{\mu}) \propto \exp \left\{ -\frac{1}{2} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)^T (\mathbf{Q} + \mathbf{G}) (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) \right\}, \tag{33}$$

where $\boldsymbol{\alpha}^* = (\mathbf{Q} + \mathbf{G})^{-1} \mathbf{Q}\boldsymbol{\lambda}$. That is, $\boldsymbol{\alpha}$ is approximately normally distributed with mean $\boldsymbol{\alpha}^*$ and covariance matrix $(\mathbf{Q} + \mathbf{G})^{-1}$, where \mathbf{Q} and \mathbf{G} are specified in equations (26) and (31), respectively.

Moving our attention to the posterior of σ_1^2 and σ_2^2 , we have to pay attention and collect the terms that depend on Δ and \mathbf{G} , since the latter also depends on Δ :

$$\pi(\sigma_1^2, \sigma_2^2 | \boldsymbol{\alpha}, \boldsymbol{\mu}, \mathbf{r}_1, \dots, \mathbf{r}_n) \propto |\Delta|^{-\frac{1}{2}} |\mathbf{J}^T \Delta^{-1} \mathbf{J}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \boldsymbol{\alpha}^T \mathbf{G} \boldsymbol{\alpha} \right\}$$

Written in scalar form, this distribution is:

$$\begin{aligned} \pi(\sigma_1^2, \sigma_2^2 | \boldsymbol{\alpha}, \boldsymbol{\mu}, \mathbf{r}_1, \dots, \mathbf{r}_n) &\propto (\sigma_1^2)^{-\frac{1}{2}(m-1)} \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^m (\alpha_i - \bar{\alpha}_v)^2 \right\} \\ &\quad \times (\sigma_2^2)^{-\frac{1}{2}(d-m-1)} \exp \left\{ -\frac{1}{2\sigma_2^2} \sum_{i=m+1}^d (\alpha_i - \bar{\alpha}_c)^2 \right\} \end{aligned}$$

That is, given α and μ , the quantities σ_1^2 and σ_2^2 are a posteriori independent. Furthermore, σ_1^2 has an Inverse Gamma distribution with shape $\frac{1}{2}(m - 3)$ and scale $\frac{1}{2} \sum_{i=1}^m (\alpha_i - \bar{\alpha}_v)^2$. Similarly, σ_2^2 has an Inverse Gamma distribution with shape $\frac{1}{2}(d - m - 3)$ and scale $\frac{1}{2} \sum_{i=m+1}^d (\alpha_i - \bar{\alpha}_c)^2$, where $\bar{\alpha}_v = \frac{1}{m} \sum_{i=1}^m \alpha_i$ and $\bar{\alpha}_c = \frac{1}{d-m} \sum_{i=m+1}^d \alpha_i$.

The conditional posterior of μ given $\alpha, \sigma_1^2, \sigma_2^2$ is proportional to the product of the likelihood function of μ and Σ in equation (27) and the prior density that represents the investor's views in equation (28):

$$\begin{aligned} \pi(\mu|\alpha, \sigma_1^2, \sigma_2^2, \mathbf{r}_1, \dots, \mathbf{r}_n) &\propto \exp\left\{-\frac{1}{2} \sum_{i=1}^n (\mathbf{r}_i - \mu)^T \Sigma^{-1} (\mathbf{r}_i - \mu)\right\} \\ &\times \exp\left\{-\frac{1}{2} (\mathbf{P}\mu - \mathbf{q})^T \Omega^{-1} (\mathbf{P}\mu - \mathbf{q})\right\} \propto \exp\left\{-\frac{n}{2} (\bar{\mathbf{r}} - \mu)^T \Sigma^{-1} (\bar{\mathbf{r}} - \mu)\right\} \\ &\quad \times \exp\left\{-\frac{1}{2} (\mathbf{P}\mu - \mathbf{q})^T \Omega^{-1} (\mathbf{P}\mu - \mathbf{q})\right\} \\ &\propto \exp\left\{-\frac{n}{2} (\mu^T \Sigma^{-1} \mu - 2\bar{\mathbf{r}}^T \Sigma^{-1} \mu + \bar{\mathbf{r}}^T \Sigma^{-1} \bar{\mathbf{r}})\right\} \\ &\quad \times \exp\left\{-\frac{1}{2} (\mu^T \mathbf{P}^T \Omega^{-1} \mathbf{P}\mu - 2\mathbf{q}^T \Omega^{-1} \mathbf{P}\mu + \mathbf{q}^T \Omega^{-1} \mathbf{q})\right\} \\ \pi(\mu|\alpha, \sigma_1^2, \sigma_2^2, \mathbf{r}_1, \dots, \mathbf{r}_n) &\propto \exp\left\{-\frac{1}{2} (\mu - \mu_{\text{post}})^T \Sigma_{\text{post}} (\mu - \mu_{\text{post}})\right\}, \text{ where} \\ \Sigma_{\text{post}} &= (n\Sigma^{-1} + \mathbf{P}^T \Omega^{-1} \mathbf{P})^{-1} \\ \mu_{\text{post}} &= \Sigma_{\text{post}} (n\Sigma^{-1} \bar{\mathbf{r}} + \mathbf{P}^T \Omega^{-1} \mathbf{q}) \end{aligned}$$

That is the exact conditional posterior distribution of μ given α, σ_1^2 and σ_2^2 is an m dimensional normal distribution with mean μ_{post} and covariance matrix Σ_{post} .

3.3 Implementation

Now that we have derived our posteriors, we are ready to implement them using a Gibbs Sampler. A Metropolis-Hastings algorithm will be needed for sampling α , for which we require the exact posterior distribution as well as an approximated posterior distribution from which a random sample can be taken. We consider the approximate posterior for α , given σ_1^2, σ_2^2 and μ in equation (33):

$$\pi^*(\alpha|\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n, \sigma_1^2, \sigma_2^2, \mu) \propto \exp\left\{-\frac{1}{2} (\alpha - \alpha^*)^T (\mathbf{Q} + \mathbf{G}) (\alpha - \alpha^*)\right\},$$

The exact posterior density is proportional to the product of the likelihood function specified in equation (24) and the prior in equation (30):

$$\pi(\alpha|\mathbf{r}_1, \dots, \mathbf{r}_n, \sigma_1^2, \sigma_2^2, \mu) \propto \exp\left\{-\frac{n}{2} \text{Tr}(\mathbf{A} + \mathbf{S}e^{-\mathbf{A}}) - \frac{1}{2} \alpha^T \mathbf{G} \alpha\right\}$$

The Metropolis-Hastings step at t^{th} iteration in the Gibbs Sampler would involve simulating a candidate value from the approximate posterior distribution: $\tilde{\alpha} \propto N(\alpha^*, (\mathbf{Q} + \mathbf{G})^{-1})$ and we would accept it with probability $\min(\rho, 1)$, where

$$\rho = \frac{\pi(\tilde{\alpha}|\mathbf{r}_1, \dots, \mathbf{r}_n, \sigma_1^{2(t)}, \sigma_2^{2(t)}, \mu^{(t)})}{\pi(\alpha^{(t)}|\mathbf{r}_1, \dots, \mathbf{r}_n, \sigma_1^{2(t)}, \sigma_2^{2(t)}, \mu^{(t)})} \cdot \frac{\pi^*(\alpha^{(t)}|\mathbf{r}_1, \dots, \mathbf{r}_n, \sigma_1^{2(t)}, \sigma_2^{2(t)}, \mu^{(t)})}{\pi^*(\tilde{\alpha}|\mathbf{r}_1, \dots, \mathbf{r}_n, \sigma_1^{2(t)}, \sigma_2^{2(t)}, \mu^{(t)})}$$

The great advantage of the Leonard-Hsu prior is that we can actually sample a covariance matrix starting from just 2 initial values: σ_1^2 and σ_2^2 . Those 2 values give us \mathbf{A} and \mathbf{G} (using equation (31)) and together with \mathbf{Q} , which actually just depends on the sample covariance matrix, we can start the Gibbs-Sampler with the Metropolis-Hastings step for sampling α . Therefore, we arrive at the following Gibbs Sampler:

Algorithm 2 Gibbs Sampler $\log(\Sigma)$

- 1: $\alpha^{(t+1)} = \begin{cases} \bar{\alpha} \sim N\left(\left(\mathbf{Q}^{(t)} + \mathbf{G}^{(t)}\right)^{-1} \mathbf{Q}^{(t)} \lambda^{(t)}, \left(\mathbf{Q}^{(t)} + \mathbf{G}^{(t)}\right)^{-1}\right) \text{w.p. } \min(\rho, 1) \\ \alpha^{(t)} \text{ otherwise} \end{cases}$
- 2: Since $\alpha = \text{Vec}^*(\log(\Sigma)) \Rightarrow \begin{cases} \text{compute } \Sigma^{(t+1)} = \exp\{\text{Vec}^{*-1}(\alpha^{(t+1)})\} \\ \text{keep } \Sigma^{(t)} \end{cases}$
- 3: $\begin{cases} \sigma_1^{2(t+1)} \sim IG\left(\frac{m-3}{2}, \frac{1}{2} \sum_{i=1}^m (\alpha_i^{(t+1)} - \bar{\alpha}_v^{(t+1)})^2\right) \\ \sigma_2^{2(t+1)} \sim IG\left(\frac{d-m-3}{2}, \frac{1}{2} \sum_{i=m+1}^d (\alpha_i^{(t+1)} - \bar{\alpha}_c^{(t+1)})^2\right) \end{cases} \Rightarrow$
 $\Rightarrow \Delta^{(t+1)} = \begin{bmatrix} \sigma_1^{2(t+1)} \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \sigma_2^{2(t+1)} \mathbf{I}_{d-n} \end{bmatrix}$
- 4: Let $\Sigma\mu = \left(n\Sigma^{(t+1)-1} + \mathbf{P}^T \Omega^{-1} \mathbf{P}\right)^{-1}, \mu^{(t+1)} \sim N\left(\Sigma\mu \left(n\Sigma^{(t+1)-1} \bar{\mathbf{r}} + \mathbf{P}^T \Omega^{-1} \mathbf{q}\right), \Sigma\mu\right)$
- 5: Compute $\mathbf{S}^{(t+1)} = \frac{\sum_{i=1}^n (\mathbf{r}_i - \mu^{(t+1)}) (\mathbf{r}_i - \mu^{(t+1)})^T}{n}, \lambda^{(t+1)} = \text{Vec}^*\left(\log\left(\mathbf{S}^{(t+1)}\right)\right), d_j^{(t+1)}$ and $\mathbf{e}_j^{(t+1)}$ the eigenvalue and normalized eigenvector of $\mathbf{S}^{(t+1)}$ respectively.
- 6: Compute $\xi_{ij}^{(t+1)}$ by identifying the coefficients of the entries of the $\log\left(\Sigma^{(t+1)}\right)$ matrix from the equation $\text{Vec}^*\left(\log\left(\Sigma^{(t+1)}\right)\right)^T \mathbf{f}_{ij}^{(t+1)} = \mathbf{e}_i^{(t+1)T} \log\left(\Sigma^{(t+1)}\right) \mathbf{e}_j^{(t+1)}$
- 7: Compute $\xi_{ij}^{(t+1)} = \frac{d_i^{(t+1)} d_j^{(t+1)}}{d_i^{(t+1)} d_j^{(t+1)} (\log[d_i^{(t+1)}] - \log[d_j^{(t+1)}])^2}$
- 8: Compute $\mathbf{Q}^{(t+1)} = \frac{n}{2} \sum_{i=1}^m \mathbf{f}_i^{(t+1)} \mathbf{f}_i^{(t+1)T} + n \sum_{i=1}^m \sum_{j=1}^m \xi_{ij}^{(t+1)} \mathbf{f}_j^{(t+1)} \mathbf{f}_j^{(t+1)T}$
- 9: Compute

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

3.4 Results

Just like before, we would like to conduct sensitivity analysis for the confidence in the investor views (the diagonal elements of the Ω 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 the same 4 stocks (AAPL, AMZN, GOOG, MSFT), but since this work was done slightly more recently, the daily returns from January 2nd 2014 to December 29th 2017 were considered. The data can be obtained using the *quantmod* package in R. We will use the following inputs (again the columns are in order AAPL, AMZN, GOOG, MSFT and the rows represent the views):

$$\mathbf{q} = \begin{bmatrix} 0.02 \\ 0.05 \end{bmatrix}, \mathbf{P} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

This approach is more computationally expensive than just having an Inverse Wishart prior on Σ . Therefore, the sensitivity analysis was ran in parallel on multiple cores (each core running the Gibbs Sampler for 1 pair (ω_1, ω_2)) and the range itself was split into 4 ranges:

- (1) $\omega_1 \in \{10^{-6}, 4 \cdot 10^{-6}, 7 \cdot 10^{-6}, 10^{-5}\}$ and $\omega_2 \in \{10^{-6}, 4 \cdot 10^{-6}, 7 \cdot 10^{-6}, 10^{-5}\}$
- (2) $\omega_1 \in \{10^{-5}, 4 \cdot 10^{-5}, 7 \cdot 10^{-5}, 10^{-4}\}$ and $\omega_2 \in \{10^{-5}, 4 \cdot 10^{-5}, 7 \cdot 10^{-5}, 10^{-4}\}$
- (3) $\omega_1 \in \{10^{-6}, 4 \cdot 10^{-6}, 7 \cdot 10^{-6}, 10^{-5}\}$ and $\omega_2 \in \{10^{-5}, 4 \cdot 10^{-5}, 7 \cdot 10^{-5}, 10^{-4}\}$
- (4) $\omega_1 \in \{10^{-5}, 4 \cdot 10^{-5}, 7 \cdot 10^{-5}, 10^{-4}\}$ and $\omega_2 \in \{10^{-6}, 4 \cdot 10^{-6}, 7 \cdot 10^{-6}, 10^{-5}\}$

The burn period was set to 10^3 and the iterations to 10^4 . Albeit those seem relatively small, convergence is actually achieved very fast when ω_i are small.

If we look at Figure 3, we notice that in this version of the model, the distance converges to 0 very fast as o1 (ω_1 in the model) and o2 (ω_2 in the model) go to 0. Also, we notice that as o1 and o2 get bigger, it converges very fast to a stabilizing distance. This is consistent with our intuition since if we are very confident in our views, the model should put a lot more importance on them, while if we are not confident at all in our views, the model should just take into consideration the history. Indeed, if we use only the history, the unbiased estimator for μ is the sample mean of the returns ($\bar{\mathbf{r}}$) and therefore the distance becomes $|\mathbf{P}\bar{\mathbf{r}} - \mathbf{q}| = 0.05388875$.

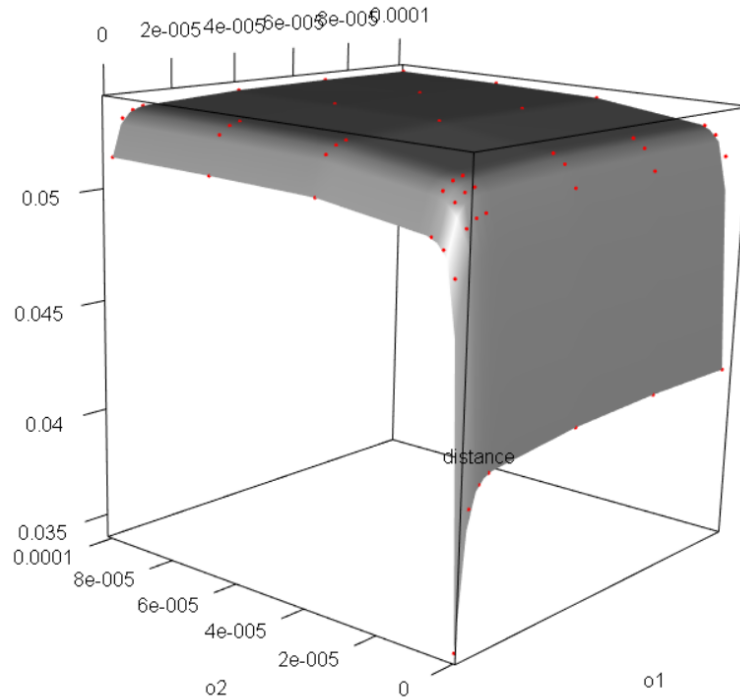


Figure 3. Distances for $\log(\Sigma)$ prior

3.5 Portfolio Performance in January 2018

We will proceed by looking at the performance over the month of January 2018 of a portfolio obtained using this model trained on the same daily returns between January 2nd 2014 and December 29th 2017. For those types of analyses, an initial investment of \$100 000 was considered, without any commissions, capital buffers for short selling, etc. Just as we did before, in order to obtain the portfolio, we would estimate using Gibbs Sampling the posterior mean ($\hat{\mu}$) and the posterior covariance ($\hat{\Sigma}$). The portfolio weights which maximize posterior portfolio mean while minimizing posterior variance (risk) are $\mathbf{w} = \frac{1}{2.5} \hat{\Sigma}^{-1} \hat{\mu}$. Using those weights we compute the profits or loses that we would obtain over the month of January 2018 (daily returns between January 2nd 2018 and January 30th 2018) with an initial investment of \$100 000. Here, one could use a different investment horizon also.

The same \mathbf{P} , \mathbf{q} , grid for ω_i , burn period, iteration period were used as before. Figure 4 is a 3D plot of the sensitivity of the profits to changes in investor’s confidence. We observe a profit that is approximately between \$10 000 and \$58 000. In order to interpret this curve, we would have to know what actually happened in the month of January 2018 using the views inputted. More specifically, over the month of January 2018, $\mathbf{Pr}_{\text{Jan2018}} = \begin{bmatrix} 0.23996743 \\ 0.01366718 \end{bmatrix}$. Albeit the 1st view inputted is a 10th of what happened in reality (AMZN outperformed AAPL by almost 24% in January 2018), the profits curve still gives a higher importance on this view than on the 2nd view. Indeed, profits increase drastically as we decrease ω_1 and keep ω_2 constant.

A 24% gain on AAPL in a month is an extreme scenario, let us consider a different stock instead of AMZN. We will replace AMZN with FB (Facebook) and we will keep all the inputs the same as before, except that we will input 3 different values for \mathbf{q} . In Figure 5 we will present the results for profits when the investor considers $\mathbf{q} = \begin{bmatrix} 0.02 \\ 0.05 \end{bmatrix}$ (a random guess), in Figure 6 the “well informed” investor (exactly what happened during the month of January 2018) $\mathbf{q} = \begin{bmatrix} 0.06212815 \\ 0.01366718 \end{bmatrix}$ and in Figure 7 the “poorly informed” investor (exactly the opposite of what happened during the month of January 2018) $\mathbf{q} = \begin{bmatrix} -0.06212815 \\ -0.01366718 \end{bmatrix}$:

- Since $\mathbf{Pr}_{\text{Jan2018}} = \begin{bmatrix} 0.06212815 \\ 0.01366718 \end{bmatrix}$, the view in which $\mathbf{q} = \begin{bmatrix} 0.02 \\ 0.05 \end{bmatrix}$ has returns that are much closer to what happened

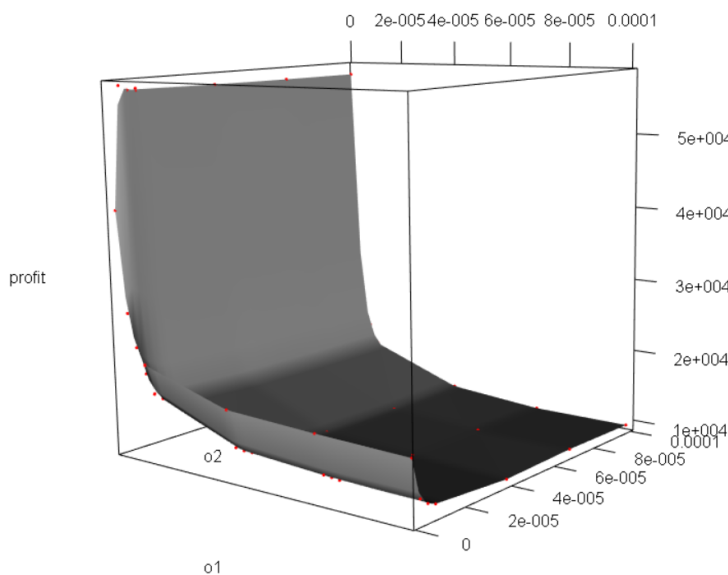


Figure 4. Profits with AMZN in and $\mathbf{q} = [0.02, 0.05]^T$

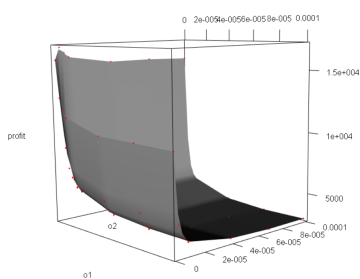


Figure 5. Profits FB instead of AMZN and $\mathbf{q} = [0.02, 0.05]^T$

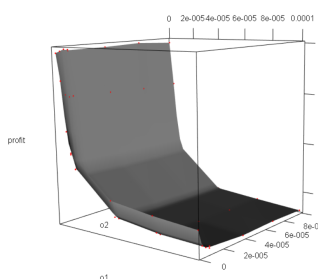


Figure 6. Profits FB instead of AMZN and view exactly like reality

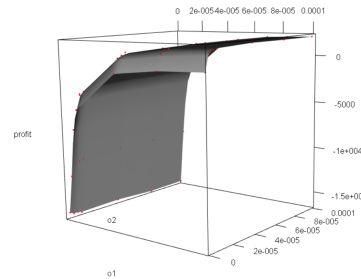


Figure 7. Profits FB instead of AMZN and view opposite of reality

in reality than when we considered AMZN instead of FB (first view in reality was 24% with AMZN in). This is reflected in Figure 5, where we notice that now the second view has a greater influence on the profits curve than what we have seen in Figure 4.

- If we compare Figures 6 and 7, we notice that they seem to be a reflection of each other with respect to a plane parallel to the "o1 vs o2" plane. This would make sense since the only difference between the two is that in Figure 6 we have $\mathbf{q} = \begin{bmatrix} 0.06212815 \\ 0.01366718 \end{bmatrix}$ (exactly reality) and in Figure 7 we have $\mathbf{q} = -\begin{bmatrix} 0.06212815 \\ 0.01366718 \end{bmatrix}$ (opposite of reality).

4. Conclusion

We have seen that our model follows our intuition: the more confident the investor is in their views, the closer the model will follow them and the less confident the investor is in their views, the more the model will follow history. Moreover, in the version containing the Leonard-Hsu prior, the profit curve when the investor is lucky and inputs views exactly as they will happen (please see Figure 6) is a mirror image of the one when they input the opposite as what will happen (please see Figure 7).

In our next paper we will introduce another full Bayesian version of Black-Litterman, but this time for large data-sets. The motivation to move to a large number of variables is driven by the fact that the investor might want to use the data for the whole market ($S\&P500$), despite the fact that they might have very few views (for example, 2 as we presented in this paper). Our next version will introduce a Bayesian factor model in order to reduce the dimension. This is done because the matrix \mathbf{Q} , defined in equation (26), is of size $d \times d = \frac{m(m+1)}{2} \times \frac{m(m+1)}{2}$ and is randomly generated at each iteration in a Gibbs Sampler. Therefore, if one considers the whole $S\&P500$, the size of this matrix in terms of memory would be of

around 106GB. Hence, after applying factor models, we will introduce a prior on the covariance matrix of the common factors instead of introducing a prior directly on the covariance matrix of the returns. The dimension of the covariance matrix of the common factors is $q \times q$ (q =number of factors), which is much smaller than $m \times m$ (m =number of stocks), the dimension of the covariance matrix of the returns.

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.
- Bellman, R. (1997). *Introduction to matrix analysis, second edition*. Society for Industrial and Applied Mathematics.
- He, G., & Litterman, R. (2002). The intuition behind black-litterman model portfolios. *SSRN Electronic Journal*.
- Janecek, K. (2004). What is a realistic aversion to risk for real-world individual investors ?.
- Leonard, T., & Hsu, J. S. J. (1992, Dec). Bayesian inference for a covariance matrix. *The Annals of Statistics*, 20(4), 1669-1696.

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/>).