

Bayesian Alternatives to the Black-Litterman Model

Mihnea S. Andrei¹ and John S.J. Hsu¹

¹Statistics and Applied Probability, University of California, Santa Barbara

¹mandrei@ucsb.edu, hsu@pstat.ucsb.edu

November 27, 2018

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 adding a typical Inverse Wishart prior on the covariance matrix of the returns (section 2). Lastly, we will use Leonard and Hsu's (1992)[5] idea of adding a prior on the logarithm of the covariance matrix (section 3). Sensitivity simulations for the level of confidence that the investor has in their own personal views were performed and performance of the models was assessed on a test data set consisting of returns over the month of January 2018.

1 The Original Black-Litterman Model

The Black-Litterman model was developed in the early 1990's and has been widely used for asset allocation. This model attempts to combine the market equilibrium ¹ with the investor's personal views. It can be shown that the optimal portfolio for an investor is the sum of a portfolio proportional to the market equilibrium portfolio and a weighted sum of portfolios reflecting their views. Please see section 1.2 for an example of how personal views are created and section 1.3 for more details on the model.

¹Please refer to subsection 1.1 for more details on how the market equilibrium is computed

1.1 Estimating the Market Equilibrium

The market is in equilibrium when all investors hold the market portfolio, w_{eq} . It is when the demand for the assets in this portfolio equals the supply. If we denote by π the market equilibrium returns, then the CAPM equation is $\pi = \lambda \Sigma w_{eq}$. Here, λ is the investor's risk aversion coefficient and Σ is the covariance matrix of the returns on the assets in the portfolio[3].

1.2 Example of Personal Views

Let us see how personal views are inputted in the traditional model. For example, let us consider 4 assets: AAPL, AMZN, GOOG and MSFT. Also, let us consider that we believe that AAPL will outperform AMZN by 2% and we also believe that GOOG will have returns that amount to 5%. The columns in P represent the 4 stocks in the order in which we enumerated them previously. Each row in P and q represents a personal view:

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

Each personal view has associated with it an uncertainty that the investor has with respect to the view. The measures of confidence are entered as diagonal entries in a matrix Ω . As we will see in the next section when we present the assumptions of the model (equations (1)), Ω is a covariance matrix. Hence, on the main diagonal we will have the variances of the returns for our personal views. Therefore, a small value reflects a high confidence in the view and vice versa (one approach of obtaining Ω is by simply manually imputing them and another approach is to estimate it).

1.3 The Black-Litterman Approach

Now that we have seen what are the individual pieces of the model, we are ready to present the mathematical formulation. We will consider that the investor is looking at n assets and has k different views on those assets.

The return of the assets is considered to be random, $r \sim N(\mu, \Sigma)$.

We create a prior on the mean of this return: $\mu \sim N(\pi, \tau \Sigma)$. However, the proof would work in the exact same way for any general covariance matrix (please see Appendix A for the proof). The variable π represents the market equilibrium returns and it is obtained by using an equation equivalent to the CAPM (4): $\pi = \lambda \Sigma w_{eq}$, with λ been the investor's risk aversion parameter, w_{eq} been the market equilibrium weights and Σ the covariance matrix. Also, here τ is considered to be a parameter that reflects

the uncertainty in the CAPM prior. Notice that the smaller the τ , the closer our μ will be to the market equilibrium returns π .

Besides this prior, we also have the investor's personal views: $P\mu \sim N(q, \Omega)$, where the same notation as in the previous section 1.2 is used [3].

Hence, the model is represented by the following 3 distributions, where the last 2 are priors:

$$\begin{aligned} r &\sim N(\mu, \Sigma) \\ \mu &\sim N(\pi, \tau\Sigma) \\ P\mu &\sim N(q, \Omega) \end{aligned} \tag{1}$$

By combining the 2 priors (the last 2 distributions from above) and using that as a new prior on the distribution of the return, one can show that the posterior of the return is (please see the appendix A for the proof):

$$r \sim N(\bar{\mu}, M^{-1} + \Sigma)$$

Where the notation used is,

$$M^{-1} = ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1} \tag{2}$$

$$\bar{\mu} = ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1} ((\tau\Sigma)^{-1}\pi - P^T\Omega^{-1}q) \tag{3}$$

Also, let $\bar{\Sigma} = M^{-1} + \Sigma$

Coming back to the example of personal views given in the previous section, if the investor thinks that GOOG will outperform the other 3 companies, it is enough to put a long weight (positive weight) of 1 on GOOG and short weights (negative weights) for the other 3. In the Black-Litterman model, the return is considered to be random and we have just seen that the posterior distribution is also normal, but with a mean represented by equation (3). This equation appropriately takes into account market volatility and correlations also. Let us further look at the weights, w , that one would obtain when using the posterior of the returns. The typical approach to the problem that an investor with risk aversion parameter λ has when trying to maximize the returns of the portfolio while minimizing the risk is to maximize the function $w^T\bar{\mu} - \frac{\lambda}{2}w^T\bar{\Sigma}w$ with respect to w . By taking the derivative with respect to w , we obtain for the optimal weights an equation equivalent to CAPM:

$$\bar{\mu} = \lambda\bar{\Sigma}w \Leftrightarrow w = \frac{1}{\lambda}\bar{\Sigma}^{-1}\bar{\mu} \tag{4}$$

Using equation (3) and the identity $(\Sigma + M^{-1})^{-1} = M - M(M + \Sigma^{-1})^{-1}M$, one can show that:

$$\begin{aligned} w^* &= \frac{1}{1 + \tau}(w_{eq} + P^T \times \Delta) \\ \Delta &= \tau \Omega^{-1} \frac{q}{\lambda} - A^{-1} P \frac{\Sigma}{1 + \tau} w_{eq} - A^{-1} P \frac{\Sigma}{1 + \tau} P^T \tau \Omega^{-1} \frac{q}{\lambda} \\ A &= \frac{\Omega}{\tau} + P \frac{\Sigma}{1 + \tau} P^T \end{aligned}$$

The first equation from above shows that the Black-Litterman optimal portfolio weights are the equilibrium weights plus a weighted sum of the personal views (P^T has as columns the views). Moreover, the three terms in the second equation have interpretations also:

- The first term shows that the more importance is given to a view either by having a high return q_k (the k^{th} entry of q), or by having a smaller uncertainty ω_k (this is the k^{th} diagonal entry in Ω and hence it will be equal to $\frac{1}{\omega_k}$ in Ω^{-1}), the more weight it will carry into the optimal portfolio
- The second term shows that a weight is penalized if the value of the product between the covariance and the market equilibrium return is high. This would indicate that the view carries less additional information.
- The last term shows that an optimized weight is penalized if the covariance between different portfolio views is high. This makes sense since it would mean that different views add little new information.

One can also observe the fact that if an investor has a different risk aversion parameter, $\hat{\lambda}$, he can obtain the optimized portfolio weights by using the equation $\hat{w}^* = \frac{\lambda}{\hat{\lambda}} w^*$

2 Inverse Wishart prior on Σ

2.1 Introduction

First of all, we have to observe what we would like to potentially improve on the original approach. But what are the shortcomings of the original approach?

- From the proof presented in Appendix A, we notice that it is not a true Bayesian approach since, as mentioned in the previous section, the distributions represented by the last 2 equations in (1) are combined to find a new prior for μ . This is used as a prior for the return.
- The original approach does not consider any current returns and we would like to model returns over a certain period of time (which will be the investor's investment horizon). Hence, we would like to have $r_1, r_2, \dots, r_m \sim N_n(\mu, \Sigma)$
- The covariance matrix Σ is estimated from historical data in the original approach. Hence, we would like to put prior on the distribution of Σ and the most common approach for covariance matrices when the data is from a normal distribution is to use an Inverse Wishart conjugate prior.

Let $r_1, r_2, \dots, r_m \sim N_n(\mu, \Sigma)$ be the observed returns over a period of length m for the n assets that the investor is considering. Also, we would like to consider a prior on the mean of those returns that is projected directly by our personal views P . Therefore, we would have, similarly to the traditional model, $P\mu \sim N(q, \Omega)$. The difference is that we would like to add an additional Inverse Wishart prior $\Sigma \sim W^{-1}(\nu, \Sigma_0)$. Since the returns $r \in \mathbb{R}^n$, while the prior is on $P\mu \in \mathbb{R}^k$ (k is the number of views), the prior is not fully specified, which would suggest the idea that we would need to create an invertible P . Also, once we have an invertible P , we can follow two approaches:

- Obtain the distribution of μ , which could be easily done if P is invertible.
- From the very beginning transform the returns into the personal view space: $r_i^* = Pr_i$. This procedure will still require P to be invertible since after obtaining the posterior in the transformed space, we have to be able to transform back.

Hence, either way, we would need to have a matrix P that is invertible and this brings us to the following discussion.

2.2 Creating an Invertible P

The matrix of our personal views is very likely not to be invertible since, as we have seen in Section 1.2, we can have relative views (the ones for which

the rows sum up to 0) and we can have absolute views (only one 1 in a row). Moreover, the k views that we will have (the number of rows in P) will be smaller than the n assets that we are considering to trade (the number of columns in P). In this section, we will present a method in which we can add rows to P such that the resulting square matrix 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 iff its row reduced echelon form is the identity matrix. This gives us the idea of taking our matrix P and adding rows to it in order to make it invertible:

- For each column in 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 one 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 matrix P from section 1.2, the above procedure gives us:

$$P = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} = P^* = \begin{bmatrix} P \\ P_2 \end{bmatrix}$$

Please notice that we denoted by P^* the augmented invertible matrix based on P , and by P_2 the part that was added to P .

We have seen at the beginning of the section that the model equations are:

$$\begin{aligned} r_1, r_2, \dots, r_m | \mu, \Sigma &\sim N_n(\mu, \Sigma) \\ P\mu &\sim N_k(q, \Omega) \\ \Sigma &\sim W^{-1}(\nu, \Sigma_0) \end{aligned}$$

As mentioned previously, we have 2 approaches. We choose to transform the returns using our personal views and, therefore, let $r_i^* = P^*r_i$. Hence, $r_1^*, r_2^*, \dots, r_m^* \sim N_n(\mu^*, \Sigma^*)$, where $\mu^* = P^*\mu$ and $\Sigma^* = P^*\Sigma P^{*T}$. Now, after the transformation onto the space of personal views, the 3 equations become:

$$\begin{aligned}
r_1^*, r_2^*, \dots, r_m^* &\sim N_n(\mu^*, \Sigma^*) \\
\mu^* &\sim N_n(q^*, \Omega^*) \\
\Sigma^* &\sim W^{-1}(\nu, \Sigma_0)
\end{aligned}$$

However, just like μ^*, Σ^* were clearly depending on μ, Σ , we also have that q^*, Ω^* depend on the originals q, Ω :

$$q^* = E[\mu^*] = E[P^* \mu] = E\left[\begin{bmatrix} P \\ P_2 \end{bmatrix} \mu\right] = \begin{bmatrix} q \\ q_2 \end{bmatrix} \quad (5)$$

$$\Omega^* = Var(P^* \mu) = Var\left(\begin{bmatrix} P \\ P_2 \end{bmatrix} \mu\right) = \begin{bmatrix} \Omega & PVar(\mu)P_2^T \\ P_2Var(\mu)P^T & P_2Var(\mu)P_2^T \end{bmatrix} \quad (6)$$

2.3 Derivation of Posterior Distributions

Now that we have found a method to augment P to a matrix P^* that is invertible and we also managed to create corresponding q^* and Ω^* , the problem is posed in a more typical Bayesian framework:

$$r_1^*, r_2^*, \dots, r_m^* | \mu^*, \Sigma^* \sim N_n(\mu^*, \Sigma^*) \quad (7)$$

$$\mu^* \sim N(q^*, \Omega^*) \quad (8)$$

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

From (7), we obtain that the joint density of our returns is:

$$f(r_1^*, \dots, r_m^* | \mu^*, \Sigma^*) \propto \det(\Sigma^*)^{-\frac{m}{2}} \exp\left\{-\frac{1}{2} \sum_{i=1}^m (r_i^* - \mu^*)^T \Sigma^{*-1} (r_i^* - \mu^*)\right\}$$

From (8), we obtain that the density for μ^* is:

$$f(\mu^*) \propto \det(\Omega^*)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} (\mu^* - q^*)^T \Omega^{*-1} (\mu^* - q^*)\right\}$$

Similarly, using (9), we obtain that the density for Σ^* is:

$$f(\Sigma^*) \propto \det(\Sigma^*)^{-\frac{\nu+n+1}{2}} \exp\left\{-\frac{1}{2} Tr\left(\Sigma_0 \Sigma^{*-1}\right)\right\}$$

Hence, by multiplying the above 3 equations, we obtain that the joint density for all of them is:

$$f(r_1^*, \dots, r_m^*, \mu^*, \Sigma^*) \propto \det(\Sigma^*)^{-\frac{\nu+m+n+1}{2}} \exp \left\{ -\frac{1}{2} \text{Tr} \left(\Sigma_0 \Sigma^{*-1} \right) \right\} \det(\Omega^*)^{-\frac{1}{2}} \cdot \exp \left\{ -\frac{1}{2} \left((\mu^* - q^*)^T \Omega^{*-1} (\mu^* - q^*) + \sum_{i=1}^m (r_i^* - \mu^*)^T \Sigma^{*-1} (r_i^* - \mu^*) \right) \right\} \quad (10)$$

Let us focus on the parenthesis in the second exponential and let us prove the following result.

Lemma 1. *The following equality holds, where $\bar{r}^* = \frac{\sum_{i=1}^m r_i^*}{m}$:*

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

Proof. We will start by manipulating the right hand side:

$$\begin{aligned} RHS &= \sum_{i=1}^m (r_i^{*T} \Sigma^{*-1} r_i^* - r_i^{*T} \Sigma^{*-1} \bar{r}^* - \bar{r}^{*T} \Sigma^{*-1} r_i^* + \bar{r}^{*T} \Sigma^{*-1} \bar{r}^*) + \\ &\quad m \bar{r}^{*T} \Sigma^{*-1} \bar{r}^* - m \bar{r}^{*T} \Sigma^{*-1} \mu^* - m \mu^{*T} \Sigma^{*-1} \bar{r}^* + m \mu^{*T} \Sigma^{*-1} \mu^* \end{aligned}$$

But since $m \bar{r}^* = \sum_{i=1}^m r_i^* \Rightarrow m \bar{r}^{*T} = \sum_{i=1}^m r_i^{*T}$, we obtain that:

$$\begin{aligned} RHS &= \sum_{i=1}^m (r_i^{*T} \Sigma^{*-1} r_i^* - r_i^{*T} \Sigma^{*-1} \bar{r}^* - \bar{r}^{*T} \Sigma^{*-1} r_i^*) + 2m \bar{r}^{*T} \Sigma^{*-1} \bar{r}^* - \\ &\quad - \left(\sum_{i=1}^m r_i^{*T} \right) \Sigma^{*-1} \mu^* - \mu^{*T} \Sigma^{*-1} \left(\sum_{i=1}^m r_i^* \right) + \sum_{i=1}^m \mu^{*T} \Sigma^{*-1} \mu^* = \\ &= \sum_{i=1}^m (r_i^{*T} \Sigma^{*-1} r_i^* - r_i^{*T} \Sigma^{*-1} \bar{r}^* - \bar{r}^{*T} \Sigma^{*-1} r_i^*) + 2m \bar{r}^{*T} \Sigma^{*-1} \bar{r}^* - \\ &\quad - \sum_{i=1}^m r_i^{*T} \Sigma^{*-1} \mu^* - \sum_{i=1}^m \mu^{*T} \Sigma^{*-1} r_i^* + \sum_{i=1}^m \mu^{*T} \Sigma^{*-1} \mu^* \end{aligned}$$

We observe that Σ^{*-1} and \bar{r}^* do not depend on the sum. Hence, we can factor them out:

$$\begin{aligned}
RHS &= \sum_{i=1}^m (r_i^{*T} \Sigma^{*-1} r_i^* - r_i^{*T} \Sigma^{*-1} \mu^* - \mu^{*T} \Sigma^{*-1} r_i^* + \mu^{*T} \Sigma^{*-1} \mu^*) + \\
&\quad + 2m\bar{r}^{*T} \Sigma^{*-1} \bar{r}^* - \left(\sum_{i=1}^m r_i^{*T} \right) \Sigma^{*-1} \bar{r}^* - \bar{r}^{*T} \Sigma^{*-1} \left(\sum_{i=1}^m r_i^* \right) = \\
&= \sum_{i=1}^m (r_i^* - \mu^*)^T \Sigma^{*-1} (r_i^* - \mu^*) + 2m\bar{r}^{*T} \Sigma^{*-1} \bar{r}^* - m\bar{r}^{*T} \Sigma^{*-1} \bar{r}^* - \\
&\quad - m\bar{r}^{*T} \Sigma^{*-1} \bar{r}^* = \sum_{i=1}^m (r_i^* - \mu^*)^T \Sigma^{*-1} (r_i^* - \mu^*)
\end{aligned}$$

□

Let us make a notation before we proceed: $s^2 = \frac{\sum_{i=1}^m (r_i^* - \bar{r}^*)^T \Sigma^{*-1} (r_i^* - \bar{r}^*)}{m-1}$

Now, by using **Lemma 1**, we are ready to come back to the parenthesis in the second exponential from the joint density of $(r_1^*, \dots, r_m^*, \mu^*, \Sigma^*)$ (equation (10)):

$$\begin{aligned}
&(\mu^* - q^*)^T \Omega^{*-1} (\mu^* - q^*) + \sum_{i=1}^m (r_i^* - \mu^*)^T \Sigma^{*-1} (r_i^* - \mu^*) = \\
&= (m-1)s^2 + m(\bar{r}^* - \mu^*)^T \Sigma^{*-1} (\bar{r}^* - \mu^*) + (\mu^* - q^*)^T \Omega^{*-1} (\mu^* - q^*) = \\
&= (m-1)s^2 + (\bar{r}^* - \mu^*)^T (m\Sigma^{*-1}) (\bar{r}^* - \mu^*) + (\mu^* - q^*)^T \Omega^{*-1} (\mu^* - q^*) \tag{11}
\end{aligned}$$

Lemma 2. (Completing the square) For any $A \in \mathbb{R}^{p \times p}$ positive definite, $B \in \mathbb{R}^{p \times p}$ positive semi-definite and $a, b \in \mathbb{R}^p$ the following identity holds:

$$\begin{aligned}
(y - a)^T A (y - a) + (y - b)^T B (y - b) &= (y - y^*)^T (A + B) (y - y^*) + \\
&\quad + (a - b)^T H (a - b),
\end{aligned}$$

where $y^* = (A + B)^{-1} (Aa + Bb)$ and $H = A(A + B)^{-1} B$. If, furthermore, B is positive definite, then $H = (A^{-1} + B^{-1})^{-1}$. [6]

Since both of our normal distributions are not degenerated because we can have inverses for both Σ^* and Ω^* , we conclude that they do not have any eigenvalues equal to 0. Moreover, since they are covariance matrices, we know that they are positive semi-definite. Therefore their eigenvalues are greater than or equal to 0. But since they can't be 0, we observe that

they have to be strictly greater than 0. This implies that both matrices are positive definite and therefore we can use the second formula for H in **Lemma 2**.

Now we are ready to apply this result to equation (11) for $y = \mu^*$, $a = \bar{r}^*$, $b = q^*$, $A = m\Sigma^{*-1}$ and $B = \Omega^{*-1}$:

$$(10) \Leftrightarrow (11) \Leftrightarrow (m-1)s^2 + (\mu^* - \bar{\mu}^*)^T (m\Sigma^{*-1} + \Omega^{*-1})(\mu^* - \bar{\mu}^*) + (\bar{r}^* - q^*)^T H(\bar{r}^* - q^*),$$

where $\bar{\mu}^* = (m\Sigma^{*-1} + \Omega^{*-1})^{-1}(m\Sigma^{*-1}\bar{r}^* + \Omega^{*-1}q^*)$ and $H = (\frac{1}{m}\Sigma^* + \Omega^*)^{-1}$.

If we go back with this result in the joint density represented by equation (10), we obtain that:

$$\begin{aligned} f(r_1^*, \dots, r_m^*, \mu^*, \Sigma^*) &\propto \\ &\propto \det(\Sigma^*)^{-\frac{m}{2}} \exp \left\{ -\frac{1}{2}(\mu^* - \bar{\mu}^*)^T (m\Sigma^{*-1} + \Omega^{*-1})(\mu^* - \bar{\mu}^*) \right\} \cdot \\ &\quad \cdot \exp \left\{ -\frac{1}{2}(\bar{r}^* - q^*)^T H(\bar{r}^* - q^*) + (m-1)s^2 \right\} \cdot \\ &\quad \cdot \det(\Omega^*)^{-\frac{1}{2}} \det(\Sigma^*)^{\frac{\nu+n+1}{2}} \exp \left\{ -\frac{1}{2} \text{Tr}(\Sigma_0 \Sigma^{*-1}) \right\} \end{aligned}$$

Since the only part that depends on μ^* is the first line of the above equation, we conclude that:

$$\begin{aligned} f(\mu^* | r_1^*, \dots, r_m^*, \Sigma^*) &\propto \exp \left\{ -\frac{1}{2}(\mu^* - \bar{\mu}^*)^T (m\Sigma^{*-1} + \Omega^{*-1})(\mu^* - \bar{\mu}^*) \right\} \Rightarrow \\ \mu^* | r_1^*, \dots, r_m^*, \Sigma^* &\sim N_n(\bar{\mu}^*, \bar{\Sigma}^*), \text{ where} \\ \bar{\mu}^* &= (m\Sigma^{*-1} + \Omega^{*-1})^{-1} (m\Sigma^{*-1}\bar{r}^* + \Omega^{*-1}q^*) \\ \bar{\Sigma}^* &= (m\Sigma^{*-1} + \Omega^{*-1})^{-1} \end{aligned} \quad (12)$$

In order to find the posterior of Σ^* , it is easier to start from the original joint density represented by equation (10). By collecting the terms that depend on Σ^* we obtain:

$$\begin{aligned} f(\Sigma^* | r_1^*, \dots, r_m^*, \mu^*) &\propto \det(\Sigma^*)^{-\frac{\nu+m+n+1}{2}} \\ &\exp \left\{ -\frac{1}{2} \left(\sum_{i=1}^m (r_i^* - \mu^*)^T \Sigma^{*-1} (r_i^* - \mu^*) + \text{Tr}(\Sigma_0 \Sigma^{*-1}) \right) \right\} \end{aligned} \quad (13)$$

We notice that this is quite close to another Inverse Wishart distribution, the only step left that we have to make is to manipulate the exponential. Let us notice that

$$\begin{aligned} \sum_{i=1}^m (r_i^* - \mu^*)^T \Sigma^{*-1} (r_i^* - \mu^*) \in \mathbb{R} &\Rightarrow \sum_{i=1}^m (r_i^* - \mu^*)^T \Sigma^{*-1} (r_i^* - \mu^*) = \\ &= \text{Tr} \left(\sum_{i=1}^m (r_i^* - \mu^*)^T \Sigma^{*-1} (r_i^* - \mu^*) \right) = \sum_{i=1}^m \text{Tr} \left((r_i^* - \mu^*)^T \Sigma^{*-1} (r_i^* - \mu^*) \right) \end{aligned}$$

But inside the *Trace* matrices are cyclically commutative as long as the dimensions agree:

$$\begin{aligned} \sum_{i=1}^m \text{Tr} \left((r_i^* - \mu^*)^T \Sigma^{*-1} (r_i^* - \mu^*) \right) &= \sum_{i=1}^m \text{Tr} \left((r_i^* - \mu^*) (r_i^* - \mu^*)^T \Sigma^{*-1} \right) = \\ &= \text{Tr} \left(\sum_{i=1}^m (r_i^* - \mu^*) (r_i^* - \mu^*)^T \Sigma^{*-1} \right) \end{aligned}$$

Finally, by using this result and equation (13), we obtain:

$$\begin{aligned} (13) &\Leftrightarrow f(\Sigma^* | r_1^*, \dots, r_m^*, \mu^*) \propto \det(\Sigma^*)^{-\frac{\nu+m+n+1}{2}} \\ &\exp \left\{ -\frac{1}{2} \text{Trace} \left(\left(\Sigma_0 + \sum_{i=1}^m (r_i^* - \mu^*) (r_i^* - \mu^*)^T \right) \Sigma^{*-1} \right) \right\} \end{aligned}$$

We notice that this is the kernel of an Inverse Wishart distribution. Therefore, we can conclude that:

$$\Sigma^* | r_1^*, \dots, r_m^*, \mu^* \sim W^{-1} \left(\nu + m, \Sigma_0 + \sum_{i=1}^m (r_i^* - \mu^*) (r_i^* - \mu^*)^T \right) \quad (14)$$

Now that we have the posterior distributions, we can implement a Gibbs Sampler, which we will see in the following section, where we will also look at how the parameters of the model were estimated.

2.4 Implementation

For implementation purposes, 4 stocks were chosen: Apple(AAPL), Amazon(AMZN), Google(GOOG) and Microsoft(MSFT). Closing prices for the 4 from 1/2/2015 until 5/1/2017 were considered and the returns were computed. Now, this data is split into 2 parts, one representing the *current*

data (the last m returns r_1, \dots, r_m , here $m = 21$) and the rest representing *historical data* used to estimate the parameters in the model. The reason why $m = 21$ was chosen is 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. Hence, in this example, the trading period for such an investor would be over a month. Next step is to augment P as discussed in section 2.2. Once P^* is created, we can just create our transformed returns $r_i^* = P^* r_i$. For this example, the personal views were (the columns represent AAPL, AMZN, GOOG, MSFT, respectively):

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

If we look at the second assumption in the model represented by equation (8), we notice that q^* and Ω^* are, respectively, the mean and covariance matrix for μ^* , which is in turn a mean of returns from a particular month (again, in this example $m = 21$, approximately a month). Hence, one solution for estimating the parameters would be to take the returns from each month in the historical data and to compute their means. This way, we would have estimates for the monthly mean returns $\hat{\mu}_i^*$, with i an integer between 1 and the number of months in the historical data. Once we obtain those, we can estimate \hat{q}^* and $\hat{\Omega}^*$ by taking the mean and the covariance of $\hat{\mu}_i^*$.

But, we have to remember that we need to reflect our personal views in the estimation presented above. In equations (5) and (6), we have showed how one should combine the estimates from the procedure just presented with the investor's personal views:

- Equation (5) shows that we should take the \hat{q}^* obtained through the above estimation and replace the first k entries with q (k , as mentioned at the beginning, was the number of personal views).
- Equation (6) shows that we should take the obtained $\hat{\Omega}^*$ and replace the top left $k \times k$ matrix with our personal choice of Ω .

Now that the parameters of our model are estimated, a typical Gibbs Sampler was used based on the posteriors represented by equations (14) and (12).

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 $\mu^{*(t)}$, call it $\hat{\mu}^*$, and the average of the simulated $\Sigma^{*(t)}$, call it $\hat{\Sigma}^*$. However, one has to remember that those

Algorithm 1 Gibbs Sampler

- 1: $\Sigma^{*(t+1)} | r_1^*, \dots, r_m^*, \mu^{*(t)} \sim W^{-1} \left(\nu + m, \Sigma_0 + \sum_{i=1}^m (r_i^* - \mu^{*(t)})(r_i^* - \mu^{*(t)})^T \right)$
2:

$$\begin{aligned} \mu^{*(t+1)} | r_1^*, \dots, r_m^*, \Sigma^{*(t+1)} &\sim N_n \left(\overline{\mu}^{*(t+1)}, \overline{\Sigma}^{*(t+1)} \right), \text{ where} \\ \overline{\mu}^{*(t+1)} &= \left(m \Sigma^{*(t+1)-1} + \Omega^{*-1} \right)^{-1} \left(m \Sigma^{*(t+1)-1} \bar{r}^* + \Omega^{*-1} q^* \right) \\ \overline{\Sigma}^{*(t+1)} &= \left(m \Sigma^{*(t+1)-1} + \Omega^{*-1} \right)^{-1} \end{aligned}$$

were transformed using P^* , hence now we would have to transform them back into the original space: $\hat{\mu} = P^{*-1} \hat{\mu}^*$, $\hat{\Sigma} = P^{*-1} \hat{\Sigma}^* P^{*-T}$. Just like in the original model, in order to get the weights, one would use an equation similar to the CAPM one presented in section 1.3: $w = \frac{1}{\lambda} \hat{\Sigma}^{-1} \hat{\mu}$. Here, $\lambda = 2.5$, as chosen in the original model. Also there has been extensive research when it comes to choosing λ . For trading stocks a risk aversion coefficient between 2 and 3 is reasonable.[4] Finally, we are ready to compare the results obtained under the original model with the ones obtained from this one.

2.5 Results Comparison

Before we delve into how we compare the 2 approaches, let us make the observation that in order to make any kind of comparison, one has to make sure that the same data sets were used and the parameters were estimated in the same way. Albeit the same personal views were imputed (same P , Ω , q), the two approaches differ in the fact that the alternative one has a prior on Σ and the original one makes use of the market equilibrium returns, which is estimated using $\pi = \lambda \Sigma w_{eq}$. In the following table, we can look at the setup for both side by side:

Alternative	Original
$r_1^*, r_2^*, \dots, r_m^* \sim N_n(\mu^*, \Sigma^*)$	$r \sim N(\mu, \Sigma)$
$\mu^* \sim N(q^*, \Omega^*)$	$\mu \sim N(\pi, \tau \Sigma)$

Instead of the market equilibrium, the alternative approach simply has another parameter, which is estimated as mentioned in section 2.4 (also the alternative has a prior on Σ and takes into consideration current data). Besides this difference, the two are using the same data sets and the same parameters. Now, the question becomes how should one compare the two.

One obvious approach would be to see how the two would perform if one would use them on the real market, which will be presented in the results section for the models that will follow later in this paper. However, it is of more interest to us to check how close to our personal opinion is the posterior mean obtained from the Gibbs Sampler.

Remark 1. *Since for both models we have that $P\mu \sim N(q, \Omega)$, the smaller the uncertainty in our views (the diagonal entries of Ω), the smaller the standard deviation and, hence, the more certain the investor is about that particular view.*

Hence, from the above remark, we will look at how $P\hat{\mu}$ behaves as we look at small values for the diagonal entries of Ω . But how should one define "small"? As we have seen in section 2.4, the expected returns for the views were $q = \begin{bmatrix} 0.02 \\ 0.05 \end{bmatrix}$. Hence, even a value of 10^{-4} is quite large since this would be the variance of our view and, therefore, the standard deviation would become 10^{-2} . Hence, a 95% confidence interval for the first view would be $(0, 0.04)$. If one tries to input even smaller ω , the Inverse Wishart random generator gives a non-singularity error. Hence, we conclude that we compare the models on values of the diagonal of the matrix Ω that are between 0 and 10^{-4} . Albeit we can't input smaller ω , for the purposes of checking the following remark, we changed q to $q = \begin{bmatrix} 0.2 \\ 0.5 \end{bmatrix}$. Hence, for both models an exhaustive method was implemented that would compute for each pair of diagonal entries in Ω a posterior mean $\hat{\mu}$. Once this is obtained, the distance $|P\hat{\mu} - q|$ can be calculated for both models.

Remark 2. *Since $P\mu \sim N(q, \Omega)$, we have that $\lim_{\Omega \rightarrow O_2} P\mu = q$ a.s.*

Therefore, as the diagonal entries of Ω get smaller and smaller we expect to get closer and closer to q .

Remark 3. *If we look at the posterior of μ^* we have that:*

$$\begin{aligned} \mu^* | r_1^*, \dots, r_m^*, \Sigma^* &\sim N_n(\bar{\mu}^*, \bar{\Sigma}^*), \text{ where} \\ \bar{\mu}^* &= \left(m\Sigma^{*-1} + \Omega^{*-1} \right)^{-1} \left(m\Sigma^{*-1}\bar{r}^* + \Omega^{*-1}q^* \right) \\ \bar{\Sigma}^* &= \left(m\Sigma^{*-1} + \Omega^{*-1} \right)^{-1} \end{aligned}$$

If we consider a small $\Omega^ \Rightarrow \Omega^{*-1}$ is large and therefore the whole term $m\Sigma^{*-1} + \Omega^{*-1} \approx \Omega^{*-1} \Rightarrow (m\Sigma^{*-1} + \Omega^{*-1})^{-1} \approx \Omega^*$. Similarly, $(m\Sigma^{*-1}\bar{r}^* +$*

$\Omega^{*-1}q^*) \approx \Omega^{*-1}q^*$ for small enough Ω^* . Hence, we would expect that the mean of the simulated $\mu^{*(t)}$ is close to q^* . Or, with the notation already used, $\hat{\mu}^* \approx q^*$. Hence, by using the previous remark also, we obtain that $P(P^{*-1}\hat{\mu}^*) \approx q$.

The following graphs have as 2 of the axis the 2 diagonal entries in Ω and the third one represents the distance $|P\bar{\mu} - q| = |P(P^{*-1}\bar{\mu}^*) - q|$:

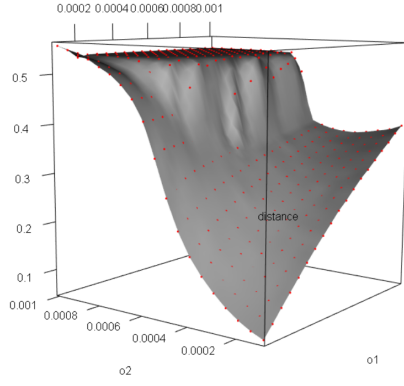


Figure 1: Results of Ω for the alternative model

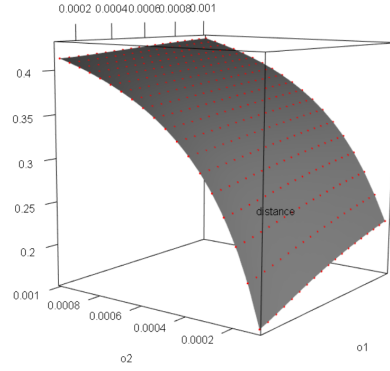


Figure 2: Results of Ω for original model

We notice from the z-axis, which represents the distance mentioned above, that the modified model more closely follows the personal views. We can look at some specific values of the distance for different pairs of ω_1 and ω_2 in table 1.

However, we would like to see if the structure of $P\hat{\mu}$ is similar to q . For this we keep the two entries in Ω equal, we exhaustively search over small ω s.t. $\Omega = \omega\mathbb{I}$ and we plot the 2 entries of $P\hat{\mu}$ together with the respective ω . Please note that the blue point in figures 2.5 represent the exact value of $q = \begin{bmatrix} 0.2 \\ 0.5 \end{bmatrix}$, which would be obtained for $\omega = 0$.

By comparing the 2 figures, we notice that not only the point simulations represented by the red points are closer, but the whole curve (which was obtained by interpolation) seems to be closer to the theoretical value represented by the blue point. Also, we notice that in both cases, as ω increases, $P\hat{\mu}$ gets further away from q , which is what theoretically should happen.

ω_1	ω_2	Original	Alternative
10^{-4}	0.0001	0.154	0.052
10^{-4}	0.00015	0.2	0.063
10^{-4}	0.0002	0.235	0.077
10^{-4}	0.00025	0.263	0.118
10^{-4}	0.0003	0.286	0.118
10^{-4}	0.00035	0.305	0.145
10^{-4}	0.0004	0.321	0.177
10^{-4}	0.00045	0.335	0.222
10^{-4}	0.0005	0.347	0.282
10^{-4}	0.00055	0.357	0.354

Table 1: Table with specific distance values

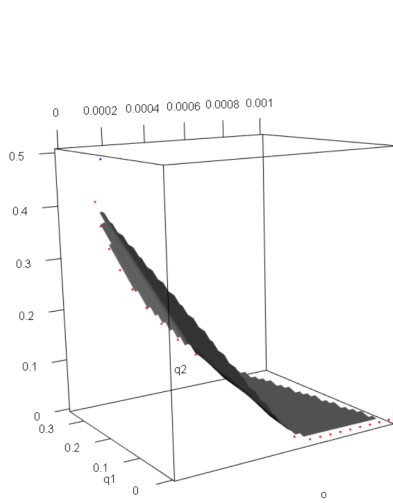


Figure 3: Results of Ω for the alternative model

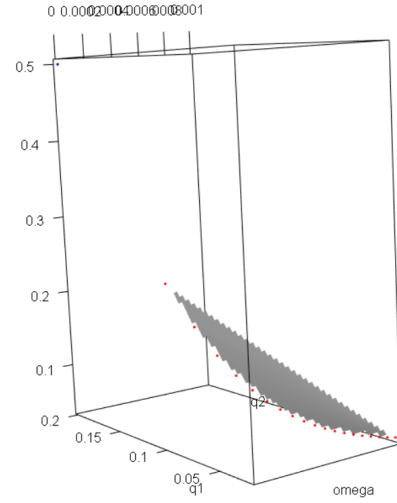


Figure 4: Results of Ω for original model

2.6 But do we need an invertible P ?

In section 2.2, we introduced a method of creating an invertible matrix P by adding rows. Ideally, the matrix of views should not be modified in any way. Also, in general, one can't create an invertible matrix by adding rows, especially when the number of views is close to the number of stocks

selected. Also, the method of augmenting P is not unique. Therefore, we would like to be able to derive the posteriors when P is unchanged from what the investor is inputting. Hence, in this section, we will consider the same setup as before, the only difference being the fact that P is not even square:

$$\begin{aligned} r_1, r_2, \dots, r_m | \mu, \Sigma &\sim N_n(\mu, \Sigma) \\ P\mu &\sim N_k(q, \Omega) \\ \Sigma &\sim W^{-1}(\nu, \Sigma_0) \end{aligned}$$

Since P shows up in the second equation of our model assumptions, the only posterior that will change from what we had previously will be that for μ . Hence, in the joint distribution, we will consider only the terms depending on μ :

$$\begin{aligned} f(\mu | r_1, \dots, r_m) &\propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^m (r_i - \mu)^T \Sigma^{-1} (r_i - \mu) \right\} \\ &\cdot \exp \left\{ -\frac{1}{2} (P\mu - q)^T \Omega^{-1} (P\mu - q) \right\} \end{aligned}$$

For the first exponential we can use Lemma 1. This yields:

$$\begin{aligned} f(\mu | r_1, \dots, r_m) &\propto \exp \left\{ -\frac{1}{2} ((m-1)s^2 + m(\bar{r} - \mu)^T \Sigma^{-1} (\bar{r} - \mu)) \right\} \\ &\cdot \exp \left\{ -\frac{1}{2} (q - P\mu)^T \Omega^{-1} (q - P\mu) \right\} \end{aligned}$$

We remember that $s^2 = \frac{\sum_{i=1}^m (r_i - \bar{r})^T \Sigma^{-1} (r_i - \bar{r})}{m-1}$ and hence this term does not depend on μ . Now, let us focus on the remaining terms in the exponential:

$$\begin{aligned} &(\bar{r} - \mu)^T (m\Sigma^{-1}) (\bar{r} - \mu) + (q - P\mu)^T \Omega^{-1} (q - P\mu) = \\ &= \bar{r}^T (m\Sigma^{-1}) \bar{r} - 2\bar{r}^T (m\Sigma^{-1}) \mu + \mu^T (m\Sigma^{-1}) \mu + q^T \Omega^{-1} q - 2q^T \Omega^{-1} P\mu + \\ &+ \mu^T P^T \Omega^{-1} P\mu = \mu^T (m\Sigma^{-1} + P^T \Omega^{-1} P) \mu - 2(\bar{r}^T (m\Sigma^{-1}) + q^T \Omega^{-1} P) \mu + \\ &\quad + \bar{r}^T (m\Sigma^{-1}) \bar{r} + q^T \Omega^{-1} q \end{aligned}$$

Since only the first two terms depend on μ , we obtain that:

$$f(\mu|r_1, \dots, r_m, \Sigma) \propto \exp \left\{ -\frac{1}{2} \mu^T (m\Sigma^{-1} + P^T \Omega^{-1} P) \mu \right\} \cdot \exp \left\{ -\frac{1}{2} 2(m\Sigma^{-1} \bar{r} + P^T \Omega^{-1} q)^T \mu \right\}$$

Lemma 3. *Let M be a symmetric and invertible matrix, then the following identity holds:*

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

Proof. We just need to expand the quadratic term:

$$\begin{aligned} (x - M^{-1}b)^T M (x - M^{-1}b) &= x^T M x - 2b^T M^{-1} M x + b^T M^{-1} M M^{-1} b = \\ &= x^T M x - 2b^T x + b^T M^{-1} b \end{aligned}$$

□

Hence, if we apply this lemma for $x = \mu$, $M = m\Sigma^{-1} + P^T \Omega^{-1} P$ and $b = m\Sigma^{-1} \bar{r} + P^T \Omega^{-1} q$, we obtain that the exponential in the distribution of the posterior of μ is (the $-\frac{1}{2}$ still sits in front of the formula, we just omit it in the following for simplicity of writing):

$$\begin{aligned} &(\mu - (m\Sigma^{-1} + P^T \Omega^{-1} P)^{-1} (m\Sigma^{-1} \bar{r} + P^T \Omega^{-1} q))^T (m\Sigma^{-1} + P^T \Omega^{-1} P) \cdot \\ &\cdot (\mu - (m\Sigma^{-1} + P^T \Omega^{-1} P)^{-1} (m\Sigma^{-1} \bar{r} + P^T \Omega^{-1} q)) - b^T M^{-1} b \end{aligned}$$

Lastly, we notice that b and M do not depend on μ , and hence, the posterior of μ is dictated by the first big term, which is actually the density of a normal distribution:

$$\boxed{\mu|r_1, \dots, r_m, \Sigma \sim N(\mu_{post}, \Sigma_{post})}, \text{ where}$$

$$\boxed{\mu_{post} = (m\Sigma^{-1} + P^T \Omega^{-1} P)^{-1} (m\Sigma^{-1} \bar{r} + P^T \Omega^{-1} q)}$$

$$\boxed{\Sigma_{post} = (m\Sigma^{-1} + P^T \Omega^{-1} P)^{-1}}$$

This posterior is very close to the one obtained by using the first approach (represented by equation (12)), the only difference being the fact that in this new approach the matrix P shows up. This is because here we did not change the investor inputted matrix P , while in the previous approach we augmented P in order for it to be invertible.

2.7 Implementation

Implementing this model is straightforward since it is very similar to the previous version. The only difference is the fact that in the posterior for μ we have P appearing, while in the previous model there was no P since we were adding rows to it so that it becomes invertible. We remind ourselves that this was the first approach because we can take the inverse and easily find the prior distribution of μ from the prior distribution of $P\mu$. Using the derived posteriors, the Gibbs Sampler is:

Algorithm 2 Gibbs Sampler

1: $\Sigma^{(t+1)} | r_1, \dots, r_m, \mu^{(t)} \sim W^{-1} (\nu + m, \Sigma_0 + \sum_{i=1}^m (r_i - \mu^{(t)})(r_i - \mu^{(t)})^T)$
2:

$$\begin{aligned} \mu^{(t+1)} | r_1, \dots, r_m, \Sigma^{(t+1)} &\sim N \left(\mu_{post}^{(t+1)}, \Sigma_{post}^{(t+1)} \right), \\ \mu_{post}^{(t+1)} &= (m\Sigma^{(t+1)^{-1}} + P^T\Omega^{-1}P)^{-1} (m\Sigma^{(t+1)^{-1}}\bar{r} + P^T\Omega^{-1}q) \\ \Sigma_{post}^{(t+1)} &= \left(m\Sigma^{(t+1)^{-1}} + P^T\Omega^{-1}P \right)^{-1} \end{aligned}$$

2.8 Results

Just like before, we will try to look at the sensitivity of our model to different confidence levels. Remarks 1 and 2 made when we presented the results for the previous model still hold. Since in Ω we have on the main diagonal (call them ω_i) the variances in our views $P\mu$, the smaller the ω_i , the more certain we are in view i . This should also be reflected in our posterior: if we provide very large ω_i , it means that we are very uncertain about the views and the model should take into consideration the history a lot more, while if we provide very small values for ω_i , it means that we are very certain about the views and the model should take them into consideration a lot more than the history.

Just like before, in order to quantify and visualize the model's sensitivity to different confidence levels, we will look at the distance $|P\mu_{post} - q|$ (which will be on one of the axis in our plots) over different combinations of ω_i . The same 4 stocks from before were chosen (AAPL,AMZN,GOOG,MSFT), but since this work is more recent, the daily returns are from 1/2/2014 to 12/29/2017. The views are (rows are views and the columns represent the 4 stocks in the order AAPL,AMZN,GOOG,MSFT):

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

When it comes to the confidence levels in the 2 views, one can input values as small as 10^{-7} without encountering any numerical issues, like we did previously when we were augmenting the matrix P . Hence, one doesn't need to make any change to the model when implementing it or when inputting any value. We take a grid of equally spaced points (ω_1, ω_2) between 10^{-7} and $2 \cdot 10^{-5}$. The burn period was set to 10^3 and the number of iterations in the Gibbs Sampler was set to 10^4 .

However, one could also use the same views, but considering the daily returns for the whole *S&P500* instead of just for 4 stocks. For this, we need the daily returns of companies actively traded in *S&P500* over the period mentioned before. We won't have to change q at all, but P has more columns since they would represent the stocks in the famous index and it will still have 2 rows for the same 2 views. One would fill out P by making sure that in the first row and the column corresponding to AAPL we will have a -1 , in the first row and the column corresponding to AMZN we will have a 1 and similarly for the second row. Of course, the dimension of some of the matrices and vectors will be much bigger and therefore, all computations will be more expensive. Hence, this version was parallelized and the number of iterations in the Gibbs Sampler decreased to 10^3 (as we will see, even with so few iterations, convergence for the mean is achieved, but convergence for the covariance matrix is not). The interval 10^{-7} to 10^{-5} for the confidence levels was split into 4 parts, in the following way:

$$(\omega_1, \omega_2) \in \begin{cases} (10^{-6} \rightarrow 10^{-5}, 10^{-6} \rightarrow 10^{-5}) \\ (10^{-5} \rightarrow 10^{-4}, 10^{-5} \rightarrow 10^{-4}) \\ (10^{-6} \rightarrow 10^{-5}, 10^{-5} \rightarrow 10^{-4}) \\ (10^{-5} \rightarrow 10^{-4}, 10^{-6} \rightarrow 10^{-5}) \end{cases}$$

Each one of the 4 ranges from above was divided into a grid of 7^2 points. Each point was ran on one core, taking a little more than 4 hours to run.

- When $\omega_1 = 10^{-6}$, a 95% confidence interval for the first view would be $(0.018, 0.022)$, which would show that the investor is very confident.
- When $\omega_1 = 10^{-4}$, a 95% confidence interval for the first view would be $(0, 0.04)$, which would show that the investor is not as confident.

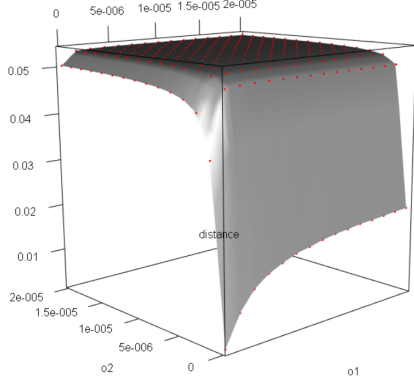


Figure 5: $|P\mu_{post} - q|$ when taking only the 4 stocks

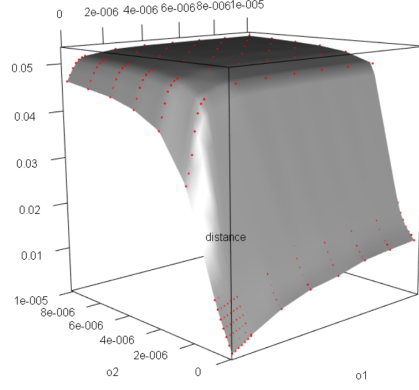


Figure 6: $|P\mu_{post} - q|$ when taking *S&P500*

In the figures presented, we notice that both curves have similar shapes, albeit the one on the right converges faster to 0 as o_i become smaller (ω_i in our model). Also, the curve on the right seems to be underneath the one on the left. Intuitively, this is because there is a lot more information in our prior for Σ when we take the whole *S&P500*. Moreover, both have very similar shapes. The distances go to 0 as ω_i go to 0. This is in tune with our intuition of how the model should behave like: as one gets more and more confident in their inputted views, the model should put a lot of importance on them and not on the historical data. Vice-versa, in both figures the distance seems to converge to a certain value as ω_i become bigger and bigger. Again, this is what we would think that the model should do since large ω_i , suggests that one is uncertain about the personal view and therefore, the history should play a more important role. Indeed, if we would only take the historical returns, an unbiased estimate for μ is \bar{r} and the distance becomes $|P\bar{r} - q| = 0.05388875$, which is what the plots seem to tend to converge to.

We will move our focus towards looking at the profits (or losses) that one would obtain when using the model to trade over the month of January 2018 (testing data consisting of daily returns between 1/2/2018 and 1/30/2018) using an initial capital of 100,000\$ (this does not include any capital requirements for short selling). We remember that in order to get portfolio weights we use the same approach as before. From Gibbs Sampling we estimate μ_{post} and Σ_{post} and we use the CAPM equation 4: $w = \frac{1}{2.5} \Sigma_{post}^{-1} \mu_{post}$.

Albeit when we took the whole *S&P500* the number of iterations in the

Gibbs Sampler was small, we notice from the above analysis that we still get very good estimates for μ_{post} since the posterior distance behaves exactly like our intuition suggests it should do. The running averages for the mean also converge fast for small ω_i . However, because of the size of Σ_{post} and because of the fact that one has to take its inverse in order to compute the portfolio weights w , the number of iterations is not enough to give accurate predictions of profits. Nevertheless, for completeness, the average profit when considering the whole *S&P500* is 13191.39\$ with a standard deviation of 2908.134\$.

We will now present the profits obtained when using only 4 stocks. We notice that the first view has a bigger impact on the profits curve than the second view. Moreover, as the confidence in the first view increases (as ω_1 goes to 0), the profits sky rocket. This is because over the month of January 2018 AMZN outperformed AAPL by 23.997% and our view was indeed that AMZN will overrun AAPL (albeit by only 2%, a 10th of what actually happened in reality).

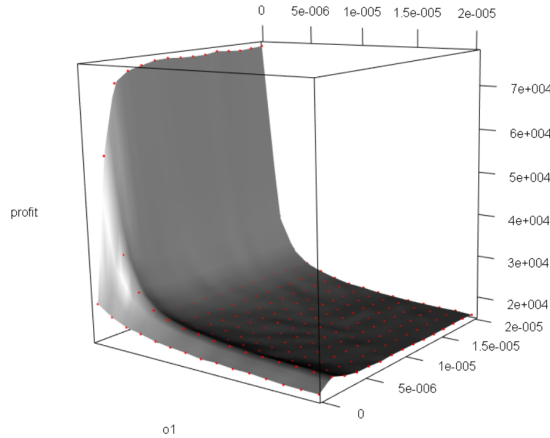


Figure 7: Profit when taking only 4 stocks

AMZN outperforming AAPL by nearly 24% in one month is uncommon. Therefore, next we will present the same results, the only change made is that we replace AMZN with FB (Facebook). The same data sets were used and all other inputs stay exactly the same as we just presented at the beginning of this section, except q . We will also look at how the model behaves when the investor inputs a personal view exactly like what happened during the month of January 2018 (very "informed" investor) and exactly the opposite of what happened during the month of January (very "uninformed"

investor). Therefore, we will also look at what happens when we choose $q = \begin{bmatrix} 0.06212815 \\ 0.01366718 \end{bmatrix}$ and $q = -\begin{bmatrix} 0.06212815 \\ 0.01366718 \end{bmatrix}$, respectively.

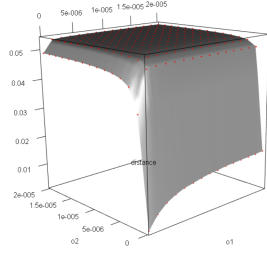


Figure 8: 4 stocks, FB in and $q = [0.02, 0.05]^T$

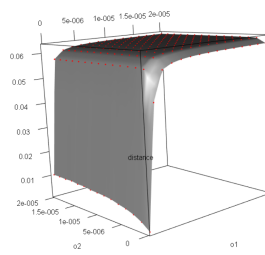


Figure 9: 4 stocks, FB in and view exactly like reality

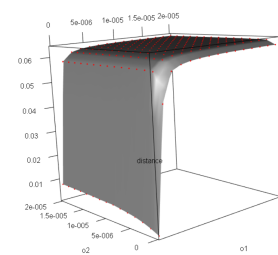


Figure 10: 4 stocks, FB in and view opposite of reality

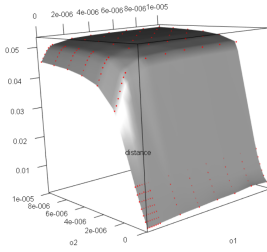


Figure 11: *S&P500*, FB in and $q = [0.02, 0.05]^T$

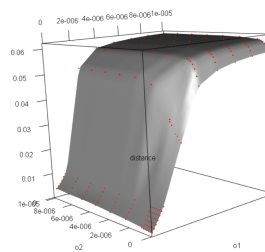


Figure 12: *S&P500*, FB in and view exactly like reality

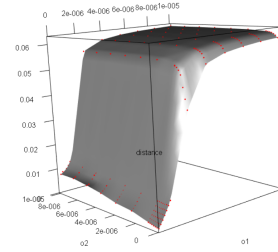


Figure 13: *S&P500* FB in and view opposite of reality

Again, just like before, we notice that, as ω_i get smaller and smaller, when taking into account the whole *S&P500*, the curve seems to be under and closer to 0 than the one when taking into account only 4 stocks. This might be because the prior on the covariance matrix containing the whole *S&P500* has more information than the one which only has 4 stocks. Moreover, for the same q , the curves have a similar orientation and general shape. Hence, this confirms the belief that albeit a small number of iterations was used for the Gibbs Sampler that takes into account the whole *S&P500*, the estimated posterior mean is still accurate. However, as mentioned before, the estimate for Σ_{post}^{-1} when it's size is big is not accurate enough to have very reliable profit estimates.

Nevertheless, for completeness of this analysis, we proceed by leaving

all the inputs mentioned before unchanged and keeping $q = \begin{bmatrix} 0.02 \\ 0.05 \end{bmatrix}$. When taking into account the whole *S&P500*, the average profit over the before mentioned range of simulated pairs (ω_1, ω_2) is 11,619.97\$ with a standard deviation of 2,852.246\$. In the next plot we can observe the profits obtained when considering just the 4 stocks mentioned before.

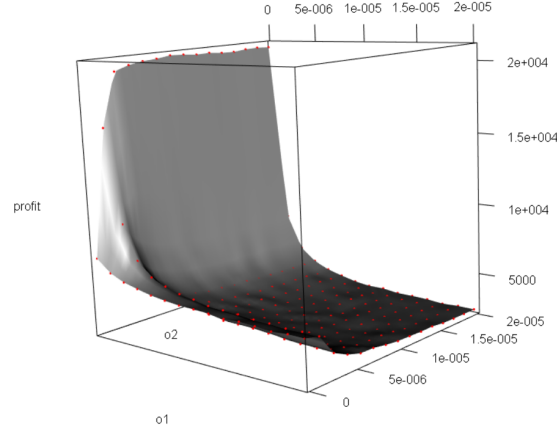


Figure 14: Profit 4 stocks, FB in and $q = [0.02, 0.05]^T$

From the figure, one can see that the first view has a higher influence on the profits than the second view. This is because if we let ω_2 constant the resulting curve increases a lot faster than the curve obtained by keeping ω_1 constant.

3 Prior on $\log(\Sigma)$

3.1 Introduction

Just like when introducing the approach with an Inverse Wishart prior, let us see what we would like to improve on it:

- Ideally, the matrix of views P should not be augmented or changed in any way. It should be left just like the user inputted it.
- One should try a different prior than the Inverse Wishart, which has been used numerous times.

Therefore, two of the assumptions will be unchanged:

$$\begin{aligned} r_1, r_2, \dots, r_m | \mu, \Sigma &\sim N_n(\mu, \Sigma) \\ P\mu &\sim N_k(q, \Omega) \end{aligned}$$

A very interesting idea for a different prior on the covariance matrix is presented by Leonard and Hsu (1992)[5]. As the title of this section is hinting, this prior will actually be on $\log(\Sigma)$. In order to better understand Leonard and Hsu's idea, let us look at the distribution:

$$f(r_1, \dots, r_m | \mu, \Sigma) = (2\pi)^{-\frac{mn}{2}} \det(\Sigma)^{-\frac{m}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^m (r_i - \mu)^T \Sigma^{-1} (r_i - \mu) \right\}$$

Let $A = \log(\Sigma)$, λ_{Ai} and $\lambda_{\Sigma i}$ (with $i = \overline{1, n}$) be the eigenvalues of A and Σ respectively. Since $A = \log(\Sigma)$ we obtain that $\lambda_{Ai} = \log(\lambda_{\Sigma i}) \Rightarrow \lambda_{\Sigma i} = e^{\lambda_{Ai}}$. Finally, by remembering that the determinant is the product of the eigenvalues and that the trace of a matrix is the sum of the eigenvalues, we notice that $\det(\Sigma) = \prod_{i=1}^n \lambda_{\Sigma i} = \prod_{i=1}^n e^{\lambda_{Ai}} = e^{Tr(A)}$. By using this in the joint distribution of the returns and by noticing that $(r_i - \mu)^T \Sigma^{-1} (r_i - \mu) \in \mathbb{R}$ we obtain:

$$\begin{aligned} f(r_1, \dots, r_m | \mu, \Sigma) &= (2\pi)^{-\frac{mn}{2}} \exp \left\{ -\frac{1}{2} Tr \left(\sum_{i=1}^m (r_i - \mu)^T \Sigma^{-1} (r_i - \mu) \right) \right\} \\ \exp \left\{ -\frac{m}{2} Tr(A) \right\} &= (2\pi)^{-\frac{mn}{2}} \exp \left\{ -\frac{1}{2} Tr \left(\sum_{i=1}^m (r_i - \mu)(r_i - \mu)^T \Sigma^{-1} \right) \right\} \\ \exp \left\{ -\frac{m}{2} Tr(A) \right\} &= (2\pi)^{-\frac{mn}{2}} \exp \left\{ -\frac{m}{2} Tr(A + S e^{-A}) \right\} \end{aligned}$$

Here, $S = \frac{1}{m} \sum_{i=1}^m (r_i - \mu)^T \Sigma^{-1} (r_i - \mu)$. Before we continue, let us define an operator and make a few notations.

Definition 1. Let A be a $n \times n$ matrix, $A = (a_{ij})_{i,j=\overline{1,n}}$, then we define an operator that stacks in a vector the entries parallel to the main diagonal:

$$Vec^*(A) = [a_{11} \ a_{22} \ \dots \ a_{nn} \mid a_{12} \ a_{23} \ \dots \ a_{n-1n} \mid \dots \mid a_{1n}]^T$$

We notice that if A is $n \times n$, $Vec^*(A)$ is $\frac{n(n+1)}{2} \times 1$. This definition brings us to the following notations:

Notation 1.

$$\begin{aligned} \lambda &= Vec^*(\log(S)), \alpha = Vec^*(\log(\Sigma)) \\ \Lambda &= \log(S), A = \log(\Sigma), d = \frac{n(n+1)}{2} \end{aligned}$$

The idea that Leonard and Hsu had was to approximate $f(r_1, \dots, r_m | \mu, \Sigma)$ by approximating e^{-A} . The approximation makes use of the fact that $X(\omega) = e^{-A\omega}$ satisfies a Volterra integral equation[2]:

$$X(t) = S^{-t} - \int_0^t S^{s-t} (A - \Lambda) X(v) dv, 0 < t < \infty,$$

By letting $t = 1$, by iterative substitution of $X(v)$ and by using the spectral decomposition of matrix S we obtain that the approximation is (please see Appendix B for the proof):

$$f^*(r_1, \dots, r_m | \alpha) = (2\pi e)^{-\frac{mn}{2}} \det(S)^{-\frac{m}{2}} \exp \left\{ -\frac{1}{2} (\alpha - \lambda)^T Q (\alpha - \lambda) \right\} \quad (15)$$

In order to see how to compute Q , we first have to introduce a couple more notations. If we let e_i, d_i to be the i^{th} normalized eigenvector with its corresponding eigenvalue, respectively, then f_{ij} is obtained by looking at the equation $Vec^*(\log(\Sigma))^T f_{ij} = e_i^T \log(\Sigma) e_j$ and identifying the coefficients of the entries in the $\log(\Sigma)$ matrix. With those f_{ij} , we can finally compute Q :

$$\begin{aligned} \xi_{ij} &= \frac{(d_i - d_j)^2}{d_i d_j (\log(d_i) - \log(d_j))^2} \\ Q &= \frac{m}{2} \sum_{i=1}^n f_{ii} f_{ii}^T + m \sum_{i < j}^n \xi_{ij} f_{ij} f_{ij}^T \end{aligned}$$

Remark 4. *The approximate distribution is: $\alpha|r_1, \dots, r_m \approx \sim N(\lambda, Q^{-1})$*

Now we are ready to move on to the next section and resent the assumptions of the model.

3.2 The Model

As mentioned in the previous section, we will have a prior on the $\log(\Sigma)$. But how would one construct an intuitive distribution? The simplest distribution that one could work with is the multivariate normal, in which the variance terms on the main diagonal have a mean θ_1 and a variance σ_1^2 and the covariance terms, which are on the off diagonal, have another mean θ_2 and another variance σ_2^2 . Hence, we arrive at the following model:

$$r_1, \dots, r_m | \mu, \Sigma \sim N(\mu, \Sigma) \quad (16)$$

$$P\mu \sim N(q, \Omega) \quad (17)$$

$$\alpha | \theta, \Delta = \text{Vec}^*(\log(\Sigma)) | \theta, \Delta \sim N(J\theta, \Delta) \quad (18)$$

Where we have the following notations:

Notation 2.

$$J = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix}, \Delta = \begin{bmatrix} \sigma_1^2 I_n & \mathbb{O} \\ \mathbb{O} & \sigma_2^2 I_{d-n} \end{bmatrix}, \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$$

3.3 Derivation of Posterior Distributions

If we let θ to have a uniform prior ($\theta \propto 1$) by integrating it out from the density in equation (18), we obtain:

Proposition 1.

$$\begin{aligned} f(\alpha | \sigma_1^2, \sigma_2^2) &= \int_{\theta} \det(\Delta)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\alpha - J\theta)^T \Delta^{-1} (\alpha - J\theta) \right\} d\theta = \\ &= 2\pi \det(\Delta)^{-\frac{1}{2}} \det(J^T \Delta^{-1} J)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \alpha^T G \alpha \right\}, \text{ where} \\ G &= (I_d - J(J^T \Delta^{-1} J)^{-1} J^T \Delta^{-1})^T \Delta^{-1} (I_d - J(J^T \Delta^{-1} J)^{-1} J^T \Delta^{-1}) \end{aligned}$$

For the proof, please see the Appendix C.

Now, by using this distribution together with the approximation obtained from the Volterra integral of the distribution of returns denoted by equation (15) and with the prior on $P\mu$ represented by equations (17), we can finally obtain the approximate joint distribution:

$$\begin{aligned} f(\alpha, \mu, \sigma_1^2, \sigma_2^2, r_1, \dots, r_m) \approx & \propto \det(\Delta)^{-\frac{1}{2}} \det(J^T \Delta^{-1} J)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \alpha^T G \alpha \right\} \cdot \\ & \cdot \det(S)^{-\frac{m}{2}} \exp \left\{ -\frac{1}{2} (\alpha - \lambda)^T Q (\alpha - \lambda) \right\} \cdot \\ & \cdot \det(\Omega)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (P\mu - q)^T \Omega^{-1} (P\mu - q) \right\} \end{aligned} \quad (19)$$

We will first proceed with finding the posterior of α . Hence, we have to collect all the terms depending on α . Since one of those is the approximation obtained from the Volterra integral, the posterior is going to be an approximate distribution:

$$f^*(\alpha | r_1, \dots, r_m, \sigma_1^2, \sigma_2^2, \mu) \approx \propto \exp \left\{ -\frac{1}{2} (\alpha^T G \alpha + (\alpha - \lambda)^T Q (\alpha - \lambda)) \right\}$$

We can apply **Lemma 2 (Completing the square)** with $y = \alpha, a = 0, A = G, b = \lambda, B = Q$ and we obtain that:

$$\boxed{\alpha | r_1, \dots, r_m, \sigma_1^2, \sigma_2^2, \mu \approx \sim N(\alpha^*, (Q + G)^{-1}), \text{ where } \alpha^* = (Q + G)^{-1} Q \lambda}$$

Moving to the posterior of σ_1^2, σ_2^2 , we have to collect the terms depending on Δ , which also includes G . We note that the term obtained from the Volterra integral approximation of the matrix exponential does not show up in this posterior. Hence, this will be an exact distribution:

$$f(\sigma_1^2, \sigma_2^2 | \alpha, \mu, r_1, \dots, r_m) \propto \det(\Delta)^{-\frac{1}{2}} \det(J^T \Delta^{-1} J)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \alpha^T G \alpha \right\}$$

However, one can write the above distribution in scalar form. By applying **Lemma 4** which can be found in **Appendix C**, one finds that the joint posterior distribution of σ_1^2, σ_2^2 is equal to:

$$f(\sigma_1^2, \sigma_2^2 | \alpha, \mu, r_1, \dots, r_m) \propto (\sigma_1^2)^{-\frac{n-1}{2}} (\sigma_2^2)^{-\frac{d-n-1}{2}} \exp \left\{ -\frac{1}{2} \alpha^T G \alpha \right\}$$

Furthermore, by applying **Lemma 5** which can also be found in **Appendix C**, we obtain that the scalar version for the equation is:

$$f(\sigma_1^2, \sigma_2^2 | \alpha, \mu, r_1, \dots, r_m) \propto (\sigma_1^2)^{-\frac{n-1}{2}} \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^n (\alpha_i - \bar{\alpha}_v)^2 \right\} \\ (\sigma_2^2)^{-\frac{d-n-1}{2}} \exp \left\{ -\frac{1}{2\sigma_2^2} \sum_{i=n+1}^d (\alpha_i - \bar{\alpha}_c)^2 \right\}$$

Here, $\bar{\alpha}_v$ are the averages of the log of the variance terms and $\bar{\alpha}_c$ are the averages of the log of the covariance terms:

$$\bar{\alpha}_v = \frac{\sum_{i=1}^n \alpha_i}{n} \text{ and } \bar{\alpha}_c = \frac{\sum_{i=n+1}^d \alpha_i}{d-n}$$

Hence, both posteriors of σ_1^2 and σ_2^2 are following Inverse Gamma distributions and they are independent:

$$\boxed{\sigma_1^2 | \alpha, \mu, r_1, \dots, r_m \sim \Gamma^{-1} \left(\frac{n-3}{2}, \frac{1}{2} \sum_{i=1}^n (\alpha_i - \bar{\alpha}_v)^2 \right)}$$

$$\boxed{\sigma_2^2 | \alpha, \mu, r_1, \dots, r_m \sim \Gamma^{-1} \left(\frac{d-n-3}{2}, \frac{1}{2} \sum_{i=n+1}^d (\alpha_i - \bar{\alpha}_c)^2 \right)}$$

We are finally ready to compute the posterior for μ also by collecting the terms that depend on it. We notice that the term obtained from the Volterra integral approximation of the matrix exponential does not show up in the posterior. Therefore, like the posteriors of σ_1^2 and σ_2^2 , this will be an exact distribution. Moreover, we notice that the first two equations in the assumptions of our model (equations (16) and (17)) are the same as when we used an Inverse Wishart prior. Therefore, the derivation for the posterior for μ will be the same, yielding:

$$\boxed{\mu | \alpha, \sigma_1^2, \sigma_2^2, r_1, \dots, r_m \sim N(\mu_{post}, \Sigma_{post})}, \text{ where}$$

$$\boxed{\mu_{post} = (m\Sigma^{-1} + P^T \Omega^{-1} P)^{-1} (m\Sigma^{-1} \bar{r} + P^T \Omega^{-1} q)}$$

$$\boxed{\Sigma_{post} = (m\Sigma^{-1} + P^T \Omega^{-1} P)^{-1}}$$

3.4 Implementation

Now that we have derived our posteriors, we are ready to implement it, using a Gibbs Sampler. The only difference from before is that we will use a Metropolis-Hastings algorithm for sampling α , for which we need the exact posterior distribution. This will be proportional to the distribution obtained from collecting all terms with an α from the joint distribution represented by equation (19):

$$\exp \left\{ -\frac{1}{2} \alpha^T G \alpha \right\} \boxed{\det(S)^{-\frac{m}{2}} \exp \left\{ -\frac{1}{2} (\alpha - \lambda)^T Q (\alpha - \lambda) \right\}}$$

We have seen that it results in the posterior:

$$\alpha | r_1, \dots, r_m, \sigma_1^2, \sigma_2^2, \mu \approx \sim N(\alpha^*, (Q + G)^{-1}), \text{ where } \alpha^* = (Q + G)^{-1} Q \lambda$$

$$\pi^*(\alpha | r_1, \dots, r_m, \sigma_1^2, \sigma_2^2, \mu) \propto \exp \left\{ -\frac{1}{2} (\alpha - \alpha^*)^T (Q + G) (\alpha - \alpha^*) \right\}$$

This is an approximation since the boxed part is an approximation of the pdf of a multivariate normal using the Volterra integral equation. If we replace this with the exact distribution, we would obtain:

$$\pi(\alpha | r_1, \dots, r_m, \sigma_1^2, \sigma_2^2) \propto \exp \left\{ -\frac{m}{2} \text{Trace} (A + S e^{-A}) - \frac{1}{2} \alpha^T G \alpha \right\}$$

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

$$\rho = \frac{\pi(\tilde{\alpha} | r_1, \dots, r_m, \sigma_1^{2(t)}, \sigma_2^{2(t)}, \mu^{(t)})}{\pi(\alpha^{(t)} | r_1, \dots, r_m, \sigma_1^{2(t)}, \sigma_2^{2(t)}, \mu^{(t)})} \cdot \frac{\pi^*(\alpha^{(t)} | r_1, \dots, r_m, \sigma_1^{2(t)}, \sigma_2^{2(t)}, \mu^{(t)})}{\pi^*(\tilde{\alpha} | r_1, \dots, r_m, \sigma_1^{2(t)}, \sigma_2^{2(t)}, \mu^{(t)})}$$

It is useful at this point to remember that because of the notation introduced in 1, we have a connection between π^* and π since there is one between A and α , namely:

$$\alpha = \text{Vec}^*(A)$$

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

Algorithm 3 Gibbs Sampler $\log(\Sigma)$

- 1: $\alpha^{(t+1)} = \begin{cases} \tilde{\alpha} \sim N\left((Q^{(t)} + G^{(t)})^{-1} Q^{(t)} \lambda^{(t)}, (Q^{(t)} + G^{(t)})^{-1}\right) \text{ w.p. } \min(\rho, 1) \\ \alpha^{(t)} \text{ otherwise} \end{cases}$
- 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: $\begin{cases} \sigma_1^{2(t+1)} \sim \Gamma^{-1}\left(\frac{n-3}{2}, \frac{1}{2} \sum_{i=1}^n \left(\alpha_i^{(t+1)} - \overline{\alpha_v^{(t+1)}}\right)^2\right) \\ \sigma_2^{2(t+1)} \sim \Gamma^{-1}\left(\frac{d-n-3}{2}, \frac{1}{2} \sum_{i=n+1}^d \left(\alpha_i^{(t+1)} - \overline{\alpha_c^{(t+1)}}\right)^2\right) \end{cases} \Rightarrow$
 $\Rightarrow \Delta^{(t+1)} = \begin{bmatrix} \sigma_1^{2(t+1)} I_n & \mathbb{O} \\ \mathbb{O} & \sigma_2^{2(t+1)} I_{d-n} \end{bmatrix}$
- 4: Let $\Sigma_\mu = \left(m\Sigma^{(t+1)^{-1}} + P^T \Omega^{-1} P\right)^{-1} \Rightarrow \mu^{(t+1)} \sim N\left(\Sigma_\mu \left(m\Sigma^{(t+1)^{-1}} \bar{r} + P^T \Omega^{-1} q\right), \Sigma_\mu\right)$
- 5: Compute $S^{(t+1)} = \frac{\sum_{i=1}^m (r_i - \mu^{(t+1)}) (r_i - \mu^{(t+1)})^T}{m}$, $\lambda^{(t+1)} = \text{Vec}^*(\log(S^{(t+1)}))$, $d_j^{(t+1)}$ and $e_j^{(t+1)}$ the eigenvalue and normalized eigenvector of $S^{(t+1)}$ respectively.
- 6: Compute $f_{ij}^{(t+1)}$ by identifying the coefficients of the entries of the $\log(\Sigma)$ matrix from the equation $\text{Vec}^*(\log(\Sigma^{(t)})) f_{ij}^{(t+1)} = e_i^{(t+1)} \log(\Sigma^{(t)}) e_j^{(t+1)}$
- 7: Compute $\xi_{ij}^{(t+1)} = \frac{(d_i^{(t+1)} - d_j^{(t+1)})^2}{d_i^{(t+1)} d_j^{(t+1)} (\log(d_i^{(t+1)}) - \log(d_j^{(t+1)}))^2}$
- 8: Compute $Q^{(t+1)} = \frac{m}{2} \sum_{i=1}^n f_{ii}^{(t+1)} f_{ii}^{(t+1)T} + m \sum_{i < j} \xi_{ij}^{(t+1)} f_{ij}^{(t+1)} f_{ij}^{(t+1)T}$
- 9: Compute

$$G^{(t+1)} = \left(I_d - J(J^T \Delta^{(t+1)^{-1}} J)^{-1} J^T \Delta^{(t+1)^{-1}}\right)^T \Delta^{(t+1)^{-1}} \cdot \left(I_d - J(J^T \Delta^{(t+1)^{-1}} J)^{-1} J^T \Delta^{(t+1)^{-1}}\right)$$

3.5 Results

Just like we did before, in this section we will depict the sensitivity of the model to changes in confidence levels (ω_i) in terms of both the distance of the posterior to investor's view and the profits obtained if one would use this model to trade.

Before we delve into the actual results for this version of the model, we notice that remarks (1) and (2) both hold. Basically, this means that as the diagonal entries in Ω get smaller, the more confident we are in the views because we have the assumption that $P\mu \sim N(q, \Omega)$. Same assumption points out the fact that the smaller Ω is, the closer $P\mu$ should be to q . Hence, a very small Ω shows the fact that the investor is very confident in this view and, therefore, the posterior should also be close to q . Therefore, the smaller our Ω is, the closer $P\mu_{\text{post}}$ should be to q . In the first part of this section we will present some plots similar to the ones presented before. We will take 2 views and do an exhaustive search over possible combinations of pairs of values for the 2 diagonal entries of Ω (which are depicted as 2 axis) and compute the same distance as before: $|P\mu_{\text{post}} - q|$ (which is depicted as 1 axis).

We chose the same 4 stocks (AAPL, AMZN, GOOG, MSFT), and we will use the same data set as when we presented the results in section 2.8: daily returns from 1/2/2014 to 12/29/2017. We will use the following inputs (again the columns are in order AAPL, AMZN, GOOG, MSFT and the rows represent the views):

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

Just like when we had a P non-square and an Inverse Wishart prior, in this version of the model, one can use smaller confidence levels than when we were just using an Inverse Wishart prior and the augmented matrix P . This time one can choose ω_i (which were defined as the entries in the main diagonal of Ω) of the order 10^{-7} without getting any numerical issues. For the results presented here, we let (ω_1, ω_2) range between 10^{-6} to 10^{-4} .

However, one can imagine that this approach is more computationally expensive than just having an Inverse Wishart prior on Σ . Therefore, the exhaustive search was run 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:

$$(\omega_1, \omega_2) \in \begin{cases} (10^{-6} \rightarrow 10^{-5}, 10^{-6} \rightarrow 10^{-5}) \\ (10^{-5} \rightarrow 10^{-4}, 10^{-5} \rightarrow 10^{-4}) \\ (10^{-6} \rightarrow 10^{-5}, 10^{-5} \rightarrow 10^{-4}) \\ (10^{-5} \rightarrow 10^{-4}, 10^{-6} \rightarrow 10^{-5}) \end{cases}$$

Each one of those ranges was split into an equally spaced grid of 4^2 points, each one of those being ran on 1 core.

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.

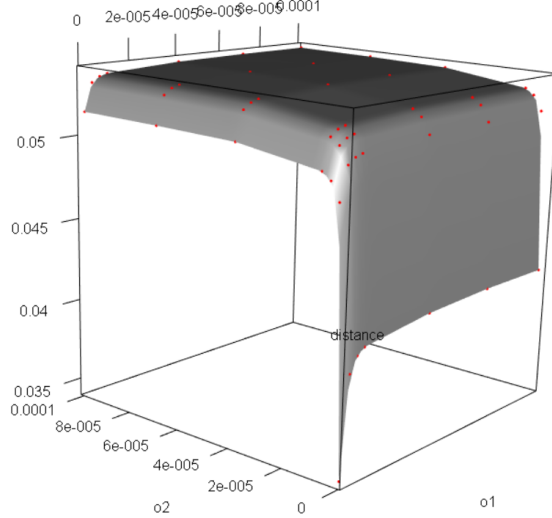


Figure 15: Distances for $\log(\Sigma)$ prior

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{r}) and therefore the distance becomes $|P\bar{r} - q| = 0.05388875$.

We also notice that the second view (corresponding to $o2$) has more influence on the posterior than the first view. This is because the 3D curve would leave a 2D line on a section parallel to the "o2 vs distance" plane that converges to 0 as o2 gets very small much faster than a section parallel to the "o1 vs distance" plane would when o1 gets very small.

We will proceed by looking at profits (losses) that we would obtain by using this model trained on the same daily returns between 1/2/2014 and 12/29/2017. We would estimate using Gibbs Sampling the posterior mean (μ_{post}) and the posterior covariance (Σ_{post}) and we use the CAPM equation (4) to obtain the weights to be $w = \frac{1}{2.5} \Sigma_{post}^{-1} \mu_{post}$. With those weights we compute the profits that we would obtain over the month January 2018 (just like before, daily returns between 1/2/2018 and 1/30/2018) with an initial investment of 100,000\$. Here, one could use a different investment horizon also.

The same P , q , grid for ω_i , burn period, iteration period were used as before. The following is a 3D plot of the sensitivity of the profits to changes in confidence:

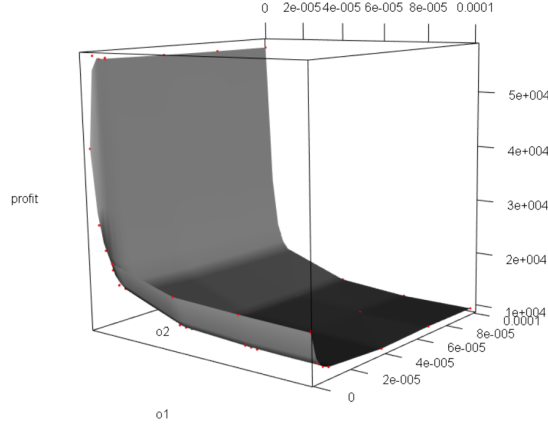


Figure 16: Profits with AMZN in and $q = [0.02, 0.05]^T$

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, $Pr_{Jan2018} = \begin{bmatrix} 0.23996743 \\ 0.01366718 \end{bmatrix}$. Albeit the inputted 1st view is a 10th of what happened in reality (AMZN outper-

formed AAPL by almost 24% in January 2018), the model puts a higher importance on it than on the 2nd view. Indeed, the profits increase drastically as we decrease ω_1 and keep ω_2 constant. Profits do not increase much as we decrease ω_2 and keep ω_1 constant.

Just like we did before, since 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 vary q . In the following 3 figures we will present the results for profits when the investor considers $q = \begin{bmatrix} 0.02 \\ 0.05 \end{bmatrix}$,

$q = \begin{bmatrix} 0.06212815 \\ 0.01366718 \end{bmatrix}$ which is exactly what happened during the month of January 2018 (the "well informed" investor) and $q = \begin{bmatrix} -0.06212815 \\ -0.01366718 \end{bmatrix}$ which is exactly the opposite of what happened during the month of January 2018 (the "poorly informed" investor):

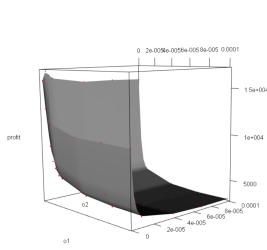


Figure 17: Profits
FB instead of AMZN
and $q = [0.02, 0.05]^T$

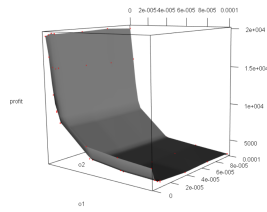


Figure 18: Profits
FB instead of AMZN
and view exactly like
reality

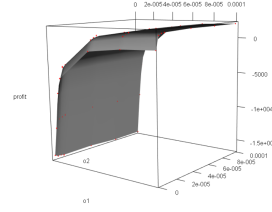


Figure 19: Profits
FB instead of AMZN
and view opposite of
reality

- Since $Pr_{Jan2018} = \begin{bmatrix} 0.06212815 \\ 0.01366718 \end{bmatrix}$, the view in which $q = \begin{bmatrix} 0.02 \\ 0.05 \end{bmatrix}$ has returns that are much closer to what happened in reality than when we had AMZN instead of FB (especially the first view is closer). We notice that the second view has a greater influence on the profits than what we have seen in figure 16 and this can be clearly noticed in figure 17 from above.
- If the investor has a view exactly like the reality (figure 18), the first view has more influence on the profits as ω_1 gets smaller and smaller.
- Moreover, if we compare figures 18 and 19, 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 18 we have a $q = \begin{bmatrix} 0.06212815 \\ 0.01366718 \end{bmatrix}$ and in figure 19 we have a $q = -\begin{bmatrix} 0.06212815 \\ 0.01366718 \end{bmatrix}$.

3.6 Limitations

In the previous section, we haven't presented any results for the whole *S&P500*. This is because we have encountered both memory allocation and running time problems. Both arise from the size of the matrices which makes all matrix computations and sampling from multivariate distributions time consuming. The biggest issue is with the construction of the matrix Q . We remind ourselves that we have to compute f_{ij} by looking at the equation $Vec^*(\log(\Sigma))^T f_{ij} = e_i^T \log(\Sigma) e_j$ and identifying the coefficients of the entries in the $\log(\Sigma)$ matrix. With those f_{ij} , we can finally compute Q :

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

$$Q = \frac{m}{2} \sum_{i=1}^n f_{ii} f_{ii}^T + m \sum_{i < j}^n \xi_{ij} f_{ij} f_{ij}^T$$

It is easy to compute ξ_{ij} and the elegant way to compute the f 's is by coding a 4 way tensor and applying the function $Vec^*(\cdot)$ to 2 of its entries (one can see the pattern more easily by taking a small dimensional example). However, this is not the fastest way since one can actually fill out each entry in Q directly. In both situations, the dimensionality problem still exists. When we take into consideration the whole *S&P500*, the number of rows and columns are of size $d = \frac{500 \cdot 501}{2}$, but since Q is symmetric we would have to store a little more than half of the entries in Q (albeit this approach makes all the formulas in the posterior a lot messier). Even so, the size of such an object is approximately 53 GB. Even with the biggest server at *UCSB*, for which a node has 1 TB of RAM memory, we could only run this in parallel on at most 20 cores.

The memory allocation problem combined with a running time that is a lot bigger than just the 4 hours that took to run the simulations presented in section 2.8 makes this approach computationally not feasible for a large data set.

We have looked at a couple of ideas to remedy the problem:

- Writing the matrix Q to the disk. Unfortunately, one would need a high speed connection (for example SSD) to be able to write it fast enough that it doesn't make the running time even longer. This is of paramount importance since we have to compute Q at each iteration of the Gibbs Sampler.
- We have looked at parallelizing the Gibbs Sampler itself (which is a Markov Chain). More precisely, in the general setting of Markov Chains, we have looked at independently starting at m initial points and, from each initial point, starting independent Markov Chains. It has been shown[1] that for one single Markov Chain that satisfies Doob's conditions, the ergotic average converges geometrically:

$$P\left(\frac{1}{n}\sum_{k=1}^n f(X_k) > \epsilon \middle| X_0 = x_0\right) \leq A(\epsilon)\rho(\epsilon)^n, \text{ where}$$

$$(\exists)d_0, t_0 \text{ s.t. } \rho(\epsilon) = \Phi(d_0, t_0)^{\frac{1}{d_0}} + \eta \text{ with } \eta \text{ s.t. } \rho(\epsilon) < 1,$$

$$\Phi(d_0, t_0) = \sup_{x_0} E\left[e^{t_0 \sum_{k=1}^{d_0} f(X_k)} \middle| x_0\right]$$

By using this result, one can easily show that for running m Markov Chains in parallel we obtain the following bound:

$$P\left(\frac{1}{m}\sum_{i=1}^m \frac{1}{n}\sum_{k=1}^n f(X_{ik}) > \epsilon \middle| x_0\right) \leq e^{-t_0^* mn \epsilon} A^*(\epsilon)^m \rho^*(\epsilon)^{mn}$$

Here, the existence of d_0^*, t_0^* and the definitions of $A^*(\cdot), \rho^*(\cdot)$ are in the same way as before. The problem is that we cannot compare the right hand sides of the 2 inequalities from above because the $A(\cdot), A^*(\cdot)$ and $\rho(\cdot), \rho^*(\cdot)$ are different since this is a proof of existence.

3.7 Current Work

The running time and memory allocation problems encountered when using the whole market would suggest that one has to reduce the dimensionality. Moreover, there is a strong connection between the original Black-Litterman model and CAPM (which can be seen as a factor analysis model in statistics). This gave us the idea of adding a fully Bayesian specified factor model to the Bayesian alternatives presented in this paper. All the posteriors have already been derived for those.

A

Proof of Original Approach

As we have seen, the original model is represented by the following 3 distributions, where the last 2 are priors:

$$\begin{aligned} r &\sim N(\mu, \Sigma) \\ \mu &\sim N(\pi, \tau\Sigma) \\ q|\mu &\sim N(P\mu, \Omega) \end{aligned}$$

The last 2 equations are combined. In order to get the joint likelihood, we would have to multiply the probability density functions of the last 2 distributions:

$$\begin{aligned} f(\mu, q) &\propto \exp \left\{ -\frac{1}{2} [(\mu - \pi)^T (\tau\Sigma)^{-1} (\mu - \pi) + (q - P\mu)^T \Omega^{-1} (q - P\mu)] \right\} = \\ &= \exp \left\{ -\frac{1}{2} \begin{bmatrix} \mu - \pi \\ q - P\mu \end{bmatrix}^T \begin{bmatrix} \tau\Sigma & 0 \\ 0 & \Omega \end{bmatrix}^{-1} \begin{bmatrix} \mu - \pi \\ q - P\mu \end{bmatrix} \right\} \end{aligned}$$

$$\text{Let } V = \begin{bmatrix} \tau\Sigma & 0 \\ 0 & \Omega \end{bmatrix} \text{ and } \alpha = \begin{bmatrix} \mu - \pi \\ q - P\mu \end{bmatrix}.$$

We will try to find a matrix A such that when we compute $\alpha' = A\alpha$ and $V' = AVA^T$ we will obtain that the joint distribution from above is equal to $\exp \left\{ -\frac{1}{2} \alpha'^T V'^{-1} \alpha' \right\}$.

Let

$$\begin{aligned} A &= \begin{bmatrix} I_n & -((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1} P^T\Omega^{-1} \\ 0 & I_k \end{bmatrix} \begin{bmatrix} I_n & 0 \\ P & I_k \end{bmatrix} = \\ &= \begin{bmatrix} I_n - ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1} P^T\Omega^{-1}P & -((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1} P^T\Omega^{-1} \\ P & I_k \end{bmatrix} \end{aligned}$$

But since $I_n = ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1} ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)$, we obtain that the first entry in the above matrix can be written as

$$\begin{aligned} ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1} ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P - P^T\Omega^{-1}P) &= \\ ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1} (\tau\Sigma)^{-1} & \end{aligned}$$

Hence, we can get the final form of A is:

$$A = \begin{bmatrix} ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1} (\tau\Sigma)^{-1} & -((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1} P^T\Omega^{-1} \\ P & I_k \end{bmatrix}$$

Note: $\det(A) = 1$. The determinant of a 2×2 block matrix is computed using the same formula as that for a normal 2×2 matrix, but we have to be careful when we look at the order in which we multiply:

$$\begin{aligned} \det(A) &= \det\{((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1}(\tau\Sigma)^{-1} + \\ &\quad + ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1}P^T\Omega^{-1}P\} = \\ &= \det\{((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1}((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)\} = \det(I) = 1 \end{aligned}$$

Finally, we are ready to find our α' and V' :

$$\begin{aligned} \alpha' &= A\alpha = A \begin{bmatrix} \mu - \pi \\ q - P\mu \end{bmatrix} = \\ &= \begin{bmatrix} ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1}(\tau\Sigma)^{-1}(\mu - \pi) - ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1}P^T\Omega^{-1}(q - P\mu) \\ P\mu - P\pi + q - P\mu \end{bmatrix} \end{aligned}$$

Now by factoring the common term in the 1st entry of the vector, we obtain:

$$\begin{aligned} &((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1}((\tau\Sigma)^{-1}\mu - (\tau\Sigma)^{-1}\pi - P^T\Omega^{-1}q + P^T\Omega^{-1}P\mu) = \\ &= ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1}[(\tau\Sigma)^{-1} + P^T\Omega^{-1}P]\mu - ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1}P^T\Omega^{-1}q = \\ &= \mu - ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1}((\tau\Sigma)^{-1}\pi - P^T\Omega^{-1}q) \end{aligned}$$

And if we let $\bar{\mu} = ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1}((\tau\Sigma)^{-1}\pi - P^T\Omega^{-1}q)$, we obtain that

$$\alpha' = A\alpha = \begin{bmatrix} \mu - \bar{\mu} \\ q - P\pi \end{bmatrix} \quad (20)$$

Now we are ready to move to the computation of V' :

$$\begin{aligned} V' &= AVA^T = \begin{bmatrix} ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1} & -((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1}P^T \\ \tau P\Sigma & \Omega \end{bmatrix} A^T \\ &= \begin{bmatrix} ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1}(\tau\Sigma)^{-1} & P^T \\ -((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1}P^T\Omega^{-1} & I_k \end{bmatrix} A^T \end{aligned}$$

And now, in order to compute A^T we can use the fact that for any 2 matrices $(AB)^T = B^T A^T$ and also the fact that Σ is symmetric, hence Σ^{-1} is also symmetric. Moreover, the same reasoning can be done for Ω . Hence, $(P^T\Omega^{-1}P)^T = P^T\Omega^{-1}P$. Now if we look at the first entry

first column in the 2×2 block matrix, we notice that it is actually equal to $(\tau\Sigma)^{-1} ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1}$. Similarly the second row first column in the block matrix is actually $-\Omega^{-1}P ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1}$.

Finally, we can compute V' . The first entry in the first column will be:

$$\begin{aligned} & ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1} (\tau\Sigma)^{-1} ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1} + \\ & + ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1} P^T\Omega^{-1}P ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1} = \\ & = ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1} ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P) ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1} = \\ & = ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1} \end{aligned}$$

Hence, the whole matrix V' will be:

$$V' = \begin{bmatrix} ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1} & 0 \\ 0 & \Omega + \tau P\Sigma P^T \end{bmatrix}$$

Now that we have both the V' and α' , it is a lot easier to notice that the joint will be:

$$\begin{aligned} f(q, \mu) & \propto \exp \left\{ -\frac{1}{2} \alpha^T V^{-1} \alpha \right\} = \exp \left\{ -\frac{1}{2} \alpha'^T A^{-T} A^T V'^{-1} A A^{-1} \alpha' \right\} = \\ & = \exp \left\{ -\frac{1}{2} \alpha'^T V'^{-1} \alpha' \right\} \end{aligned}$$

But this is the kernel of a multivariate normal. Also, from equation (20) and from the above form of the matrix V' , we obtain that :

$$\begin{aligned} \mu & \sim N \left(\bar{\mu}, ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1} \right), \\ \bar{\mu} & = ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1} ((\tau\Sigma)^{-1}\pi - P^T\Omega^{-1}q) \end{aligned}$$

If we denote by M^{-1} the covariance matrix and we take the above as been the new prior on the expected returns: $r|\mu \sim N(\mu, \Sigma)$, we obtain that the joint is:

$$\begin{aligned} f(r, \mu) & \propto \exp \left\{ -\frac{1}{2} ((\mu - \bar{\mu})^T M(\mu - \bar{\mu}) + (r - \mu)^T \Sigma^{-1}(r - \mu)) \right\} = \\ & = \exp \left\{ -\frac{1}{2} ((\mu - \bar{\mu})^T M(\mu - \bar{\mu}) + (\mu - r)^T \Sigma^{-1}(\mu - r)) \right\} \end{aligned}$$

Using lemma 2 with $y = \mu, a = \bar{\mu}, b = r, A = M, B = \Sigma^{-1}$, we can conclude that:

$$f(r, \mu) \propto \exp \left\{ -\frac{1}{2}(\mu - \mu^*)^T (M + \Sigma^{-1})(\mu - \mu^*) \right\} \cdot \exp \left\{ (r - \bar{\mu})^T \left(M^{-1} + (\Sigma^{-1})^{-1} \right)^{-1} (r - \bar{\mu}) \right\}$$

Hence the posterior of the returns follows also a normal distribution:

$$r \sim N(\bar{\mu}, M^{-1} + \Sigma)$$

B

Proof of Approximation using Volterra Integrals

As mentioned before, Bellman in his book *Introduction to Matrix Analysis* shows an even more general result than what we need. The matrix exponential $x(t) = e^{(A_0 + cB_0)t}$ satisfies the Volterra integral equation:

$$X(t) = e^{A_0 t} + c \int_0^t e^{A_0(t-s)} B_0 X(s) ds, 0 < t < \infty$$

Now if we let in the above equation $A_0 = -\Lambda$, $B_0 = \Lambda - A$, $c = 1$ and remembering that $\Lambda = \log(S)$ we obtain:

$$X(t) = S^{-t} - \int_0^t S^{s-t} (A - \Lambda) X(v) dv, 0 < t < \infty,$$

Since we want to approximate e^{-A} , we let in the above equation $t = 1$ and we repeatedly replace X :

$$\begin{aligned} e^{-A} &= X(1) = S^{-1} - \int_0^1 S^{s-1} (A - \Lambda) S^{-s} ds = \\ &= S^{-1} - \int_0^1 S^{s-1} (A - \Lambda) \left(S^{-s} - \int_0^s S^{u-s} (A - \Lambda) X(u) du \right) ds = \\ &= S^{-1} - \int_0^1 S^{s-1} (A - \Lambda) S^{-s} ds + \int_0^1 \int_0^s S^{s-1} (A - \Lambda) S^{u-s} (A - \Lambda) \cdot \\ &\quad \cdot \left(S^{-u} - \int_0^u S^{v-u} (A - \Lambda) X(v) dv \right) dud s \approx \\ &\approx S^{-1} - \int_0^1 S^{s-1} (A - \Lambda) S^{-s} ds + \int_0^1 \int_0^s S^{s-1} (A - \Lambda) S^{u-s} (A - \Lambda) S^{-u} dud s \end{aligned}$$

Where this is an approximation because the triple and higher order integrals were ignored. The conditional pdf of the returns is:

$$f(r_1, \dots, r_m | \mu, \Sigma) \propto \exp\left\{-\frac{m}{2} \text{Tr}(A + Se^{-A})\right\}$$

Hence, from the Volterra approximation, by multiplying by S and taking the trace, we obtain:

$$\begin{aligned} \text{Tr}(Se^{-A}) &\approx n - \int_0^1 \text{Tr}(S^s(A - \Lambda)S^{-s}) ds + \\ &+ \int_0^1 \int_0^s \text{Tr}(S^s(A - \Lambda)S^{u-s}(A - \Lambda)S^{-u}) du ds \end{aligned}$$

The first integral is easier to compute:

$$\int_0^1 \text{Tr}(S^s(A - \Lambda)S^{-s}) ds = \int_0^1 \text{Tr}(A - \Lambda) ds = \text{Tr}(A - \Lambda)$$

The second integral requires more calculations. Before we delve into them, let us write the spectral decomposition of S as $S = E_0 D_0 E_0^T$. If we define the matrix \log through the Taylor series expansion, and by using the fact that E_0 is orthonormal, we obtain that the spectral decomposition of $\log(S)$ is $\Lambda = \log(S) = E_0 \log(D_0) E_0^T$. Also, let us make another notation: $B = E_0^T(A - \Lambda)E_0 \Rightarrow E_0 B E_0^T = A - \Lambda$:

$$\begin{aligned} \text{Tr}(S^s(A - \Lambda)S^{u-s}(A - \Lambda)S^{-u}) &= \text{Tr}\left((A - \Lambda)S^{u-s}(A - \Lambda)S^{-(u-s)}\right) = \\ &= \text{Tr}\left(E_0 B D_0^{u-s} B D_0^{-(u-s)} E_0^T\right) = \text{Tr}\left(B D_0^{u-s} B D_0^{-(u-s)}\right) \end{aligned}$$

In order to compute the integral of this *Trace* term, we will try to put it in scalar form:

$$B D_0^{u-s} = \begin{bmatrix} b_{11}d_1^{u-s} & b_{12}d_2^{u-s} & \dots & b_{1n}d_n^{u-s} \\ \vdots & \vdots & \dots & \vdots \\ b_{n1}d_1^{u-s} & b_{n2}d_2^{u-s} & \dots & b_{nn}d_n^{u-s} \end{bmatrix}$$

For the matrix $B D_0^{-(u-s)}$ we obtain a similar result, the only difference is that d_i^{u-s} is replaced by $\frac{1}{d_i^{u-s}}$. Also, from the spectral decomposition, please note that d_i are the eigenvalues of S .

Since we need the $Tr \left(BD_0^{u-s} BD_0^{-(u-s)} \right)$, we will only compute the diagonal entries of this matrix:

$$\begin{aligned} & \text{diag} \left(BD_0^{u-s} BD_0^{-(u-s)} \right) = \\ & \begin{bmatrix} b_{11}^2 + b_{12}b_{21} \left(\frac{d_2}{d_1} \right)^{u-s} + b_{13}b_{31} \left(\frac{d_3}{d_1} \right)^{u-s} + \cdots + b_{1n}b_{n1} \left(\frac{d_n}{d_1} \right)^{u-s} \\ b_{21}b_{12} \left(\frac{d_1}{d_2} \right)^{u-s} + b_{22}^2 + \cdots + b_{2n}b_{n2} \left(\frac{d_n}{d_2} \right)^{u-s} \\ \vdots \\ b_{n1}b_{1n} \left(\frac{d_n}{d_1} \right)^{u-s} + b_{n2}b_{2n} \left(\frac{d_n}{d_2} \right)^{u-s} + \cdots + b_{nn}^2 \end{bmatrix} \end{aligned}$$

But we know that B is symmetric. Therefore, we obtain that:

$$\begin{aligned} & \int_0^1 \int_0^s Tr \left(BD_0^{u-s} BD_0^{-(u-s)} \right) dud s = \sum_{i=1}^n \int_0^1 \int_0^s b_{ii}^2 dud s + \\ & + \sum_{i \neq j}^n \int_0^1 \int_0^s b_{ij}^2 \left(\frac{d_i}{d_j} \right)^{u-s} dud s, \text{ where we have that} \\ & \sum_{i=1}^n \int_0^1 \int_0^s b_{ii}^2 dud s = \sum_{i=1}^n \frac{b_{ii}^2}{2} \text{ and also} \\ & \sum_{i \neq j}^n \int_0^1 \int_0^s b_{ij}^2 \left(\frac{d_i}{d_j} \right)^{u-s} dud s = \sum_{i \neq j}^n \int_0^1 b_{ij}^2 \left(\frac{d_i}{d_j} \right)^{u-s} \cdot \\ & \cdot \frac{1}{\log(d_i) - \log(d_j)} \Big|_0^s ds = \sum_{i \neq j} \frac{b_{ij}^2}{\log(d_i) - \log(d_j)} \int_0^1 1 - \left(\frac{d_i}{d_j} \right)^{-s} ds = \\ & = \sum_{i \neq j} \frac{b_{ij}^2}{\log(d_i) - \log(d_j)} \left(1 - \left(\frac{d_j}{d_i} \right)^s \frac{1}{\log(d_j) - \log(d_i)} \right) \Big|_0^1 = \\ & = \sum_{i \neq j} \frac{b_{ij}^2}{\log(d_i) - \log(d_j)} + \sum_{i \neq j} b_{ij}^2 \frac{\frac{d_j}{d_i} - 1}{(\log(d_i) - \log(d_j))^2} = \\ & = \sum_{i < j} \left(\frac{b_{ij}^2}{\log(d_i) - \log(d_j)} + \frac{b_{ji}^2}{\log(d_j) - \log(d_i)} \right) + \sum_{i < j} b_{ij}^2 \frac{\frac{d_j}{d_i} + \frac{d_i}{d_j} - 2}{(\log(d_i) - \log(d_j))^2} = \\ & = 0 + \sum_{i < j} b_{ij}^2 \frac{\frac{d_j}{d_i} + \frac{d_i}{d_j} - 2}{(\log(d_i) - \log(d_j))^2} \end{aligned}$$

Finally, by adding the two double integrals, we obtain that

$$\begin{aligned} \int_0^1 \int_0^s \text{Tr} \left(B D_0^{u-s} B D_0^{-(u-s)} \right) du ds &= n - \text{Tr}(A) + \text{Tr}(\Lambda) + \\ &+ \frac{1}{2} \sum_{i=0}^n b_{ii}^2 + \sum_{i < j} b_{ij}^2 \frac{\frac{d_j}{d_i} + \frac{d_i}{d_j} - 2}{(\log(d_i) - \log(d_j))^2} \end{aligned}$$

With the notation of the ξ_{ij} introduced in the paper, we obtain the Volterra approximation represented by equation (15).

C

Proof of Proposition 1

The following equality holds:

$$\begin{aligned} f(\alpha | \sigma_1^2, \sigma_2^2) &= \int_{\theta} \det(\Delta)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\alpha - J\theta)^T \Delta^{-1} (\alpha - J\theta) \right\} d\theta = \\ &= 2\pi \det(\Delta)^{-\frac{1}{2}} \det(J^T \Delta^{-1} J)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \alpha^T G \alpha \right\}, \text{ where} \\ G &= (I_d - J(J^T \Delta^{-1} J)^{-1} J^T \Delta^{-1})^T \Delta^{-1} (I_d - J(J^T \Delta^{-1} J)^{-1} J^T \Delta^{-1}) \end{aligned}$$

Proof. Before we actually attempt to compute the integral, we would like to put all the quantities in scalar form since this would make our life easier. This brings us to the following two lemmas:

Lemma 4. $\det(\Delta)^{-\frac{1}{2}} \det(J^T \Delta^{-1} J)^{-\frac{1}{2}} = \frac{1}{\sqrt{n(d-n)}} (\sigma_1^2)^{-\frac{n-1}{2}} (\sigma_2^2)^{-\frac{d-n-1}{2}}$

Proof.

$$J^T \Delta^{-1} J = \begin{bmatrix} \frac{1}{\sigma_1^2} & \dots & \frac{1}{\sigma_1^2} & 0 & \dots & 0 \\ 0 & \dots & 0 & \frac{1}{\sigma_2^2} & \dots & \frac{1}{\sigma_2^2} \end{bmatrix} J = \begin{bmatrix} \frac{n}{\sigma_1^2} & 0 \\ 0 & \frac{d-n}{\sigma_2^2} \end{bmatrix}$$

Hence, we obtain that $\det(J^T \Delta^{-1} J)^{-\frac{1}{2}} = \frac{1}{\sqrt{n(d-n)}} (\sigma_1^2)^{\frac{1}{2}} (\sigma_2^2)^{\frac{1}{2}}$ Also, clearly since Δ is diagonal, we obtain that:

$$\det(\Delta)^{-\frac{1}{2}} = (\sigma_1^2)^{-\frac{n}{2}} (\sigma_2^2)^{-\frac{d-n}{2}}$$

Multiplying the two determinants, we obtain the desired result. □

Now let us turn our attention to writing in scalar form the term in the exponential:

Lemma 5. $\alpha^T G \alpha = \frac{1}{\sigma_1^2} \sum_{i=1}^n (\alpha_i - \bar{\alpha}_v)^2 + \frac{1}{\sigma_2^2} \sum_{i=n+1}^d (\alpha_i - \bar{\alpha}_c)^2$, where $\bar{\alpha}_v$ is the average of the α 's on the main diagonal (i.e. those that originate from the log of the variance terms of the returns) and $\bar{\alpha}_c$ is the average of all the α 's that are on the off diagonal (i.e. those that originate from the log of the covariance terms of the returns).

Proof. First of all, one can notice that the formula for G can be simplified for calculation purposes:

$$\begin{aligned} G &= (I_d - J(J^T \Delta^{-1} J)^{-1} J^T \Delta^{-1})^T \Delta^{-1} (I_d - J(J^T \Delta^{-1} J)^{-1} J^T \Delta^{-1}) = \\ &= \Delta^{-1} - \Delta^{-1} J(J^T \Delta^{-1} J)^{-1} J^T \Delta^{-1} - \Delta^{-1} J(J^T \Delta^{-1} J)^{-1} J^T \Delta^{-1} + \\ &\quad + \Delta^{-1} J(J^T \Delta^{-1} J)^{-1} J^T \Delta^{-1} = \Delta^{-1} - \Delta^{-1} J(J^T \Delta^{-1} J)^{-1} J^T \Delta^{-1} \end{aligned}$$

We remember that we have computed $J^T \Delta^{-1} J$ in lemma 4:

$$\begin{aligned} J^T \Delta^{-1} J &= \begin{bmatrix} \frac{n}{\sigma_1^2} & 0 \\ 0 & \frac{d-n}{\sigma_2^2} \end{bmatrix} \text{ and } \Delta^{-1} J = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 \\ \vdots & \vdots \\ \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \\ \vdots & \vdots \\ 0 & \frac{1}{\sigma_2^2} \end{bmatrix} \Rightarrow \\ \Rightarrow \Delta^{-1} J (J^T \Delta^{-1} J) &= \begin{bmatrix} \frac{1}{n} & 0 \\ \vdots & \vdots \\ \frac{1}{n} & 0 \\ 0 & \frac{1}{d-n} \\ \vdots & \vdots \\ 0 & \frac{1}{d-n} \end{bmatrix} \Rightarrow \\ \Rightarrow \Delta^{-1} J (J^T \Delta^{-1} J) J^T \Delta^{-1} &= \begin{bmatrix} \frac{1}{n\sigma_1^2} & \cdots & \frac{1}{n\sigma_1^2} & 0 & \cdots & 0 \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ \frac{1}{n\sigma_1^2} & \cdots & \frac{1}{n\sigma_1^2} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \frac{1}{(d-n)\sigma_2^2} & \cdots & \frac{1}{(d-n)\sigma_2^2} \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & 0 & \frac{1}{(d-n)\sigma_2^2} & \cdots & \frac{1}{(d-n)\sigma_2^2} \end{bmatrix} \end{aligned}$$

Now we just have to subtract this matrix from Δ^{-1} , which is just diagonal, and we can finally compute the desired quantity:

$$\begin{aligned}\alpha^T G \alpha = & \sum_{1 \leq i \neq j \leq n} \frac{1}{n\sigma_1^2} \alpha_i \alpha_j + \sum_{i=1}^n \frac{n-1}{n\sigma_1^2} \alpha_i^2 + \sum_{n+1 \leq i \neq j \leq d} \frac{1}{(d-n)\sigma_2^2} \alpha_i \alpha_j + \\ & + \sum_{i=n+1}^d \frac{d-n-1}{(d-n)\sigma_2^2} \alpha_i^2\end{aligned}$$

By looking at this equation and the one that we have to prove, we realize that if we would manage to show the following identity, we would also prove the lemma:

$$\sum_{1 \leq i \neq j \leq n} \frac{1}{n\sigma_1^2} \alpha_i \alpha_j + \sum_{i=1}^n \frac{n-1}{n\sigma_1^2} \alpha_i^2 = \frac{1}{\sigma_1^2} \sum_{i=1}^n (\alpha_i - \bar{\alpha}_v)^2$$

Let us start from the right hand side:

$$\begin{aligned}\sum_{i=1}^n (\alpha_i - \bar{\alpha}_v)^2 &= \sum_{i=1}^n \alpha_i^2 - n\bar{\alpha}^2 = \sum_{i=1}^n \alpha_i^2 - \frac{1}{n} \left(\sum_{i=1}^n \alpha_i \right)^2 = \\ &= \sum_{i=1}^n \alpha_i^2 - \frac{1}{n} \sum_{i=1}^n \alpha_i^2 - \frac{1}{n} \sum_{i \neq j} \alpha_i \alpha_j = \sum_{i=1}^n \frac{n-1}{n} \alpha_i^2 - \frac{1}{n} \sum_{i \neq j} \alpha_i \alpha_j\end{aligned}$$

□

Now we finally have all the necessary identities to write the integral in our proposition in scalar form. We would have to prove that:

$$\begin{aligned}& \int_{\theta_1} \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^n (\alpha_i - \theta_1)^2 \right\} d\theta_1 \int_{\theta_2} \exp \left\{ -\frac{1}{2\sigma_2^2} \sum_{i=n+1}^q (\alpha_i - \theta_2)^2 \right\} d\theta_2 = \\ &= 2\pi \frac{\sigma_1}{\sqrt{n}} \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^n (\alpha_i - \bar{\alpha}_v)^2 \right\} \frac{\sigma_2}{\sqrt{d-n}} \exp \left\{ -\frac{1}{2\sigma_2^2} \sum_{i=n+1}^q (\alpha_i - \bar{\alpha}_c)^2 \right\}\end{aligned}$$

Hence, if we manage to show the following identity, we would manage to prove the proposition also:

$$\int_{\theta_1} \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^n (\alpha_i - \theta_1)^2 \right\} d\theta_1 = \sqrt{2\pi} \frac{\sigma_1}{\sqrt{n}} \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^n (\alpha_i - \bar{\alpha}_v)^2 \right\}$$

Let us start from the left hand side and subtract and add the average $\overline{\alpha_v}$ in each term of the sum from the exponential:

$$\begin{aligned}
LHS &= \int_{\theta_1} \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^n (\alpha_i - \theta_1)^2 \right\} d\theta_1 = \\
&= \int_{\theta_1} \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^n (\alpha_i - \overline{\alpha_v} + \overline{\alpha_v} - \theta_1)^2 \right\} d\theta_1 = \\
&= \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^n (\alpha_i - \overline{\alpha_v})^2 \right\} \cdot \\
&\cdot \int_{\theta_1} \exp \left\{ -\frac{1}{2\sigma_1^2} \left(\sum_{i=1}^n (\overline{\alpha_v} - \theta_1)^2 + 2 \sum_{i=1}^n (\alpha_i - \overline{\alpha_v})(\overline{\alpha_v} - \theta_1) \right) \right\} d\theta_1 = \\
&= \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^n (\alpha_i - \overline{\alpha_v})^2 \right\} \cdot \\
&\cdot \int_{\theta_1} \exp \left\{ -\frac{1}{2\sigma_1^2} \left(n(\overline{\alpha_v} - \theta_1)^2 + 2(\overline{\alpha_v} - \theta_1) \sum_{i=1}^n (\alpha_i - \overline{\alpha_v}) \right) \right\} d\theta_1 = \\
&= \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^n (\alpha_i - \overline{\alpha_v})^2 \right\} \int_{\theta_1} \exp \left\{ -\frac{1}{2 \left(\frac{\sigma_1}{\sqrt{n}} \right)^2} (\overline{\alpha_v} - \theta_1)^2 \right\} d\theta_1
\end{aligned}$$

Now we recognize that the term inside the integral is close to the density of a normal distribution. Hence, this gives us the idea of doing the change of variables:

$$\begin{aligned}
y_1 &= \frac{\theta_1 - \overline{\alpha_v}}{\frac{\sigma_1}{\sqrt{n}}} \Rightarrow dy_1 = \frac{\sqrt{n}}{\sigma_1} d\theta_1 \Rightarrow d\theta_1 = \frac{\sigma_1}{\sqrt{n}} dy_1 \Rightarrow \\
\Rightarrow LHS &= \sqrt{2\pi} \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^n (\alpha_i - \overline{\alpha_v})^2 \right\} \int_{\theta_1} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} y_1^2 \right\} \frac{\sigma_1}{\sqrt{n}} dy_1 = \\
&= \sqrt{2\pi} \frac{\sigma_1}{\sqrt{n}} \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^n (\alpha_i - \overline{\alpha_v})^2 \right\}
\end{aligned}$$

As mentioned before, a similar identity can be showed for the second integral that depends solely on θ_2 and this completes the proof of the proposition.

References

- [1] Baum, L.E., Katz, M. and Read, R.R. (Feb 1962) *Convergence rates for the Law of Large Numbers*. American Mathematical Society, Vol. 102, No. 2, pp. 187-199
- [2] Bellman, R. (1970) *Introduction to Matrix Analysis*. McGraw-Hill, New York.
- [3] Guangliang, H. and Litterman, R. (Feb. 2017) *The Intuition Behind Black-Litterman Model Portfolios*. SSRN, 22 Oct. 2002, Web.
- [4] Janecek, K. (2004) *What is a Realistic Aversion to Risk for Real-world Individual Investors*. Semanticsscholar, Web.
- [5] Leonard, T. and Hsu, John S.J. (Dec. 1992) *Bayesian Inference for a Covariance Matrix*. The Annals of Statistics, vol. 20, no. 4, pp. 1669-1696
- [6] Leonard, T. and Hsu, John S.J. (1999) *Bayesian Methods: An Analysis for Statisticians and Interdisciplinary Researchers*. Cambridge UK, Cambridge UP.

□