Stockholms universitet

Bayesian Estimation of optimal portfolio: Theory and Practice

Theodoros Theologitis

Masteruppsats i matematisk statistik
Master Thesis in Mathematical Statistics

Masteruppsats 2016:8
Matematisk statistik
December 2016

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Matematiska institutionen

Stockholm University

Mathematical Statistics
Stockholm University
Master Thesis **2016:8**
http://www.math.su.se

Bayesian Estimation of optimal portfolio: Theory and Practice

Theodoros Theologitis*

December 2016

Abstract

In this thesis, we consider the estimation of the weights of the minimum variance portfolio from a Bayesian point of view. Standard methods, which are based on determining sample moments, contain estimation errors which are not taken into consideration in the investment process. This often leads to a false asset allocation and might cause extreme risks to the investor. The Bayesian framework accounts for those errors by treating the parameters as random variables. Four different prior models will be used and compared through their posterior distributions and the point estimates of their posterior means. A simulation study is performed in order to test the point estimates of the posterior means through the L_2 deviations from the means of a true model. The data are generated from the multivariate normal distribution. In an empirical study we analyze the posterior distributions of the weights under all considered models through a domestic portfolio. Given the posterior distributions we can make probabilistic statements for the weights by creating the credible intervals and calculating posterior probabilities.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: theologitis.t@gmail.com. Supervisor: Taras Bodnar.

Acknowledgements

I would like to express my gratitude to my supervisor, Associate Professor Taras Bodnar (University of Stockholm), for his invaluable guidance and support through the writing of this thesis. Without him, this project would not have been completed.

I would like to thank my family, Katerina and Kiki, for always being there for me.

Last but not least, I would like to thank Melina for her continuous encouragement and for making my time in Sweden more pleasant. Special thanks go to my good friends in Greece.

Contents

List of Abbreviations

IW	Inverse Wishart
MN	Multivariate Normal
MLE	Maximum Likelihood Estimation/Estimator
MVP	Minimum Variance Portfolio

List of Figures

List of Tables

Chapter 1

Introduction

When Harry Markowitz set the fundamentals of the *Modern Portfolio Theory* (see Markowitz 1952), above all, he stressed the importance of portfolio diversification. His methodology gave us the tools on how to derive both the set of portfolios that have the minimum possible variance (minimum risk) and those sets that lie on the *efficient frontier*, e.g for a given level of risk there is a certain value of expected return. Strictly speaking, those ideas constitute an integral part of what is called as *Mean-Variance Analysis* which assumes that an investor, when choosing a portfolio, only cares about expected return and risk. Furthermore, an investor is assumed to be risk averse meaning that his goal is to minimize the variance of his portfolio return for a higher expected return.

However, the *Mean Variance* framework is too generic since everything should work well if only we "knew" exactly the parameters (mean and covariance matrix) of the distributions of asset returns. In practice, this information is not available, since future data can be very different from what we have observed, and by replacing the unknown parameters with their estimates, this is where estimation risk comes into play.

Michaud (1989) states that securities with extreme expected returns, correlations and variances are more likely to be either overestimated or underestimated in the classical *Mean Variance* optimization procedure ("estimation-error maximizers"). In addition, Stein (1956) showed that the sample mean is not a good estimator for n-variate problems when $n > 2$. Following Steins' statement, Merton (1980) showed that estimates of covariances from historical data are more accurate over time than estimates of the expected returns. Chopra et al. (1993) conducted three tests by slightly modifying the true parameters (means and variances) and found that errors in means are approximately 11 times more important than errors in variances.

As multiple sources of financial literature have discussed and stressed the importance of estimation risk, several approaches have been used to account for the problem. Portfolio resampling is one of them. Jorion (1992) performed a simulation study in which he calculated the sample means and sample covariance matrix from historical data and based on these estimates he simulated numerous samples from *multivariate normal* distribution. Those samples were used to represent observed data and by calculating again the sample means and covariance matrices he constructed multiple sets of means and covariance matrices and, thus, created a distribution of optimal weights.

There are also robust approaches to portfolio selection in which the utility func-

tion was formed in such a way to incorporate explicitly estimation risk (see Goldfarb et al. 2003) in the optimization process.

On the other hand, the Bayesian approach accounts for estimation risk since the unknown parameters are treated as random variables. A prior belief for the unknown parameters blended with the observed data produce an entire distribution. From a decision-making point of view, it is not unlikely that an investor has a specific belief about the optimal portfolio allocation. In addition, equally weighted portfolio seems to have good out-of-sample performance (see DeMiguel et al. 2009).

This paper focuses on the estimation of a special case of the mean-variance optimal portfolios, known as the *Minimum Variance Portfolio*. The Bayesian framework and the analytic derivations of the posterior distributions of the weights from Bodnar et al. (2017) are used. In total, four different models are analyzed. In chapter 2, a general background about some important concepts in portfolio theory is presented. At the end of the chapter, MVP derivation is provided based on the classical mean-variance optimization problem. In chapter 3, we go through some aspects of Bayesian estimation and introduce the priors that will be used throughout the paper. Chapter 4 serves as a simulation study in order to test the point estimates of the posterior means of the weights under all considered priors through the L_2 loss function. The data used to perform the study were simulated from a conjugate and an informative prior. In chapter 5, a portfolio consisted of five stocks that are traded in the Stockholm stock exchange is used and analyzed within the Bayesian framework. Through the simulation study, the aim of this paper is to test the performance of the mean estimates of the posterior distributions under different assumptions. Within the empirical study, the goal is to check how much information we manage to incorporate through the informative priors, which assume specific prior beliefs about the optimal allocation or based on historical estimates, compared to uninformative priors, which only assume some vague information about the parameters and let, mainly, the observed data affect the posterior distribution (see Rachev et al. 2008 p.102-103 and Carlin et al. 2008 ch.2.2.3).

Chapter 2

Optimal Portfolio Selection

2.1 Portfolio definition

What is a portfolio?

- It is a distribution of some initial capital across a given set of financial assets or/and bonds.

More formally, a portfolio of k-securities is a vector $\big(x_1(t), x_2(t), \ldots, x_k(t)\big)$ indicating the number of shares of the respective security held by an investor at time t. The value of such a portfolio at time t is

$$V(t) = x_1(t) \cdot S_1(t) + x_2(t) \cdot S_2(t) + \cdots + x_k \cdot S_k(t)$$

where $S_i(t)$ is the price of the i-th security at time t.

2.2 Portfolio weights

A portfolio constructed from k different securities can be described in terms of their weights,

$$\omega_i = \frac{x_i S_i(0)}{V(0)}, \quad i = 1, 2, \ldots, k$$

where x_i is the number of shares of the i-th security, $S_i(0)$ is the price of security i at time $t = 0$, and $V(0)$ is the value of the portfolio at time $t = 0$ or in other words, the initial capital invested in the portfolio. Note that the sum of the weights is always 1 since

$$\omega_1 + \omega_2 + \cdots + \omega_k = \frac{x_1 S_1(0)}{V(0)} + \frac{x_2 S_2(0)}{V(0)} + \cdots + \frac{x_k S_k(0)}{V(0)} = \frac{V(0)}{V(0)} = 1$$

In matrix notation, it can be written as

$$\boldsymbol{\omega}^T \mathbf{1} = 1$$

where $\mathbf{1} = [11 \cdots 1]^T$ is a one-column matrix with all k entries equal to 1 and $\boldsymbol{\omega} = [\omega_1 \quad \omega_2 \quad \cdots \quad \omega_k]^T$. Observe that weights of the portfolio can be either greater than 1 or negative if **short sales** are allowed. A **short position** or, in other words, a negative weight means that you borrow a risky asset and sell it today, but you

then have to purchase back the asset and return it to the initial lender. Hence, you anticipate a decrease in the value of the borrowed stock in order to make profit when you close the short position. You can, also, borrow a risk-free asset (e.g money) and invest more capital in risky-assets but you then, as well, have to return back the capital together with the aggregated interest rate.

2.3 Simple and logarithmic return

Let $S(t)$ be the price of an asset (e.g stock, bond) at time t and $S(t-1)$ be the respective price at time $t-1$. The **simple net return** is defined as

$$R(t) = \frac{S(t) - S(t-1)}{S(t-1)}.$$

Rewriting $R(t) = \frac{S(t)}{S(t-1)} - 1$, the **simple gross return** is given by

$$1 + R(t) = \frac{S(t)}{S(t-1)}.$$

In finance, it is very usual to calculate the **logarithmic/continuously compounded return** which is given by

$$r(t) = \ln\big(1 + R(t)\big) = \ln\bigg(\frac{S(t)}{S(t-1)}\bigg).$$

Both kinds of returns have its pros and cons which are analyzed in detail by Hudson et al. (2010). In this paper, we will work with the logarithmic returns.

2.4 Risk and Expected Return

Suppose that the simple returns of the k assets at time t are $R_1(t), R_2(t), \ldots, R_k(t)$. The return of this portfolio is given by

$$R_p(t) = \omega_1 R_1(t) + \omega_2 R_2(t) + \cdots + \omega_k R_k(t) = \sum_{i=1}^{k} \omega_i \cdot R_i(t).$$

Of course, the logarithmic return of a portfolio is not a linear function of the log returns of its components and the previous equality only holds approximately, i.e $r_p(t) \approx R_p(t)$ (see Zivot 2014, p. 19). The expected return is given by

$$\mu_p = E\big(R_p(t)\big) = E\big(\sum_{i=1}^{k} \omega_i \cdot R_i(t)\big) = \sum_{i=1}^{k} \omega_i E\big(R_i(t)\big) = \sum_{i=1}^{k} \omega_i \mu_i = \boldsymbol{\omega}^T \mathbf{m}$$

where $\mu_i = E\big(R_i(t)\big)$ is the expected return of asset i at time t and $\boldsymbol{m}$ contains all the expected returns arranged into a one-column matrix as
$\boldsymbol{m} = [\mu_1 \quad \mu_2 \quad \ldots \quad \mu_k]^T$.

The risk of such a portfolio can be quantified by computing the variance of the portfolio return.

$$\begin{aligned}\sigma_p^2 = \mathrm{Var}(R_p) &= \mathrm{Var}\Big(\sum_{i=1}^{k} \omega_i \cdot R_i(t)\Big) = \mathrm{Cov}\Big(\sum_{i=1}^{k} \omega_i \cdot R_i(t), \sum_{j=1}^{k} \omega_j \cdot R_j(t)\Big) \\ &= \sum_{i=1}^{k}\sum_{j=1}^{k} \omega_i\omega_j \mathrm{Cov}\big(R_i(t), R_j(t)\big) \\ &= \boldsymbol{\omega}^T \boldsymbol{\Sigma} \boldsymbol{\omega}\end{aligned}$$

where $\boldsymbol{\Sigma}$ is the $k \times k$ *covariance matrix* with entries $\sigma_{ij} = \mathrm{Cov}(R_i, R_j)$.

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \dots & \sigma_{kk} \end{bmatrix}$$

Note that the diagonal elements of Σ, $\sigma_{ii} = \mathrm{Cov}(R_i, R_i) = \mathrm{Var}(R_i)$ for $i = 1, 2, \dots, k$ are just the variances of returns.

2.5 Minimum Variance Portfolio

The *Minimum Variance Portfolio* has the smallest variance among all the feasible portfolios. A cautious investor will choose such a portfolio against a portfolio which gives a higher expected return but with increased risk.

2.5.1 Finding the MVP

To find this portfolio we have to solve the following minimization problem

$$\begin{aligned}&\underset{(\omega_1,\dots,\omega_k)}{\text{minimize}}\ \sigma_p^2 = \sum_{i=1}^{k} \omega_i^2 \mathrm{Var} R_i + 2 \cdot \sum_{i=1}^{k}\sum_{j=i+1}^{k} \omega_i\omega_j \mathrm{Cov}(R_i, R_j) \\ &\text{such that} \quad \omega_1 + \omega_2 + \dots + \omega_k = 1\end{aligned}$$

The Lagrange function for this problem is

$$\begin{aligned}L(\omega_1, \omega_2, \dots, \omega_k, \lambda) = &\sum_{i=1}^{k} \omega_i^2 \mathrm{Var} R_i + 2 \cdot \sum_{i=1}^{k}\sum_{j=i+1}^{k} \omega_i\omega_j \mathrm{Cov}(R_i, R_j) \\ &+ \lambda(\omega_1 + \omega_2 + \dots + \omega_k - 1)\end{aligned}$$

and the first order conditions for a minimum are

$$\begin{aligned}
\frac{\partial L}{\partial \omega_1} &= 2\omega_1 \mathrm{Var} R_1 + 2 \cdot \sum_{i \neq 1}^{k} \omega_i \mathrm{Cov}(R_1, R_i) + \lambda = 0, \\
\frac{\partial L}{\partial \omega_2} &= 2\omega_2 \mathrm{Var} R_2 + 2 \cdot \sum_{i \neq 2}^{k} \omega_i \mathrm{Cov}(R_2, R_i) + \lambda = 0, \\
&\vdots \\
\frac{\partial L}{\partial \omega_k} &= 2\omega_k \mathrm{Var} R_k + 2 \cdot \sum_{i=1}^{k-1} \omega_i \mathrm{Cov}(R_k, R_i) + \lambda = 0 \\
\frac{\partial L}{\partial \lambda} &= \omega_1 + \omega_2 + \cdots + \omega_k - 1 = 0
\end{aligned}$$

The minimization problem can be expressed, using matrix notation, in the following form

$$\underset{(\omega_1,\ldots,\omega_k)}{\text{minimize}}\, \boldsymbol{\omega}^T \boldsymbol{\Sigma} \boldsymbol{\omega} \quad \text{such that} \quad \boldsymbol{\omega}^T \mathbf{1} = 1 \tag{2.1}$$

The Lagrange function is now given by

$$L(\boldsymbol{\omega}, \lambda) = \boldsymbol{\omega}^T \boldsymbol{\Sigma} \boldsymbol{\omega} - \lambda(\boldsymbol{\omega}^T \mathbf{1} - 1)$$

The first order conditions for the minimization problem (2.1) can be written as

$$\frac{\partial L(\boldsymbol{\omega}, \lambda)}{\partial \boldsymbol{\omega}^T} = 2 \cdot \boldsymbol{\Sigma} \boldsymbol{\omega} + \lambda \cdot \mathbf{1} = 0 \tag{2.2}$$

$$\frac{\partial L(\boldsymbol{\omega}, \lambda)}{\partial \lambda} = \boldsymbol{\omega}^T \mathbf{1} - 1 = 0 \tag{2.3}$$

Using (2.2), we solve with respect to $\boldsymbol{\omega}$ and get:

$$\boldsymbol{\omega} = -\frac{1}{2} \cdot \lambda \cdot \boldsymbol{\Sigma}^{-1} \mathbf{1} \tag{2.4}$$

Next, we combine (2.3), (2.4) and solve with respect to λ.

$$\begin{aligned}
1 = \mathbf{1}^T \boldsymbol{\omega} = -\frac{1}{2} \cdot \lambda \cdot \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \implies \\
\lambda = -\frac{2}{\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}}
\end{aligned}$$

Substituting the value of λ to (2.2) completes the solution to the problem and gives us the weights of the *minimum variance portfolio* as follows :

$$\begin{aligned}
2 \cdot \boldsymbol{\Sigma} \boldsymbol{\omega} = -\lambda \cdot \mathbf{1} = \frac{2 \cdot \mathbf{1}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} \implies \\
\boldsymbol{\omega}_{\mathrm{MVP}} = \frac{\boldsymbol{\Sigma}^{-1} \mathbf{1}}{\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}}
\end{aligned} \tag{2.5}$$

The expected return μ_{MVP} and the risk σ^2_{MVP} of the *minimum variance portfolio* are given as

$$\mu_{\mathrm{MVP}} = \boldsymbol{\omega}^T \boldsymbol{m} = \left(\frac{\boldsymbol{\Sigma}^{-1} \mathbf{1}}{\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}} \right)^T \cdot \boldsymbol{m} = \frac{\mathbf{1}^T \left(\boldsymbol{\Sigma}^{-1} \right)^T \boldsymbol{m}}{\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}} = \frac{\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{m}}{\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}} \tag{2.6}$$

and

$$\sigma^2_{\text{MVP}} = \boldsymbol{\omega}^T \boldsymbol{\Sigma} \boldsymbol{\omega} = \left(\frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}} \right)^T \cdot \frac{\overbrace{\boldsymbol{\Sigma} \cdot \boldsymbol{\Sigma}^{-1}}^{\boldsymbol{I}}\mathbf{1}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}} = \frac{\mathbf{1}^T\left(\boldsymbol{\Sigma}^{-\mathbf{1}}\right)^{\mathbf{T}}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}} \cdot \frac{\boldsymbol{I} \cdot \mathbf{1}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}$$
$$= \frac{\cancel{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}}{\cancel{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}} \cdot \mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}$$
$$= \frac{1}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}. \tag{2.7}$$

Note that $\boldsymbol{I}$ stands for the *identity matrix* which is a $k \times k$ square matrix with ones on the main diagonal and zeros everywhere else, i.e

$$\boldsymbol{I} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

Chapter 3

Estimation of Optimal Portfolio

Following the formulation of the *minimum variance portfolio*, the solution of the optimization problem in (2.5) needs to be exploited. In addition, the lowest variance bound σ^2_{MVP}(see equation (2.7)) can only be implemented if we can, somehow, calculate the covariance matrix $\boldsymbol{\Sigma}$ of the stock returns. This has been the main topic of many researches and the question of which technique produces the best out-of sample portfolio returns is rather difficult to answer. Frequentists argue that given a large amount of data, the mean vector of returns $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ can be estimated by their sample counterparts. On the other hand, Bayesians treat data as fixed and the parameters of question as random quantities. In what follows, we will present the *Frequentist approach* regarding the estimation of expected return and risk and then we will introduce the *Bayesian* framework suggesting different priors in order to model the asset returns and the weights.

3.1 Frequentist approach

In Frequentist inference, the likelihood function plays a central role. Given the data, our goal is to maximize the likelihood function, e.g find the value of the parameter for which the likelihood function is maximal. This procedure is called *Maximum Likelihood Estimation* (MLE) and results in finding point estimates together with the respective standard error (or covariance matrix) with which we can create confidence intervals. In portfolio theory, the most popular choice for modelling the properties of asset returns is the *normal distribution*. Suppose that there are k assets and $\boldsymbol{r}_t = (r_{1t}, r_{2t}, \ldots, r_{kt})$ are the returns at time t for $t = 1, 2, \ldots, n$. The returns are assumed to have *multivariate normal distribution* $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu}$ is a $k \times 1$ vector and $\boldsymbol{\Sigma}$ is a $k \times k$ matrix. The maximum likelihood estimators of the multivariate normal distribution replace the expected returns $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ in equations (2.6), (2.7). Note that although the MLE of the mean

$$\widehat{\boldsymbol{\mu}}_{\text{mle}} = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{r_t}$$

is an unbiased estimator, the MLE of the variance

$$\widehat{\boldsymbol{\Sigma}}_{mle} = \frac{1}{n}\sum_{t=1}^{n} (\boldsymbol{r_t} - \widehat{\boldsymbol{\mu}})(\boldsymbol{r_t} - \widehat{\boldsymbol{\mu}})^T$$

is biased. That is the reason why, in practice,

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n-1}\sum_{t=1}^{n}(\boldsymbol{r_t} - \widehat{\boldsymbol{\mu}})(\boldsymbol{r_t} - \widehat{\boldsymbol{\mu}})^T$$

is often applied. Hence, the unknown parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are estimated from the available data.

3.2 Estimation from the Bayesian perspective

Within the Bayesian context, the whole point and the main objective in order to draw inference is to derive the *posterior distribution*. In contrast to the *frequentist* approach (which considers the parameter in question fixed but unknown), the parameter we want to estimate, say $\boldsymbol{\theta}$, is considered as random variable. The key idea is that we are trying to assign a probability to all different values of the parameter we want to estimate. In order to achieve this goal, e.g to find the *posterior distribution*, we make use of the following famous formula known as *Bayes' Theorem*:

$$P(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{P(\boldsymbol{y}|\boldsymbol{\theta}) \cdot P(\boldsymbol{\theta})}{P(\boldsymbol{y})} \tag{3.1}$$

where $P(\boldsymbol{\theta}|\boldsymbol{y})$ is the **posterior** distribution given the **observed** data $\boldsymbol{y}$, $P(\boldsymbol{y}|\boldsymbol{\theta}) = L(\boldsymbol{\theta}|\boldsymbol{y})$ is the **likelihood** function. It is a prior specification, an assumption for the process which generates the data. $P(\boldsymbol{\theta})$ is the **prior** distribution of the parameter $\boldsymbol{\theta}$. It represents the pre-experimental knowledge of the parameter value $\boldsymbol{\theta}$ or, in other words, it is our subjective belief about the parameter $\boldsymbol{\theta}$ before we observe the data. $P(\boldsymbol{y})$ is the marginal probability of the data $\boldsymbol{y}$.
We can rewrite equation (3.1) in the following form :

$$P(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{L(\boldsymbol{\theta}|\boldsymbol{y}) \cdot P(\boldsymbol{\theta})}{\int P(\boldsymbol{y}|\boldsymbol{\theta}) \cdot P(\boldsymbol{\theta}) d\boldsymbol{\theta}} = \frac{L(\boldsymbol{\theta}|\boldsymbol{y}) \cdot P(\boldsymbol{\theta})}{\int L(\boldsymbol{\theta}|\boldsymbol{y}) \cdot P(\boldsymbol{\theta}) d\boldsymbol{\theta}} \tag{3.2}$$

Since the denominator does not depend on θ we can again rewrite (3.2) as :

$$P(\boldsymbol{\theta}|\boldsymbol{y}) \propto L(\boldsymbol{\theta}|\boldsymbol{y}) \cdot P(\boldsymbol{\theta}) \tag{3.3}$$

Equation (3.3) provides us with the posterior distribution of θ up to some unknown constant.
Through the estimation process, we consider the general setup for the portfolio weights, as discussed in (Bodnar et al. 2017, p. 293) and (Bodnar et al. 2008, p. 131), by considering arbitrary linear combinations of the **MVP** weights. Let $\boldsymbol{L}$ be an arbitrary $p \times k$ matrix of constants with $p < k$. Let $\boldsymbol{l}_i \in \mathcal{R}^k$, $i = 1, \ldots, p$, $1 \leq p \leq k-1$, and $\boldsymbol{L}^T = (\boldsymbol{l}_1, \ldots, \boldsymbol{l}_p)$.
We define

$$\boldsymbol{\theta} = \boldsymbol{L}\mathbf{w}_{\text{MVP}} = \frac{\mathbf{L\Sigma^{-1}1}}{\mathbf{1^T\Sigma^{-1}1}} = \left(\frac{\mathbf{l_1^T\Sigma^{-1}1}}{\mathbf{1^T\Sigma^{-1}1}}, \ldots, \frac{\mathbf{l_p^T\Sigma^{-1}1}}{\mathbf{1^T\Sigma^{-1}1}}\right) \tag{3.4}$$

The setting above allows us not only to analyse a specific weight of our portfolio but also to make comparisons of one weight against the other.

3.2.1 Bayesian Point Estimates

In the Bayesian setting, there is a greater freedom since one can find the **mean**, **mode** or the **median** of the posterior distribution $P(\boldsymbol{\theta}|\boldsymbol{y})$. In addition, similar to the Frequentists' MLE approach, one can find the parameter $\widehat{\boldsymbol{\theta}}$ that maximizes the posterior distribution, what we call **Most Probable Estimator**, i.e

$$\widehat{\boldsymbol{\theta}} = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} P(\boldsymbol{\theta}|\boldsymbol{y}),$$

where $\boldsymbol{\Theta}$ is the set of all possible values of $\boldsymbol{\theta}$. Nevertheless, the most common Bayesian point estimate is the **mean** of the posterior which is defined as

$$\mu_p = E(\boldsymbol{\theta}|\boldsymbol{y}) = \int_{\boldsymbol{\Theta}} \boldsymbol{\theta} \cdot P(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}.$$

3.2.2 Credible Intervals

Given that we have obtained the posterior distribution $P(\boldsymbol{\theta}|\boldsymbol{y})$ and a significance level α, an **equal-tailed** (i.e symmetric) $100(1-\alpha)\%$ credible interval is (θ_L, θ_U) such that

$$\frac{\alpha}{2} = \int_{-\infty}^{\theta_L} P(\boldsymbol{\theta}|\boldsymbol{y})\, d\boldsymbol{\theta} = \int_{\theta_U}^{\infty} P(\boldsymbol{\theta}|\boldsymbol{y})\, d\boldsymbol{\theta}.$$

This means that we have to find the interval between the quantiles $q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}}$ of the posterior distribution. In addition, the **Highest Posterior Density**(HPD) credible set is a two-tailed interval assuming the smallest possible interval on a posterior density for a given significance level α. For a more formal definition address to Carlin et al. (2008)[ch. 2.3.2]. Credible sets are for Bayesians what confidence intervals are for Frequentists. Despite the analogy, the inference about these two concepts is completely different. On the one hand, credible sets can be seen as a probabilistic statement for the parameter $\boldsymbol{\theta}$. Thus, based on the posterior distribution, the $100(1-\alpha)\%$ credible set (θ_L, θ_U) given the observed data, means that **there is a $100(1-\alpha)\%$ probability that the parameter $\boldsymbol{\theta}$ lies in** (θ_L, θ_U).

On the other hand, a confidence interval, from the Frequentist perspective, is **not** a probabilistic statement. Instead, the interpretation is that given we were to repeat the experiment numerous times, $100(1-\alpha)\%$ of the created intervals would include the true parameter $\boldsymbol{\theta}$.

3.2.3 Choice of the prior

The **first prior** that is going to be used has been applied in portfolio theory by Klein et al. (1976), Barry (1974) and Brown (1976). It is an uninformative *diffuse* prior in the sense that the investor has no prior knowledge of how the parameters vary that he would wish to incorporate. It allows us to model the average vector of returns $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$. The density of the so-called *Jeffrey's* prior is given by

$$p_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{k+1}{2}}$$

where k is the number of assets.
The **second prior** that will be used is an informative *conjugate* prior meaning that our personal belief towards the parameters is incorporated via this specific choice of model. The conjugate prior for the mean vector of returns, conditional on $\boldsymbol{\Sigma}$, is a multivariate normal distribution, e.g $\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim \mathcal{N}(\boldsymbol{\mu}_c, \frac{1}{\kappa_c}\boldsymbol{\Sigma})$ and can be written as

$$p_c(\boldsymbol{\mu}|\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-1/2}\exp\Big\{ -\frac{\kappa_c}{2}(\boldsymbol{\mu}-\boldsymbol{\mu_c})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu_c})\Big\} \tag{3.5}$$

where $\boldsymbol{\mu}_c$ is the prior mean and κ_c is a parameter reflecting the prior precision of $\boldsymbol{\mu}_c$.
The conjugate prior for the unknown covariance matrix is the inverse *Wishart* distribution, e.g $\boldsymbol{\Sigma} \sim \mathcal{IW}(\boldsymbol{S_c}, \nu_c - k - 1)$ and can be expressed as

$$p_c(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\nu_c/2} \cdot \exp\Big\{ -\frac{1}{2}\text{tr}[\boldsymbol{S}_c\boldsymbol{\Sigma}^{-1}]\Big\} \tag{3.6}$$

where ν_c is a precision parameter on $\boldsymbol{\Sigma}$ and $\boldsymbol{S}_c$ is a known prior $k \times k$ matrix of $\boldsymbol{\Sigma}$. Combining (3.5) and (3.6) we get the following joint prior for both parameters as

$$p_c(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(\nu_c+1)/2} \cdot \exp\Big\{ -\frac{\kappa_c}{2}(\boldsymbol{\mu}-\boldsymbol{\mu_c})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu_c}) - \frac{1}{2}\text{tr}[\boldsymbol{S}_c\boldsymbol{\Sigma}^{-1}]\Big\} \tag{3.7}$$

At this point, the following Lemma from Bodnar et al. (2017) will be introduced. Its importance is crucial since, in section 4.2, the derivation of the "true" weights of the chosen model are related to the distinct choice of the prior precision matrix $\boldsymbol{S}_c$. More precisely, when we simulate a covariance matrix $\boldsymbol{\Sigma}$ from IW distribution, the choice of the prior matrix $\boldsymbol{S}_c$ defines the weights of our model.
Lemma 1. Let $\boldsymbol{\Sigma}|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \sim \mathcal{IW}_k(\tau_0, \boldsymbol{V}_0)$ with $\boldsymbol{V}_0 = \boldsymbol{V}_0(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ and let $\boldsymbol{L}$ be a $p \times k$ matrix of constants.
Then

$$\frac{\boldsymbol{L}\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}\Big|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \sim t_p\Big(\tau_0 - k - 1; \frac{\boldsymbol{L}\boldsymbol{V}_0^{-1}\mathbf{1}}{\mathbf{1}^T\boldsymbol{V}_0^{-1}\mathbf{1}}; \frac{1}{\tau_0 - k - 1}\frac{\boldsymbol{L}\boldsymbol{R}_0\boldsymbol{L}^T}{\mathbf{1}^T\boldsymbol{V}_0^{-1}\mathbf{1}}\Big) \tag{3.8}$$

Proof. (see Bodnar et al. 2017, Appendix A)

In the following prior models we will make statements directly for the linear combination of the portfolio weights. From the investor's perspective that approach seems more natural since it is rational to have a prior belief about the optimal allocation of the wealth than a prior belief about the expected returns. Before we proceed to the last priors we will present the general setting of reparameterization (see Bodnar et al. 2017, ch. 4.1) since it will be used in order to derive the covariance matrix for the simulation study based on the informative prior in Chapter 4.
Let $\tilde{\boldsymbol{L}} = \begin{pmatrix}\boldsymbol{L}^T & \mathbf{1}\end{pmatrix}^T = \begin{pmatrix}\boldsymbol{L} \\ \mathbf{1}^T\end{pmatrix}$, $\tilde{\boldsymbol{\Sigma}} = \tilde{\boldsymbol{L}}\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{L}}^T$, $\zeta = \mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}$ and $\boldsymbol{\theta}$ defined from previous section as $\boldsymbol{\theta} = \frac{\boldsymbol{L}\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}} = \frac{\boldsymbol{L}\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\zeta}$. $\boldsymbol{\Psi}$ is defined in the following way

$$\boldsymbol{\Psi} = \boldsymbol{L}\boldsymbol{\Sigma}^{-1}\boldsymbol{L}^T - \frac{\boldsymbol{L}\boldsymbol{\Sigma}^{-1}\mathbf{1}\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{L}^T}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}.$$

Solving with respect to $\boldsymbol{L\Sigma}^{-1}\boldsymbol{L}^T$, we get

$$\boldsymbol{L\Sigma}^{-1}\boldsymbol{L}^T = \boldsymbol{\Psi} + \frac{\boldsymbol{L\Sigma}^{-1}\mathbf{1}\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{L}^T}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}} = \boldsymbol{\Psi} + \frac{\zeta\not{^2}\cdot\boldsymbol{\theta}\boldsymbol{\theta}^T}{\not{\zeta}} = \boldsymbol{\Psi} + \zeta\cdot\boldsymbol{\theta}\boldsymbol{\theta}^T$$

So,

$$\begin{aligned}\tilde{\boldsymbol{\Sigma}} = \tilde{\boldsymbol{L}}\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{L}}^T &= \begin{pmatrix}\boldsymbol{L}\\ \mathbf{1}^T\end{pmatrix}\boldsymbol{\Sigma}^{-1}\begin{pmatrix}\boldsymbol{L}^T & \mathbf{1}\end{pmatrix} = \begin{pmatrix}\boldsymbol{L}\\ \mathbf{1}^T\end{pmatrix}\begin{pmatrix}\boldsymbol{\Sigma}^{-1}\boldsymbol{L}^T & \boldsymbol{\Sigma}^{-1}\mathbf{1}\end{pmatrix}\\ &= \begin{bmatrix}\boldsymbol{L\Sigma}^{-1}\boldsymbol{L}^T & \boldsymbol{L\Sigma}^{-1}\mathbf{1}\\ \mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{L}^T & \mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}\end{bmatrix} = \begin{bmatrix}\boldsymbol{\Psi}+\zeta\cdot\boldsymbol{\theta}\boldsymbol{\theta}^T & \zeta\boldsymbol{\theta}\\ \zeta\boldsymbol{\theta}^T & \zeta\end{bmatrix}\\ &= \zeta\begin{bmatrix}\frac{\boldsymbol{\Psi}}{\zeta}+\boldsymbol{\theta}\boldsymbol{\theta}^T & \boldsymbol{\theta}\\ \boldsymbol{\theta}^T & 1\end{bmatrix}. \end{aligned} \tag{3.9}$$

The **third** prior will be the Jeffrey's non-informative prior expressed by the following model as

$$p_n(\boldsymbol{\theta},\boldsymbol{\Psi},\zeta) \propto \zeta^{\frac{p}{2}-1}|\boldsymbol{\Psi}|^{-\frac{p}{2}-1} \tag{3.10}$$

For more details (see Bodnar et al. 2017, Ch. 4.1).
Finally, the **fourth** prior that will be used is an informative prior under a hierarchical Bayesian model similar to the one developed by Tunaru (2002). The informative prior is given by

$$\boldsymbol{\theta} \sim \mathcal{N}_p\big(\boldsymbol{w}_I, \frac{1}{\zeta}\boldsymbol{\Psi}^{-1}\big) \tag{3.11}$$

$$\boldsymbol{\Psi} \sim W_p(\nu_I, \boldsymbol{S}_I) \tag{3.12}$$

$$\zeta \sim \text{Gamma}(\delta_1, 2\delta_2) \tag{3.13}$$

where $\boldsymbol{w}_I$ is the prior mean, ν_I is a prior precision parameter on $\boldsymbol{\Psi}$, $\boldsymbol{S}_I$ is the known matrix and δ_1, δ_2 are prior constants. The joint prior is expressed as

$$\begin{aligned}p_I(\boldsymbol{\theta},\boldsymbol{\Psi},\boldsymbol{\zeta}) \propto &\left|\frac{1}{\zeta}\boldsymbol{\Psi}^{-1}\right|^{-\frac{1}{2}}\exp\Big\{-\frac{\zeta}{2}(\boldsymbol{\theta}-\boldsymbol{w}_I)^T\boldsymbol{\Psi}(\boldsymbol{\theta}-\boldsymbol{w}_I)\Big\}\\ &\times\zeta^{\delta_1-1}|\boldsymbol{\Psi}|^{\frac{\nu_I-p-1}{2}}\exp\Big\{-\frac{1}{2}\text{tr}[\boldsymbol{S}_I^{-1}\boldsymbol{\Psi}]-\frac{\zeta}{2\delta_2}\Big\}.\end{aligned}$$

3.2.4 Posterior distribution based on the asset returns

Proposition 1. Let $\boldsymbol{X_1},\ldots,\boldsymbol{X_n}|\boldsymbol{\mu},\boldsymbol{\Sigma}$ be independent and identically distributed with $\boldsymbol{X_i}|\boldsymbol{\mu},\boldsymbol{\Sigma} \sim \mathcal{N}_k(\boldsymbol{\mu},\boldsymbol{\Sigma})$. Let $\boldsymbol{L}$ be a $p\times k$ matrix of constants, $p<k$ and $\mathbf{1}$ denotes the vector of ones. Then
(a) Under the diffuse prior $p_d(\boldsymbol{\mu},\boldsymbol{\Sigma})$ the posterior for $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\theta}|\boldsymbol{X_1},\ldots,\boldsymbol{X_n} \sim t_p\left(n-1;\widehat{\boldsymbol{\theta}};\frac{1}{n-1}\frac{\boldsymbol{LR}_d\boldsymbol{L}^T}{\mathbf{1}^T\boldsymbol{S}^{-1}\mathbf{1}}\right) \tag{3.14}$$

where $\boldsymbol{R}_d = \boldsymbol{S}^{-1}-\boldsymbol{S}^{-1}\mathbf{1}\mathbf{1}^T\boldsymbol{S}^{-1}/\mathbf{1}^T\boldsymbol{S}^{-1}\mathbf{1}$, $\widehat{\boldsymbol{\theta}} = \frac{\boldsymbol{LS^{-1}1}}{\mathbf{1}^T\boldsymbol{S^{-1}1}}$.

(b) Under the conjugate prior $p_c(\boldsymbol{\mu},\boldsymbol{\Sigma})$ the posterior for $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\theta}|\boldsymbol{X_1},\ldots,\boldsymbol{X_n} \sim t_p\left(\nu_c+n-k-1;\frac{\boldsymbol{LV}_c^{-1}\boldsymbol{L}^T}{\mathbf{1}^T\boldsymbol{V}_c^{-1}\mathbf{1}};\frac{1}{\nu_c+n-k-1}\frac{\boldsymbol{LR}_c\boldsymbol{L}^T}{\mathbf{1}^T\boldsymbol{V}_c^{-1}\mathbf{1}}\right) \tag{3.15}$$

where

$$\begin{aligned}
\boldsymbol{r}_c &= \frac{n\overline{\boldsymbol{X}} + \kappa_c \boldsymbol{\mu}_c}{n + \kappa_c}, \\
\boldsymbol{V}_c &= (n-1)\boldsymbol{S} + \boldsymbol{S}_c + (n+\kappa_c)\boldsymbol{r}_c\boldsymbol{r}_c^T + n\overline{\boldsymbol{X}}\,\overline{\boldsymbol{X}}^T + \kappa_c\boldsymbol{\mu}_c\boldsymbol{\mu}_c^T, \\
\boldsymbol{R}_c &= \boldsymbol{V}_c^{-1} - \frac{\boldsymbol{V}_c^{-1}\mathbf{1}\mathbf{1}^T\boldsymbol{V}_c^{-1}}{\mathbf{1}^T\boldsymbol{V}_c^{-1}\mathbf{1}}.
\end{aligned}$$

Proofs. (see Bodnar et al. 2017, Appendix A)

3.2.5 Posterior distribution based on the MVP weights

Proposition 2. Let $\boldsymbol{X_1}, \ldots, \boldsymbol{X_n}|\boldsymbol{\mu}, \boldsymbol{\Sigma}$ be independent and identically distributed with $\boldsymbol{X_i}|\boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let $\boldsymbol{L}$ be a $p \times k$ matrix of constants with $p < k$. Then the posterior for the MV portfolio weights $\boldsymbol{\theta}$ under the jeffrey's non-informative prior $p_n(\boldsymbol{\theta}, \boldsymbol{\Psi}, \zeta)$ is given by

$$\boldsymbol{\theta}|\boldsymbol{X_1}, \ldots, \boldsymbol{X_n} \sim t_p\left(n-k+p; \widehat{\boldsymbol{\theta}}; \frac{1}{n-k+p}\frac{\boldsymbol{L}\boldsymbol{R}_d\boldsymbol{L}^T}{\mathbf{1}^T\boldsymbol{S}^{-1}\mathbf{1}}\right) \tag{3.16}$$

Proof. (see Bodnar et al. 2017, 4.1 Non-informative Prior)
This result is almost identical to the posterior obtained under the diffuse prior based on expected returns. The only difference lies on the degrees of freedom.

Proposition 3. Let $\boldsymbol{X_1}, \ldots, \boldsymbol{X_n}|\boldsymbol{\mu}, \boldsymbol{\Sigma}$ be independent and identically distributed with $\boldsymbol{X_i}|\boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let $\boldsymbol{L}$ be a $p \times k$ matrix of constants with $p < k$. Then the posterior for $\boldsymbol{\theta}$ under the informative prior $p_I(\boldsymbol{\theta}, \boldsymbol{\Psi}, \zeta)$ is given by

$$\begin{aligned}
p_I(\boldsymbol{\theta}|\boldsymbol{X_1}, \ldots, \boldsymbol{X_n}) \propto &\left[(\boldsymbol{\theta}-\boldsymbol{w}_I)^T(\boldsymbol{S}_I^{-1} + (n-1)(\boldsymbol{L}\boldsymbol{R}_d\boldsymbol{L}^T)^{-1})^{-1}(\boldsymbol{\theta}-\boldsymbol{w}_I)\right]^{\frac{n-k+2p+2\delta_1}{2}} \\
&\times U\left(\frac{n-k+2p+2\delta_1}{2}; \frac{p+2\delta_1-\nu_I+1}{2}; g(\boldsymbol{\theta})\right)
\end{aligned} \tag{3.17}$$

where $U(\cdot,\cdot,\cdot)$ is a confluent hypergeometric function introduced by Abramowitz et al. (1964) and

$$g(\boldsymbol{\theta}) = \frac{n-1}{2}\frac{\left((\boldsymbol{\theta}-\widehat{\theta})^T(\boldsymbol{L}\boldsymbol{R}_d\boldsymbol{L}^T)^{-1}(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}) + (\mathbf{1}^T\boldsymbol{S}^{-1}\mathbf{1})^{-1}\right) + \frac{\delta_2^{-1}}{n-1}}{(\boldsymbol{\theta}-\boldsymbol{w}_I)^T\left(\boldsymbol{S}_I^{-1} + (n-1)(\boldsymbol{L}\boldsymbol{R}_d\boldsymbol{L}^T)^{-1}\right)^{-1}(\boldsymbol{\theta}-\boldsymbol{w}_I)}.$$

Throughout the paper, instead of the previous result, we will use the following stochastic representation formula for $\boldsymbol{\theta}$ (see Bodnar et al. 2017, p. 298):

$$\boldsymbol{\theta} = \boldsymbol{r}_I + \zeta^{-\frac{1}{2}}\big(\boldsymbol{V}_I(\tau)\big)^{\frac{1}{2}}\boldsymbol{z}_0 \tag{3.18}$$

where

$$\begin{aligned}
\boldsymbol{z}_0 &\sim \mathcal{N}_p(\mathbf{0}_p, \boldsymbol{I}_p), \\
\zeta|\tau &\sim \text{Gamma}\left(\frac{n-k+2p+2\delta_1}{2}, \frac{2}{h_I(\tau)}\right), \\
\tau &\sim \text{Gamma}\left(\frac{n-k+p+\nu_I-1}{2}, 2\right)
\end{aligned}$$

and

$$
\begin{aligned}
\boldsymbol{P}_1 &= \left(\boldsymbol{S}_I^{-1} + (n-1)(\boldsymbol{L}\boldsymbol{R}_d\boldsymbol{L}^T)^{-1} \right)^{-1}, \\
\boldsymbol{P}_2 &= (n-1)(\boldsymbol{L}\boldsymbol{R}_d\boldsymbol{L}^T)^{-1}, \\
r &= \delta_2^{-1} + (n-1)(\mathbf{1}^T\boldsymbol{S}^{-1}\mathbf{1})^{-1}, \\
\boldsymbol{V}_I(\tau) &= (\tau\boldsymbol{P}_1 + \boldsymbol{P}_2)^{-1}, \\
\boldsymbol{r}_I(\tau) &= (\tau\boldsymbol{P}_1 + \boldsymbol{P}_2)^{-1}(\tau\boldsymbol{P}_1\boldsymbol{w}_I + \boldsymbol{P}_2\widehat{\boldsymbol{\theta}}), \\
\boldsymbol{h}_I(\tau) &= r + \tau\boldsymbol{w}_I^T\boldsymbol{P}_1\boldsymbol{w}_I + \widehat{\boldsymbol{\theta}}^T\boldsymbol{P}_2\widehat{\boldsymbol{\theta}} - \boldsymbol{r}_I(\tau)^T\big(\boldsymbol{V}_I(\tau)\big)^{-1}\boldsymbol{r}_I(\tau).
\end{aligned}
$$

Chapter 4

Numerical Study

Simulated data from the MN distribution are used and they represent the weekly log-returns of different stocks for a period of one year. Based on the simulated data and assuming that we want to examine five different stocks, their respective weights are evaluated under four different models. Almost each model produces different point estimates. The goal of the study was to assume specific priors for the mean vector of expected returns $\boldsymbol{\mu}$ of the stocks, their respective covariance matrix $\boldsymbol{\Sigma}$ and for the linear combinations of the weights $\boldsymbol{\theta}$ in two different scenarios.

In the first scenario, the mean vector of expected returns and the covariance matrix are assumed to follow distributions from the conjugate prior (see (3.5), (3.6)), e.g the vector of expected returns follows a MN and the covariance matrix follows an IW distribution. In addition, the data are generated from MN with an IW distributed covariance matrix, and a MN distributed mean vector given this specific covariance matrix.

In the second scenario, the parameters $\boldsymbol{\theta}$ (linear combination of weights), $\boldsymbol{\Psi}$, ζ are simulated from an informative prior (see (3.11), (3.12), (3.13)) and they are used together with (3.9) and $\tilde{\boldsymbol{\Sigma}} = \tilde{\boldsymbol{L}}\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{L}}^T$ for the generation of the covariance matrix $\boldsymbol{\Sigma}$. Then, having $\boldsymbol{\Sigma}$ we just needed a mean vector in order to simulate the data from MN distribution.

In the first case, the prior precision matrix $\boldsymbol{S}_c$, which accounts for the generation of the covariance matrix $\boldsymbol{\Sigma}$, uniquely defines the "true" weights of the MVP (see (3.8)). Note that, in the second case, the "true" weights are directly used as input in (3.11) and need not be calculated as it was done in the first scenario. Moreover, the prior weights are then simulated based on the true weights.

Throughout the simulation study, we have assumed that the number of assets is five, e.g $k = 5$. In order to derive the distributions of the five weights of each model separately, p is set to 1. Thus, $\boldsymbol{L}$ (see (3.4)) is defined as the following 1×5 matrix:

$$\boldsymbol{L} = (1, 0, 0, 0, 0).$$

In an analogous way, inference for the remaining 4 weights is drawn by setting $\boldsymbol{L} = (0, 1, 0, 0, 0)$, $\boldsymbol{L} = (0, 0, 1, 0, 0)$ and so forth.

Furthermore, since we know the "true" parameters, we can assess the performance of all models in terms of their point estimates. Hence, the quadratic loss function is introduced and the best point estimate will be the one that minimizes the expected quadratic loss function.

4.1 Loss Function & Expected Loss Function

Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k) \in \boldsymbol{\Theta}$ be a vector of the true values of the parameter and let $\widehat{\boldsymbol{\theta}}_M = \big(\widehat{\theta}_{1,M}, \widehat{\theta}_{2,M}, \ldots, \widehat{\theta}_{k,M}\big)$ be the vector of estimated parameters under a specific model. The **Quadratic** or $\boldsymbol{L}_2$ loss function is defined as

$$L(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_M) = ||\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_M||^2 = \sum_{i=1}^{k} \big(\theta_i - \widehat{\theta}_{i,M}\big)^2$$

and the approximated **Expected Quadratic** or $\boldsymbol{L}_2$ **loss function** is given by

$$R(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_M) = E\big(L(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_M)\big) = \frac{1}{N} \sum_{j=1}^{N} \sum_{i=1}^{k} \big(\theta_{i,j} - \widehat{\theta}_{i,j,M}\big)^2$$

where N represents the number of simulations.

4.2 Simulation from Conjugate prior

As mentioned in the beginning of the chapter, the first case that is considered is that of conjugate priors where the investor has informative beliefs about the mean vector and the covariance matrix of returns. In the Bayesian framework, the `IW` distribution is the conjugate prior to the covariance matrix of the multivariate normal distribution (see Nydick 2012, Ch 2.4) , whereas the multivariate normal (conditional on $\boldsymbol{\Sigma}$) is the conjugate prior to the mean vector of the multivariate normal distribution (see Murphy 2007, p.17-18). Hence, we simulate a 5×5 covariance matrix from the `IW` distribution with 50 degrees of freedom, e.g we have $k = 5$ assets and we collect the weekly returns for a period of one year ($n = 50$). Given the covariance matrix $\boldsymbol{\Sigma}$, the vector $\boldsymbol{\mu}$ of expected returns is simulated from `MN` distribution. After following those steps, our dataset is provided by simulating values from `MN` with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

For the informative prior we need to make assumptions about the prior constants. Thus, we set $\delta_1 = 1$, $\delta_2 = 0.5$, $\nu_I = n$, $S_I = 1$, and $w_I = 0.2$, i.e all prior weights are be equal (equally weighted portfolio). In addition, the point estimate of each weight, under the stochastic representation formula (see (3.18)), is drawn based on 20000 simulations of z_0, τ and ζ and then taking the average (see Bodnar et al. 2017, p. 298). This procedure is applied four times for $\boldsymbol{L} = (1, 0, 0, 0, 0)$, $\boldsymbol{L} = (0, 1, 0, 0, 0)$ etc, resulting in four point estimates. Since the sum of the weights must be equal to one, the fifth weight can be derived as a function of the other weights.

For the conjugate prior we assume $\nu_c = k_c = n + 6$, $\mu_c = \mathbf{0}_k$, where n represents the number of collected data for each asset, $\mathbf{0}_k$ is a k-dimensional vector of zeros. $\boldsymbol{S}_c$ was chosen aiming to cover two basic directions. First of all, we wanted to allow for flexibility so that $\boldsymbol{S}_c$ would be as realistic as possible. A convenient way to achieve that was to use the relationship that connects a correlation matrix with a covariance matrix in the following way:

$$\boldsymbol{R}_{cc} = \boldsymbol{D}^{-1/2} \boldsymbol{S}_c \boldsymbol{D}^{-1/2} \implies \tag{4.1}$$

$$\boldsymbol{S}_c = \boldsymbol{D}^{1/2} \boldsymbol{R}_{cc} \boldsymbol{D}^{1/2} \tag{4.2}$$

where the correlation matrix $\boldsymbol{R}_{cc}$ is a $k \times k$ symmetric matrix with diagonal elements equal to 1. Since the non-diagonal entries take values in the interval (-1, 1), random numbers were simulated from `Uniform` (-1, 1) distribution. $\boldsymbol{D}^{1/2}$ is a $k \times k$ diagonal matrix with elements equal to the standard deviations. Hence, a random diagonal matrix of variances $\boldsymbol{D}$ was used as input. The results of the simulations that have been set under the described context above are presented in the first two sub-tables of Table 4.1. In the first sub-table, the entries of the diagonal matrix $\boldsymbol{D}$ were set equal to 0.1, 1.8, 2.4, 0.5, 0.9 (as it can be seen from the diagonal elements of $\boldsymbol{S}_c$). In the second sub-table, all the diagonal entries of $\boldsymbol{D}$ were set equal to 2. Note that the generated portfolios allow for short sales.

Next, we wanted to test the L_2 deviations based on specific portfolios. Thus, a diagonal matrix $\boldsymbol{S}_c$ was used directly as input to the `IW` distribution. This is a heuristic approach because, by lemma 1 in (3.8), if we take a diagonal matrix $\boldsymbol{S}_c$ with equal elements then the derived portfolio is just the equally weighted portfolio, i.e $w_i = 0.2$ for $i = 1, 2, \ldots, 5$. Again, if we take non equal diagonal elements for $\boldsymbol{S}_c$ then we get a non equally-weighted portfolio. The results of these slightly modified simulations are presented in the bottom sub-tables of Table 4.1. In any case, the matrix $\boldsymbol{S}_c$ must be **positive definite**, meaning that not only it has to be symmetric but it also has to have positive eigenvalues.

In all considered scenarios, multiple datasets from conjugate priors have been generated based on 1000 simulations. In order to test the performance of each model, comparison of the point estimates of weights against the "true" weights is performed via the expected quadratic loss function. Note that since the diffuse and the non informative prior have the same point estimates, we only needed to calculate three instead of four expected quadratic loss functions. Moreover, it should be mentioned that since the "true" covariance matrix is simulated from the conjugate prior, we would expect that the smallest deviation corresponds to the case of the model which assumes conjugate priors. The results in Table 4.1 validates our expectation but with one exception. In the last sub-table the informative prior performs better but this is due to the fact that the chosen prior matrix $\boldsymbol{S}_c$, which is a diagonal matrix with all entries equal to 1, provides an equally weighted portfolio as the "true" portfolio which coincides with our assumption for the prior weights of the informative prior.

Table 4.1: L-2 deviations from the "true" weights based on 1000 simulations from the conjugate priors. Each sub-table represents a different choice of the prior precision matrix $\boldsymbol{S_c}$ and hence a different set of "true" weights.

Prior matrix	True weights	Model	
$\boldsymbol{S_c} = \begin{pmatrix} 0.10 & -0.1483 & -0.0818 & 0.0025 & 0.0704 \\ -0.1483 & 1.80 & 0.3551 & 0.7575 & 0.1907 \\ -0.0818 & 0.3551 & 2.40 & -0.4091 & -0.0588 \\ 0.0025 & 0.7575 & -0.4091 & 0.50 & 0.1283 \\ 0.0704 & 0.1907 & -0.0588 & 0.1283 & 0.90 \end{pmatrix}$	$\omega_1 = 0.5481$ $\omega_2 = -0.1328$ $\omega_3 = 0.1404$ $\omega_4 = 0.4412$ $\omega_5 = 0.0030$	Diff. Conj. Inf. Non-Inf.	0.3278 0.0833 0.3323 0.3278
$\boldsymbol{S_c} = \begin{pmatrix} 2 & -0.6990 & -0.3339 & 0.0227 & 0.4693 \\ -0.6990 & 2 & 0.3417 & 1.5969 & 0.2996 \\ -0.3339 & 0.3417 & 2 & -0.7469 & 0.0800 \\ 0.0227 & 1.5969 & -0.7469 & 2 & 0.3825 \\ 0.4693 & 0.2996 & 0.0800 & 0.3825 & 2 \end{pmatrix}$	$\omega_1 = -0.1563$ $\omega_2 = -1.1325$ $\omega_3 = 0.8204$ $\omega_4 = 1.3360$ $\omega_5 = 0.1325$	Diff. Conj. Inf. Non-Inf.	0.3483 0.0921 3.5697 0.3483
$\boldsymbol{S_c} = \begin{pmatrix} 0.25 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 9.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 9.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 9 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 9 \end{pmatrix}$	$\omega_1 = 0.9$ $\omega_2 = 0.025$ $\omega_3 = 0.025$ $\omega_4 = 0.025$ $\omega_5 = 0.025$	Diff. Conj. Inf. Non-Inf.	0.0083 0.0021 0.5738 0.0083
$\boldsymbol{S_c} = \begin{pmatrix} 1 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.000 & 1 & 0.000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 1 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 1 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1 \end{pmatrix}$	$\omega_1 = 0.2$ $\omega_2 = 0.2$ $\omega_3 = 0.2$ $\omega_4 = 0.2$ $\omega_5 = 0.2$	Diff. Conj. Inf. Non-Inf.	0.0337 0.0086 3.58e-05 0.0337

4.3 Simulation from Informative prior

In this part of the simulation study, the "true" weights are directly used as input in (3.11). In order to simulate a set of informative weights we also need to generate values for $\boldsymbol{\Psi}$ and ζ. Our main goal still remains, as in the previous section, to generate a covariance matrix under the informative's prior setup by using (3.9). The mean vector of returns is generated from `MN` distribution given the derived covariance matrix and, consequently, our dataset is provided by simulating data from `MN` distribution given both the generated mean vector and the covariance matrix. Also, recall that $\tilde{\boldsymbol{L}}$ was defined as $\tilde{\boldsymbol{L}} = \begin{pmatrix} \boldsymbol{L} \\ \mathbf{1}^T \end{pmatrix}$ and $\boldsymbol{L}$ was defined as the $(k-1) \times k$ matrix of constants, i.e it can be written as

$$\boldsymbol{L} = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ (k-1) \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & \cdots & (k-1) & k \end{array} \\ \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \end{array}.$$

Thus, $\tilde{\boldsymbol{L}}$ can now be written as

$$\tilde{\boldsymbol{L}} = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ (k-1) \\ k \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & \cdots & (k-1) & k \end{array} \\ \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix} \end{array}.$$

The generation of the covariance matrix was completed in 3 steps.
In step 1, we generate ζ, $\boldsymbol{\Psi}$ and $\boldsymbol{\theta}$ in the following way:

$$\begin{aligned} \zeta &\sim \text{Gamma}(\delta_1, 2\delta_2) \\ \boldsymbol{\Psi} &\sim W_4(\nu_I, \boldsymbol{S}_I) \\ \boldsymbol{\theta} &\sim \mathcal{N}_4(\boldsymbol{w_I}, \frac{1}{\zeta}\boldsymbol{\Psi^{-1}}) \end{aligned}$$

In step 2, $\tilde{\boldsymbol{\Sigma}}$ is calculated by plugging in the values from step 1 in (3.9).
In step 3, the covariance matrix can now be derived by using $\tilde{\boldsymbol{\Sigma}} = \tilde{\boldsymbol{L}}\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{L}}^T$ and solving with respect to $\boldsymbol{\Sigma}$, i.e

$$\begin{aligned} \tilde{\boldsymbol{\Sigma}} = \tilde{\boldsymbol{L}}\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{L}}^T &\implies \\ \tilde{\boldsymbol{L}}^{-1}\tilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{L}}^T &\implies \\ \tilde{\boldsymbol{L}}^{-1}\tilde{\boldsymbol{\Sigma}}(\tilde{\boldsymbol{L}}^T)^{-1} = \boldsymbol{\Sigma}^{-1} &\implies \\ \boldsymbol{\Sigma} = \left(\tilde{\boldsymbol{L}}^{-1}\tilde{\boldsymbol{\Sigma}}(\tilde{\boldsymbol{L}}^T)^{-1}\right)^{-1} & \end{aligned}$$

Since we wanted to provide results for several choices of hyper-parameters, we have considered four different couples of δ_1, δ_2 for one choice of the matrix $\boldsymbol{S}_I$. The prior ("true") weights were selected in order to include L_2 deviations from :

1. an equally weighted portfolio, i.e $w_{I,i} = 0.2$ for $i = 1, \ldots, 5$
2. a slightly skewed portfolio, i.e $w_I = (0.5, 0.125, \ldots, 0.125)$
3. a heavily skewed portfolio, i.e $w_I = (0.9, 0.025, \ldots, 0.025)$
4. a portfolio allowing for short sales, i.e $w_I = (1.4, -0.1, \cdots - 0.1)$

In all considered cases, the informative prior outperforms the other models as it can be seen in Table 4.2.

Table 4.2: L-2 deviations from the prior/true weights based on 1000 simulations from the Informative prior. The diagonal matrix $\boldsymbol{S_I}$ has been set such that the main diagonal entries are equal to 0.1.

	Prior/True weights	Diffuse	Conjugate	Inf.	Non-Inf.
$\delta_1 = 1, \delta_2 = .5$	$\omega_I = (0.2, 0.2, \ldots, 0.2)$	27.4355	15.9909	10.9963	27.4355
	$\omega_I = (0.5, 0.125, \ldots, 0.125)$	26.7057	15.1960	10.7461	26.7057
	$\omega_I = (0.9, 0.025, \ldots, 0.025)$	29.8530	17.4218	11.9897	29.8530
	$\omega_I = (1.4, -0.1, \ldots, -0.1)$	27.1242	16.1115	11.0367	27.1242
$\delta_1 = 2, \delta_2 = 1$	$\omega_I = (0.2, 0.2, \ldots, 0.2)$	1.2165	0.7043	0.5168	1.2165
	$\omega_I = (0.5, 0.125, \ldots, 0.125)$	1.2593	0.7656	0.5329	1.2593
	$\omega_I = (0.9, 0.025, \ldots, 0.025)$	1.2517	0.7109	0.5348	1.2517
	$\omega_I = (1.4, -0.1, \ldots, -0.1)$	1.2598	0.7607	0.5469	1.2598
$\delta_1 = 7.5, \delta_2 = .5$	$\omega_I = (0.2, 0.2, \ldots, 0.2)$	0.4347	0.2511	0.1869	0.4347
	$\omega_I = (0.5, 0.125, \ldots, 0.125)$	0.4403	0.2540	0.1883	0.4403
	$\omega_I = (0.9, 0.025, \ldots, 0.025)$	0.4346	0.2609	0.1870	0.4346
	$\omega_I = (1.4, -0.1, \ldots, -0.1)$	0.4433	0.2864	0.1899	0.4433
$\delta_1 = 9, \delta_2 = .25$	$\omega_I = (0.2, 0.2, \ldots, 0.2)$	0.6952	0.3975	0.2958	0.6952
	$\omega_I = (0.5, 0.125, \ldots, 0.125)$	0.6958	0.3979	0.2922	0.6958
	$\omega_I = (0.9, 0.025, \ldots, 0.025)$	0.6882	0.4035	0.2923	0.6882
	$\omega_I = (1.4, -0.1, \ldots, -0.1)$	0.6778	0.4216	0.2870	0.6778

4.4 Comparing two measurement methods for the Informative prior

However convenient the stochastic representation formula (3.18) might seem, it also has its drawbacks. For all the parts of the simulation study we have made the assumption that $\boldsymbol{S_I}$ is a diagonal matrix and hence by extracting the diagonal elements (scalars) we were able to calculate each weight separately. But, in case that we wanted to assume for a non diagonal matrix $\boldsymbol{S_I}$ our method would come to a dead end. This is equivalent to saying that the dependencies between the various weights through the inverse wishart distributed covariance matrix would be lost. In order to overcome this obstacle one has to calculate the vector of weights **at once**. This can be achieved by setting

$$\boldsymbol{L} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Dealing with this issue is by no means negligible but disadvantages also exist. Calculating the weights at once is computationally expensive meaning that it takes much more time for the code to run. With this method, we would need to compute roughly three times as many matrices as the chosen sample size for the random generated numbers (τ, ζ, z_0). The aim of this section is to compare the procedure that was already used in the previous sections, with the method we just introduced. Simulated data from the informative prior are used. The method generating the data is

exactly the same as the one considered in the previous section. The expected loss functions, though, were only computed for the informative posterior estimates based on the two methods. In the following table **method 1** stands for the method which treats the weights separately whereas **method 2** handles the weights at once. $\boldsymbol{S}_I$ is, again, a diagonal matrix with entries equal to 1. The prior weights are considered as $w_{I,i} = 0.2$ for $i = 1, \ldots, 5$. In addition, we modify the sample size of the random generated numbers τ, ζ, $\boldsymbol{z}_0$ in order to examine if there is any significant improvement of the estimates. The first method which treats the weights separately slightly outperforms the second in all considered cases. Furthermore, increasing either the number of simulations or the sample size of the hyperparameters used in stochastic representation formula does not seem to affect the precision of the estimates. Table 4.3 presents the respective results from the simulations of data from informative priors. The first method performs slightly better in terms of the precision of the point estimates in all kinds of modifications that we set.

Table 4.3: L_2 deviations of the informative prior based on simulations of data from informative prior. Two different methods for measuring the posterior informative point estimates are compared.

Informative		sample size of $\boldsymbol{z}_0$, ζ, τ	# of simulations
Method 1	$10^{-2}\times3.088$	200	1000
Method 2	$10^{-2}\times3.925$		
Method 1	$10^{-2}\times2.936$	400	1000
Method 2	$10^{-2}\times4.893$		
Method 1	$10^{-2}\times3.123$	600	2000
Method 2	$10^{-2}\times7.460$		
Method 1	$10^{-2}\times3.911$	800	3000
Method 2	$10^{-2}\times5.067$		
Method 1	$10^{-2}\times4.538$	1000	4000
Method 2	$10^{-2}\times6.244$		

Chapter 5

Empirical Study

5.1 Financial Data and sample estimates

In order to perform the empirical part of the thesis, we have considered the daily logarithmic returns of five companies that are listed on *Nasdaq Stockholm* (Stockholm Stock Exchange). Hence, a domestic portfolio is analyzed. The portfolio is composed of companies which belong to the index (H&M, Volvo, Nordea) meaning that they are traded frequently, together with other two companies (Clas Ohlson, SAS) that are traded infrequently (low volume). The data were downloaded from `https://finance.yahoo.com` and were directly loaded into `RStudio` using the `get.hist.quote` command (see Zeileis et al. 2005, p.20-21). More precisely, the *adjusted closing* prices of the stocks were downloaded for a time horizon starting from 3/1/2011 until 31/12/2015. The dataset was divided into two sub-periods. The period from 3/1/2011 - 31/12/2013 served as a prerun meaning that the sample estimates of the covariance matrix and the vector of expected returns were used as inputs in order to derive the posterior distributions. In addition, the MVP was calculated based on the sample covariance matrix. In the following tables 5.1, 5.2, 5.3, we present the relevant estimates. The data from 1/1/2014 - 31/12/2015 were used as the in-sample period and can be interpreted as the data for which we updated our prior belief about the prior distribution of the parameters. The sample estimates of the mean vector of returns and the covariance matrix are presented in tables 5.4, 5.5.

Table 5.1: Sample means of daily logarithmic returns starting from 3/1/2011 until 31/12/2013.

	H&M	Clas Ohlson	Nordea	SAS	Volvo
$\boldsymbol{\mu}_{\text{prior}} \times 10^4$	5.005	2.972	2.129	-5.361	-3.580

Table 5.2: Sample Variance-Covariance matrix $\boldsymbol{S}_{\text{prior}}$ of daily logarithmic returns starting from 3/1/2011 until 31/12/2013. The values presented have been multiplied with 10^4.

	H&M	Clas Ohlson	Nordea	SAS	Volvo
H&M	1.963				
Clas Ohlson	1.098	3.681			
Nordea	1.546	1.438	3.146		
SAS	1.395	1.127	1.962	12.874	
Volvo	1.705	1.815	2.461	2.233	4.131

Table 5.3: Sample weights of the MVP for the period starting from 1/3/2011 until 31/12/2013.

	H&M	Clas Ohlson	Nordea	SAS	Volvo
$\boldsymbol{\omega}_{\text{prior}}$	0.6589	0.2322	0.1414	0.0297	-0.0623

Note that the prior sample estimates are meaningful only for the models which assume informative priors. Hence, as already mentioned, the estimated parameters were used as input in the prior distributions by setting $\boldsymbol{\mu}_c = \boldsymbol{\mu}_{\text{prior}}$, $\boldsymbol{S}_c = \boldsymbol{S}_{\text{prior}}$ and $\boldsymbol{w}_I = \boldsymbol{\omega}_{\text{prior}}$.

Table 5.4: Sample means $\overline{\boldsymbol{X}}$ of daily logarithmic returns starting from 1/1/2014 until 31/12/2015.

	H&M	Clas Ohlson	Nordea	SAS	Volvo
$\overline{\boldsymbol{X}} \times 10^4$	1.605	5.943	1.612	7.509	-0.130

Table 5.5: Sample Variance-Covariance matrix $\boldsymbol{S}$ of daily logarithmic returns starting from 1/1/2014 until 31/12/2015. The values presented have been multiplied with 10^4.

	H&M	Clas Ohlson	Nordea	SAS	Volvo
H&M	1.446				
Clas Ohlson	0.519	2.248			
Nordea	0.866	0.650	2.146		
SAS	0.393	0.365	0.753	6.377	
Volvo	0.926	0.639	1.326	0.854	2.989

In-sample estimates for the average daily logarithmic returns produce better results compared to their prior period counterparts. Note that there is only one stock (Volvo) with negative average return. In addition, the variance-covariance matrix of the prior period contains greater values and implies a more unstable performance for our portfolio of stocks. Nevertheless, in both periods, the covariances are positive meaning that the chosen stocks move towards the same direction.

5.2 Posterior distributions

All the plots of the posterior distributions are presented in figure 5.1. In table 5.6, we also present the Bayesian point estimates of the mean under all different models. The diffuse and the non informative prior appear, not surprisingly, quite similar. Note that the location (mean) parameters, which were calculated based on the MVP for the in-sample data, are identical. The superiority of the informative prior lies on the fact that it has managed to incorporate our pre-experimental knowledge about the optimal allocation resulting in much wider densities around the point estimates of the mean. On the contrary, the conjugate prior has, certainly, failed to fit our prior beliefs. There is no indication that the densities have moved towards the prior weights. In addition, the probability of getting the prior weights of, for example, Volvo or SAS is zero or close to zero. The prior information for both the informative and the conjugate prior corresponds to the minimum variance portfolio which was calculated from historical data covering the period from 1/1/2011 - 31/12/2013.

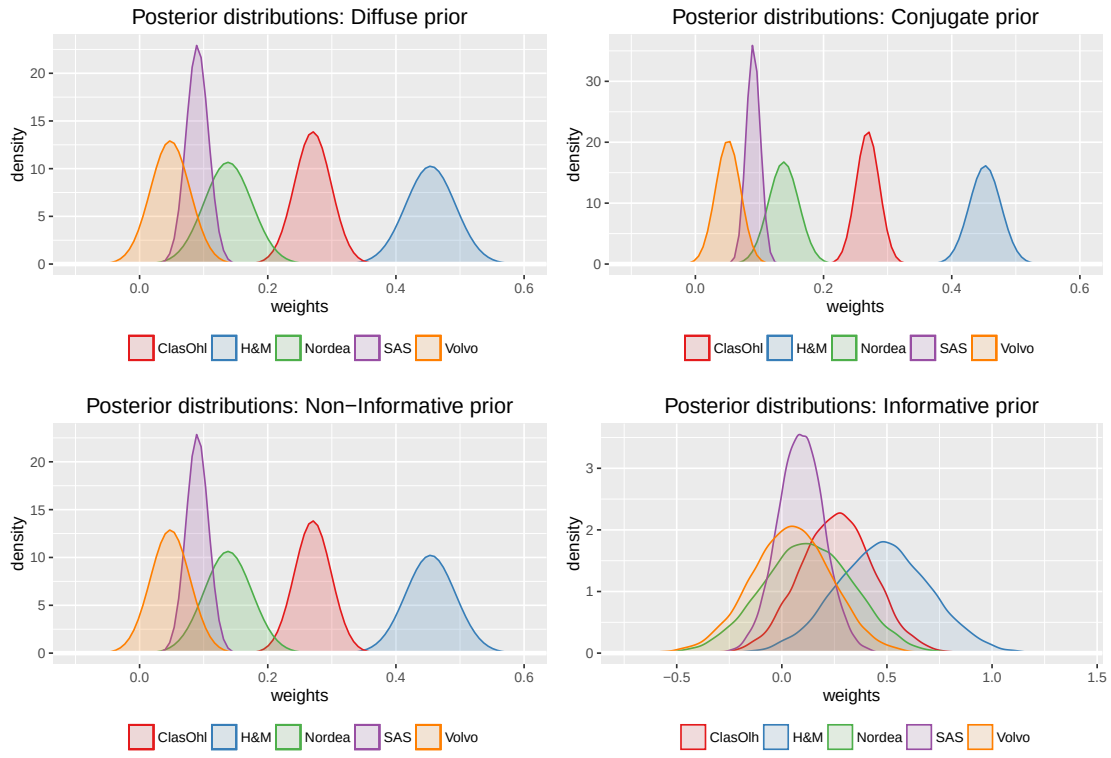

Figure 5.1: Posterior densities for the portfolio weights of H&M, Clas Ohlsson, Nordea, SAS, Volvo under the four different models for the period from 1/1/2014 - 31/12/2015. The prior parameters for the case of Conjugate and Informative prior have been calculated using historical data from 1/1/2011 - 31/12/2013.

Table 5.6: Posterior point estimates of the mean under all different priors. Note that the last column represents the weights derived from the classical frequentist approach which coincide with the weights from the diffuse/uninformative prior.

Weights	Diff.	Conj.	Inf.	Frequentist
H&M	0.4540	0.4521	0.4816	0.4540
Clas Ohl.	0.2702	0.2689	0.2605	0.2702
Nordea	0.1379	0.1377	0.1236	0.1379
SAS	0.09	0.0903	0.0901	0.09
Volvo	0.0478	0.0510	0.0441	0.0478

Next, based on the posterior distributions, the 95% credible sets, i.e the intervals that the true parameter is contained with probability of 95%, are presented in figure 5.2. For the informative prior, the credible sets were provided based on the sample data, i.e 20000 simulated data for each weight. The informative prior has produced the widest confidence intervals, something which makes sense given the posterior distributions from figure 5.1. On the other hand, the credible intervals arised from the conjugate prior are more narrow than those based on the uninformative priors which goes against the main goal of incorporating prior beliefs.

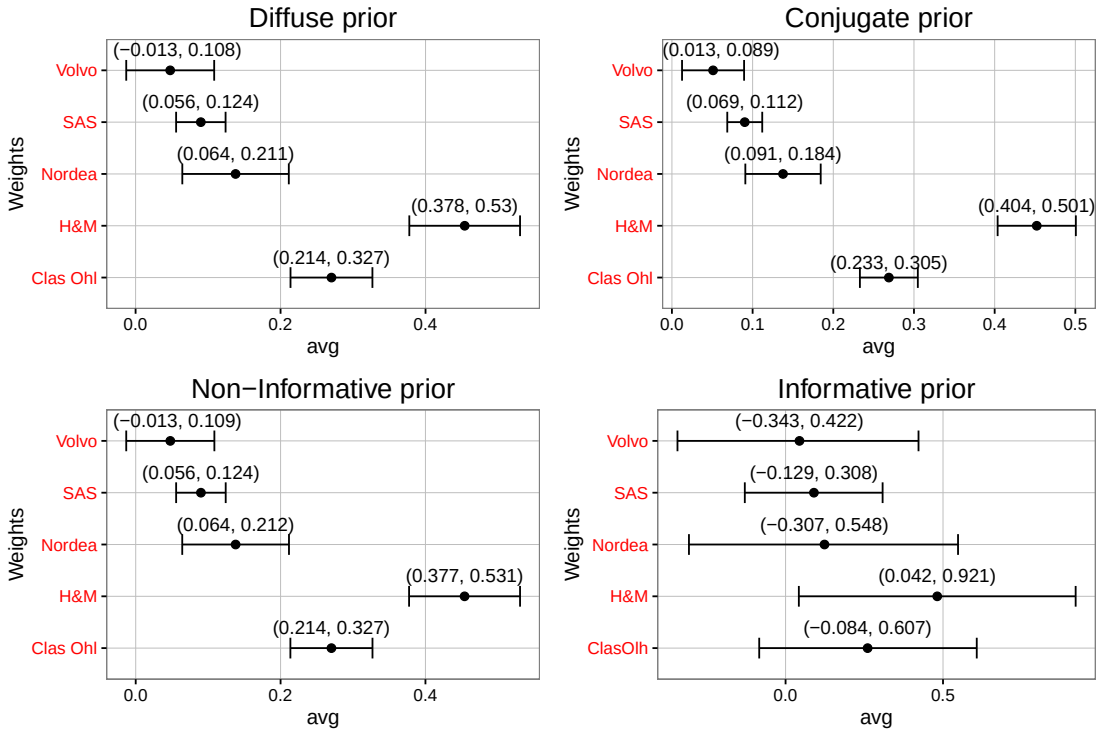

Figure 5.2: 95% Bayesian credible intervals for the portfolio weights of H&M, Clas Ohlsson, Nordea, SAS, Volvo under all different models.

From a practitioner's point of view, the probability of putting a positive/negative weight on an asset might be of interest. In table 5.7, the probabilities of those complementary events are presented. In the case of the informative prior, all the weights, more or less, are likely to be assigned a negative value. For all other priors,

only the weights of Volvo have a realistic chance of being negative. Note that the probabilities of the informative prior were calculated from the empirical distribution of the simulated data according to

$$F_N(x) = \frac{\#\text{ of data} \leq x}{N}$$

where N is the total number of data. In our case the interpretation is that the probability of getting a negative value is equal to the proportion of the simulated values that are less or equal to zero. Furthermore, one might want to test the weights of one asset against the other. This is done by setting, for example, $\boldsymbol{L} = (1, -1, 0, 0, 0)$, $\boldsymbol{L} = (1, 0, -1, 0, 0)$ and so forth meaning that we are testing the differences of the first weight to all the others. In table 5.3, the distributions of the differences between the weight of H&M and the other weights are presented. Among the priors, only the informative prior provides a substantial chance of observing that the differences of the random variables of the differences of the weights (seen as random variables) are negative (see figure 5.3 and table 5.8), i.e $P(Y = W_{\text{H\&M}} - W_j \leq 0) > 0$ where j represents the rest of the weights. Another thing that could possibly be of interest for an investor is the correlation of the different weights. In proposition 1-(a) and 1-(b), the posterior covariance matrices of the diffuse and the conjugate prior up to a proportionality constant are defined as $\boldsymbol{R}_d$ and $\boldsymbol{R}_c$ respectively. Hence, it is possible to derive the correlation matrices of the weights by using the setting (4.1). The correlation between all pairs of weights range from moderate negative to very weak negative close to zero (see figure 5.4). Of course, a negative correlation implies that the weights move towards opposite directions.

Table 5.7: Posterior probabilities for all the models that the weights of the assets are either greater or less than zero.

		Posterior Probabilities	
Models	Weights	$\omega \leq 0$	$\omega > 0$
Diffuse	H&M	0	1
	Clas Ohlson	0	1
	Nordea	$10^{-4}\times 1.246$	0.9998
	SAS	$10^{-7}\times 1.553$	0.9999
	Volvo	$10^{-2}\times 6.124$	0.9387
Conjugate	H&M	0	1
	Clas Ohlson	0	1
	Nordea	$10^{-9}\times$ 4.576	0.9999
	SAS	$10^{-16}\times 3.451$	0.9999
	Volvo	$10^{-3}\times 4.7$	0.9952
Informative	H&M	0.0154	0.9846
	Clas Ohlson	0.0698	0.9302
	Nordea	0.28565	0.71435
	SAS	0.20535	0.79465
	Volvo	0.4063	0.5937
Non-Informative	H&M	0	1
	Clas Ohlson	0	1
	Nordea	$10^{-4}\times 1.299$	0.9998
	SAS	$10^{-7}\times 1.680$	0.9999
	Volvo	$10^{-2}\times 6.178$	0.9382

Table 5.8: Posterior Probabilities for the differences of the random variables of the weights under the informative prior. Four differences are compared and correspond to the comparison of the H&M weight against the other weights.

		Posterior Probabilities	
Models	Differences	$\omega_{\text{H\&M}} - \omega_j \leq 0$	$\omega_{\text{H\&M}} - \omega_j > 0$
Informative	H&M vs Clas	0.19195	0.80805
	H&M vs Nordea	0.0861	0.9139
	H&M vs SAS	0.05755	0.94245
	H&M vs Volvo	0.03935	0.96065

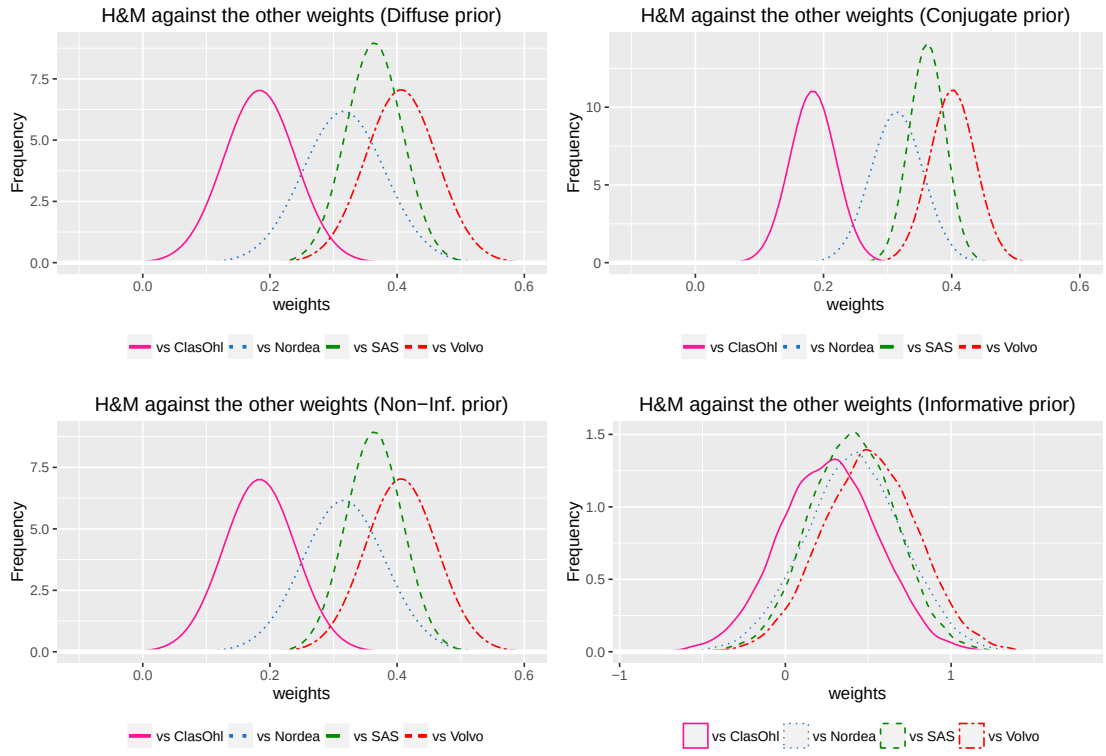

Figure 5.3: Posterior distributions of the difference of the H&M weight and the other weights, i.e the distributions of the difference between pairs of random variables.

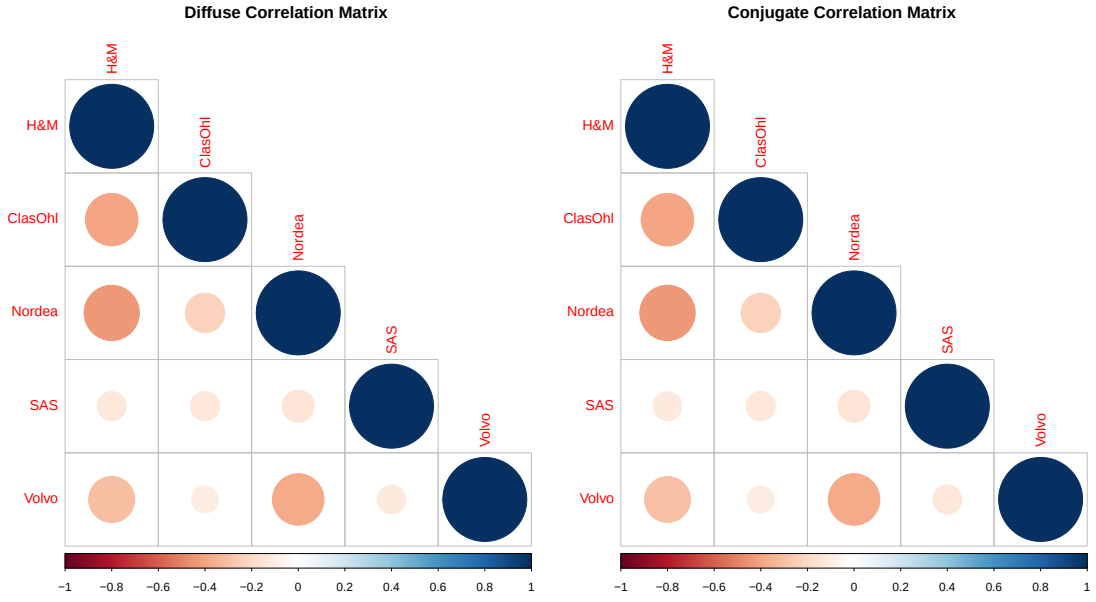

Figure 5.4: Correlation matrices for the diffuse and conjugate posterior distributions.

Chapter 6

Conclusion

In this paper, we have looked into a Bayesian method that can be used in portfolio theory. More specifically, we have investigated the *minimum variance portfolio* under a Bayesian approach. Throughout the thesis, four different models have been compared.

In the simulation study, the aim was to compare the mean point estimates of the weights under the posterior distributions by generating data from different priors. Towards that direction, we also wanted to test the robustness of the priors which generated the data by considering different values for the prior parameters. In section 4.2, since we have simulated data from the conjugate prior we would expect that the posterior mean of the weights under the conjugate prior would outperform the other posterior means in terms of the expected quadratic loss function, i.e the average deviations of the mean estimates from the "true" parameters based on 1000 simulations. Indeed, the smallest values of the deviations correspond to the model which coincides with the true model generating the data, i.e the conjugate prior. The only occasion in which different results are produced is when the 'true' parameters coincide with the prior assumption of the informative prior for the weights. We could also state that this is a strong indication that the informative prior has managed to embody our prior beliefs. In addition, since our sample size is small (# of obs = 50) it is actually a good thing that the informative prior has not adjusted to the data. In section 4.3, the data were generated from the informative prior and, thus, the posterior mean estimates of the weights under the informative prior were closer to the "true" weights compared to the posterior mean estimates of the other models.

In chapter 5, the goal was to analyze the posterior densities of the weights of a domestic portfolio based on two years of daily data. A couple of models (informative ones) required specification for some of their parameters. This was provided by estimating parameters from a prior period which served as our pre-experimental knowledge. The diffuse and the non-informative prior produced almost identical results. Note that the conjugate prior did not live up to our expectations. The posterior distributions had smaller standard deviations from the uninformative priors resulting in more narrow curves failing to incorporate estimates from the prior period (compare, for example, the weights of H&M and Volvo in table 5.3 to the posterior distributions under the conjugate prior in figure 5.2). Probably the big sample size (daily data for two financial years) forced our prior belief to vanish. On the other hand, the posterior distributions provided by the informative prior were very wide. The point estimates of the means were in between the prior MVP and

the in-sample MVP estimates suggesting that the informative prior has successfully incorporated prior knowledge.

Appendix A

A.1 Shifted Student-t distribution

A random variable X which has a t-distribution with *mean* μ, *scale* parameter σ and ν *degrees of freedom* $\big(X \sim T(\mu, \sigma, \nu)\big)$ is just a linear transformation of a standard Student t random variable expressed as:

$$g(T) = X = \mu + \sigma \cdot T.$$

The **cumulative distribution** function after the transformation is given as follows:

$$\begin{aligned} F_X(x) &= P(X \leq x) = P(\mu + \sigma \cdot T \leq x) \\ &= P\left(T \leq \frac{x-\mu}{\sigma}\right) \\ &= F_T\left(\frac{x-\mu}{\sigma}\right) \end{aligned} \tag{A.1}$$

The **probability density** function is derived by making use of the transformation theorem as follows:

$$f_X(x) = f_T\big(g^{-1}(x)\big) \cdot \left|\frac{\partial}{\partial x}\big(g^{-1}(x)\big)\right| \tag{A.2}$$

Since $g(T) = X = \mu + \sigma \cdot T$, the inverse function is given by $g^{-1}(T) = \frac{T-\mu}{\sigma}$. Therefore, (A.2) can be written as:

$$f_X(x) = f_T\left(\frac{x-\mu}{\sigma}\right) \cdot \frac{1}{\sigma} \tag{A.3}$$

The **inverse cumulative distribution** function corresponding to $p = F_X(x) = F_T\big(\frac{x-\mu}{\sigma}\big)$ is defined as:

$$\begin{aligned} \frac{x-\mu}{\sigma} = F_T^{-1}(p) \implies x = \sigma \cdot F_T^{-1}(p) + \mu \\ \implies F_X^{-1}(p) = \sigma \cdot F_T^{-1}(p) + \mu \end{aligned} \tag{A.4}$$

In `R`, in order to calculate the cumulative distribution, the inverse cumulative distribution and the probability density we used the following functions:

```
pt_ls <- function(x, df, mu, a, lower.tail=T){
pt((x - mu)/a, df, lower.tail=lower.tail)}
qt_ls <- function(prob, df, mu, a) qt(prob, df)*a + mu
dt_ls <- function(x, df, mu, a) 1/a * dt((x - mu)/a, df).
```

A.2 Cholesky Decomposition

In sections 4.4 and 5.2, we came across the calculation of the square root of a real valued symmetric and square matrix. This can be achieved by deriving the cholesky decomposition of this matrix. Every positive definite matrix $\boldsymbol{A}$ can be factorized as $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^T$. $\boldsymbol{L}$ is a lower triangular matrix with positive diagonal elements and is called the *cholesky* factor of $\boldsymbol{A}$. Let $\boldsymbol{A}$ be a $n \times n$ square matrix, i.e

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{bmatrix} = \boldsymbol{L}\boldsymbol{L}^T,$$

and let

$$\boldsymbol{L} = \begin{bmatrix} l_{11} & 0 & 0 & \dots & 0 \\ l_{21} & l_{22} & 0 & \dots & 0 \\ l_{31} & l_{32} & l_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & l_{n3} & \dots & l_{nn} \end{bmatrix}, \boldsymbol{L}^T = \begin{bmatrix} l_{11} & l_{21} & l_{31} & \dots & l_{n1} \\ 0 & l_{22} & l_{32} & \dots & l_{n2} \\ 0 & 0 & l_{33} & \dots & l_{n3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & l_{nn} \end{bmatrix}.$$

Multiplying $\boldsymbol{L}$ by $\boldsymbol{L}^T$ we get

$$\boldsymbol{L}\boldsymbol{L}^T = \begin{bmatrix} l_{11}^2 & l_{11}l_{21} & l_{11}l_{31} & \dots & l_{11}l_{n1} \\ l_{11}l_{21} & l_{21}^2 + l_{22}^2 & l_{21}l_{31} + l_{22}l_{32} & \dots & l_{21}l_{n1} + l_{22}l_{n2} \\ l_{11}l_{31} & l_{21}l_{31} + l_{22}l_{32} & l_{31}^2 + l_{32}^2 + l_{33}^2 & \dots & l_{31}l_{n1} + l_{32}l_{n2} + l_{33}l_{n3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{11}l_{n1} & l_{21}l_{n1} + l_{22}l_{n2} & l_{31}l_{n1} + l_{32}l_{n2} + l_{33}l_{n3} & \dots & l_{n1}^2 + l_{n2}^2 + \dots + l_{nn}^2 \end{bmatrix}.$$

Thus, by matching the elements of $\boldsymbol{A}$ to the elements of $\boldsymbol{L}\boldsymbol{L}^T$ we get that all the diagonal elements of $\boldsymbol{L}$ can be derived using the following formula

$$l_{kk} = \sqrt{a_{kk} - \sum_{j=1}^{k-1} l_{kj}}, \text{ for } k = 1, 2, \dots, n,$$

whereas the other non diagonal elements of $\boldsymbol{L}$ can be calculated using

$$l_{ki} = \frac{a_{ki} - \sum_{j=1}^{i-1} l_{ij}l_{kj}}{l_{ii}}, \text{for } k = 1, 2, \dots, n \text{ and } i < k.$$

In `R`, we used the `base` function `chol` which returns the upper diagonal matrix $\boldsymbol{L}^T$.

A.3 Multivariate Normal distribution

In chapter 4, MN distribution was used in various occasions. First of all, we simulated the mean vector of returns from MN, since a MN mean vector is the conjugate prior to a MN likelihood. Secondly, the simulated data, also, followed a MN distribution. Thus, it would be appropriate to provide some relevant background. Recall that the **univariate normal** distribution is given by

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\Big\{ -\frac{1}{2}\Big(\frac{x-\mu}{\sigma}\Big)^2 \Big\}$$

where μ is the mean and σ^2 is the variance. Isolating the exponent part we note that

$$\text{exponent} = -\frac{1}{2}\Big(\frac{x-\mu}{\sigma}\Big)^2 = -\frac{1}{2}(x-\mu)\big(\sigma^2\big)^{-1}(x-\mu).$$

Now for the multivariate case, the exponent part will be

$$(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}),$$

where $\boldsymbol{\mu}$ is a mean vector and Σ is a matrix of covariances. Hence, the **multivariate** analogue of the univariate case is given by

$$f(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\Big\{ -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \Big\}$$

where μ is a $n\times 1$ vector, $\boldsymbol{\Sigma}$ is a $n\times n$ matrix and $|\boldsymbol{\Sigma}|$ is the determinant of matrix $\boldsymbol{\Sigma}$. For more information regarding the MN distribution one can address to Do (2008). In R, in order to simulate numbers from MN distribution we used the function mvrnorm from the MASS package.

A.4 Wishart and inverse Wishart distributions

The Wishart distribution is a family of distributions defined over **positive definite** random matrices. More precisely,
if $\boldsymbol{S}$ is a $p\times p$ matrix and can be written as $\boldsymbol{S} = \boldsymbol{X}^T\boldsymbol{X}$ where $\boldsymbol{X}$ is a $n\times p$ matrix of data for which each row $\boldsymbol{x}_i \sim \mathcal{N}_p(\boldsymbol{0}_p, \boldsymbol{V})$ for $i = 1, 2, \ldots, n$, *then* $\boldsymbol{S}$ is said to have a Wishart distribution with $n > p-1$ *degrees of freedom* and *scale matrix* $\boldsymbol{V}$ $(p\times p)$. We denote

$$\boldsymbol{S} \sim \mathcal{W}_p(\boldsymbol{V}, n).$$

Note that $\boldsymbol{X}^T\boldsymbol{X}$ is a scatter matrix which is used as estimate for a covariance matrix. Hence, matrix $\boldsymbol{V}$ can be thought of as a covariance matrix.
In addition, if $\boldsymbol{S} \sim \mathcal{W}_p(\boldsymbol{V}, n)$ then $\boldsymbol{S}^{-1}$ is said to follow an inverse Wishart distribution denoted by

$$\boldsymbol{S}^{-1} \sim \mathcal{W}_p^{-1}(\boldsymbol{V}^{-1}, n+p+1).$$

For a more formal definition of the *Wishart* and *inverse Wishart* distributions together with their properties address to Nydick (2012)[ch. 1,2].
In R, for the generation of *Wishart* and *inverse Wishart* distributed matrices we used the function rWishart from stats package and the function riwish from the MCMCpack package respectively.

A.5 Kernel Density Estimation

In section 5.2, we plotted the posterior distribution of the informative prior using the stochastic representation formula (3.18). This was done by applying the kernel density estimation method which is a non parametric density estimation technique based on sample data from some distribution with unknown density. It is similar to how one creates histograms. More formally,

$$\widehat{f_h} = \frac{1}{nb} \sum_{i=1}^{n} K\left(\frac{x - x_i}{b}\right)$$

where $K(.)$ is a chosen kernel and $b > 0$ is called the bandwidth. One choice of density, the one we have used, is a gaussian kernel i.e $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$. Note that the bandwidth was set as the default value obtained by *Silverman's* rule of thumb (see Silverman 1986, p.48) as

$$b \approx 1.06 \cdot \widehat{\sigma} \cdot n^{-1/5}$$

where $\widehat{\sigma}$ is the sample standard deviation and n is the size of the sample. In `R`, we applied density estimation by calling `geom_density` from `ggplot2` package.

Appendix B

R code

The code below was used for the calculations of the expected quadratic L-2 loss functions for the top two sub-tables of table 4.1.

```
##########################################
#### Input: D_c-Diagonal matrix of  ####
####         the Variances of S_c    ####
#### Ouput: deviations, W_true, S_c ####
####         num. of iteration for   ####
####         correlation matrix R_xx ####
##########################################
L2DevConj<-function(D_c){
  library(MCMCpack)#Simulate from multivariate normal
  library(matrixcalc)#check positive definite matrix
  ##############################################################
  #### Define the (pxk) matrices L, with p=1 and k=5 ####
  ##############################################################
  L_1<-matrix(c(1,0,0,0,0),1,5);L_2<-matrix(c(0,1,0,0,0),1,5)
  L_3<-matrix(c(0,0,1,0,0),1,5);L_4<-matrix(c(0,0,0,1,0),1,5)
  L_5<-matrix(c(0,0,0,0,1),1,5)
  L<-list(L_1,L_2,L_3,L_4,L_5)
  ##################################
  #### define some constants ####
  ##################################
  delta1<-1; delta2<-0.5
  p<-nrow(L_1); S_I<-1; w_I<-rep(0.2, 5)
  ones<-matrix(1,5, 1)
  k_c<-nu_c<-56;k<-5
  set.seed(3)
  ##############################################
  #### Generate the correlation matrix    ####
  #### which will be used together with   ####
  #### D_c for the generation of the      ####
  #### prior precision covariance matrix  ####
  #### S_c. We will force S_c to be       ####
  #### positive definite by modifying the ####
  #### correlation matrix                 ####
  ##############################################
  R_xx<-matrix(0, 5, 5)
  S_c<-matrix(0, 5, 5)
  m<-0
  while(!is.positive.definite(S_c)){
    for(i in 1:5){
      R_xx[i,i:5]<-c(1, runif(5-i, -1, 1))
```

```
      R_xx[i:5, i]<-R_xx[i,i:5]
    }
    S_c<-chol(D_c)%*%R_xx%*%chol(D_c)
    S_c<-round(S_c, 5);m<-m+1
      }
#### Lemma 1 ####
W_true<-solve(S_c)%*%ones/as.numeric(t(ones)%*%solve(S_c)%*%ones)
#########################################################
#### Create the variables that we will use ####
#### in order to store the values of the 5 ####
#### weights for every simulation          ####
#########################################################
THETADIFF<-matrix(0,5,1000)
thetaconj<-matrix(0,5,1000)
THETAINF<-matrix(0,5,1000)
#########################################################
#### the variable in which we will store       ####
#### the squared differences for every weight, ####
#### e.g the quadratic loss function           ####
#########################################################
loss.table<-lapply(1:3,function(x) matrix(0,5,1000))
#################################
#### L-2 loss function    ####
#### for 1000 simulations ####
#################################
for(i in 1:1000){
  ########################################
  #### Generate a Covariance matrix ####
  ########################################
  Sigma.sim <- riwish(nu_c-k-1,S_c)
  ##############################################
  #### Generate \mu vector and the data ####
  ##############################################
  mus.sim<-mvrnorm(1,rep(0,5),Sigma.sim/k_c)
  data.sim<-mvrnorm(50,mus.sim,Sigma.sim)
  covar<-cov(data.sim)
  nu_I<-n<-nrow(data.sim)
  ############################
  #### estimated weights ####
  ############################
  DEN_est<-t(ones)%*%solve(covar)%*%ones
  NUM_est<-solve(covar)%*%ones
  W_est<-NUM_est/as.numeric(DEN_est)
  ####################################
  #### calculate the matrix R_d ####
  ####################################
  R_D<-solve(covar)-(solve(covar)%*%ones%*%t(ones)%*%solve(covar))/
  as.numeric(DEN_est)
  ##########################
  #### diffuse weights ####
  ##########################
  thetadif1.sim<-as.numeric(L_1%*%W_est);thetadif2.sim<-as.numeric(L
  _2%*%W_est)
  thetadif3.sim<-as.numeric(L_3%*%W_est);thetadif4.sim<-as.numeric(L
  _4%*%W_est)
  thetadif5.sim<-as.numeric(L_5%*%W_est)
  ###############################################
  #### Save the weights of each simulation ####
```

```
    #### based on the dif. prior                ####
    ##################################################
    THETADIFF[,i]<-rbind(thetadif1.sim,thetadif2.sim,thetadif3.sim,
    thetadif4.sim,thetadif5.sim)
    ##############################
    #### conjugate weights ####
    ##############################
    X_BAR<-matrix(colMeans(data.sim),5,1)
    R_C<-(n*X_BAR)/(n+k_c)
    V_C<-(n-1)*covar+S_c+(n+k_c)*R_C%*%t(R_C)+n*X_BAR%*%t(X_BAR)
    DENOM<-t(ones)%*%solve(V_C)%*%ones
    thetacon1.sim<-as.numeric(L_1%*%solve(V_C)%*%ones/DENOM)
    thetacon2.sim<-as.numeric(L_2%*%solve(V_C)%*%ones/DENOM)
    thetacon3.sim<-as.numeric(L_3%*%solve(V_C)%*%ones/DENOM)
    thetacon4.sim<-as.numeric(L_4%*%solve(V_C)%*%ones/DENOM)
    thetacon5.sim<-as.numeric(L_5%*%solve(V_C)%*%ones/DENOM)
    ##################################################
    #### Save the weights of each simulation ####
    #### based on the conj. prior               ####
    ##################################################
    thetaconj[,i]<-rbind(thetacon1.sim,thetacon2.sim,thetacon3.sim,
    thetacon4.sim,thetacon5.sim)
    ##############################
    #### informative prior ####
    ##############################
    for(j in 1:4){
      R<-1/delta2+(n-1)/as.numeric(DEN_est)
      Z_0<-rnorm(20000)
      TAU<-rgamma(20000, shape = (n-k+p+nu_I-1)/2, scale = 2)
      P_two<-(n-1)/as.numeric(L[[j]]%*%R_D%*%t(L[[j]]))
      P_one<-1/(1/S_I+P_two)
      V.I<-1/(TAU*P_one+P_two)
      r.I<-V.I*(TAU*P_one*w_I[j]+P_two*THETADIFF[j,i])
      h.I<-R+TAU*w_I[j]*P_one*w_I[j]+THETADIFF[j,i]*P_two*THETADIFF[j,i]-
    r.I/V.I*r.I
      ZETA<-rgamma(20000, shape = (n-k+2*p+2*delta1)/2, scale = 2/h.I)
      #### stochastic representation formula ####
      THETAINF[j,i]<-mean(r.I+ZETA^(-1/2)*sqrt(V.I)*Z_0)
    }
    THETAINF[5, i]<-1-sum(THETAINF[, i])
    ####
    loss.table[[1]][,i]<-(THETADIFF[,i]-W_true)^2
    loss.table[[2]][,i]<-(thetaconj[,i]-W_true)^2
    loss.table[[3]][,i]<-(THETAINF[,i]- W_true)^2
  }
  ########################################
  #### expected L-2 loss function ####
  ########################################
  return(list(deviations=data.frame(diffuse = sum(colSums(loss.
    table[[1]]))/1000,
                      conjugate = sum(colSums(loss.table[[2]]))/1000,
                      informative = sum(colSums(loss.table[[3]]))/1000), W
    _true= W_true, S_c=S_c, num.of.dif.correlation.matrices=m))
}
```

Bibliography

Abramowitz, M and IA Stegun (1964). *Handbook of functions with formulas, graphs and mathematical tables*.

Barry, Christopher B (1974). "Portfolio analysis under uncertain means, variances, and covariances". In: *The Journal of Finance* 29.2, pp. 515–522.

Bodnar, Taras, Stepan Mazur, and Yarema Okhrin (2017). "Bayesian Estimation of the Global Minimum Variance Portfolio". In: *European Journal of Operational Research* 256, pp. 292–307.

Bodnar, Taras and Wolfgang Schmid (2008). "A test for the weights of the global minimum variance portfolio in an elliptical model". In: *Metrika* 67.2, pp. 127–143.

Brown, Stephen Jeffery (1976). "Optimal portfolio choice under uncertainty: A Bayesian approach". PhD thesis. University of Chicago, Graduate School of Business.

Carlin, Bradley P and Thomas A Louis (2008). *Bayesian methods for data analysis*. CRC Press.

Chopra, Vijay K and William T Ziemba (1993). "The effect of errors in means, variances, and covariances on optimal portfolio choice". In:

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal (2009). "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?" In: *Review of Financial Studies* 22.5, pp. 1915–1953.

Do, Chuong B (2008). "The multivariate Gaussian distribution". In: *Section Notes, Lecture on Machine Learning, CS* 229.

Goldfarb, Donald and Garud Iyengar (2003). "Robust portfolio selection problems". In: *Mathematics of Operations Research* 28.1, pp. 1–38.

Hudson, Robert and Andros Gregoriou (2010). "Calculating and comparing security returns is harder than you think: A comparison between logarithmic and simple returns". In: *Available at SSRN 1549328*.

Jorion, Philippe (1992). "Portfolio optimization in practice". In: *Financial Analysts Journal* 48.1, pp. 68–74.

Klein, Roger W and Vijay S Bawa (1976). "The effect of estimation risk on optimal portfolio choice". In: *Journal of Financial Economics* 3.3, pp. 215–231.

Markowitz, Harry (1952). "Portfolio selection". In: *The journal of finance* 7.1, pp. 77–91.

Merton, Robert C (1980). "On estimating the expected return on the market: An exploratory investigation". In: *Journal of financial economics* 8.4, pp. 323–361.

Michaud, Richard O (1989). "The Markowitz optimization enigma: is' optimized'optimal?" In: *Financial Analysts Journal* 45.1, pp. 31–42.

Murphy, Kevin P (2007). "Conjugate Bayesian analysis of the Gaussian distribution". In: *def* $1.2\sigma 2$, p. 16.

Nydick, Steven W (2012). "The wishart and inverse wishart distributions". In: *J. Stat.*

Rachev, Svetlozar T et al. (2008). *Bayesian methods in finance*. Vol. 153. John Wiley & Sons.

Silverman, Bernard W (1986). *Density estimation for statistics and data analysis*. Vol. 26. CRC press.

Stein, Charles (1956). "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution". In: *Proceedings of the Third Berkeley symposium on mathematical statistics and probability*. Vol. 1. 399, pp. 197–206.

Tunaru, Radu (2002). "Hierarchical Bayesian models for multiple count data". In: *Austrian Journal of statistics* 31.3, pp. 221–229.

Zeileis, Achim and Gabor Grothendieck (2005). "zoo: S3 Infrastructure for Regular and Irregular Time Series". In: *Journal of Statistical Software* 14.1, pp. 1–27. ISSN: 1548-7660. DOI: 10.18637/jss.v014.i06. URL: https://www.jstatsoft.org/index.php/jss/article/view/v014i06.

Zivot, Eric (2014). *Return Calculations*. Accessed: 2016-11-23, p. 35. URL: http://faculty.washington.edu/ezivot/econ424/returnCalculations.pdf.