Stockholms universitet

Prior Impact on Optimal Portfolio Selection

Max Sjödin

Masteruppsats i matematisk statistik
Master Thesis in Mathematical Statistics

Masteruppsats 2021:16
Matematisk statistik
Juni 2021

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Master Thesis **2021:16**
http://www.math.su.se

Prior Impact on Optimal Portfolio Selection

Max Sjödin*

June 2021

Abstract

In this thesis we solve the problem of optimal portfolio selection from the Bayesian perspective. We consider four priors: the diffuse, the conjugate, the hierarchical, and the objective-based prior. For the diffuse and the conjugate prior we use the stochastic representation in order to draw samples from the posterior predictive distribution. For the hierarchical and the objective-based prior we derive the conditional posterior distributions, of the parameters of the asset returns, in order to draw samples from the posterior predictive distribution. An extensive comparison study is performed via Monte Carlo simulation in order to assess the performance based on the suggested performance measures. The Bayesian efficient frontier, the set of optimal portfolios, is constructed and compared to the sample efficient frontier, which is known to be overoptimistic, and the population efficient frontier. Theoretically and using real data from the Stockholm market we show that most of the Bayesian approaches outperform the frequentist approach.

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

Acknowledgement

I want to thank my supervisor Taras Bodnar, professor in financial mathematics at Stockholm University, for having answered all my questions on Bayesian statistics and for providing valuable feedback during the writing process of this thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The field of optimal portfolio theory began with the research done by Markowitz (1952)[26]. In this paper he stresses the importance of diversification and the procedure of finding reasonable values for parameters of the asset returns. He believed that this procedure should combine statistical methods and practical judgement. Merton (1972)[27] showed that the set of all optimal portfolios, known as the efficient frontier, creates a parabola in the mean-variance space. As further research was conducted, the efficient frontier has become well known and its properties well examined. However, the practical problem of estimating the parameters, of the efficient frontier, still remains. The estimated efficient frontier, based on frequentist statistics, is known to be optimistic. This was shown by Broadie (1993)[15], who, using small amounts of historical data in the parameter estimation, exposed his model to estimation errors and suggested a method for adjusting the optimistic bias.

While these studies contribute to the research of optimal portfolio selection and the efficient frontier from the frequentist point of view, Bayesian statistics grew popular over the last few years for a number of reasons. Bayesian theory is regarded to resemble the way humans use information. Most importantly, the Bayesian framework allows the incorporation of subjective beliefs on the outcome of future events, this violates the fundamental principles of frequentist statistics (see, e.g., Avramov and Zhou (2010)[2]). Anderson and Cheng (2016)[1] proposed a Bayesian-averaging portfolio choice strategy with excellent out-of-sample performance. Brandt (2010)[14] gives an excellent review of modern approaches to portfolio selection, including Bayesian approaches.

In this thesis we will focus on four different priors: the diffuse prior; the conjugate prior; the hierarchical prior; and the objective-based prior. The diffuse and conjugate prior are regarded as well established in the Bayesian literature and in research applying Bayesian methods to portfolio selection (see, e.g., Frost and Savarino (1986)[17]; Gelman, Carlin, Stern, and Rubin (2004)[18]; Klein and Bawa (1976)[25]; Rachev et al. (2008)[29]; Sekerke (2015)[30]; Zellner (1971)[34]). The hierarchical prior was suggested by Greyserman et al. (2006)[20]. They showed that a fully hierarchical model produced promising results warranting more study. The objective-based prior was suggested by Tu and Zhou (2010)[33]. Bodnar, Mazur, and Okhrin, (2017)[10] derived explicit formulas for the posterior distribution of linear combinations of the global minimum variance portfolio weights for the diffuse, conjugate, and hierarchical prior.

This thesis will be structured as follows. In chapter 2 we describe the basics of optimal portfolio selection, derive the expression for the minimum variance portfolio, and describe necessary theory regarding the efficient frontier. In chapter 3 we discuss the priors considered and derive ways to draw samples from the posterior predictive distribution. In chapter 4 the results of the simulation study as well as the empirical illustration are discussed.

Chapter 2

Optimal portfolio selection

2.1 Portfolio definition and weights

A portfolio is a collection of financial investments, which may contain a wide range of assets. Let S_i denote the price of the i-th security at time t, and m_i denote the number of shares in the i-th security at time t. Then, the value of this portfolio, consisting of k-securities, is the number of shares in each security multiplied by their price, giving the value of this portfolio at time t as

$$V = m_1 S_1 + m_2 S_2 + \cdots + m_k S_k.$$

The portfolio weight of a particular security is measured in the percentage that the security makes up, that is

$$w_i = \frac{m_i S_i}{V}, i = 1, 2, ..., k,$$

where w_i is the weight of the i-th security. The portfolio weights for the whole portfolio is calculated using equation (2.1) and (2.2),

$$w_1 + w_2 + \cdots + w_k = \frac{m_i S_i}{V} + \frac{m_2 S_2}{V} + \cdots + \frac{m_k S_k}{V} = \frac{V}{V} = 1,$$

which sum to one by definition. In matrix form, this can be written as

$$\boldsymbol{w}^T \mathbf{1} = 1,$$

where $\mathbf{1} = (1, 1, ..., 1)^T$ and $\boldsymbol{w} = (w_1, w_2, ..., w_k)^T$. Throughout the remainder of this thesis all vectors are column vectors by default.

2.2 Expected return

The simple return of security i at time t is denoted $x_{t,i}$. The return of a portfolio, with weights $\boldsymbol{w}$, is given by

$$x_{t,p} = \sum_{i=1}^{k} w_i \cdot x_{t,i}.$$

It follows from (2.5) by the additivity of mathematical expectation that the expected return of this portfolio is given by

$$\begin{aligned}\mu_p = E[x_{t,p}] = E[\sum_{i=1}^{k} w_i \cdot x_{t,i}] &= \sum_{i=1}^{k} w_i \cdot E[x_{t,i}] \\ &= \sum_{i=1}^{k} w_i \mu_i = \boldsymbol{\mu}^T \boldsymbol{w},\end{aligned}$$

where $\mu_i = E[x_{t,i}]$ is the expected return of security i that is assumed to be constant over time. The vector containing the expected returns of the k securities at time t is expressed as $\boldsymbol{\mu} = (\mu_1 \ \ \mu_2 \ldots \mu_k)^T$.

2.3 Risk

In the portfolio theory, the risk of a particular portfolio refers to the variance of the portfolio return, $x_{t,p}$. The variance of a portfolio, consisting of k securities, is given by

$$
\begin{aligned}
\sigma_p^2 = Var(x_{t,p}) &= Var\left(\sum_{i=1}^{k} w_i \cdot x_{t,i}\right) \\
&= Cov\left(\sum_{i=1}^{k} w_i \cdot x_{t,i}, \sum_{j=1}^{k} w_j \cdot x_{t,j}\right) \\
&= \sum_{i=1}^{k}\sum_{j=1}^{k} w_i w_j Cov\left(x_{t,i}, x_{t,j}\right) \\
&= \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w},
\end{aligned}
$$

where $\boldsymbol{\Sigma}$ is the covariance matrix with elements $\sigma_{ij} = Cov(x_{t,i}, x_{t,j})$. The matrix $\boldsymbol{\Sigma}$ has the structure

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

where the diagonal elements are the variances of the securities.

2.4 Global minimum variance portfolio

In optimal portfolio theory one of the goals is to minimize the risk of a portfolio. The global minimum variance portfolio is unique in that it estimates the lowest possible variance. The weights of the global minimum variance portfolio are derived by solving the following minimization problem,

$$
\underset{\boldsymbol{w}}{min}\, \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w} \;\text{ such that }\; \boldsymbol{w}^T \mathbf{1} = 1.
$$

The minimization problem can be solved using the method of Lagrange multipliers which is a strategy for finding the local maxima and minima of a function subject to equality constraints. The solution to the minimization problem is

$$\boldsymbol{w}_{GMV} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}.$$

The proof is left in appendix A.1 for the interested reader. The expected return of the global minimum variance portfolio is given by

$$\mu_{GMV} = \boldsymbol{\mu}^T\boldsymbol{w}_{GMV} = \frac{\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}.$$

The variance of the global minimum variance portfolio is given by

$$\begin{aligned}\sigma^2_{GMV} = \boldsymbol{w}^T_{GMV}\boldsymbol{\Sigma}\boldsymbol{w}_{GMV} &= \left(\frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}\right)^T \frac{\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}} \\ &= \frac{\mathbf{1}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}},\end{aligned}$$

where we use that $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1} = \boldsymbol{I}_k$, the k-dimensional identity matrix, by definition of the covariance matrix.

2.5 Efficient frontier

The efficient frontier is the set of optimal portfolios that offer the highest expected return, for a certain level of variance. This results in a parabola in the mean-variance space. We do consider short sales, which means that the weights may be smaller than 0 and larger than 1. We only consider portfolio allocations without a riskless security. This means that the efficient frontier is fully determined by maximizing the mean-variance utility, i.e.

$$\boldsymbol{w}^T\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{w}^T\boldsymbol{\Sigma}\boldsymbol{w} \text{ subject to } \boldsymbol{w}^T\mathbf{1} = 1, \tag{2.1}$$

where γ represents the risk aversion of the investor. By varying the risk aversion coefficient $\gamma > 0$, all portfolios on the efficient frontier can be obtained. The solution to (2.1) is given by

$$\boldsymbol{w}_{MV,\gamma} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}} + \gamma^{-1}\boldsymbol{R}\boldsymbol{\mu} \text{ with } \boldsymbol{R} = \boldsymbol{\Sigma}^{-1} - \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}\mathbf{1}^T\boldsymbol{\Sigma}^{-1}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}.$$

The expected return and variance of the MV portfolio is given by

$$R_{MV,\gamma} = \frac{\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}}{\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}} + \gamma^{-1} \boldsymbol{\mu}^T \boldsymbol{R} \boldsymbol{\mu} \quad \text{and} \quad V_{MV,\gamma} = \frac{1}{\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}} + \gamma^{-2} \boldsymbol{\mu}^T \boldsymbol{R} \boldsymbol{\mu}. \tag{2.2}$$

Note that (2.2) is a parametric equation in our case where we only consider portfolio allocations without a riskless asset. Solving (2.2) with respect to γ we obtain the parabola

$$(R - R_{GMV})^2 = \boldsymbol{\mu}^T \boldsymbol{R} \boldsymbol{\mu} (V - V_{GMV}),$$

where R_{GMV} is the expected return and V_{GMV} is the variance, of the global minimum variance portfolio. The global minimum variance portfolio can be obtained from the expected quadratic utility by letting γ tend to infinity.

2.6 Sample efficient frontier

In reality $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown and need to be estimated. The common approach, based on frequentist statistics, estimates parameters of the asset returns with their sample estimates. The sample estimates of the mean vector and the covariance matrix are given by

$$\hat{\boldsymbol{\mu}} = \overline{\boldsymbol{x}}_{(t-1)} \quad \text{and} \quad \hat{\boldsymbol{\Sigma}} = d_n \boldsymbol{S}_{(t-1)}$$

where

$$\begin{aligned}
\overline{\boldsymbol{x}}_{(t-1)} &= \frac{1}{n} \sum_{i=t-n}^{t-1} \boldsymbol{x}_i, \\
d_n &= \frac{1}{n-1} \quad \text{and} \\
\boldsymbol{S}_{(t-1)} &= \sum_{i=t-n}^{t-1} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{(t-1)})(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{(t-1)})^T.
\end{aligned}$$

Using the sample estimates we obtain the sample portfolio weights

$$\boldsymbol{w}_{S,\gamma} = \frac{\boldsymbol{S}_{(t-1)}^{-1}\mathbf{1}}{\mathbf{1}^T\boldsymbol{S}_{(t-1)}^{-1}\mathbf{1}} + \gamma^{-1}\boldsymbol{Q}_{(t-1)}\overline{\boldsymbol{x}}_{(t-1)} \quad \text{with} \quad \boldsymbol{Q}_{(t-1)} = \boldsymbol{S}_{(t-1)}^{-1} - \frac{\boldsymbol{S}_{(t-1)}^{-1}\mathbf{1}\mathbf{1}^T\boldsymbol{S}_{(t-1)}^{-1}}{\mathbf{1}^T\boldsymbol{S}_{(t-1)}^{-1}\mathbf{1}},$$

with the sample expected return and variance given by

$$R_{S,\gamma} = \frac{\mathbf{1}^T\boldsymbol{S}_{(t-1)}^{-1}\overline{\boldsymbol{x}}_{(t-1)}}{\mathbf{1}^T\boldsymbol{S}_{(t-1)}^{-1}\mathbf{1}} + \gamma^{-1}d_n^{-1}\overline{\boldsymbol{x}}_{(t-1)}^T\boldsymbol{Q}_{(t-1)}\overline{\boldsymbol{x}}_{(t-1)}$$

and

$$V_{S,\gamma} = \frac{d_n}{\mathbf{1}^T\boldsymbol{S}_{(t-1)}^{-1}\mathbf{1}} + \gamma^{-2}d_n^{-1}\overline{\boldsymbol{x}}_{(t-1)}^T\boldsymbol{Q}_{(t-1)}\overline{\boldsymbol{x}}_{(t-1)}.$$

The sample efficient frontier, which was derived by Bodnar and Schmid (2008)[12], (2009)[13], Kan and Smith (2008)[24], is expressed as

$$(R - R_{GMV,S})^2 = \frac{\overline{\boldsymbol{x}}_{(t-1)}^T\boldsymbol{Q}_{(t-1)}\overline{\boldsymbol{x}}_{(t-1)}}{d_n}(V - V_{GMV,S}),$$

where

$$R_{GMV,S} = \frac{\mathbf{1}^T\boldsymbol{S}_{(t-1)}^{-1}\overline{\boldsymbol{x}}_{(t-1)}}{\mathbf{1}^T\boldsymbol{S}_{(t-1)}^{-1}\mathbf{1}} \quad \text{and} \quad V_{GMV,S} = \frac{d_n}{\mathbf{1}^T\boldsymbol{S}_{(t-1)}^{-1}\mathbf{1}}$$

are the sample expected return and variance of the global minimum variance portfolio.

Chapter 3

Bayesian portfolio selection

When we consider the problem of optimal portfolio selection from the Bayesian perspective we are interested in the posterior predictive distribution. We consider the parameters of the asset returns to be random variables. The goal is to assign probabilities to the possible values of the parameters. This is done by deriving the posterior distribution of the parameters. In order to derive the posterior distribution we consider Bayes' Theorem:

$$P(\boldsymbol{\theta}|\boldsymbol{x}_{(t-1)}) = \frac{P(\boldsymbol{x}_{(t-1)}|\boldsymbol{\theta}) \cdot P(\boldsymbol{\theta})}{P(\boldsymbol{x}_{(t-1)})} \tag{3.1}$$

where $P(\boldsymbol{\theta}|\boldsymbol{x}_{(t-1)})$ is the posterior distribution of $\boldsymbol{\theta}$ given the observed data $\boldsymbol{x}_{(t-1)}$. The likelihood function $P(\boldsymbol{x}_{(t-1)}|\boldsymbol{\theta}) = L(\boldsymbol{\theta}|\boldsymbol{x}_{(t-1)})$ is the measure of support provided by data for each possible value of the parameter $\boldsymbol{\theta}$. The prior distribution $P(\boldsymbol{\theta})$ represents the prior knowledge of the parameter $\boldsymbol{\theta}$, this reflects our subjective beliefs about the parameter $\boldsymbol{\theta}$ before any data are observed. Lastly, $P(\boldsymbol{x}_{(t-1)})$ is the marginal probability of the observed data $\boldsymbol{x}_{(t-1)}$. We can rewrite equation (3.1) as

$$P(\boldsymbol{\theta}|\boldsymbol{x}_{(t-1)}) = \frac{L(\boldsymbol{\theta}|\boldsymbol{x}_{(t-1)}) \cdot P(\boldsymbol{\theta})}{\int L(\boldsymbol{\theta}|\boldsymbol{x}_{(t-1)}) \cdot P(\boldsymbol{\theta}) d\boldsymbol{\theta}} \propto L(\boldsymbol{\theta}|\boldsymbol{x}_{(t-1)}) \cdot P(\boldsymbol{\theta}). \tag{3.2}$$

Since we integrate over $\boldsymbol{\theta}$ in the denominator, this expression is proportional to the posterior distribution. From the posterior distribution (3.2) we derive the posterior predictive distribution as follows:

$$P(\boldsymbol{x}_{(t)}|\boldsymbol{x}_{(t-1)}) = \int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} P(\boldsymbol{x}_{(t)}|\boldsymbol{\theta})P(\boldsymbol{\theta}|\boldsymbol{x}_{(t-1)})d\boldsymbol{\theta}. \tag{3.3}$$

Throughout this thesis we assume that the vectors of asset returns are independent and identically distributed conditional on the mean vector and the covariance matrix. In this chapter we provide details on the priors for the parameters of the asset returns and ways to draw samples from the posterior predictive distribution for all considered priors.

3.1 Considered priors

3.1.1 Diffuse prior

The first prior we consider is the diffuse prior, also known as the non-informative Jeffreys prior, on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The diffuse prior has been applied in theory by Barry (1974)[3]; Brown (1976)[16]; and Klein and Bawa (1976)[25]. This prior infers no initial information regarding the distribution of the characteristics of the assets. It is given by

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

When measuring the performance of methods based on the diffuse prior relative to methods based on frequentist statistics, the Bayesian approach tends to perform equal to or better than the frequentist approach. Stambaugh (1997)[32] showed that when the assets have varying histories the Bayesian approach may use this information which leads to different results.

3.1.2 Conjugate prior

The second prior we consider is the conjugate prior. This prior is based on the (extended) Black-Litterman model (cf. Black and Litterman (1992)[7]). In order to incorporate expert knowledge, Black and Litterman (1992)[7] suggested to employ the normal prior for the vector of expected returns $\boldsymbol{\mu}$. This approach is known in financial literature as the Black-Litterman model. We consider an extension of this model by also including a prior on the covariance $\boldsymbol{\Sigma}$. This leads to an informative prior which considers a normal distributed prior for $\boldsymbol{\mu}$ and an inverse Wishart distributed prior for the covariance matrix $\boldsymbol{\Sigma}$.

Formally, the conjugate prior can be expressed as

$$p_c(\boldsymbol{\mu}|\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-1/2} exp\left\{-\frac{m_c}{2}(\boldsymbol{\mu}-\boldsymbol{r}_c)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{r}_c)\right\},$$

and

$$p_c(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-d_c/2} exp\left\{-\frac{1}{2}tr\left[\boldsymbol{S}_c\boldsymbol{\Sigma}^{-1}\right]\right\},$$

where the additional model parameters $\boldsymbol{r}_c, m_c, d_c, \boldsymbol{S}_c$ are known as hyperparameters. The normal prior for $\boldsymbol{\mu}$ is k-dimensional normal distribution with mean vector $\boldsymbol{r}_c$ and covariance matrix $\boldsymbol{\Sigma}/m_c$, while the inverse Wishart prior for $\boldsymbol{\Sigma}$ is parametrised with d_c degrees of freedom and parameter matrix $\boldsymbol{S}_c$. The prior mean vector $\boldsymbol{r}_c$ reflects the prior beliefs about $\boldsymbol{\mu}$ and the matrix $\boldsymbol{S}_c$ reflects the prior beliefs about the covariance $\boldsymbol{\Sigma}$. The other hyperparameters m_c and d_c are known as precision parameters for $\boldsymbol{r}_c$ and $\boldsymbol{S}_c$, respectively. The joint prior for both parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is given by

$$\begin{aligned} p_c(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(d_c+1)/2} exp\bigg\{ &-\frac{m_c}{2}(\boldsymbol{\mu}-\boldsymbol{r}_c)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{r}_c) \\ &-\frac{1}{2}tr\left[\boldsymbol{S}_c\boldsymbol{\Sigma}^{-1}\right]\bigg\}. \end{aligned}$$

The conjugate prior has been researched extensively, most notably by Frost and Savarino (1986)[17], Rachev et al. (2009)[29], Avramov and Zhou (2010)[2], Bodnar et al. (2017)[10]. An interesting application was proposed by Frost and Savarino (1986)[17] who considered identical expected return, variance and covariance for all assets. This method showed promising results.

3.1.3 Hierarchical prior

The third prior we consider is the hierarchical prior which was suggested by Greyserman et al. (2006)[20]. The hierarchical prior is given by

$$
\begin{aligned}
p_h(\boldsymbol{\mu}|\xi, \eta, \boldsymbol{\Sigma}) &\propto |\boldsymbol{\Sigma}|^{-1/2} exp\left\{-\frac{\kappa_h}{2}(\boldsymbol{\mu} - \xi\mathbf{1})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \xi\mathbf{1})\right\}, \\
p_h(\boldsymbol{\Sigma}) &\propto \frac{\eta^{-k(d_h-k-1)/2}}{|\boldsymbol{\Sigma}|^{d_h/2}} exp\left\{-\frac{1}{2\eta} tr\left[\boldsymbol{S}_h\boldsymbol{\Sigma}^{-1}\right]\right\}, \\
p_h(\xi) &\propto 1, \\
p_h(\eta) &\propto \eta^{-(\varepsilon_1+1)} exp\left\{-\frac{\varepsilon_2}{\eta}\right\},
\end{aligned}
$$

where κ_h is a prior precision parameter on $\boldsymbol{\mu}$; d_h is a similar prior precision parameter on $\boldsymbol{\Sigma}$; $\boldsymbol{S}_h$ is a known prior matrix of $\boldsymbol{\Sigma}$; ε_1 and ε_2 are prior constants. The joint prior of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, ξ, and η is expressed as

$$
\begin{aligned}
p_h(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi, \eta) \propto\ & |\boldsymbol{\Sigma}|^{-1/2} exp\left\{-\frac{\kappa_h}{2}(\boldsymbol{\mu} - \xi\mathbf{1})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \xi\mathbf{1})\right\} \\
& \times \frac{\eta^{-k(d_h-k-1)/2}}{|\boldsymbol{\Sigma}|^{d_h/2}} exp\left\{-\frac{1}{2\eta} tr\left[\boldsymbol{S}_h\boldsymbol{\Sigma}^{-1}\right]\right\} \\
& \times \eta^{-(\varepsilon_1+1)} exp\left\{-\frac{\varepsilon_2}{\eta}\right\} \\
\propto\ & |\boldsymbol{\Sigma}|^{-(d_h+1)/2} exp\left\{-\frac{1}{2\eta} tr\left[\boldsymbol{S}_h\boldsymbol{\Sigma}^{-1}\right]\right\} \\
& \times \eta^{-k(d_h-k-1)/2-\varepsilon_1-1} exp\bigg\{ - \frac{\kappa_h}{2}(\boldsymbol{\mu} \\
& - \xi\mathbf{1})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \xi\mathbf{1}) - \frac{\varepsilon_2}{\eta}\bigg\}.
\end{aligned}
$$

Most notably the hierarchical prior, in contrast to the conjugate and objective-based prior, shrinks all elements of the mean vector by an equal amount.

3.1.4 Objective-based prior

The fourth prior we consider is the objective-based prior suggested by Tu and Zhou (2010)[33]. The objective-based prior is given by

$$p_{ob}(\boldsymbol{\mu}|\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-1/2} exp\left\{-\frac{s^2}{2\sigma_{ob}^2}(\boldsymbol{\mu} - \gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob})\right\}$$
$$p_{ob}(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-d_{ob}/2} exp\left\{-\frac{1}{2}tr\left[\boldsymbol{S}_{ob}\boldsymbol{\Sigma}^{-1}\right]\right\},$$

where $s^2 = \frac{1}{k}tr(\boldsymbol{\Sigma})$ is the average of the diagonal elements of $\boldsymbol{\Sigma}$; σ_{ob}^2 indicates the uncertainty about $\boldsymbol{\mu}$; γ represents the investor's risk aversion; $\boldsymbol{w}_{ob}$, d_{ob}, and $\boldsymbol{S}_{ob}$ are prior constants. The joint prior distribution of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is given by

$$p_{ob}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(d_{ob}+1)/2} exp\left\{-\frac{1}{2}tr\left[\boldsymbol{S}_{ob}\boldsymbol{\Sigma}^{-1}\right]\right\}$$
$$\times\, exp\left\{-\frac{s^2}{2\sigma_{ob}^2}(\boldsymbol{\mu} - \gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob})\right\}.$$

3.2 Stochastic representation and efficient frontier

3.2.1 Diffuse prior

The following theorem gives us a way to draw samples from the posterior predictive distribution of the diffuse prior. The theorem was derived and proved by Bauder et al. (2020)[5].

Theorem 1. Let $\boldsymbol{X_1}, \boldsymbol{X_2}, ...$ be infinitely exchangeable and multivariate centred spherically symmetric. Let $P(\boldsymbol{\theta}) = |\boldsymbol{F}|^{1/2}$ be Jeffreys' prior where $|\boldsymbol{A}|$ denotes the determinant of a squared matrix $\boldsymbol{A}$ and $\boldsymbol{F} = -E(\partial^2 log(f(\boldsymbol{x}_{(t-1)}| \boldsymbol{\theta}))/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T)$ is the Fisher information matrix. Assume $n > k$. Then the stochastic representation of the random variable $\hat{X}_{p,t}$ whose density is the posterior predictive distribution (3.3) is given by

$$\hat{X}_{p,t} \stackrel{\mathrm{d}}{=} \boldsymbol{w}^T \overline{\boldsymbol{x}}_{(t-1)} + \sqrt{\boldsymbol{w}^T \boldsymbol{S}_{(t-1)} \boldsymbol{w}} \times \left(\frac{\tau_1}{\sqrt{n(n-k)}} + \sqrt{1 + \frac{\tau_1^2}{n-k}} \frac{\tau_2}{\sqrt{n-k+1}} \right),$$

where

$$\overline{\boldsymbol{x}}_{(t-1)} = \frac{1}{n} \sum_{i=t-n}^{t-1} \boldsymbol{x}_i \text{ and}$$
$$\boldsymbol{S}_{(t-1)} = \sum_{i=t-n}^{t-1} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{(t-1)})(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{(t-1)})^T,$$

and τ_1 and τ_2 are independent with $\tau_1 \sim t_{n-k}$ and $\tau_2 \sim t_{n-k+1}$.

From this expression we can generate the sample $\hat{X}_{p,t}^{(1)}, \ldots, \hat{X}_{p,t}^{(B)}$ which can be used to calculate important characteristics of the posterior predicitve distribution. *Theorem* 1 also provides us with an analytical expression of the expected value and variance of the posterior predictive distribution, this is formulated in the following corollary:

Corollary 1. Under the conditions of Theorem 1, let $n - k > 2$. Then:

$$E(\boldsymbol{w}^T \boldsymbol{X}_t | \boldsymbol{x}_{(t-1)}) = \boldsymbol{w}^T \overline{\boldsymbol{x}}_{(t-1)}$$

and

$$Var(\boldsymbol{w}^T \boldsymbol{X}_t | \boldsymbol{x}_{(t-1)}) = c_{k,n} \boldsymbol{w}^T \boldsymbol{S}_{(t-1)} \boldsymbol{w} \text{ with}$$
$$c_{k,n} = \frac{1}{n-k-1} + \frac{2n-k-1}{n(n-k-1)(n-k-2)}.$$

The proof to *Corollary* 1 is given by Bauder et al. (2020)[5]. Based on the results of *Corollary* 1 we can construct an optimal portfolio by maximizing the mean-variance utility function given by equation (2.1):

$$
\begin{aligned}
U(\boldsymbol{w}) &= E(\boldsymbol{w}^T \boldsymbol{X}_t | \boldsymbol{x}_{(t-1)}) - \frac{\gamma}{2} Var(\boldsymbol{w}^T \boldsymbol{X}_t | \boldsymbol{x}_{(t-1)}) \\
&= \boldsymbol{w}^T \overline{\boldsymbol{x}}_{(t-1)} - \frac{c_{k,n}\gamma}{2} \boldsymbol{w}^T \boldsymbol{S}_{(t-1)} \boldsymbol{w}.
\end{aligned} \tag{3.4}
$$

The quantity $\gamma > 0$ represents the investor's risk aversion. Maximizing the mean-variance utility closely resembles the optimization problem studied by Ingersoll (1987)[22] and Okhrin and Schmid (2006)[28]). The only difference is that the risk aversion coefficient γ is multiplied with the constant $c_{k,n}$. The solution to the optimization problem (3.4) is given by

$$
\begin{aligned}
\boldsymbol{w}_{MV,\gamma} &= \frac{\boldsymbol{S}_{(t-1)}^{-1} \mathbf{1}}{\mathbf{1}^T \boldsymbol{S}_{(t-1)}^{-1} \mathbf{1}} + \gamma^{-1} c_{k,n}^{-1} \boldsymbol{Q}_{(t-1)} \overline{\boldsymbol{x}}_{(t-1)} \quad \text{with} \\
\boldsymbol{Q}_{(t-1)} &= \boldsymbol{S}_{(t-1)}^{-1} - \frac{\boldsymbol{S}_{(t-1)}^{-1} \mathbf{1} \mathbf{1}^T \boldsymbol{S}_{(t-1)}^{-1}}{\mathbf{1}^T \boldsymbol{S}_{(t-1)}^{-1} \mathbf{1}}
\end{aligned}
$$

together with the expected return and the variance expressed as

$$
R_{MV,\gamma} = \frac{\mathbf{1}^T \boldsymbol{S}_{(t-1)}^{-1} \overline{\boldsymbol{x}}_{(t-1)}}{\mathbf{1}^T \boldsymbol{S}_{(t-1)}^{-1} \mathbf{1}} + \gamma^{-1} c_{k,n}^{-1} \overline{\boldsymbol{x}}_{(t-1)}^T \boldsymbol{Q}_{(t-1)} \overline{\boldsymbol{x}}_{(t-1)} \tag{3.5}
$$

and

$$
V_{MV,\gamma} = \frac{c_{k,n}}{\mathbf{1}^T \boldsymbol{S}_{(t-1)}^{-1} \mathbf{1}} + \gamma^{-2} c_{k,n}^{-1} \overline{\boldsymbol{x}}_{(t-1)}^T \boldsymbol{Q}_{(t-1)} \overline{\boldsymbol{x}}_{(t-1)}, \tag{3.6}
$$

respectively, where we use that $\boldsymbol{Q}_{(t-1)} \mathbf{1} = 0$ and $\boldsymbol{Q}_{(t-1)} \boldsymbol{S}_{(t-1)} \boldsymbol{Q}_{(t-1)} = \boldsymbol{Q}_{(t-1)}$ in (3.6). Equation (3.5) and (3.6) determines the set of all optimal portfolios obtained as solutions to the optimization problem (3.4). Solving these equations with respect to γ leads to a set of optimal portfolios in the mean-variance space, called the efficient frontier. This set of optimal portfolios for the diffuse prior is given by

$$
(R - R_{GMV})^2 = \frac{\overline{\boldsymbol{x}}_{(t-1)}^T \boldsymbol{Q}_{(t-1)} \overline{\boldsymbol{x}}_{(t-1)}}{c_{k,n}} (V - V_{GMV}), \tag{3.7}
$$

where

$$R_{GMV} = \frac{\mathbf{1}^T \boldsymbol{S}_{(t-1)}^{-1} \overline{\boldsymbol{x}}_{(t-1)}}{\mathbf{1}^T \boldsymbol{S}_{(t-1)}^{-1} \mathbf{1}} \quad \text{and} \quad V_{GMV} = \frac{c_{k,n}}{\mathbf{1}^T \boldsymbol{S}_{(t-1)}^{-1} \mathbf{1}}$$

are the expected return and variance of the global minimum variance portfolio. The slope parameter of the efficient frontier is given by

$$s = \frac{\overline{\boldsymbol{x}}_{(t-1)}^T \boldsymbol{Q}_{(t-1)} \overline{\boldsymbol{x}}_{(t-1)}}{c_{k,n}}.$$

Equation (3.7) specifies a parabola in the mean-variance space.

3.2.2 Conjugate prior

The following theorem gives us a way, similarly to that of *Theorem* 1, to draw samples from the posterior predictive distribution of the conjugate prior. This theorem was also derived and proved by Bauder et al. (2020)[5].

Theorem 2. Let $\boldsymbol{X_1}, \boldsymbol{X_2}, ...$ be infinitely exchangeable and multivariate centred spherically symmetric. Assume $n + d_c - 2k > 0$. Then, under the application of the conjugate prior, the stochastic representation of the random variable $\hat{X}_{p,t}$ whose density is the posterior predictive distribution (3.3) is given by

$$\begin{aligned} \hat{X}_{p,t} \stackrel{\mathrm{d}}{=} & \boldsymbol{w}^T \overline{\boldsymbol{x}}_{(t-1),c} + \sqrt{\boldsymbol{w}^T \boldsymbol{S}_{(t-1),c} \boldsymbol{w}} \Bigg(\frac{\eta_1}{\sqrt{(n+m_c)(n+d_c-2k)}} \\ & + \sqrt{1 + \frac{\eta_1^2}{n+d_c-2k}} \frac{\eta_2}{\sqrt{n+d_c-2k+1}} \Bigg), \end{aligned}$$

where

$$\begin{aligned} \overline{\boldsymbol{x}}_{(t-1),c} &= \frac{n\overline{\boldsymbol{x}}_{(t-1)} + m_c \boldsymbol{r}_c}{n+m_c} \quad \text{and} \\ \boldsymbol{S}_{(t-1),c} &= \boldsymbol{S}_{(t-1)} + \boldsymbol{S}_c + nm_c \frac{(\boldsymbol{r}_c - \overline{\boldsymbol{x}}_{(t-1),c})(\boldsymbol{r}_c - \overline{\boldsymbol{x}}_{(t-1),c})^T}{n+m_c}, \end{aligned} \tag{3.8}$$

and η_1 and η_2 are independent with $\eta_1 \sim t_{n+d_c-2k}$ and $\eta_2 \sim t_{n+d_c-2k+1}$.

Theorem 2 also provides us with an analytical expression of the expected value and variance of the posterior predictive distribution, this is formulated in the following corollary:

Corollary 2. Under the conditions of Theorem 2, let $n + d_c - 2k > 2$. Then:

$$E(\boldsymbol{w}^T \boldsymbol{X}_t | \boldsymbol{x}_{(t-1)}) = \boldsymbol{w}^T \overline{\boldsymbol{x}}_{(t-1),c} \tag{3.9}$$

and

$$Var(\boldsymbol{w}^T \boldsymbol{X}_t | \boldsymbol{x}_{(t-1)}) = q_{k,n} \boldsymbol{w}^T \boldsymbol{S}_{(t-1),c} \boldsymbol{w} \tag{3.10}$$

with

$$\begin{aligned} q_{k,n} = & \frac{1}{n + d_c - 2k - 1} \\ & + \frac{2n + m_c + d_c - 2k - 1}{(n + m_c)(n + d_c - 2k - 1)(n + d_c - 2k - 2)}. \end{aligned}$$

The proof to *Corollary* 2 is given by Bauder et al. (2020)[5]. Substituting equation (3.9) and (3.10) in (3.4) we solve the optimization problem for the conjugate prior. The solution to the optimization problem is given by

$$\begin{aligned} \boldsymbol{w}_{MV,\gamma} &= \frac{\boldsymbol{S}^{-1}_{(t-1),c} \mathbf{1}}{\mathbf{1}^T \boldsymbol{S}^{-1}_{(t-1),c} \mathbf{1}} + \gamma^{-1} q^{-1}_{k,n} \boldsymbol{Q}_{(t-1),c} \overline{\boldsymbol{x}}_{(t-1),c} \quad \text{with} \\ \boldsymbol{Q}_{(t-1),c} &= \boldsymbol{S}^{-1}_{(t-1),c} - \frac{\boldsymbol{S}^{-1}_{(t-1),c} \mathbf{1} \mathbf{1}^T \boldsymbol{S}^{-1}_{(t-1),c}}{\mathbf{1}^T \boldsymbol{S}^{-1}_{(t-1),c} \mathbf{1}} \end{aligned}$$

together with the expected return and the variance expressed as

$$R_{MV,\gamma} = \frac{\mathbf{1}^T \boldsymbol{S}^{-1}_{(t-1),c} \overline{\boldsymbol{x}}_{(t-1),c}}{\mathbf{1}^T \boldsymbol{S}^{-1}_{(t-1),c} \mathbf{1}} + \gamma^{-1} q^{-1}_{k,n} \overline{\boldsymbol{x}}^T_{(t-1),c} \boldsymbol{Q}_{(t-1),c} \overline{\boldsymbol{x}}_{(t-1),c} \tag{3.11}$$

and

$$V_{MV,\gamma} = \frac{q_{k,n}}{\mathbf{1}^T \boldsymbol{S}^{-1}_{(t-1),c} \mathbf{1}} + \gamma^{-2} q^{-1}_{k,n} \overline{\boldsymbol{x}}^T_{(t-1),c} \boldsymbol{Q}_{(t-1),c} \overline{\boldsymbol{x}}_{(t-1),c}. \tag{3.12}$$

Although the solution looks similar to that of the diffuse prior they are in fact very different due to the definition of $\overline{\boldsymbol{x}}_{(t-1),c}$ and $\boldsymbol{S}_{(t-1),c}$ in (3.8). Solving equation (3.11) and (3.12) with respect to γ leads to the set of optimal portfolios, in the mean-variance space, called the efficient frontier. This set of optimal portfolios, for the conjugate prior, is given by

$$(R - R_{GMV})^2 = \frac{\overline{\boldsymbol{x}}^T_{(t-1),c}\boldsymbol{Q}_{(t-1),c}\overline{\boldsymbol{x}}_{(t-1),c}}{q_{k,n}}(V - V_{GMV}), \tag{3.13}$$

where

$$R_{GMV} = \frac{\mathbf{1}^T\boldsymbol{S}^{-1}_{(t-1),c}\overline{\boldsymbol{x}}_{(t-1),c}}{\mathbf{1}^T\boldsymbol{S}^{-1}_{(t-1),c}\mathbf{1}} \quad \text{and} \quad V_{GMV} = \frac{q_{k,n}}{\mathbf{1}^T\boldsymbol{S}^{-1}_{(t-1),c}\mathbf{1}}$$

are the expected return and variance of the global minimum variance portfolio. The slope parameter of the efficient frontier is given by

$$s = \frac{\overline{\boldsymbol{x}}^T_{(t-1),c}\boldsymbol{Q}_{(t-1),c}\overline{\boldsymbol{x}}_{(t-1),c}}{q_{k,n}}.$$

Equation (3.13) specifies a parabola in the mean-variance space.

3.3 Conditional posterior distribution

3.3.1 Hierarchical prior

In order to draw inference from the posterior predictive distribution of the hierarchical prior, we derive the conditional posterior distributions for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The posterior distribution of the hierarchical prior is expressed as

$$\begin{aligned}
p_h(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi, \eta | \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \propto\ & L(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n | \boldsymbol{\mu}, \boldsymbol{\Sigma}) p_h(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi, \eta) \\
& \propto |\boldsymbol{\Sigma}|^{-(d_h+n+1)/2} exp\left\{-\frac{1}{2} tr\left[(\eta^{-1}\boldsymbol{S}_h + (n-1)\boldsymbol{S})\boldsymbol{\Sigma}^{-1}\right]\right\} \\
& \times \eta^{-k(d_h-k-1)/(2-\varepsilon_1-1)} exp\left\{-\frac{\varepsilon_2}{\eta}\right\} \\
& \times exp\bigg\{ -\frac{\kappa_h}{2}(\boldsymbol{\mu} - \xi\mathbf{1})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \xi\mathbf{1}) \\
& - \frac{n}{2}(\boldsymbol{\mu} - \overline{\boldsymbol{X}})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \overline{\boldsymbol{X}})\bigg\},
\end{aligned} \tag{3.14}$$

where

$$\begin{aligned}
L(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto\ & |\boldsymbol{\Sigma}|^{-n/2} \\
\times exp & \left\{-\frac{n}{2}(\boldsymbol{\mu} - \overline{\boldsymbol{X}})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \overline{\boldsymbol{X}}) - \frac{n-1}{2} tr\left[\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\right]\right\}.
\end{aligned}$$

From equation (3.14) we derive the conditional posterior distribution for $\boldsymbol{\mu}$ given $\boldsymbol{\Sigma}$, ξ and η, and the conditional posterior distribution for $\boldsymbol{\Sigma}$ given ξ and η (see appendix A.2.1). The conditional posterior distribution of $\boldsymbol{\mu}$ is expressed as

$$\begin{aligned}
p_h(\boldsymbol{\mu} | \boldsymbol{\Sigma}, \xi, \eta, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \propto\ & |\boldsymbol{\Sigma}|^{-1/2} \\
& \times exp\bigg\{ -\frac{\kappa_h + n}{2} (\boldsymbol{\mu} - \boldsymbol{r}_h)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \boldsymbol{r}_h) \bigg\},
\end{aligned} \tag{3.15}$$

where

$$\boldsymbol{r}_h = \frac{\kappa_h \xi \mathbf{1} + n\overline{\boldsymbol{X}}}{\kappa_h + n}.$$

This is the kernel of a multivariate normal distribution, that is

$$\boldsymbol{\mu}|\boldsymbol{\Sigma},\xi,\eta,\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n \sim N(\boldsymbol{r}_h,\frac{1}{\kappa_h+n}\boldsymbol{\Sigma}). \tag{3.16}$$

Integrating out $\boldsymbol{\mu}$ from (3.14) we get conditional posterior distribution of $\boldsymbol{\Sigma}$ expressed as

$$p_h(\boldsymbol{\Sigma}|\xi,\eta,\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n) \propto |\boldsymbol{\Sigma}|^{-(d_h+n)/2} exp\Bigg\{-\frac{1}{2}tr\Bigg[\Bigg(\eta^{-1}\boldsymbol{S}_h+(n-1)\boldsymbol{S}$$
$$+\frac{\kappa_n n}{\kappa_n+n}(\overline{\boldsymbol{X}}-\xi\mathbf{1})(\overline{\boldsymbol{X}}-\xi\mathbf{1})^T\Bigg)\boldsymbol{\Sigma}^{-1}\Bigg]\Bigg\}, \tag{3.17}$$

which is the kernel of a inverse-Wishart distribution (see Gupta and Nagar (2000)[21]), that is

$$\boldsymbol{\Sigma}|\xi,\eta,\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n \sim IW_k(d_h+n,\boldsymbol{\Psi}_h) \tag{3.18}$$

where

$$\boldsymbol{\Psi}_h=\eta^{-1}\boldsymbol{S}_h+(n-1)\boldsymbol{S}+\frac{\kappa_n n}{\kappa_n+n}(\overline{\boldsymbol{X}}-\xi\mathbf{1})(\overline{\boldsymbol{X}}-\xi\mathbf{1})^T.$$

The marginal posterior distributions of ξ is given by

$$p_h(\xi|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n)\propto 1,$$

which we choose to be uniformly distributed on the interval $(-0.01, 0.01)$. Finally, the marginal posterior distribution of η is given by

$$p_h(\eta|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n)\propto \eta^{-k(d_h-k-1)/2-\varepsilon_1-1} exp\left\{-\frac{\varepsilon_2}{\eta}\right\}.$$

which is the kernel of a inverse-Gamma distribution, that is

$$\eta \sim Inverse-Gamma(\varepsilon_1,\varepsilon_2).$$

The following process, were we assume the asset returns to follow a normal distribution, allows us to draw samples from the posterior predictive distribution:

1. Draw $\tilde{\boldsymbol{\mu}}_h$ and $\tilde{\boldsymbol{\Sigma}}_h$ from the conditional posterior distributions (3.16) and (3.18), respectively.
2. Then, draw $\boldsymbol{Y}_{h,t}$ from $N_k(\tilde{\boldsymbol{\mu}}_h, \tilde{\boldsymbol{\Sigma}}_h)$.

We repeat these two steps N_h times to obtain the sample $\boldsymbol{Y}_{h,t}^{(1)}, \ldots, \boldsymbol{Y}_{h,t}^{(N_h)}$, which is a sample from the posterior predictive distribution. The sample $\boldsymbol{Y}_{h,t}^{(1)}, \ldots, \boldsymbol{Y}_{h,t}^{(N_h)}$ allows us to estimate the efficient frontier, see section 2.6. Sample efficient frontier.

3.3.2 Objective-based prior

Similarly to the derivations in the case of the hierarchical prior we derive the conditional posterior distributions for $\boldsymbol{\mu}$ and the marginal posterior distribution for $\boldsymbol{\Sigma}$ in order to draw inference from the posterior predictive distribution of the objective-based prior. The posterior distribution of the objective-based prior is expressed as

$$\begin{aligned} p_{ob}(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) &\propto L(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n|\boldsymbol{\mu}, \boldsymbol{\Sigma}) p_{ob}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &\propto |\boldsymbol{\Sigma}|^{-(d_{ob}+n+1)/2} exp\left\{-\frac{1}{2} tr\left[(\boldsymbol{S}_{ob} + (n-1)\boldsymbol{S})\boldsymbol{\Sigma}^{-1}\right]\right\} \\ &\times exp\bigg\{-\frac{s^2}{2\sigma_{ob}^2}(\boldsymbol{\mu} - \gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}) \\ &- \frac{n}{2}(\boldsymbol{\mu} - \overline{\boldsymbol{X}})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \overline{\boldsymbol{X}})\bigg\}. \end{aligned} \tag{3.19}$$

From equation (3.19) we derive the conditional posterior distribution for $\boldsymbol{\mu}$ and marginal posterior for $\boldsymbol{\Sigma}$ (see appendix A.2.2). The conditional posterior distribution of $\boldsymbol{\mu}$ is expressed as

$$p_{ob}(\boldsymbol{\mu}|\boldsymbol{\Sigma}, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \propto |\boldsymbol{\Sigma}|^{-1/2} exp\bigg\{-\frac{m_{ob}}{2}\left(\boldsymbol{\mu} - \boldsymbol{r}_{ob}\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} - \boldsymbol{r}_{ob}\right)\bigg\} \tag{3.20}$$

where

$$
\begin{aligned}
m_{ob} &= \frac{s^2}{\sigma_{ob}^2} + n \\
\boldsymbol{r}_{ob} &= \frac{\frac{s^2}{\sigma_{ob}^2}\gamma \boldsymbol{\Sigma} \boldsymbol{w}_{ob} + n\overline{\boldsymbol{X}}}{\frac{s^2}{\sigma_{ob}^2} + n}.
\end{aligned}
$$

This is the kernel of a multivariate normal distribution, that is

$$
\boldsymbol{\mu} | \boldsymbol{\Sigma}, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \sim N(\boldsymbol{r}_{ob}, \frac{1}{m_{ob}} \boldsymbol{\Sigma}). \tag{3.21}
$$

Integrating out $\boldsymbol{\mu}$ from (3.19) we get marginal posterior distribution of $\boldsymbol{\Sigma}$ expressed as

$$
\begin{aligned}
p_{ob}(\boldsymbol{\Sigma} | \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) &\propto |\boldsymbol{\Sigma}|^{-(d_{ob}+n)/2} exp \left\{ -\frac{1}{2} tr \left[(\boldsymbol{S}_{ob} + (n-1)\boldsymbol{S}) \boldsymbol{\Sigma}^{-1} \right] \right\} \\
&\times exp \Bigg\{ -\frac{1}{2} \left(\frac{n \cdot \frac{s^2}{\sigma_{ob}^2}}{m_{ob}} \left(\overline{\boldsymbol{X}} - \gamma \boldsymbol{\Sigma} \boldsymbol{w}_{ob} \right)^T \boldsymbol{\Sigma}^{-1} \left(\overline{\boldsymbol{X}} - \gamma \boldsymbol{\Sigma} \boldsymbol{w}_{ob} \right) \right) \Bigg\}.
\end{aligned} \tag{3.22}
$$

The following process, were we assume the asset returns to follow a normal distribution, allows us to draw samples from the posterior predictive distribution:

1. Draw $\tilde{\boldsymbol{\mu}}_{ob}$ from the conditional posterior distributions (3.21) and $\tilde{\boldsymbol{\Sigma}}_{ob}$ by implementing the rejection sampling algorithm (see appendix A.3).
2. Then, draw $\boldsymbol{Y}_{ob,t}$ from $N_k(\tilde{\boldsymbol{\mu}}_{ob}, \tilde{\boldsymbol{\Sigma}}_{ob})$.

We repeat these two steps N_{ob} times to obtain the sample $\boldsymbol{Y}_{ob,t}^{(1)}, \ldots, \boldsymbol{Y}_{ob,t}^{(N_{ob})}$, which is a sample from the posterior predictive distribution. The sample $\boldsymbol{Y}_{ob,t}^{(1)}, \ldots, \boldsymbol{Y}_{ob,t}^{(N_{ob})}$ allows us to estimate the efficient frontier, see section 2.6. Sample efficient frontier.

Chapter 4

Numerical study

In this chapter we assess the performance of the different priors as well as the method based on the frequentist statistics within a numerical study (see, e.g. Jobson and Korkie (1981)[23], Okhrin and Schmid (2006)[28], Bodnar et al. (2017b)[11], (2019)[9]). The overestimation of the sample efficient frontier is well known (c.f., Basak et al. (2005)[4], Siegel and Woodgate (2007)[31], Bodnar and Bodnar (2010)[8]) and therefore we expect the sample efficient frontier to be optimistic.

4.1 Simulation study

In this section, we provide a detailed analysis of the prior impact on optimal portfolio selection based on an extensive Monte-Carlo simulation study. The results of Proposition 4.6 of Bernard and Smith (2000)[6] ensure that the conditional multivariate normal distribution satisfies the assumptions of infinitely exchangeability and of multivariate centred spherical symmetry. Using this result, we assume that asset returns independent and identically distributed as $\boldsymbol{X}_t|\boldsymbol{\mu}, \boldsymbol{\Sigma} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. That is, $\boldsymbol{X}_t$ is conditionally multivariate normal distributed given the mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)^T$ and covariance matrix $\boldsymbol{\Sigma}$. In each repetition we generate the elements of $\boldsymbol{\mu}$ from the uniform distribution on the interval $[-0.01, 0.01]$, that is $\mu_i \sim Unif(-0.01, 0.01)$. For the covariance matrix we consider the decomposition $\boldsymbol{\Sigma} = \boldsymbol{DRD}$, where $\boldsymbol{R}$ is the correlation matrix and $\boldsymbol{D} = diag(\delta_1, \ldots, \delta_k)$ is the diagonal matrix of standard deviations.

Table 4.1: Average absolute deviation (AD) of the estimated portfolio expected return and of the estimated portfolio variance from their population values for low volatility.

		AD portfolio expected return				AD portfolio variance			
		n=50	n=75	n=100	n=130	n=50	n=75	n=100	n=130
k = 5	Sample	0.1209	0.0883	0.0790	0.0640	0.0024	0.0018	0.0016	0.0013
k = 5	Diffuse	**0.1008**	**0.0811**	**0.0711**	**0.0605**	**0.0020**	**0.0016**	**0.0014**	**0.0012**
k = 5	Conjugate	0.3333	0.2206	0.1765	0.1408	0.0067	0.0044	0.0035	0.0028
k = 5	Hierarchical	0.5774	0.5472	0.5658	0.5425	0.0115	0.0109	0.0113	0.0108
k = 5	Objective-based	0.6016	0.3705	0.2966	0.2171	0.0120	0.0074	0.0059	0.0043
k = 10	Sample	0.4018	0.2831	0.2241	0.1748	0.0080	0.0057	0.0045	0.0035
k = 10	Diffuse	**0.2407**	**0.2028**	**0.1699**	**0.1379**	**0.0048**	**0.0041**	**0.0034**	**0.0028**
k = 10	Conjugate	0.8376	0.5309	0.3910	0.2995	0.0167	0.0106	0.0078	0.0060
k = 10	Hierarchical	1.3452	1.3599	1.3573	1.3085	0.0269	0.0272	0.0271	0.0262
k = 10	Objective-based	1.5212	0.9915	0.7329	0.5398	0.0304	0.0198	0.0147	0.0108
k = 25	Sample	4.0719	1.9430	1.2754	0.9442	0.0814	0.0389	0.0255	0.0189
k = 25	Diffuse	**0.8712**	**0.6229**	**0.4935**	**0.4272**	**0.0174**	**0.0125**	**0.0099**	**0.0085**
k = 25	Conjugate	2.1114	1.1508	0.8259	0.6371	0.0422	0.0230	0.0165	0.0127
k = 25	Hierarchical	3.7844	3.7926	3.7410	3.7703	0.0757	0.0758	0.0748	0.0754
k = 25	Objective-based	6.5587	3.5531	2.4233	1.8079	0.1312	0.0711	0.0485	0.0362
k = 40	Sample	26.7445	7.3245	4.1978	2.7878	0.5349	0.1465	0.0840	0.0558
k = 40	Diffuse	2.2564	1.1963	0.8822	0.7189	0.0451	0.0239	0.0176	0.0144
k = 40	Conjugate	**1.3652**	**1.0978**	**0.8668**	**0.7149**	**0.0273**	**0.0220**	**0.0173**	**0.0143**
k = 40	Hierarchical	6.2525	6.2715	6.2433	6.2205	0.1250	0.1254	0.1249	0.1244
k = 40	Objective-based	17.6229	8.1610	5.0860	3.5604	0.3525	0.1632	0.1017	0.0712

Notes: The smallest values are depicted in bold. The risk aversion coefficient is set to $\gamma = 50$. For the population covariance matrix we consider low volatility, i.e. $\delta_i \sim Unif(0.002, 0.005)$.

The correlation matrix is set to $\boldsymbol{R} = (1-\rho)\boldsymbol{I}_k + \rho\boldsymbol{J}_k$ with $\rho = 0.6$, $\boldsymbol{I}_k$ is the k-dimensional identity matrix, and $\boldsymbol{J}_k$ is the k-dimensional matrix of ones. We consider two choices of volatility; low volatility with $\delta_i \sim Unif(0.002\ , 0.005)$ and high volatility with $\delta_i \sim Unif(0.005, 0.02)$. In each repetition we generate a sample $\boldsymbol{X}_t^{(1)}, ..., \boldsymbol{X}_t^{(n)}$. We put $k \in \{5, 10, 25,\ 40\}$, $n \in \{50, 75, 10, 130\}$, and $\gamma = 50$.

In the case of the conjugate prior the precision parameters are $m_c = d_c = 50$, while the prior mean $\boldsymbol{r}_c = \boldsymbol{\mu} + 0.5\boldsymbol{\epsilon}$ and the prior on $\boldsymbol{\Sigma}$ is $\boldsymbol{S}_c = \boldsymbol{\Sigma} + 0.5\boldsymbol{\Delta}$ with $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_k)^T$ and $\boldsymbol{\Delta} = diag(\delta_1^2, \ldots, \delta_k^2)$ where $\epsilon_i \sim Unif(-0.01, 0.01)$ and $\delta_i \sim Unif(0.001, 0.005)$. For the hierarchical prior the precision parameter

Table 4.2: Average absolute deviation (AD) of the estimated portfolio expected return and of the estimated portfolio variance from their population values for high volatility.

		AD portfolio expected return				AD portfolio variance			
		n=50	n=75	n=100	n=130	n=50	n=75	n=100	n=130
k = 5	Sample	0.0140	0.0103	0.0085	0.0076	0.0003	**0.0002**	**0.0002**	0.0002
k = 5	Diffuse	**0.0115**	**0.0090**	**0.0079**	**0.0071**	**0.0002**	**0.0002**	**0.0002**	**0.0001**
k = 5	Conjugate	0.0453	0.0298	0.0212	0.0173	0.0009	0.0006	0.0004	0.0003
k = 5	Hierarchical	0.0430	0.0382	0.0349	0.0344	0.0009	0.0008	0.0007	0.0007
k = 5	Objective-based	0.0587	0.0371	0.0262	0.0216	0.0011	0.0007	0.0005	0.0004
k = 10	Sample	0.0428	0.0305	0.0222	0.0181	0.0009	0.0006	0.0004	0.0004
k = 10	Diffuse	**0.0254**	**0.0215**	**0.0165**	**0.0148**	**0.0005**	**0.0004**	**0.0003**	**0.0003**
k = 10	Conjugate	0.1059	0.0678	0.0472	0.0352	0.0021	0.0014	0.0009	0.0007
k = 10	Hierarchical	0.1154	0.1104	0.1010	0.0967	0.0023	0.0022	0.0020	0.0019
k = 10	Objective-based	0.1572	0.1005	0.0699	0.0517	0.0031	0.0020	0.0014	0.0010
k = 25	Sample	0.4138	0.1973	0.1377	0.0930	0.0083	0.0039	0.0028	0.0019
k = 25	Diffuse	**0.0878**	**0.0625**	**0.0501**	**0.0411**	**0.0018**	**0.0012**	**0.0010**	**0.0008**
k = 25	Conjugate	0.2755	0.1428	0.1019	0.0710	0.0055	0.0029	0.0020	0.0014
k = 25	Hierarchical	0.3603	0.3477	0.3440	0.3207	0.0072	0.0069	0.0069	0.0064
k = 25	Objective-based	0.6991	0.3561	0.2483	0.1688	0.0140	0.0071	0.0050	0.0034
k = 40	Sample	2.8357	0.7337	0.4399	0.2966	0.0567	0.0147	0.0088	0.0059
k = 40	Diffuse	0.2214	**0.1169**	**0.0900**	**0.0741**	0.0044	**0.0023**	**0.0018**	**0.0015**
k = 40	Conjugate	**0.2074**	0.1300	0.1012	0.0825	**0.0041**	0.0026	0.0020	0.0017
k = 40	Hierarchical	0.5993	0.6045	0.5951	0.5808	0.0120	0.0121	0.0119	0.0116
k = 40	Objective-based	2.1010	0.8130	0.5108	0.3501	0.0420	0.0163	0.0102	0.0070

Notes: The smallest values are depicted in bold. The risk aversion coefficient is set to $\gamma = 50$. For the population covariance matrix we consider high volatility, i.e. $\delta_i \sim Unif(0.005, 0.02)$.

$d_h = \kappa_h = 50$ and $\boldsymbol{S}_h = \boldsymbol{S}_c = \boldsymbol{\Sigma}+0.5\boldsymbol{\Delta}$. The parameter $\xi \sim Unif(-0.01, 0.01)$ and $\eta \sim Inverse - Gamma(\varepsilon_1, \varepsilon_2)$ with $\varepsilon_1 = 0.0001$ and $\varepsilon_2 = 0.0001$. For the objective-based prior s^2 is the average of the diagonal elements of $\boldsymbol{\Sigma}$, $\sigma^2_{ob} = 50$, and $d_{ob} = 50$. The prior matrix $\boldsymbol{S}_{ob} = \boldsymbol{S}_c = \boldsymbol{\Sigma} + 0.5\boldsymbol{\Delta}$ and $\boldsymbol{w}_{ob} = \mathbf{1}_k/k$ which equates to a equally weighted portfolio.

The results of Theorem 1 and Theorem 2 give us an analytical expression of the expected value and variance of the posterior predictive distribution for the diffuse and conjugate prior, respectively. For the hierarchical prior we need to estimate the expected portfolio return and portfolio variance.

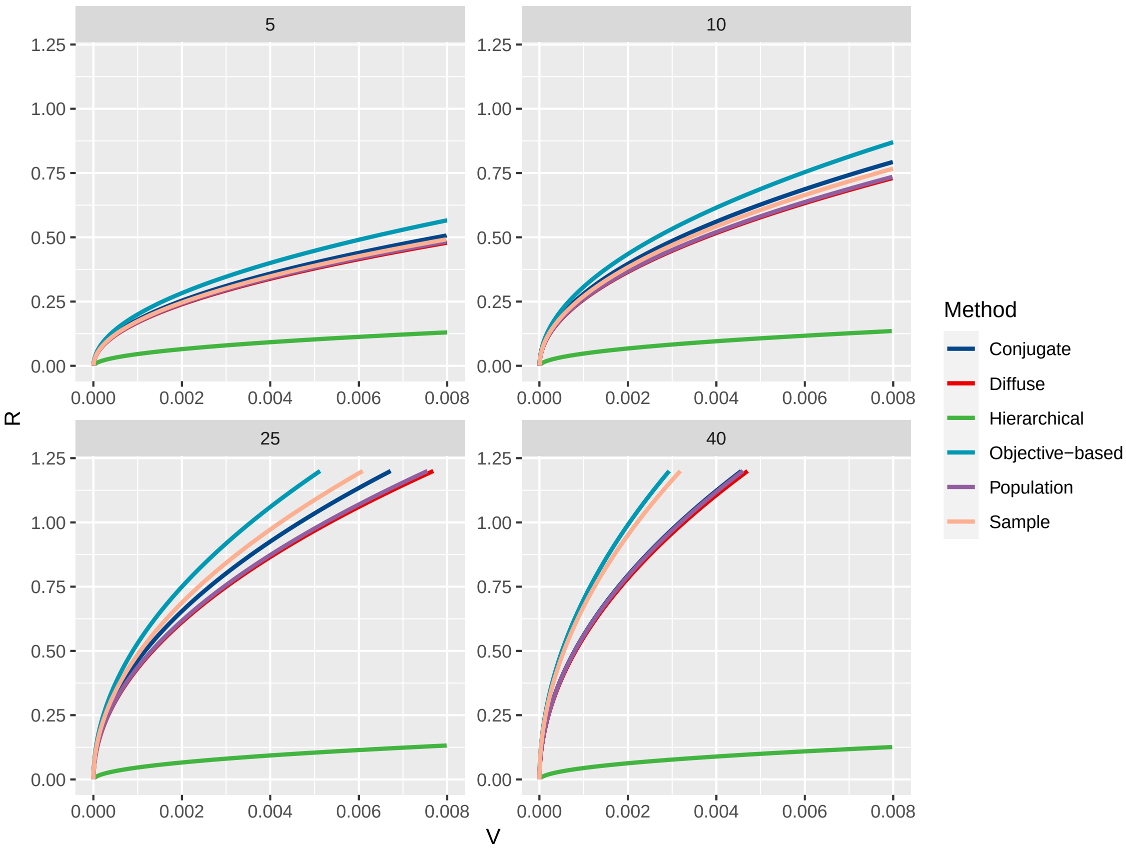

Figure 4.1: The efficient frontier for all considered methods, were $n = 130$ and $k \in \{5, 10, 25, 40\}$ for low volatility.

This is done by first drawing $\tilde{\boldsymbol{\mu}}_h$ and $\tilde{\boldsymbol{\Sigma}}_h$ from the conditional posterior distributions (3.15) and (3.17), respectively. Then, draw $\boldsymbol{Y}_{h,t}$ from $N_k(\tilde{\boldsymbol{\mu}}_h, \tilde{\boldsymbol{\Sigma}}_h)$. We repeat these two steps N_h times and estimate the sample mean and variance using the sample $\boldsymbol{Y}_{h,t}^{(1)}, \dots, \boldsymbol{Y}_{h,t}^{(N_h)}$. This procedure is performed for each repetition.

The process for the objective-based prior is similar to that of the hierarchical prior. The only difference is that we need to draw $\tilde{\boldsymbol{\Sigma}}_{ob}$ by implementing the rejection sampling algorithm (see appendix A.3). The draws from this rejection algorithm are in fact a sample from the target density $f(\boldsymbol{X})$, there is no approximation involved (see Givens and Hoeting(2012)[19]).

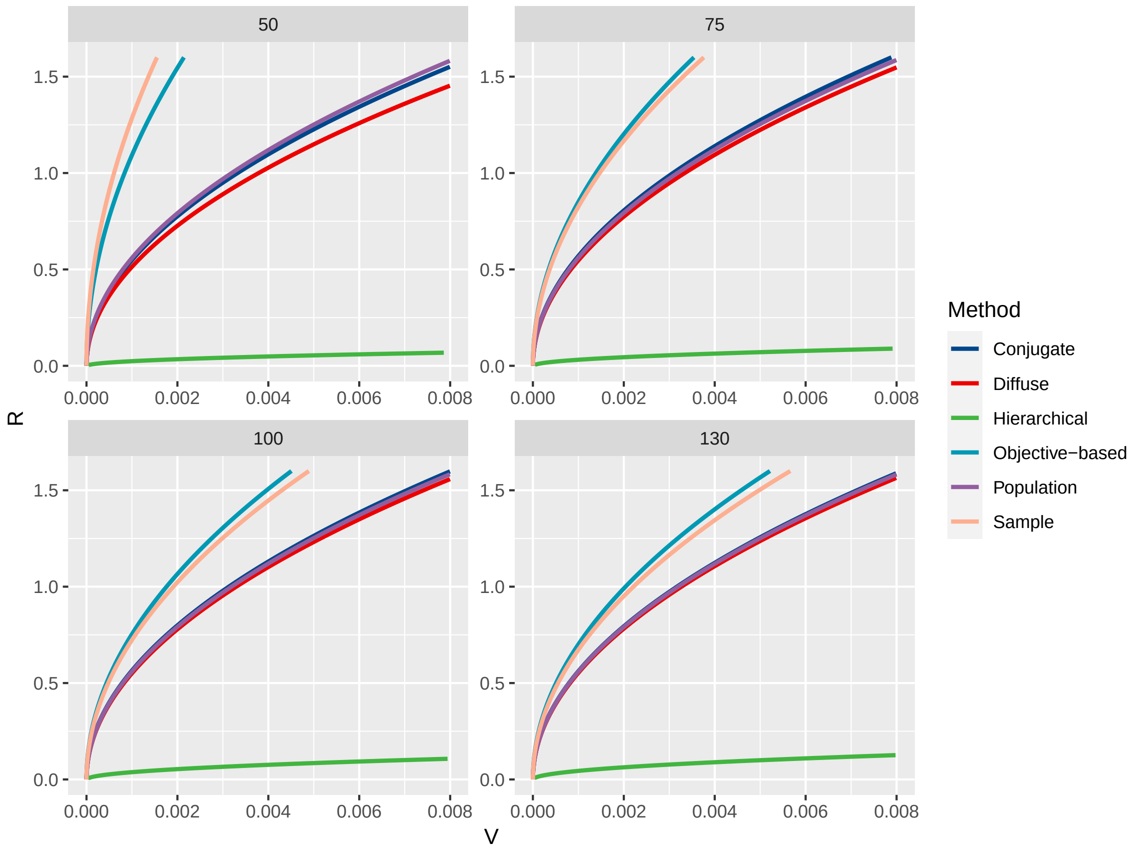

Figure 4.2: The efficient frontier for all considered methods, were $k = 40$ and $n \in \{50, 75, 100, 130\}$ for low volatility.

Then, given $\tilde{\boldsymbol{\mu}}_{ob}$ and $\tilde{\boldsymbol{\Sigma}}_{ob}$ from the marginal posterior distributions (3.20) and (3.22), respectively, we draw $\boldsymbol{Y}_{ob,t}$ from $N_k(\tilde{\boldsymbol{\mu}}_{ob}, \tilde{\boldsymbol{\Sigma}}_{ob})$. We repeat these two steps N_{ob} times and estimate the sample mean and variance using the sample $\boldsymbol{Y}_{ob,t}^{(1)}, \dots, \boldsymbol{Y}_{ob,t}^{(N_h)}$. The results are based on $B = 1000$ independent repetitions and $N_h = N_{ob} = 10000$ for each repetition.

As a measure of performance, the average absolute deviance from the estimator to the true population value was computed for portfolio expected return and variance. The values are summarised in Table 4.1 for low volatility, and Table 4.2 for high volatility. In the case of low volatility we observe that, when $k \in \{5, 10, 25\}$, the diffuse estimator leads to the best performance in terms of point estimation of the expected return and variance.

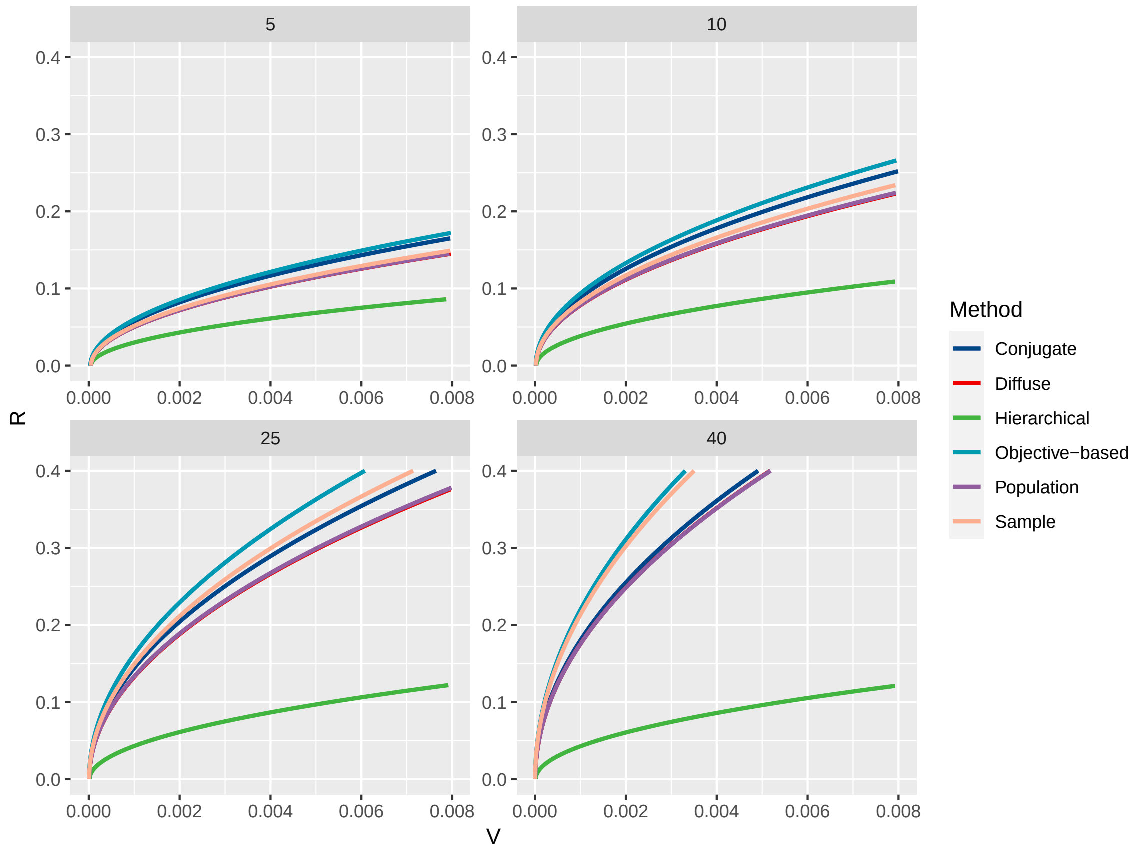

Figure 4.3: The efficient frontier for all considered methods, were $k = 40$ and $n \in \{50, 75, 100, 130\}$ for high volatility.

However, as the portfolio dimension increases the conjugate estimator improves considerably. For $k = 40$, the conjugate estimator leads to the best point estimation of the expected return and variance, for all considered sample sizes. The objective-based estimator performs worst out of the five estimator when the sample size is $n \in \{50, 75\}$, as the sample size increases beyond 75 the objective-based estimator only outperforms the hierarchical estimator. For a given portfolio dimension, the hierarchical estimator deviate by a similar amount regardless of the sample size. For high volatility, varying k and n, similar observations are made about the deviations of the hierarchical estimator.

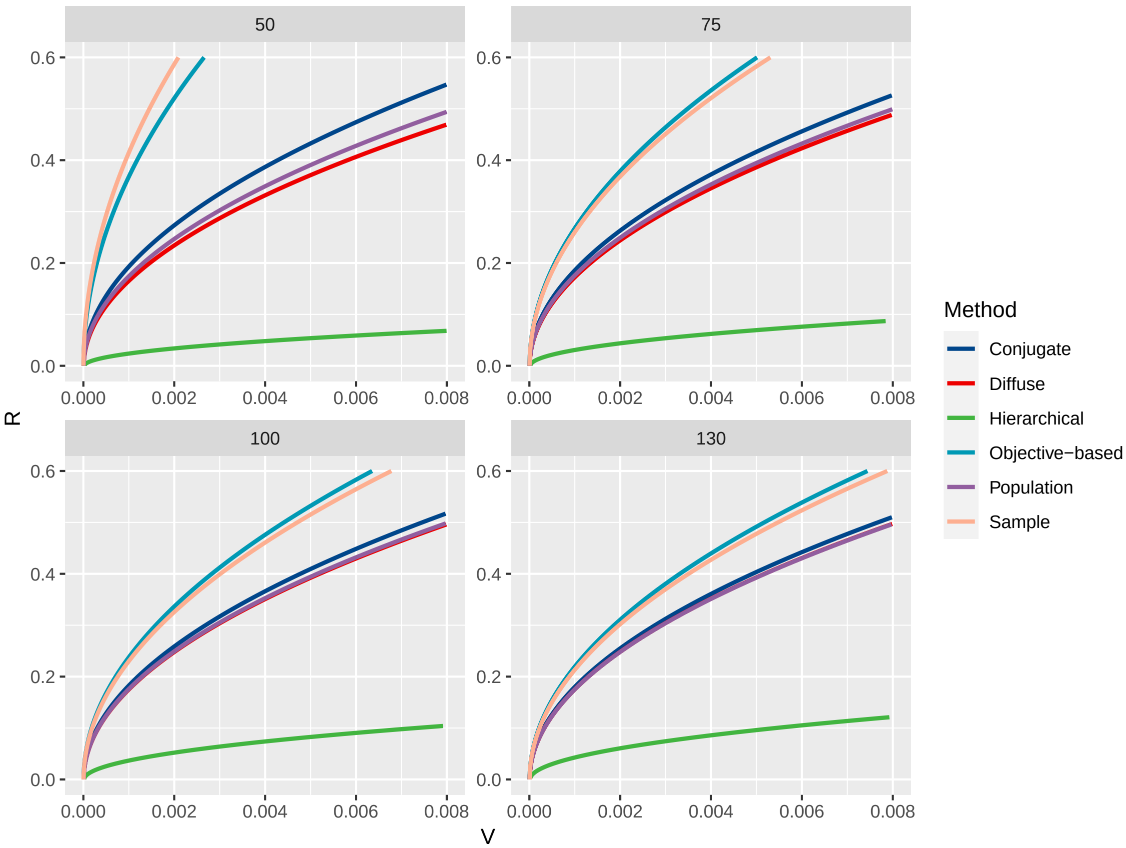

Figure 4.4: The efficient frontier for all considered methods, were $n = 130$ and $k \in \{5, 10, 25, 40\}$ for high volatility.

For $k = 5$ and $n \in \{75, 100\}$ the sample estimator performs as good as the diffuse estimator, in terms of point estimation of the portfolio variance. While the conjugate estimator outperformed the diffuse estimator when $k = 40$ for low volatility, this is only the case when $n = 50$ in the case of high volatility.

Regarding the efficient frontier for low volatility, varying k with fixed $n = 130$ in Figure 4.1 and varying n with fixed $k = 40$ in Figure 4.2, the following is observed. The objective-based estimator shows the most overestimation of the population efficient frontier for all considered portfolio dimensions and sample sizes, the only exception is when the portfolio dimension is large and the sample size $n = 50$. In this case the sample estimator exhibits even larger overestimation, compared to the objective-based estimator. The diffuse estimator tends to underestimate the population efficient frontier, if only by a negligible amount. The underestimation of the diffuse estimator tends to decrease as the sample size increases. As the portfolio dimension increases or the sample size decreases, the sample estimator performs worse as its overestimation, of the population efficient frontier, increases. In contrast, increasing the portfolio dimension or the sample size, the conjugate estimator performs better, showing less overestimation. For all considered portfolio dimensions and sample sizes, the hierarchical estimator, in contrast to most estimators, underestimates the population efficient frontier tremendously. And as the portfolio dimension increases, so does the underestimation of the hierarchical estimator. The reason for this large underestimation may be that we shrink all elements of the mean vector by the same amount. As such, the slope coefficient of the efficient frontier will be shrunk to zero. This is exactly what is observed in the figures.

4.2 Empirical illustration

4.2.1 Data

The data are comprised of stocks from OMX Stockholm 30 combined with a number of stocks from the Stockholm market, thus allowing for portfolios consisting of more than 30 assets. The stocks from the Stockholm market are selected by most traded. We consider portfolio dimensions $k \in \{5, 10, 25, 40\}$ for daily and weekly returns. For given sample size n, we use two samples. For daily returns, the first sample ends on the 29th of March 2021 and begins n-days earlier. For weekly returns, the first sample ends on the 29th of March 2021 and begins n-weeks earlier. For daily returns, the second sample ends right before the first sample begins, and starts n-days earlier. For weekly returns, the second sample ends right before the first sample begins, and starts n-weeks earlier. The first sample is used to asses the performance of the considered priors and the second sample is used to determine the value of the hyperparameters $\boldsymbol{\mu}_c$, $\boldsymbol{S}_c$, $\boldsymbol{S}_h$, and $\boldsymbol{S}_{ob}$. Letting $\hat{\boldsymbol{x}}$ denote the sample mean and $\hat{\boldsymbol{S}}$ denote the sample covariance of the second sample, the hyperparameters previously mentioned are, $\boldsymbol{r}_c = \hat{\boldsymbol{x}} + 0.5\boldsymbol{\epsilon}$ and $\boldsymbol{S}_c = \boldsymbol{S}_h = \boldsymbol{S}_{ob} = \hat{\boldsymbol{S}} + 0.5\boldsymbol{\Delta}$, with $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_k)^T$ and $\boldsymbol{\Delta} = diag(\delta_1^2, \dots, \delta_k^2)$ where $\epsilon_i \sim Unif(-0.01, 0.01)$ and $\delta_i \sim Unif(0.001, 0.005)$. We consider samples of

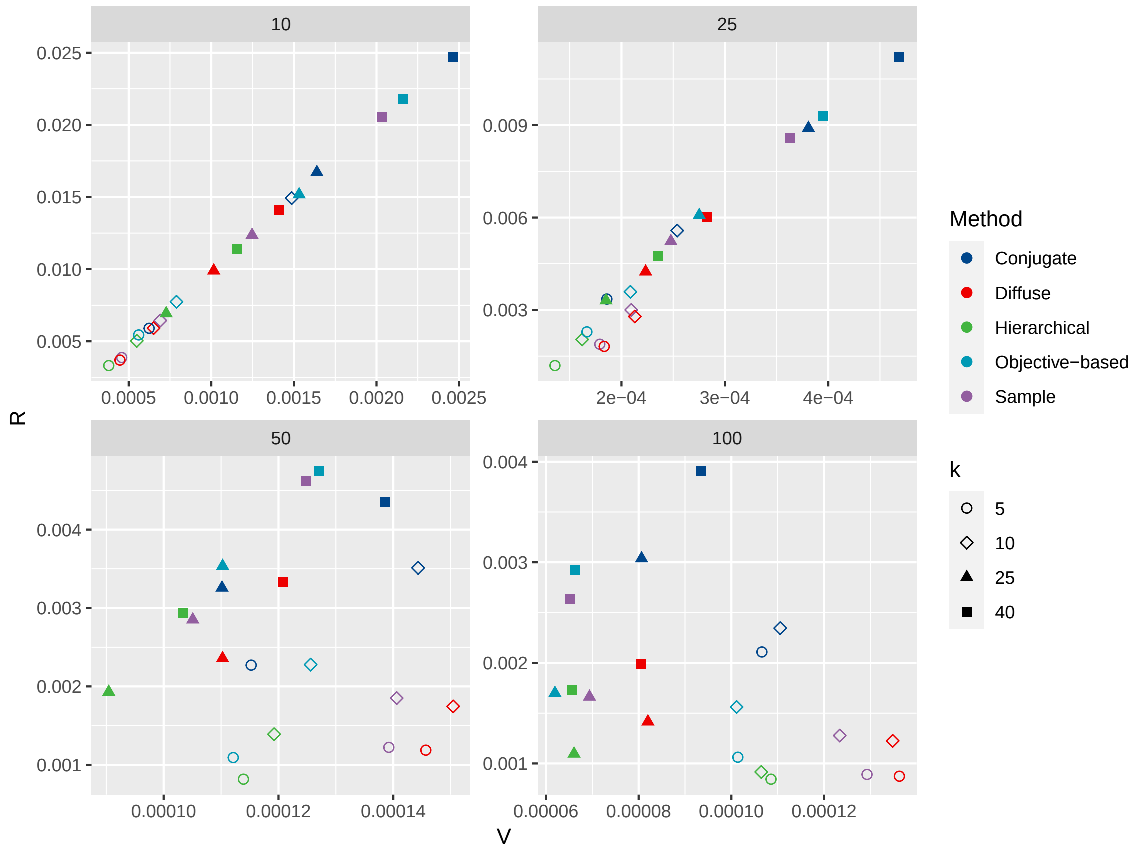

Figure 4.5: The expected portfolio return and portfolio variance for all considered methods, were $n = 130$, $\gamma \in \{10, 25, 50, 100\}$, and $k \in \{5, 10, 25, 40\}$ for daily data.

size $n \in \{52, 78, 104, 130\}$, corresponding to two and a half months up to six and a half months of daily returns, and one year up to two and a half years of weekly returns.

4.2.2 Results for daily data

In Figure 4.5, fixing $n = 130$, considering different portfolio dimensions $k \in \{5, 10, 25, 40\}$, and different levels of risk aversion $\gamma \in \{10, 25, 50, 100\}$, we observe the following.

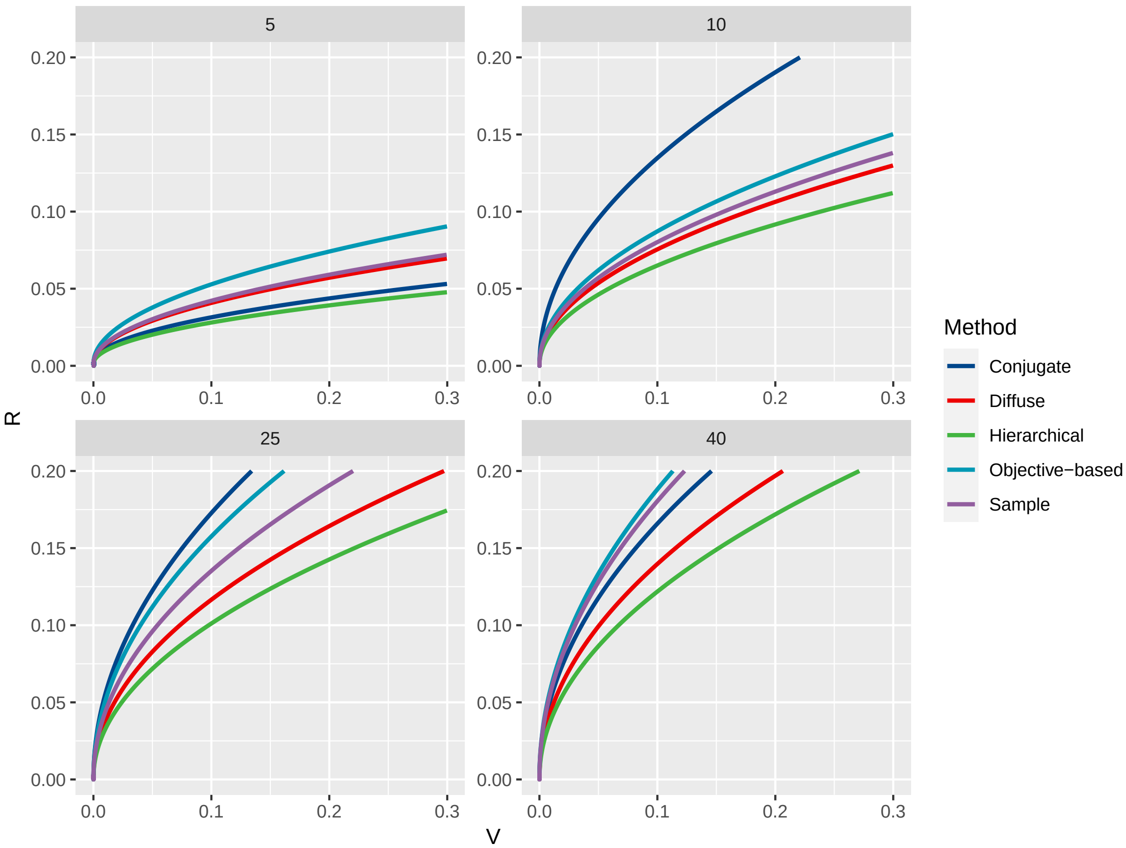

Figure 4.6: The efficient frontier for all considered methods, were $n = 130$ and $k \in \{5, 10, 25, 40\}$ for daily data.

For all portfolio dimensions and levels of risk aversion, the conjugate estimator estimates the highest expected return and the hierarchical estimator estimates the lowest expected return. For low levels of risk aversion, i.e. $\gamma \in \{10, 25\}$, the same can be observed in terms of the portfolio variance, the conjugate estimator estimates the highest variance while the hierarchical estimator estimates the lowest variance. When the levels of risk aversion is low, i.e. $\gamma \in \{10, 25\}$, the objective-based estimator estimates the second highest expected return and variance among all estimators. However, when the levels of risk aversion is high, i.e. $\gamma \in \{50, 100\}$, and the portfolio dimension small, i.e. $k \in \{5, 10\}$, the objective-based estimator estimates of the expected return is among the highest and the variance is the among the

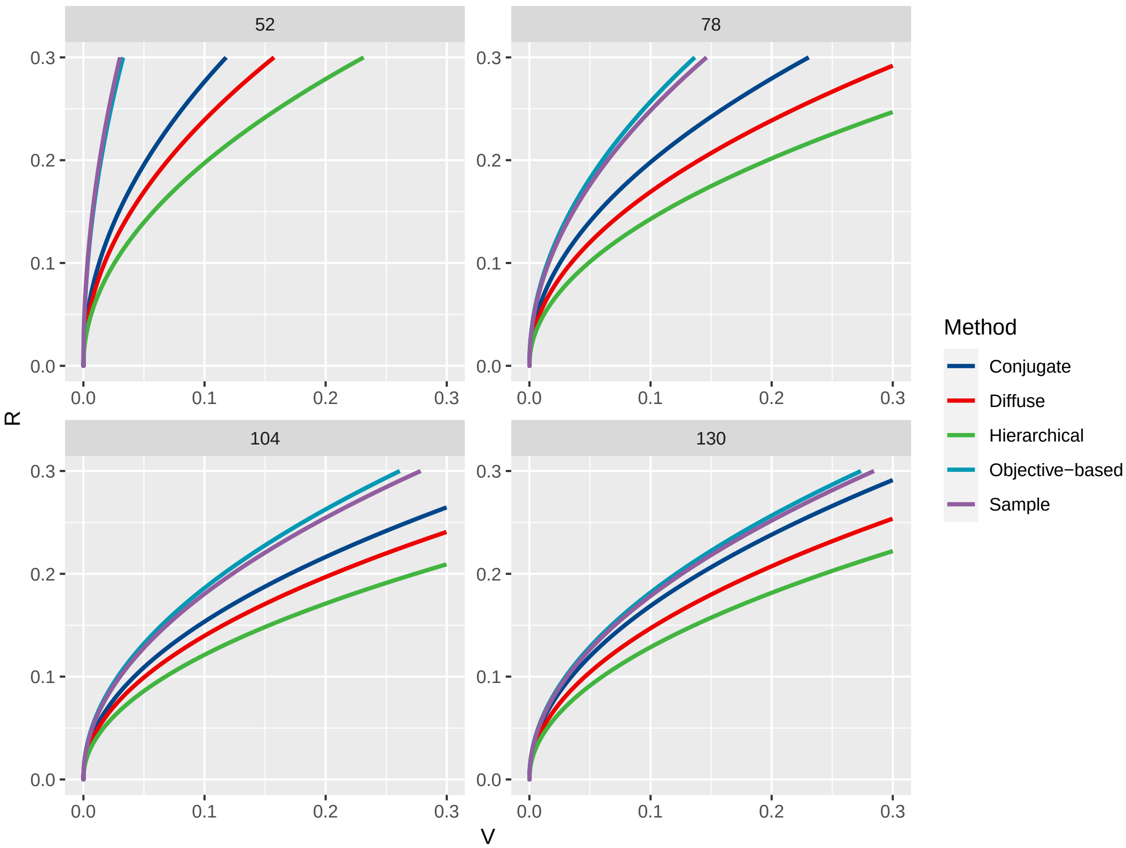

Figure 4.7: The efficient frontier for all considered methods, were $k = 40$ and $n \in \{52, 78, 104, 130\}$ for daily data.

lowest. That is, for the investor who is risk averse and considers small portfolio dimensions, the objective-based estimator promises high expected portfolio return combined with low portfolio variance. For a given level of risk aversion, the diffuse estimator is closely related to the sample estimator, the difference being that the diffuse estimator tends to estimate lower expected return and higher variance, compared to the sample estimator.

Regarding the efficient frontier, Figure 4.6 shows the estimated efficient frontier for a fixed $n = 130$ and considering different portfolio dimensions $k \in \{5, 10, 25, 40\}$. The diffuse efficient frontier always lies below the sample and the objective-based efficient frontier, and deviate stronger when the portfolio dimension gets larger. The conjugate efficient frontier also tends to lie above the diffuse efficient frontier, the only exception being when $k = 5$. For a given portfolio dimension, the objective-based estimator estimates the highest, or second highest, expected return for a given level of variance. It is remarkable that, for all considered portfolio dimensions, the hierarchical efficient frontier estimates the lowest expected return for a given level of variance.

Figure 4.7 shows the estimated efficient frontier for a fixed $k = 40$ and considering different length of historical daily returns $n \in \{52, 78, 104, 130\}$. For longer histories, i.e. $n \in \{78, 104, 130\}$, the order of the efficient frontiers, from highest to lowest, is as follows: objective-based, sample, conjugate, diffuse, and hierarchical. When $n = 52$ the sample efficient frontier lie slightly above the objective-based efficient frontier. For all n the sample and objective-based efficient frontier lie close to each other and exhibit the strongest overestimation. However, the sample and objective-based efficient frontier show weaker overestimation with growing sample size.

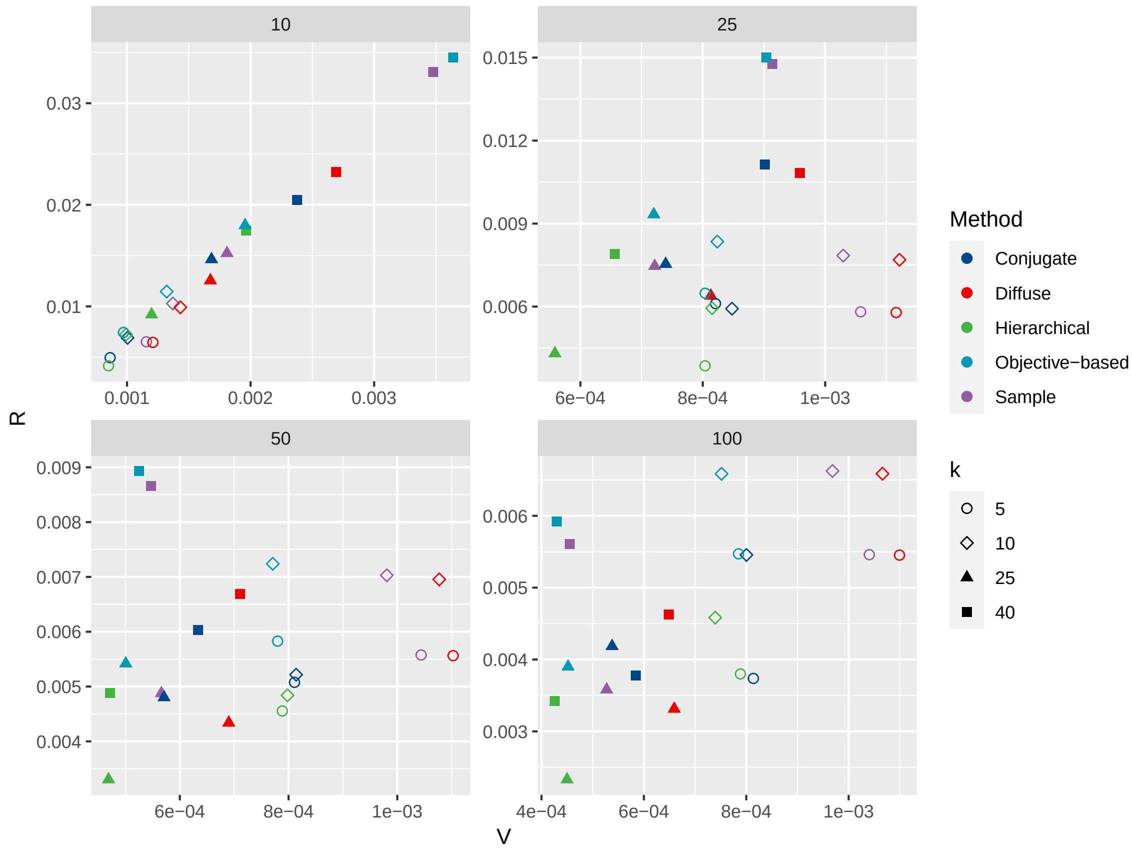

Figure 4.8: The expected portfolio return and portfolio variance for all considered methods, were $n = 130$, $\gamma \in \{10, 25, 50, 100\}$, and $k \in \{5, 10, 25, 40\}$ for weekly data.

4.2.3 Results for weekly data

In Figure 4.8, fixing $n = 130$, considering different portfolio dimensions $k \in \{5, 10, 25, 40\}$, and different levels of risk aversion $\gamma \in \{10, 25, 50, 100\}$, we observe the following.

The hierarchical estimator estimates the lowest, or second lowest, expected return and variance for all considered portfolio dimensions and levels of risk aversion. The objective-based estimator estimates the highest, or second highest, expected return and the lowest, or second lowest, variance among all estimators. The diffuse estimator estimates the highest variance for most

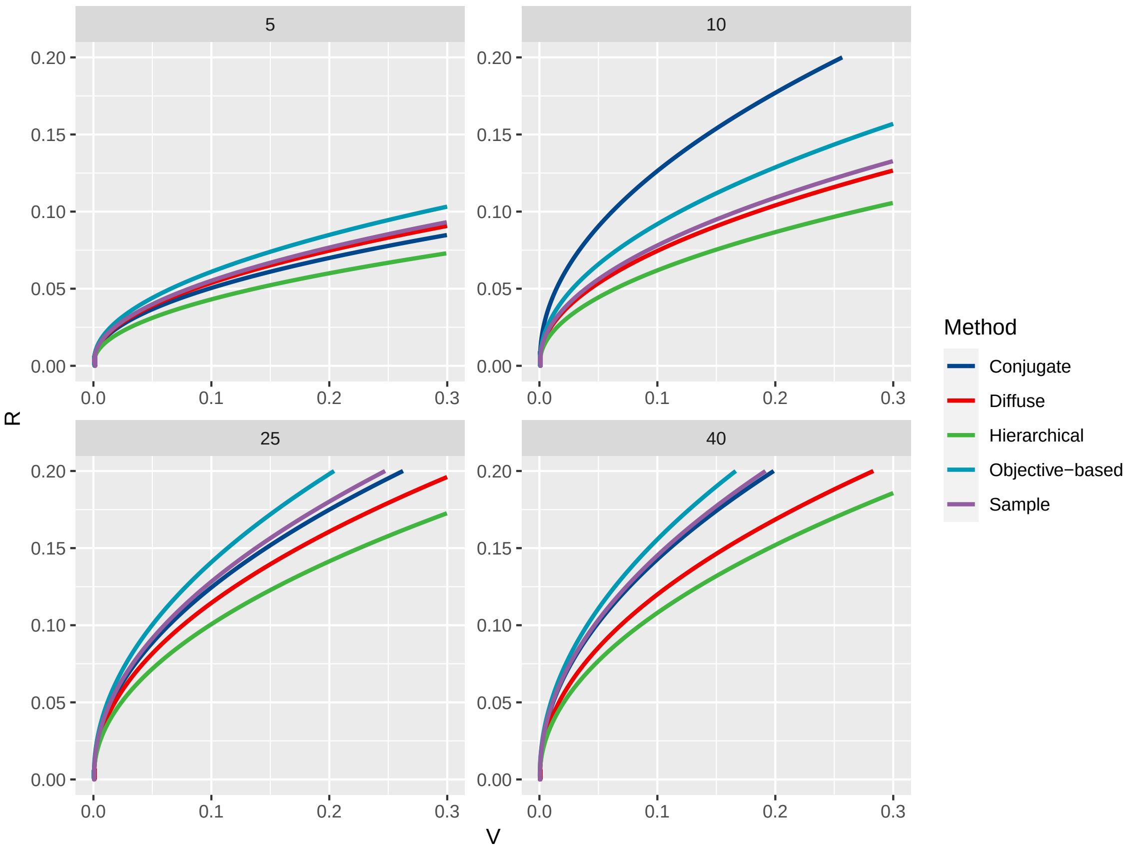

Figure 4.9: The efficient frontier for all considered methods, were $n = 130$ and $k \in \{5, 10, 25, 40\}$ for weekly data.

portfolio dimensions and levels of risk aversion. For small portfolio dimensions, i.e. $k \in \{5, 10\}$, and risk aversion $\gamma \in \{25, , 50, 100\}$ the estimated variance for the sample estimator and the diffuse estimator is significantly higher than the other estimators. Similarly to the results for daily returns the diffuse estimator estimates lower expected return and higher variance, compared to the sample estimator.

Regarding the efficient frontier, Figure 4.9 shows the estimated efficient frontier for a fixed $n = 130$ and considering different portfolio dimensions $k \in \{5, 10, 25, 40\}$. Similarly to the results from daily return, the diffuse efficient frontier lies below the sample efficient frontier, and the hierarchical efficient frontier estimates the lowest return for a given variance. As the

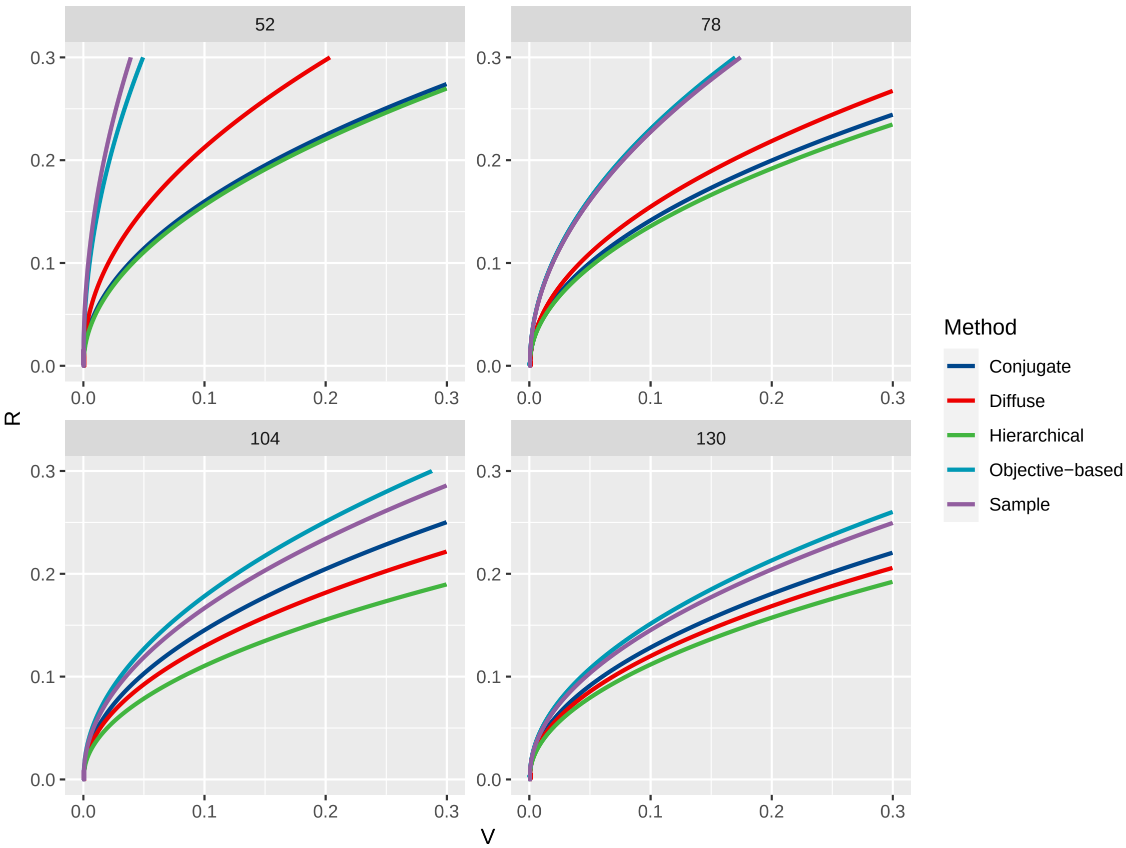

Figure 4.10: The efficient frontier for all considered methods, were $k = 40$ and $n \in \{52, 78, 104, 130\}$ for weekly data.

portfolio dimension increases the diffuse and the hierarchical efficient frontier deviate from the other efficient frontiers, estimating the lowest expected returns for a given level of variance. The objective-based estimator tends to estimate the highest expected return for a given level of variance, and the conjugate estimator tends to lie close to the sample estimator. However, this is not the case for $k = 10$, in this case the conjugate estimator stands out by estimating significantly higher expected return for a given level of variance, compared to all other methods.

In Figure 4.10, fixing $k = 40$ and consider different length of historical weekly returns $n \in \{52, 78, 104, 130\}$ for the efficient frontier. For smaller samples, i.e. $n \in \{52, 78\}$, the conjugate efficient frontier lies close to the hierarchical efficient frontier. As the sample size increases the diffuse efficient frontier lies closer to the hierarchical efficient frontier. When $n \in \{104, 130\}$ the diffuse estimator estimates the second lowest expected return for a given level of variance. Regardless of portfolio dimension or sample size, the hierarchical efficient frontier lies below all other efficient frontiers. For all n the sample and objective-based efficient frontier lies close to each other and exhibit the strongest overestimation. This is especially prominent when the length of historical weekly returns is short.

Chapter 5

Conclusion

In this thesis we consider the problem of optimal portfolio choice from the Bayesian perspective. We consider four different estimators; the diffuse, the conjugate, the hierarchical, and the objective-based prior. For the diffuse and conjugate estimators we use the stochastic representation in order to draw samples from the posterior predictive distribution. While, for the hierarchical and the objective-based estimator we derive the conditional posterior distributions, for the parameters of the asset returns, in order to draw samples from the posterior predictive distribution. These are compared to each other and to an estimator based on frequentist statistics. In order to compare the different estimators we perform an extensive comparison study via Monte Carlo simulation. Also, an empirical illustration, based on stocks from the Stockholm market, is performed.

In the simulation study the aim was to assess the performance of the considered estimators. This is done by measuring the average absolute deviation of the point estimates, of the expected portfolio return and the portfolio variance, from the true population values. Also, we derive the efficient frontier, the set of optimal portfolios, for varying portfolio dimensions and sample sizes. For most portfolio dimensions and sample sizes the diffuse estimator deviated the least. However, in some cases, when the portfolio dimension is large, the conjugate estimator outperformed the diffuse estimator. The objective-based estimator showed the most overestimation of the population efficient frontier. On the other hand, the hierarchical estimator showed tremendous underestimation of the population efficient frontier. The reason for this large underestimation may be that we shrink all elements of the mean vector by the same amount. As such, the slope coefficient of the efficient frontier will be shrunk to zero. This is exactly what is observed in the figures.

In addition to the simulation study an empirical illustration was done based on stocks from the Stockholm market. In the empirical illustration we observed that, in most cases, the estimators perform as expected based on the observations from the simulation study. However, the hierarchical estimator performed much better, in that it seems to show much less underestimation, relative to the other estimators. Based on the simulation study, depending on the portfolio dimension, either the diffuse or the conjugate estimator showed the best performance. It is possible that the hierarchical estimator, when applied to real data, is favourable due to its modest estimation of the expected return for a given level of variance.

Conclusively, we found that the diffuse, the conjugate, and the hierarchical estimator showed less overestimation, compared to sample estimator. In the simulation study, the hierarchical estimator showed immense underestimation while the objective-based estimator showed great overestimation, of the population efficient frontier. For future studies it could be of interest to investigate how the hierarchical estimator would perform for different values of the prior parameters, both in a simulation setting but primarily when applied to real data.

Appendix A

Derivations

A.1 Global minimum variance portfolio

Proof. Global minimum variance portfolio

In order to find the weights, $\boldsymbol{w}_{GMV}$, of the minimum variance portfolio such that $\boldsymbol{w}_{GMV}^T \mathbf{1} = 1$ we use the method of Lagrange multipliers. The minimization problem to be solved can be expressed as the following

$$\underset{\boldsymbol{w}}{min}\, \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w} \;\text{ subject to }\; \boldsymbol{w}^T \mathbf{1} = 1$$

The Lagrangian function that corresponds to this minimization problem, subject to the equality constraint, is given by

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

The solution to this minimization problem is obtained by solving the first order conditions with respect to $\boldsymbol{w}$. The first order conditions are

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

$$\frac{\partial L(\boldsymbol{w}, \lambda)}{\partial \lambda} = \boldsymbol{w}^T \mathbf{1} - 1 = 0. \tag{A.2}$$

Solving (A.1) with respect to $\boldsymbol{w}$ we get

$$\boldsymbol{w} = -\frac{\lambda \cdot \boldsymbol{\Sigma}^{-1}\mathbf{1}}{2}.$$

Using this expression of $\boldsymbol{w}$ we solve (A.2) with respect to λ and get

$$\lambda = -\frac{2}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}.$$

Substituting this expression of lambda in (A.1) we get desired weights of the minimum variance portfolio

$$\begin{aligned} 2 \cdot \boldsymbol{\Sigma}\boldsymbol{w} &= \frac{2}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}} \cdot \mathbf{1} \implies \\ \boldsymbol{w}_{GMV} &= \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}. \end{aligned}$$

■

A.2 Conditional posterior distribution

A.2.1 Hierarchical prior

Derivations of the conditional posterior predictive distributions, of the parameters of the asset returns, for hierarchical prior. The posterior distribution is given by

$$
\begin{aligned}
p_h(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi, \eta | \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) &\propto L(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n | \boldsymbol{\mu}, \boldsymbol{\Sigma}) p_h(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi, \eta) \\
&\propto |\boldsymbol{\Sigma}|^{-(d_h+n+1)/2} exp\left\{-\frac{1}{2} tr\left[(\eta^{-1}\boldsymbol{S}_h + (n-1)\boldsymbol{S})\boldsymbol{\Sigma}^{-1}\right]\right\} \\
&\times \eta^{-k(d_h-k-1)/2-\varepsilon_1-1} exp\left\{-\frac{\varepsilon_2}{\eta}\right\} \\
&\times exp\bigg\{ -\frac{\kappa_h}{2}(\boldsymbol{\mu} - \xi\mathbf{1})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \xi\mathbf{1}) \\
&- \frac{n}{2}(\boldsymbol{\mu} - \overline{\boldsymbol{X}})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \overline{\boldsymbol{X}})\bigg\}, \qquad \text{(A.3)}
\end{aligned}
$$

where $\mathbf{1}$ is the k-dimensional vector of ones and

$$
\begin{aligned}
L(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n | \boldsymbol{\mu}, \boldsymbol{\Sigma}) &\propto |\boldsymbol{\Sigma}|^{-n/2} \\
\times\, exp&\left\{-\frac{n}{2}(\boldsymbol{\mu} - \overline{\boldsymbol{X}})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \overline{\boldsymbol{X}}) - \frac{n-1}{2} tr\left[\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\right]\right\}.
\end{aligned}
$$

To derive the conditional posterior of $\boldsymbol{\mu}$ we focus on the third exponential term of (A.3) since that is the only term that depends on $\boldsymbol{\mu}$. We can rewrite the third exponential term of (A.3) as

$$
\begin{aligned}
&exp\bigg\{ -\frac{\kappa_h}{2}(\boldsymbol{\mu} - \xi\mathbf{1})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \xi\mathbf{1}) - \frac{n}{2}(\boldsymbol{\mu} - \overline{\boldsymbol{X}})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \overline{\boldsymbol{X}})\bigg\} \\
&= exp\bigg\{ -\frac{\kappa_h}{2}\left(\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - 2\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1}\xi\mathbf{1} + \xi\mathbf{1}^T \boldsymbol{\Sigma}^{-1}\xi\mathbf{1}\right) \\
&\qquad - \frac{n}{2}\left(\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - 2\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{X}} + \overline{\boldsymbol{X}}^T \boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{X}}\right)\bigg\} \\
&= exp\bigg\{ -\frac{1}{2}\Big[(\kappa_h + n)\, \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - 2\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1}\left(\kappa_h\xi\mathbf{1} + n\overline{\boldsymbol{X}}\right) \\
&\qquad + \kappa_h\xi\mathbf{1}^T \boldsymbol{\Sigma}^{-1}\xi\mathbf{1} + n\overline{\boldsymbol{X}}^T \boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{X}}\Big]\bigg\}
\end{aligned}
$$

$$
\begin{aligned}
&= exp\Bigg\{ -\frac{\kappa_h + n}{2} \left(\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - 2\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{r}_h + \boldsymbol{r}_h^T \boldsymbol{\Sigma}^{-1} \boldsymbol{r}_h\right) \\
&\qquad - \frac{1}{2}\Bigg[n\overline{\boldsymbol{X}}^T \boldsymbol{\Sigma}^{-1} \overline{\boldsymbol{X}} + \kappa_h \xi \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \xi \mathbf{1} \\
&\qquad - (\kappa_h + n) \left(\frac{\kappa_h \xi \mathbf{1} + n\overline{\boldsymbol{X}}}{\kappa_h + n}\right)^T \boldsymbol{\Sigma}^{-1} \left(\frac{\kappa_h \xi \mathbf{1} + n\overline{\boldsymbol{X}}}{\kappa_h + n}\right) \Bigg] \Bigg\} \\
&= exp\Bigg\{ -\frac{\kappa_h + n}{2} (\boldsymbol{\mu} - \boldsymbol{r}_h)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \boldsymbol{r}_h) \\
&\qquad - \frac{1}{2}\left(\frac{1}{\kappa_h + n}\right) \Bigg[(\kappa_h + n)\, n\overline{\boldsymbol{X}}^T \boldsymbol{\Sigma}^{-1} \overline{\boldsymbol{X}} \\
&\qquad + (\kappa_h + n)\, \kappa_h (\xi \mathbf{1})^T \boldsymbol{\Sigma}^{-1} (\xi \mathbf{1}) - (\kappa_h \xi \mathbf{1})^T \boldsymbol{\Sigma}^{-1} (\kappa_h \xi \mathbf{1}) \\
&\qquad - 2n\overline{\boldsymbol{X}}^T \boldsymbol{\Sigma}^{-1} (\kappa_h \xi \mathbf{1}) - n\overline{\boldsymbol{X}}^T \boldsymbol{\Sigma}^{-1} n\overline{\boldsymbol{X}} \Bigg] \Bigg\} \\
&= exp\Bigg\{ -\frac{\kappa_h + n}{2} (\boldsymbol{\mu} - \boldsymbol{r}_h)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \boldsymbol{r}_h) \\
&\qquad - \frac{1}{2}\left(\frac{\kappa_h n}{\kappa_h + n}\right) \Bigg[\overline{\boldsymbol{X}}^T \boldsymbol{\Sigma}^{-1} \overline{\boldsymbol{X}} \\
&\qquad + (\xi \mathbf{1})^T \boldsymbol{\Sigma}^{-1} (\xi \mathbf{1}) - 2\overline{\boldsymbol{X}}^T \boldsymbol{\Sigma}^{-1} (\xi \mathbf{1}) \Bigg] \Bigg\} \\
&= exp\Bigg\{ -\frac{\kappa_h + n}{2} (\boldsymbol{\mu} - \boldsymbol{r}_h)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \boldsymbol{r}_h) \\
&\qquad - \frac{1}{2}\left(\frac{\kappa_n n}{\kappa_n + n} (\overline{\boldsymbol{X}} - \xi \mathbf{1})^T \boldsymbol{\Sigma}^{-1} (\overline{\boldsymbol{X}} - \xi \mathbf{1})\right) \Bigg\}
\end{aligned}
$$

where

$$
\boldsymbol{r}_h = \frac{\kappa_h \xi \mathbf{1} + n\overline{\boldsymbol{X}}}{\kappa_h + n}.
$$

Using this, the posterior distribution can be expressed as

$$
\begin{aligned}
p_h(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi, \eta | \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \propto\ & |\boldsymbol{\Sigma}|^{-(d_h+n+1)/2} \\
& \times exp\left\{-\frac{1}{2}tr\left[(\eta^{-1}\boldsymbol{S}_h + (n-1)\boldsymbol{S})\boldsymbol{\Sigma}^{-1}\right]\right\} \\
& \times \eta^{-k(d_h-k-1)/2-\varepsilon_1-1} exp\left\{-\frac{\varepsilon_2}{\eta}\right\} \\
& \times exp\Bigg\{-\frac{\kappa_h+n}{2}\left(\boldsymbol{\mu}-\boldsymbol{r}_h\right)^T\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}-\boldsymbol{r}_h\right) \\
& -\frac{1}{2}\left(\frac{\kappa_n n}{\kappa_n+n}(\overline{\boldsymbol{X}}-\xi\boldsymbol{1})^T\boldsymbol{\Sigma}^{-1}(\overline{\boldsymbol{X}}-\xi\boldsymbol{1})\right)\Bigg\}. \qquad \text{(A.4)}
\end{aligned}
$$

Rearranging the terms of (A.4) we get the following expression

$$
\begin{aligned}
p_h(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi, \eta | \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \propto\ & |\boldsymbol{\Sigma}|^{-(d_h+n)/2} \\
& \times exp\left\{-\frac{1}{2}tr\left[(\eta^{-1}\boldsymbol{S}_h + (n-1)\boldsymbol{S})\boldsymbol{\Sigma}^{-1}\right]\right\} \\
& \times exp\left\{-\frac{1}{2}\left(\frac{\kappa_n n}{\kappa_n+n}(\overline{\boldsymbol{X}}-\xi\boldsymbol{1})^T\boldsymbol{\Sigma}^{-1}(\overline{\boldsymbol{X}}-\xi\boldsymbol{1})\right)\right\} \\
& \times |\boldsymbol{\Sigma}|^{-1/2} exp\left\{-\frac{\kappa_h+n}{2}\left(\boldsymbol{\mu}-\boldsymbol{r}_h\right)^T\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}-\boldsymbol{r}_h\right)\right\} \\
& \times \eta^{-k(d_h-k-1)/2-\varepsilon_1-1} exp\left\{-\frac{\varepsilon_2}{\eta}\right\} \qquad \text{(A.5)}
\end{aligned}
$$

Hence, the conditional posterior distribution of $\boldsymbol{\mu}$ is given by

$$
p_h(\boldsymbol{\mu} | \boldsymbol{\Sigma}, \xi, \eta, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \propto |\boldsymbol{\Sigma}|^{-1/2} exp\left\{-\frac{\kappa_h+n}{2}\left(\boldsymbol{\mu}-\boldsymbol{r}_h\right)^T\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}-\boldsymbol{r}_h\right)\right\}
$$

which is the kernel of a multivariate normal distribution, that is

$$
\boldsymbol{\mu} | \boldsymbol{\Sigma}, \xi, \eta, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \sim N(\boldsymbol{r}_h, \frac{1}{\kappa_h+n}\boldsymbol{\Sigma}).
$$

Integrating out $\boldsymbol{\mu}$ from (A.5) we get conditional posterior distribution of $\boldsymbol{\Sigma}$ expressed as

$$
\begin{aligned}
p_h(\boldsymbol{\Sigma}|\xi,\eta, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \propto |\boldsymbol{\Sigma}|^{-(d_h+n)/2} \\
& \times exp\left\{-\frac{1}{2} tr\left[(\eta^{-1}\boldsymbol{S}_h + (n-1)\boldsymbol{S})\boldsymbol{\Sigma}^{-1}\right]\right\} \\
& \times exp\bigg\{-\frac{1}{2}\left(\frac{\kappa_n n}{\kappa_n + n}(\overline{\boldsymbol{X}} - \xi\mathbf{1})^T\boldsymbol{\Sigma}^{-1}(\overline{\boldsymbol{X}} - \xi\mathbf{1})\right)\bigg\} \\
& \times \eta^{-k(d_h-k-1)/2-\varepsilon_1-1} exp\left\{-\frac{\varepsilon_2}{\eta}\right\}
\end{aligned}
$$

$$
\begin{aligned}
&\propto |\boldsymbol{\Sigma}|^{-(d_h+n)/2} exp\left\{-\frac{1}{2} tr\left[(\eta^{-1}\boldsymbol{S}_h + (n-1)\boldsymbol{S})\boldsymbol{\Sigma}^{-1}\right]\right\} \\
&\times exp\bigg\{-\frac{1}{2} tr\left(\frac{\kappa_n n}{\kappa_n + n}(\overline{\boldsymbol{X}} - \xi\mathbf{1})(\overline{\boldsymbol{X}} - \xi\mathbf{1})^T\boldsymbol{\Sigma}^{-1}\right)\bigg\} \\
&\times \eta^{-k(d_h-k-1)/2-\varepsilon_1-1} exp\left\{-\frac{\varepsilon_2}{\eta}\right\}.
\end{aligned}
$$

Hence, the conditional posterior distribution of $\boldsymbol{\Sigma}$ is given by

$$
\begin{aligned}
p_h(\boldsymbol{\Sigma}|\xi, \eta, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \propto |\boldsymbol{\Sigma}|^{-(d_h+n)/2} exp\bigg\{-\frac{1}{2} tr\bigg[\Big(\eta^{-1}\boldsymbol{S}_h + (n-1)\boldsymbol{S} \\
+ \frac{\kappa_n n}{\kappa_n + n}(\overline{\boldsymbol{X}} - \xi\mathbf{1})(\overline{\boldsymbol{X}} - \xi\mathbf{1})^T\Big)\boldsymbol{\Sigma}^{-1}\bigg]\bigg\},
\end{aligned}
$$

which is the kernel of inverse Wishart, that is

$$
\boldsymbol{\Sigma}|\xi, \eta, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \sim IW_k(d_h + n, \boldsymbol{\Psi}_h)
$$

where

$$
\boldsymbol{\Psi}_h = \eta^{-1}\boldsymbol{S}_h + (n-1)\boldsymbol{S} + \frac{\kappa_n n}{\kappa_n + n}(\overline{\boldsymbol{X}} - \xi\mathbf{1})(\overline{\boldsymbol{X}} - \xi\mathbf{1})^T.
$$

Finally, the marginal posterior distributions of ξ and η are given by

$$
\begin{aligned}
&p_h(\xi|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n) \propto 1\\
&p_h(\eta|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n) \propto \eta^{-k(d_h-k-1)/2-\varepsilon_1-1} exp\left\{-\frac{\varepsilon_2}{\eta}\right\}.
\end{aligned}
$$

A.2.2 Objective-based prior

Derivations of the conditional posterior distribution of $\boldsymbol{\mu}$ and the marginal posterior predictive distribution of $\boldsymbol{\Sigma}$, for objective-based prior. The posterior distribution is given by

$$
\begin{aligned}
p_{ob}(\boldsymbol{\mu},\boldsymbol{\Sigma}|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n) &\propto L(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n|\boldsymbol{\mu},\boldsymbol{\Sigma})p_{ob}(\boldsymbol{\mu},\boldsymbol{\Sigma})\\
&\propto |\boldsymbol{\Sigma}|^{-(d_{ob}+n+1)/2} exp\left\{-\frac{1}{2}tr\left[(\boldsymbol{S}_{ob}+(n-1)\boldsymbol{S})\boldsymbol{\Sigma}^{-1}\right]\right\}\\
&\times exp\bigg\{-\frac{s^2}{2\sigma^2_{ob}}(\boldsymbol{\mu}-\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob})\\
&-\frac{n}{2}(\boldsymbol{\mu}-\overline{\boldsymbol{X}})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\overline{\boldsymbol{X}})\bigg\}, \qquad \text{(A.6)}
\end{aligned}
$$

where

$$
\begin{aligned}
L(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n|\boldsymbol{\mu},\boldsymbol{\Sigma}) &\propto |\boldsymbol{\Sigma}|^{-n/2}\\
&\times exp\left\{-\frac{n}{2}(\boldsymbol{\mu}-\overline{\boldsymbol{X}})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\overline{\boldsymbol{X}})-\frac{n-1}{2}tr\left[\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\right]\right\}.
\end{aligned}
$$

To derive the conditional posterior of $\boldsymbol{\mu}$ we focus on the second exponential term of (A.6) since that is the only term that depends on $\boldsymbol{\mu}$. We can rewrite the second exponential term of (A.6) as

$$
\begin{aligned}
&exp\Bigg\{ -\frac{s^2}{2\sigma_{ob}^2}(\boldsymbol{\mu}-\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}) \\
&\qquad -\frac{n}{2}(\boldsymbol{\mu}-\overline{\boldsymbol{X}})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\overline{\boldsymbol{X}})\Bigg\} \\
&= exp\Bigg\{ -\frac{s^2}{2\sigma_{ob}^2}\left(\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}-2\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}+(\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob})^T\boldsymbol{\Sigma}^{-1}(\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob})\right) \\
&\qquad -\frac{n}{2}\left(\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}-2\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{X}}+\overline{\boldsymbol{X}}^T\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{X}}\right)\Bigg\} \\
&= exp\Bigg\{ -\frac{1}{2}\Bigg[\left(\frac{s^2}{\sigma_{ob}^2}+n\right)\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}-2\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\left(\frac{s^2}{\sigma_{ob}^2}\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}+n\overline{\boldsymbol{X}}\right) \\
&\qquad +\left(\frac{s^2}{\sigma_{ob}^2}\right)(\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob})^T\boldsymbol{\Sigma}^{-1}(\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob})+n\overline{\boldsymbol{X}}^T\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{X}}\Bigg]\Bigg\} \\
&= exp\Bigg\{ -\frac{m_{ob}}{2}\left(\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}-2\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{r}_{ob}+\boldsymbol{r}_{ob}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{r}_{ob}\right) \\
&\qquad -\frac{1}{2}\Bigg(n\overline{\boldsymbol{X}}^T\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{X}}+\frac{s^2}{\sigma_{ob}^2}(\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob})^T\boldsymbol{\Sigma}^{-1}(\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}) \\
&\qquad -\left(\frac{s^2}{\sigma_{ob}^2}+n\right)\left(\frac{\frac{s^2}{\sigma_{ob}^2}\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}+n\overline{\boldsymbol{X}}}{\frac{s^2}{\sigma_{ob}^2}+n}\right)^T\boldsymbol{\Sigma}^{-1}\left(\frac{\frac{s^2}{\sigma_{ob}^2}\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}+n\overline{\boldsymbol{X}}}{\frac{s^2}{\sigma_{ob}^2}+n}\right)\Bigg)\Bigg\} \\
&= exp\Bigg\{ -\frac{m_{ob}}{2}(\boldsymbol{\mu}-\boldsymbol{r}_{ob})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{r}_{ob}) \\
&\qquad -\frac{1}{2}\left(\frac{1}{\frac{s^2}{\sigma_{ob}^2}+n}\right)\Bigg[\left(\frac{s^2}{\sigma_{ob}^2}+n\right)n\overline{\boldsymbol{X}}^T\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{X}} \\
&\qquad +\left(\frac{s^2}{\sigma_{ob}^2}+n\right)\frac{s^2}{\sigma_{ob}^2}(\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob})^T\boldsymbol{\Sigma}^{-1}(\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}) \\
&\qquad -\left(\frac{s^2}{\sigma_{ob}^2}\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}+n\overline{\boldsymbol{X}}\right)^T\boldsymbol{\Sigma}^{-1}\left(\frac{s^2}{\sigma_{ob}^2}\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}+n\overline{\boldsymbol{X}}\right)\Bigg]\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
= exp\Bigg\{ & -\frac{m_{ob}}{2}\left(\boldsymbol{\mu}-\boldsymbol{r}_{ob}\right)^T\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}-\boldsymbol{r}_{ob}\right) \\
& -\frac{1}{2}\left(\frac{1}{\frac{s^2}{\sigma_{ob}^2}+n}\right)\Bigg[\left(\frac{s^2}{\sigma_{ob}^2}+n\right)n\overline{\boldsymbol{X}}^T\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{X}} \\
& +\left(\frac{s^2}{\sigma_{ob}^2}+n\right)\frac{s^2}{\sigma_{ob}^2}(\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob})^T\boldsymbol{\Sigma}^{-1}(\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}) \\
& -\left(\frac{s^2}{\sigma_{ob}^2}\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}\right)^T\boldsymbol{\Sigma}^{-1}\left(\frac{s^2}{\sigma_{ob}^2}\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}\right) \\
& -2n\overline{\boldsymbol{X}}^T\boldsymbol{\Sigma}^{-1}\left(\frac{s^2}{\sigma_{ob}^2}\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}\right)-n\overline{\boldsymbol{X}}^T\boldsymbol{\Sigma}^{-1}n\overline{\boldsymbol{X}}\Bigg]\Bigg\} \\
= exp\Bigg\{ & -\frac{m_{ob}}{2}\left(\boldsymbol{\mu}-\boldsymbol{r}_{ob}\right)^T\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}-\boldsymbol{r}_{ob}\right) \\
& -\frac{1}{2}\left(\frac{n\frac{s^2}{\sigma_{ob}^2}}{\frac{s^2}{\sigma_{ob}^2}+n}\right)\Bigg[\overline{\boldsymbol{X}}^T\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{X}} \\
& +(\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob})^T\boldsymbol{\Sigma}^{-1}(\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob})-2\overline{\boldsymbol{X}}^T\boldsymbol{\Sigma}^{-1}\left(\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}\right)\Bigg]\Bigg\} \\
= exp\Bigg\{ & -\frac{m_{ob}}{2}\left(\boldsymbol{\mu}-\boldsymbol{r}_{ob}\right)^T\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}-\boldsymbol{r}_{ob}\right) \\
& -\frac{1}{2}\left(\frac{n\frac{s^2}{\sigma_{ob}^2}}{\frac{s^2}{\sigma_{ob}^2}+n}\right)\Bigg[(\overline{\boldsymbol{X}}-\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob})^T\boldsymbol{\Sigma}^{-1}(\overline{\boldsymbol{X}}\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob})\Bigg]\Bigg\}
\end{aligned}
$$

where

$$
\begin{aligned}
m_{ob} &= \frac{s^2}{\sigma_{ob}^2}+n \\
\boldsymbol{r}_{ob} &= \frac{\frac{s^2}{\sigma_{ob}^2}\gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}+n\overline{\boldsymbol{X}}}{\frac{s^2}{\sigma_{ob}^2}+n}.
\end{aligned}
$$

Using this, the posterior distribution can be expressed as

$$p_{ob}(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{X}_1, \dots, \boldsymbol{X}_n) \propto |\boldsymbol{\Sigma}|^{-(d_{ob}+n+1)/2} exp\left\{-\frac{1}{2} tr\left[(\boldsymbol{S}_{ob} + (n-1)\boldsymbol{S})\boldsymbol{\Sigma}^{-1}\right]\right\}$$
$$\times\, exp\Bigg\{ -\frac{m_{ob}}{2}\left(\boldsymbol{\mu} - \boldsymbol{r}_{ob}\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} - \boldsymbol{r}_{ob}\right)$$
$$-\frac{1}{2}\left(\frac{n \cdot \frac{s^2}{\sigma^2_{ob}}}{m_{ob}}\left(\overline{\boldsymbol{X}} - \gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}\right)^T \boldsymbol{\Sigma}^{-1}\left(\overline{\boldsymbol{X}} - \gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}\right)\right)\Bigg\}. \tag{A.7}$$

Rearranging the terms of (A.7) we get the following expression

$$p_{ob}(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{X}_1, \dots, \boldsymbol{X}_n) \propto |\boldsymbol{\Sigma}|^{-(d_{ob}+n)/2} exp\left\{-\frac{1}{2} tr\left[(\boldsymbol{S}_{ob} + (n-1)\boldsymbol{S})\boldsymbol{\Sigma}^{-1}\right]\right\}$$
$$\times\, exp\Bigg\{ -\frac{1}{2}\left(\frac{n \cdot \frac{s^2}{\sigma^2_{ob}}}{m_{ob}}\left(\overline{\boldsymbol{X}} - \gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}\right)^T \boldsymbol{\Sigma}^{-1}\left(\overline{\boldsymbol{X}} - \gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}\right)\right)\Bigg\}$$
$$\times\, |\boldsymbol{\Sigma}|^{-1/2} exp\Bigg\{ -\frac{m_{ob}}{2}\left(\boldsymbol{\mu} - \boldsymbol{r}_{ob}\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} - \boldsymbol{r}_{ob}\right)\Bigg\} \tag{A.8}$$

Hence, the conditional posterior distribution of $\boldsymbol{\mu}$ is given by

$$p_{ob}(\boldsymbol{\mu}|\boldsymbol{\Sigma}, \boldsymbol{X}_1, \dots, \boldsymbol{X}_n) \propto |\boldsymbol{\Sigma}|^{-1/2} exp\Bigg\{ -\frac{m_{ob}}{2}\left(\boldsymbol{\mu} - \boldsymbol{r}_{ob}\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} - \boldsymbol{r}_{ob}\right)\Bigg\}$$

which is the kernel of a multivariate normal distribution, that is

$$\boldsymbol{\mu}|\boldsymbol{\Sigma}, \boldsymbol{X}_1, \dots, \boldsymbol{X}_n \sim N(\boldsymbol{r}_{ob}, \frac{1}{m_{ob}}\boldsymbol{\Sigma}).$$

Integrating out $\boldsymbol{\mu}$ from (A.8) we get that the marginal posterior distribution of $\boldsymbol{\Sigma}$ is expressed as

$$p_{ob}(\boldsymbol{\Sigma}|\boldsymbol{X}_1, \dots, \boldsymbol{X}_n) \propto |\boldsymbol{\Sigma}|^{-(d_{ob}+n)/2} exp\left\{-\frac{1}{2} tr\left[(\boldsymbol{S}_{ob} + (n-1)\boldsymbol{S})\boldsymbol{\Sigma}^{-1}\right]\right\}$$
$$\times\, exp\Bigg\{ -\frac{1}{2}\left(\frac{n \cdot \frac{s^2}{\sigma^2_{ob}}}{m_{ob}}\left(\overline{\boldsymbol{X}} - \gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}\right)^T \boldsymbol{\Sigma}^{-1}\left(\overline{\boldsymbol{X}} - \gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}\right)\right)\Bigg\}.$$

A.3 Rejection sampling algorithm

The objective of the rejection sampling algorithm is to obtain a random draw from a distribution $f(\boldsymbol{X})$, which can be calculated up to a proportionality constant. This is done by sampling candidates from a distribution which we are able to obtain a random sample from, and then correcting the sampling probabilities through random rejection of some candidates. The distribution we want to draw random samples from is the marginal posterior distribution of $\boldsymbol{\Sigma}$, for the objective-based prior, which is expressed in equation (3.22), that is

$$\begin{aligned} f(\boldsymbol{X}) =& p_{ob}(\boldsymbol{\Sigma}|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \propto |\boldsymbol{\Sigma}|^{-(d_{ob}+n)/2} \\ & \times exp\left\{-\frac{1}{2}tr\left[(\boldsymbol{S}_{ob} + (n-1)\boldsymbol{S})\boldsymbol{\Sigma}^{-1}\right]\right\} \\ & \times exp\Bigg\{-\frac{1}{2}\left(\frac{n\cdot\frac{s^2}{\sigma^2_{ob}}}{m_{ob}}\left(\overline{\boldsymbol{X}} - \gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}\right)^T\boldsymbol{\Sigma}^{-1}\left(\overline{\boldsymbol{X}} - \gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}\right)\right)\Bigg\}. \end{aligned}$$

Using that

$$|\boldsymbol{\Sigma}|^{-(d_{ob}+n)/2} exp\left\{-\frac{1}{2}tr\left[(\boldsymbol{S}_{ob} + (n-1)\boldsymbol{S})\boldsymbol{\Sigma}^{-1}\right]\right\}$$

is the kernel of an inverse-Wishart distribution, and that

$$exp\Bigg\{-\frac{1}{2}\left(\frac{n\cdot\frac{s^2}{\sigma^2_{ob}}}{m_{ob}}\left(\overline{\boldsymbol{X}} - \gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}\right)^T\boldsymbol{\Sigma}^{-1}\left(\overline{\boldsymbol{X}} - \gamma\boldsymbol{\Sigma}\boldsymbol{w}_{ob}\right)\right)\Bigg\}$$

only takes on values between 0 and 1, we can draw a random sample from $f(\boldsymbol{X})$ as follows.

1. Sample $\boldsymbol{Y} \sim IW(d_{ob} + n, \boldsymbol{S}_{ob} + (n-1)\boldsymbol{S})$.
2. Sample $U \sim Unif(0, 1)$.

3. Reject $\boldsymbol{Y}$ if

$$U > exp\left\{ -\frac{1}{2}\left(\frac{n \cdot \frac{s^2(\mathbf{Y})}{\sigma_{ob}^2}}{m_{ob}} \left(\overline{\boldsymbol{X}} - \gamma \boldsymbol{Y}\boldsymbol{w}_{ob}\right)^T \boldsymbol{Y}^{-1} \left(\overline{\boldsymbol{X}} - \gamma \boldsymbol{Y}\boldsymbol{w}_{ob}\right)\right)\right\},$$

   where $s^2(\mathbf{Y}) = \frac{1}{k}tr(\boldsymbol{Y})$. In this case, we reject $\boldsymbol{Y}$ and return to step 1.

4. Otherwise keep $\boldsymbol{Y}$ and consider this a an element of the target random sample drawn from $f(\boldsymbol{X})$.

Bibliography

[1] E. W. Anderson and A.-R. Cheng, *Robust Bayesian Portfolio Choices*, Review of Financial Studies, 29 (2016), pp. 1330–1375.

[2] D. Avramov and G. Zhou, *Bayesian Portfolio Analysis*, Annual Review of Financial Economics, 2 (2010), pp. 25–47.

[3] C. Barry and B. Christopher, *Portfolio Analysis under Uncertain Means, Variances, and Covariances*, Journal of Finance, 29 (1974), pp. 515–522.

[4] G. K. Basak, R. Jagannathan, and R. Ma, *Estimating the risk in sample efficient portfolios*, PhD thesis, Northwestern University, Evanston, IL, 2005. Technical report.

[5] D. Bauder, T. Bodnar, N. Parolya, and W. Schmid, *Bayesian mean-variance analysis: Optimal portfolio selection under parameter uncertainty*, Quantitative Finance, (2021).

[6] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*, Wiley, 2000.

[7] F. Black and R. Litterman, *Global portfolio optimization*, Financial Analysts Journal, 48 (1992), pp. 28–43.

[8] O. Bodnar and T. Bodnar, *On the unbiased estimator of the efficient frontier*, International Journal of Theoretical and Applied Finance (IJTAF), 13 (2010), pp. 1065–1073.

[9] T. Bodnar, S. Dmytriv, N. Parolya, and W. Schmid, *Tests for the weights of the global minimum variance portfolio in a high-dimensional setting*, IEEE Transactions on Signal Processing, 67 (2017), pp. 4479–4493.

[10] T. Bodnar, S. Mazur, and Y. Okhrin, *Bayesian estimation of the global minimum variance portfolio*, European Journal of Operational Research, 256 (2017), pp. 292–307.

[11] T. Bodnar, N. Parolya, and W. Schmid, *Estimation of the global minimum variance portfolio in high dimensions*, European Journal of Operational Research, 266 (2017).

[12] T. Bodnar and W. Schmid, *Estimation of optimal portfolio compositions for Gaussian returns*, Statistics & Decisions, 26 (2008), pp. 179–201.

[13] T. Bodnar and W. Schmid, *Econometrical analysis of the sample efficient frontier*, European Journal of Finance, 15 (2009), pp. 317–335.

[14] M. W. Brandt, *Chapter 5 - portfolio choice problems*, in Handbook of Financial Econometrics: Tools and Techniques, vol. 1 of Handbooks in Finance, North-Holland, San Diego, 2010, pp. 269–336.

[15] M. Broadie, *Computing efficient frontiers using estimated parameters*, Annals of Operations Research, 45 (1993), pp. 21–58.

[16] S. Brown, *Optimal portfolio choice under uncertainty: A Bayesian approach*, University of Chicago Ph.D thesis, (1976).

[17] P. A. Frost and J. E. Savarino, *An Empirical Bayes Approach to Efficient Portfolio Selection*, Journal of Financial and Quantitative Analysis, 21 (1986), pp. 293–305.

[18] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, Chapman and Hall/CRC, 2nd ed. ed., 2004.

[19] G. H. Givens and J. A. Hoeting, *Computational statistics*, John Wiley & Sons, Hoboken, NJ, USA, 2 ed., 2012.

[20] A. Greyserman, D. Jones, and W. Strawderman, *Portfolio Selection Using Hierarchical Bayesian Analysis and MCMC Methods*, Journal of Banking & Finance, 30 (2006), pp. 669–678.

[21] A. Gupta and D. Nagar, *Matrix Variate Distributions*, Chapman and Hall/CRC, 2000.

[22] J. E. Ingersoll, *Theory of Financial Decision Making*, Rowman and Littlefield Publishing, 1987.

[23] D. Jobson and R. B. Korkie, *Performance hypothesis testing with the sharpe and treynor measures*, Journal of Finance, 36 (1981), pp. 889–908.

[24] R. Kan and D. Smith, *The Distribution of the Sample Minimum-Variance Frontier*, Management Science, 54 (2008), pp. 1364–1380.

[25] R. W. Klein and V. S. Bawa, *The effect of estimation risk on optimal portfolio choice*, Journal of Financial Economics, 3 (1976), pp. 215–231.

[26] H. Markowitz, *Portfolio selection*, Journal of Finance, 7 (1952), pp. 77–91.

[27] R. Merton, *An Analytic Derivation of the Efficient Portfolio Frontier*, Journal of Financial and Quantitative Analysis, 7 (1972).

[28] Y. Okhrin and W. Schmid, *Distributional properties of portfolio weights*, Journal of Econometrics, 134 (2006), pp. 235–256.

[29] S. J. Rachev, T. Svetlozar, C. Maté, J. S. J. Hsu, B. S. Bagasheva, and F. J. Fabozzi, *Bayesian Methods in Finance*, John Wiley and Sons, 02 2008.

[30] M. Sekerke, *Bayesian risk management: A guide to model risk and sequential learning in financial markets*, Wiley, Hoboken, NJ, 2015.

[31] A. Siegel and A. Woodgate, *Performance of portfolios optimized with estimation error*, Management Science, 53 (2007), pp. 1005–1015.

[32] R. F. Stambaugh, *Analyzing investments whose histories differ in length*, Journal of Financial Economics, 45 (1997), pp. 285–331.

[33] J. Tu and G. Zhou, *Incorporating Economic Objectives into Bayesian Priors: Portfolio Choice under Parameter Uncertainty*, Journal of Financial and Quantitative Analysis, 45 (2010), pp. 959–986.

[34] A. Zellner, *An introduction to Bayesian inference in econometrics*, John Wiley & Sons, New York, NY, 1971.