Bayesian portfolio selection and risk estimation

Vilhelm Niklasson


# Bayesian portfolio selection and risk estimation

Vilhelm Niklasson

Address: Matematiska Institutionen, Stockholms Universitet, 106 91 Stockholm
E-mail address: niklasson@math.su.se

Printed in Sweden by Eprint AB, Stockholm, 2022

Distributor: Department of Mathematics, Stockholm University

# Abstract

This thesis concerns portfolio theory from a Bayesian perspective and it includes two papers related to this theme. In the first paper, optimal portfolio weights are derived from a Bayesian perspective to the problem of minimizing the portfolio risk in terms of value at risk (VaR) or conditional value at risk (CVaR) given a certain level of expected return. The so called efficient frontiers, i.e., the set of optimal portfolios with the highest possible expected return given a certain level of VaR or CVaR, are also presented. Both a noninformative prior and an informative prior are considered and the resulting portfolios are compared to the ones obtained from the conventional method which is based on the sample estimates. Using simulated and empirical stock returns, we conclude that the Bayesian approach outperforms the conventional method in terms of out-of-sample VaR estimation when the global minimum VaR portfolio is considered. Moreover, we show within a simulation study that the efficient frontiers obtained from the Bayesian procedure are generally more conservative and closer to the true efficient frontier when the latter is known.

In the second paper, we consider the problem of determining VaR or CVaR of a portfolio when market conditions are quickly changing. When the market conditions are stable over time, it is possible to model future returns by using an equal belief in all historical observations. However, it is well known that volatility tends to cluster, particularly when considering daily or more frequent data observations. We use a Bayesian approach to create a model where the prior belief about returns resembles the recent period and where the degree of belief depends on how much the recent period deviates from the long term period. The new model is compared to some classical homoscedastic and heteroscedastc models in terms of VaR estimation using both simulated and empirical stock returns. We conclude that the new model performs well in the Basel backtest, particularly during turbulent market conditions where other models struggle.

# List of papers

The following papers, referred to in the text by their Roman numerals, are included in this thesis.

PAPER I: **Bayesian portfolio selection using VaR and CVaR**
Taras Bodnar, Mathias Lindholm, Vilhelm Niklasson, Erik Thorsén *Applied mathematics and computation*, 427. Advance online publication https://doi.org/10.1016/j.amc.2022.127120.

PAPER II: **Volatility sensitive Bayesian estimation of portfolio VaR and CVaR**
Taras Bodnar, Vilhelm Niklasson, Erik Thorsén *Working paper*.

---

**Author's contribution**: In the first paper, V. Niklasson was involved in deriving the theoretical results and writing the paper with all other authors. In the second paper, V. Niklasson contributed with constructing and analysing the model as well as writing the paper in collaboration with all other authors. V. Niklasson also carried out the computer calculations in both articles.

**Paper not included in the thesis**: V. Niklasson has also co-authored the paper 'Fast automatic deforestation detectors and their extensions for other spatial objects' together with Jesper Muren, Dmitry Otryakhin and Maxim Romashin. The paper has been submitted for publication.

# Contents

# Acknowledgements

First of all, I would like to thank my supervisor Taras Bodnar for coming up with the topic of this thesis and for all the rewarding discussions. I really appreciate that you encourage me to explore my own ideas and that you support me with your knowledge. I would also like to address a special thanks to Erik Thorsén who has has been a great co-author, friend and office mate. The discussions with you have always been very helpful. I would also like to thank my co-supervisor Mathias Lindholm for his clever ideas and all the rest of my colleagues at the Department of Mathematics at Stockholm University. Finally, I would like to thank my supporting family and Sofia.

# 1. Introduction

This chapter gives an introduction to the theory used in the papers of this thesis, a summary of the papers and some suggestions on future research. The major topics covered are portfolio theory, Bayesian statistics and their connection.

## 1.1 Portfolio theory

The foundation of portfolio theory was presented by Markowitz (1952) for which he received the Nobel memorial prize in economic sciences in 1990. He provided a general framework for determining efficient portfolios, i.e., optimal combinations of financial assets. Let $\boldsymbol{X}_t$ denote the $k$-dimensional vector of the asset returns at time $t$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Markowitz's idea was that an investor should use the portfolio weights $\boldsymbol{w}$ that minimize the portfolio variance $V_P(\boldsymbol{w}) = \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}$ given a certain level of expected portfolio return $R_P(\boldsymbol{w}) = \boldsymbol{w}^\top \boldsymbol{\mu}$. Mathematically, this mean-variance (MV) problem can be formulated as[1]

$$\underset{\boldsymbol{w}: R_P(\boldsymbol{w})=R_0,\, \boldsymbol{w}^\top \mathbf{1}=1}{\operatorname{argmin}} \quad V_P(\boldsymbol{w}), \tag{1.1.1}$$

where $R_0$ denotes the target return and $\boldsymbol{w}^\top \mathbf{1} = 1$ means that the investor is fully invested in the available assets, i.e., there is no risk-free asset.

The method of Lagrange multipliers can be used to solve the optimization problem (1.1.1) and its solution is often expressed in terms of the solution to the so called global minimum variance (GMV) portfolio (see, e.g., p. 53 in Ingersoll, 1987). The GMV portfolio is the efficient portfolio with the lowest possible variance and it is obtained by removing the constraint $R_P(\boldsymbol{w}) = R_0$ in (1.1.1). The GMV portfolio satisfies

$$\boldsymbol{w}_{GMV} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}}, \quad R_{GMV} = \frac{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}}, \quad V_{GMV} = \frac{1}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}}, \tag{1.1.2}$$

[1]Markowitz actually initially assumed that all the weights should be non-negative, but (1.1.1) shows a common generalization of the initial problem.

where $\boldsymbol{w}_{GMV}$, $R_{GMV}$ and $V_{GMV}$ denote the weights, expected return and variance of return, respectively, of the GMV portfolio. The solution to the general MV problem (1.1.1) is then given by

$$\boldsymbol{w}_{MV,R_0} = \boldsymbol{w}_{GMV} + \frac{R_0 - R_{GMV}}{V_{GMV}} \boldsymbol{Q}\boldsymbol{\mu}, \quad \boldsymbol{Q} = \boldsymbol{\Sigma}^{-1} - \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}}{\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}}. \tag{1.1.3}$$

The set of minimal achievable variances together with target expected returns for all $R_0 \geq R_{GMV}$ constitute the so called efficient frontier. This frontier shows the optimal trade-off betweeen return and risk (measured in terms of the variance) that an investor can achieve and it can be used as a benchmark and guidance for practitioners. In the mean-variance space, the efficient frontier corresponds to the upper part of a parabola which was shown by Merton (1972). The equation of the parabola is given by

$$(R - R_{GMV})^2 = s(V - V_{GMV}), \tag{1.1.4}$$

where $s = \boldsymbol{\mu}^\top \boldsymbol{Q}\boldsymbol{\mu}$ is the slope parameter. Figure 1.1 shows an example of the efficient frontier for 20 randomly selected stocks from S&P 500 by the end of 2020 based on sample estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

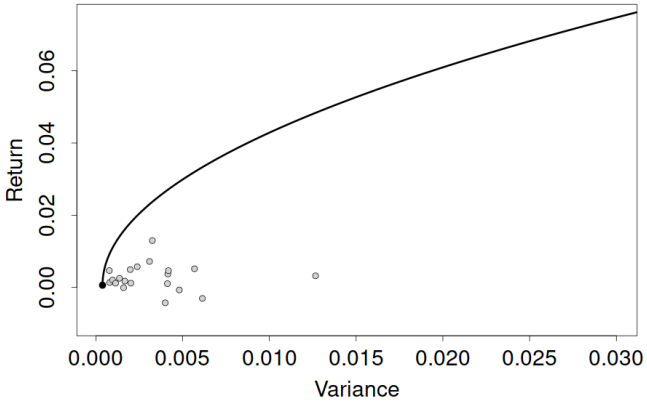


**Figure 1.1:** Mean-variance efficient frontier based on 20 stocks. The grey dots correspond to expected returns and variances of individual stocks and the black dot corresponds to the GMV portfolio.

The traditional MV-problem (1.1.1) is also commonly formulated in terms of a trade-off problem as (see, e.g., Okhrin and Schmid, 2006)

$$\underset{\boldsymbol{w}:\boldsymbol{w}^\top\mathbf{1}=1}{\operatorname{argmax}} R_P(\boldsymbol{w}) - \frac{\gamma}{2} V_P(\boldsymbol{w}), \tag{1.1.5}$$

where $\gamma > 0$ is the trade-off coefficient. Varying $\gamma$ in (1.1.5) will have similar effect as varying $R_0$ in (1.1.1). For each $\gamma$ there exists a corresponding value of $R_0$ that will yield the same optimal solution. The solution to (1.1.5) can be expressed as

$$\boldsymbol{w}_{MV,\gamma} = \boldsymbol{w}_{GMV} + \gamma^{-1}\boldsymbol{Q}\boldsymbol{\mu}. \tag{1.1.6}$$

During the last decades it has become popular to use other measures of risk than the variance, which was traditionally used in Markovitz framework. This has been motivated by new regulations, the inability of the variance to capture tail risk and its lack of other desirable properties. Alexander and Baptista (2002, 2004) studied the impact of using value at risk (VaR) and conditional value at risk (CVaR) in the portfolio optimization problem. These are risk measures that depend on the tail (loss) behaviour of the return distribution (see, e.g., p. 207 in Föllmer and Schied, 2011).

**Definition 1.1.1** (VaR and CVaR)**.** Let $X_t$ be a return distribution at time $t$ and set $Y_t = -X_t$. VaR at level $\alpha \in (0,1)$ is the $\alpha$ quantile of $Y_t$, i.e.,

$$\mathrm{VaR}_\alpha(X_t) = -\inf\{x \in \mathbb{R} : F_X(x) > 1-\alpha\} = F_Y^{-1}(\alpha) \tag{1.1.7}$$

and CVaR at level $\alpha$ is the average loss when VaR at level $\alpha$ is exceeded, i.e.,

$$\mathrm{CVaR}_\alpha(X_t) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_\gamma(X_t)\,d\gamma. \tag{1.1.8}$$

Alexander and Baptista (2002, 2004) formulated the portfolio mean-VaR and mean-CVaR optimization problems at level $\alpha$ as

$$\underset{\boldsymbol{w}:\, R_P(\boldsymbol{w})=R_0,\, \boldsymbol{w}^\top \mathbf{1}=1}{\operatorname{argmin}} \quad \mathrm{VaR}_{P,\alpha}(\boldsymbol{w}), \tag{1.1.9}$$

and

$$\underset{\boldsymbol{w}:\, R_P(\boldsymbol{w})=R_0,\, \boldsymbol{w}^\top \mathbf{1}=1}{\operatorname{argmin}} \quad \mathrm{CVaR}_{P,\alpha}(\boldsymbol{w}), \tag{1.1.10}$$

where $\mathrm{VaR}_{P,\alpha}(\boldsymbol{w})$ and $\mathrm{CVaR}_{P,\alpha}(\boldsymbol{w})$ denote the portfolio VaR and CVaR at level $\alpha$, respectively. They showed that a portfolio which is mean-VaR or mean-CVaR efficient will also be mean-variance efficient under the assumption of normally distributed returns, but the reverse is not always true. Moreover, they proved that the global minimum VaR and CVaR portfolios will only exist if $\alpha$ is large enough.

In practise when doing portfolio optimization it is necessary to estimate the unknown parameters of the return distribution. Unfortunately, this turns out to be a difficult task and the common plug-in procedure where the true parameters are replaced by sample estimates will often result in poorly

performing portfolios (see, e.g., Frankfurter, Phillips, and Seagle, 1971). To mitigate this issue, a wide variety of solutions have been suggested. A popular approach is the shrinkage method by Ledoit and Wolf (2004) in which the sample covariance matrix is "shrunk" towards some target matrix. It is also possible to use robust portfolio optimization procedures where the worst case scenarios for the estimates are considered (see, e.g., Fabozzi, Kolm, Pachamanova, and Focardi, 2007). Another solution, which is the topic of this thesis, is to use Bayesian statistics.

## 1.2 Bayesian statistics

Bayesian statistics is a branch of statistics where the unknown parameters of a model are considered to be random variables and where prior belief is used together with observed data to make statistical conclusions. Let $\boldsymbol{\Theta}$ denote the parameters of a model and denote the observed data by $\boldsymbol{x}$. The fundamental building block of Bayesian statistics is Bayes' rule which states that (see, e.g., p. 7 in Gelman, Carlin, Stern, and Rubin, 2013)

$$f(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{f(\boldsymbol{\theta})f(\boldsymbol{x}|\boldsymbol{\theta})}{f(\boldsymbol{x})}, \tag{1.2.1}$$

where $f(\boldsymbol{\theta})$ and $f(\boldsymbol{\theta}|\boldsymbol{x})$ denote the prior and posterior distributions of $\boldsymbol{\Theta}$, respectively, and $f(\boldsymbol{x}|\boldsymbol{\theta})$ denotes the likelihood. The posterior distribution of $\boldsymbol{\Theta}$ can be considered the updated belief about the model parameters after having observed $\boldsymbol{X} = \boldsymbol{x}$. Note that the denominator in (1.2.1) does not depend on $\boldsymbol{\Theta}$ and hence it is common to consider the unnormalized version of the posterior distribution of $\boldsymbol{\Theta}$, namely

$$f(\boldsymbol{\theta}|\boldsymbol{x}) \propto f(\boldsymbol{\theta})f(\boldsymbol{x}|\boldsymbol{\theta}). \tag{1.2.2}$$

The posterior distribution of $\boldsymbol{\Theta}$ can be used to make conclusions about the model parameters, including their uncertainty. Inference about future (unobserved) values $\tilde{\boldsymbol{x}}$ can also be made from a Bayesian perspective by using

$$f(\tilde{\boldsymbol{x}}|\boldsymbol{x}) = \int f(\tilde{\boldsymbol{x}}, \boldsymbol{\theta}|\boldsymbol{x})d\boldsymbol{\theta} = \int f(\tilde{\boldsymbol{x}}|\boldsymbol{\theta})f(\boldsymbol{\theta}|\boldsymbol{x})d\boldsymbol{\theta}, \tag{1.2.3}$$

where $f(\tilde{\boldsymbol{x}}|\boldsymbol{x})$ is referred to as the posterior predictive distribution of $\tilde{\boldsymbol{X}}$. This distribution can, for instance, be used to make point predictions of future observations by using its mean or mode.

Based on the above, it is clear that Bayesian inference has two crucial steps. First, the probability distributions of the data and parameters must be

determined. Such choices are often based on prior knowledge of the underlying scientific problem. Second, the posterior distributions are calculated based on observed data. This potentially means that a multi-dimensional integral must be evaluated and numerical methods such as Markov chain Monte Carlo (MCMC) have been developed for this purpose (see, Hastings, 1970; Metropolis et al., 1953).

Sometimes it is possible to choose the involved probability distributions so that the prior distribution of $\boldsymbol{\Theta}$ becomes (in some sense) noninformative or it may be possible to specify them so that the posterior distributions become easier to evaluate. A well known noninformative prior distribution of $\boldsymbol{\Theta}$ is the Jeffreys prior. This prior distribution is proportional to the square root of the determinant of the Fisher information matrix $I(\boldsymbol{\theta})$ which is defined as

$$I(\boldsymbol{\theta}) = -E\left[\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\log\big(f(\boldsymbol{x}|\boldsymbol{\theta})\big)\,|\boldsymbol{\theta}\right]. \tag{1.2.4}$$

This prior has the key feature that it is invariant to parameter transformation. This means that for a parameterization $\boldsymbol{\theta}$ and an alternative parameterization $\boldsymbol{\phi}$, where it is assumed that $\boldsymbol{\theta}$ is a continuously differentiable function of $\boldsymbol{\phi}$, it holds that

$$f(\boldsymbol{\phi}) = f(\boldsymbol{\theta})|\det(\nabla\boldsymbol{g})| \propto \sqrt{\det(I(\boldsymbol{\phi}))}, \tag{1.2.5}$$

where $\nabla\boldsymbol{g}$ denotes the Jacobian matrix of $\boldsymbol{\theta} = \boldsymbol{g}(\boldsymbol{\phi})$. Hence we can go from the Jeffreys prior using one parameterization to another parameterization by just multiplying by the absolute value of the determinant of the Jacobian matrix.

Another commonly used prior distribution of $\boldsymbol{\Theta}$ is the conjugate prior. This prior distribution has the property that the posterior distribution of $\boldsymbol{\Theta}$ belongs to the same probability distribution family as the prior distribution. Hence, whether or not a prior distribution is considered to be a conjugate prior will depend on the likelihood function. For some likelihood functions it is possible to simply determine a conjugate prior, for instance when the likelihood function is a multivariate normal distribution.

**Definition 1.2.1** (Multivariate normal distribution)**.** The $k$-dimensional random vector $\boldsymbol{X}$ follows a multivariate normal distribution with $k$-dimensional mean vector $\boldsymbol{\mu}$ and symmetric positive definite $k \times k$ covariance matrix $\boldsymbol{\Sigma}$ if its probability density function is given by (see, e.g., p. 579 in Gelman, Carlin, Stern, and Rubin, 2013)

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{k}{2}}\det(\boldsymbol{\Sigma})^{\frac{1}{2}}}e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}. \tag{1.2.6}$$

This distribution is denoted $N_k(\boldsymbol{\mu},\boldsymbol{\Sigma})$.

In the case of a multivariate normal likelihood function with unknown mean and covariance matrix, the conjugate prior is a normal-inverse-Wishart distribution (see, e.g., p. 73 in Gelman, Carlin, Stern, and Rubin, 2013). This distribution is a composition of the multivariate normal distribution and the inverse-Wishart distribution where the latter can be derived from the inverse of a Wishart distribution. The Wishart distributon with $n$ degrees of freedom is the distribution of the sample covariance matrix of the multivariate normal distribution with know mean vector based on $n$ observations.

**Definition 1.2.2** (Wishart distribution and inverse Wishart distribution)**.** The $k \times k$ random matrix $\boldsymbol{X}$ follows a Wishart distribution with $n$ degrees of freedom and symmetric positive definite $k \times k$ scale matrix $\boldsymbol{V}$ if its probability density function is given by (see, e.g., p. 579 in Gelman, Carlin, Stern, and Rubin, 2013)

$$f(\boldsymbol{x}) = \frac{1}{2^{\frac{np}{2}} \det(\boldsymbol{V})^{\frac{n}{2}} \boldsymbol{\Gamma}_k\left(\frac{n}{2}\right)} \det(\boldsymbol{x})^{\frac{n-k-1}{2}} e^{-\frac{1}{2}\operatorname{tr}(\boldsymbol{V}^{-1}\boldsymbol{x})}, \tag{1.2.7}$$

where $\boldsymbol{\Gamma}_k$ is the $k$-dimensional gamma function defined as

$$\boldsymbol{\Gamma}_k\left(\frac{n}{2}\right) = \pi^{\frac{k(k-1)}{4}} \prod_{j=1}^{k} \Gamma\left(\frac{n}{2} - \frac{j-1}{2}\right), \tag{1.2.8}$$

and $\Gamma$ is the gamma function which is the common generalization of the factorial function. A random matrix follows an inverse Wishart distribution if its inverse follows a Wishart distribution. The Wishart distribution is denoted $W_k(\boldsymbol{V}, n)$ and the inverse Wishart distribution is denoted $IW_k(\boldsymbol{V}^{-1}, n)$.

**Definition 1.2.3** (Normal-inverse-Wishart distribution)**.** The composition $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu}$ is a $k$-dimensional random vector and $\boldsymbol{\Sigma}$ is a symmetric positive definite $k \times k$ random matrix follows a normal-inverse-Wishart distribution with parameters $\boldsymbol{m}_0$, $r_0$, $\boldsymbol{S}_0$ and $d_0$ if (see, e.g., p. 73 in Gelman, Carlin, Stern, and Rubin, 2013)

$$\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim N_k(\boldsymbol{m}_0, \frac{1}{r_0}\boldsymbol{\Sigma}), \quad \text{and} \quad \boldsymbol{\Sigma} \sim IW_k(\boldsymbol{S}_0, d_0), \tag{1.2.9}$$

and its probability density function is the product of the involved multivariate normal distribution and inverse-Wishart distribution. This distribution is denoted $NIW(\boldsymbol{m}_0, r_0, \boldsymbol{S}_0, d_0)$.

The parameters $\boldsymbol{m}_0, r_0, \boldsymbol{S}_0, d_0$ in the normal-inverse-Wishart prior distribution can be used to control the prior belief about the unknown $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$

if the likelihood function is a multivariate normal distribution. The parameters $\boldsymbol{m}_0, r_0, \boldsymbol{S}_0, d_0$ are commonly referred to as hyperparameters since they impact the distribution of parameters on a higher level. More precisely, we can think that the prior belief about the mean corresponds to $r_0$ observations with sample mean $\boldsymbol{m}_0$ and the prior belief about the covariance matrix is based on $d_0$ observations with sum of pairwise deviation products equal to $\boldsymbol{S}_0$. Practitioners may have opinions about these hyperparameters based on some specific knowledge, otherwise it is possible to specify them by using the empirical Bayes method or a full hierarchical Bayesian model. The empirical Bayes method is based on the idea to specify the hyperparameters to their most likely values based on the observed data whereas a full hierarchical Bayesian model also imposes distributions on the hyperparameters.

## 1.3 Bayesian portfolio theory

Bayesian statistics has been applied in the filed of portfolio theory since around the 1960s with pioneering works by Bawa, Brown, and Klein (1979), Winkler (1973), Winkler and Barry (1975), and Zellner and Chetty (1965). The Bayesian approach has potentially many advantages, for instance it makes it easy to specify prior beliefs and model uncertainty. An extensive survey of Bayesian portfolio analysis was made by Avramov and Zhou (2010). In the early applications, noninformative priors or data-based priors were mainly used. Later on, more involved methods have been developed to specify priors. A popular such method was presented by Black and Litterman (1992) in their somewhat informal Bayesian portfolio allocation method. They assume that expected returns are difficult to model and that an investor with no particular view should hold the market portfolio, i.e., a value weighted portfolio, and 'backed out' the expected return from those weights. An investor with a relative option could then adjust this expected return in the prior. Other authors, such as Pástor, 2000 and Pástor and Stambaugh (2000), have used asset pricing models to specify prior beliefs.

In general, the optimal portfolio weights in a Bayesian setup are obtained by using the posterior predictive distribution. As previously, let $\boldsymbol{X}_t$ denote the $k$-dimensional vector of asset returns at time $t$ and let $\boldsymbol{x}_{(t-1)} = (\boldsymbol{x}_{t-n}, \ldots, \boldsymbol{x}_{t-1})$ be the observation matrix, i.e., realizations of $\boldsymbol{X}_{t-n}, \ldots, \boldsymbol{X}_{t-1}$ between $t-n$ and $t-1$, with probability density parameters $\boldsymbol{\theta}$. Moreover, let $X_{P,t} = \boldsymbol{w}^\top \boldsymbol{X}_t$ be the portfolio return at time $t$ given the weights $\boldsymbol{w}$ and denote its posterior predictive mean $E[X_{P,t}|\boldsymbol{x}_{(t-1)}]$ by $R_{P,t}(\boldsymbol{w})$ and its posterior predictive variance $\mathrm{Var}(X_{P,t}|\boldsymbol{x}_{(t-1)})$ by $V_{P,t}(\boldsymbol{w})$. In the traditional Bayesian

mean-variance setting, the optimal portfolio weights are obtained from

$$\underset{\boldsymbol{w}:\, R_{P,t}(\boldsymbol{w})=R_0,\, \boldsymbol{w}^\top \mathbf{1}=1}{\operatorname{argmin}} \quad V_{P,t}(\boldsymbol{w}), \tag{1.3.1}$$

where $R_0$ is the expected target return. Using the trade-off formulation, the problem can also be expressed as

$$\underset{\boldsymbol{w}:\boldsymbol{w}^\top \mathbf{1}=1}{\operatorname{argmax}} R_{P,t}(\boldsymbol{w}) - \frac{\gamma}{2} V_{P,t}(\boldsymbol{w}), \tag{1.3.2}$$

where $\gamma > 0$ is the trade-off coefficient. The difference compared to the conventional problems (1.1.1) and (1.1.5) may seem insignificant, but unlike the conventional method the Bayesian approach automatically accounts for estimation uncertainty by using the posterior predictive distribution.

Under the assumption that assets returns are independent and identically distributed (i.i.d.) and follow a multivariate normal distribution when conditioning on the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, Bauder, Bodnar, Parolya, and Schmid (2021) derived the optimal portfolio weights and efficient frontiers using both the Jeffreys prior and the conjugate normal-inverse-Wishart prior. Let

$$\bar{\boldsymbol{x}}_{t-1} = \frac{1}{n} \sum_{i=t-n}^{t-1} \boldsymbol{x}_i, \quad \boldsymbol{S}_{t-1} = \sum_{i=t-n}^{t-1} (\boldsymbol{x}_i - \bar{\boldsymbol{x}}_{t-1})(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_{t-1})^\top, \tag{1.3.3}$$

$$c_{k,n} = \frac{n+1}{n(n-k-2)}. \tag{1.3.4}$$

Using the Jeffreys prior and assuming $n > k$, the optimal portfolio weights that solve (1.3.2) are

$$\boldsymbol{w}_{MV,\gamma} = \frac{\boldsymbol{S}_{t-1}^{-1}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{S}_{t-1}^{\top}\mathbf{1}} + \gamma^{-1} c_{k,n}^{-1} \boldsymbol{Q}_{t-1} \bar{\boldsymbol{x}}_{t-1} \tag{1.3.5}$$

with

$$\boldsymbol{Q}_{t-1} = \boldsymbol{S}_{t-1}^{-1} - \frac{\boldsymbol{S}_{t-1}^{-1}\mathbf{1}\mathbf{1}^\top \boldsymbol{S}_{t-1}^{-1}}{\mathbf{1}^\top \boldsymbol{S}_{t-1}^{-1}\mathbf{1}}. \tag{1.3.6}$$

Moreover, the efficient frontier satisfies

$$(R - R_{GMV,J})^2 = \frac{\bar{\boldsymbol{x}}_{t-1}^\top \boldsymbol{Q}_{t-1} \bar{\boldsymbol{x}}_{t-1}}{c_{k,n}} (V - V_{GMV,J}), \tag{1.3.7}$$

with

$$R_{GMV,J} = \frac{\mathbf{1}^\top \boldsymbol{S}_{t-1}^{-1} \bar{\boldsymbol{x}}_{t-1}}{\mathbf{1}^\top \boldsymbol{S}_{t-1}^{-1}\mathbf{1}}, \quad V_{GMV,J} = \frac{c_{k,n}}{\mathbf{1}^\top \boldsymbol{S}_{t-1}^{-1}\mathbf{1}}. \tag{1.3.8}$$

Bauder, Bodnar, Parolya, and Schmid (2021) also showed that similar results hold for the normal-inverse-Wishart conjugate prior, but in that case the hyperparameters will impact the results.

In general, the effect of using the Jeffreys prior instead of the sample estimates is that the minimum achievable variance will be larger for a certain level of expected return. The difference becomes particularly pronounced when $k$ is close to $n$, in which case $c_{k,n}$ will become very large and increase the variance of the GMV portfolio. The conjugate prior has the effect of shrinking the sample mean towards some prior mean and also impacting the covariance matrix towards what is specified by the prior hyperparameters.

Figure 1.2 shows how the conventional efficient frontier based on the sample estimates compares to the efficient frontier using the Jeffreys and conjugate prior based on the empirical Bayes method when the same setup as in Figure 1.1 is used. As can be seen, the Bayesian mean-variance efficient frontiers are located below the conventional efficient frontier which is known to be too optimistic (see, e.g., Siegel and Woodgate, 2007).

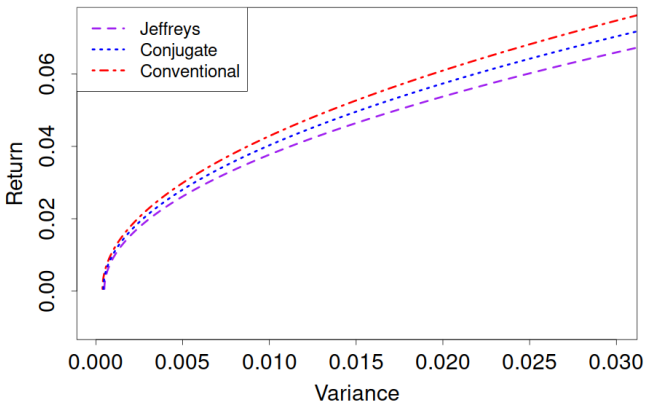


**Figure 1.2:** Mean-variance efficient frontiers based on 20 stocks using different estimation methods.

The papers in this thesis use the same setup as Bauder, Bodnar, Parolya, and Schmid (2021) but VaR and CVaR are used as risk measures instead of the variance. This means that we extend the results of Alexander and Baptista (2002, 2004) to a Bayesian framework where the parameter uncertainty is taken into account. Moreover, we develop and evaluate a new way to specify the the hyperparameters in the conjugate prior which makes it possible to quickly adapt to changing market conditions. Such changes occur from time to time, not least recently due to the Covid-19 outbreak.

## 1.4 Summary of papers

Below follows a summary of the papers in the thesis.

### 1.4.1. Paper I

In the first paper, we solve the mean-VaR and mean-CVaR portfolio optimization problems from a Bayesian perspective using the posterior predictive distribution of future portfolio returns. To do this, we first derive expressions of the posterior predictive portfolio distribution using the Jeffreys prior and the conjugate prior, assuming that the asset returns are multivariate normally distributed when conditioning on the mean vector and covariance matrix. We show that this posterior predictive distribution is a $t$-distribution with degrees of freedom depending on the number of assets, observations and potentially some hyperparameter. This is an important result which extends and simplifies the theoretical contribution in Bauder, Bodnar, Parolya, and Schmid (2021). We then derive the portfolio VaR and CVaR at time $t$ and show that they can be expressed on the common form

$$Q_{t-1}(\boldsymbol{w}) = -\boldsymbol{w}^\top \bar{\boldsymbol{x}}_{t-1} + q_\alpha \sqrt{r_{k,n}} \sqrt{\boldsymbol{w}^\top \boldsymbol{S}_{t-1} \boldsymbol{w}}, \tag{1.4.1}$$

where $\bar{\boldsymbol{x}}_{t-1}$, $\boldsymbol{S}_{t-1}$ and $r_{k,n}$ depend on the prior and historical observations until time $t-1$, $q_\alpha$ depends on if VaR or CVaR is considered and $\boldsymbol{w}$ denotes the portfolio weights. Next we solve the optimization problem

$$\underset{\boldsymbol{w}: R_{t-1}(\boldsymbol{w})=R_0,\, \boldsymbol{w}^\top \mathbf{1}=1}{\operatorname{argmin}} \quad Q_{t-1}(\boldsymbol{w}), \tag{1.4.2}$$

where $R_{t-1}(\boldsymbol{w})$ denotes the target expected portfolio return given the information until time $t-1$. We also determine the weights of the global minimum VaR and CVaR portfolios and derive expressions of the mean-VaR and mean-CVaR efficient frontiers and show that these correspond to the upper part of a hyperbola.

Simulated and empirical data are used to illustrate how the new method performs in comparison with the conventional plug-in method based on the sample estimates. We show that the Bayesian procedure typically results in better VaR estimates and its mean-VaR efficient frontiers are less overoptimistic and closer to the true (population) efficient frontier during simulation. Figure 1.3 shows an example of the mean-VaR efficient frontier for $\alpha = 0.99$ based on empirical returns of 20 randomly selected stocks in the S&P 500 index by the end of 2020.

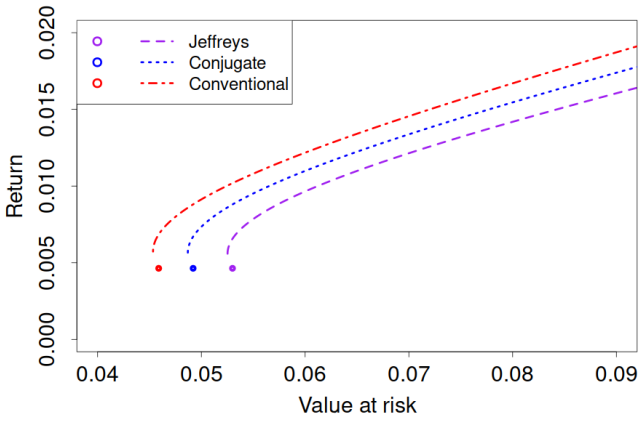


**Figure 1.3:** Mean-VaR efficient frontiers based on 20 stocks using different estimation methods. The locations of the GMV portfolios are marked by circles.

### 1.4.2. Paper II

In the second paper, we use a similar setup as in Paper I but we develop and evaluate a new method to specify the hyperparameters $\boldsymbol{S}_0$ and $d_0$ in the normal-inverse-Wishart conjugate prior. The aim is to specify them so that we capture volatility clustering, i.e., periods of high and low volatility (standard deviation) of asset returns. We do this by specifying $\boldsymbol{S}_0$ and $d_0$ based on two different rolling window periods. More precisely, $\boldsymbol{S}_0$, which corresponds to our prior belief about the covariances, is set to be similar to the recent period, and $d_0$, which can be thought of as our degree of certainty in the prior, will depend on how much the recent portfolio variance deviates from the long term variance. The new volatility sensitive method is referred to as VS$(\cdot,\cdot,\cdot)$ where the first parameter corresponds to the length of the recent rolling window period and the second and third parameters control how much the degree of certainty should increase if the recent period is a high or low volatility period, respectively.

We evaluate the new method by using the Basel backtest of VaR and we compare it to other well known methods, including the conjugate method based on empirical Bayes (EB) and the DCC-GARCH method which is a heteroscedastic model. Using simulated and empirical data, we show that the new method performs very well, particularly during turbulent market periods. This is illustrated in Figure 1.4 where VaR is estimated for $\alpha = 0.99$ on a daily basis for an equally weighted portfolio of size 10 during 2020. We observe that many of the standard estimation methods have too many VaR exceedances, especially around March when the volatility was very high. Our new method seems to be able to capture such rapid changes better.

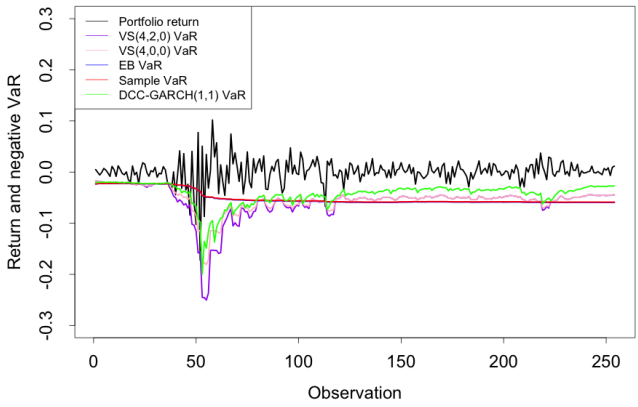


**Figure 1.4:** Returns and negative VaR estimates at the 99 % level for a portfolio of size 10 in 2020 using different estimation methods.

## 1.5 Future research

Bayesian statistics makes it easy to account for parameter uncertainty and include prior information. These are attractive features, not least in portfolio theory. Even if a lot of research already has been made in this field, there are still plenty of research topics that have not been explored in detail.

In the future it would be of interest to study more how to include prior information in a Bayesian stock return model by using different sources of information. For instance, the VIX index could be a good indicator of future volatility and the trade volume could also contain information which is not directly observed in historical returns. Being able to include such information in a model could potentially have great benefits compared to a model that only uses stock returns as a source of information.

It would also be of interest to consider more complicated distributions than the ones used in the papers of this thesis, for instance skewed distributions. Markov Chain Monte Carlo (MCMC) is a great tool which makes it easy to sample from posterior distributions even when they are difficult to evaluate theoretically. These kind of methods have already been used to model financial data (see, e.g., Chib and Greenberg, 1996), but possibly not in connection with other interesting topics such as robustness analysis.

Finally, Bayesian time series models, such as Bayesian structural time series (Scott and Varian, 2014), have shown good potential. How to construct such models in a multivariate setting that allows for fast portfolio optimization is a challenging task that could be explored more in the future.

# Bibliography


Alexander, Gordon J and Alexandre M Baptista (2002). "Economic implications of using a mean-VaR model for portfolio selection: A comparison with mean-variance analysis". In: *Journal of Economic Dynamics and Control* 26.7, pp. 1159–1193.

Alexander, Gordon J and Alexandre M Baptista (2004). "A comparison of VaR and CVaR constraints on portfolio selection with the mean-variance model". In: *Management Science* 50.9, pp. 1261–1273.

Avramov, Doron and Guofu Zhou (2010). "Bayesian portfolio analysis". In: *The Annual Review of Financial Economics* 2.1, pp. 25–47.

Bauder, David, Taras Bodnar, Nestor Parolya, and Wolfgang Schmid (2021). "Bayesian mean–variance analysis: Optimal portfolio selection under parameter uncertainty". In: *Quantitative Finance* 21, pp. 221–242.

Bawa, Vijay S, Stephen J Brown, and Roger W Klein (1979). *Estimation risk and optimal portfolio choice*. North-Holland.

Black, Fischer and Robert Litterman (1992). "Global portfolio optimization". In: *Financial Analysts Journal* 48, pp. 28–43.

Chib, Siddhartha and Edward Greenberg (1996). "Markov chain Monte Carlo simulation methods in econometrics". In: *Econometric theory* 12.3, pp. 409–431.

Fabozzi, Frank J, Petter N Kolm, Dessislava A Pachamanova, and Sergio M Focardi (2007). "Robust portfolio optimization". In: *The Journal of portfolio management* 33.3, pp. 40–48.

Frankfurter, George M, Herbert E Phillips, and John P Seagle (1971). "Portfolio selection: the effects of uncertain means, variances, and covariances". In: *Journal of Financial and Quantitative Analysis* 6.5, pp. 1251–1262.

Föllmer, Hans and Alexander Schied (2011). *Stochastic Finance: An Introduction in Discrete Time*. Walter de Gruyter.

Gelman, Andrew, John B Carlin, Hal S Stern, and Donald B Rubin (2013). *Bayesian data analysis*. Chapman and Hall/CRC.

Hastings, W Keith (1970). "Monte Carlo sampling methods using Markov chains and their applications". In.

Ingersoll, Jonathan E (1987). *Theory of financial decision making*. Vol. 3. Rowman & Littlefield.

Ledoit, Olivier and Michael Wolf (2004). "Honey, I shrunk the sample covariance matrix". In: *The Journal of Portfolio Management* 30.4, pp. 110–119.

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

Merton, Robert C (1972). "An analytic derivation of the efficient Portfolio frontier". In: *The Journal of Financial and Quantitive Analysis* 7.4, pp. 1851–1872.

Metropolis, Nicholas et al. (1953). "Equation of state calculations by fast computing machines". In: *The journal of chemical physics* 21.6, pp. 1087–1092.

Okhrin, Yarema and Wolfgang Schmid (2006). "Distributional properties of portfolio weights". In: *Journal of Econometrics* 134.1, pp. 235–256.

Pástor, L'uboš (2000). "Portfolio selection and asset pricing models". In: *The Journal of Finance* 55.1, pp. 179–223.

Pástor, L'uboš and Robert F Stambaugh (2000). "Comparing asset pricing models: an investment perspective". In: *Journal of Financial Economics* 56.3, pp. 335–381.

Scott, Steven L and Hal R Varian (2014). "Predicting the present with Bayesian structural time series". In: *International Journal of Mathematical Modelling and Numerical Optimisation* 5.1-2, pp. 4–23.

Siegel, Andrew F and Artemiza Woodgate (2007). "Performance of portfolios optimized with estimation error". In: *Management Science* 53, pp. 1005–1015.

Winkler, Robert L (1973). "Bayesian models for forecasting future security prices". In: *Journal of Financial and Quantitative Analysis*, pp. 387–405.

Winkler, Robert L and Christopher B Barry (1975). "A Bayesian model for portfolio selection and revision". In: *The Journal of Finance* 30.1, pp. 179–192.

Zellner, Arnold and V Karuppan Chetty (1965). "Prediction and decision problems in regression models from the Bayesian point of view". In: *Journal of the American Statistical Association* 60.310, pp. 608–616.

Matematiska Institutionen, Stockholms Universitet, 106 91 Stockholm
E-mail address: niklasson@math.su.se