



Mathematical Statistics  
Stockholm University

## **Bayesian Quantile-Based Portfolio Selection**

Taras Bodnar  
Mathias Lindholm  
Vilhelm Niklasson  
Erik Thorsén

**Research Report 2020:11**

ISSN 1650-0377

**Postal address:**

Mathematical Statistics  
Dept. of Mathematics  
Stockholm University  
SE-106 91 Stockholm  
Sweden

**Internet:**

<http://www.math.su.se>



Mathematical Statistics  
Stockholm University  
Research Report **2020:11**,  
<http://www.math.su.se>

# Bayesian Quantile-Based Portfolio Selection

Taras Bodnar, Mathias Lindholm,

Vilhelm Niklasson and Erik Thorsén

December 2020

## Abstract

We study the optimal portfolio allocation problem from a Bayesian perspective using value at risk (VaR) and conditional value at risk (CVaR) as risk measures. By applying the posterior predictive distribution for the future portfolio return, we derive relevant quantiles needed in the computations of VaR and CVaR, and express the optimal portfolio weights in terms of observed data only. This is in contrast to the conventional method where the optimal solution is based on unobserved quantities which are estimated, leading to sub-optimality. We also obtain the expressions for the weights of the global minimum VaR and CVaR portfolios, and specify conditions for their existence. It is shown that these portfolios may not exist if the confidence level used for the VaR or CVaR computation are too low. Moreover, analytical expressions for the mean-VaR and mean-CVaR efficient frontiers are presented and the extension of theoretical results to general coherent risk measures is provided. One of the main advantages of the suggested Bayesian approach is that the theoretical results are derived in the finite-sample case and thus they are exact and can be applied to large-dimensional portfolios.

By using simulation and real market data, we compare the new Bayesian approach to the conventional method by studying the performance and existence of the global minimum VaR portfolio and by analysing the estimated efficient frontiers. It is concluded that the Bayesian approach outperforms the conventional one, in particular at predicting the out-of-sample VaR.

# 1 Introduction

The economic theory underlying optimal portfolio selection was pioneered by Markowitz (1952). In his seminal paper, the optimal allocation of the available assets was determined by their expected returns and the covariance matrix of the asset returns. In practise, however, both the vector of expected returns and the covariance matrix are unknown and have to be estimated by using historical data. These estimates were traditionally treated as the true parameters of the data-generating process and plugged into the equations for the weights of optimal portfolios. Unfortunately, this causes a challenging problem since the optimal portfolio weights appear to be sensitive to misspecification of the input parameters, especially the expected returns, and thus estimation errors can lead to poorly performing portfolios (see, Chopra and Ziemba 1993, Frankfurter et al. 1971, Klein and Bawa 1976, Merton 1980, Simaan 2014, Bodnar et al. 2018).

The problem concerning parameter uncertainty was one of the reasons why the Bayesian approach was introduced in portfolio theory during the 1970s (Winkler and Barry 1975). In the Bayesian setting, the parameters of the distribution of the asset returns are modelled as random variables with a prior distribution summarizing the information about the asset returns which is not present in the historical data. The posterior distribution, derived from the likelihood function and the used priors, provides updated knowledge on the model parameters conditionally on the observed data. In addition to reducing estimation risk, a Bayesian setting also makes it possible to employ useful prior information and implement fast numerical algorithms (see, Avramov and Zhou 2010, Rachev et al. 2008). Several different ways of specifying a prior have been suggested in the literature (see, e.g., Avramov and Zhou 2010, Bodnar et al. 2017, Tu and Zhou 2010). In general, a prior can be either informative or non-informative. Early applications of Bayesian statistics in finance were mainly based on uninformative or model-based priors (Bawa et al. 1979). However, during the 80s and 90s, some more sophisticated priors were developed. Two such highly influential priors are the hyperparameter prior by Jorion (1986), which was inspired by the Bayes-Stein shrinkage approach, and the informative prior by Black and Litterman (1992), which relies on market equilibrium arguments.

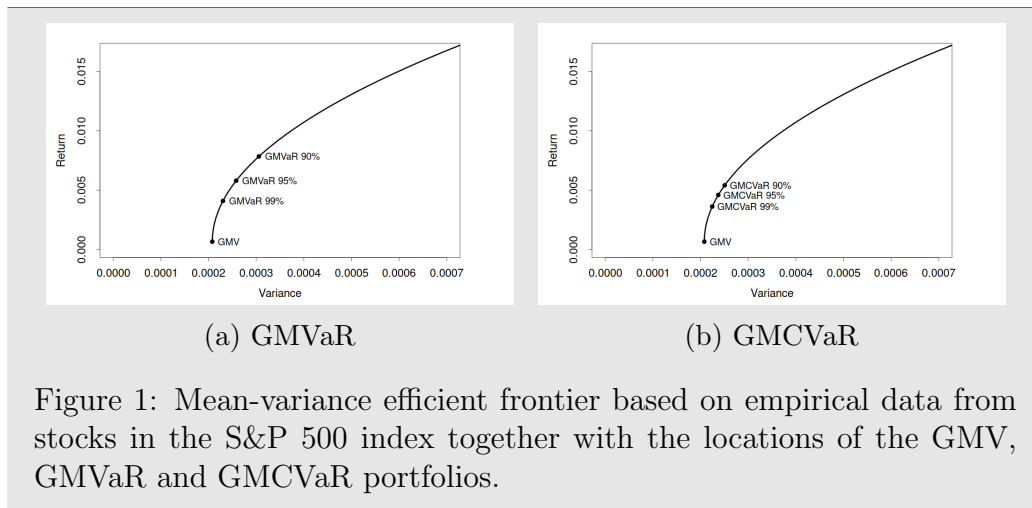
In addition to specifying a prior, an assumption must be made about the return distribution. Such a modelling choice is crucial both in the frequentist and in the Bayesian setup. The standard assumption of unconditional normality has been heavily criticized since actual market returns tend to have a higher density both close to the mean and far away from it (Cont 2001). Because of this, more general assumptions allowing for heavier tails have

been suggested (see, e.g., Adcock 2014, Bauder et al. 2020b). For instance, Bauder et al. (2020b) studied a Bayesian setting where they used the general assumption that the asset returns are exchangeable and multivariate centered spherically symmetric (see also, Brugière 2020). They derived a stochastic representation of the posterior predictive distribution, i.e., the distribution of the next return conditioned on the previous returns, under this assumption. By using this representation in the portfolio optimization problem, the common suboptimality issue in the standard plug-in approach was avoided since the estimates were no longer treated as the true parameters. Using a stochastic representation is a well-established tool in computational statistics (Givens and Hoeting 2013) and it reduces the need for demanding Markov Chain Monte Carlo simulations. Moreover, a stochastic representation makes it easier to find quantiles which can be used to measure the risk.

In the standard Markowitz setup, the risk is measured in terms of the variance of the portfolio return (Markowitz 1952). Recent regulations have forced banks and insurance companies to use other measures of risk, such as value at risk (VaR) and conditional value at risk (CVaR). The former is still in use in the Solvency II directive while the latter is enforced by the Basel III and IV standards. Both VaR and CVaR are quantile-based risk measures focusing on downside risk, meaning that the risk is determined from a quantile in the right tail of a loss distribution. A lot of attention has been given to these two risk measures over the last decades concerning their applicability, computations and back-testing (see, e.g., Baumol 1963, Jorion 1997, Pritsker 1997, Meng and Taylor 2020, Staino and Russo 2020). The two measures differ substantially in the way the quantiles are selected and it is usually argued that CVaR is a better measure of risk in comparison to VaR since it does not violate the desirable property of subadditivity (Artzner et al. 1999).

The usage of VaR and CVaR in portfolio optimization has also become increasingly popular over the last decades. For instance, Rockafellar and Uryasev (2000) and Babat et al. (2018) studied the minimization of CVaR in a portfolio using linear programming. Alexander and Baptista (2002, 2004) derived the weights of a portfolio when using VaR or CVaR in the objective function in the portfolio optimization problem under the assumption of multivariate normally distributed asset returns. They noted that the optimal allocation strategy does not depend on whether the variance, VaR or CVaR is used as a risk measure. However, the portfolio that globally minimizes VaR or CVaR may not coincide with the one which globally minimizes the variance. This is illustrated in Figure 1 where the mean-variance efficient frontier is plotted together with the locations of the global minimum variance (GMV), global minimum VaR (GMVaR) and global minimum CVaR (GM-

CVaR) portfolios based on weekly returns for 50 randomly selected stocks in the S&P 500 index. In this figure, the efficient frontier and the optimal portfolios are all calculated using the conventional method, i.e., using sample estimates and treating them as the true parameters of the data-generating model. The fact that the GMVaR and GMCVaR portfolios are located above the GMV portfolio on the mean-variance efficient frontier motivates the study of these portfolios also for an investor who constructs the portfolio following the mean-variance analysis, since their expected returns determine the lower bounds for an investor who wants to be efficient also from a mean-VaR or mean-CVaR perspective. Given the new regulations, this should be highly desirable.



This paper contributes to the current literature on portfolio theory by combining a Bayesian framework and a quantile-based asset allocation strategy. This gives new insights into how an optimal portfolio can be constructed in practice. We further compare the new Bayesian approach to its frequentist counterpart and demonstrate some of the advantages of the former through a simulation study and an empirical illustration. Throughout the report, we consider two different well-known priors (one informative and one non-informative) and we use general (non-normal) assumptions on the asset returns as Bauder et al. (2020b) and Brugière (2020).

The rest of the paper is organized as follows. In Section 2, we present a Bayesian model of portfolio and asset returns and discuss its time series properties. Moreover, the posterior predictive distribution is derived together with stochastic representations related to it. In Section 3, we use the obtained stochastic representations to establish expressions for VaR and CVaR of a

portfolio and we also extend the analysis to more general risk measures. Section 4 deals with portfolio optimization under parameter uncertainty. The Bayesian quantile-based portfolio optimization problem is solved and conditions for the existence of a global minimum VaR and CVaR portfolio are presented. The mean-VaR and mean-CVaR efficient frontiers are also derived. Section 5 contains a simulation study where our Bayesian approach is compared to the conventional method by analysing the performance and existence of the global minimum VaR portfolio as well as by comparing the estimated efficient frontiers. The comparison is continued in Section 6 with an empirical study based on market data from stocks in the S&P 500 index. Section 7 contains a conclusion and discussion based on our findings. Finally, all of the proofs are moved to the appendix together with a description of the data.

## 2 Bayesian model of portfolio and asset returns

In this section, we introduce our Bayesian model and derive the posterior predictive distributions of the portfolio return using two different priors. We also make a time series interpretation of the model used to describe the stochastic properties of the asset returns.

### 2.1 Posterior predictive distribution of portfolio return

Let  $\mathbf{X}_t$  be a  $k$ -dimensional vector of logarithmic asset returns at time  $t$ , i.e., each element  $X_{t,i}$ ,  $i = 1, \dots, k$  of  $\mathbf{X}_t$  is defined as

$$X_{t,i} := \log \left( \frac{P_{t,i}}{P_{t-1,i}} \right),$$

where  $P_{t,i}$  denotes the price at time  $t$  of asset  $i$ . Throughout the paper we assume that the asset returns  $\mathbf{X}_1, \mathbf{X}_2, \dots$  are infinitely exchangeable and multivariate centered spherically symmetric (see, Bernardo and Smith 2009, for the definition and properties). This assumption in particular implies that the asset returns are neither normally nor independently distributed. Moreover, let  $\mathbf{x}_{(t-1)} = (\mathbf{x}_{t-n}, \dots, \mathbf{x}_{t-1})$  be the observation matrix of the asset returns  $\mathbf{x}_{t-n}, \dots, \mathbf{x}_{t-1}$  taken from time  $t-n$  until  $t-1$  whose distribution depends on the parameter vector  $\boldsymbol{\theta}$ . Bayes theorem provides the posterior

distribution of  $\boldsymbol{\theta}$  expressed as

$$\pi(\boldsymbol{\theta}|\mathbf{x}_{(t-1)}) \propto f(\mathbf{x}_{(t-1)}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}),$$

where  $\pi(\cdot)$  is the prior distribution and  $f(\cdot|\boldsymbol{\theta})$  is the likelihood function.

At time  $t$  the return of the portfolio with weights  $\mathbf{w} = (w_1, \dots, w_k)$ , where it is assumed that  $\mathbf{1}^\top \mathbf{w} = 1$ , is given by

$$X_{P,t} := \mathbf{w}^\top \mathbf{X}_t.$$

The posterior predictive distribution of  $X_{P,t}$ , i.e., the conditional distribution of  $X_{P,t}$  given  $\mathbf{x}_{(t-1)}$ , is computed by (see, Bernardo and Smith 2009)

$$f(x_{P,t}|\mathbf{x}_{(t-1)}) = \int_{\boldsymbol{\theta} \in \Theta} f(x_{P,t}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{x}_{(t-1)})d\boldsymbol{\theta}, \quad (2.1)$$

where  $\Theta$  denotes the parameter space and  $f(\cdot|\boldsymbol{\theta})$  is the conditional density of  $X_{P,t}$  given  $\boldsymbol{\theta}$ . The posterior predictive distribution (2.1) determines the distribution of the future portfolio return at time  $t$  given the information available in the historical data of asset returns up to  $t - 1$ . It can be used to construct a point prediction of the future portfolio return, such as the posterior predictive mean or mode, together with the uncertainty, expressed as the posterior predictive variance. Furthermore, a prediction interval for the portfolio return can be obtained as a posterior predictive credible interval.

The integration in (2.1) makes it, usually, impossible to find an analytical expression for the posterior predictive distribution. In such cases, the point prediction of the future portfolio return together with its uncertainty can be obtained by using the Markov Chain Monte Carlo method (see, e.g., Krüger et al. 2020). Moreover, Zellner and Ando (2010) argued that a direct Monte Carlo approach provides an efficient computational method to compute a Bayesian estimation and to determine its uncertainty. Another way to deal with the integration issue is by directly simulating from the posterior predictive distribution by using a stochastic representation (see, e.g., Bauder et al. 2020b).

Employing the non-informative Jeffreys prior and the informative conjugate prior, Bauder et al. (2020b) derived a stochastic representation of a random variable following the posterior predictive distribution (2.1) under the assumptions of infinitely exchangeability and multivariate centered spherical symmetry. In this case, the model parameters are the mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$  of the asset returns (see, Section 2.2 for the detailed discussion). The Jeffreys prior and the conjugate prior are given by

$$\pi_J(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(k+1)/2}, \quad (2.2)$$



and

$$\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim N_k\left(\mathbf{m}_0, \frac{1}{r_0}\boldsymbol{\Sigma}\right), \quad \text{and} \quad \boldsymbol{\Sigma} \sim IW_k(d_0, \mathbf{S}_0), \quad (2.3)$$

respectively, where  $|\cdot|$  stands for the determinant and  $IW_k(d_0, \mathbf{S}_0)$  denotes the inverse Wishart distribution with  $d_0$  degrees of freedom and parameter matrix  $\mathbf{S}_0$  (see, Gupta and Nagar 2000, for the definition and properties). The quantities  $\mathbf{m}_0$ ,  $r_0$ ,  $d_0$ , and  $\mathbf{S}_0$  are the hyperparameters of the conjugate prior where  $\mathbf{m}_0$  and  $\mathbf{S}_0$  reflect the investor's prior belief about the mean vector and covariance matrix, whereas  $r_0$  and  $d_0$  represent the precision of these beliefs. The two priors (2.2) and (2.3) have been widely used in financial literature where the Jeffreys prior is usually also referred to as the diffuse prior (see, e.g., Barry 1974, Bodnar et al. 2017, Brown 1976, Stambaugh 1997), and the conjugate prior is also commonly related to the Black-Litterman model (see, Bauder et al. 2020b, Black and Litterman 1992, Frost and Savarino 1986, Kolm and Ritter 2017).

The stochastic representations derived by Bauder et al. (2020b) fully determine the posterior predictive distribution. They are expressed in terms of two independent standard  $t$ -distributed random variables whose degrees of freedom depend on the assigned prior. Using these findings, the posterior predictive distribution is deduced and is presented in Theorem 2.1. Let  $t(q, a, b^2)$  denote the univariate  $t$ -distribution with  $q$  degrees of freedom, location parameter  $a$  and scale parameter  $b$ . Then, we obtain the following results.

**Theorem 2.1.** *Let asset returns  $\mathbf{X}_1, \mathbf{X}_2, \dots$  be infinitely exchangeable and multivariate centered spherically symmetric. Then*

1. *under the Jeffreys prior, it holds for  $n > k$  that the posterior predictive distribution defined in (2.1) is  $t(d_{k,n,J}, \mathbf{w}^\top \bar{\mathbf{x}}_{t-1,J}, r_{k,n,J} \mathbf{w}^\top \mathbf{S}_{t-1,J} \mathbf{w})$  with  $d_{k,n,J} = n - k$ ,*

$$r_{k,n,J} = \frac{n+1}{n(n-k)}, \quad \bar{\mathbf{x}}_{t-1,J} = \frac{1}{n} \sum_{i=t-n}^{t-1} \mathbf{x}_i, \quad (2.4)$$

and

$$\mathbf{S}_{t-1,J} = \sum_{i=t-n}^{t-1} (\mathbf{x}_i - \bar{\mathbf{x}}_{t-1,J})(\mathbf{x}_i - \bar{\mathbf{x}}_{t-1,J})^\top; \quad (2.5)$$

2. *under the conjugate prior, it holds for  $n + d_0 - 2k > 0$  that the posterior predictive distribution defined in (2.1) is  $t(d_{k,n,C}, \mathbf{w}^\top \bar{\mathbf{x}}_{t-1,C}, r_{k,n,C} \mathbf{w}^\top \mathbf{S}_{t-1,C} \mathbf{w})$  with  $d_{k,n,C} = n + d_0 - 2k$ ,*

$$r_{k,n,C} = \frac{n+r_0+1}{(n+r_0)(n+d_0-2k)}, \quad \bar{\mathbf{x}}_{t-1,C} = \frac{n\bar{\mathbf{x}}_{t-1,J} + r_0\mathbf{m}_0}{n+r_0}, \quad (2.6)$$

and

$$\mathbf{S}_{t-1,C} = \mathbf{S}_{t-1,J} + \mathbf{S}_0 + nr_0 \frac{(\mathbf{m}_0 - \bar{\mathbf{x}}_{t-1,C})(\mathbf{m}_0 - \bar{\mathbf{x}}_{t-1,C})^\top}{n + r_0}. \quad (2.7)$$

The proof of Theorem 2.1 is given in the appendix.

Since the structure of the posterior predictive distributions are similar, we introduce new notations which allow us to combine the two findings of Theorem 2.1 into a single result. Let

$$\begin{cases} d_{k,n} = d_{k,n,J}, r_{k,n} = r_{k,n,J}, \bar{\mathbf{x}}_{t-1} = \bar{\mathbf{x}}_{t-1,J}, \mathbf{S}_{t-1} = \mathbf{S}_{t-1,J}, \\ d_{k,n} = d_{k,n,C}, r_{k,n} = r_{k,n,C}, \bar{\mathbf{x}}_{t-1} = \bar{\mathbf{x}}_{t-1,C}, \mathbf{S}_{t-1} = \mathbf{S}_{t-1,C}, \end{cases} \quad (2.8)$$

where the first row corresponds to the Jeffreys prior and the second row corresponds to the conjugate prior. From Theorem 2.1 we deduce that the posterior predictive distribution under both priors can be expressed as  $t(d_{k,n}, \mathbf{w}^\top \bar{\mathbf{x}}_{t-1}, r_{k,n} \mathbf{w}^\top \mathbf{S}_{t-1} \mathbf{w})$  with  $d_{k,n}$ ,  $r_{k,n}$ ,  $\bar{\mathbf{x}}_{t-1}$ , and  $\mathbf{S}_{t-1}$  as in (2.8). Let  $\widehat{X}_{P,t}$  denote a random variable which follows the posterior predictive distribution, i.e., whose distribution coincides with the conditional distribution of portfolio return  $X_{P,t}$  given the information available up to time  $t-1$ . Application of Theorem 2.1 together with (2.8) then gives the following stochastic representation of  $\widehat{X}_{P,t}$

$$\widehat{X}_{P,t} \stackrel{d}{=} \mathbf{w}^\top \bar{\mathbf{x}}_{t-1} + \tau \sqrt{r_{k,n}} \sqrt{\mathbf{w}^\top \mathbf{S}_{t-1} \mathbf{w}}, \quad (2.9)$$

where  $\tau \sim t(d_{k,n})$  with  $t(d_{k,n})$  denoting the standard  $t$ -distribution with  $d_{k,n}$  degrees of freedom, i.e., the  $t$ -distribution with zero location parameter and scale parameter equal to one. Representation (2.9) immediately implies that

$$\mathbb{E}(\widehat{X}_{P,t}) = \mathbb{E}(X_{P,t} | \mathbf{x}_{(t-1)}) = \mathbf{w}^\top \bar{\mathbf{x}}_{t-1}, \quad (2.10)$$

$$\text{Var}(\widehat{X}_{P,t}) = \text{Var}(X_{P,t} | \mathbf{x}_{(t-1)}) = \frac{d_{k,n} r_{k,n}}{d_{k,n} - 2} \mathbf{w}^\top \mathbf{S}_{t-1} \mathbf{w}, \quad (2.11)$$

for  $d_{k,n} > 1$  and  $d_{k,n} > 2$ , respectively.

## 2.2 Time series representation of asset returns

In the financial literature it is common to describe asset returns in terms of a time series model (see, e.g., Tsay 2010). In order to make comparisons to such models easier, we now give a time series representation of the considered Bayesian model.

The assumption of infinite exchangeability implies following De Finetti's theorem that there exists a probability measure conditionally on which the

asset returns are independent and identically distributed (see, Kingman et al. 1978). Moreover, the assumption of multivariate centered spherically symmetry determines the conditional distribution of the asset returns. Namely, we get that the asset returns conditionally on the mean vector  $\boldsymbol{\mu}$  and on the covariance matrix  $\boldsymbol{\Sigma}$  are independent and normally distributed (see, Proposition 4.6 in Bernardo and Smith 2009). That is

$$\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_t | \boldsymbol{\mu}, \boldsymbol{\Sigma} \text{ are independent with } \mathbf{X}_i | \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \quad (2.12)$$

Model (2.12) should not be confused with the assumption that the asset returns are independent and normally distributed as it is usually presented in financial and econometric literature. The two models coincide only when  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  possess probability distributions concentrated in single points. This would imply that the investor knows  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  with certainty, which is not the case in practice.

In order to show the considerable difference between the model (2.12) and the model assuming independent and normally distributed asset returns, we next derive the marginal distribution of  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_t$  by using the findings of Theorem 2.1 and the stochastic representation (2.9). The results will be obtained by assigning the Jeffreys prior to the model parameters  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  as well as the conjugate prior.

In Theorem 2.1 we showed that given a sample of size  $n > k$ , the posterior predictive distribution of  $X_{P,t} = \mathbf{w}^\top \mathbf{X}_t$  given  $\mathbf{X}_{t-n}, \dots, \mathbf{X}_{t-1}$  has a univariate  $t$ -distribution under both priors (see, also the stochastic representation (2.9)). Since  $\mathbf{w}$  is an arbitrary vector with the only restriction  $\mathbf{1}^\top \mathbf{w} = 1$ , choosing  $n = t - 1$  we get

$$\mathbf{X}_t | \mathbf{X}_1, \dots, \mathbf{X}_{t-1} \sim t_k(d_{k,t-1}, \bar{\mathbf{x}}_{1:t-1}, r_{k,t-1} \mathbf{S}_{1:t-1}), \quad (2.13)$$

and, similarly for  $j = k + 2, \dots, t - 1$  it holds that

$$\mathbf{X}_{j+1} | \mathbf{X}_1, \dots, \mathbf{X}_j \sim t_k(d_{k,j}, \bar{\mathbf{x}}_{1:j}, r_{k,j} \mathbf{S}_{1:j}), \quad (2.14)$$

where  $d_{k,j}$  and  $r_{k,j}$  for  $j = k + 2, \dots, t - 1$  are defined as in (2.8);  $\bar{\mathbf{x}}_{1:j}$  and  $\mathbf{S}_{1:j}$  are defined similarly to  $\bar{\mathbf{x}}_{t-1}$  and  $\mathbf{S}_{t-1}$ , where the sums in (2.4) and (2.5) are from 1 to  $j$ . The symbol  $t_k(\cdot, \cdot, \cdot)$  stands for the  $k$ -dimensional multivariate  $t$ -distribution.

The last result can also be written in the time series context in the following way

$$\mathbf{X}_{j+1} = \bar{\mathbf{x}}_{1:j} + \sqrt{r_{k,j}} \mathbf{S}_{1:j}^{1/2} \boldsymbol{\epsilon}_j \quad \text{for } j = k + 2, \dots, t - 1, \quad (2.15)$$

where  $\varepsilon_j \sim t_k(d_{k,j}, \mathbf{0}, \mathbf{I})$  is the error term and  $\mathbf{S}_{1:j}^{1/2}$  denotes a square root of  $\mathbf{S}_{1:j}$ . Finally, we get that the joint density of  $\mathbf{X}_1, \dots, \mathbf{X}_{k+2}$  is given by (see Appendix B in Bauder et al. 2020a)

$$f(\mathbf{x}_1, \dots, \mathbf{x}_{k+2}) \propto |\mathbf{S}_{1:k+2}|^{-\frac{d_{k,k+2}+k-1}{2}}. \quad (2.16)$$

The derived time series model for  $\mathbf{X}_1, \dots, \mathbf{X}_t$  indicates that a complicated nonlinear dependence structure is present in the conditional model (2.12). Moreover, model (2.12) can be used to capture the time-dependent structure in the dynamics of both the conditional mean vector and the conditional covariance matrix of the asset returns which is usually observed in practice (see, e.g., Tsay 2010). Finally, the error terms in the time series model for  $\mathbf{X}_1, \dots, \mathbf{X}_t$  in (2.15) are multivariate  $t$ -distributed with degrees of freedom depending on the amount of available information. This gives rise to an additional source of non-stationarity and allows for heavier tails than in a model where asset returns are assumed to be independent and normally distributed corresponding to the conventional approach.

### 3 Quantile-based risk measures

The quantile based risk-measures value at risk (VaR) and conditional value at risk (CVaR) are introduced in the following section and we derive their analytical expressions in our Bayesian setting using the posterior predictive distribution. Moreover, we extend the analysis to more general risk measures.

#### 3.1 Posterior predictive VaR and posterior predictive CVaR

The results of Theorem 2.1 provides an easy way to sample from the posterior predictive distribution as well as to find its quantiles. The two most common quantile-based risk measures used in the literature are VaR and CVaR which we define in this section using the posterior predictive distribution of the portfolio return.

Denoting by  $\widehat{X}_{P,t}$  a random variable whose distribution coincides with the posterior predictive distribution of the portfolio return  $X_{P,t}$  given  $\mathbf{x}_{(t-1)}$  and using that the posterior predictive distribution is absolutely continuous, the two quantile-based risk measures at confidence level  $\alpha \in (0.5, 1)$  are defined by

$$\text{VaR}_{\alpha,t-1}(\widehat{X}_{P,t}) := F_{Y,t-1}^{-1}(\alpha)$$

and

$$\text{CVaR}_{\alpha,t-1}(\widehat{X}_{P,t}) := \mathbb{E}[Y|Y \geq \text{VaR}_{\alpha,t-1}(\widehat{X}_{P,t})],$$

respectively, where  $Y := -\widehat{X}_{P,t}$  is the portfolio loss with cumulative distribution function  $F_{Y,t-1}(\cdot)$ .

In the following, it is implicit that the probability and expectation in the definitions of the VaR and CVaR are formulated in terms of the posterior predictive distribution and they are conditioned on previous asset returns  $\mathbf{x}_{(t-1)}$ . Moreover, the posterior predictive distribution is free of parameters and is fully determined by the observed data. This constitutes the main advantage of the suggested approach, namely it takes the parameter uncertainty into account before the optimal portfolio is constructed, i.e., the optimisation problem is formulated and solved. We discuss this point in detail in Section 4.

By definition,  $\text{VaR}_{\alpha,t-1}(\widehat{X}_{P,t})$  satisfies

$$P\left(\widehat{X}_{P,t} \leq -\text{VaR}_{\alpha,t-1}(\widehat{X}_{P,t})\right) = 1 - \alpha. \quad (3.1)$$

Rewriting the left hand side of (3.1) yields

$$\begin{aligned} P\left(\mathbf{w}^\top \bar{\mathbf{x}}_{t-1} + \tau \sqrt{r_{k,n}} \sqrt{\mathbf{w}^\top \mathbf{S}_{t-1} \mathbf{w}} \leq -\text{VaR}_{\alpha,t-1}(\widehat{X}_{P,t})\right) \\ = P\left(\tau \leq \frac{-\text{VaR}_{\alpha,t-1}(\widehat{X}_{P,t}) - \mathbf{w}^\top \bar{\mathbf{x}}_{t-1}}{\sqrt{r_{k,n}} \sqrt{\mathbf{w}^\top \mathbf{S}_{t-1} \mathbf{w}}}\right), \end{aligned}$$

where  $\tau$  is standard  $t$ -distributed with degrees of freedom  $d_{k,n}$  defined in (2.8) depending on the prior. Hence,

$$\frac{-\text{VaR}_{\alpha,t-1}(\widehat{X}_{P,t}) - \mathbf{w}^\top \bar{\mathbf{x}}_{t-1}}{\sqrt{r_{k,n}} \sqrt{\mathbf{w}^\top \mathbf{S}_{t-1} \mathbf{w}}} = d_{1-\alpha},$$

where  $d_{1-\alpha}$  is the  $(1 - \alpha)$  quantile of the  $t$ -distribution with  $d_{k,n}$  degrees of freedom which satisfies  $d_\alpha = -d_{1-\alpha}$  due to the symmetry of the  $t$ -distribution. Thus,

$$\text{VaR}_{\alpha,t-1}(\widehat{X}_{P,t}) = -\mathbf{w}^\top \bar{\mathbf{x}}_{t-1} + d_\alpha \sqrt{r_{k,n}} \sqrt{\mathbf{w}^\top \mathbf{S}_{t-1} \mathbf{w}}. \quad (3.2)$$

Similarly, it follows by the definition of CVaR that

$$\begin{aligned} \text{CVaR}_{\alpha,t-1}(\widehat{X}_{P,t}) &= \mathbb{E}\left(-\widehat{X}_{P,t} | -\widehat{X}_{P,t} \geq \text{VaR}_{\alpha,t-1}(\widehat{X}_{P,t})\right) \\ &= -\mathbf{w}^\top \bar{\mathbf{x}}_{t-1} + k_\alpha \sqrt{r_{k,n}} \sqrt{\mathbf{w}^\top \mathbf{S}_{t-1} \mathbf{w}}, \end{aligned} \quad (3.3)$$

with

$$\begin{aligned} k_\alpha &= \mathbb{E}(-\tau | -\tau \geq d_\alpha) = \frac{1}{1-\alpha} \int_{d_\alpha}^{\infty} t f_{d_{k,n}}(t) dt \\ &= \frac{1}{1-\alpha} \frac{\Gamma\left(\frac{d_{k,n}+1}{2}\right)}{\Gamma\left(\frac{d_{k,n}}{2}\right)} \frac{d_{k,n}}{\sqrt{\pi d_{k,n}}} \frac{1}{d_{k,n}-1} \left(1 + \frac{d_\alpha^2}{d_{k,n}}\right)^{-\frac{d_{k,n}-1}{2}}, \end{aligned}$$

where  $f_{d_{k,n}}(t)$  denotes the density of the  $t$ -distribution with  $d_{k,n}$  degrees of freedom and we use that the distribution of  $-\tau$  coincides with  $\tau$  due to the symmetry of the  $t$ -distribution.

The expressions for VaR and CVaR given in (3.2) and (3.3), respectively, can be presented in the following form

$$Q_{t-1}(\mathbf{w}) = -\mathbf{w}^\top \bar{\mathbf{x}}_{t-1} + q_\alpha \sqrt{r_{k,n}} \sqrt{\mathbf{w}^\top \mathbf{S}_{t-1} \mathbf{w}}, \quad (3.4)$$

where  $q_\alpha = d_\alpha$  when considering VaR and  $q_\alpha = k_\alpha$  when considering CVaR. This general formulation will be used extensively in the rest of the paper in order to handle VaR and CVaR cases simultaneously.

Since  $\alpha \in (0.5, 1)$  and the  $t$ -distribution is symmetric, we get that  $d_\alpha > 0$ . Moreover, we have that  $k_\alpha > 0$  by definition. As a result,  $q_\alpha > 0$  which together with the convexity of  $\sqrt{\mathbf{w}^\top \mathbf{S}_{t-1} \mathbf{w}}$  implies that

**Theorem 3.1.** *Under the conditions of Theorem 2.1,  $Q_{t-1}(\mathbf{w})$  is convex with respect to  $\mathbf{w}$ .*

The proof of Theorem 3.1 is given in the appendix.

## 3.2 Extension to general risk measures

The stochastic representation in equation (2.9) shows that the general presentation of the posterior predictive VaR and of the posterior predictive CVaR as given in (3.4) can be extended to other risk measures used in portfolio theory and risk management. Such a result will provide a possibility to formulate and to solve the portfolio choice problem under more general setups considered in financial mathematics which is based on the context of coherent risk functionals (see, e.g., Artzner et al. 1999). Namely, instead of considering the VaR and the CVaR to specify the portfolio risk, one can choose any risk measure which is relevant, translation invariant and positive homogeneous. These properties are all present in the definition of a coherent risk measure. For a risk functional  $\rho$ , this can be formulated as

- (a) *Relevance:* For all  $X_{P,t} \leq 0$  we have  $\rho(X_{P,t}) \geq 0$ .

(b) *Translation invariance*: For any real scalar  $a$ ,  $\rho(X_{P,t} + a) = \rho(X_{P,t}) - a$ .

(c) *Positive homogeneity*: For any positive scalar  $\lambda$ ,  $\rho(\lambda X_{P,t}) = \lambda \rho(X_{P,t})$ .

The relevance property provides us with a foundation for the interpretation of risk measures. From an investor's point of view, a risk measure cannot assign a negative risk to a certain loss. That is often, if not always, how we think about risk. The property of translation invariance states that if one adds more cash to the portfolio or invest a larger proportion of the capital in something that is deemed as risk-free, then the risk-exposure should decrease by the corresponding amount. The property of positive homogeneity can be thought of as using leverage to acquire another (possibly more aggressive) position in the market. Such a position will increase the risk accordingly.

When the asset returns  $\mathbf{X}_1, \mathbf{X}_2, \dots$  are infinitely exchangeable and multivariate centered spherically symmetric, then the analytical expression of (posterior) general risk measures is deduced from (2.9) and it is given by

$$\begin{aligned} \rho_{t-1}(\widehat{X}_{P,t}) &= \rho_{t-1}(\mathbf{w}^\top \bar{\mathbf{x}}_{t-1} + \tau \sqrt{r_{k,n}} \sqrt{\mathbf{w}^\top \mathbf{S}_{t-1} \mathbf{w}}) \\ &= -\mathbf{w}^\top \bar{\mathbf{x}}_{t-1} + \rho_{t-1}(\tau) \sqrt{r_{k,n}} \sqrt{\mathbf{w}^\top \mathbf{S}_{t-1} \mathbf{w}}, \end{aligned} \quad (3.5)$$

which coincides with (3.4) when we set  $\rho_{t-1}(\tau) = q_\alpha$ . To this end, we note that  $\rho_{t-1}(\tau)$  does not depend on the portfolio weights  $\mathbf{w}$ . As a result, finding the optimal portfolio by optimizing (3.5) is the same problem as finding the optimal portfolio based on (3.4). Since the conditions of relevance, translation invariance and positive homogeneity are present in the definition of a coherent risk measure, any (posterior) coherent risk measure should satisfy (3.5) and, consequently, the results of the next section can be applied.

## 4 Portfolio optimization

We now turn to the theory related to portfolio optimization and present the solutions of the quantile-based optimal portfolio choice problems from both the Bayesian and the frequentist points of view.

First, we describe existent results related to the conventional portfolio optimization problems. This is followed by theory related to the Bayesian quantile-based approach which relies on some previous findings. Special focus is put on the global minimum VaR and global minimum CVaR portfolios. The section ends with a derivation of the efficient frontier in the mean-quantile space.

## 4.1 Portfolio optimization problems and existing solutions

The mean-variance optimization problem of Markowitz (1952) and its solution provide a fundamental concept to practical asset allocation. It was originally formulated in terms of the population parameters of the asset return distribution: mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ . As a result, the solution of Markowitz's optimization problem depends on these unknown quantities which have to be estimated before the implementation in practice.

The population expected portfolio return of the portfolio with weights  $\mathbf{w}$  and its population variance are given by  $R_P(\mathbf{w}) = \mathbf{w}^\top \boldsymbol{\mu}$  and  $V_P(\mathbf{w}) = \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}$ , respectively. Then Markowitz's optimization problem is given by

$$\min_{\mathbf{w}: R_P(\mathbf{w})=R_0, \mathbf{w}^\top \mathbf{1}=\mathbf{1}} V_P(\mathbf{w}), \quad (4.1)$$

where  $\mathbf{1}$  denotes the  $k$ -dimensional vector of ones. The solution of (4.1) is expressed as

$$\mathbf{w}_{MV} = \mathbf{w}_{GMV} + \frac{R_0 - R_{GMV}}{s} \mathbf{M} \boldsymbol{\mu} \quad (4.2)$$

with

$$\mathbf{M} = \boldsymbol{\Sigma}^{-1} - \frac{\boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{1}^\top \boldsymbol{\Sigma}^{-1}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}}, \quad (4.3)$$

and where

$$\mathbf{w}_{GMV} = \frac{\boldsymbol{\Sigma}^{-1} \mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}} \quad (4.4)$$

are the weights of the global minimum variance (GMV) portfolio whose population expected return and population variance are given by

$$R_{GMV} = \frac{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}} \quad \text{and} \quad V_{GMV} = \frac{1}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}}. \quad (4.5)$$

The quantity  $s$  is the slope parameter of the mean-variance efficient frontier, the set of all optimal portfolios in the mean-variance space, and it is given by

$$s = \boldsymbol{\mu}^\top \mathbf{M} \boldsymbol{\mu}. \quad (4.6)$$

The efficient frontier itself is a parabola given by (see, e.g., Bodnar and Schmid 2009, Kan and Smith 2008, Merton 1972)

$$(R - R_{GMV})^2 = s(V - V_{GMV}). \quad (4.7)$$

Recently, Markowitz's optimization problem has been reformulated from the Bayesian perspectives by Bauder et al. (2020b). Using the portfolio expected return  $R_{t-1}(\mathbf{w}) = \mathbb{E}(X_{P,t} | \mathbf{x}_{(t-1)})$  and the portfolio variance  $V_{t-1}(\mathbf{w}) =$



$\text{Var}(X_{P,t}|\mathbf{x}_{(t-1)})$  computed from the posterior predictive distribution as in (2.10)-(2.11), Markowitz's optimization problem from the Bayesian point of view is given by

$$\min_{\mathbf{w}: R_{t-1}(\mathbf{w})=R_0, \mathbf{w}^\top \mathbf{1}=1} V_{t-1}(\mathbf{w}), \quad (4.8)$$

with the solution expressed as

$$\mathbf{w}_{MV,t-1} = \mathbf{w}_{GMV,t-1} + \frac{R_0 - R_{GMV,t-1}}{s_{t-1}} \mathbf{M}_{t-1} \bar{\mathbf{x}}_{t-1} \quad (4.9)$$

with

$$\mathbf{M}_{t-1} = \mathbf{S}_{t-1}^{-1} - \frac{\mathbf{S}_{t-1}^{-1} \mathbf{1} \mathbf{1}^\top \mathbf{S}_{t-1}^{-1}}{\mathbf{1}^\top \mathbf{S}_{t-1}^{-1} \mathbf{1}}, \quad (4.10)$$

$$\mathbf{w}_{GMV,t-1} = \frac{\mathbf{S}_{t-1}^{-1} \mathbf{1}}{\mathbf{1}^\top \mathbf{S}_{t-1}^{-1} \mathbf{1}}, \quad (4.11)$$

and

$$R_{GMV,t-1} = \frac{\mathbf{1}^\top \mathbf{S}_{t-1}^{-1} \bar{\mathbf{x}}_{t-1}}{\mathbf{1}^\top \mathbf{S}_{t-1}^{-1} \mathbf{1}}, \quad V_{GMV,t-1} = \frac{d_{k,n} r_{k,n}}{d_{k,n} - 2} \frac{1}{\mathbf{1}^\top \mathbf{S}_{t-1}^{-1} \mathbf{1}}. \quad (4.12)$$

The quantity  $s_{t-1}$  is one of the factors determining the slope of the Bayesian efficient frontier in the mean-variance space and it is given by

$$s_{t-1} = \bar{\mathbf{x}}_{t-1}^\top \mathbf{M}_{t-1} \bar{\mathbf{x}}_{t-1}. \quad (4.13)$$

Also from the Bayesian perspective, the efficient frontier is a parabola expressed as (see, Bauder et al. 2020b)

$$(R - R_{GMV,t-1})^2 = \frac{d_{k,n} - 2}{d_{k,n} r_{k,n}} s_{t-1} (V - V_{GMV,t-1}). \quad (4.14)$$

In contrast to the population optimal portfolios and the efficient frontier, the Bayesian optimal portfolio and the Bayesian efficient frontier are presented in terms of the historical data that are observable up to time  $t - 1$ , when the optimal portfolio for the next period is constructed. Moreover, the Bayesian portfolio allocation is based on the predictive posterior distribution and incorporates the parameter uncertainty in the decision process before the weights of optimal portfolios are computed.

Assuming asset returns to be normally distributed, Alexander and Baptista (2002, 2004) extended Markowitz's optimization problem by replacing the population portfolio variance in (4.1) with the population VaR and CVaR, respectively, given by

$$Q_P(\mathbf{w}) = -\mathbf{w}^\top \boldsymbol{\mu} + q_{P;\alpha} \sqrt{\mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}}, \quad (4.15)$$

where  $q_{P;\alpha} = z_\alpha$  for the VaR and  $q_{P;\alpha} = \frac{\exp(-z_\alpha^2/2)}{(1-\alpha)\sqrt{2\pi}}$  for the CVaR where  $z_\alpha$  denotes the  $\alpha$ -quantile of the standard normal distribution.

The quantile-based optimization problems of Alexander and Baptista (2002, 2004) are given by

$$\min_{\mathbf{w}: R_P(\mathbf{w})=R_0, \mathbf{w}^\top \mathbf{1}=1} Q_P(\mathbf{w}). \quad (4.16)$$

If the constraint on the expected return is omitted in (4.16), then the solutions of (4.16) are the weights of the population optimal portfolios with the smallest values of VaR (or CVaR) at confidence level  $\alpha$  given by (see, Bodnar et al. 2012)

$$\mathbf{w}_{GMQ} = \mathbf{w}_{GMV} + \frac{\sqrt{V_{GMV}}}{\sqrt{q_{P;\alpha}^2 - s}} \mathbf{M}\boldsymbol{\mu}. \quad (4.17)$$

Similarly to the mean-variance portfolio, the weights (4.17) of the population minimum VaR (or CVaR) portfolio cannot be computed. First, the unknown population parameters  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  should be estimated by using historical data of asset returns and, then, the estimator of  $\mathbf{w}_{GMQ}$  is constructed as a proxy of the true portfolio weights. This two-step procedure of constructing an optimal portfolio usually leads to sub-optimal solutions since the parameter uncertainty is ignored in its construction. In the next subsection, we deal with the problem from the viewpoint of Bayesian statistics which allows to incorporate the parameter uncertainty directly into the decision process before the optimization problem is solved.

## 4.2 Bayesian quantile-based optimal portfolios

For a general quantile-based risk measure, the extension of the Alexander and Baptista optimization problem (4.16) from the Bayesian perspectives is given by

$$\min_{\mathbf{w}: R_{t-1}(\mathbf{w})=R_0, \mathbf{w}^\top \mathbf{1}=1} Q_{t-1}(\mathbf{w}), \quad (4.18)$$

where  $R_{t-1}(\mathbf{w})$  and  $Q_{t-1}(\mathbf{w})$  are computed by using the posterior predictive distribution as discussed in Section 2.1 and Section 3. For the special choices of the function  $Q_{t-1}(\mathbf{w})$  as discussed in Section 3.1, we get the optimization problems that minimize the predictive portfolio VaR and the predictive portfolio CVaR.

The solution of the optimization problem (4.18) can be presented in the

following way

$$\begin{aligned}
& \underset{\mathbf{w}: R_{t-1}(\mathbf{w})=R_0, \mathbf{w}^\top \mathbf{1}=1}{\operatorname{argmin}} Q_{t-1}(\mathbf{w}) \\
&= \underset{\mathbf{w}: R_{t-1}(\mathbf{w})=R_0, \mathbf{w}^\top \mathbf{1}=1}{\operatorname{argmin}} -R_{t-1}(\mathbf{w}) + q_\alpha \sqrt{\frac{d_{k,n} - 2}{d_{k,n}}} \sqrt{V_{t-1}(\mathbf{w})} \\
&= \underset{\mathbf{w}: R_{t-1}(\mathbf{w})=R_0, \mathbf{w}^\top \mathbf{1}=1}{\operatorname{argmin}} V_{t-1}(\mathbf{w}),
\end{aligned}$$

provided that  $d_{k,n} > 2$ . Hence, on the one hand, all solutions of the quantile-based optimization problem (4.18) are also the solutions of the mean-variance optimization problem (4.8) and belong to the efficient frontier (4.14). On the other hand, all four optimization problems (4.1), (4.8), (4.16), and (4.18) possess a solution only if  $R_0$  is properly chosen. For example, (4.1) and (4.8) have solutions if and only if  $R_0 > R_{GMV}$  and  $R_0 > R_{GMV,t-1}$ , respectively, while for solving (4.16) one requires that

$$q_{P;\alpha}^2 - s > 0. \quad (4.19)$$

Below in Theorem 4.1, we formulate the conditions needed for the existence of the Bayesian optimal portfolio in the sense of minimizing  $Q_{t-1}(\mathbf{w})$ .

It has to be noted that the conditions of solution existence formulated in the case of the population optimization problems (4.1) and (4.16) depend on the unknown population parameters of the data generating process and thus they cannot be validated in practice. In contrast, the Bayesian formulation of the optimization problems makes it possible to specify the existence conditions in terms of the previously observed data  $\mathbf{x}_{(t-1)}$  and, thus, to check them before the optimization problem is solved. Finally, the conditions on the existence of the solutions in the quantile-based optimization problems (4.16) and (4.18) depend on the chosen confidence level  $\alpha$ , although the solutions themselves are independent of it.

Similarly to the mean-variance optimization problems, in order to determine under which conditions imposed on  $R_0$  the solutions of the quantile-based optimization problem exist, one has to find the optimal portfolio with the smallest possible value of the objective function  $Q_{t-1}(\mathbf{w})$ , that is when the constraint  $R_{t-1}(\mathbf{w}) = R_0$  is dropped from the optimization problem (4.18). The expected return of this portfolio will provide the smallest possible value for which the optimization problem (4.18) possesses a solution. To this end, we note that this is also the portfolio which a completely risk averse investor may be interested in. The following theorem expresses the variance and return of such a portfolio.

**Theorem 4.1.** *Let  $d_{k,n} > 2$ . Then, under the conditions of Theorem 2.1, the global minimum quantile (GMQ)-based optimal portfolio exists if and only if*

$$q_\alpha^2 > r_{k,n}^{-1} s_{t-1}, \quad (4.20)$$

where  $s_{t-1}$  is defined in (4.13). Moreover, its posterior predictive expected return and variance are given by

$$R_{GMQ,t-1} = R_{GMV,t-1} + \frac{r_{k,n}^{-1} s_{t-1}}{\sqrt{q_\alpha^2 - r_{k,n}^{-1} s_{t-1}}} \sqrt{\frac{d_{k,n} - 2}{d_{k,n}}} \sqrt{V_{GMV,t-1}}, \quad (4.21)$$

and

$$V_{GMQ,t-1} = \frac{q_\alpha^2}{q_\alpha^2 - r_{k,n}^{-1} s_{t-1}} V_{GMV,t-1}, \quad (4.22)$$

where  $V_{GMV,t-1}$  and  $R_{GMV,t-1}$  are given in (4.12).

The statement of Theorem 4.1 is proved in the appendix. Its results determine the lower bound for possible values of  $R_0$  that can be used in the optimization problem (4.18). Since  $R_{GMQ,t-1} > R_{GMV,t-1}$ , we get that the set of optimal portfolios which solve (4.18) does not coincide with the set of the Bayesian mean-variance optimal portfolios which lie on the upper part of the efficient frontier given by the parabola (4.14) in the mean-variance space.

The findings of Theorem 4.1 lead to the expression of the smallest possible value of  $Q_{t-1}(\mathbf{w})$  for the selected confidence level  $\alpha$  expressed as

$$Q_{GMQ,t-1} = -R_{GMQ,t-1} + q_\alpha \sqrt{\frac{d_{k,n} - 2}{d_{k,n}}} \sqrt{V_{GMQ,t-1}}. \quad (4.23)$$

Finally, the weights of the global minimum quantile-based portfolio are deduced from the findings of Theorem 4.1 and they are presented in Theorem 4.2.

**Theorem 4.2.** *Let  $d_{k,n} > 2$  and the inequality (4.20) holds. Then, under the conditions of Theorem 2.1, the weights of the global minimum quantile-based optimal portfolio are given by*

$$\mathbf{w}_{GMQ,t-1} = \mathbf{w}_{GMV,t-1} + \frac{r_{k,n}^{-1} \sqrt{V_{GMV,t-1}}}{\sqrt{q_\alpha^2 - r_{k,n}^{-1} s_{t-1}}} \sqrt{\frac{d_{k,n} - 2}{d_{k,n}}} \mathbf{M}_{t-1} \bar{\mathbf{x}}_{t-1}. \quad (4.24)$$

### 4.3 Bayesian efficient frontier in mean-quantile space

Earlier in this section we proved that the solutions of the quantile-based portfolio optimization problem (4.18) belong to the Bayesian efficient frontier (4.14) in the mean-variance space. We now characterise the location of the Bayesian quantile-based optimal portfolio in the mean-quantile (mean-Q) space. It has to be noted that the population mean-VaR efficient frontier was carried out by Alexander and Baptista (2002) under the assumption that the asset returns are multivariate normally distributed. We extend these findings in Theorem 4.3 whose proof is given in the appendix.

**Theorem 4.3.** *Let  $d_{k,n} > 2$  and  $s_{t-1} > 0$ . Then, under the conditions of Theorem 2.1, the Bayesian efficient frontier in the mean-Q space is a hyperbola given by*

$$Q = q_\alpha \sqrt{\frac{(R - R_{GMV,t-1})^2}{r_{k,n}^{-1} s_{t-1}} + \frac{d_{k,n} - 2}{d_{k,n}} V_{GMV,t-1}} - R. \quad (4.25)$$

Expressions for the mean-variance efficient frontier using the Bayesian setup was derived in Bauder et al. (2020b). It holds that the mean-variance efficient frontier (4.14) is a parabola in the mean-variance space and a hyperbola in the mean-standard deviation space for  $s_{t-1} > 0$ . These findings are in line with the results in Merton (1972), where the same conclusions were drawn for the population efficient frontier. In Theorem 4.3, we prove that the efficient frontier in the mean-Q space is also a hyperbola under the same condition  $s_{t-1} > 0$ . It is interesting to note that since  $\mathbf{M}_{t-1}$  is positive semi-definite with  $\mathbf{M}_{t-1} \mathbf{1} = \mathbf{0}$  by construction, it always holds that  $s_{t-1} \geq 0$  with  $s_{t-1} = 0$  only if the elements of the vector  $\bar{\mathbf{x}}_{t-1}$  are all equal. Another important observation is that both efficient frontiers (4.14) and (4.25) are determined by the same set of quantities  $R_{GMV,t-1}$ ,  $V_{GMV,t-1}$ , and  $s_{t-1}$  which are computed from the historical data of asset returns.

*Remark 4.4.* Using the proof of Theorem 4.3, we also obtain the analytical expression of the population efficient frontier in the mean-Q space, thus complementing the findings of Alexander and Baptista (2002) who presents this frontier in the empirical study without deriving its closed-form expression. It holds that the population efficient frontier in the mean-Q space is a hyperbola expressed as

$$Q = q_{P;\alpha} \sqrt{\frac{(R - R_{GMV})^2}{s} + V_{GMV}} - R. \quad (4.26)$$

It is fully determined by the same set of constants  $R_{GMV}$ ,  $V_{GMV}$ , and  $s$  as the population efficient frontier (4.7), which is also a hyperbola in the mean standard-deviation space.

## 5 Simulation study

In the following section, we analyse how the Bayesian approaches compare to the conventional method via simulations. We will do so by studying VaR prediction using the global minimum VaR (GMVaR) portfolio and by looking at how frequently the conditions (4.19) and (4.20) are satisfied. The comparison for other quantile-based risk measures can be done similarly by emphasising that the coherent risk measures have the same structure as given in (3.5). Utilizing (3.5) it is interesting to note that any coherent risk measure can be rewritten as VaR at confidence level  $\beta = F_{d_{k,n}}(\rho_{t-1}(\tau))$  for the Bayesian approaches, where  $F_{d_{k,n}}(\cdot)$  stands for the cumulative distribution function of the univariate  $t$ -distribution with  $d_{k,n}$  degrees of freedom and  $\tau$  is a  $t$ -distributed random variable with  $d_{k,n}$  degrees of freedom. For the conventional method, one can use the same procedure, where the  $t$ -distribution is replaced by the standard normal distribution. Finally, we compare the different estimation methods by illustrating their corresponding efficient frontiers. When doing this comparison we also include the global minimum variance (GMV) portfolio in the analysis to see where it is located in the mean-VaR space.

### 5.1 Setup of simulation study

Throughout the simulation study, the asset returns are generated from a multivariate normal distribution. This distribution satisfies the assumptions of infinite exchangeability and multivariate centered spherical symmetry when conditioning on the parameters (see, e.g., Proposition 4.6 in Bernardo and Smith 2009). In order to not restrict the analysis to certain parameters, the mean vector and covariance matrix are randomized in each new simulation iteration. We draw  $\boldsymbol{\mu}$  from the uniform distribution on  $[-0.003, 0.005]$ , i.e.,  $\mu_i \sim U(-0.003, 0.005)$ , and the covariance matrix is constructed by writing it as  $\boldsymbol{\Sigma} = \mathbf{D}\mathbf{R}\mathbf{D}$  where  $\mathbf{R}$  is a correlation matrix with  $(\mathbf{R})_{ij} = 0.3$  if  $i \neq j$  and  $\mathbf{D}$  is a diagonal matrix with entries given by  $(\mathbf{D})_{ii} \sim U(0.03, 0.04)$ . We also consider different sample sizes and portfolio sizes by using  $n \in \{100, 200\}$  and  $k = cn$  for  $c \in \{0.1, 0.3, 0.5, 0.7\}$ . Moreover, we use  $\alpha \in \{0.95, 0.99\}$  to study the impact of the confidence level in the GMVaR computations. For each parameter setup, we consider 10000 independent simulation runs when studying the performance and existence of the GMVaR portfolios. In each such simulation iteration, the out-of-sample performance of the portfolios is evaluated for one period ahead. We then aggregate the obtained results in all simulation runs. To this end, only those results where the conventional and Bayesian GMVaR conditions (4.19) and (4.20) are satisfied simultaneously

are considered.

Since the true parameters of the asset return distribution are known during simulation, it is possible to make comparisons with the population GMVaR portfolios as well as the population efficient frontier. The population GMVaR portfolios are constructed from the same mathematical formulas as when using the conventional method but they are based on the true parameters. Hence the population portfolios can be used as benchmarks for the corresponding Bayesian and conventional portfolios which are all based on parameter estimates. Similarly, the population efficient frontier can be used as a reference for the estimated efficient frontiers. Finally, the hyperparameters  $\mathbf{m}_0$  and  $\mathbf{S}_0$  of the conjugate prior are determined by using the empirical Bayesian approach (see, e.g., Bauder et al. 2020a) where we set  $d_0 = r_0 = n$ .

## 5.2 Existence and performance

The GMVaR existence conditions (4.19) and (4.20) are not satisfied in several simulation runs when the portfolio dimension becomes large in comparison to the sample size. More precisely, in the following cases  $\{n = 100, k = 50, \alpha = 0.95\}$ ,  $\{n = 100, k = 70, \alpha = 0.95\}$ ,  $\{n = 200, k = 100, \alpha = 0.95\}$ ,  $\{n = 200, k = 140, \alpha = 0.95\}$ ,  $\{n = 100, k = 70, \alpha = 0.99\}$  and  $\{n = 200, k = 140, \alpha = 0.99\}$ , the conventional condition (4.19) is not met for 247, 8432, 1486, 9995, 1027 and 4155 out of the 10000 simulation runs, respectively. For the Jeffreys prior, the corresponding numbers are 0, 16, 0, 51, 0 and 0, and they are 4, 581, 4, 2694, 1 and 0 for the conjugate prior. For all other values of  $\{n, k, \alpha\}$  the existence conditions are satisfied. They are always fulfilled for the population GMVaR portfolio. Based on these findings we conclude that both Bayesian GMVaR portfolios are more likely to exist than its conventional counterpart with the Bayesian GMVaR portfolio under the Jeffreys prior demonstrating the lowest frequencies of non-existence in all of the considered cases. Our results also show that it is more likely that the GMVaR conditions are not satisfied when  $\alpha$  is small or when  $c$  is large. Finally, we point out that when  $\{n = 200, k = 140, \alpha = 0.95\}$  the conventional GMVaR portfolio does not exist in 9995 out of 10000 simulation iterations. For this reason, the values of the performance measures for this configuration are not presented in Tables 1 and 2.

We use two measures to analyze the performance of the GMVaR portfolios. The first measure is the relative frequency of times the estimated VaR is exceeded, i.e.,

$$\frac{1}{N} \sum_{i=1}^N \mathbf{1}\{-X_{\text{GMVaRP},i} \geq \widehat{\text{VaR}}_{\alpha}(X_{\text{GMVaRP},i})\},$$

where  $N$  is the number of simulations,  $\mathbf{1}$  is the indicator function,  $X_{\text{GMVaR},i}$  is the actual return of the estimated GMVaR portfolio for simulation  $i$  and  $\widehat{\text{VaR}}_{\alpha}(X_{\text{GMVaR},i})$  is its predicted VaR. The latter two are calculated using equations (4.23) and (4.24) in the Bayesian cases and (4.15) and (4.17) in the conventional and population cases. By the definition of VaR, an exceedance rate close to  $1 - \alpha$  means a good prediction of the VaR.

In Table 1 we observe that the relative VaR exceedance of the population GMVaR portfolios are always close to the target confidence level. The only source of noise is from the number of simulation runs. Such results do not hold for the three estimated GMVaR portfolios. The relative exceedance frequencies are close to the target confidence level when the portfolio dimension is small with respect to the sample size. For other values of  $k$  and  $n$ , the predicted VaRs underestimate the true values. These results are in line with recent findings in portfolio theory, i.e., that the sample optimal portfolios are overoptimistic and tend to underestimate the risk. To this end we note that although the Bayesian approaches underestimate the risk, they still perform considerably better than the conventional approach. Especially, when  $k/n \geq 0.5$  the relative exceedance rate for the conventional approach is almost twice as large as the ones obtained for the Bayesian methods when  $\alpha = 0.95$  and it is almost three times larger for  $\alpha = 0.99$ .



Table 1: Relative VaR exceedance frequencies for the population GMVaR portfolio and its three estimates

Parameter setup			GMVaR portfolio			
$\alpha$	$n$	$k$	Jeffreys	Conjugate	Conventional	Population
0.95	100	10	0.0663	0.0741	0.0793	0.0510
		30	0.1101	0.1375	0.1688	0.0510
		50	0.1676	0.2237	0.3001	0.0490
		70	0.1996	0.2691	0.3967	0.0510
	200	20	0.0640	0.0694	0.0760	0.0471
		60	0.1114	0.1389	0.1682	0.0490
		100	0.1649	0.2199	0.2984	0.0525
	140	—	—	—	—	
0.99	100	10	0.0154	0.0190	0.0216	0.0114
		30	0.0348	0.0510	0.0720	0.0095
		50	0.0663	0.1049	0.1755	0.0098
		70	0.1160	0.1909	0.3300	0.0110
	200	20	0.0135	0.0158	0.0193	0.0091
		60	0.0330	0.0475	0.0667	0.0101
		100	0.0676	0.1074	0.1723	0.0115
		140	0.1158	0.1814	0.3179	0.0091

The relative VaR exceedance frequency measures how accurate the GM-VaR estimates are by its definition. It does not measure how far away the estimated GMVaR is from the population GMVaR. To investigate this behaviour of our predictions we use a second measure which is the average absolute deviation of the estimated GMVaR to its population value, i.e.,

$$\frac{1}{N} \sum_{i=1}^N |\widehat{\text{VaR}}_{\alpha}(X_{\text{GMVaRP},i}) - \text{VaR}_{\alpha}(X_{\text{GMVaRP},i})|,$$

where  $\text{VaR}_{\alpha}(X_{\text{GMVaRP},i})$  is the VaR of the population GMVaR portfolio. Since the population GMVaR portfolio is based on the true parameter values, the VaR of the population GMVaR portfolio coincides with true value of VaR. Hence, it can be used as a benchmark and the average absolute deviation should ideally be close to zero.

Table 2 shows the results of the GMVaR portfolio comparison using the average absolute deviation as a performance measure. Like in Table 1, we observe the same performance when the portfolio size is considerably smaller than the sample size. The Bayesian approach based on the Jeffreys prior shows the smallest deviations although the different portfolios are very close

in their performance. For larger portfolio sizes, the Bayesian methods are significantly better than the conventional approach. They possess smaller values of the performance criteria as well as the computed standard deviations are smaller than those obtained for the conventional procedure. The differences become very large in the extreme case when  $k/n = 0.7$ .

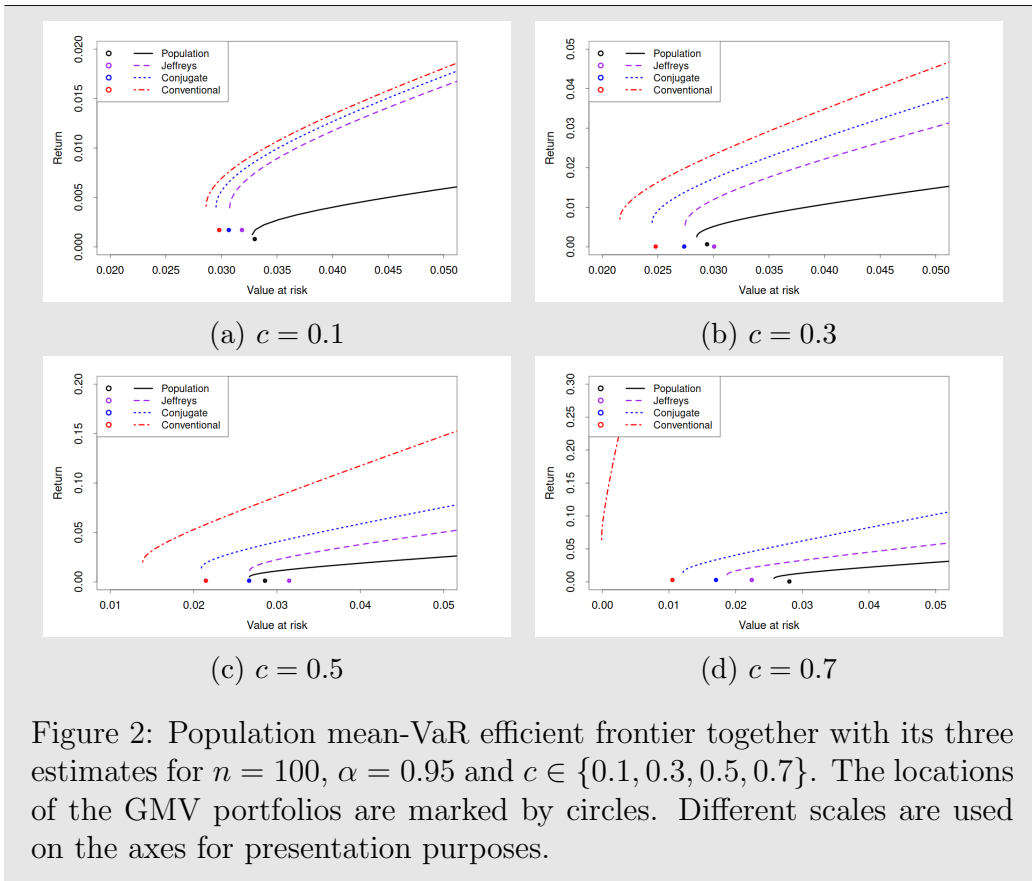
Table 2: Average absolute deviation of the VaR of the estimated GMVaR portfolios to the VaR of the population GMVaR portfolio. Values inside the brackets represent the standard deviations.

Parameter setup			GMVaR portfolio		
$\alpha$	$n$	$k$	Jeffreys	Conjugate	Conventional
0.95	100	10	0.0027 (0.0021)	0.0029 (0.0022)	0.0032 (0.0024)
		30	0.0030 (0.0023)	0.0048 (0.0030)	0.0076 (0.0033)
		50	0.0038 (0.0028)	0.0079 (0.0038)	0.0148 (0.0042)
		70	0.0038 (0.0029)	0.0084 (0.0039)	0.0213 (0.0039)
	200	20	0.0019 (0.0014)	0.0021 (0.0016)	0.0026 (0.0018)
		60	0.0023 (0.0017)	0.0045 (0.0023)	0.0075 (0.0024)
		100	0.0030 (0.0022)	0.0078 (0.0028)	0.0157 (0.0035)
		140	—	—	—
0.99	100	10	0.0035 (0.0027)	0.0035 (0.0027)	0.0041 (0.0029)
		30	0.0035 (0.0027)	0.0051 (0.0034)	0.0088 (0.0038)
		50	0.0040 (0.0031)	0.0080 (0.0043)	0.0161 (0.0041)
		70	0.0052 (0.0039)	0.0110 (0.0053)	0.0266 (0.0054)
	200	20	0.0023 (0.0017)	0.0025 (0.0018)	0.0032 (0.0022)
		60	0.0024 (0.0018)	0.0047 (0.0026)	0.0085 (0.0027)
		100	0.0030 (0.0022)	0.0079 (0.0032)	0.0158 (0.0031)
		140	0.0034 (0.0025)	0.0103 (0.0036)	0.0266 (0.0040)

To conclude, the suggested Bayesian methods perform better at estimating the GMVaR than the conventional approach. Using the Jeffreys prior seems to give the best results, although using the conjugate prior is also beneficial compared to the conventional method. However, all of the methods are underestimating the GMVaR, especially when  $c$  is large. This is indicated by VaR exceedance frequencies much higher than  $1 - \alpha$  and large deviations to the population GMVaR. However, even if none of the methods perform very well for large-dimensional portfolios, this is the situation where we see the greatest benefit of using the Bayesian approaches. This point is further studied in the next section where we investigate the influence of parameter uncertainty on the estimation of the whole mean-VaR efficient frontier.

### 5.3 Comparison of efficient frontiers

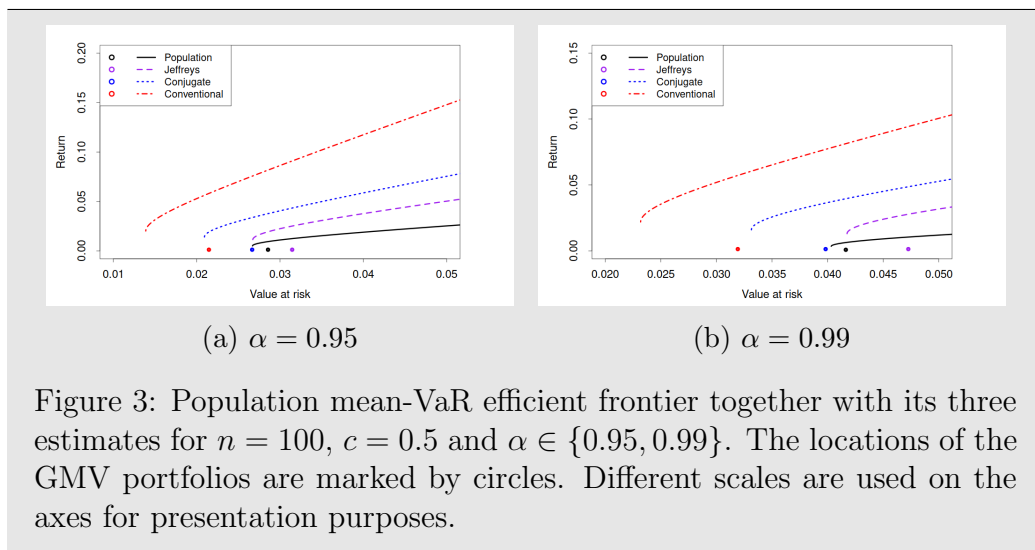
In order to get a better understanding of the impact of parameter uncertainty on the quantile-based portfolio selection, we use the theoretical findings of Section 4.3 and plot the population mean-VaR efficient frontier together with its three estimates in Figures 2 to 4. The estimates of the mean-VaR efficient frontier are computed for a single simulation run as described at the beginning of this section by using (4.25) and (4.26) for the Bayesian and conventional estimates, respectively. It should be noted that the figures present the most common results which are also observed for other simulation runs. All of the figures also show where the portfolio which globally minimizes the variance is located in the mean-VaR space using each of the methods.



The mean-VaR efficient frontiers and the locations of the GMV portfolios are depicted in Figure 2 for  $\alpha = 0.95$ ,  $n = 100$ , and  $c \in \{0.1, 0.3, 0.5, 0.7\}$ . We observe that all methods overestimate the location of the true efficient frontier in the mean-VaR space. Such a behaviour is similar to the one pre-

viously documented for the Markowitz efficient frontier in the mean-variance space by Broadie (1993), Siegel and Woodgate (2007), Bodnar and Bodnar (2010), Bauder et al. (2019) among others. Namely, ignoring the parameter uncertainty leads to overoptimistic investment opportunities where the investors expect more return for the same level of risk than the population efficient frontier determines. The situation becomes even worse when the conventional mean-VaR frontier is constructed for  $c = 0.5$  and especially for  $c = 0.7$ . The conventional efficient frontier deviates drastically from the population efficient frontier. Among the two Bayesian efficient frontiers, the one based on the Jeffreys prior leads to the curves that are closest to the population frontier for all considered portfolio sizes. We also observe the positive effect of portfolio diversification in Figure 2. Increasing the portfolio dimension leads to the reduction of the VaR of the GMVaR portfolio. Also, we note the positive effect on the slope parameter of the efficient frontier which becomes larger.

Figure 2 also illustrates that the portfolios that minimize the variance are not located on the mean-VaR efficient frontiers. This is an expected but important observation which illustrates that an investor who is mean-variance efficient may not always be mean-VaR efficient.



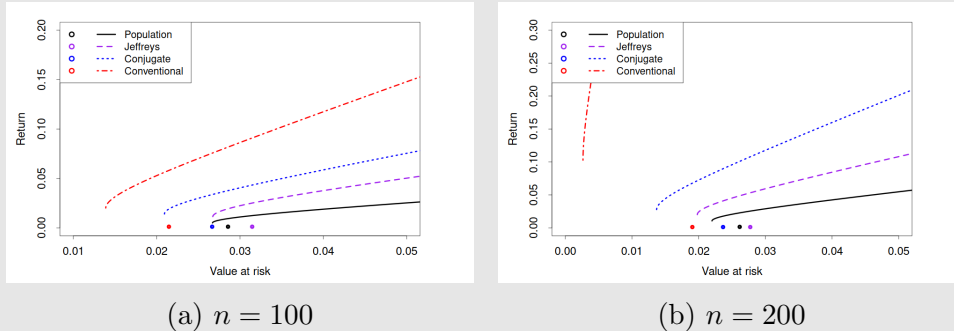


Figure 4: Population mean-VaR efficient frontier together with its three estimates for  $c = 0.5$ ,  $\alpha = 0.95$  and  $n \in \{100, 200\}$ . The locations of the GMV portfolios are marked by circles. Different scales are used on the axes for presentation purposes.

Figure 3 and 4 demonstrate that the conclusions drawn from the results of Figure 2 are also valid for other values of  $\alpha$  and  $n$ . In both figures the Bayesian approach with Jeffreys prior provides the best fit of the population efficient frontier followed by the Bayesian estimate based on the conjugate prior. Also, we observe that the increase of the portfolio dimension with the simultaneous increase of the sample size leads the reduction of the VaR of the GMVaR portfolio and to the increase in the slope parameter of the efficient frontier. Moreover, the GMV portfolios are again shown to not be mean-VaR efficient.

## 6 Empirical illustration

We now continue the comparison between the Bayesian and conventional methodologies through an application on actual market data. As in the simulation study, we study the performance and existence of the GMVaR portfolio and investigate the behaviour of their efficient frontiers. Once again, we consider the cases  $n \in \{100, 200\}$ ,  $k = cn$  for  $c \in \{0.1, 0.3, 0.5, 0.7\}$  and  $\alpha \in \{0.95, 0.99\}$ .

### 6.1 Data description

We use weekly returns on stocks included in the S&P 500 index for the period from the 1st of January, 2010 to the 28th of March, 2020. In order to circumvent the possible bias of selecting stocks which outperform or underperform

the rest of the market, we consider all stocks included in the S&P 500 index by our end date that were already part of the the index by our chosen start date. The lack of public information makes it difficult to know exactly when a certain stock was added to this index, but based on Wikipedia contributors (2020) we have chosen to consider 221 stocks that were present in the index before the 1st of January, 2010. A complete list of the stocks is provided in Table 4 in Appendix B.

In the empirical analysis, we randomly choose 500 portfolios of size  $k$  from the list of stocks for each possible value of  $\{n, k, \alpha\}$ . Once the stocks have been selected, they are kept for the whole time period but the weights of the GMVaR portfolio are re-calculated each week and the performance is evaluated on a weekly basis and then averaged across all sampled portfolios. It should be noted that the performance results are only based on portfolios which satisfy the conventional and Bayesian GMVaR conditions (4.19) and (4.20), respectively, for all estimates of the GMVaR portfolio simultaneously. Finally, the hyperparameters when using the conjugate prior are specified as in the simulation study, i.e., by employing the empirical Bayesian approach and setting  $d_0 = r_0 = n$ .

## 6.2 Results of the empirical illustration

Regarding the existence of the estimates of the GMVaR portfolio, it is more likely that the Bayesian GMVaR condition (4.20) is satisfied than the conventional condition (4.19). The Bayesian GMVaR portfolios exist all the time whereas the conventional GMVaR portfolio does not always exist when  $\{n = 100, k = 70, \alpha = 0.95\}$ ,  $\{n = 200, k = 140, \alpha = 0.95\}$  and  $\{n = 100, k = 70, \alpha = 0.99\}$ . For those values, the conventional GMVaR condition fails during the time period for 369, 5 and 1 portfolios, respectively, out of the 500 portfolios. Hence, we observe that the conditions are more likely to be satisfied when  $\alpha$  is large or when  $c$  is small.

As in the simulation study, we consider the relative VaR exceedance frequency when evaluating the performance of each estimate of the GMVaR portfolio. This value should ideally be close to  $1 - \alpha$ . The result is summarized in Table 3.

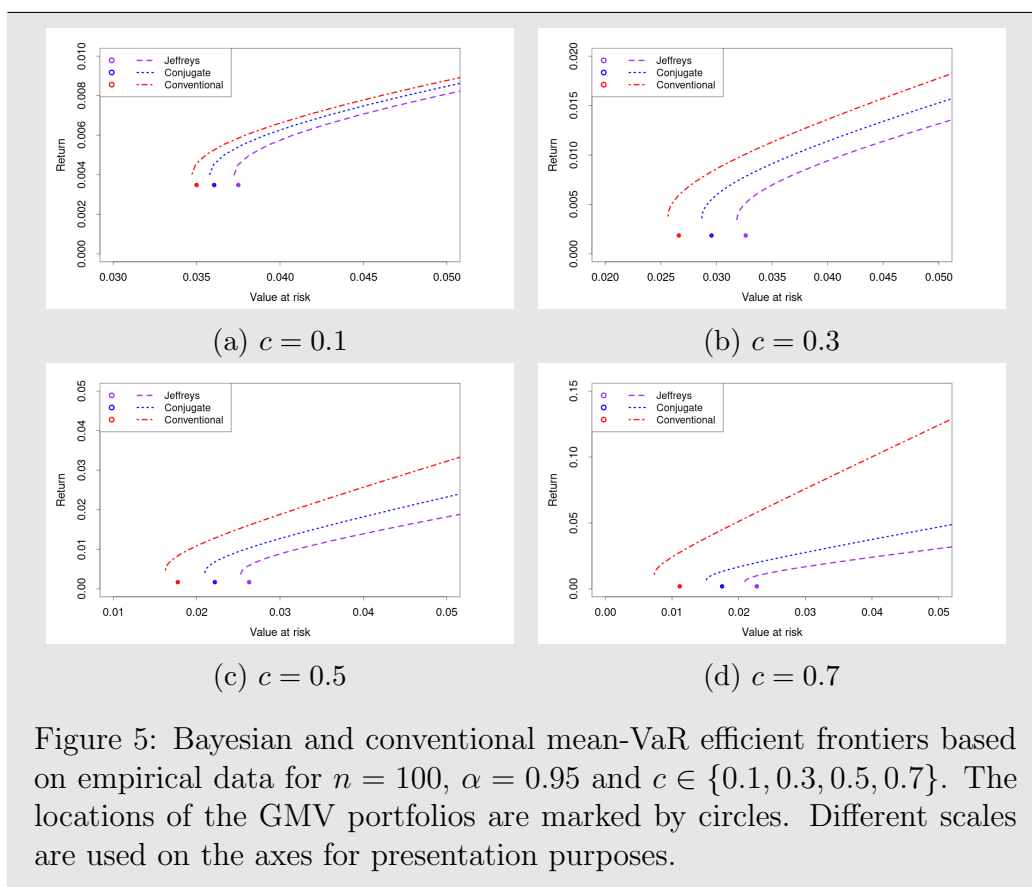
Table 3: Relative VaR exceedance frequencies for the three estimates of the GMVaR portfolio.

Parameter setup			GMVaR portfolio		
$\alpha$	$n$	$k$	Jeffreys	Conjugate	Conventional
0.95	100	10	0.0771	0.0828	0.0872
		30	0.1140	0.1356	0.1602
		50	0.1617	0.2071	0.2698
		70	0.2227	0.2900	0.4040
	200	20	0.0794	0.0846	0.0890
		60	0.1144	0.1354	0.1600
		100	0.1625	0.2047	0.2655
		140	0.2294	0.2920	0.3966
0.99	100	10	0.0361	0.0394	0.0424
		30	0.0532	0.0664	0.0833
		50	0.0802	0.1137	0.1669
		70	0.1283	0.1911	0.3057
	200	20	0.0375	0.0404	0.0429
		60	0.0521	0.0646	0.0818
		100	0.0816	0.1135	0.1650
		140	0.1359	0.1948	0.3046

As can be seen, the Bayesian approaches are clearly outperforming the conventional method by having exceedance frequencies closer to  $1 - \alpha$ . Using the Jeffreys prior gives the best results in all situations that we consider. However, as in the simulation study, all of the methods are underestimating VaR since the exceedance frequency is always higher than  $1 - \alpha$ . This is especially pronounced when  $c$  is large, i.e., in the case of a large-dimensional portfolio. Even if none of the methods performs very well for such situations, the Bayesian approaches, especially the one based on the Jeffreys prior, provides a considerable improvement in comparison to the conventional method by reducing the relative exceedance frequency by 45% for  $\alpha = 0.95$  and by 55% for  $\alpha = 0.99$  when  $k/n = 0.7$ . The application of the Bayesian approach based on the conjugate prior also results in much lower relative exceedance frequencies compared to the conventional method, although this Bayesian GMVaR portfolio performs always worse than the one based on the Jeffreys prior.

Similar results to those observed in Table 3 are also present in Figure 5 where we plot the estimated mean-VaR efficient frontier for the end date using  $\alpha = 0.95$ ,  $n = 100$ , and  $c = k/n \in \{0.1, 0.3, 0.5, 0.7\}$ . Both Bayesian

efficient frontiers are always located under the conventional efficient frontier. While the three efficient frontiers almost coincide when  $c = 0.1$ , the difference between the conventional and Bayesian approaches becomes pronounced when  $c$  becomes larger, particularly when  $c = 0.7$ . Moreover, we again see that the GMV portfolios are not mean-VaR efficient. All of this is in line with the observations made for Figure 2 in the simulation study. Varying  $\alpha$  and  $n$  using the empirical data will also result in the same relationships between the efficient frontiers as shown in Figures 3 and 4 in the simulation study, indicating the considerable overoptimism present in the construction of the conventional efficient frontier.



## 7 Conclusion

The traditional mean-variance analysis has been a paramount foundation for the extension to portfolio analysis based on one-sided risk measures which are popular in financial mathematics. However, the conventional approaches



related to the construction of optimal portfolios usually ignore the parameter uncertainty in the construction of an optimal portfolio. It is common to define optimal portfolios by a two-step procedure where first an optimization problem is solved and then the optimal portfolios are estimated by replacing the unknown quantities in the solutions by the corresponding sample counterparts.

The Bayesian methodology pose a fundamental difference to the conventional approaches in its viewpoint on what we want to optimize: Investors care about their *future risk in taking a position*, not the risk of having the a certain position today. In light of data, today's outcome is already determined and usually not interesting. The Bayesian framework use the predictive posterior distribution to cope with this. That is, the Bayesian methodology answers the problem in a straightforward manner while the conventional method simply ignores it.

We contribute to the existent literature by formulating and solving quantile-based portfolio allocation problems from the perspective of Bayesian statistics. This approach is advantageous since it allows to take the parameter uncertainty into account before the optimization problem is solved. The development of the general risk functionals from the Bayesian perspectives appears to be a very promising subject of research with great potential of future development. The risk functionals can be defined through all information available up to the point in time when a portfolio is constructed or a decision on the risk of the current position should be made. As a result, no unknown or unobservable quantities are present in their definitions. This is a very appealing property since it takes all uncertainties into account before the risk functional is determined.

In the frequentist setting, the general risk measure of an optimal portfolio choice problem will have a similar structure as under the Bayesian setup when asset returns are elliptically contoured distributed (see, e.g., Gupta et al. 2013, for a definition and properties thereof). However, both the portfolio expected return and the portfolio variance are determined by unknown parameters of the distribution of asset returns which must be estimated in any practical application. This procedure would lead to an important task, namely to include the parameter uncertainty in the definition of the general risk measure. This challenging task has not properly been treated in the literature up to now when frequentist methods are employed, while the Bayesian approach provides an intelligent automatic solution.

Results of the simulation study and of the empirical application leads to the conclusion that the Bayesian approaches to portfolio construction provide a good alternative to the conventional procedures and they are usually preferable in most of the considered cases. The Bayesian approaches outperform

the conventional one in terms of providing a better VaR prediction. This holds uniformly, independently of the portfolio dimension, sample size, and the confidence level used in the computation of the VaR. Only when the portfolio dimension is relatively small to the sample size does the conventional method perform similarly to the Bayesian approaches. Such a behavior is expected since the priors used in the derivation of Bayesian inference can be interpreted as a regularisation and it might not be necessary employ that in such cases. Although using Jeffreys prior gave the best results in our study, a more careful calibration of the hyperparameters of the conjugate prior could have made that one more beneficial. In practice, the hyperparameters would be specified using knowledge from experts within fundamental market analysis.

We also find that the conventional mean-VaR efficient frontier considerably overestimates the location of the true mean-VaR frontier. Although the Bayesian approaches reduce the underestimation of the VaR considerably and shrink the estimates of the efficient frontier, they still show significant overoptimism when the portfolio dimension is large in comparison to the sample size, i.e., when a large-dimensional optimal portfolio is constructed. Further research in this direction is needed which might lead to interesting results completing the existing findings in the direction of large-dimensional portfolio construction (see, e.g., Fan et al. 2012, Hautsch et al. 2015, Bodnar et al. 2019, Cai et al. 2020).

## References

- Adcock, C. J. (2014). Mean–variance–skewness efficient surfaces, stein’s lemma and the multivariate extended skew-student distribution. *European Journal of Operational Research*, 234(2):392–401.
- Alexander, G. J. and Baptista, A. M. (2002). Economic implications of using a mean-var model for portfolio selection: A comparison with mean-variance analysis. *Journal of Economic Dynamics and Control*, 26(7):1159–1193.
- Alexander, G. J. and Baptista, A. M. (2004). A comparison of var and cvar constraints on portfolio selection with the mean-variance model. *Management Science*, 50(9):1261–1273.
- Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.
- Avramov, D. and Zhou, G. (2010). Bayesian portfolio analysis. *The Annual Review of Financial Economics*, 2(1):25–47.
- Babat, O., Vera, J. C., and Zuluaga, L. F. (2018). Computing near-optimal value-

- at-risk portfolios using integer programming techniques. *European Journal of Operational Research*, 266(1):304–315.
- Barry, C. B. (1974). Portfolio analysis under uncertain means, variances, and covariances. *Journal of Finance*, 29:515–522.
- Bauder, D., Bodnar, R., Bodnar, T., and Schmid, W. (2019). Bayesian estimation of the efficient frontier. *Scandinavian Journal of Statistics*, 46:802–830.
- Bauder, D., Bodnar, T., Parolya, N., and Schmid, W. (2020a). Bayesian inference of the multi-period optimal portfolio for an exponential utility. *Journal of Multivariate Analysis*, 175. 104544.
- Bauder, D., Bodnar, T., Parolya, N., and Schmid, W. (2020b). Bayesian mean–variance analysis: optimal portfolio selection under parameter uncertainty. *Quantitative Finance*, page to appear.
- Baumol, W. J. (1963). An expected gain-confidence limit criterion for portfolio selection. *Management Science*, 10(1):174–182.
- Bawa, V. S., Brown, S. J., and Klein, R. W. (1979). *Estimation risk and optimal portfolio choice*. North-Holland.
- Bernardo, J. M. and Smith, A. F. (2009). *Bayesian theory*, volume 405. John Wiley & Sons.
- Black, F. and Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, 48:28–43.
- Bodnar, O. and Bodnar, T. (2010). On the unbiased estimator of the efficient frontier. *International Journal of Theoretical and Applied Finance*, 13:1065–1073.
- Bodnar, T., Dmytriv, S., Parolya, N., and Schmid, W. (2019). Tests for the weights of the global minimum variance portfolio in a high-dimensional setting. *IEEE Transactions on Signal Processing*, 67(17):4479–4493.
- Bodnar, T., Mazur, S., and Okhrin, Y. (2017). Bayesian estimation of the global minimum variance portfolio. *European Journal of Operational Research*, 256:292–307.
- Bodnar, T., Parolya, N., and Schmid, W. (2018). Estimation of the global minimum variance portfolio in high dimensions. *European Journal of Operational Research*, 266(1):371–390.
- Bodnar, T. and Schmid, W. (2009). Econometrical analysis of the sample efficient frontier. *The European Journal of Finance*, 15(3):317–335.
- Bodnar, T., Schmid, W., and Zabolotsky, T. (2012). Minimum var and minimum cvar optimal portfolios: estimators, confidence regions, and tests. *Statistics & Risk Modeling with Applications in Finance and Insurance*, 29(4):281–313.
- Broadie, M. (1993). Computing efficient frontiers using estimated parameters. *Annals of Operations Research*, 45:21–58.
- Bronstein, I. N., Semendyayev, K. A., Musiol, G., and Mühlig, H. (2015). *Handbook of Mathematics*. Springer Science & Business Media.

- Brown, S. J. (1976). *Optimal portfolio choice under uncertainty: a Bayesian approach*. PhD thesis, University of Chicago.
- Brugière, P. (2020). *Quantitative Portfolio Management*. Springer.
- Cai, T. T., Hu, J., Li, Y., and Zheng, X. (2020). High-dimensional minimum variance portfolio estimation based on high-frequency data. *Journal of Econometrics*, 214(2):482–494.
- Chopra, V. K. and Ziemba, W. T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management*, 19(2):6–11.
- Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223–236.
- Fan, J., Zhang, J., and Yu, K. (2012). Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association*, 107(498):592–606.
- Frankfurter, G. M., Phillips, H. E., and Seagle, J. P. (1971). Portfolio selection: the effects of uncertain means, variances, and covariances. *Journal of Financial and Quantitative Analysis*, 6(5):1251–1262.
- Frost, P. A. and Savarino, J. E. (1986). An empirical bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis*, 21(3):293–305.
- Givens, G. H. and Hoeting, J. A. (2013). *Computational statistics*. John Wiley & Sons.
- Gupta, A. K. and Nagar, D. K. (2000). *Matrix variate distributions*. Chapman and Hall/CRC, Boca Raton.
- Gupta, A. K., Varga, T., and Bodnar, T. (2013). *Elliptically contoured models in statistics and portfolio theory*. Springer.
- Hautsch, N., Kyj, L. M., and Malec, P. (2015). Do high-frequency data improve high-dimensional portfolio allocations? *Journal of Applied Econometrics*, 30:263–290.
- Jorion, P. (1986). Bayes-stein estimation for portfolio analysis. *Journal of Financial and Quantitative analysis*, 21(3):279–292.
- Jorion, P. (1997). *Value at risk: the new benchmark for controlling market risk*. Irwin Professional Pub.
- Kan, R. and Smith, D. R. (2008). The distribution of the sample minimum-variance frontier. *Management Science*, 54(7):1364–1380.
- Kingman, J. F. et al. (1978). Uses of exchangeability. *The Annals of Probability*, 6(2):183–197.
- Klein, R. W. and Bawa, V. S. (1976). The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics*, 3(3):215–231.
- Kolm, P. and Ritter, G. (2017). On the bayesian interpretation of black–litterman. *European Journal of Operational Research*, 258(2):564–572.

- Kotz, S. and Nadarajah, S. (2004). *Multivariate t-distributions and their applications*. Cambridge University Press.
- Krüger, F., Lerch, S., Thorarinsdottir, T. L., and Gneiting, T. (2020). Predictive inference based on markov chain monte carlo output. *arXiv preprint arXiv:1608.06802*.
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91.
- Meng, X. and Taylor, J. W. (2020). Estimating value-at-risk and expected shortfall using the intraday low and range data. *European Journal of Operational Research*, 280(1):191–202.
- Merton, R. C. (1972). An analytic derivation of the efficient portfolio frontier. *The Journal of Financial and Quantitative Analysis*, 7(4):1851–1872.
- Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8(4):323–361.
- Pritsker, M. (1997). Evaluating value at risk methodologies: accuracy versus computational time. *Journal of Financial Services Research*, 12(2-3):201–242.
- Rachev, S. T., Hsu, J. S., Bagasheva, B. S., and Fabozzi, F. J. (2008). *Bayesian methods in finance*. John Wiley & Sons.
- Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42.
- Siegel, A. F. and Woodgate, A. (2007). Performance of portfolios optimized with estimation error. *Management Science*, 53:1005–1015.
- Simaan, Y. (2014). The opportunity cost of mean–variance choice under estimation risk. *European Journal of Operational Research*, 234(2):382–391.
- Staino, A. and Russo, E. (2020). Nested conditional value-at-risk portfolio selection: A model with temporal dependence driven by market-index volatility. *European Journal of Operational Research*, 280(2):741–753.
- Stambaugh, R. F. (1997). Analyzing investments whose histories differ in length. *Journal of Financial Economics*, 45:285–331.
- Tsay, R. S. (2010). *Analysis of financial time series*. John Wiley & Sons.
- Tu, J. and Zhou, G. (2010). Incorporating economic objectives into bayesian priors: Portfolio choice under parameter uncertainty. *Journal of Financial and Quantitative Analysis*, pages 959–986.
- Wikipedia contributors (2020). List of s&p 500 companies – wikipedia, the free encyclopedia. [Online; accessed 14-July-2020].
- Winkler, R. L. and Barry, C. B. (1975). A bayesian model for portfolio selection and revision. *The Journal of Finance*, 30(1):179–192.
- Zellner, A. and Ando, T. (2010). A direct monte carlo approach for bayesian analysis of the seemingly unrelated regression model. *Journal of Econometrics*, 159(1):33–45.

## Acknowledgement

This research was partly supported by the Swedish Research Council (VR) via the project ‘‘Bayesian Analysis of Optimal Portfolios and Their Risk Measures’’.

## A Proofs of theoretical results

In order to get the stochastic representations presented below we also need the following result.

**Lemma A.1.** *Let a random variable  $z$  possess the following stochastic representation*

$$z \stackrel{d}{=} \frac{\tau_1}{\sqrt{vd}} + \sqrt{1 + \frac{\tau_1^2}{d}} \frac{\tau_2}{\sqrt{d+1}}, \quad (\text{A.1})$$

where  $d > 0$ ,  $\tau_1$  and  $\tau_2$  are independent with  $\tau_1 \sim t(d)$  and  $\tau_2 \sim t(d+1)$ . Then,  $z$  follows a *t-distribution with  $d$  degrees of freedom, location parameter 0, and scale parameter  $\sqrt{(v+1)/vd}$ .*

*Proof of Lemma A.1.* Since  $\tau_1$  and  $\tau_2$  are independent with  $\tau_2 \sim t(d+1)$ , the conditional distribution of  $z$  given  $\tau_1$  is a *t-distribution with  $d+1$  degrees of freedom, location parameter  $\tau_1/\sqrt{vd}$  and scale parameter  $g(\tau_1)/\sqrt{d+1}$  with  $g(\tau_1) = \sqrt{1 + \tau_1^2/d}$ .* Thus the joint distribution of  $z$  and  $\tau_1$  is given by

$$\begin{aligned} f(z, \tau_1) &= f(z|\tau_1)f(\tau_1) \\ &= \frac{\Gamma\left(\frac{d+2}{2}\right)}{\Gamma\left(\frac{d+1}{2}\right)} \frac{1}{\sqrt{\pi}} \frac{1}{g(\tau_1)} \left(1 + \left(\frac{z - \frac{\tau_1}{\sqrt{vd}}}{g(\tau_1)}\right)^2\right)^{-\frac{d+2}{2}} \frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \frac{1}{\sqrt{\pi d}} \left(1 + \frac{\tau_1^2}{d}\right)^{-\frac{d+1}{2}} \\ &\propto \frac{1}{g(\tau_1)} \left(1 + \left(\frac{z - \frac{\tau_1}{\sqrt{vd}}}{g(\tau_1)}\right)^2\right)^{-\frac{d+2}{2}} g(\tau_1)^{-(d+1)} = \left(g(\tau_1)^2 + \left(z - \frac{\tau_1}{\sqrt{vd}}\right)^2\right)^{-\frac{d+2}{2}} \\ &= \left(1 + \frac{1}{d}[z, \tau_1] \begin{bmatrix} d & -\frac{\sqrt{vd}}{v} \\ -\frac{\sqrt{vd}}{v} & \frac{v+1}{v} \end{bmatrix} \begin{bmatrix} z \\ \tau_1 \end{bmatrix}\right)^{-\frac{d+2}{2}}. \end{aligned}$$

The last expression is the kernel of a multivariate *t-distribution with  $d$  degrees of freedom, location vector  $\boldsymbol{\nu} = \mathbf{0}$  and dispersion matrix  $\boldsymbol{\Omega}$  given by*

$$\boldsymbol{\Omega} = \begin{bmatrix} \frac{v+1}{vd} & \frac{\sqrt{vd}}{vd} \\ \frac{\sqrt{vd}}{vd} & 1 \end{bmatrix}.$$

Hence, the marginal distribution of  $z$  is also a  $t$ -distribution with  $d$  degrees of freedom, location 0 and scale  $\sqrt{(v+1)/vd}$  (see, e.g., Kotz and Nadarajah 2004).  $\square$

*Proof of Theorem 2.1.* Bauder et al. (2020b) characterized the posterior predictive distribution of the portfolio return by deriving the stochastic representation of  $\widehat{X}_{P,t}$  given by

$$\widehat{X}_{P,t} \stackrel{d}{=} m + \sqrt{s} \left( \frac{\tau_1}{\sqrt{vd}} + \sqrt{1 + \frac{\tau_1^2}{d} \frac{\tau_2}{\sqrt{d+1}}} \right)$$

with  $m = \mathbf{w}^\top \bar{\mathbf{x}}_{t-1,J}$ ,  $s = \mathbf{w}^\top \mathbf{S}_{t-1,J} \mathbf{w}$ ,  $v = n$ , and  $d = n - k$  under the Jeffreys prior and with  $m = \mathbf{w}^\top \bar{\mathbf{x}}_{t-1,I}$ ,  $s = \mathbf{w}^\top \mathbf{S}_{t-1,I} \mathbf{w}$ ,  $v = n + r_0$  and  $d = n + d_0 - 2k$  under the conjugate prior. The application of Lemma A.1 leads to the statement of the theorem.  $\square$

*Proof of Theorem 3.1.* The statement of the theorem follows from the fact that  $\mathbf{w}^\top \bar{\mathbf{x}}_{t-1}$  is linear in  $\mathbf{w}$  and that, since  $\mathbf{S}_{t-1}$  is positive definite,  $\sqrt{\mathbf{w}^\top \mathbf{S}_{t-1} \mathbf{w}}$  can be regarded as the Euclidean norm of  $\mathbf{S}_{t-1}^{1/2} \mathbf{w}$  where  $\mathbf{S}_{t-1}^{1/2}$  is the symmetric square root of  $\mathbf{S}_{t-1}$ . Since  $q_\alpha > 0$  and the Euclidean norm is convex, the result follows.  $\square$

*Proof of Theorem 4.1.* Let  $c_{k,n} = \frac{d_{k,n} r_{k,n}}{d_{k,n} - 2}$ . Since the solution of

$$\min_{\mathbf{w}: \mathbf{w}^\top \mathbf{1} = 1} Q_{t-1}(\mathbf{w}),$$

belongs to the Bayesian efficient frontier (4.14) in the mean-variance space, it can be found by solving the univariate optimization problem given by

$$\begin{aligned} \min_{V: V \geq V_{GMV,t-1}} & -R_{GMV,t-1} - (c_{k,n})^{-1/2} \sqrt{s_{t-1}} \sqrt{V - V_{GMV,t-1}} \\ & + q_\alpha (c_{k,n})^{-1/2} \sqrt{r_{k,n}} \sqrt{V} \end{aligned} \quad (\text{A.2})$$

where  $R_{GMV,t-1}$  and  $V_{GMV,t-1}$  are given in (4.12) and  $s_{t-1}$  is defined in (4.13). The solution of (A.2) solves

$$q_\alpha \sqrt{r_{k,n}} \frac{1}{\sqrt{V}} = \sqrt{s_{t-1}} \frac{1}{\sqrt{V - V_{GMV,t-1}}} \quad (\text{A.3})$$

and it is given by

$$V_{GMQ,t-1} = \frac{q_\alpha^2}{q_\alpha^2 - r_{k,n}^{-1} s_{t-1}} V_{GMV,t-1}. \quad (\text{A.4})$$

where it obviously holds that  $V_{GMQ,t-1} > V_{GMV,t-1}$  as soon as  $q_\alpha^2 - r_{k,n}^{-1}s_{t-1} > 0$ , which coincides with the second order condition needed to ensure that  $V_{GMQ,t-1}$  is the solution of (A.2).

Finally,  $R_{GMQ,t-1}$  is obtained from (4.14) and it is given by

$$\begin{aligned} R_{GMQ,t-1} &= R_{GMV,t-1} + \sqrt{c_{k,n}^{-1}s_{t-1}} \sqrt{\frac{q_\alpha^2}{q_\alpha^2 - r_{k,n}^{-1}s_{t-1}} V_{GMV,t-1} - V_{GMV,t-1}} \\ &= R_{GMV,t-1} + \frac{r_{k,n}^{-1}s_{t-1}}{\sqrt{q_\alpha^2 - r_{k,n}^{-1}s_{t-1}}} \sqrt{\frac{d_{k,n} - 2}{d_{k,n}}} \sqrt{V_{GMV,t-1}}. \end{aligned}$$

□

*Proof of Theorem 4.3.* From (4.14) and (3.4), we get

$$\frac{(R - R_{GMV,t-1})^2}{a_{t-1}} + V_{GMV,t-1} = V \quad \text{and} \quad V = \left(\frac{R + Q}{b}\right)^2 \quad (\text{A.5})$$

where

$$a_{t-1} = \frac{d_{k,n} - 2}{d_{k,n}r_{k,n}}s_{t-1} \quad \text{and} \quad b = q_\alpha \sqrt{\frac{d_{k,n} - 2}{d_{k,n}}}.$$

From (A.5), we get

$$\begin{aligned} Q &= b \sqrt{\frac{(R - R_{GMV,t-1})^2}{a_{t-1}} + V_{GMV,t-1} - R} \\ &= q_\alpha \sqrt{\frac{(R - R_{GMV,t-1})^2}{r_{k,n}^{-1}s_{t-1}} + \frac{d_{k,n} - 2}{d_{k,n}} V_{GMV,t-1} - R}. \end{aligned}$$

Finally, we note that the Bayesian efficient frontier in the mean-Q space can be rewritten as

$$R^2 - 2RR_{GMV,t-1} + R_{GMV,t-1}^2 - \frac{a_{t-1}}{b^2}R^2 - 2\frac{a_{t-1}}{b^2}RQ - \frac{a_{t-1}}{b^2}Q^2 + a_{t-1}V_{GMV,t-1} = 0,$$

which is a hyperbola in the mean-Q space for  $s_{t-1} > 0$  (see, e.g., Section 3.5.2.11 in Bronshtein et al. 2015) since

$$-\frac{a_{t-1}}{b^2} \left(1 - \frac{a_{t-1}}{b^2}\right) - \frac{a_{t-1}^2}{b^4} = -\frac{a_{t-1}}{b^2} = -\frac{s_{t-1}}{r_{k,n}q_\alpha^2} < 0.$$

□



## B List of stocks

Table 4 presents the list of stocks considered in Section 6.

Table 4: Stocks considered in the empirical illustration.

MMM	ABT	ADBE	AES	AFL	A	APD	AKAM
ALL	GOOG	AMZN	AEE	AXP	AIG	AMT	AMP
ABC	AMGN	APH	ADI	ANTM	AON	APA	AIV
AAPL	AMAT	ADM	AIZ	T	ADSK	ADP	AZO
AVB	AVY	BLL	BAC	BK	BAX	BDX	BBY
BIIB	BKNG	BXP	BSX	BF.B	CHRW	COG	COF
CAH	CCL	CBRE	CNP	CTL	CF	SCHW	CI
CINF	CTAS	CSCO	C	CTXS	CLX	CME	CMS
CTSH	CMA	CAG	STZ	COST	CSX	CMI	DVA
XRAY	DVN	DFS	DOV	DUK	ETFC	EMN	ECL
EA	EMR	EOG	EFX	EQR	EL	EXPE	EXPD
FAST	FDX	FIS	FISV	FLIR	FLS	FMC	GPS
GIS	GPC	GILD	GL	GS	GWW	HRB	HAS
PEAK	HES	HD	HON	HRL	HST	HPQ	ITW
INTC	ICE	IPG	IFF	INTU	ISRG	IVZ	IRM
J	SJM	JNJ	JPM	JNPR	KEY	KIM	LB
LH	LEN	LLY	LNC	LIN	LMT	LOW	MRO
MMC	MAS	MA	MCD	MDT	MCHP	MU	MSFT
TAP	MYL	NDAQ	NOV	NTAP	NWL	NEM	NEE
NKE	NBL	JWN	NOC	NUE	NVDA	ORLY	OXY
ORCL	PCAR	PH	PBCT	PKI	PM	PXD	PNC
PFG	PGR	PLD	PRU	PSA	PHM	PWR	DGX
RL	RF	RSG	RHI	ROP	ROST	CRM	SLB
SHW	SPG	SNA	LUV	SWK	SYK	SYY	TGT
FTI	TXT	TIF	TJX	TRV	TFC	UNH	UPS
UNM	VFC	VAR	VTR	VRSN	VZ	V	VMC
WMT	WBA	DIS	WEC	WFC	WELL	WDC	WMB
WYNN	XLNX	YUM	ZBH	ZION			