European Journal of Operational Research 256 (2017) 292–307

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Interfaces with Other Disciplines

Bayesian estimation of the global minimum variance portfolio[☆]

Taras Bodnar[a], Stepan Mazur[b], Yarema Okhrin[c,*]

[a] Department of Mathematics, Stockholm University, Roslagsvägen 101, Stockholm SE-10691, Sweden
[b] Department of Statistics, Lund University, Tycho Brahes väg 1, Lund SE-22007, Sweden
[c] Department of Statistics, University of Augsburg, Augsburg D-86159, Germany

ARTICLE INFO

Article history:
Received 20 April 2015
Accepted 25 May 2016
Available online 21 June 2016

Keywords:
Global minimum variance portfolio
Posterior distribution
Credible interval
Wishart distribution

ABSTRACT

In this paper we consider the estimation of the weights of optimal portfolios from the Bayesian point of view under the assumption that the conditional distributions of the logarithmic returns are normal. Using the standard priors for the mean vector and the covariance matrix, we derive the posterior distributions for the weights of the global minimum variance portfolio. Moreover, we reparameterize the model to allow informative and non-informative priors directly for the weights of the global minimum variance portfolio. The posterior distributions of the portfolio weights are derived in explicit form for almost all models. The models are compared by using the coverage probabilities of credible intervals. In an empirical study we analyze the posterior densities of the weights of an international portfolio.

1. Introduction

Starting with the seminal paper of Markowitz (1952) the classical mean-variance portfolio theory has drawn much attention in academic literature. Generally speaking, the theory allows us to determine the optimal port/folio weights which guarantee the lowest risk for a given expected portfolio return. Under Gaussian asset returns, the problem is equivalent to minimizing the expected quadratic utility of the future wealth (c.f., Bodnar, Parolya, & Schmid, 2013). In practice, however, the model frequently led to investment opportunities with modest ex-post profits and high risk. To clarify this and to develop improved trading strategies several issues were addressed, which can be roughly separated in two partly overlapping branches in literature. The first strand of research analyses the estimation risk in portfolio weights, which arises if we replace the unknown parameters of the distribution of asset returns with their sample counterparts. The results on the finite sample distributions can be used in different ways. First, we can develop a test to check if the weights of a particular asset significantly deviate from prespecified values, e.g. test for efficiency (see Ang & Bekaert, 2002; Bodnar & Schmid, 2008; Britten-Jones, 1999; Jobson & Korkie, 1989; Stambaugh, 1997). Second, we can test the significance of the investment in a given asset, e.g. significance of international diversification (see French & Poterba, 1991). Third, we may assess the sensitivity of portfolio weights to changes in the parameters of the asset returns as in Best and Grauer (1991); Bodnar (2009); Chopra and Ziemba (1993), and many others.

The main contribution of Markowitz from the financial perspective is the recognition of the importance of diversification. From a statistical point of view, the portfolio theory stresses the importance of the variance as a measure of risk and particularly the importance of the structure of the covariance matrix for diversification purposes. Markowitz's approach allows us to determine the minimum variance set of portfolios and the sets of efficient portfolios. While the minimum variance set consists of those portfolios which possess the minimum variance for a chosen level of the expected return, the efficient set contains the portfolios with the highest level of the expected return for each level of risk. As a result, the choice of an optimal portfolio depends on the investor's attitude towards risk, i.e. on his/her level of risk aversion. Markowitz (2014) showed both theoretically and empirically that the mean-variance method and the expected utility approach lead to similar optimal portfolios, whereas (Liesiö & Salo, 2012) developed a portfolio selection framework which uses the set inclusion to capture incomplete information about scenario probabilities and utility functions. This approach identifies all of the non-dominated project portfolios in view of this information as well as it offers the decision support for the rejection and selection of projects.

☆ The authors appreciate the financial support of the German Science Foundation (DFG), projects BO3521/2 and OK103/1, "Wishart Processes in Statistics and Econometrics: Theory and Applications". The first author is partly supported by the German Science Foundation (DFG) via the Research Unit 1735 "Structural Inference in Statistics: Adaptation and Efficiency". He also appreciates the financial support of the German Science Foundation (DFG) via the projects BO 3521/3-1 and SCHM 859/13-1 "Bayesian Estimation of the Multi-Period Optimal Portfolio Weights and Risk Measures".

* Corresponding author. Tel.: +49 821 5984152; fax: +49 821 5984227.
E-mail address: yarema.okhrin@wiwi.uni-augsburg.de (Y. Okhrin).

http://dx.doi.org/10.1016/j.ejor.2016.05.044
0377-2217/

Levy and Levy (2014) analyzed the impact of estimation error in portfolio optimization, while (Bodnar, Parolya, & Schmid, 2015a; 2015c) present analytical solutions to multi-period portfolio choice problems based on the quadratic and exponential utility functions.

The global minimum variance (GMV) portfolio is a specific optimal portfolio which possesses the smallest variance among all portfolios on the efficient frontier. This portfolio corresponds to the fully-risk averse investor who aims to minimize the variance without taking the expected return into consideration. The importance of the GMV portfolio in financial applications was well motivated by Merton (1980) who pointed out that the estimates of the variances and the covariances of the asset returns are much more accurate than the estimates of the means. Later, Best and Grauer (1991) showed that the sample efficient portfolio is extremely sensitive to changes in the asset means, whereas (Chopra & Ziemba, 1993) concluded for a real data set that errors in means are over ten times as damaging as errors in variances and over twenty times as errors in covariances. For this reason many authors assume equal means for the portfolio asset returns or, in other terms, the GMV portfolio. This is one reason why this is extensively discussed in literature (Chan et al. 1999). Moreover, the GMV portfolio has the lowest variance of any feasible portfolio. More evidence regarding the practical application of the GMV portfolio can be found in Haugen (1999).

In contrary to the above approaches, the second strand of research opts for the Bayesian framework. The Bayesian setting resembles the decision making of market participants and the human way of information utilization. Similarly, investors use the past experiences and memory (historical event, trends) for decisions at a given time point. These subjective beliefs flow into the decision making process in a Bayesian setup via specific priors. From this point of view the Bayesian framework is potentially more attractive in portfolio theory (see Avramov & Zhou, 2010). The first applications of Bayesian statistics in portfolio analysis were completely based on uninformative or data-based priors, see Winkler (1973); Winkler and Barry (1975). Bawa, Brown, and Klein (1979) provided an excellent review on early examples of Bayesian studies on portfolio choice. These contributions stimulated a steady growth of interest in Bayesian tools for asset allocation problems. Barberis (2000); Jorion (1986); Kandel and Stambaugh (1996); Pastor (2000) used the Bayesian framework to analyze the impact of the underlying asset pricing or predictive model for asset returns on the optimal portfolio choice. Bodnar, Parolya, and Schmid (2015b); Golosnoy and Okhrin (2007); 2008); Kan and Zhou (2007); Wang (2005) concentrated on shrinkage estimation, which allows to shift the portfolio weights to prespecified values, reflecting the prior beliefs of investors. Brandt (2010) gives a state of the art review of modern portfolio selection techniques, paying particular attention to Bayesian approaches.

In the majority of the mentioned papers, the authors defined specific priors for the model parameters and the subsequent evaluation of posterior distributions or asset allocation decisions was performed numerically. The reason is that the involved integral expressions are too complex for analytic derivation. In this paper we derive explicit formulas for the posterior distributions of the global minimum-variance portfolio weights for several non-informative and informative priors on the parameters of asset returns. Furthermore, using a specific reparameterization we obtain non-informative and informative priors for the portfolio weights directly. This appears to be more consistent with the decision processes of investors. The corresponding posterior distributions are presented too. The established results are evaluated within a simulation study, which assesses the coverage probabilities of credible intervals, and within an empirical study, where we concentrate on the posterior distributions of the weights of an internationally diversified portfolio.

The rest of the paper is structured as follows. Bayesian estimation of the GMV portfolio using preliminary results is presented in Section 2. The posterior distributions for the GMV portfolio are derived and summarized in Theorem 1. In Section 3 we propose informative and non-informative prior distributions for the weights of the GMV portfolio and the corresponding posterior distributions (Theorems 2 and 3). In Section 4 the credible intervals and credible sets for the previous posterior distributions are obtained. The results of numerical and empirical studies are given in Section 5, while Section 6 summarizes the paper. The appendix (Section 7) contains the proof of Theorem 1 and additional technical results.

2. Bayesian portfolio selection

We consider a portfolio of k assets. Let $\mathbf{X}_i = (X_{1i}, \dots, X_{ki})^T$ be the k-dimensional random vector of log-returns at time $i = 1, \dots, n$. For small values of returns, the simple and the log-returns behave similarly. Let $\mathbf{w} = (w_1, \dots, w_k)^T$ be the vector of portfolio weights, where w_j denotes the weight of the jth asset, and let $\mathbf{1}$ be the vector of ones. Assuming that dynamics of the terminal wealth is governed by the standard Brownian motion[1], we obtain the log-normal distribution as the distribution of the wealth. This leads to Gaussian portfolio log-returns, which are equal to the sum of the log-returns of the underlying assets, i.e. $\mathbf{X}_{\mathbf{w}} = \sum_{i=1}^k w_i \mathbf{X}_i$ (see Dhaene, Vanduffel, Goovaerts, Kaas, & Vyncke, 2005). Note that below we assume a conditional normal distribution of the asset returns given the mean vector and the covariance matrix, which is a much weaker assumption than the unconditional Gaussian distribution. Let the mean vector of the asset returns be denoted by $\boldsymbol{\mu}$ and the covariance matrix by a positive definite matrix $\boldsymbol{\Sigma}$. The GMV portfolio is the unique solution of the optimization problem

$$\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \rightarrow min \quad \text{s.t.} \quad \mathbf{w}^T \mathbf{1} = 1. \tag{1}$$

In general we allow for short sales and therefore for negative weights. The solution of (1) is given by

$$\mathbf{w}_{GMV} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T \boldsymbol{\Sigma}^{-1}\mathbf{1}}. \tag{2}$$

Since $\boldsymbol{\Sigma}$ is an unknown parameter, the formula in (2) is infeasible for practical purposes. Given a sample of size n of historical vectors of returns $\mathbf{x}_1, \dots, \mathbf{x}_n$, we can compute the sample covariance matrix

$$\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T,$$

where $\overline{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$. The sample estimator of the GMV portfolio weights is constructed by replacing $\boldsymbol{\Sigma}$ with $\mathbf{S}$ in (2) and it is given by

$$\widehat{\mathbf{w}}_{GMV} = \frac{\mathbf{S}^{-1}\mathbf{1}}{\mathbf{1}^T \mathbf{S}^{-1}\mathbf{1}}. \tag{3}$$

In this paper we take a more general setup by considering arbitrary linear combinations of the GMV portfolio weights. Let $\mathbf{L}$ be an arbitrary $p \times k$ matrix of constants, $p < k$, and define

$$\boldsymbol{\theta} = \mathbf{L}\mathbf{w}_{GMV} = \frac{\mathbf{L}\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T \boldsymbol{\Sigma}^{-1}\mathbf{1}}. \tag{4}$$

[1] In fact when the length of the subsequent time intervals becomes smaller we obtain in the limit that $(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$ effectively becomes increments of a multidimensional Brownian motion and by keeping the portfolio weights constant during the investment horizon (by continuously rebalancing) we obtain that the terminal wealth is lognormally distributed. In this regard, note that the assumption of constant portfolio weights appears as a trading constraint. There is also a rich literature on optimal mean-variance portfolios when there are no such trading constraints (in which case the optimal terminal wealth is no longer lognormally distributed); see, Bernard and Vanduffel (2014); Goetzmann, Ingersoll, Spiegel, and Welch (2007)

The sample estimator of $\boldsymbol{\theta}$ is given by

$$\widehat{\boldsymbol{\theta}} = \mathbf{L}\widehat{\mathbf{w}}_{GMV} = \frac{\mathbf{L}\mathbf{S}^{-1}\mathbf{1}}{\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1}}. \tag{5}$$

In practice, the investors concentrate on the point estimators $\widehat{\boldsymbol{\theta}}$ without realizing the estimation risk induced by estimated parameters $\bar{\mathbf{x}}$ and $\mathbf{S}$. This risk is extremely damaging for asset allocation since it renders wrong or misspecified portfolios (see Best & Grauer, 1991). In order to assess the estimation risk we must consider $\widehat{\mathbf{w}}_{GMV}$ and $\hat{\boldsymbol{\theta}}$ as a random quantity. As $\bar{\mathbf{x}}$ and $\mathbf{S}$ deviate from the true parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, so can the estimated portfolio deviate from the weights of the true optimal portfolio leading to poor out-sample performance in practice. The variation in the parameters can also have others sources than pure estimation reasons. In the time series framework it is frequently observed that the parameters are not constant over time. These dynamics are frequently modeled either by an appropriate time series process or by a regime switching process. Although this type of dynamics is difficult to implement here directly, it allows for some additional information which should be exploited for portfolio decisions.

Thus a very important objective is not only to quantify and formalize the information about the parameters, but also to take it into account already while computing the optimal portfolio composition. Methodologically the Bayesian framework offers a convenient and appropriate set of tools. Within this framework we rely on our beliefs or prior information about the parameters of the model and formalize these beliefs in form of prior distributions. The most frequently applied priors for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in the financial literature are the diffuse prior (see, e.g., Barry, 1974; Brown, 1976, and Klein & Bawa (1976)), the conjugate prior (Frost & Savarino, 1986), and the hierarchical prior (Greyserman, Jones, & Strawderman, 2006) which we introduce next. The diffuse prior is an uninformative prior, which implies that the statistician has no additional information about the stochastic nature of the unknown parameters. The conjugate prior is an informative prior and we assume that the mean returns follow a normal distribution and the covariance matrix follows an inverse Wishart distribution. The hierarchical prior is a more complex prior which allows for additional distributional assumptions about the precision of the priors for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

For every prior we can compute the corresponding posterior distribution of the portfolio weights which takes the prior distribution of the parameters into account. This means that we provide not only the point estimator of the optimal portfolio weights as it is usual in practice, but also its whole distribution. The mean of this distribution provides us with a new Bayesian estimator of the portfolio weights which accounts for the investor's prior beliefs. These results allow us to run tests for portfolio weights and construct credible sets. The latter are the analogue of confidence intervals and they determine regions where the true portfolio weights lie with high probability. We can use these findings to test the significance of the investment in a particular asset. Detailed discussion and results are provided in Section 3.

From the financial perspective it might be difficult to formulate and to motivate a specific prior for the parameters but it is common to have some beliefs about the optimal portfolio composition. For example one might formulate the prior beliefs in form of the equally weighted portfolio, which shows superior out-of-sample long-run performance as reported frequently. Alternatively, the prior portfolio composition might be proportional to the market capitalizations of the underlying assets or some prespecified portfolio targeted by an investment fund. This valuable information shall complement the mean-variance portfolio. Relying on this idea we develop the Bayesian estimation of the GMV portfolio with priors for the portfolio weights. The results of this analysis are summarized in Section 4. By formalizing the beliefs regarding the desired portfolio in form of a prior distribution of portfolio weights, we provide a methodology for constructing the posterior distribution of the GMV portfolio weights. The next section provides details on the assumptions and the main results on posterior distributions.

3. Priors for the parameters of the asset returns

In this section we provide details on priors for the parameters of the asset returns, derive the posterior distribution of the GMV portfolio weights, and give expressions for the point estimates.

Diffuse prior: We start with the standard diffuse prior on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, applied in portfolio theory by Barry (1974); Brown (1976), and Klein and Bawa (1976). The diffuse prior implies that the investor has no initial information regarding the distribution of the characteristics of the assets. Thus the Bayesian estimator for the portfolio weights reflects this uncertainty. The prior densities of this non-informative prior is given by

$$p_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{k+1}{2}}. \tag{6}$$

Bayesian models based on the diffuse prior are usually not worse in comparison to classical methods of portfolio selection. However, when some of the k risky assets have longer histories than others, then Bayesian approaches may exploit this additional information and lead to different results (see Stambaugh, 1997).

Conjugate prior: The second considered prior is the conjugate prior. In contrast to the diffuse prior (6), the conjugate prior is an informative prior which considers a normal prior for $\boldsymbol{\mu}$ (conditional on $\boldsymbol{\Sigma}$) and an inverse Wishart prior for $\boldsymbol{\Sigma}$. Thus the investor has a prior opinion about the targeted values of expected returns and the covariance matrix. Formally this is expressed as

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

and

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

where $\boldsymbol{\mu}_c$ is the prior mean; κ_c is a parameter reflecting the prior precision of $\boldsymbol{\mu}_c$; ν_c is a similar prior precision parameter on $\boldsymbol{\Sigma}$; $\mathbf{S}_c$ is a known prior matrix of $\boldsymbol{\Sigma}$. Then the joint prior for both parameters is

$$p_c(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(\nu_c+1)/2} \exp\left\{-\frac{\kappa_c}{2}(\boldsymbol{\mu}-\boldsymbol{\mu}_c)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_c) - \frac{1}{2}\mathrm{tr}[\mathbf{S}_c\boldsymbol{\Sigma}^{-1}]\right\}. \tag{7}$$

Frost and Savarino (1986) proposed an interesting application of the conjugate prior where all securities possess identical expected returns, variances and pairwise correlation coefficients - the so-called $1/N$ rule. They showed that the conjugate prior works better than a non-informative prior as well as better than the strategies obtained from the frequentist point of view.

Hierarchical prior: Next, we consider the hierarchial Bayes model which was suggested by Greyserman et al. (2006). They demonstrated that a fully hierarchical Bayes procedure produces promising results warranting more study. The investor quantifies his uncertainty regarding the priors using additional stochastic parameters. The priors are given by

$$p_h(\boldsymbol{\mu}|\xi,\eta,\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{\kappa_h}{2}(\boldsymbol{\mu}-\xi\mathbf{1})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\xi\mathbf{1})\right\}$$

$$p_h(\boldsymbol{\Sigma}) \propto \frac{\eta^{-k(\nu_h-k-1)/2}}{|\boldsymbol{\Sigma}|^{\nu_h/2}} \exp\left\{-\frac{1}{2\eta}\mathrm{tr}[\mathbf{S}_h\boldsymbol{\Sigma}^{-1}]\right\}$$

$$p_h(\xi) \propto 1$$

$$p_h(\eta) \propto \eta^{-\varepsilon_1-1} \exp\left\{-\frac{\varepsilon_2}{\eta}\right\},$$

where κ_h is a parameter reflecting the prior precision of $\boldsymbol{\mu}$; ν_h is a similar prior precision parameter on $\boldsymbol{\Sigma}$; $\mathbf{S}_h$ is a known prior matrix of $\boldsymbol{\Sigma}$; ε_1 and ε_2 are prior constants.

Then the joint prior of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, ξ, and η is expressed as

$$\begin{aligned} p_h(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi, \eta) &\propto |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{\kappa_h}{2}(\boldsymbol{\mu}-\xi\mathbf{1})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\xi\mathbf{1})\right\} \\ &\times \frac{\eta^{-k(\nu_h-k-1)/2}}{|\boldsymbol{\Sigma}|^{\nu_h/2}} \exp\left\{-\frac{1}{2\eta}\mathrm{tr}[\mathbf{S}_h\boldsymbol{\Sigma}^{-1}]\right\} \\ &\times \eta^{-\varepsilon_1-1}\exp\left\{-\frac{\varepsilon_2}{\eta}\right\} \\ &\propto |\boldsymbol{\Sigma}|^{-(\nu_h+1)/2}\exp\left\{-\frac{1}{2\eta}\mathrm{tr}[\mathbf{S}_h\boldsymbol{\Sigma}^{-1}]\right\} \\ &\times \eta^{-k(\nu_h-k-1)/2-\varepsilon_1-1}\exp\left\{-\frac{\kappa_h}{2}(\boldsymbol{\mu} \right. \\ &\left. -\xi\mathbf{1})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\xi\mathbf{1})-\frac{\varepsilon_2}{\eta}\right\}. \end{aligned} \tag{8}$$

Let $t_p(m, \mathbf{a}, \mathbf{B})$ and $f_{t_p(m,\mathbf{a},\mathbf{B})}(\cdot)$ denote the distribution and the density of p-dimensional t-distribution with m degrees of freedom, location vector $\mathbf{a}$, and dispersion matrix $\mathbf{B}$. In Theorem 1 we present the posterior distributions of $\boldsymbol{\theta}$ under the diffuse, the conjugate and the hierarchial priors.

Theorem 1. *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n|\boldsymbol{\mu}, \boldsymbol{\Sigma}$ be independently and identically distributed with $\mathbf{X}_1|\boldsymbol{\mu}, \boldsymbol{\Sigma} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let $\mathbf{L}$ be a $p \times k$ matrix of constants, $p < k$ and $\mathbf{1}$ denotes the vector of ones. Then*

(a) *Under the diffuse prior $p_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ the posterior for $\boldsymbol{\theta}$ is given by*

$$\boldsymbol{\theta}|\mathbf{X}_1, \ldots, \mathbf{X}_n \sim t_p\left(n-1; \widehat{\boldsymbol{\theta}}; \frac{1}{n-1}\frac{\mathbf{L}\mathbf{R}_d\mathbf{L}^T}{\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1}}\right), \tag{9}$$

where $\mathbf{R}_d = \mathbf{S}^{-1} - \mathbf{S}^{-1}\mathbf{1}\mathbf{1}^T\mathbf{S}^{-1}/\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1}$.

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

$$\boldsymbol{\theta}|\mathbf{X}_1, \ldots, \mathbf{X}_n \sim t_p\left(\nu_c+n-k \right. \\ \left. -1; \frac{\mathbf{L}\mathbf{V}_c^{-1}\mathbf{1}}{\mathbf{1}^T\mathbf{V}_c^{-1}\mathbf{1}}; \frac{1}{\nu_c+n-k-1}\frac{\mathbf{L}\mathbf{R}_c\mathbf{L}^T}{\mathbf{1}^T\mathbf{V}_c^{-1}\mathbf{1}}\right), \tag{10}$$

where

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

(c) *Under the hierarchial prior $p_h(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi, \eta)$ the posterior for $\boldsymbol{\theta}$ is given by*

$$p_h(\boldsymbol{\theta}|\mathbf{X}_1, \ldots, \mathbf{X}_n) \propto \int_{-\infty}^{+\infty}\int_0^{+\infty} f_{t_p\left(\nu_h+n-k-1; \frac{\mathbf{L}\mathbf{V}_h^{-1}\mathbf{1}}{\mathbf{1}^T\mathbf{V}_h^{-1}\mathbf{1}}; \frac{1}{\nu_h+n-k-1}\frac{\mathbf{L}\mathbf{R}_h\mathbf{L}^T}{\mathbf{1}^T\mathbf{V}_h^{-1}\mathbf{1}}\right)}(\boldsymbol{\theta}) \\ \times \eta^{-k(\nu_h-k-1)/2-\varepsilon_1-1}\exp\left\{-\frac{\varepsilon_2}{\eta}\right\}\mathbf{d}\xi\,\mathbf{d}\eta, \tag{11}$$

where

$$\begin{aligned} \mathbf{r}_h &= \mathbf{r}_h(\xi) = \frac{n\overline{\mathbf{X}}+\kappa_h\xi\mathbf{1}}{n+\kappa_h}, \\ \mathbf{V}_h &= \mathbf{V}_h(\xi,\eta) = (n-1)\mathbf{S}+\eta^{-1}\mathbf{S}_h \\ &\quad -(n+\kappa_h)\mathbf{r}_h\mathbf{r}_h^T+n\overline{\mathbf{X}}\,\overline{\mathbf{X}}^T+\kappa_h\xi^2\mathbf{1}\mathbf{1}^T, \\ \mathbf{R}_h &= \mathbf{R}_h(\xi,\eta) = \mathbf{V}_h^{-1}-\mathbf{V}_h^{-1}\mathbf{1}\mathbf{1}^T\mathbf{V}_h^{-1}/\mathbf{1}^T\mathbf{V}_h^{-1}\mathbf{1}. \end{aligned}$$

The results of Theorem 1 show that under the diffuse and the conjugate priors the posterior distributions for the linear combinations of the GMV portfolio weights are multivariate t-distributions. Also, the posterior for the linear combinations of the GMV portfolio weights under the hierarchial prior is presented by using a two-dimensional integral and the well-known univariate density functions. Moreover, using (11) we get the stochastic representation of $\boldsymbol{\theta}$ expressed as

$$\boldsymbol{\theta} \stackrel{d}{=} \frac{\mathbf{L}\mathbf{V}_h^{-1}(\xi,\eta)\mathbf{1}}{\mathbf{1}^T\mathbf{V}_h^{-1}(\xi,\eta)\mathbf{1}} + \frac{1}{\sqrt{\nu_h+n-k-1}}\left(\frac{\mathbf{L}\mathbf{R}_h(\xi,\eta)\mathbf{L}^T}{\mathbf{1}^T\mathbf{V}_h^{-1}(\xi,\eta)\mathbf{1}}\right)^{1/2}\mathbf{t}_0, \tag{12}$$

where $\xi \sim \text{Uniform}(-\infty, +\infty)$, $\eta \sim \text{Inverse-Gamma}(\varepsilon_1, \varepsilon_2)$, $\mathbf{t}_0 \sim t_p(\nu_h+n-k-1, \mathbf{0}, \mathbf{I})$ and ξ, η, $\mathbf{t}_0$ are mutually independent. The symbol $\stackrel{d}{=}$ denotes equality in distribution.

Applying the properties of the multivariate t-distribution we obtain that the Bayesian estimators of $\boldsymbol{\theta}$ under the diffuse prior (5) and under the conjugate prior (6) are

$$\widehat{\boldsymbol{\theta}}_d = \widehat{\boldsymbol{\theta}} \quad \text{and} \quad \widehat{\boldsymbol{\theta}}_c = \frac{\mathbf{L}\mathbf{V}_c^{-1}\mathbf{1}}{\mathbf{1}^T\mathbf{V}_c^{-1}\mathbf{1}}, \tag{13}$$

respectively. Under the hierarchial prior the Bayesian estimator of $\boldsymbol{\theta}$ is given by

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_h &= \int_{R^p}\int_{-\infty}^{+\infty}\int_0^{+\infty}\boldsymbol{\theta}\,p_h(\boldsymbol{\theta}|\mathbf{X}_1,\ldots,\mathbf{X}_n)\mathbf{d}\boldsymbol{\theta}\mathbf{d}\xi\mathbf{d}\eta \\ &= \int_{-\infty}^{+\infty}\int_0^{+\infty}\eta^{-k(\nu_h-k-1)/2-\varepsilon_1-1}\exp\left\{-\frac{\varepsilon_2}{\eta}\right\}\frac{\mathbf{L}\mathbf{V}_h^{-1}(\xi,\eta)\mathbf{1}}{\mathbf{1}^T\mathbf{V}_h^{-1}(\xi,\eta)\mathbf{1}}\mathbf{d}\xi\mathbf{d}\eta. \end{aligned} \tag{14}$$

The last integral can be computed numerically. Alternatively, $\widehat{\boldsymbol{\theta}}_h$ can be approximated by using the stochastic representation (12). This is performed by generating a sample of independent pseudo random variables ξ and η with $\xi \sim \text{Uniform}(-\infty, +\infty)$ and $\eta \sim \text{Inverse-Gamma}(\varepsilon_1, \varepsilon_2)$, calculating $\boldsymbol{\theta}$ for each repetition using (12), and then taking the average.

The above results reveal the impact on the prior not only on the posterior distribution, but also on the point estimators. We observe that the point estimator based on the diffuse prior coincides with the classical estimator. This is consistent with our expectations, since the diffuse prior add no information, but merely reflect the uncertainty. The conjugate and hierarchical priors are informative priors. This leads to corrected versions of the point estimators which reflect the additional information in the priors.

Objective-based prior: Next, we consider the objective-based prior on $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ suggested by Tu and Zhou (2010). It is given by

$$\begin{aligned} p_{ob}(\boldsymbol{\mu}|\boldsymbol{\Sigma}) &\propto |\boldsymbol{\Sigma}|^{-1/2} \\ &\times \exp\left\{-\frac{s^2}{2\sigma_{ob}^2}(\boldsymbol{\mu}-\gamma\boldsymbol{\Sigma}\mathbf{w}_{ob})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\gamma\boldsymbol{\Sigma}\mathbf{w}_{ob})\right\} \\ p_{ob}(\boldsymbol{\Sigma}) &\propto |\boldsymbol{\Sigma}|^{-\nu_{ob}/2}\exp\left\{-\frac{1}{2}\mathrm{tr}[\mathbf{S}_{ob}\boldsymbol{\Sigma}^{-1}]\right\}, \end{aligned}$$

where γ is the coefficient of relative risk aversion; $\mathbf{w}_{ob}$ is a suitable prior constant; σ_{ob}^2 is a scale parameter that indicates the degree of uncertainty about $\boldsymbol{\mu}$; s^2 is the average of the diagonal elements of $\boldsymbol{\Sigma}$; ν_{ob} and $\mathbf{S}_{ob}$ are prior constants. Then the joint prior distribution of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is given by

$$\begin{aligned} p_{ob}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &\propto |\boldsymbol{\Sigma}|^{-(\nu_{ob}+1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}[\mathbf{S}_{ob}\boldsymbol{\Sigma}^{-1}]\right\} \\ &\times \exp\left\{-\frac{s^2}{2\sigma_{ob}^2}(\boldsymbol{\mu}-\gamma\boldsymbol{\Sigma}\mathbf{w}_{ob})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\gamma\boldsymbol{\Sigma}\mathbf{w}_{ob})\right\}, \end{aligned} \tag{15}$$

which leads to the posterior distribution of $(\boldsymbol{\mu},\boldsymbol{\Sigma})$ expressed as

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

where (see Appendix A)

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

Integrating out $\boldsymbol{\mu}$ we get the posterior distribution of $\boldsymbol{\Sigma}$ expressed as

$$
\begin{aligned}
p_{ob}(\boldsymbol{\Sigma}|\mathbf{X}_1,\dots,\mathbf{X}_n) \propto |\boldsymbol{\Sigma}|^{-(\nu_{ob}+n)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}[(\mathbf{S}_{ob}+(n-1)\mathbf{S})\boldsymbol{\Sigma}^{-1}]\right\}\\
\times\exp\left\{-\frac{1}{2}\left[n\overline{\mathbf{X}}^T\boldsymbol{\Sigma}^{-1}\overline{\mathbf{X}}+\frac{s^2}{\sigma_{ob}^2}\mathbf{w}_{ob}^T\boldsymbol{\Sigma}\mathbf{w}_{ob}\right.\right.\\
\left.\left.-\left(n+\frac{s^2}{\sigma_{ob}^2}\right)\mathbf{r}_{ob}^T\boldsymbol{\Sigma}^{-1}\mathbf{r}_{ob}\right]\right\},
\end{aligned}
\tag{16}
$$

where

$$
\mathbf{r}_{ob}=\frac{\frac{s^2}{\sigma_{ob}^2}\gamma\boldsymbol{\Sigma}\mathbf{w}_{ob}+n\overline{\mathbf{X}}}{\frac{s^2}{\sigma_{ob}^2}+n}.
$$

Unfortunately, using the objective-based prior (15) we are not able to derive the analytical expression for the posterior distribution for $\boldsymbol{\theta}$. Theoretically, the posterior of $\boldsymbol{\theta}$ can be obtained by making the transformation

$$
\boldsymbol{\Sigma}\rightsquigarrow\begin{pmatrix}\boldsymbol{\theta}\\ \mathbf{v}\end{pmatrix},
$$

where $\boldsymbol{\theta}\in R^p$ and $\mathbf{v}\in R^{\frac{k(k-1)}{2}-p}$, and integrating out $\mathbf{v}$. However, because $p_{ob}(\boldsymbol{\Sigma}|\mathbf{X}_1,\dots,\mathbf{X}_n)$ is a complicated function of $\boldsymbol{\theta}$, this leads to a difficult multiple integral with respect to $\mathbf{v}$. As a result, the Bayesian estimation of $\boldsymbol{\theta}$ is obtained via simulations based on (16).

Tu and Zhou (2010) demonstrated that the portfolio strategies based on the objective-based prior work better than the strategies under other priors. In particular, they proposed the application of the objective-based prior to the portfolio weights of the general mean-variance portfolio and reported good results.

4. Priors for the GMV portfolio weights

In the previous section we concentrated on statistical models for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, which were subsequently used to derive the posterior distributions of a linear combination of portfolio weights. Thus we specified prior information on $k+k(k+1)/2$ parameters to make an inference about $\boldsymbol{\theta}$ of dimension p. In this section we reparameterize the model to make statements directly about the priors of the portfolio weights. This procedure is also more natural from a decision making perspective since investors sometimes have some perception of the optimal or preferred portfolio composition.

More specifically, we consider a reparameterized model for the asset returns which is used to derive an informative prior and a non-informative prior for the linear combinations of the GMV portfolio weights. We provide explicit formulas for the corresponding posterior distributions in the next step. It is noted that the posteriors derived under the reparameterized model are usually the same as the posteriors obtained from the original model since for any one-to-one mapping $\boldsymbol{\varphi}=\boldsymbol{\varphi}(\boldsymbol{\theta})$, the posterior $p(\boldsymbol{\varphi}|\mathbf{X}_1,\dots,\mathbf{X}_n)$ obtained from the reparameterized model $p(\mathbf{X}_1,\dots,\mathbf{X}_n|\boldsymbol{\varphi},\boldsymbol{\lambda})$ must be coherent with the posterior $p(\boldsymbol{\theta}|\mathbf{X}_1,\dots,\mathbf{X}_n)$ calculated from the original model $p(\mathbf{X}_1,\dots,\mathbf{X}_n|\boldsymbol{\theta},\boldsymbol{\lambda})$. Moreover, if the model has a sufficient statistic $\mathbf{t}=\mathbf{t}(\mathbf{X})$, then the posterior $p(\boldsymbol{\theta}|\mathbf{X}_1,\dots,\mathbf{X}_n)$ derived from the full model $p(\mathbf{X}_1,\dots,\mathbf{X}_n|\boldsymbol{\theta},\boldsymbol{\lambda})$ is the same as the posterior $p(\boldsymbol{\theta}|\mathbf{t})$ obtained from the equivalent model $p(\mathbf{t}|\boldsymbol{\theta},\boldsymbol{\lambda})$ (cf. Bernardo, 2005, p.5)).

4.1. Non-informative prior

We begin with the Jeffreys non-informative prior. Similarly, as for other non-informative priors, we assume no initial information about the weights. Formally, this lack of beliefs cannot be stated directly and is specified for an arbitrary parameter vector as the square root of the determinant of the Fisher information matrix. Using this prior we compute the corresponding posterior distribution for the weights of the GMV portfolio. Let $\widetilde{\mathbf{L}}=(\mathbf{L}^T,\mathbf{1})^T$, $\widetilde{\boldsymbol{\Sigma}}=\widetilde{\mathbf{L}}\boldsymbol{\Sigma}^{-1}\widetilde{\mathbf{L}}^T$, $\zeta=\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}$, $\widetilde{\mathbf{S}}=\widetilde{\mathbf{L}}\mathbf{S}^{-1}\widetilde{\mathbf{L}}^T$, and $\boldsymbol{\Psi}=\mathbf{L}\boldsymbol{\Sigma}^{-1}\mathbf{L}^T-\frac{\mathbf{L}\boldsymbol{\Sigma}^{-1}\mathbf{1}\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{L}^T}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}$. Since

$$
\widetilde{\boldsymbol{\Sigma}}=\begin{bmatrix}\mathbf{L}\boldsymbol{\Sigma}^{-1}\mathbf{L}^T & \mathbf{L}\boldsymbol{\Sigma}^{-1}\mathbf{1}\\ \mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{L}^T & \mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}\end{bmatrix}=\zeta\begin{bmatrix}\boldsymbol{\Psi}/\zeta+\boldsymbol{\theta}\boldsymbol{\theta}^T & \boldsymbol{\theta}\\ \boldsymbol{\theta}^T & 1\end{bmatrix}
\tag{17}
$$

we get that

$$
|\widetilde{\boldsymbol{\Sigma}}|=\zeta\left|\mathbf{L}\boldsymbol{\Sigma}^{-1}\mathbf{L}^T-\frac{\mathbf{L}\boldsymbol{\Sigma}^{-1}\mathbf{1}\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{L}^T}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}\right|=\zeta|\boldsymbol{\Psi}|.
\tag{18}
$$

Since $(n-1)\mathbf{S}|\boldsymbol{\Sigma}\sim W_k(n-1,\boldsymbol{\Sigma})$ (k-dimensional Wishart distribution with $n-1$ degrees of freedom and covariance matrix $\boldsymbol{\Sigma}$) and $\mathrm{rank}(\widetilde{\mathbf{L}})=p+1$ we get from Theorem 3.2.11 of Muirhead (1982) that

$$
(n-1)\widetilde{\mathbf{S}}^{-1}|\boldsymbol{\Sigma}\sim W_{p+1}\left(n+p-k,\widetilde{\boldsymbol{\Sigma}}^{-1}\right).
$$

From the properties of the Wishart distribution (see Muirhead, 1982) it holds that

$$
(n-1)^{-1}\widetilde{\mathbf{S}}|\boldsymbol{\Sigma}\sim IW_{p+1}\left(n-k+2(p+1),\widetilde{\boldsymbol{\Sigma}}\right).
$$

This shows that $\widetilde{\mathbf{S}}$ is a sufficient statistic for $\widetilde{\boldsymbol{\Sigma}}$. Then the posterior $p_n(\boldsymbol{\theta}|\mathbf{X}_1,\dots,\mathbf{X}_n)$ obtained from the full model coincides with the posterior $p_n(\boldsymbol{\theta}|(n-1)^{-1}\widetilde{\mathbf{S}})$ calculated from the equivalent model (cf. Bernardo, 2005, p. 5)).

Next, we rewrite the likelihood function in terms of $(\boldsymbol{\theta},\boldsymbol{\Psi},\zeta)$. It holds that

$$
L\left((n-1)^{-1}\widetilde{\mathbf{S}}\middle|\boldsymbol{\theta},\boldsymbol{\Psi},\zeta\right)\propto|\widetilde{\boldsymbol{\Sigma}}|^{(n-k+p)/2}\mathrm{etr}\left[-\frac{n-1}{2}\widetilde{\boldsymbol{\Sigma}}\widetilde{\mathbf{S}}^{-1}\right].
\tag{19}
$$

Using (18) and

$$
\begin{aligned}
\mathrm{tr}[\widetilde{\mathbf{S}}^{-1}\widetilde{\boldsymbol{\Sigma}}]&=\zeta\,\mathrm{tr}\left(\begin{bmatrix}\widetilde{\mathbf{S}}_{11}^{(-)} & \widetilde{\mathbf{S}}_{12}^{(-)}\\ \widetilde{\mathbf{S}}_{21}^{(-)} & \widetilde{S}_{22}^{(-)}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Psi}/\zeta+\boldsymbol{\theta}\boldsymbol{\theta}^T & \boldsymbol{\theta}\\ \boldsymbol{\theta}^T & 1\end{bmatrix}\right)\\
&=\mathrm{tr}[\widetilde{\mathbf{S}}_{11}^{(-)}\boldsymbol{\Psi}]+\zeta\left(\mathrm{tr}[\widetilde{\mathbf{S}}_{11}^{(-)}\boldsymbol{\theta}\boldsymbol{\theta}^T]+2\mathrm{tr}[\widetilde{\mathbf{S}}_{12}^{(-)}\boldsymbol{\theta}^T]+\widetilde{S}_{22}^{(-)}\right),
\end{aligned}
$$

where

$$
\widetilde{\mathbf{S}}_{11}^{(-)}=\left(\mathbf{L}\mathbf{S}^{-1}\mathbf{L}^T-\frac{\mathbf{L}\mathbf{S}^{-1}\mathbf{1}\mathbf{1}^T\mathbf{S}^{-1}\mathbf{L}^T}{\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1}}\right)^{-1},
$$

$$
\widetilde{\mathbf{S}}_{12}^{(-)}=-\widetilde{\mathbf{S}}_{11}^{(-)}\frac{\mathbf{L}\mathbf{S}^{-1}\mathbf{1}}{\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1}},\quad\widetilde{\mathbf{S}}_{21}^{(-)}=\left[\widetilde{\mathbf{S}}_{12}^{(-)}\right]^T,
$$

$$
\widetilde{S}_{22}^{(-)}=(\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1})^{-1}+\frac{\mathbf{1}^T\mathbf{S}^{-1}\mathbf{L}^T\widetilde{\mathbf{S}}_{11}^{(-)}\mathbf{L}\mathbf{S}^{-1}\mathbf{1}}{(\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1})^2},
$$

we get

$$\log L((n-1)^{-1}\widetilde{\mathbf{S}}|\boldsymbol{\theta},\boldsymbol{\Psi},\zeta) \propto \frac{n-k+p}{2}\log|\boldsymbol{\Psi}| + \frac{n-k+p}{2}\log\zeta$$
$$-\frac{n-1}{2}\mathrm{tr}[\widetilde{\mathbf{S}}_{11}^{(-)}\boldsymbol{\Psi}] - \frac{\zeta(n-1)}{2}\left(\mathrm{tr}[\widetilde{\mathbf{S}}_{11}^{(-)}\boldsymbol{\theta}\boldsymbol{\theta}^T] + 2\mathrm{tr}[\widetilde{\mathbf{S}}_{12}^{(-)}\boldsymbol{\theta}^T] + \widetilde{S}_{22}^{(-)}\right). \tag{20}$$

Let $\boldsymbol{\phi} = (\boldsymbol{\theta}^T, \mathbf{vech}(\boldsymbol{\Psi})^T, \zeta)^T$. Then the Fisher information matrix $\mathbf{I}(\boldsymbol{\theta},\boldsymbol{\Psi},\zeta)$ for the parameters $(\boldsymbol{\theta},\boldsymbol{\Psi},\zeta)$ is given by (see Appendix B)

$$\mathbf{I}(\boldsymbol{\theta},\boldsymbol{\Psi},\zeta = -E\left[\frac{\partial^2 \log L((n-1)^{-1}\widetilde{\mathbf{S}}|\boldsymbol{\theta},\boldsymbol{\Psi},\zeta)}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}^T}\right]$$
$$= \begin{bmatrix} (n-k+p)\zeta\boldsymbol{\Psi}^{-1} & \mathbf{0}_{p\times p(p+1)/2} & \mathbf{0}_p \\ \mathbf{0}_{p(p+1)/2\times p} & \frac{n-k+p}{2}\mathbf{G}_p^T(\boldsymbol{\Psi}^{-1}\otimes\boldsymbol{\Psi}^{-1})\mathbf{G}_p & \mathbf{0}_{p(p+1)/2} \\ \mathbf{0}_p^T & \mathbf{0}_{p(p+1)/2}^T & \frac{n-k+p}{2}\zeta^{-2} \end{bmatrix},$$

where $\mathbf{G}_p$ is the duplication matrix defined by $\mathbf{vec}(\mathbf{B}) = \mathbf{G}_p\mathbf{vech}(\mathbf{B})$ for any symmetric $\mathbf{B}(p\times p)$; **vec** denotes the operator which transforms a matrix into a vector by stacking the columns of the matrix; **vech** stands for the operator that takes a symmetric $p\times p$ matrix and stacks the lower triangular half into a single vector of length $p(p+1)/2$; $\mathbf{0}_{p\times p}$ is the $p\times p$ null matrix and $\mathbf{0}_p$ denotes the p-dimensional null vector.

Since (see, e.g. Magnus & Neudecker, 2007)

$$|(\mathbf{G}_p^T\mathbf{G}_p)^{-1}\mathbf{G}_p^T(\boldsymbol{\Psi}^{-1}\otimes\boldsymbol{\Psi}^{-1})\mathbf{G}_p(\mathbf{G}_p^T\mathbf{G}_p)^{-1}| = 2^{-p(p-1)/2}|\boldsymbol{\Psi}|^{-(p+1)}$$

we get that

$$|\mathbf{I}(\boldsymbol{\theta},\boldsymbol{\Psi},\zeta)| \propto \zeta^{p-2}|\boldsymbol{\Psi}|^{-p-2}.$$

Hence, the Jeffreys prior for $(\boldsymbol{\theta},\boldsymbol{\Psi},\zeta)$ is given by

$$p_n(\boldsymbol{\theta},\boldsymbol{\Psi},\zeta) \propto \zeta^{p/2-1}|\boldsymbol{\Psi}|^{-p/2-1}. \tag{21}$$

Using the Jeffreys prior (21) we obtain the posterior distribution of $(\boldsymbol{\theta},\boldsymbol{\Psi},\zeta)$ expressed as

$$p_n\left(\boldsymbol{\theta},\boldsymbol{\Psi},\zeta\,\Big|(n-1)^{-1}\widetilde{\mathbf{S}}\right) \propto L\left((n-1)^{-1}\widetilde{\mathbf{S}}\Big|\boldsymbol{\theta},\boldsymbol{\Psi},\zeta\right)p_n\left(\boldsymbol{\theta},\boldsymbol{\Psi},\zeta\right)$$
$$\propto \zeta^{(n-k+2p)/2-1}$$
$$\times\exp\left\{-\frac{\zeta(n-1)}{2}\left(\mathrm{tr}[\widetilde{\mathbf{S}}_{11}^{(-)}\boldsymbol{\theta}\boldsymbol{\theta}^T] + 2\mathrm{tr}[\widetilde{\mathbf{S}}_{12}^{(-)}\boldsymbol{\theta}^T] + \widetilde{S}_{22}^{(-)}\right)\right\}$$
$$\times|\boldsymbol{\Psi}|^{(n-k)/2-1}\mathrm{etr}\left\{-\frac{n-1}{2}\widetilde{\mathbf{S}}_{11}^{(-)}\boldsymbol{\Psi}\right\}.$$

Integrating out $\boldsymbol{\Psi}$ and ζ, the posterior distribution for $\boldsymbol{\theta}$ equals

$$p_n\left(\boldsymbol{\theta}\Big|(n-1)^{-1}\widetilde{\mathbf{S}}\right) \propto \left(\mathrm{tr}[\widetilde{\mathbf{S}}_{11}^{(-)}\boldsymbol{\theta}\boldsymbol{\theta}^T] + 2\mathrm{tr}[\widetilde{\mathbf{S}}_{12}^{(-)}\boldsymbol{\theta}^T] + \widetilde{S}_{22}^{(-)}\right)^{-(n-k+2p)/2}$$
$$\propto \left(\widetilde{S}_{22}^{(-)} - \left(\widetilde{\mathbf{S}}_{12}^{(-)}\right)^T\left(\widetilde{\mathbf{S}}_{11}^{(-)}\right)^{-1}\widetilde{\mathbf{S}}_{12}^{(-)}\right.$$
$$\left.+\left(\boldsymbol{\theta}+\left(\widetilde{\mathbf{S}}_{11}^{(-)}\right)^{-1}\widetilde{\mathbf{S}}_{12}^{(-)}\right)^T\widetilde{\mathbf{S}}_{11}^{(-)}\left(\boldsymbol{\theta}+\left(\widetilde{\mathbf{S}}_{11}^{(-)}\right)^{-1}\widetilde{\mathbf{S}}_{12}^{(-)}\right)\right)^{-(n-k+2p)/2}$$
$$\propto t_p\left(n-k+p; -\left(\widetilde{\mathbf{S}}_{11}^{(-)}\right)^{-1}\widetilde{\mathbf{S}}_{12}^{(-)};\right.$$
$$\left.\times\frac{\widetilde{S}_{22}^{(-)} - \left(\widetilde{\mathbf{S}}_{12}^{(-)}\right)^T\left(\widetilde{\mathbf{S}}_{11}^{(-)}\right)^{-1}\widetilde{\mathbf{S}}_{12}^{(-)}}{n-k+p}\left(\widetilde{\mathbf{S}}_{11}^{(-)}\right)^{-1}\right).$$

Rewriting the location vector and the dispersion matrix of the multivariate t-distribution by using

$$-(\widetilde{\mathbf{S}}_{11}^{(-)})^{-1}\widetilde{\mathbf{S}}_{12}^{(-)} = \widehat{\boldsymbol{\theta}}, \tag{22}$$

$$\widetilde{\mathbf{S}}_{11}^{(-)} = (\mathbf{L}\mathbf{R}_d\mathbf{L}^T)^{-1}, \tag{23}$$

$$\widetilde{S}_{22}^{(-)} - \left(\widetilde{\mathbf{S}}_{12}^{(-)}\right)^T\left(\widetilde{\mathbf{S}}_{11}^{(-)}\right)^{-1}\widetilde{\mathbf{S}}_{12}^{(-)} = (\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1})^{-1} \tag{24}$$

leads to the following result.

Theorem 2. *Let $\mathbf{X}_1,\ldots,\mathbf{X}_n|\boldsymbol{\mu},\boldsymbol{\Sigma}$ be independently and identically distributed with $\mathbf{X}_i|\boldsymbol{\mu},\boldsymbol{\Sigma}\sim N_k(\boldsymbol{\mu},\boldsymbol{\Sigma})$. Let $\mathbf{L}$ be a $p\times k$ matrix of constants with $p<k$. Then the posterior for the GMV portfolio weights $\boldsymbol{\theta}$ under the Jeffreys non-informative prior $p_n(\boldsymbol{\theta},\boldsymbol{\Psi},\zeta)$ is given by*

$$\boldsymbol{\theta}|\mathbf{X}_1,\ldots,\mathbf{X}_n \sim t_p\left(n-k+p; \widehat{\boldsymbol{\theta}}; \frac{1}{n-k+p}\frac{\mathbf{L}\mathbf{R}_d\mathbf{L}^T}{\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1}}\right). \tag{25}$$

Theorem 2 shows that the posterior for the GMV portfolio weights under the Jeffreys non-informative prior $p_n(\boldsymbol{\theta},\boldsymbol{\Psi},\zeta)$ has a p-variate t-distribution with $n-k+p$ degrees of freedom, location vector $\widehat{\boldsymbol{\theta}}$ and dispersion matrix $\frac{1}{n-k+p}\frac{\mathbf{L}\mathbf{R}_d\mathbf{L}}{\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1}}$. This result is similar to the one obtained for the diffuse prior. The difference is present in the degrees of freedom of the t-distribution only.

Applying the properties of the multivariate t-distribution we get that the Bayesian estimation of $\boldsymbol{\theta}$ under the non-informative prior (21) is

$$\widehat{\boldsymbol{\theta}}_n = \widehat{\boldsymbol{\theta}},$$

which is the same as under the diffuse prior (6).

4.2. Informative prior

In this section we construct an informative prior for the GMV weights obtained under a hierarchical Bayesian model. Tunaru (2002) developed a multiple response model for counts which is specified hierarchically and belongs to the fully Bayesian family. A similar hierarchical model is considered here.

The suggested informative prior is given by

$$\boldsymbol{\theta} \sim N_p\left(\mathbf{w}_I, \frac{1}{\zeta}\boldsymbol{\Psi}^{-1}\right)$$
$$\boldsymbol{\Psi} \sim W_p(\nu_I, \mathbf{S}_I)$$
$$\zeta \sim \mathrm{Gamma}(\delta_1, 2\delta_2),$$

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

$$p_I\left(\boldsymbol{\theta},\boldsymbol{\Psi},\zeta\right) \propto \left|\frac{1}{\zeta}\boldsymbol{\Psi}^{-1}\right|^{-1/2}\exp\left\{-\frac{\zeta}{2}(\boldsymbol{\theta}-\mathbf{w}_I)^T\boldsymbol{\Psi}(\boldsymbol{\theta}-\mathbf{w}_I)\right\}$$
$$\times\zeta^{\delta_1-1}|\boldsymbol{\Psi}|^{(\nu_I-p-1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}[\mathbf{S}_I^{-1}\boldsymbol{\Psi}] - \frac{\zeta}{2\delta_2}\right\}. \tag{26}$$

Relying in this prior and integrating out the $\boldsymbol{\Psi}$ and ζ we obtain the posterior distribution of the portfolio weights, which is summarized in the following theorem. The proof can be found in the appendix.

Theorem 3. *Let $\mathbf{X}_1,\ldots,\mathbf{X}_n|\boldsymbol{\mu},\boldsymbol{\Sigma}$ be independently and identically distributed with $\mathbf{X}_i|\boldsymbol{\mu},\boldsymbol{\Sigma}\sim N_k(\boldsymbol{\mu},\boldsymbol{\Sigma})$. Let $\mathbf{L}$ be a $p\times k$ matrix of constants with $p<k$. Then the posterior for $\boldsymbol{\theta}$ under the informative prior $p_I(\boldsymbol{\theta},\boldsymbol{\Psi},\zeta)$ is given by*

$$p_I(\boldsymbol{\theta}|\mathbf{X}_1,\ldots,\mathbf{X}_n) \propto [(\boldsymbol{\theta}-\mathbf{w}_I)^T(\mathbf{S}_I^{-1}$$
$$+(n-1)(\mathbf{L}\mathbf{R}_d\mathbf{L}^T)^{-1})^{-1}(\boldsymbol{\theta}-\mathbf{w}_I)]^{(n-k+2p+2\delta_1)/2}$$
$$\times U((n-k+2p+2\delta_1)/2; (p+2\delta_1$$
$$-\nu_I+1)/2; g(\boldsymbol{\theta})) \tag{27}$$

where $g(\boldsymbol{\theta})$ is given in (43).

Theorem 3 shows that the posterior for the GMV portfolio weights under the informative prior $p_I(\boldsymbol{\theta},\boldsymbol{\Psi},\zeta)$ is given using a well-known special mathematical function. Using (27), the

Bayesian estimator of $\boldsymbol{\theta}$ is obtained

$$\widehat{\boldsymbol{\theta}}_I = \int_{R^p} \boldsymbol{\theta} p_I(\boldsymbol{\theta}|\mathbf{X}_1, \ldots, \mathbf{X}_n)\mathbf{d}\boldsymbol{\theta}, \tag{28}$$

which is a p-dimensional integral. This integral can be evaluated numerically.

Next, we derive another expression for $\widehat{\boldsymbol{\theta}}_I$ which is based on a one-dimensional integral independent of p. Using

$$b^{-a} \propto \int_0^{+\infty} \tau^{a-1} e^{-b\tau/2}\mathbf{d}\tau$$

and (41), the posterior distribution under the informative prior is given by

$$\begin{aligned}
p_I\Big(\boldsymbol{\theta}, \zeta \Big| (n-1)\widetilde{\mathbf{S}}\Big) &\propto \int_0^{+\infty} \tau^{(n-k+p+\nu_I-1)/2} \zeta^{(n-k+2p+2\delta_1-2)/2} \\
&\times \exp\Bigg\{ -\frac{(n-1)\zeta}{2} \Bigg(\frac{\delta_2^{-1}}{n-1} \\
&+ \mathbf{1}^T\mathbf{S}^{-1}\mathbf{1} + (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})^T (\mathbf{L}\mathbf{R}_d\mathbf{L}^T)^{-1} (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}) \Bigg) \Bigg\} \\
&\times \exp\Bigg\{ -\frac{\tau}{2} [1 + \zeta (\boldsymbol{\theta} - \mathbf{w}_I)^T (\mathbf{S}_I^{-1} \\
&+ (n-1)(\mathbf{L}\mathbf{R}_d\mathbf{L}^T)^{-1})^{-1} (\boldsymbol{\theta} - \mathbf{w}_I)] \Bigg\} \mathbf{d}\tau.
\end{aligned}$$

Let

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

Then

$$\begin{aligned}
&p_I(\boldsymbol{\theta}, \zeta, \tau | (n-1)^{-1}\widetilde{\mathbf{S}}) \\
&\propto \exp\left\{ -\frac{1}{2} (\boldsymbol{\theta} - \mathbf{r}_I(\tau))^T \left(\frac{1}{\zeta} \mathbf{V}_I(\tau) \right)^{-1} (\boldsymbol{\theta} - \mathbf{r}_I(\tau)) \right\} \\
&\times \zeta^{(n-k+2p+2\delta_1-2)/2} \exp\left\{ -\frac{\zeta}{2} \mathbf{h}_I(\tau) \right\} \\
&\times \tau^{(n-k+p+\nu_I-1)/2} \exp\left\{ -\frac{\tau}{2} \right\}.
\end{aligned} \tag{29}$$

Using (29) we get a very useful stochastic representation for $\boldsymbol{\theta}$ expressed as

$$\boldsymbol{\theta} \stackrel{d}{=} \mathbf{r}_I(\tau) + \zeta^{-1/2} (\mathbf{V}_I(\tau))^{1/2} \mathbf{z}_0, \tag{30}$$

where

$$\mathbf{z}_0 \sim N_p(\mathbf{0}_p, \mathbf{I}_p), \tag{31}$$

$$\zeta | \tau \sim \text{Gamma}\left((n-k+2p+2\delta_1)/2, \frac{2}{h_I(\tau)} \right), \tag{32}$$

$$\tau \sim \text{Gamma}((n-k+p+\nu_I-1)/2, 2). \tag{33}$$

The application of (30) leads to

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}_I &= E(\boldsymbol{\theta}|\mathbf{X}_1, \ldots, \mathbf{X}_n) = E(E(\boldsymbol{\theta}|\tau, \zeta, \mathbf{X}_1, \ldots, \mathbf{X}_n)|\mathbf{X}_1, \ldots, \mathbf{X}_n) \\
&= E(\mathbf{r}_I(\tau)|\mathbf{X}_1, \ldots, \mathbf{X}_n) \\
&= \int_0^{+\infty} (\tau\mathbf{P}_1 + \mathbf{P}_2)^{-1} (\mathbf{P}_2\widehat{\boldsymbol{\theta}} + \tau\mathbf{P}_1\mathbf{w}_I) f_{\text{Gamma}((n-k+p+\nu_I+1)/2,2)}(\tau)\mathbf{d}\tau,
\end{aligned}$$

which is a one-dimensional integral and can easily approximated numerically. Finally, we note that $\widehat{\boldsymbol{\theta}}_I$ ca also be approximated by using the stochastic representation (30). This is achieved by drawing a sample of $\mathbf{z}_0$, ζ, and τ with the joint distribution as specified in (31)–(33), calculating $\boldsymbol{\theta}$ from (30), and then taking the average.

5. Credible sets

In this section we derive credible sets for the GMV portfolio weights based on the posterior distributions obtained in the previous sections. Note that credible sets in Bayesian statistics are analogous to confidence sets in frequentist statistics. Thus it is an interval estimator derived form the posterior distribution.

5.1. Credible intervals for a GMV portfolio weight

Without loss of generality we deal with the first weight of the GMV portfolio only and note that the credible intervals for other weights can be obtained similarly. Let $\mathbf{L} = \mathbf{e}_1^T = (1, 0, \ldots, 0)$. Then under the diffuse prior (6) the posterior for $\theta = \mathbf{e}_1^T\boldsymbol{\Sigma}^{-1}\mathbf{1}/\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}$ is expressed as

$$\theta | \mathbf{X}_1, \ldots, \mathbf{X}_n \sim t\left(n-1; \frac{\mathbf{e}_1^T\mathbf{S}^{-1}\mathbf{1}}{\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1}}; \frac{1}{n-1} \frac{\mathbf{e}_1^T\mathbf{R}_d\mathbf{e}_1}{\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1}} \right). \tag{34}$$

Let $t_{m;\,\beta}$ be the β-quantile of the t-distribution with m degrees of freedom. The application of (9) leads to the $(1-\alpha)$-credible interval C_d for the first weight of the GMV portfolio given by

$$\begin{aligned}
C_d = \Bigg[&\frac{\mathbf{e}_1^T\mathbf{S}^{-1}\mathbf{1}}{\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1}} - \frac{1}{\sqrt{n-1}} \frac{\sqrt{\mathbf{e}_1^T\mathbf{R}_d\mathbf{e}_1}}{\sqrt{\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1}}} t_{n-1;\alpha/2};\ \frac{\mathbf{e}_1^T\mathbf{S}^{-1}\mathbf{1}}{\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1}} \\
&+ \frac{1}{\sqrt{n-1}} \frac{\sqrt{\mathbf{e}_1^T\mathbf{R}_d\mathbf{e}_1}}{\sqrt{\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1}}} t_{n-1;1-\alpha/2} \Bigg].
\end{aligned} \tag{35}$$

Similarly, under the conjugate prior (7) the $(1-\alpha)$-credible interval C_c of θ is

$$\begin{aligned}
C_c = \Bigg[&\frac{\mathbf{e}_1^T\mathbf{V}_c^{-1}\mathbf{1}}{\mathbf{1}^T\mathbf{V}_c^{-1}\mathbf{1}} - \frac{1}{\sqrt{\nu_c+n-k-1}} \frac{\sqrt{\mathbf{e}_1^T\mathbf{R}_c\mathbf{e}_1}}{\sqrt{\mathbf{1}^T\mathbf{V}_c^{-1}\mathbf{1}}} t_{\nu_c+n-k-1;\alpha/2}; \\
&\frac{\mathbf{e}_1^T\mathbf{V}_c^{-1}\mathbf{1}}{\mathbf{1}^T\mathbf{V}_c^{-1}\mathbf{1}} + \frac{1}{\sqrt{\nu_c+n-k-1}} \frac{\sqrt{\mathbf{e}_1^T\mathbf{R}_c\mathbf{e}_1}}{\sqrt{\mathbf{1}^T\mathbf{V}_c^{-1}\mathbf{1}}} t_{\nu_c+n-k-1;1-\alpha/2} \Bigg],
\end{aligned} \tag{36}$$

while under the non-informative prior (21) it is given by

$$\begin{aligned}
C_n = \Bigg[&\frac{\mathbf{e}_1^T\mathbf{S}^{-1}\mathbf{1}}{\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1}} - \frac{1}{\sqrt{n-k+p}} \frac{\sqrt{\mathbf{e}_1^T\mathbf{R}_d\mathbf{e}_1}}{\sqrt{\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1}}} t_{n-k+p;\alpha/2}; \\
&\frac{\mathbf{e}_1^T\mathbf{S}^{-1}\mathbf{1}}{\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1}} + \frac{1}{\sqrt{n-k+p}} \frac{\sqrt{\mathbf{e}_1^T\mathbf{R}_d\mathbf{e}_1}}{\sqrt{\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1}}} t_{n-k+p;1-\alpha/2} \Bigg].
\end{aligned} \tag{37}$$

Under the hierarchial prior (8) and the informative prior (26), the $(1-\alpha)$-credible intervals C_h and C_I for the GMV portfolio weight are given by

$$C_h = \left[q_{h;\alpha/2}(\theta|\mathbf{X}_1, \ldots, \mathbf{X}_n); q_{h;1-\alpha/2}(\theta|\mathbf{X}_1, \ldots, \mathbf{X}_n) \right] \tag{38}$$

and

$$C_I = \left[q_{I;\alpha/2}(\theta|\mathbf{X}_1, \ldots, \mathbf{X}_n); q_{I;1-\alpha/2}(\theta|\mathbf{X}_1, \ldots, \mathbf{X}_n) \right], \tag{39}$$

where $q_{h;\beta}(\theta; \mathbf{X}_1, \ldots, \mathbf{X}_n)$ is the β-quantile of the posterior for a GMV portfolio weight (11) under the hierarchial prior (8); $q_{I;\beta}(\theta; \mathbf{X}_1, \ldots, \mathbf{X}_n)$ is the β-quantile of the posterior for a GMV portfolio weight (27) under the informative prior (26). The quantiles for both posteriors $p_h(\theta|\mathbf{X}_1, \ldots, \mathbf{X}_n)$ and $p_I(\theta|\mathbf{X}_1, \ldots, \mathbf{X}_n)$ are obtained via simulations by using the stochastic representations (12) and (30), respectively.

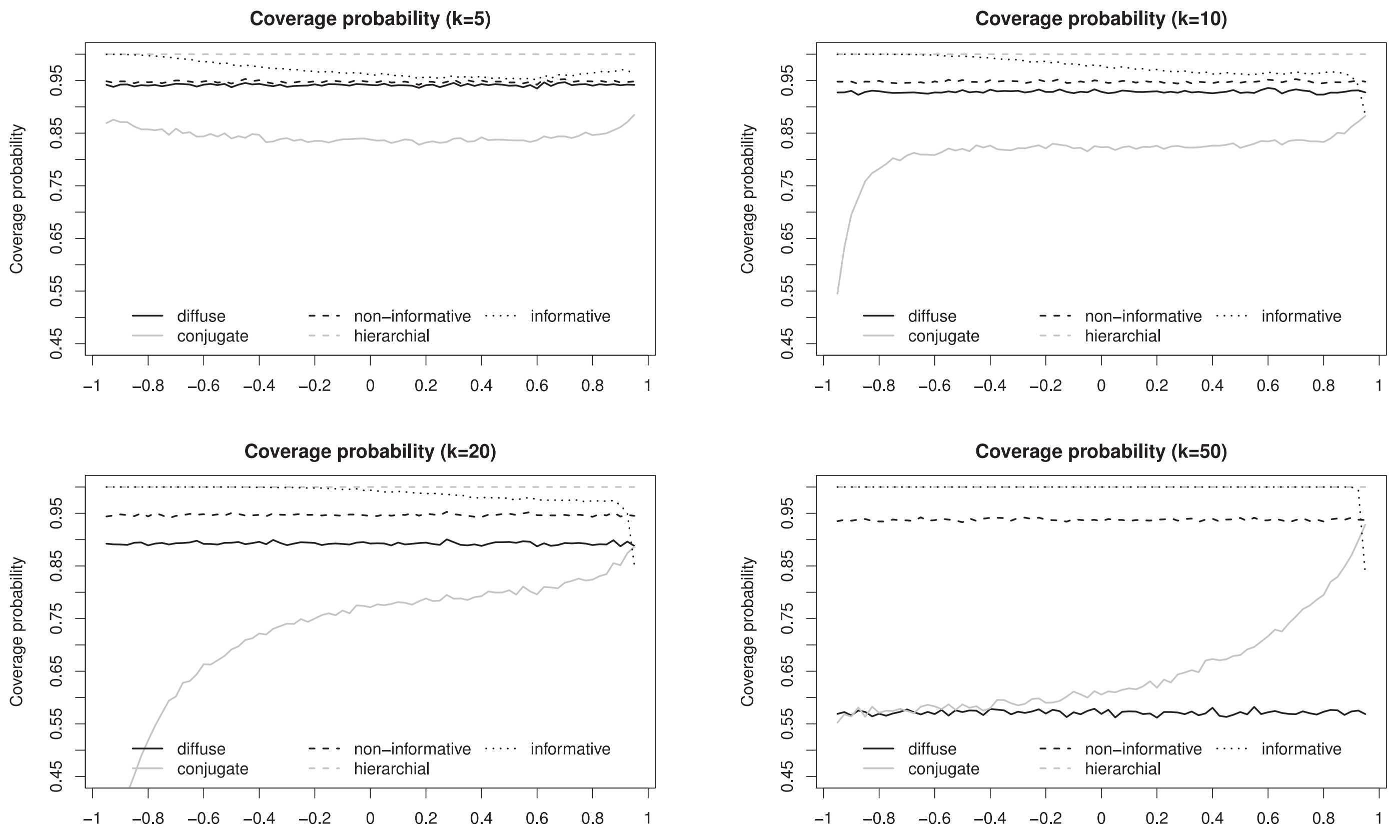

Fig. 1. Coverage probabilities for 95 percent credible intervals based on different priors as a function of ρ in different dimensions k.

5.2. Elliptical credible sets

Let $F_{i,j}$ denote the F-distribution with i and j degrees of freedom. In Theorem 1a we prove that $\boldsymbol{\theta}$ follows a p-variate multivariate t-distribution with $n-1$ degrees of freedom, location parameter $\widehat{\boldsymbol{\theta}}$ and scale parameter $\frac{1}{n-1}\frac{\mathbf{LR}_d\mathbf{L}^T}{\mathbf{1}^T\mathbf{S1}}$ under the diffuse prior (6). This result provides a motivation for considering the following elliptical credible set expressed as

$$\left\{\mathbf{r}\in R^p : \frac{n-1}{p}(\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1})(\widehat{\boldsymbol{\theta}}_d-\mathbf{r})^T\left(\mathbf{LR}_d\mathbf{L}^T\right)^{-1}(\widehat{\boldsymbol{\theta}}_d-\mathbf{r})\le F_{p,n-1;1-\alpha}\right\},$$

where $F_{i,j;\,\beta}$ denotes the β-quantile of F-distribution with i and j degrees of freedom.

The above result follows from the fact that $\boldsymbol{\theta}|\mathbf{X}_1,\ldots,\mathbf{X}_n\sim t_p(n-1,\widehat{\boldsymbol{\theta}},\frac{1}{n-1}\frac{\mathbf{LR}_d\mathbf{L}^T}{\mathbf{1}^T\mathbf{S1}})$ and consequently $T_d=\frac{n-1}{p}(\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1})(\widehat{\boldsymbol{\theta}}_d-\mathbf{r})^T(\mathbf{LR}_d\mathbf{L}^T)^{-1}(\widehat{\boldsymbol{\theta}}_d-\mathbf{r})\sim F_{p,n-1}$.

Similarly, the elliptical credible set under the conjugate prior (7) is given by

$$\left\{\mathbf{r}\in R^p : \frac{\nu_c+n-k-1}{p}(\mathbf{1}^T\mathbf{V}_c^{-1}\mathbf{1})(\widehat{\boldsymbol{\theta}}_c-\mathbf{r})^T\left(\mathbf{LR}_c\mathbf{L}^T\right)^{-1}(\widehat{\boldsymbol{\theta}}_c-\mathbf{r})\le F_{p,\nu_c+n-k-1;1-\alpha}\right\},$$

while under the non-informative prior (21) it is given by

$$\left\{\mathbf{r}\in R^p : \frac{n-k+p}{p}(\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1})(\widehat{\boldsymbol{\theta}}_n-\mathbf{r})^T\left(\mathbf{LR}_d\mathbf{L}^T\right)^{-1}(\widehat{\boldsymbol{\theta}}_n-\mathbf{r})\le F_{p,n-k+p;1-\alpha}\right\}.$$

Finally, using the stochastic representations (12) and (30) for $\boldsymbol{\theta}$ under the hierarchial prior (8) and under the informative prior (26), the elliptical credible sets are obtained numerically by applying the bootstrap method (see Davison and Hinkley, 1997, p.174).

6. Numerical and empirical illustrations

6.1. Numerical study

In this section we assess the performance of different priors within a numerical study. We compute the coverage probabilities of credible intervals for the portfolio weights based on the posterior distributions from the previous sections. For this purpose we compute the 95 percent credible intervals explicitly if the corresponding quantiles come from a known distribution. Alternatively, as in the case of hierarchical and informative priors, the quantiles are computed via simulations using the respective stochastic representation with the number of repetitions equal to 1000. In the next step we simulate 10000 samples of asset returns, compute the corresponding portfolio weights and count the fraction of times the weights are covered by the credible intervals.

The comparison is done for $p=1$, $\mathbf{L}=\mathbf{e}_1^T$, $\boldsymbol{\mu}=0.01\cdot(1,2,\ldots,k)^T$ and $\boldsymbol{\Sigma}=(\rho^{|i-j|})_{i,j=1,\ldots,k}$, where ρ takes values between -1 and 1. Since the dimension of the portfolio is particularly of interest we consider $k\in\{5,\ 10,\ 20,\ 50\}$. The sample size n is set to 60, which is a typical value in financial literature and corresponds to roughly two months of daily data or a year of weekly data, respectively. In all considered cases we take the following parameters for the conjugate prior $\nu_c=\kappa_c=n$, $\boldsymbol{\mu}_c=\mathbf{0}_k$ and $\mathbf{S}_c=\mathbf{I}_k$; for the hierarchial prior $\epsilon_1=0.0001$, $\epsilon_2=0.0001$ (as in Greyserman et al., 2006), $\kappa_h=\nu_h=n$ and $\mathbf{S}_h=\mathbf{I}_k$; for the informative prior $\delta_1=1$ and $\delta_2=0.5$, $\nu_I=n$, $w_I=1/k$, $S_I=1$.

The coverage probabilities as functions of ρ for different values of k are plotted in Fig. 1. The informative and the hierarchical priors in particular yield obviously too wide credible intervals,

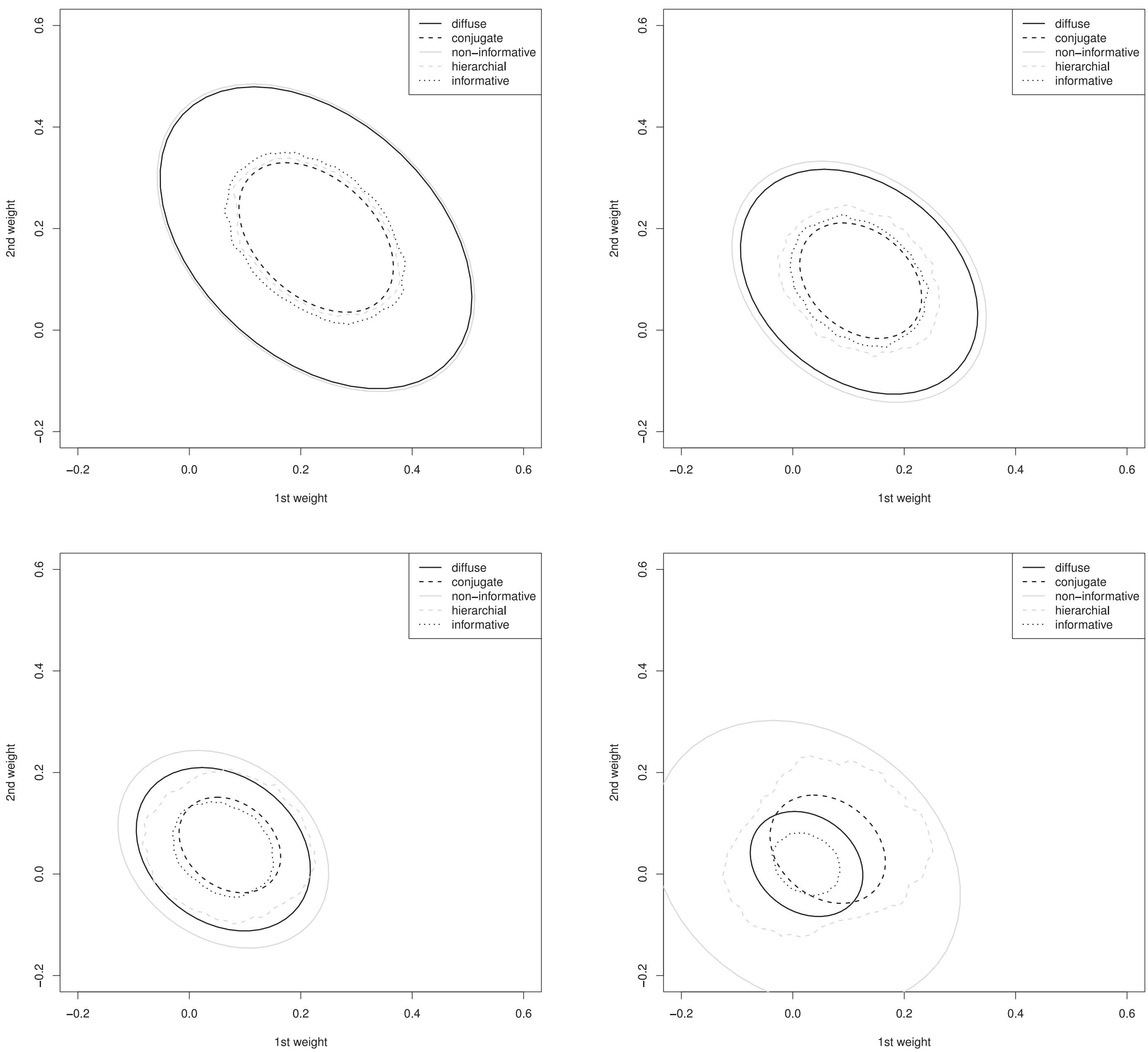

Fig. 2. 95 percent credible sets for the first two assets based on different priors in different dimensions k.

causing the coverage probability to be almost one. This holds in all dimensions and for all values of ρ in case of the hierarchical prior, whereas an extreme behavior of the informative prior is present only for large values of k and negative correlation. The conjugate prior causes too narrow credible intervals leading to coverage probabilities much lower than 95 percent. The higher is k, the larger is the discrepancy. The diffuse prior shows a stable behavior with respect to ρ and heavily undershoots the true coverage probability only for high k. In contrast to the previous priors, the non-informative prior does uniformly the best job with a minor bias even for $k = 50$.

To assess the superiority of the informative prior we compare the precision of the methods if the data stem from the model based on informative prior. In practice, this assumption implies that the prior beliefs of the investor about the GMV portfolio weights are correct. Since the prior information about the GMV portfolio weights is directly present only in the informative prior, we chose this model to simulate the data. We run 10^5 replications and within each replication we simulate a path of returns starting with the informative prior. For each data set we then estimate the portfolio weights using all five kinds of priors distributions. Additionally, we consider different prior weights $\mathbf{w}_I$: a) an equally weighted portfolio with $w_{I,i} = 0.2$ for all i; b) a slightly skewed portfolio with $w_{I,1} = 0.6$ and $w_{I,i} = 0.1$ for $i > 1$; c) a heavily skewed portfolio $w_{I,1} = 0.8$ and $w_{I,i} = 0.05$ for $i > 1$; d) a portfolio with short sales with $w_{I,1} = -1.2$ and $w_{I,i} = 0.05$ for $i > 1$. All other parameters are selected as described above. As a measure of performance we consider the L^2-deviation of the estimators from the true weights averaged over all replications. The results are summarized in Table 1. We conclude that the additional information in the prior is obviously advantageous and leads to weights which are closer to the true weights. The conjugate

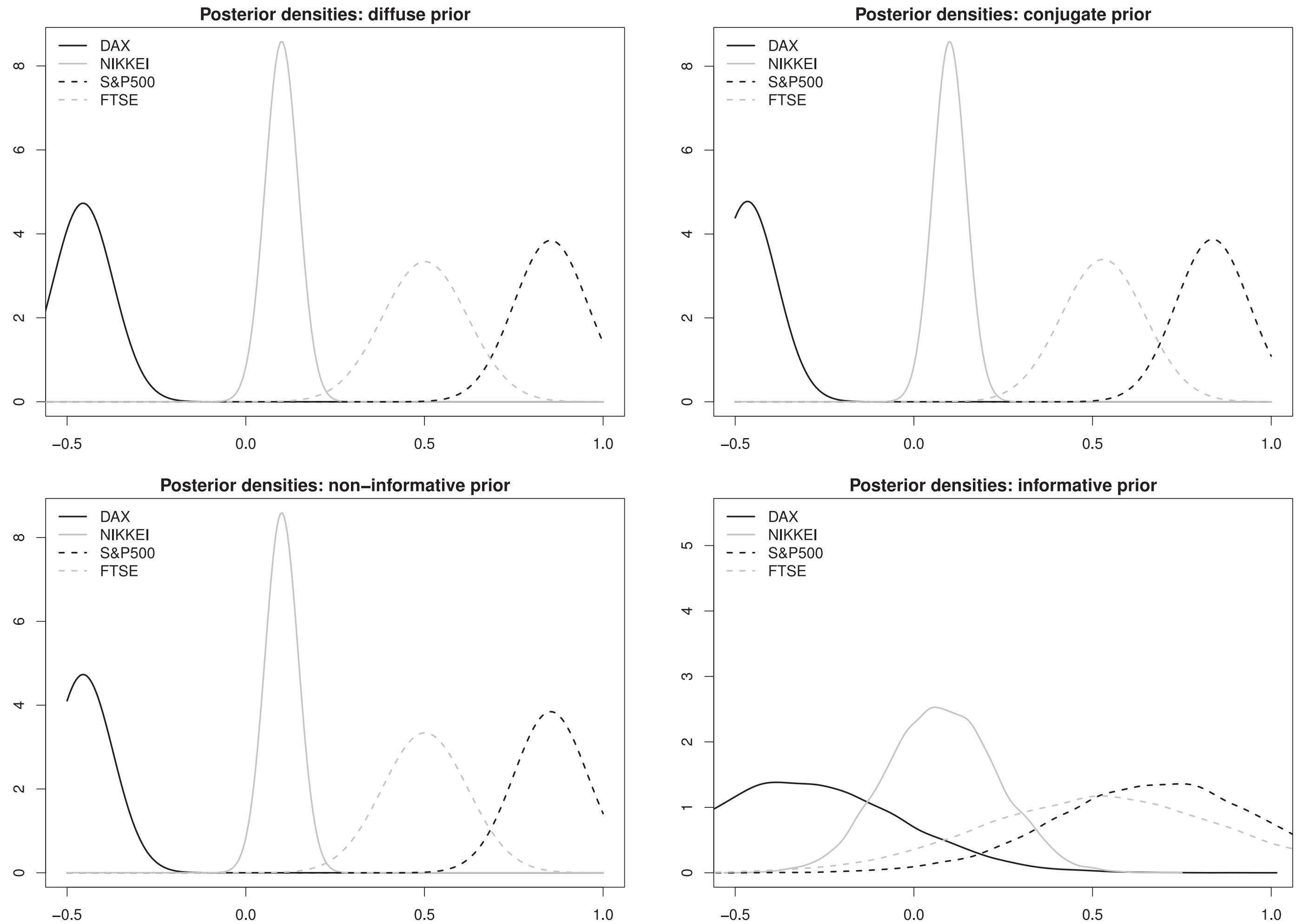

Fig. 3. Posterior densities for the portfolio weights of DAX, NIKKEI, S&P500 and FTSE for the period from 16.03.2010 till 27.01.2015 based on the diffuse (top left), the conjugate prior (top right), the non-informative (bottom left), and the informative prior (bottom right). The priors are based on historical observations from 26.04.2005 till 09.03.2010.

Table 1
L^2-deviations from the prior weights based on 10^5 simulations from the informative prior.

	Diffuse	Conj.	Hier.	Non-inf.	Inf.
$w_I = c(0.2, \dots, 0.2)$	11.239	2.378	3.644	11.239	1.484
$w_I = c(0.8, 0.05, \dots, 0.05)$	11.094	2.455	3.630	11.094	1.492
$w_I = c(0.6, 0.1, \dots, 0.1)$	11.415	2.388	3.601	11.415	1.459
$w_I = c(-1.2, 0.05, \dots, 0.05)$	11.297	2.554	3.658	11.297	1.495

and hierarchical prior are a little worse. But the non-informative priors show the worst performance due to neglected additional information. This implies that the additional information in form of priors should be used if available. Since the Bayes estimators for the GMV portfolio weights resulting from the application of the diffuse prior and the non-informative prior coincide with the sample estimator obtained by using methods of frequentist statistics, the results of Table 1 demonstrate that the application of the Bayes estimators using informative priors outperforms considerably the frequentist approach if the prior information about the weights is available.

The coverage probabilities illustrate the performance of the individual methods, but do not give any insights regarding the size of the credible sets. For this purpose we visualize the credible sets for the first two assets in Fig. 2. We use the same parameter constellations as above. Note that for the diffuse, conjugate and non-informative priors the credible sets are ellipses centered at the corresponding point estimates. For the hierarchical and informative priors we run a simulation study and plot the level lines of the two-dimensional 95 percent quantile. We observe that the diffuse and non-informative priors generate much wider credible sets. This is explained by the absence of additional information in such priors. If the priors are informative, the credible sets reflect this additional information and become narrower. The estimators based on informative priors reveal a drift in the ellipses, which is due to the shrinkage-type point estimators. This becomes, however, evident only for larger dimensions. Another artefact of this behavior is potentially bimodal form of the credible sets for the hierarchical prior. Note, that the credible sets are narrow compared to the confidence sets reported in Okhrin and Schmid (2007).

6.2. Empirical illustration

Within the empirical illustration we consider the weekly logarithmic returns for four international financial indices DAX, NIKKEI, S&P500 and FTSE for the period from 22.01.1985 till 27.01.2015 resulting in 1567 observation points. The empirical study is twofold. First, we assess the posterior distribution of the GMV portfolio weights. Second, we evaluate a trading strategy based on Bayesian

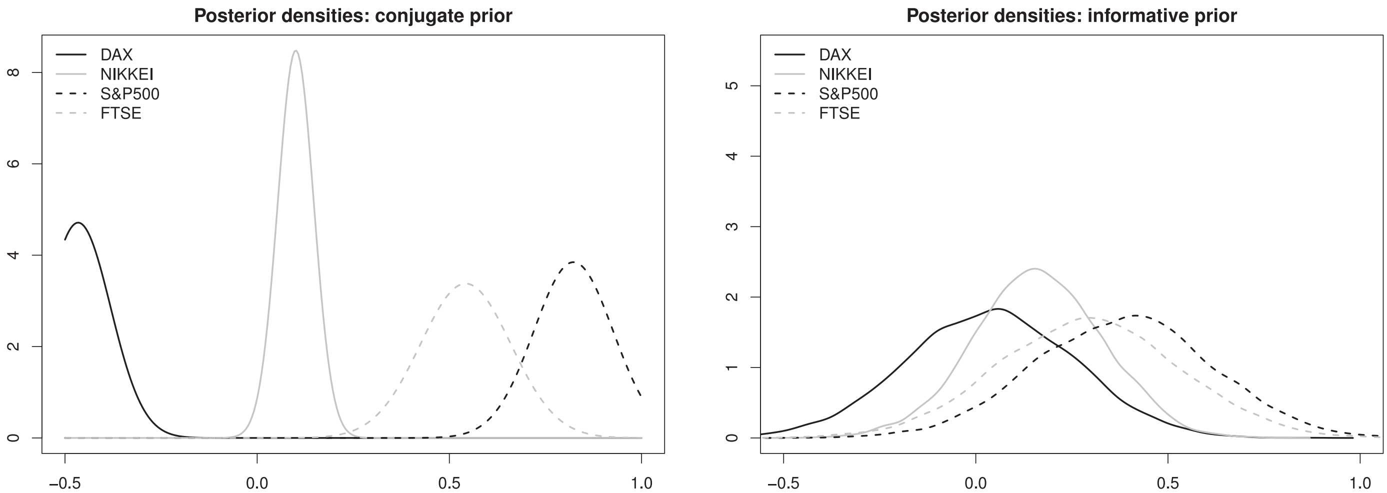

Fig. 4. Posterior densities for the portfolio weights of DAX, NIKKEI, S&P500 and FTSE for the period from 16.03.2010 till 27.01.2015 based on the the conjugate prior (left) and the informative prior (right). The priors correspond to the equally weighted portfolio.

estimates for the weight. Within the study we consider both the priors for the asset returns and the priors for portfolio weights. To diversify the study and to show the robustness of the results we choose two types of priors. The first prior mimics the classical statistical approach when a historical sample is used to estimate the parameters of priors relying on the empirical Bayes approach. Here we use a sample of length 255 (5 years of weekly data) preceding the estimation period. The second type of prior utilizes the evidence that the equally weighted portfolio shows a good performance out-of-sample. Thus here we take the equally weighted portfolio as the second prior in our study. In the case of priors for the parameters of asset returns this corresponds to equal mean returns, equal variances and equal correlations for all assets. To assess the posterior distribution we take the observation from 16.03.2010 till 27.01.2015 as the in-sample period, and the data form 26.04.2005 till 09.03.2010 as a prerun. The mean, the covariance matrix and the corresponding global minimum variance portfolio weights for the prior sample are equal to

$$\boldsymbol{\mu}_{prior} = (12.505,\ -3.120,\ -1.195,\ 4.792)' \times 10^{-4},$$

$$\mathbf{S}_{prior} = 10^{-4} \times \begin{pmatrix} 8.743 & 6.361 & 6.380 & 6.614 \\ 6.361 & 13.144 & 5.123 & 6.460 \\ 6.380 & 5.123 & 6.892 & 5.367 \\ 6.614 & 6.460 & 5.367 & 6.955 \end{pmatrix}.$$

These parameters are used as input parameters in the prior distributions, i.e. $\boldsymbol{\mu}_c = \boldsymbol{\mu}_{prior},\ \mathbf{S}_c = \mathbf{S}_h = \mathbf{S}_{prior},\ \mathbf{w}_I = \mathbf{w}_{prior}$. For the working sample the corresponding parameters are equal to:

$$\bar{\mathbf{X}} = (23.176,\ 20.377,\ 22.604,\ 7.665)' \times 10^{-4},$$

$$\mathbf{S} = 10^{-4} \times \begin{pmatrix} 8.620 & 5.072 & 4.920 & 5.814 \\ 5.072 & 10.041 & 3.564 & 3.907 \\ 4.920 & 3.564 & 4.225 & 3.942 \\ 5.814 & 3.907 & 3.942 & 5.165 \end{pmatrix}.$$

Note that the prior period covers the global financial crisis, which was followed by a relatively calm period starting from 2010. This is mirrored in the estimated parameters. The average returns in the crisis period are much lower and for two markets even negative. The volatilities appear to reflect the turmoil performance of financial markets heavily.

Keeping other hyperparameters as in the simulation study, we compute the posterior densities for each weight, thus setting $\mathbf{L}$ being equal to basis vectors $\mathbf{e}_i$ for $i = 1, \dots, 4$. Due to the poor coverage of the hierarchical prior, we drop it from the analysis here. The plots of all densities based on non-informative and informative priors are given in Fig. 3 for the historical prior and in Fig. 4 for the prior based on the equally weighted portfolio. Due to the low dimension, both priors lead obviously to very close posteriors centered around the sample weights. We expect stronger deviations with increasing k. The conjugate prior fails to incorporate the prior information appropriately and leads to a density similar to the density of the non-informative prior. In contrast to this observation, the informative prior clearly utilizes the prior information leading to shifted and wider densities. This is consistent with our expectations. The densities with the equally weighted portfolio as a prior show clearly the shift in the weights to 0.25. The same is however not observed for the conjugate prior. Here the large sample size reduces the influence of the prior.

To evaluate the goodness of the suggested estimators we simulate a real trading strategy. We compare the estimators based on the conjugate, hierarchical, non-informative and informative priors for the weights. The prior information reflects our belief into the equally weighted portfolio, which is our benchmark. The diffuse priors lead to numerically identical point estimates as the non-informative prior and thus is dropped from the analysis here. We estimate the required parameters using the previous 51 observations (one year of weekly data) at each moment of time. The portfolio is held one time period, i.e. one week. At the beginning of the next week we compute the realized portfolio return. This procedure is repeated for the complete data set. Using the obtained time series of portfolio returns, we compute the following performance measures: mean portfolio return, standard deviation of the portfolio return, Sharpe ratio, Value-at-Risk (VaR) and expected shortfall (ES) at 95 percent and 99 percent levels. The results are summarized in Table 2. The equally weighted portfolio has the highest average return, but clearly underperforms the remaining alternatives in terms of risk. Among the Bayesian strategies, the estimators based on the conjugate prior seem to have the best risk measures, but the lowest average return. To assess the dynamics of the weights we plot the corresponding times series in Fig. 5. The behavior of the weights captures the volatile periods on financial markets with rapid drops in more risky assets. The hierarchical prior leads to extremely volatile portfolio weights, leading to an unrealistic and expensive strategy. The informative prior for the weights utilizes the equally weighted

Table 2
Performance measures of the alternative trading strategies based on different estimates of the portfolio weights from 22.01.1985 till 27.01.2015. The estimation window is set to 51. The priors correspond to the equally weighted portfolio.

	Conjugate	Hierarchical	Non-inf	Informative	Eq
$\hat{\mu}_p \times 10^{-4}$	8.7145	9.2926	9.0381	9.2451	10.0631
$\hat{\sigma}_p \times 10^{-2}$	2.1293	2.2394	2.1446	2.1596	2.8007
Sharpe ratio $\times 10^{-2}$	4.0927	4.1496	4.2143	4.2809	4.4135
VaR 95 percent $\times 10^{-2}$	6−.0889	−6.6330	−6.0865	−6.9071	−7.0855
VaR 99 percent $\times 10^{-2}$	−3.3035	−3.4347	−3.3142	−3.3805	−3.5642
ES 95 percent $\times 10^{-2}$	−9.3079	−10.0754	−9.3323	−9.4014	−9.4391
ES 99 percent $\times 10^{-2}$	−5.3715	−5.5685	−5.3862	−5.4119	−5.7093

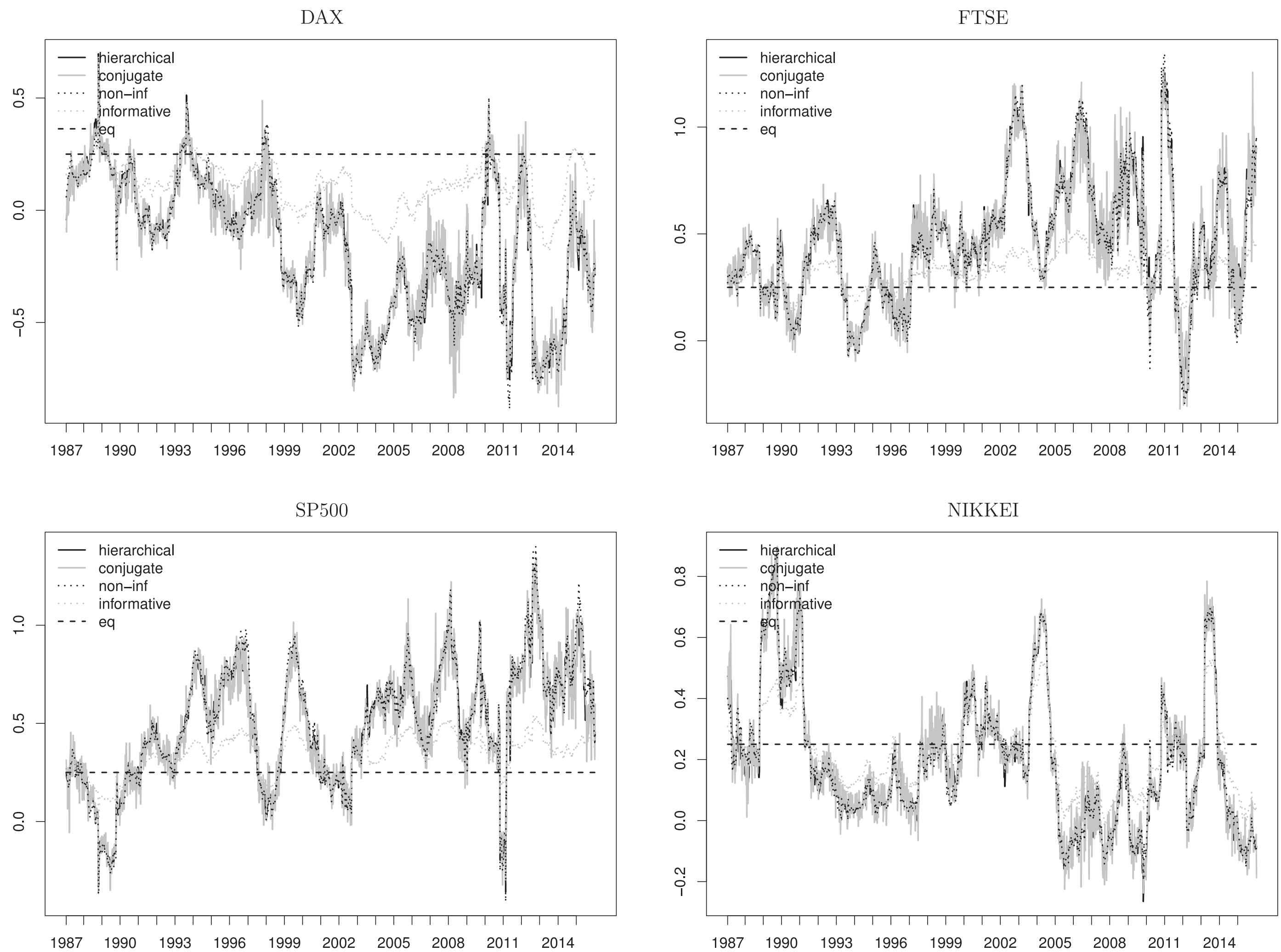

Fig. 5. Time series of alternative estimators of optimal portfolio weights. Length of the estimation window is set to 51. The priors correspond to the equally weighted portfolio.

prior and results in portfolio weights which are much closer to 0.25 (weight of the equally weighted portfolio). Note that the estimator with non-informative prior numerically coincides with the classical frequentist estimator of the portfolio weights.

7. Summary

In this paper we analyze the global minimum variance portfolio within a Bayesian framework. This setup allows us to incorporate prior beliefs of the investors and to incorporate these into the portfolio decisions. Assuming different priors for the asset returns, we derive explicit formulas for the posterior distributions of linear combinations of GMV portfolio weights. In particular, we consider non-informative (diffuse) and informative (conjugate and hierarchical) priors. Furthermore, relying on a suitable model transformation, we suggest a prior directly for the portfolio weights. The results are evaluated within a numerical study, where we assess the coverage probabilities of credible intervals, and within an empirical study, where we consider the posterior densities for the weights of an international portfolio. Additionally, we run a simulated trading strategy with real data and evaluate the strategies with a series of performance measures. Both studies showed good results of the suggested priors and revealed the need for further analysis, particularly the extension to the general mean-variance portfolio.

Acknowledgments

The authors are thankful to Professor Emanuele Borgonovo and an anonymous Reviewer for careful reading of the paper and for their suggestions which have improved an earlier version of this paper. We also thank Professor Lorenzo Peccati for his helpful suggestions to the first version of the paper and David Bauder for his comments used in the preparation of the revised version of the paper.

Appendix A. Proof of Theorem 1

First, we present an important lemma which is used in the proof of Theorem 1.

Lemma 1. *Let*

$$\boldsymbol{\Sigma}|\mathbf{X}_1,\ldots,\mathbf{X}_n \sim IW_k(\tau_0,\mathbf{V}_0)$$

with $\mathbf{V}_0=\mathbf{V}_0(\mathbf{X}_1,\ldots,\mathbf{X}_n)$ *and let* $\mathbf{L}$ *be a* $p\times k$ *matrix of constants. Then*

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

where $\mathbf{R}_0=\mathbf{V}_0^{-1}-\mathbf{V}_0^{-1}\mathbf{1}\mathbf{1}^T\mathbf{V}_0^{-1}/\mathbf{1}^T\mathbf{V}_0^{-1}\mathbf{1}$.

Proof of Lemma 1. From Theorem 3.4.1 of Gupta and Nagar (2000) it holds that $\boldsymbol{\Sigma}^{-1}|\mathbf{X}_1,\ldots,\mathbf{X}_n$ has a k-dimensional Wishart distribution with (τ_0-k-1) degrees of freedom and the covariance matrix $\mathbf{V}_0^{-1}$.

Let $\widetilde{\mathbf{L}}=(\mathbf{L}^T,\mathbf{1})^T$ and $\mathbf{A}=\widetilde{\mathbf{L}}\boldsymbol{\Sigma}^{-1}\widetilde{\mathbf{L}}^T=\{\mathbf{A}_{ij}\}_{i,j=1,2}$ with $\mathbf{A}_{11}=\mathbf{L}\boldsymbol{\Sigma}^{-1}\mathbf{L}^T$, $\mathbf{A}_{12}=\mathbf{L}\boldsymbol{\Sigma}^{-1}\mathbf{1}$, $\mathbf{A}_{21}=\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{L}^T$, and $A_{22}=\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}$. Similarly, let $\mathbf{H}=\widetilde{\mathbf{L}}\mathbf{V}_0^{-1}\widetilde{\mathbf{L}}^T=\{\mathbf{H}_{ij}\}_{i,j=1,2}$ with $\mathbf{H}_{11}=\mathbf{L}\mathbf{V}_0^{-1}\mathbf{L}^T$, $\mathbf{H}_{12}=\mathbf{L}\mathbf{V}_0^{-1}\mathbf{1}$, $\mathbf{H}_{21}=\mathbf{1}^T\mathbf{V}_0^{-1}\mathbf{L}^T$ and $H_{22}=\mathbf{1}^T\mathbf{V}_0^{-1}\mathbf{1}$.

Since $\boldsymbol{\Sigma}^{-1}|\mathbf{X}_1,\ldots,\mathbf{X}_n\sim W_k(\tau_0-k-1,\mathbf{V}_0^{-1})$ and $rank(\widetilde{\mathbf{L}})=p+1\le k$, the application of Theorem 3.2.5 by Muirhead (1982) leads to $\mathbf{A}\sim W_{p+1}(\tau_0-k-1,\mathbf{H})$. Moreover, using Theorem 3.2.10 of Muirhead (1982), we obtain

$$\frac{\mathbf{L}\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}=\left.\frac{\mathbf{A}_{12}}{A_{22}}\right|A_{22},\mathbf{X}_1,\ldots,\mathbf{X}_n\sim N_p(\mathbf{H}_{12}\mathbf{H}_{22}^{-1},\mathbf{H}_{11\cdot2}\mathbf{A}_{22}^{-1}),\tag{40}$$

where $\mathbf{H}_{11\cdot2}=\mathbf{H}_{11}-\mathbf{H}_{12}\mathbf{H}_{21}/H_{22}$.

The application of Theorem 3.2.8 by Muirhead (1982) leads to $\frac{A_{22}}{H_{22}}\sim\chi^2_{\tau_0-k-1}$. Consequently,

$$A_{22}|\mathbf{X}_1,\ldots,\mathbf{X}_n\sim\Gamma((\tau_0-k-1)/2;2H_{22}),$$

i.e. A_{22} is gamma distributed with shape parameter $(\tau_0-k-1)/2$ and scale parameter $2H_{22}$.

Hence, the posterior distribution of $\frac{\mathbf{L}\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}$ is given by

$$\begin{aligned}
&p_{\frac{\mathbf{L}\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}|\mathbf{X}_1,\ldots,\mathbf{X}_n}(\mathbf{y})\\
&=\int_0^{+\infty}p_{\frac{\mathbf{L}\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}|A_{22},\mathbf{X}_1,\ldots,\mathbf{X}_n}(\mathbf{y}|A_{22}=z)\,p_{A_{22}|\mathbf{X}_1,\ldots,\mathbf{X}_n}(z)\mathbf{d}z\\
&=\frac{(2\pi)^{-p/2}|\mathbf{H}_{11\cdot2}|^{-1/2}}{\Gamma((\tau_0-k-1)/2)(2H_{22})^{(\tau_0-k-1)/2}}\int_0^{\infty}z^{(p+\tau_0-k-1)/2-1}\\
&\times\exp\left\{-\frac{z}{2}\left[H_{22}^{-1}+(\mathbf{y}-\mathbf{H}_{12}H_{22}^{-1})^T\mathbf{H}_{11\cdot2}^{-1}(\mathbf{y}-\mathbf{H}_{12}H_{22}^{-1})\right]\right\}\mathbf{d}z\\
&=\frac{\Gamma((p+\tau_0-k-1)/2)}{\Gamma((\tau_0-k-1)/2)}\frac{\left|\frac{1}{\tau_0-k-1}\frac{\mathbf{H}_{11\cdot2}}{H_{22}}\right|^{-1/2}}{[\pi(\tau_0-k-1)]^{p/2}}\\
&\times\left[1+\frac{1}{\tau_0-k-1}\left(\mathbf{y}-\frac{\mathbf{H}_{12}}{H_{22}}\right)^T\right.\\
&\left.\times\left[\frac{1}{\tau_0-k-1}\frac{\mathbf{H}_{11\cdot2}}{H_{22}}\right]^{-1}\left(\mathbf{y}-\frac{\mathbf{H}_{12}}{H_{22}}\right)\right]^{(p+\tau_0-k-1)/2},
\end{aligned}$$

where the last expression is the density of p-dimensional t-distribution with (τ_0-k-1) degrees of freedom, location vector $\mathbf{H}_{12}H_{22}^{-1}$, and scale matrix $\frac{1}{\tau_0-k-1}\mathbf{H}_{11\cdot2}H_{22}^{-1}$. Noting that

$$\begin{aligned}
\mathbf{H}_{12}H_{22}^{-1}&=\frac{\mathbf{L}\mathbf{V}_0^{-1}\mathbf{1}}{\mathbf{1}^T\mathbf{V}_0^{-1}\mathbf{1}}\quad\text{and}\quad\mathbf{H}_{11\cdot2}H_{22}^{-1}\\
&=\left(\mathbf{L}\mathbf{V}_0^{-1}\mathbf{L}^T-\frac{\mathbf{L}\mathbf{V}_0^{-1}\mathbf{1}\mathbf{1}^T\mathbf{V}_0^{-1}\mathbf{L}^T}{\mathbf{1}^T\mathbf{V}_0^{-1}\mathbf{1}}\right)\frac{1}{\mathbf{1}^T\mathbf{V}_0^{-1}\mathbf{1}}
\end{aligned}$$

completes the proof of Lemma 1 □

Proof of Theorem 1. First, we rewrite the expression of the likelihood function which is then used in the calculation of the posteriors. It holds that

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

a) In the case of the standard diffuse prior $p_d(\boldsymbol{\mu},\boldsymbol{\Sigma})$, the posterior distribution of $(\boldsymbol{\mu},\boldsymbol{\Sigma})$ is given by

$$p_d(\boldsymbol{\mu},\boldsymbol{\Sigma}|\mathbf{X}_1,\ldots,\mathbf{X}_n)\propto L(\mathbf{X}_1,\ldots,\mathbf{X}_n|\boldsymbol{\mu},\boldsymbol{\Sigma})p_d(\boldsymbol{\mu},\boldsymbol{\Sigma}).$$

Integrating out $\boldsymbol{\mu}$ leads to

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

The application of Lemma 1 with $\tau_0=n+k$ and $\mathbf{V}_0=(n-1)\mathbf{S}$ completes the proof of Theorem 1a.

b) The posterior distribution of $(\boldsymbol{\mu},\boldsymbol{\Sigma})$ under the conjugate prior is given by

$$p_c(\boldsymbol{\mu},\boldsymbol{\Sigma}|\mathbf{X}_1,\ldots,\mathbf{X}_n)\propto L(\mathbf{X}_1,\ldots,\mathbf{X}_n|\boldsymbol{\mu},\boldsymbol{\Sigma})p_c(\boldsymbol{\mu},\boldsymbol{\Sigma}).$$

Integrating out $\boldsymbol{\mu}$ leads to

$$\begin{aligned}
p_c(\boldsymbol{\Sigma}|\mathbf{X}_1,\ldots,\mathbf{X}_n)&\propto\int_{R^k}L(\mathbf{X}_1,\ldots,\mathbf{X}_n|\boldsymbol{\mu},\boldsymbol{\Sigma})p_c(\boldsymbol{\mu},\boldsymbol{\Sigma})\mathbf{d}\boldsymbol{\mu}\\
&\propto|\boldsymbol{\Sigma}|^{-(\nu_c+n+1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}[((n-1)\mathbf{S}+\mathbf{S}_c)\boldsymbol{\Sigma}^{-1}]\right\}\\
&\times\int_{R^k}\exp\left\{-\frac{n}{2}(\overline{\mathbf{X}}_n-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{X}}_n-\boldsymbol{\mu})\right.\\
&\left.-\frac{\kappa_c}{2}(\boldsymbol{\mu}_c-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_c-\boldsymbol{\mu})\right\}\mathbf{d}\boldsymbol{\mu}\\
&\propto|\boldsymbol{\Sigma}|^{-(\nu_c+n+1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}[\mathbf{V}_c\boldsymbol{\Sigma}^{-1}]\right\}\\
&\times\int_{R^k}\exp\left\{-\frac{n+\kappa_c}{2}(\mathbf{r}_c-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{r}_c-\boldsymbol{\mu})\right\}\mathbf{d}\boldsymbol{\mu}\\
&\propto|\boldsymbol{\Sigma}|^{-(\nu_c+n)/2}\mathrm{etr}\left\{-\frac{1}{2}\mathbf{V}_c\boldsymbol{\Sigma}^{-1}\right\},
\end{aligned}$$

where

$$\mathbf{r}_c = \frac{n\overline{\mathbf{X}}_n + \kappa_c \boldsymbol{\mu}_c}{n+\kappa_c},$$

$$\mathbf{V}_c = (n-1)\mathbf{S} + \mathbf{S}_c + (n+\kappa_c)\mathbf{r}_c\mathbf{r}_c^T + n\overline{\mathbf{X}}_n\overline{\mathbf{X}}_n^T + \kappa_c\boldsymbol{\mu}_0\boldsymbol{\mu}_0^T.$$

The rest of the proof follows from Lemma 1 with $\tau_0 = \nu_c + n$ and $\mathbf{V}_0 = \mathbf{V}_c$.

c) Under the hierarchial prior $p_h(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi, \eta)$, the conditional posterior distribution of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ given $\xi, \eta, \mathbf{X}_1, \ldots, \mathbf{X}_n$ is

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

Integrating out $\boldsymbol{\mu}$ leads to

$$\begin{aligned}
&p_h(\boldsymbol{\Sigma}|\xi, \eta, \mathbf{X}_1, \ldots, \mathbf{X}_n)\\
&\quad \propto \int_{R^k} p_h(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\xi, \eta, \mathbf{X}_1, \ldots, \mathbf{X}_n)\mathbf{d}\boldsymbol{\mu}\\
&\quad \propto \eta^{-k(\nu_h-k-1)/2}|\boldsymbol{\Sigma}|^{-(\nu_h+n)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}[\mathbf{V}_h\boldsymbol{\Sigma}^{-1}]\right\},
\end{aligned}$$

where

$$\mathbf{r}_h = \mathbf{r}_h(\xi) = \frac{n\overline{\mathbf{X}} + \kappa_h\xi\mathbf{1}}{n+\kappa_h},$$

$$\begin{aligned}
\mathbf{V}_h = \mathbf{V}_h(\xi, \eta) &= (n-1)\mathbf{S} + \eta^{-1}\mathbf{S}_h\\
&\quad - (n+\kappa_h)\mathbf{r}_h\mathbf{r}_h^T + n\overline{\mathbf{X}}\,\overline{\mathbf{X}}^T + \kappa_h\xi^2\mathbf{1}\mathbf{1}^T.
\end{aligned}$$

The application of Lemma 1 with $\tau_0 = \nu_h + n$ and $\mathbf{V}_0 = \mathbf{V}_h$ and the integration over ξ, η lead to the expression presented in Theorem 1c. □

Appendix B. Derivation of the Fisher information matrix

Let $\boldsymbol{\phi} = (\boldsymbol{\theta}^T, \mathbf{vech}(\boldsymbol{\Psi})^T, \zeta)^T$. Then the Fisher information matrix $I(\boldsymbol{\theta}, \boldsymbol{\Psi}, \zeta)$ is given by

$$I(\boldsymbol{\theta}, \boldsymbol{\Psi}, \zeta) = -E\left[\frac{\partial^2}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}^T}\log L((n-1)^{-1}\widetilde{\mathbf{S}}|\boldsymbol{\theta}, \boldsymbol{\Psi}, \zeta)\right],$$

where (see (20))

$$\begin{aligned}
\log L((n-1)^{-1}\widetilde{\mathbf{S}}|\boldsymbol{\theta}, \boldsymbol{\Psi}, \zeta) &\propto \frac{n-k+p}{2}\log|\boldsymbol{\Psi}| + \frac{n-k+p}{2}\log\zeta\\
&\quad - \frac{n-1}{2}\mathrm{tr}[\widetilde{\mathbf{S}}_{11}^{(-)}\boldsymbol{\Psi}] - \frac{\zeta(n-1)}{2}\left(\mathrm{tr}[\widetilde{\mathbf{S}}_{11}^{(-)}\boldsymbol{\theta}\boldsymbol{\theta}^T] + 2\mathrm{tr}[\widetilde{\mathbf{S}}_{12}^{(-)}\boldsymbol{\theta}^T] + \widetilde{S}_{22}^{(-)}\right).
\end{aligned}$$

It holds that

$$\frac{\partial}{\partial\boldsymbol{\theta}}\log L((n-1)^{-1}\widetilde{\mathbf{S}}|\boldsymbol{\theta}, \boldsymbol{\Psi}, \zeta) = -(n-1)\zeta\widetilde{\mathbf{S}}_{11}^{(-)}\boldsymbol{\theta} - (n-1)\zeta\widetilde{\mathbf{S}}_{12}^{(-)},$$

$$\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\log L((n-1)^{-1}\widetilde{\mathbf{S}}|\boldsymbol{\theta}, \boldsymbol{\Psi}, \zeta) = -(n-1)\zeta\widetilde{\mathbf{S}}_{11}^{(-)},$$

$$\begin{aligned}
\frac{\partial}{\partial\zeta}\log L((n-1)^{-1}\widetilde{\mathbf{S}}|\boldsymbol{\theta}, \boldsymbol{\Psi}, \zeta) &= \frac{n-k+p}{2}\zeta^{-1}\\
&\quad - \frac{n-1}{2}\left(\mathrm{tr}[\widetilde{\mathbf{S}}_{11}^{(-)}\boldsymbol{\theta}\boldsymbol{\theta}^T] + 2\mathrm{tr}[\widetilde{\mathbf{S}}_{12}^{(-)}\boldsymbol{\theta}^T] + \widetilde{S}_{22}^{(-)}\right),
\end{aligned}$$

$$\frac{\partial^2}{\partial^2\zeta}\log L((n-1)\widetilde{\mathbf{S}}|\boldsymbol{\theta}, \boldsymbol{\Psi}, \zeta) = -\frac{n-k+p}{2}\zeta^{-2},$$

$$\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\zeta}\log L((n-1)^{-1}\widetilde{\mathbf{S}}|\boldsymbol{\theta}, \boldsymbol{\Psi}, \zeta) = -(n-1)\widetilde{\mathbf{S}}_{11}^{(-)}\boldsymbol{\theta} - (n-1)\widetilde{\mathbf{S}}_{12}^{(-)}.$$

From the properties of the differential of a determinant (cf. Magnus & Neudecker, 2007) we obtain

$$\mathbf{d}|\boldsymbol{\Psi}| = |\boldsymbol{\Psi}|(\mathbf{vec}(\boldsymbol{\Psi}^{\mathbf{-1}}))^T\mathbf{dvec}(\boldsymbol{\Psi}).$$

Using the relationship between **vec** and **vech** operators (see Harville, 1997, p. 365) we get

$$\mathbf{vec}(\boldsymbol{\Psi}) = \mathbf{G}_p\mathbf{vech}(\boldsymbol{\Psi}),$$

and, hence,

$$\mathbf{d}|\boldsymbol{\Psi}| = |\boldsymbol{\Psi}|(\mathbf{vech}(\boldsymbol{\Psi}^{-1}))^T\mathbf{G}_p^T\mathbf{G}_p\mathbf{dvech}(\boldsymbol{\Psi}).$$

The last equality leads to

$$\frac{\partial|\boldsymbol{\Psi}|}{\partial\mathbf{vech}(\boldsymbol{\Psi})} = |\boldsymbol{\Psi}|(\mathbf{G}_p^T\mathbf{G}_p)^T\mathbf{vech}(\boldsymbol{\Psi}^{-1})$$

and, consequently,

$$\frac{\partial\ln|\boldsymbol{\Psi}|}{\partial\mathbf{vech}(\boldsymbol{\Psi})} = \mathbf{G}_p^T\mathbf{G}_p\mathbf{vech}(\boldsymbol{\Psi}^{-1}).$$

The second order derivative is

$$\begin{aligned}
\frac{\partial^2\ln|\boldsymbol{\Psi}|}{\partial\mathbf{vech}(\boldsymbol{\Psi})\partial(\mathbf{vech}(\boldsymbol{\Psi}))^T} &= \mathbf{G}_p^T\mathbf{G}_p\frac{\partial\mathbf{vech}(\boldsymbol{\Psi}^{-1})}{\partial(\mathbf{vec}\boldsymbol{\Psi})^T}\\
&= -(\mathbf{G}_p^T\mathbf{G}_p)\mathbf{H}_p(\boldsymbol{\Psi}^{\mathbf{-1}}\otimes\boldsymbol{\Psi}^{-1})\mathbf{G}_p,
\end{aligned}$$

where the last equality follows from Harville (1997), p. 368 with $\mathbf{H}_p = (\mathbf{G}_p^T\mathbf{G}_p)^{-1}\mathbf{G}_p^T$.

Thus using the previous results for the partial derivatives of a symmetric matrix, we get

$$\begin{aligned}
\frac{\partial\log L((n-1)^{-1}\widetilde{\mathbf{S}}|\boldsymbol{\theta}, \boldsymbol{\Psi}, \zeta)}{\partial\mathbf{vech}(\boldsymbol{\Psi})} &= \frac{n-k+p}{2}\mathbf{G}_p^T\mathbf{G}_p\mathbf{vech}(\boldsymbol{\Psi}^{-1})\\
&\quad - \frac{n-1}{2}\mathbf{vech}(\widetilde{\mathbf{S}}_{11}^{(-)}),
\end{aligned}$$

$$\frac{\partial^2\log L((n-1)^{-1}\widetilde{\mathbf{S}}|\boldsymbol{\theta}, \boldsymbol{\Psi}, \zeta)}{\partial\mathbf{vech}(\boldsymbol{\Psi})\partial(\mathbf{vech}(\boldsymbol{\Psi}))^T} = -\frac{n-k+p}{2}\mathbf{G}_p^T(\boldsymbol{\Psi}^{\mathbf{-1}}\otimes\boldsymbol{\Psi}^{-1})\mathbf{G}_p$$

$$\frac{\partial^2\log L((n-1)^{-1}\widetilde{\mathbf{S}}|\boldsymbol{\theta}, \boldsymbol{\Psi}, \zeta)}{\partial\mathbf{vech}(\boldsymbol{\Psi})\partial\boldsymbol{\theta}^T} = \mathbf{0},\quad \frac{\partial^2\log L((n-1)^{-1}\widetilde{\mathbf{S}}|\boldsymbol{\theta}, \boldsymbol{\Psi}, \zeta)}{\partial\mathbf{vech}(\boldsymbol{\Psi})\partial\zeta} = \mathbf{0}.$$

The identity $(n-1)\widetilde{\mathbf{S}}^{-1} \sim W_{p+1}(n+p-k, \widetilde{\boldsymbol{\Sigma}}^{-1})$ and the properties of the Wishart distribution (see Muirhead, 1982) lead to

$$\mathrm{E}[\widetilde{\mathbf{S}}^{-1}] = \frac{n+p-k}{n-1}\widetilde{\boldsymbol{\Sigma}}^{-1} = \frac{n+p-k}{n-1}\begin{bmatrix}\boldsymbol{\Psi}^{-1} & -\boldsymbol{\Psi}^{-1}\boldsymbol{\theta}\\ -\boldsymbol{\theta}^T\boldsymbol{\Psi}^{-1} & \zeta^{-1}+\boldsymbol{\theta}^T\boldsymbol{\Psi}^{-1}\boldsymbol{\theta}\end{bmatrix}$$

Hence,

$$E(\widetilde{\mathbf{S}}_{11}^{(-)}) = \frac{n+p-k}{n-1}\boldsymbol{\Psi}^{-1};$$

$$E(\widetilde{\mathbf{S}}_{12}^{(-)}) = -\frac{n+p-k}{n-1}\boldsymbol{\Psi}^{-1}\boldsymbol{\theta};$$

$$E(\widetilde{\mathbf{S}}_{22}^{(-)}) = \frac{n+p-k}{n-1}(\zeta^{-1}+\boldsymbol{\theta}^T\boldsymbol{\Psi}^{-1}\boldsymbol{\theta})$$

As a result the Fisher information matrix is given by

$$\begin{aligned}
&\mathbf{I}(\boldsymbol{\theta}, \boldsymbol{\Psi}, \zeta)\\
&\propto -E\begin{bmatrix}-(n-k+p)\zeta\widetilde{\mathbf{S}}_{11}^{(-)} & \mathbf{0}_{p\times p(p+1)/2} & -(n-1)(\widetilde{\mathbf{S}}_{11}^{(-)}\boldsymbol{\theta}+\widetilde{\mathbf{S}}_{12}^{(-)})\\ \mathbf{0}_{p(p+1)/2\times p} & -\frac{n-k+p}{2}\mathbf{G}_p^T(\boldsymbol{\Psi}^{\mathbf{-1}}\otimes\boldsymbol{\Psi}^{-1})\mathbf{G}_p & \mathbf{0}_{p(p+1)/2}\\ -(n-1)(\widetilde{\mathbf{S}}_{11}^{(-)}\boldsymbol{\theta}+\widetilde{\mathbf{S}}_{12}^{(-)})^T & \mathbf{0}_{p(p+1)/2}^T & -\frac{n-k+p}{2}\zeta^{-2}\end{bmatrix}\\
&\propto \begin{bmatrix}(n-k+p)\zeta\boldsymbol{\Psi}^{-1} & \mathbf{0}_{p\times p(p+1)/2} & \mathbf{0}_p\\ \mathbf{0}_{p(p+1)/2\times p} & \frac{n-k+p}{2}\mathbf{G}_p^T(\boldsymbol{\Psi}^{\mathbf{-1}}\otimes\boldsymbol{\Psi}^{-1})\mathbf{G}_p & \mathbf{0}_{p(p+1)/2}\\ \mathbf{0}_p^T & \mathbf{0}_{p(p+1)/2}^T & \frac{n-k+p}{2}\zeta^{-2}\end{bmatrix}.
\end{aligned}$$

□

Proof of Theorem 3:. The posterior distribution under the informative prior (26) is given by

$$p_I\Big(\boldsymbol{\theta},\boldsymbol{\Psi},\zeta\,\Big|\,(n-1)^{-1}\widetilde{\mathbf{S}}\Big)\propto L\Big((n-1)^{-1}\widetilde{\mathbf{S}}\Big|\boldsymbol{\theta},\boldsymbol{\Psi},\zeta\Big)p_I\Big(\boldsymbol{\theta},\boldsymbol{\Psi},\zeta\Big),$$

where the likelihood function is given in (20). Thus,

$$p_I\Big(\boldsymbol{\theta},\boldsymbol{\Psi},\zeta\,\Big|\,(n-1)^{-1}\widetilde{\mathbf{S}}\Big)\propto|\boldsymbol{\Psi}|^{(n-k+\nu_I)/2}\mathrm{etr}\left\{-\frac{1}{2}\mathbf{A}\boldsymbol{\Psi}\right\}$$
$$\times\zeta^{(n-k+2p+2\delta_1-2)/2}\exp\left\{-\frac{\zeta(n-1)}{2}\left(\frac{\delta_2^{-1}}{n-1}+\mathrm{tr}[\widetilde{\mathbf{S}}_{11}^{(-)}\boldsymbol{\theta}\boldsymbol{\theta}^T]\right.\right.$$
$$\left.\left.+\ 2\mathrm{tr}[\widetilde{\mathbf{S}}_{12}^{(-)}\boldsymbol{\theta}^T]+\widetilde{S}_{22}^{(-)}\right)\right\},$$

where

$$\mathbf{A}=\mathbf{A}(\boldsymbol{\theta},\zeta)=\zeta(\boldsymbol{\theta}-\mathbf{w}_I)(\boldsymbol{\theta}-\mathbf{w}_I)^T+\mathbf{S}_I^{-1}+(n-1)(\mathbf{L}\mathbf{R}_d\mathbf{L}^T)^{-1}.$$

Integrating out $\boldsymbol{\Psi}$ and using the equalities (cf. Harville, 1997, p. 205)

$$|\mathbf{A}|=|\mathbf{S}_I^{-1}+(n-1)(\mathbf{L}\mathbf{R}_d\mathbf{L}^T)^{-1}|$$
$$\times[1+\zeta(\boldsymbol{\theta}-\mathbf{w}_I)^T(\mathbf{S}_I^{-1}+(n-1)(\mathbf{L}\mathbf{R}_d\mathbf{L}^T)^{-1})^{-1}(\boldsymbol{\theta}-\mathbf{w}_I)],$$
$$\mathrm{tr}[\widetilde{\mathbf{S}}_{11}^{(-)}\boldsymbol{\theta}\boldsymbol{\theta}^T]+2\mathrm{tr}[\widetilde{\mathbf{S}}_{12}^{(-)}\boldsymbol{\theta}^T]+\widetilde{S}_{22}^{(-)}$$
$$=\Big(\boldsymbol{\theta}+(\widetilde{\mathbf{S}}_{11}^{(-)})^{-1}\widetilde{\mathbf{S}}_{12}^{(-)}\Big)^T\widetilde{\mathbf{S}}_{11}^{(-)}\Big(\boldsymbol{\theta}+(\widetilde{\mathbf{S}}_{11}^{(-)})^{-1}\widetilde{\mathbf{S}}_{12}^{(-)}\Big)$$
$$-(\widetilde{\mathbf{S}}_{12}^{(-)})^T(\widetilde{\mathbf{S}}_{11}^{(-)})^{-1}\widetilde{\mathbf{S}}_{12}^{(-)}+\widetilde{S}_{22}^{(-)}$$

together with (22)–(24) we get

$$p_I\Big(\boldsymbol{\theta},\zeta\,\Big|\,(n-1)^{-1}\widetilde{\mathbf{S}}\Big) \tag{41}$$

$$\propto|\mathbf{A}|^{-(n-k+p+\nu_I+1)/2}\zeta^{(n-k+2p+2\delta_1-2)/2}$$
$$\times\exp\left\{-\frac{(n-1)\zeta}{2}\left(2\frac{\delta_2^{-1}}{n-1}+\mathrm{tr}[\widetilde{\mathbf{S}}_{11}^{(-)}\boldsymbol{\theta}\boldsymbol{\theta}^T]\ +\ 2\mathrm{tr}[\widetilde{\mathbf{S}}_{12}^{(-)}\boldsymbol{\theta}^T]+\widetilde{S}_{22}^{(-)}\right)\right\}$$
$$\propto[1+\zeta(\boldsymbol{\theta}-\mathbf{w}_I)^T(\mathbf{S}_I^{-1}$$
$$+(n-1)(\mathbf{L}\mathbf{R}_d\mathbf{L}^T)^{-1})^{-1}(\boldsymbol{\theta}-\mathbf{w}_I)]^{-(n-k+p+\nu_I+1)/2}$$
$$\times\zeta^{(n-k+2p+2\delta_1-2)/2}\exp\left\{-\frac{(n-1)\zeta}{2}\left(\frac{\delta_2^{-1}}{n-1}+\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1}\right.\right.$$
$$\left.\left.+(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}})^T(\mathbf{L}\mathbf{R}_d\mathbf{L}^T)^{-1}(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}})\right)\right\}. \tag{42}$$

Let $U(a,\,b,\,z)$ denote the confluent hypergeometric function (Abramowitz & Stegun, 1972) expressed as

$$U(a,b,z)=\frac{1}{\Gamma(a)}\int_0^\infty t^{a-1}\exp\{-zt\}(1+t)^{b-a-1}\mathbf{d}t$$

for $a=(n-k+2p+2\delta_1)/2,\ b=(p+2\delta_1-\nu_I+1)/2,$ and $z=g(\boldsymbol{\theta})$ with

$$g(\boldsymbol{\theta})=\frac{n-1}{2}\frac{((\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}})^T(\mathbf{L}\mathbf{R}_d\mathbf{L}^T)^{-1}(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}})+(\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1})^{-1})+\frac{\delta_2^{-1}}{n-1}}{(\boldsymbol{\theta}-\mathbf{w}_I)^T(\mathbf{S}_I^{-1}+(n-1)(\mathbf{L}\mathbf{R}_d\mathbf{L}^T)^{-1})^{-1}(\boldsymbol{\theta}-\mathbf{w}_I)}. \tag{43}$$

Then, the posterior for $\boldsymbol{\theta}$ is given by

$$p_I(\boldsymbol{\theta}|(n-1)^{-1}\widetilde{\mathbf{S}})\propto\int[1+\zeta(\boldsymbol{\theta}-\mathbf{w}_I)^T(\mathbf{S}_I^{-1}$$
$$+(n-1)(\mathbf{L}\mathbf{R}_d\mathbf{L}^T)^{-1})^{-1}(\boldsymbol{\theta}-\mathbf{w}_I)]^{-(n-k+p+\nu_I+1)/2}$$
$$\times\zeta^{(n-k+2p+2\delta_1-2)/2}\exp\left\{-\frac{(n-1)\zeta}{2}\left((\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}})^T(\mathbf{L}\mathbf{R}_d\mathbf{L})^{-1}(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}})\right.\right.$$
$$\left.\left.+\ (\mathbf{1}^T\mathbf{S}^{-1}\mathbf{1})^{-1}+\frac{\delta_2^{-1}}{n-1}\right)\right\}\mathbf{d}\zeta$$
$$\propto[(\boldsymbol{\theta}-\mathbf{w}_I)^T(\mathbf{S}_I^{-1}+(n-1)(\mathbf{L}\mathbf{R}_d\mathbf{L}^T)^{-1})^{-1}(\boldsymbol{\theta}-\mathbf{w}_I)]^{(n-k+2p+2\delta_1)/2}$$
$$\times U((n-k+2p+2\delta_1)/2;\,(p+2\delta_1-\nu_I+1)/2;\,g(\boldsymbol{\theta})).$$

This completes the proof. □

References

Abramowitz, M., & Stegun, I. (1972). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. New York: Dover.

Ang, A., & Bekaert, G. (2002). International asset allocation with regime shifts. *Review of Financial Studies, 15*(4), 1137–1197.

Avramov, D., & Zhou, G. (2010). Bayesian portfolio analysis. *Annual Review of Financial Economics, 2*, 25–47.

Barberis, N. (2000). Investing for the long run when returns are predictable. *Journal of Finance, 55*, 225–264.

Barry, C. (1974). Portfolio analysis under uncertain means, variances, and covariances. *Journal of Finance, 29*, 515–522.

Bawa, V. S., Brown, S., & Klein, R. (1979). *Estimation risk and optimal portfolio choice*. Amsterdam: North-Holland.

Bernard, C., & Vanduffel, S. (2014). Mean–variance optimal portfolios in the presence of a benchmark with applications to fraud detection. *European Journal of Operational Research, 234*(2), 469–480.

Bernardo, J. (2005). *Handbook of statistics: Vol.49* (pp. 17–90). Elsevier.

Best, M., & Grauer, R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results. *Review of Financial Studies, 4*(2), 315–342.

Bodnar, O. (2009). Sequential surveillance of the tangency portfolio weights. *International Journal of Theoretical and Applied Finance, 12*, 797–810.

Bodnar, T., Parolya, N., & Schmid, W. (2013). On the equivalence of quadratic optimization problems commonly used in portfolio theory. *European Journal of Operational Research, 229*(3), 637–644.

Bodnar, T., Parolya, N., & Schmid, W. (2015a). A closed-form solution of the multi--period portfolio choice problem for a quadratic utility function. *Annals of Operations Research, 229*(1), 121–158.

Bodnar, T., Parolya, N., & Schmid, W. (2015b). Estimation of the global minimum variance portfolio in high dimensions. *Technical Report*.

Bodnar, T., Parolya, N., & Schmid, W. (2015c). On the exact solution of the multi--period portfolio choice problem for an exponential utility under return predictability. *European Journal of Operational Research, 246*(2), 528–542.

Bodnar, T., & Schmid, W. (2008). A test for the weights of the global minimum variance portfolio in an elliptical model. *Metrika, 67*, 127–143.

Brandt, M. (2010). *Handbook of financial econometrics* (pp. 269–336). North Holland.

Britten-Jones, M. (1999). The sampling error in estimates of mean-variance efficient portfolio weights. *Journal of Finance, 54*, 655–671.

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

Chopra, V. K., & Ziemba, W. T. (1993). The effect of errors in means, variances and covariances on optimal portfolio choice. *Journal of Portfolio Management, 234*, 6–11.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press.

Dhaene, J., Vanduffel, S., Goovaerts, M., Kaas, R., & Vyncke, D. (2005). Comonotonic approximations for optimal portfolio selection problems. *The Journal of Risk and Insurance, 72*, 253–300.

French, K., & Poterba, J. (1991). Investor diversification and international equity markets. *American Economic Review, 81*, 222–226.

Frost, P., & Savarino, J. (1986). An empirical Bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis, 21*, 293–305.

Goetzmann, W., Ingersoll, J., Spiegel, M., & Welch, I. (2007). Portfolio performance manipulation and manipulation-proof performance measures. *Review of Financial Studies, 20*(5), 1503–1546.

Golosnoy, V., & Okhrin, Y. (2007). Multivariate shrinkage for optimal portfolio weights. *European Journal of Finance, 13*, 441–458.

Golosnoy, V., & Okhrin, Y. (2008). General uncertainty in portfolio selection: a case-based decision approach. *Journal of Economic Behavior and Organization, 67*, 718–734.

Greyserman, A., Jones, D., & Strawderman, W. (2006). Portfolio selection using hierarchical Bayesian analysis and mcmc methods. *Journal of Banking & Finance, 30*, 669–678.

Gupta, A., & Nagar, D. (2000). *Matrix Variate Distributions*. Chapman and Hall/CRC.

Harville, D. A. (1997). *Matrix Algebra From Statistician's Perspective*. Springer, New York.

Haugen, R. (1999). *The New Finance, The Case Against Efficient Markets*. Prentice Hall.

Jobson, J., & Korkie, B. (1989). A performance interpretation of multivariate tests of asset set intersection, spanning, and mean-variance efficiency. *Journal of Financial and Quantitative Analysis, 24*, 185–204.

Jorion, P. (1986). Bayes-stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis, 21*, 293–305.

Kan, R., & Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis, 42*, 621–656.

Kandel, S., & Stambaugh, R. F. (1996). On the predictability of stock returns: an asset-allocation perspective. *Journal of Finance, 51*, 385–424.

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

Levy, H., & Levy, M. (2014). The benefits of differential variance-based constraints in portfolio optimization. *European Journal of Operational Research, 234*(2), 372–381.

Liesiö, J., & Salo, A. (2012). Scenario-based portfolio selection of investment projects with incomplete probability and utility information. *European Journal of Operational Research, 217*(1), 162–172.

Magnus, J. R., & Neudecker, H. (2007). *Matrix differential calculus with applications in statistics and econometrics*. John Wiley & Sons Ltd., New York.
Markowitz, H. (2014). Mean-variance approximations to expected utility. *European Journal of Operational Research, 234*(2), 346–355.
Markowitz, H. M. (1952). Mean-variance analysis in portfolio choice and capital markets. *Journal of Finance, 7*, 77–91.
Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics, 8*, 323–361.
Muirhead, R. J. (1982). *Aspects of multivariate statistical theory*. Wiley, New York.
Okhrin, Y., & Schmid, W. (2007). Comparison of different estimation techniques for portfolio selection. *Advances in Statistical Analysis, 91*, 109–127.
Pastor, L. (2000). Portfolio selection and asset pricing models. *Journal of Finance, 55*, 179–223.
Stambaugh, R. (1997). Analyzing investments whose histories differ in length. *Journal of Financial Economics, 45*, 285–331.
Tu, J., & Zhou, G. (2010). Incorporating economic objectives into Bayesian priors: portfolio choice under parameter uncertainty. *Journal of Financial and Quantitative Analysis, 45*, 959–986.
Tunaru, R. (2002). Hierarchical Bayesian models for multiple count data. *Austrian Journal of Statistics, 31*, 221–229.
Wang, Z. (2005). A shrinkage approach to model uncertainty and asset allocation. *Review of Financial Studies, 18*, 673–705.
Winkler, R. (1973). Bayesian models for forecasting future security prices. *Journal of Financial and Quantitative Analysis, 8*, 387–405.
Winkler, R. L., & Barry, C. B. (1975). A Bayesian model for portfolio selection and revision. *Journal of Finance, 30*, 179–192.