# Bayesian inference of the multi-period optimal portfolio for an exponential utility

Bauder, David; Bodnar, Taras; Parolya, Nestor; Schmid, Wolfgang

**Important note**
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

# Bayesian inference of the multi-period optimal portfolio for an exponential utility

David Bauder [a], Taras Bodnar [b],[*], Nestor Parolya [c], Wolfgang Schmid [d]

[a] *Department of Mathematics, Humboldt-University of Berlin, D-10099 Berlin, Germany*
[b] *Department of Mathematics, Stockholm University, SE-10691 Stockholm, Sweden*
[c] *Institute of Applied Mathematics, Delft University of Technology, Delft, Netherlands*
[d] *Department of Statistics, European University Viadrina, D-15207 Frankfurt (Oder), Germany*

**ABSTRACT**

We consider the estimation of the multi-period optimal portfolio obtained by maximizing an exponential utility. Employing the Jeffreys non-informative prior and the conjugate informative prior, we derive stochastic representations for the optimal portfolio weights at each time point of portfolio reallocation. This provides a direct access not only to the posterior distribution of the portfolio weights but also to their point estimates together with uncertainties and their asymptotic distributions. Furthermore, we present the posterior predictive distribution for the investor's wealth at each time point of the investment period in terms of a stochastic representation for the future wealth realization. This in turn makes it possible to use quantile-based risk measures or to calculate the probability of default, i.e the probability of the investor wealth to become negative. We apply the suggested Bayesian approach to assess the uncertainty in the multi-period optimal portfolio by considering assets from the FTSE 100 in the weeks after the British referendum to leave the European Union. The behaviour of the novel portfolio estimation method in a precarious market situation is illustrated by calculating the predictive wealth, the risk associated with the holding portfolio, and the probability of default in each period.

© 2019 Elsevier Inc. All rights reserved.

## 1. Introduction

In portfolio theory, the mean–variance paradigm introduced by [36] is still a popular reference for understanding the relationship between systematic risk, return and investment behaviour. A portfolio is determined here by using the asset expected returns and their covariances. As a starting point, [36] was vastly extended in the following 70 years. While [36] focused only on a single investment period, the multi-period solution was introduced in [37]. Merton [39] showed that the mean–variance multi-period setting in the continuous time case is equivalent to expected utility maximization for an exponential utility function. The multi-period optimal portfolio choice problems for different utility functions were considered by [6,20,25,41,49].

While these studies focus on the continuous time case, [14,15,22,35], presented the results in the discrete time case for the quadratic utility function and the exponential utility function. In particular, [15] derived an analytical expression for the multi-period optimal portfolio weights under the assumption of non-tradable predictable variables and

---

a VAR(1)-structure which are described as linear combinations of the precision matrix (inverse covariance matrix) and the expected return vector. While this setting allows for flexibility in building trading strategies under quite unrestrictive assumptions, there are still shortcomings: (i) since the parameters of the asset return distribution, namely the mean vector and the covariance matrix, are unknown quantities, the optimal portfolio weights cannot be constructed in practice and they are obtained by replacing the unknown parameter of the asset return distribution by the corresponding estimates; (ii) although the distributional properties of the estimated optimal portfolio weights and corresponding inference procedures were derived in a number of literature studies for the single-period investment strategies (see, e.g., [10,12,16–18,28,43,51,52]), the problem with the overlapping estimation windows appears to be very crucial under the multi-period setting; (iii) due to the multivariate structure, the determination of the joint distribution of the estimated multi-period optimal portfolio weights is a challenging task.

To tackle all these three challenges, we opt for a Bayesian approach. The Bayesian approach is a well established method for building trading strategies in a single-period optimal portfolio choice problem, starting with [54] and [55] and continued until this day. For an overview, see, e.g., [19] where also Bayesian portfolio methods are discussed, or [3]. As [3] pointed out, the Bayesian setting is a realistic description of human decision making processes and information utilization. Both past events and experiences influence the beliefs of market participants at least up to a certain degree how an investment will develop. The investor beliefs are modelled via a prior distributions which represents the relevant information regarding the behaviour of the asset returns. While there is a plenty of possibilities to specify the prior, we focus on the non-informative diffuse prior and the informative conjugate prior (see, e.g., [27,56]) not only for computational reasons but mainly because of their popularity in the financial literature (c.f., [1,3,5,11,21,26,33,48,50]). Furthermore, their application allows to derive the corresponding posterior distributions in closed-form what enables us to access important risk measures and to construct credible sets. Finally, the Bayesian approach to the construction of optimal portfolio that is based on the posterior predictive distribution derived in the case of the factor model was developed by [2,44,45,47] among others. In contrast the results of the present paper do not use any particular structure of the asset pricing model and allow to directly access the posterior distribution of the unknown parameters of the data-generating process used to determine the weights of the multi-period optimal portfolio.

The obtained posterior distributions of the optimal portfolio weights under both employed priors are presented in terms of their stochastic representations. A stochastic representation is a well established tool in computational statistics (c.f., [29]) and in the theory of elliptically contoured distributions (see, e.g. [31]) which was already used in Bayesian statistics by [7,11,57]. Among others [57] argued that the direct Monte Carlo approach is a computationally efficient method to calculate Bayesian estimates. It turns out that the derived stochastic representations are very powerful, allowing us to access not only the posterior distribution of the multi-period optimal portfolio weights, but also to determine the predictive distribution for the wealth at each point of the holding period. Therefore, we are able to access the quantiles for the posterior predictive wealth distribution and can calculate the risk associated with the portfolio at every point over the lifetime of a portfolio, besides analytical Bayesian estimates for the weights together with their uncertainties. Besides these pleasing properties, the developed stochastic representations are highly efficient from a computational point of view since Markov chain Monte Carlo methods are no longer needed. In addition to the derivation of these results, we illustrate this method and its properties on real data. We test the model in an exhaustive study using data from the FTSE 100 index (Financial Times Stock Exchange Index), where the portfolios cover the time of Great Britains referendum to leave the European Union on 23.6.2016, more commonly regarded as "brexit", where a slim majority of British voters decided to leave the European Union. Although this result was regarded as a less likely option in advance, it was regarded as the option with the least favourable effects on the British economy and should therefore have an effect on a portfolio covering this period.

The remaining paper is structured in the following way. In Section 2, we briefly review the solution of the multi-period optimal portfolio choice problem with the exponential utility derived in [15]. The stochastic representations for the optimal portfolio weights under both priors are presented in Theorems 1 and 2 (Section 2.2), which are used to derive the corresponding Bayes estimates for the weights (Theorem 3) together with their covariance matrix (Theorem 4) as well as to prove the posterior asymptotic normality (Theorem 5). In Section 2.3, we obtain the posterior predictive distribution for the wealth during the holding period which is provided in terms of stochastic representation in Theorem 6 under both employed priors. In Section 3, the suggested Bayesian approach is applied to the brexit-data by calculating the asymptotic distributions for the optimal portfolio weights, determining the credible sets for the portfolio wealth and specifying the default probabilities at each time point. Section 4 summarizes the main results of the paper, while all technical proofs are moved to Appendix A.

## 2. Bayesian analysis of multi-period optimal portfolios

### 2.1. Analytical solution of the multi-period optimization problem

Let $\mathbf{X}_t = (X_{t,1}, X_{t,2}, \ldots, X_{t,k})^\top$ be a random vector of returns on $k$ assets taken at time point $t$. Throughout the paper we assume that the asset returns $\mathbf{X}_1, \mathbf{X}_2, \ldots$ are infinitely exchangeable and multivariate centred spherically symmetric. This assumption, in particular, implies (see, e.g., Bernardo and Smith [8, Proposition 4.6]) that the asset returns are independently and identically distributed given the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ with the conditional

distribution given by $\mathbf{X}_t|\boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ($k$-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$). It is noted that the imposed assumption implies that neither the unconditional distribution of the asset returns is normal nor that they are independently distributed. Moreover, the unconditional distribution of the asset returns appears to be heavy-tailed which is usually observed for financial data.

The quantities $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote the parameters of the asset returns distribution where $\boldsymbol{\Sigma}$ is assumed to be a positive definite matrix of size $k \times k$. We consider a multi-period portfolio choice problem with the allocation of initial wealth at time point $t = 0$ and with the subsequent update of the portfolio structure at time points $t \in \{1, 2, \ldots, T\}$. Let $\mathbf{v}_t = (v_{t,1}, \ldots, v_{t,k})^\top$ stand for the vector of portfolio weights determined at time $t$ and let $r_{f,t}$ be the return on the risk-free asset in period $t$. We assume that short-selling is allowed, i.e., the weights could also be negative. The vector $\mathbf{v}_t$ specifies the structure of the portfolio related to the risky assets, whereas the part of the wealth equal to $1 - \mathbf{1}^\top \mathbf{v}_t$ is invested into the risk-free asset where $\mathbf{1}$ denotes the $k$-dimensional vector of ones. Then the investor's wealth in period $t$ is expressed as

$$W_t = W_{t-1}(1 + (1 - \mathbf{1}^\top \mathbf{v}_{t-1})r_{f,t} + \mathbf{v}_{t-1}^\top \mathbf{X}_t) = W_{t-1}(1 + r_{f,t} + \mathbf{v}_{t-1}^\top(\mathbf{X}_t - r_{f,t}\mathbf{1})).$$

An investor seeks to maximize the utility of the final wealth, i.e. $U(W_T)$, where $U(x) = -\exp(-\gamma x)$ is the exponential utility function and the coefficient of absolute risk aversion, $\gamma > 0$, determines the investor's attitude towards risk. The optimization problem is given by

$$V(0, W_0) = \max_{\{\mathbf{v}_s\}_{s=0}^{T-1}} \mathbb{E}_0[U(W_T)], \tag{1}$$

where the maximum is taken with respect to all weights $\mathbf{v}_0, \ldots, \mathbf{v}_{T-1}$ which specify the portfolio structure during the initial period of investment as well as during all consequent reallocations. The solution of (1) is derived in recursive way starting from the last period by applying Bellman equations at $0, 1, \ldots, T - 1$. The optimization problem at time point $T - t$ is then given by

$$V(T - t, W_{T-t}) = \max_{\{\mathbf{v}_s\}_{s=T-t}^{T-1}} \mathbb{E}_{T-t} \left[ \max_{\{\mathbf{v}_s\}_{s=T-t+1}^{T-1}} \mathbb{E}_{T-t+1}[U(W_T)] \right]$$
$$= \max_{\mathbf{v}_{T-t}} \mathbb{E}_{T-t} \left[ V(T - t + 1, W_{T-t}(r_{f,T-t} + \mathbf{w}_{T-t+1}^\top(\mathbf{X}_{T-t+1} - r_{f,T-t+1}\mathbf{1}))) \right]$$

subject to the terminal condition $U(W_T) = -\exp(-\gamma W_T)$ with $\mathbf{w}_{T-t+1}$ as the optimal portfolio weights in period $T-t+1$. For details on this method, see e.g. [46], while [15] determine an analytical solution of (1) under the exponential utility. The latter results are summarized in Proposition 1.

**Proposition 1.** *Let $\mathbf{X}_t$, $t \in \{0, \ldots, T\}$ be a sequence of conditionally independently and identically distributed vectors of $k$ risky assets with $\mathbf{X}_t|\boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let $\boldsymbol{\Sigma}$ be positive definite. Then the optimal multi-period portfolio weights are given by*

$$\mathbf{w}_t = C_t \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_{f,t+1}\mathbf{1}), \, C_t = (\gamma W_t \prod_{i=t+2}^{T} R_{f,i})^{-1} \tag{2}$$

*for $t \in \{0, \ldots, T - 1\}$ where $R_{f,i} = 1 + r_{f,i}$ and $\prod_{i=T+1}^{T} R_{f,i} \equiv 1$.*

Although Proposition 1 provides a simple solution of the multi-period portfolio choice problem, (2) cannot directly be applied in practice since $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown parameters of the asset return distribution. As a result, these two quantities have to be estimated before the portfolio (2) is constructed. However, the usage of the estimated mean vector and the estimated covariance matrix instead of the population ones does not ensure that the estimated portfolio weights coincide with the true ones. Then two main questions raise: (i) how strongly deviates the estimated portfolio from the population one? and (ii) is it reasonable to invest into the estimated portfolio? Both questions have to be treated by using statistical methods and are very closely connected to the distributional properties of the estimates for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

The traditional approach of estimating the portfolio weights relies on the methods from classical statistics where the sample mean vector and the sample covariance matrix are used. Let $\mathbf{x}_{t-n+1}, \ldots, \mathbf{x}_t$ be the observation vectors of asset returns which are considered as realizations of the corresponding random vectors $\mathbf{X}_i$, $i \in \{t-n+1, \ldots, t\}$. Then the mean vector and the covariance matrix at time point $t$ are estimated by

$$\bar{\mathbf{x}}_t = \frac{1}{n} \sum_{i=t-n+1}^{t} \mathbf{x}_i, \quad \mathbf{S}_t = \frac{1}{n-1} \sum_{i=t-n+1}^{t} (\mathbf{x}_i - \bar{\mathbf{x}}_t)(\mathbf{x}_i - \bar{\mathbf{x}}_t)^\top, \tag{3}$$

respectively. The sample estimate of the multi-period optimal portfolio is obtained by replacing $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in (2) by the corresponding estimates from (3). This leads to

$$\hat{\mathbf{w}}_t = C_t \mathbf{S}_t^{-1}(\bar{\mathbf{x}}_t - r_{f,t+1}\mathbf{1}), \quad C_t = (\gamma W_t \prod_{i=t+2}^{T} R_{f,i})^{-1}, \quad t \in \{0, \ldots, T - 1\}. \tag{4}$$

Using the findings in [13], we obtain the density function, the moments and the stochastic representation of the sample multi-period optimal portfolio weights from the viewpoint of frequentistic statistics. These results provide answers on the above two presented questions and allow us to characterize the distributional properties of each vector of weights $\hat{\mathbf{w}}_t$ separately. On the other hand, they do not take into account the multi-period nature of the considered investment procedure. More precisely, it is not possible to provide the characterization of the whole multi-period optimal portfolio, since the overlapping samples are used and the dependence structure between the estimated portfolio weights becomes severe.

For that reason, we deal with the problem of estimating the multi-period optimal portfolio from the viewpoint of Bayesian statistics and consider the portfolio constructed by using (4) as a benchmark portfolio without investigating its distributional properties in detail. In contrast to the methods of the frequentistic statistics, the application of the Bayesian approach allows the sequential update of the available information which is a very important property needed for estimating the multi-period portfolio weights.

### 2.2. Bayesian estimation of portfolio weights

Let $\mathbf{x}_{t,n} = (\mathbf{x}_{t-n+1}, \ldots, \mathbf{x}_t)$ denote the observation matrix at time point $t$ which consists of $n$ asset return vectors from $t - n + 1$ to $t$. According to Bayes theorem, the beliefs regarding $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are updated in the presence of occurring data, yielding the posterior distribution $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{x}_{t,n})$ to be proportional to the product of the likelihood function $L(\mathbf{x}_{t,n}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and the prior distribution $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The posterior is then used to derive Bayesian estimates for the multi-period optimal portfolio weights as well as their characteristics, like the covariance matrix and a credible region which is an analogue to a confidence region in frequentistic statistics. Bayes theorem states that

$$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{x}_{t,n}) \propto L(\mathbf{x}_{t,n}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

The choice of the prior $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is an important step in the Bayesian decision process. Although the prior should reflect the investor's belief regarding the parameters of the asset return distribution, it also strongly affects the model's computational properties since it influences the accessibility of the posterior distribution. Several priors for the mean vector and covariance matrix of the asset returns have been suggested in literature (see, e.g., [3,5,21,26,33,48,50]) with the recent paper of [11] summarizing these results. In the following, we choose the Jeffreys non-informative prior and a conjugate informative prior for both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. These two priors are widely used in the context of Bayesian inference of optimal portfolios.

The Jeffreys non-informative prior, also known as the diffuse prior, is given by

$$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(k+1)/2} \tag{5}$$

while the conjugate prior is expressed as

$$\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim \mathcal{N}_k\left(\mathbf{m}_0, \frac{1}{r_0}\boldsymbol{\Sigma}\right), \quad \boldsymbol{\Sigma} \sim \mathcal{IW}_k(d_0, \mathbf{S}_0), \tag{6}$$

where $\mathbf{m}_0, r_0, d_0, \mathbf{S}_0$ are additional model parameters known as hyperparameters. The symbol $\mathcal{IW}_k(d_0, \mathbf{S}_0)$ denotes the inverse Wishart distribution with $d_0$ degrees of freedom and parameter matrix $\mathbf{S}_0$ (see, [30]). The prior mean $\boldsymbol{\mu}_0$ reflects our prior expectations about the expected asset returns, while $\mathbf{S}_0$ represents in the model the prior beliefs about the covariance matrix. The other two hyperparameters $r_0$ and $d_0$ are known as precision parameters for $\boldsymbol{\mu}_0$ and $\mathbf{S}_0$, respectively. Note that the prior (6) corresponds to the well-known conjugate normal–inverse-Wishart model as discussed by, e.g., [27]. In this case the posterior is accessible in an analytical form and moreover, has the same distribution as the prior with updated hyperparameters.

In Proposition 2, we present the marginal posterior of $\boldsymbol{\mu}$ as well as the conditional posterior of $\boldsymbol{\Sigma}$ given $\boldsymbol{\mu}$. These results will be later used in the derivation of Bayesian estimates for the optimal portfolio weights. In the following the symbol $t_k(d, \mathbf{a}, \mathbf{A})$ stands for the multivariate $k$-dimensional $t$-distribution with $d$ degrees of freedom, location vector $\mathbf{a}$ and dispersion matrix $\mathbf{A}$. In the case of $k = 1$, $\mathbf{a} = 0$, and $\mathbf{A} = 1$, we use the notation $t_d$ to denote the standard univariate $t$-distribution with $d$ degrees of freedom.

**Proposition 2.** *Let $\mathbf{X}_{t-n+1}, \ldots, \mathbf{X}_t$ be conditionally independently distributed with $\mathbf{X}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $i \in \{t-n+1, \ldots, t\}$ with $n > k$. Then:*

*(a) Under the diffuse prior (5), the marginal posterior distribution of $\boldsymbol{\mu}$ is given by*

$$\boldsymbol{\mu}|\mathbf{x}_{t,n} \sim t_k\left(n - k, \bar{\mathbf{x}}_{t,d}, \frac{1}{n(n-k)}\mathbf{S}_{t,d}\right), \quad \bar{\mathbf{x}}_{t,d} = \bar{\mathbf{x}}_t, \quad \mathbf{S}_{t,d} = (n-1)\mathbf{S}_t.$$

*The conditional posterior distribution of $\boldsymbol{\Sigma}$ given $\boldsymbol{\mu}$ is expressed as*

$$\boldsymbol{\Sigma}|\boldsymbol{\mu}, \mathbf{x}_{t,n} \sim \mathcal{IW}_k(n + k + 1, \mathbf{S}_{t,d}^*(\boldsymbol{\mu})), \quad \mathbf{S}_{t,d}^*(\boldsymbol{\mu}) = \mathbf{S}_{t,d} + n(\boldsymbol{\mu} - \bar{\mathbf{x}}_{t,d})(\boldsymbol{\mu} - \bar{\mathbf{x}}_{t,d})^\top.$$

(b) Under the conjugate prior (6), the marginal posterior distribution of $\boldsymbol{\mu}$ is given by

$$\boldsymbol{\mu}|\mathbf{x}_{t,n} \sim t_k\left(n + d_0 - 2k, \bar{\mathbf{x}}_{t,c}, \frac{1}{(n+r_0)(n+d_0-2k)}\mathbf{S}_{t,c}\right),$$

$$\bar{\mathbf{x}}_{t,c} = \frac{n\bar{\mathbf{x}}_t + r_0\mathbf{m}_0}{n+r_0}, \quad \mathbf{S}_{t,c} = \mathbf{S}_{t,d} + \mathbf{S}_0 + nr_0\frac{(\mathbf{m}_0 - \bar{\mathbf{x}}_{t,c})(\mathbf{m}_0 - \bar{\mathbf{x}}_{t,c})^\top}{n+r_0}.$$

The conditional posterior distribution of $\boldsymbol{\Sigma}$ given $\boldsymbol{\mu}$ is expressed as

$$\boldsymbol{\Sigma}|\boldsymbol{\mu}, \mathbf{x}_{t,n} \sim IW_k(n + d_0 + 1, \mathbf{S}_{t,c}^*(\boldsymbol{\mu})), \quad \mathbf{S}_{t,c}^*(\boldsymbol{\mu}) = \mathbf{S}_{t,c} + (n+r_0)(\boldsymbol{\mu} - \bar{\mathbf{x}}_{t,c})(\boldsymbol{\mu} - \bar{\mathbf{x}}_{t,c})^\top.$$

The proof of Proposition 2 follows from Chapter 3 in [27] who presented the expressions of the marginal posterior distributions of $\boldsymbol{\mu}$ under both the diffuse and the conjugate priors. Then, the results for the conditional posteriors of $\boldsymbol{\Sigma}$ are obtained from the joint posterior distributions using the formulae for the marginal posteriors for $\boldsymbol{\mu}$. It is remarkable that although the results for the marginal posteriors for both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are widely used in Bayesian inferences and the conditional posteriors for $\boldsymbol{\mu}$ given $\boldsymbol{\Sigma}$ have been considered previously in literature (see, e.g., [53]), the results for the conditional posteriors of $\boldsymbol{\Sigma}$ given $\boldsymbol{\mu}$ have not been discussed nor used. Next, we show that the last finding allows to derive posterior distributions for functions which includes both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

In order to assess the risk associated with estimating the optimal portfolio weights, we need to derive results about the posterior distribution of the weights presented in Proposition 1 which are given as a product of the inverse covariance matrix and the mean vector. Next, we establish very useful stochastic representations for these weights, endowing the parameters with their diffuse and conjugate priors. The results are summarized in Theorem 1, where the stochastic representations are derived for an arbitrary linear combination of optimal portfolio weights. These findings are later used for calculating the Bayesian estimates of the portfolio weights (Theorem 3) and their covariance matrix (Theorem 4). It is noted that the application of the stochastic representation to describe the distribution of random quantities has been used both in the conventional statistics (see, e.g., [29,31]) and the Bayesian statistics (see, [11]). Later on, the symbol "$\stackrel{d}{=}$" denotes equality in distribution. The proof of Theorem 1 is presented in Appendix A.

**Theorem 1.** Let $\mathbf{L}$ be a $p \times k$ matrix of constants. Then under the assumption of Proposition 2 we get:

(a) Under the diffuse prior (5), the stochastic representation of $\mathbf{Lw}_t$ is given by

$$\mathbf{Lw}_t \stackrel{d}{=} C_t\eta\mathbf{L}\mathbf{S}_{t,d}^*(\boldsymbol{\mu})^{-1}(\boldsymbol{\mu} - r_{f,t+1}) + C_t\sqrt{\eta}\Big((\boldsymbol{\mu} - r_{f,t+1})^\top\mathbf{S}_{t,d}^*(\boldsymbol{\mu})^{-1}(\boldsymbol{\mu} - r_{f,t+1})\mathbf{L}\mathbf{S}_{t,d}^*(\boldsymbol{\mu})^{-1}\mathbf{L}^\top$$

$$- \mathbf{L}\mathbf{S}_{t,d}^*(\boldsymbol{\mu})^{-1}(\boldsymbol{\mu} - r_{f,t+1})(\boldsymbol{\mu} - r_{f,t+1})^\top\mathbf{S}_{t,d}^*(\boldsymbol{\mu})^{-1}\mathbf{L}^\top\Big)^{1/2}\mathbf{z}_0,$$

where $\eta \sim \chi_n^2$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$, and $\boldsymbol{\mu}|\mathbf{x}_{t,n} \sim t_k\left(n - k, \bar{\mathbf{x}}_{t,d}, \mathbf{S}_{t,d}/(n(n-k))\right)$. Moreover, $\eta$, $\mathbf{z}_0$ and $\boldsymbol{\mu}$ are mutually independent.

(b) Under the conjugate prior (6), the stochastic representation of $\mathbf{Lw}_t$ is given by

$$\mathbf{Lw}_t \stackrel{d}{=} C_t\eta\mathbf{L}\mathbf{S}_{t,c}^*(\boldsymbol{\mu})^{-1}(\boldsymbol{\mu} - r_{f,t+1}) + C_t\sqrt{\eta}\Big((\boldsymbol{\mu} - r_{f,t+1})^\top\mathbf{S}_{t,c}^*(\boldsymbol{\mu})^{-1}(\boldsymbol{\mu} - r_{f,t+1})\mathbf{L}\mathbf{S}_{t,c}^*(\boldsymbol{\mu})^{-1}\mathbf{L}^\top$$

$$- \mathbf{L}\mathbf{S}_{t,c}^*(\boldsymbol{\mu})^{-1}(\boldsymbol{\mu} - r_{f,t+1})(\boldsymbol{\mu} - r_{f,t+1})^\top\mathbf{S}_{t,c}^*(\boldsymbol{\mu})^{-1}\mathbf{L}^\top\Big)^{1/2}\mathbf{z}_0,$$

where $\eta \sim \chi_{n+d_0-k}^2$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$, and $\boldsymbol{\mu}|\mathbf{x} \sim t_k\left(n + d_0 - 2k, \bar{\mathbf{x}}_{t,c}, \mathbf{S}_{t,c}/((n+r_0)(n+d_0-2k))\right)$. Moreover, $\eta$, $\mathbf{z}_0$ and $\boldsymbol{\mu}$ are mutually independent.

The results of Theorem 1 show that in both cases, i.e., when the mean vector and the covariance matrix are endowed by the diffuse prior and the conjugate prior, the obtained stochastic representations are very similar and the posterior distributions of the multi-period optimal portfolio weights from Proposition 1 can be described by three random variables which have standard univariate/multivariate distributions.

Another important application of Theorem 1 is that the results of this theorem also provide a hint how these distributions can be accessed in practice via simulations, namely by simulating samples from the $\chi^2$-distribution, the normal distribution, and the $t$-distribution. Although the derived stochastic representations have some nice computational properties in terms of speed, they are not computationally efficient. In the following theorem we derive further stochastic representations under both priors by applying the Sherman–Morrison–Woodbury formula (see, e.g., [40, p. 124]) on the inverse of the posterior scale matrices $\mathbf{S}_{t,d}^*(\boldsymbol{\mu})$ and $\mathbf{S}_{t,c}^*(\boldsymbol{\mu})$. The proof of the theorem is provided in the Appendix A. Let $\mathcal{F}(d_1, d_2)$ denote the $F$-distribution with $d_1$ and $d_2$ degrees of freedom.

**Theorem 2.** Under the assumption of Theorem 1 we get:

(a) Under the diffuse prior (5), the stochastic representation of $\mathbf{Lw}_t$ is given by

$$\mathbf{Lw}_t \stackrel{d}{=} C_t\eta\mathbf{L}\boldsymbol{\zeta}_d + C_t\sqrt{\eta}\left(\epsilon_d\mathbf{L}\boldsymbol{\Upsilon}_d\mathbf{L}^\top - \mathbf{L}\boldsymbol{\zeta}_d\boldsymbol{\zeta}_d^\top\mathbf{L}^\top\right)^{1/2}\mathbf{z}_0, \tag{7}$$

with

$$\epsilon_d = \epsilon_d(Q, \mathbf{u}) = (\bar{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})^\top \mathbf{S}_{t,d}^{-1}(\bar{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1}) + \frac{2}{\sqrt{n}} \frac{\sqrt{kQ/(n-k)}}{1 + kQ/(n-k)}(\bar{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})^\top \mathbf{S}_{t,d}^{-1/2}\mathbf{u}$$

$$+ \frac{1}{n}\frac{kQ/(n-k)}{1 + kQ/(n-k)} - \frac{kQ/(n-k)}{1 + kQ/(n-k)}\left((\bar{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})^\top \mathbf{S}_{t,d}^{-1/2}\mathbf{u}\right)^2,$$

$$\boldsymbol{\zeta}_d = \boldsymbol{\zeta}_d(Q, \mathbf{u}) = \mathbf{S}_{t,d}^{-1}(\bar{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1}) + \frac{1}{\sqrt{n}}\frac{\sqrt{kQ/(n-k)}}{1 + kQ/(n-k)}\mathbf{S}_{t,d}^{-1/2}\mathbf{u}$$

$$- \frac{kQ/(n-k)}{1 + kQ/(n-k)}\mathbf{S}_{t,d}^{-1/2}\mathbf{u}\mathbf{u}^\top \mathbf{S}_{t,d}^{-1/2}(\bar{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1}),$$

$$\boldsymbol{\Upsilon}_d = \boldsymbol{\Upsilon}_d(Q, \mathbf{u}) = \mathbf{S}_{t,d}^{-1} - \frac{kQ/(n-k)}{1 + kQ/(n-k)}\mathbf{S}_{t,d}^{-1/2}\mathbf{u}\mathbf{u}^\top \mathbf{S}_{t,d}^{-1/2},$$

where $\eta \sim \chi_n^2$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$, $Q \sim \mathcal{F}(k, n-k)$, and $\mathbf{u}$ uniformly distributed on the unit sphere in $\mathbb{R}^k$. Moreover, $\eta$, $\mathbf{z}_0$, $Q$, and $\mathbf{u}$ are mutually independent.

(b) Under the conjugate prior (6), the stochastic representation of $\mathbf{L}\mathbf{w}_t$ is given by

$$\mathbf{L}\mathbf{w}_t \stackrel{d}{=} C_t \eta \mathbf{L}\boldsymbol{\zeta}_c + C_t\sqrt{\eta}\left(\epsilon_c \mathbf{L}\boldsymbol{\Upsilon}_c\mathbf{L}^\top - \mathbf{L}\boldsymbol{\zeta}_c\boldsymbol{\zeta}_c^\top \mathbf{L}^\top\right)^{1/2}\mathbf{z}_0, \tag{8}$$

with

$$\epsilon_c = \epsilon_d(Q, \mathbf{u}) = (\bar{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1})^\top \mathbf{S}_{t,d}^{-1}(\bar{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1}) + \frac{2}{\sqrt{n+r_0}}\frac{\sqrt{kQ/(n+d_0-2k)}}{1 + kQ/(n+d_0-2k)}(\bar{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1})^\top \mathbf{S}_{t,d}^{-1/2}\mathbf{u}$$

$$+ \frac{1}{n+r_0}\frac{kQ/(n+d_0-2k)}{1 + kQ/(n+d_0-2k)} - \frac{kQ/(n+d_0-2k)}{1 + kQ/(n+d_0-2k)}\left((\bar{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1})^\top \mathbf{S}_{t,d}^{-1/2}\mathbf{u}\right)^2,$$

$$\boldsymbol{\zeta}_c = \boldsymbol{\zeta}_d(Q, \mathbf{u}) = \mathbf{S}_{t,c}^{-1}(\bar{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1}) + \frac{1}{\sqrt{n+r_0}}\frac{\sqrt{kQ/(n+d_0-2k)}}{1 + kQ/(n+d_0-2k)}\mathbf{S}_{t,c}^{-1/2}\mathbf{u}$$

$$- \frac{kQ/(n+d_0-2k)}{1 + kQ/(n+d_0-2k)}\mathbf{S}_{t,c}^{-1/2}\mathbf{u}\mathbf{u}^\top \mathbf{S}_{t,c}^{-1/2}(\bar{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1}),$$

$$\boldsymbol{\Upsilon}_c = \boldsymbol{\Upsilon}_d(Q, \mathbf{u}) = \mathbf{S}_{t,c}^{-1} - \frac{kQ/(n+d_0-2k)}{1 + kQ/(n+d_0-2k)}\mathbf{S}_{t,c}^{-1/2}\mathbf{u}\mathbf{u}^\top \mathbf{S}_{t,c}^{-1/2},$$

where $\eta \sim \chi_{n+d_0-k}^2$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$, $Q \sim \mathcal{F}(k, n+d_0-2k)$, and $\mathbf{u}$ uniformly distributed on the unit sphere in $\mathbb{R}^k$. Moreover, $\eta$, $\mathbf{z}_0$, $Q$, and $\mathbf{u}$ are mutually independent.

Theorem 2 provides alternative stochastic representations of the optimal portfolio weights obtained under the diffuse prior and under the conjugate prior. Although more difficult mathematical expressions are present in Theorem 2, they are more computationally efficient than the ones provided in Theorem 1. Namely, there is no need to calculate the inverse of the matrices $\mathbf{S}_{t,d}^*(\boldsymbol{\mu})$ and $\mathbf{S}_{t,c}^*(\boldsymbol{\mu})$ in each simulation run and instead, we only calculate the inverse of the matrices $\mathbf{S}_{t,d}$ and $\mathbf{S}_{t,c}$ once for the whole simulation study. This property surely speeds up the simulation study considerably. Finally, we note that the realizations of the random vector $\mathbf{u}$, which is uniformly distributed on the unit sphere in $\mathbb{R}^k$, are obtained by drawing $\mathbf{z}$ from the $k$-dimensional standard normal distribution and calculating $\mathbf{u} = \mathbf{z}/\sqrt{\mathbf{z}^\top\mathbf{z}}$. To this end, we note that continuous random quantities and data-driven objects are only present in (7) and (8). Since they all are finite with probability one, we also get that optimal portfolio weights are finite with probability one.

The results of Theorem 2 are used to derive Bayesian estimates for the weights of the multi-period optimal portfolio at the initial period of investment as well as at each time of reallocations. They are presented in Theorem 3.

**Theorem 3.** *Under the assumption of Theorem 1, we get*

(a) *Under the diffuse prior (5), the Bayes estimate for the optimal portfolio weights at time point t is given by*

$$\hat{\mathbf{w}}_{t,d} = \mathbb{E}(\mathbf{w}_t|\mathbf{x}_{t,n}) = C_t(n-1)\mathbf{S}_{t,d}^{-1}(\bar{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1}).$$

(b) *Under the conjugate prior (6), the Bayes estimate for the optimal portfolio weights at time point t is given by*

$$\hat{\mathbf{w}}_{t,c} = \mathbb{E}(\mathbf{w}_t|\mathbf{x}_{t,n}) = C_t(n+d_0-k-1)\mathbf{S}_{t,c}^{-1}(\bar{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1}).$$

The proof of the theorem is given in Appendix A. It is interesting to note that the estimate for the optimal portfolio weights obtained under the diffuse prior coincides with the expression derived in Section 2.1 for their frequentistic estimate since $\mathbf{S}_{t,d}/(n-1) = \mathbf{S}_t$.

Finally, we present the expressions for the covariance matrices of the optimal portfolio weights in Theorem 4 with the proof relegated to Appendix A. These formulas characterize the dependencies between the portfolio weight and also allow to access their Bayesian risk.

**Theorem 4.** *Under the assumption of Theorem* 1*, we get:*

(a) *Under the diffuse prior* (5)*, the covariance matrix of* $\mathbf{w}_t$ *is given by*

$$\mathbf{V}_{t,d} = \mathbb{V}ar(\mathbf{w}_t|\mathbf{x}_{t,n}) = C_t^2\left[(n-1)\mathbf{S}_{t,d}^{-1}(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})^\top\mathbf{S}_{t,d}^{-1} + \left(\frac{n^2 + k - 2}{n(n+2)} + \frac{k-1}{k}b_d\right)\mathbf{S}_{t,d}^{-1}\right],$$

*where* $b_d = n(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})^\top\mathbf{S}_{t,d}^{-1}(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})$;

(b) *Under the conjugate prior* (6)*, the covariance matrix of* $\mathbf{w}_t$ *is given by*

$$\mathbf{V}_{t,c} = \mathbb{V}ar(\mathbf{w}_t|\mathbf{x}_{t,n}) = C_t^2\Bigg[(n+d_0-k-1)\mathbf{S}_{t,c}^{-1}(\overline{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1})(\overline{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1})^\top\mathbf{S}_{t,c}^{-1}$$
$$+ \left(\frac{(n+d_0-k)^2 + k - 2}{(n+r_0)(n+d_0-k+2)} + \frac{(n+d_0-k)(k-1)}{(n+r_0)k}b_c\right)\mathbf{S}_{t,c}^{-1}\Bigg],$$

*where* $b_c = (n+r_0)(\overline{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1})^\top\mathbf{S}_{t,c}^{-1}(\overline{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1})$.

The results of Theorems 3 and 4 provide the first two moments of optimal portfolio weights and, consequently, they characterize their mean values, variances, and correlations. Although different formulas are obtained under the diffuse prior and under the conjugate prior, when the sample size increases the differences between the corresponding expressions become negligible.

More general results are provided in Theorem 5 where it is shown that $\mathbf{w}_t$ converge to the same asymptotic normal distribution under the diffuse prior and under the conjugate prior.

**Theorem 5.** *Under the assumption of Theorem* 1*, it holds as* $n \longrightarrow \infty$ *that*

$$\sqrt{n}(\mathbf{w}_t - \hat{\mathbf{w}}_t)|\mathbf{x}_{t,n} \xrightarrow{d} \mathcal{N}\left(0, C_t^2\left[\check{\mathbf{S}}_t^{-1}(\check{\mathbf{x}}_t - r_{f,t+1}\mathbf{1})(\check{\mathbf{x}}_t - r_{f,t+1}\mathbf{1})^\top\check{\mathbf{S}}_t^{-1} + \left(1 + \frac{k-1}{k}(\check{\mathbf{x}}_t - r_{f,t+1}\mathbf{1})^\top\check{\mathbf{S}}_t^{-1}(\check{\mathbf{x}}_t - r_{f,t+1}\mathbf{1})\right)\check{\mathbf{S}}_t^{-1}\right]\right)$$

*under both the diffuse prior and the conjugate prior where*

$$\check{\mathbf{x}}_t \equiv \lim_{n \to \infty} \overline{\mathbf{x}}_{t,d} = \lim_{n \to \infty} \overline{\mathbf{x}}_{t,c}, \quad \check{\mathbf{S}}_t \equiv \lim_{n \to \infty} \frac{\mathbf{S}_{t,d}}{n-1} = \lim_{n \to \infty} \frac{\mathbf{S}_{t,c}}{n+r_0}, \quad \hat{\mathbf{w}}_t \equiv \lim_{n \to \infty} \hat{\mathbf{w}}_{t,d} = \lim_{n \to \infty} \hat{\mathbf{w}}_{t,c} = C_t\check{\mathbf{S}}_t^{-1}(\check{\mathbf{x}}_t - r_{f,t+1}\mathbf{1}).$$

The proof of Theorem 5 is given in Appendix A. Its results are in line with the Bernstein–von Mises theorem (see, [8]) which shows under some regularity conditions that the posterior distribution converges to the normal one independently of the prior used when the sample size tends to infinity. In practice, the asymptotic covariance matrix of $\mathbf{w}_t$ is approximated by using $\overline{\mathbf{x}}_t$ and $\mathbf{S}_t$ instead of $\check{\mathbf{x}}_t$ and $\check{\mathbf{S}}_t$.

## 2.3. Posterior predictive distribution

In this section we derive the posterior predictive distribution of the wealth at time point $t + 1$, $\widehat{W}_{t+1}$, given the observable data $\mathbf{x}_{t,n}$ under the diffuse prior (5) and the conjugate prior (6) for the given vector of portfolio weights $\mathbf{v}_t$ and the current wealth $W_t$. Namely, the aim is to derive the posterior predictive distribution of

$$W_{t+1} = W_t(1 + r_{f,t} + \mathbf{v}_t^\top(\mathbf{X}_{t+1} - r_{f,t+1})) \tag{9}$$

given information provided by the observation matrix $\mathbf{x}_{t,n}$, i.e.,

$$f_{\widehat{W}_{t+1}}(w|\mathbf{x}_{t,n}) = \int_{\boldsymbol{\mu},\boldsymbol{\Sigma}} f_{\widehat{W}_{t+1}}(w|\boldsymbol{\mu},\boldsymbol{\Sigma},\mathbf{x}_{t,n})\pi(\boldsymbol{\mu},\boldsymbol{\Sigma}|\mathbf{x}_{t,n})\mathrm{d}\boldsymbol{\mu}\mathrm{d}\boldsymbol{\Sigma},$$

where $\pi(\boldsymbol{\mu},\boldsymbol{\Sigma}|\mathbf{x}_{t,n})$ is the posterior distribution obtained under the diffuse prior or the conjugate prior. The symbol $\widehat{W}_{t+1}$ denotes a random variable whose distribution coincides with the posterior predictive distribution of the wealth calculated at time point $t + 1$.

In Theorem 6 we present the stochastic representations of the posterior predictive distribution of $\widehat{W}_{t+1}$ with the proof given in Appendix A. The symbol $t_d$ stands for the standard univariate $t$-distribution with $d$ degrees of freedom.

**Theorem 6.** *Under the assumption of Theorem 1 we get:*

(a) *Under the diffuse prior* (5), *the stochastic representation of the posterior predictive distribution of $W_{t+1}$ is given by*

$$\widehat{W}_{t+1} \overset{d}{=} W_t \left( 1 + r_{f,t+1} + \mathbf{v}_t^\top (\overline{\mathbf{x}}_{t,d} - r_{f,t+1}) + \sqrt{\mathbf{v}_t^\top \mathbf{S}_{t,d} \mathbf{v}_t} \left( \frac{t_1}{\sqrt{n(n-k)}} + \sqrt{1 + \frac{t_1^2}{n-k}} \frac{t_2}{\sqrt{n-k+1}} \right) \right)$$

*where $t_1$ and $t_2$ are independent with $t_1 \sim t_{n-k}$ and $t_2 \sim t_{n-k+1}$;*

(b) *Under the conjugate prior* (6), *the stochastic representation of the posterior predictive distribution of $W_{t+1}$ is given by*

$$\widehat{W}_{t+1} \overset{d}{=} W_t \left( 1 + r_{f,t+1} + \mathbf{v}_t^\top (\overline{\mathbf{x}}_{t,c} - r_{f,t+1}) \right.$$

$$\left. + \sqrt{\mathbf{v}_t^\top \mathbf{S}_{t,c} \mathbf{v}_t} \left( \frac{t_1}{\sqrt{(n+r_0)(n+d_0-2k)}} + \sqrt{1 + \frac{t_1^2}{n+d_0-2k}} \frac{t_2}{\sqrt{n+d_0-2k+1}} \right) \right),$$

*where $t_1$ and $t_2$ are independent with $t_1 \sim t_{n+d_0-2k}$ and $t_2 \sim t_{n+d_0-2k+1}$.*

The results in Theorem 6 are very useful in analysing the behaviour of the investor's wealth during the whole investment period as well as at the final point $T$. It allows: (i) to calculate the probability with which the investor can become bankrupt at each time point during the whole investment horizon; (ii) to construct the prediction intervals for the wealths at each time point of the investment period; (iii) to determine risk measures, like Value-at-Risk (VaR) and conditional VaR (CVaR), of the investment strategy during the future reallocation; (iv) to specify a region where the final wealth belongs to with a high probability. We illustrate in Section 3 these results based on real data.

### 2.4. Numerical comparison to myopic strategies

In order to demonstrate the performance gain of the multi-period strategies we make a short comparison study with a myopic one. In the following we choose the global minimum variance (GMV) portfolio as benchmark myopic weights. The reason for that is a very unstable behaviour of the single period exponential portfolio weights from (1) where instead of the time-varying constants $C_t$ one needs to plug in $\gamma^{-1}$. Surprisingly, we observe a large probability of a negative wealth (default) for this strategy in many simulation scenarios resulting in a flat empirical cumulative distribution function (ECDF). Thus, the time-varying nature of the portfolio weights is crucial for an exponential portfolio, where the weights should be normalized by the wealth of the previous period. Other popular strategies could be a naive (equally weighted) portfolio or investing the whole wealth in a riskless asset. However, our simulations show no significant difference of these myopic portfolio strategies compared to the results obtained for the GMV portfolio. That is why, to this end, only the GMV portfolio is presented in the figures. We perform the following simulation procedure:

- Generate every element of $\boldsymbol{\mu}$ from the uniform distribution on the interval $(-0.01, 0.01)$ and generate $\boldsymbol{\Sigma}$ from the equicorrelated covariance model with a correlation matrix filled with correlations $\rho = 0.6$, whereas the variances are generated again uniformly from an interval $(0.01, 0.1)$. Fix $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.
- Generate asset returns from the multivariate normal distribution with the above $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.
- Construct two optimal portfolios: Bayes myopic (GMV portfolio) and Bayes multi-period portfolio with the diffuse prior. Find the final wealth in both cases.
- Repeat the procedure $B = 10^3$ times and find the ECDF of the final wealth for two strategies, which should be compared between each other for $k = 12$, $T = 13$, $\gamma = 0.8$ and $n \in \{52, 78, 104, 130\}$. Everywhere we assume that the return of a riskless asset is constant which is given by $r_f = 0.0002$.

The results presented in Fig. 1 are striking. The multiperiod strategy is strongly dominating the GMV portfolio in terms of the first order stochastic dominance, i.e., the probability to get a larger value of the final wealth is considerably larger than for the GMV portfolio. For example, for the sample size $n = 52$ the probability of getting wealth between 20 and 25 is roughly 0.2 by the Bayes multi-period strategy, whereas it is zero for the myopic GMV weights already for the wealth level greater than 2. These findings are independent of the number of assets and other values of the risk aversion coefficient $\gamma$, where the results are very similar. Interestingly, this effect is even more pronounced if the number of periods $T$ increases. This toy example motivates to use the Bayes multi-period weights instead of the single periodic ones in the case of dynamic decision making.

## 3. Empirical study

### 3.1. Data description

The data used in the empirical study consist of weekly returns on twelve stocks from the FTSE 100, namely Barclays, Glaxo Smith Kline, Standard Life, Marks and Spencer, Burberry Group plc, HSBC, LLoyds Banking, NEXT plc, Rolls-Royce
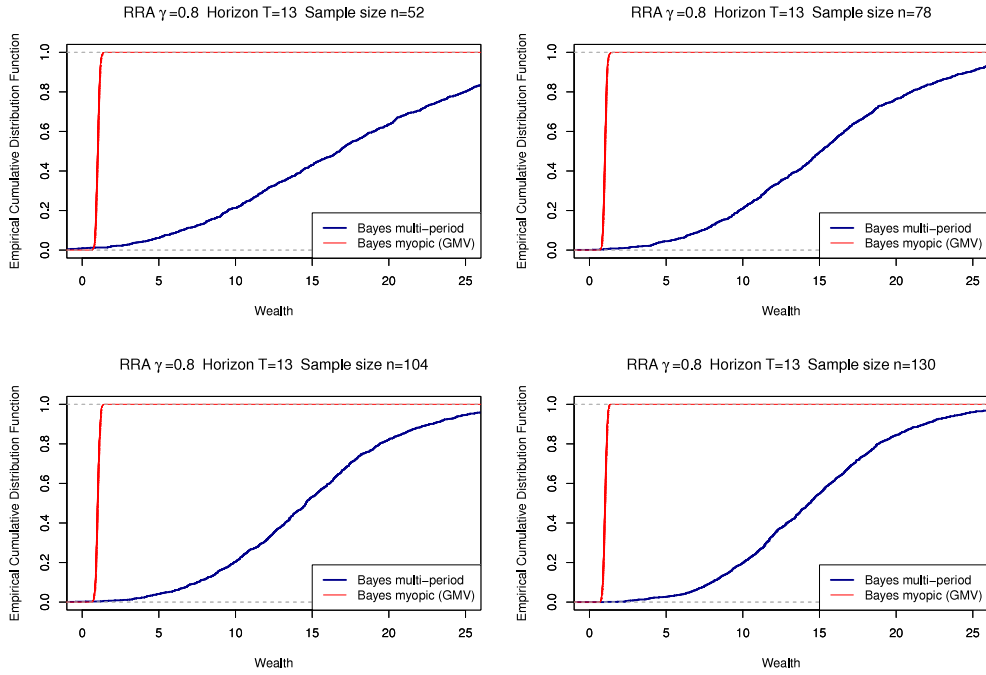
**Fig. 1.** Empirical cumulative distribution function of the final wealth for the Bayes multi-period strategy and the Bayes GMV portfolio. We set $\gamma = 0.8$, $k = 12$, $T = 13$ and $n \in \{52, 78, 104, 130\}$. The riskless rate of return $r_f$ is equal to 0.0002.

Holding, The Sage Group, Tesco plc and Unilever which represent a variety of branches with strong international activities. Since the parameters of the asset returns are not usually constant over a longer period of time, we disregard the use of monthly data which are closer to the normal distribution and choose weekly returns as a compromise between actuality and the assumption of conditional normality. As a risk-free rate we use the weekly returns on the three-months US treasury bill.

The portfolio weights are estimated using a rolling window estimation with different sample sizes of $n \in \{52, 78, 104, 130\}$ corresponding to one year up to two and a half years of weekly data in steps of six months. The portfolio runs from 6.6.2016 until 5.9.2016 ($T = 13$) covering a precarious market situation due to Great Britains referendum to leave the European Union on 23.06.2016. The gross returns of these assets are given in Fig. 2. Especially Barclays suffered a loss of nearly 10% in the week after the brexit decision but also suffered losses in the weeks prior to the brexit. HSBC announced that significant parts of her banking operations is moved from the City of London to different locations as a direct reaction to the referendum and it is rumoured that Lloyds seeks for a German banking licence as a consequence to the brexit. The returns of the Marks and Spencer share were not as affected by the brexit but the company reported that consumer confidence would be weakened in the days prior to the brexit. This also implies price uncertainty for domestic consumer products due to a decline of the pound losing almost a fifth of his value against the dollar after the brexit vote, which was emphasized for example by Tesco and Unilever. But Glaxo Smith Kline and Standard Life seem to be unaffected by the brexit decision, yielding even positive returns. Rolls Royce, after all, faced significant losses in the beginning of 2016 and is hit by the brexit vote severely, since they need to hedge a huge amount of British pounds against currency fluctuations because most of the contracts in aerospace are conducted in dollars.

### 3.2. Posterior distribution of the weights

Due to Theorem 2 it is possible to access the posterior distribution of the weights directly. The weights can be sampled using the following procedure:

1. Generate independently

   - $\eta \sim \chi_n^2$ under the diffuse prior or $\eta \sim \chi_{n+d_0-k}^2$ under the conjugate prior
   - $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$
   - $Q \sim \mathcal{F}(k, n-k)$ under the diffuse prior or $Q \sim \mathcal{F}(k, n+d_0-2k)$ under the conjugate prior
   - $\mathbf{Z} \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I}_k) \rightarrowtail \mathbf{u} = \mathbf{Z}/\sqrt{\mathbf{Z}'\mathbf{Z}}$

2. Compute the vector of portfolio weights by using the stochastic representation (7) for the diffuse prior or (8) for the conjugate prior.

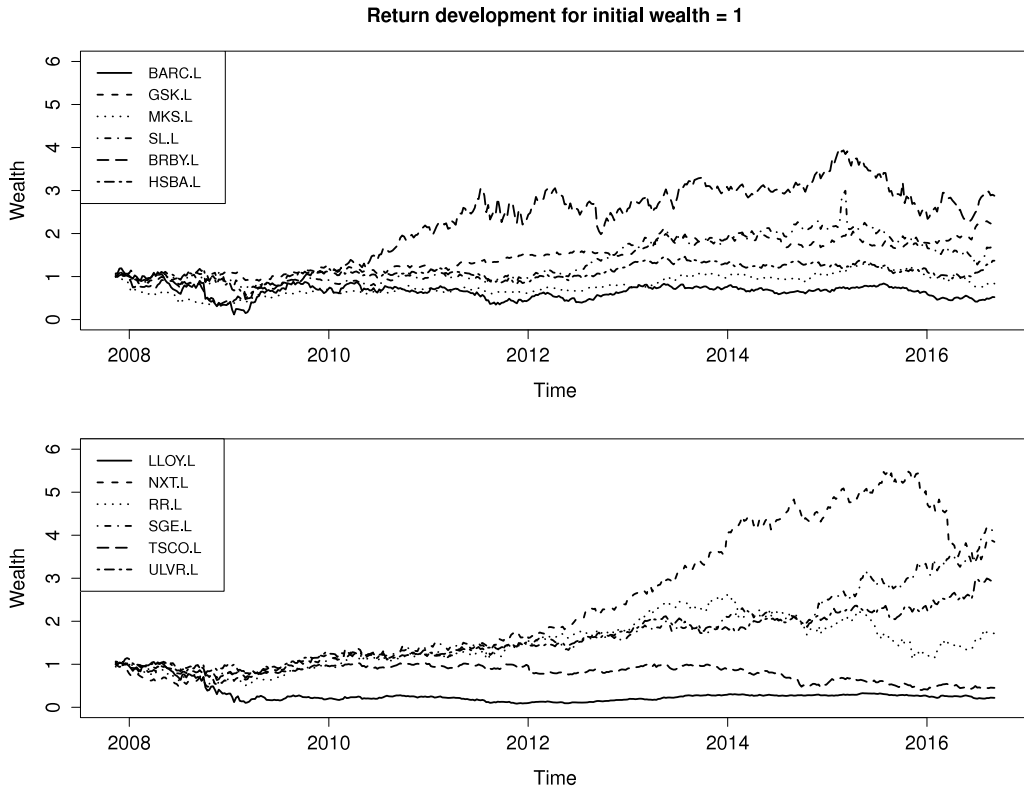3. Repeat steps (1) and (2) $B$ times.

**Fig. 2.** Development of the gross returns for the twelve assets considered in the portfolio.

The implementation of this simulation procedure leads to sequences of optimal portfolio weights of size B at each time point of the investment period, from which using their sample distribution we approximate the posterior distributions of the weights as well as their important quantiles from these distributions and the credible sets for portfolio weights. It is remarkable that all computations can easily be done by generating samples from the well known univariate distributions and high numerical precision could be achieved by choosing the corresponding value of B.

In Figs. 3 and 4, we analyse the finite-sample behaviour of the results presented in Theorem 5. Namely, we investigate the speed of convergence of the posterior distribution of the optimal portfolio weights to the corresponding asymptotic distribution which is a normal distribution according to Theorem 5 for both priors. The choice of the hyperparameters $\mathbf{m}_0$ and $\mathbf{S}_0$ in the case of the conjugate prior is of particular interest. According to the Bayesian paradigm, $\mathbf{m}_0$ and $\mathbf{S}_0$ represent the correct belief of the decision maker. In practice, however, there are several data driven methods how to replace $\mathbf{m}_0$ and $\mathbf{S}_0$ by data-dependent values $\hat{\mathbf{m}}_0$ and $\hat{\mathbf{S}}_0$. We make use of the empirical Bayes approach (see Appendix B for the derivation of the formulas) which is applied to the weekly data of the returns on the corresponding assets directly from the time period before the empirical counterparts of the portfolio weights are estimated, always with the same time window. Namely, they are given by

$$\hat{\mathbf{m}}_0 = \bar{\mathbf{x}}_{n-t}, \quad \hat{\mathbf{S}}_0 = \frac{(d_0 - k - 1)(n-1)}{n}\mathbf{S}_{n-t}$$

with the derivation moved to Appendix B. The prior parameters for $t > 1$ are estimated using a rolling window starting in the corresponding period. We set $d_0$ equal to the number of observations in the pres-sample period, i.e., $d_0 = n$.

We set $B = 10^5$ for draws from the stochastic representations of Theorem 2 and compare the standardized weight of Glaxo Smith Kline (GSK) calculated for the period $T - 1$ in the case of several sample sizes $n \in \{52, 78, 104, 130\}$. The corresponding histograms are given in Fig. 3 for the diffuse prior and in Fig. 4 for the conjugate prior. In both figures we also present the p-values of the Shapiro–Wilk test, indicating if the standardized weights follow a standard normal distribution. This hypothesis is rejected for $n = 52$ and $n = 78$ in the case of the diffuse prior for a common significance level of 5% but it cannot be rejected at this level for larger sample sizes. Stronger results are obtained in the case of the conjugate prior, where the null hypothesis cannot be rejected at 5% level for all considered sample sizes. We therefore conclude that the approximate distribution of Theorem 5 works reasonably well.
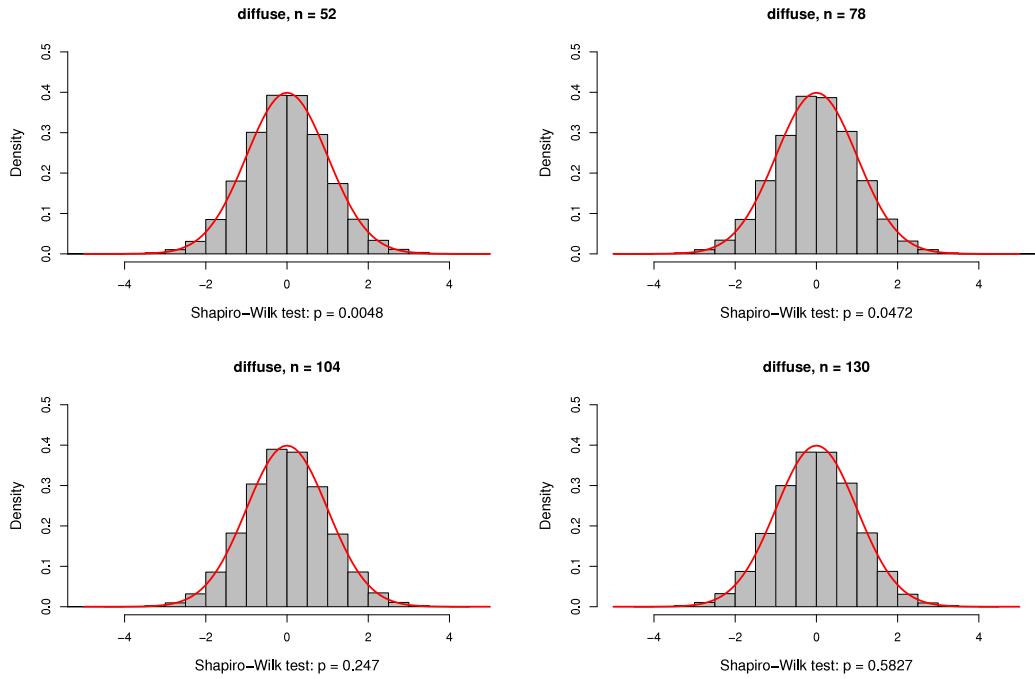
**Fig. 3.** Histograms of the standardized Glaxo Smith Kline (GSK) weight for the diffuse prior. The hypothesis that the weight is normally distributed cannot be rejected for common significance levels when the sample size is larger than $n = 100$.



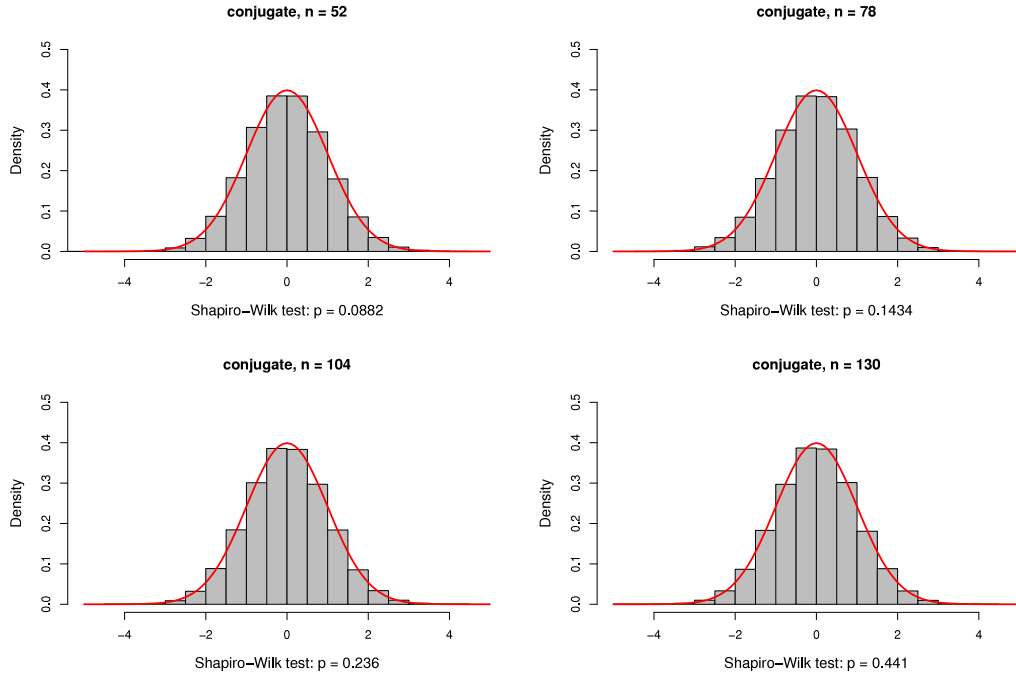**Fig. 4.** Histograms of the standardized Glaxo Smith Kline (GSK) weight for the conjugate prior. The hypothesis that the weight is normally distributed cannot be rejected for common significance levels in the case of all considered sample sizes.
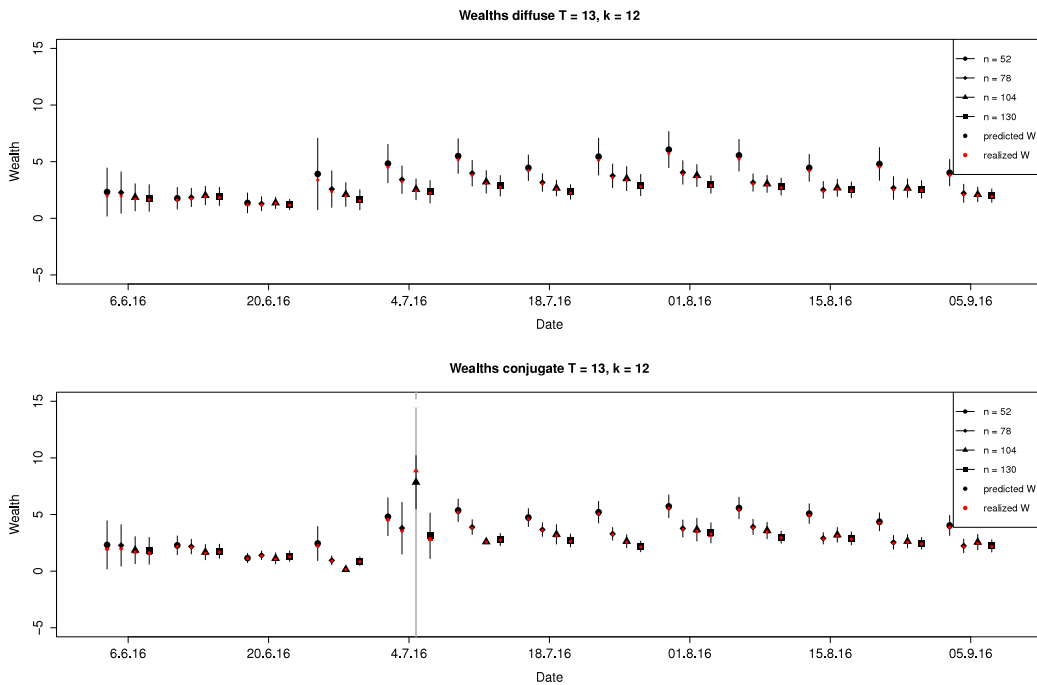
**Fig. 5.** Wealth development and 95% credible intervals for the diffuse prior (above) and for the conjugate prior (below). The wealth for smaller *n* is almost always higher compared to a portfolio estimated with larger *n*, while the credible intervals are much narrower for larger *n*.

### 3.3. Wealth development and credibility intervals

Since the main purpose of investing is making money, investors are therefore interested in how much money they made during an investment period. We focus again on the same investment period covering the brexit referendum as in the previous subsection.

During the lifetime of the portfolio, no bankruptcy occurred. But more importantly, the stochastic representation for the posterior predictive distribution given in Theorem 6 can be used to calculate credible intervals for the wealth. By generating $B = 10^5$ draws from Theorem 6 and calculating the 95% credible intervals, we generate upper and lower bounds for the wealth in the specific period. These intervals together with the predicted and realized wealths are shown in Fig. 5. We observe a difference in the width of the intervals for lower and larger sample sizes which was expected. The credible intervals are considerably smaller for $n \in \{104, 130\}$ compared to smaller *n*. Note that the sample size has to be sufficiently large in relation to the number of assets. Otherwise, the credible intervals are inflated due to massive estimation uncertainty known as the curse of dimensionality.

It might happen that both the diffuse and the conjugate priors do not perform well when the sample size increases. The reason for the diffuse prior is that the empirical counterparts might not describe the portfolio running period well, indicating a trade-off between the actuality and stability of the parameters. This problem is amplified for the conjugate prior since the prior parameters are determined using even more distant data. While the data-driven approach to the conjugate prior is somewhat realistic, it is not completely in line with the Bayesian paradigm. When the expectations and therefore the choice of hyperparameters are closer to the return behaviour after the B brexit, the results could be improved. Although this is consistent with the Bayesian paradigm, such an approach is of course not entirely practical but not impractical: using appropriate forecasting methods, other data driven methods can be applicable as long as they yield a reliable point estimate. This subjective approach emphasizes the possibility as well as the necessity to resemble realistic future market behaviour in the prior parameterization and it is left for future research.

### 3.4. Default probability

Due to the accessibility of the posterior predictive distribution, we can also calculate the default probability of our portfolio at each time point, defined as the event that our wealth becomes negative at this point in time. The predictive probability of default can easily be determined by calculating the amount of defaults in relation to all draws, in this case $B = 10^5$. The development of the defaults is given in Fig. 6. Again, we find a pattern resembling the credible intervals of the posterior predictive distribution illustrated in the previous section with no surprises.
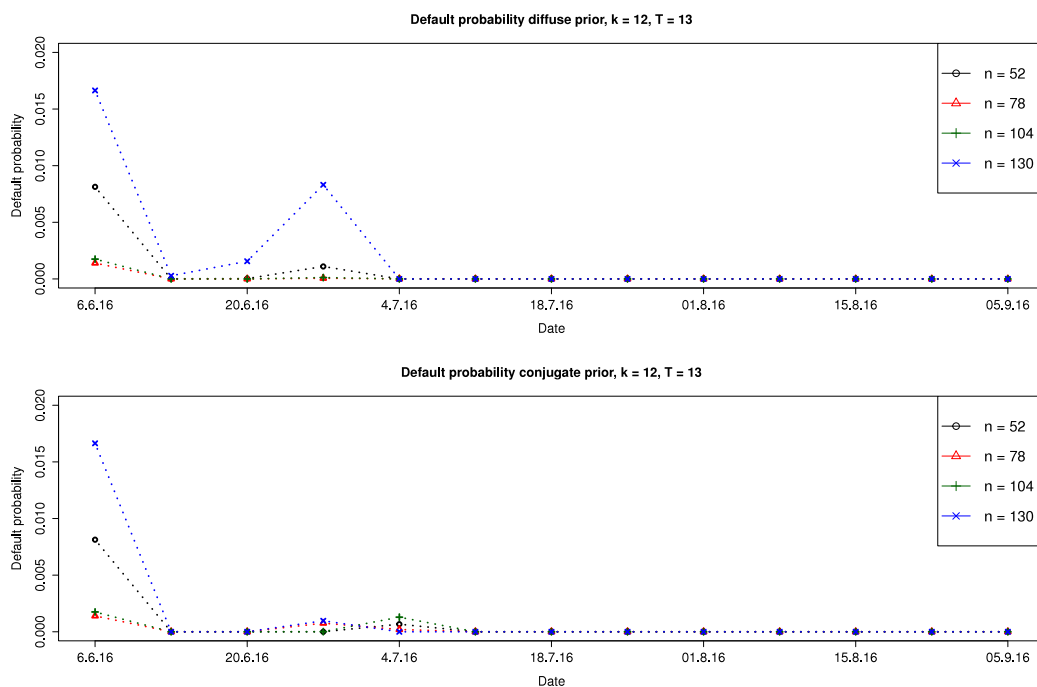
**Fig. 6.** Default probabilities for the diffuse prior (above) and for the conjugate prior (below).

Starting with the diffuse prior, we observe a slightly increased default probability on 27.6.2016, the week after the brexit referendum. With the conjugate prior, this default probability is lower in the same week. Again, the peak for $n = 130$ of the diffuse prior again resembles the trade-off between parameter stability and actuality, resulting here in a slightly increased default probability. The default probability for the conjugate prior is slightly increased in the following week compared to the diffuse prior, presumably due to parameters relying on a wider estimation window.

## 4. Summary

In this article we consider the estimation of the multi-period portfolio for an exponential utility function in a Bayesian setting. Since the portfolio weights are given as the product of two multivariate/matrix-variate random quantities, accessing the distribution of the weights is a challenging task. By choosing the non-informative and the conjugate prior, the posterior distributions of the weights have pleasing properties since the conditional distribution of the precision matrix for a given return vector follows an inverted Wishart distribution. With this insight we could use this well understood distribution (see e.g., [42]) to derive stochastic representations for the weights which is a direct access to the posterior distribution. Furthermore, these representations also provide us with Bayesian estimates for the optimal portfolio weights together with their covariance matrix. In addition to this, we derive the posterior predictive distribution for the wealth which makes it possible to calculate the quantiles of the portfolio wealth at each time point of the investment period and it is therefore highly relevant for risk purposes. The method is then applied to real data from the FTSE 100 covering the period of the brexit referendum. With these data we determine the posterior distribution of the weights, the predictive wealths in each period, the lower wealth quantiles as well as the default probability in every time period.

It turns out that the use of stochastic representations to generate the posterior distribution numerically is computationally highly efficient: the representations rely on samples from well known distributions and no MCMC methods are needed. In the empirical part of Section 3 it was demonstrated that these methods work well and are easy to implement. We have to emphasize several points: while the non-informative prior will yield results which coincide with the common frequentistic case and is as easily to apply as the classical case, the conjugate or informative prior is said to involve a potentially large degree of subjectivity — sometimes implying that the frequentistic approach or the non-informative prior would be objective. But we have to choose the sample size in all of these cases which is naturally a subjective choice with a huge effect on the performance of the portfolio as we demonstrate in Section 3. This trade-off between parameter actuality and parameter stability has to be faced by the practitioner. One advantage of the conjugate prior is of course that we can incorporate our beliefs regarding the future behaviour of the asset returns in our model which is not possible neither in the frequentistic nor in the non-informative case. This is clearly at the core of every investment decision and reflects natural decision making. Nevertheless, the hyperparameters have to be chosen carefully and a rigorous sensitivity analysis is left for future research.

There are still other open research questions regarding the multi-period portfolio choice with exponential utility function which are left for future research. The present approach can be extended to the high-dimensional case (see, e.g., [4,9]) and to the case with predictable variables as discussed in [15] under the assumption that the parameters of the asset return distribution are known. This, however, is much more difficult due to the more complicated structure of the optimal portfolio weights and the dependence structure of the asset returns. Furthermore, the multi-period optimal portfolios obtained by using other utility functions can be estimated following the approach suggested in the paper.

## Acknowledgments

## Appendix A. Proofs of the theorems

In this part of the paper we present the proofs of the theoretical results. First, we note that the derived posterior distributions under the diffuse prior and under the conjugate prior in Proposition 2 have a similar structure. For that reason, we formulate and prove some lemmas from which the results in both cases of the diffuse prior and the conjugate prior follow.

**Lemma 1.** *Let*

$$\boldsymbol{\Omega}|\boldsymbol{\nu}, \mathbf{y} \sim \mathcal{IW}_k(k_y, \mathbf{S}_y^*(\boldsymbol{\nu})), \quad \boldsymbol{\nu}|\mathbf{y} \sim t_k\left(d_y, \mathbf{m}_y, \mathbf{S}_y/d_y\right),$$

*where* $\mathbf{S}_y^*(\boldsymbol{\nu}) = \nu_y(\mathbf{S}_y + (\boldsymbol{\nu} - \mathbf{m}_y)(\boldsymbol{\nu} - \mathbf{m}_y)^\top)$ *and let* $\mathbf{M}$ *be a* $p \times k$-*dimensional matrix of constants. Then the stochastic representation of* $\mathbf{M}\boldsymbol{\Omega}^{-1}(\boldsymbol{\nu} - \mathbf{a})$ *is given by*

$$\mathbf{M}\boldsymbol{\Omega}^{-1}(\boldsymbol{\nu} - \mathbf{a}) \stackrel{d}{=} \eta \mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}(\boldsymbol{\nu} - \mathbf{a})$$
$$+ \sqrt{\eta}\left((\boldsymbol{\nu} - \mathbf{a})^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}(\boldsymbol{\nu} - \mathbf{a}) \cdot \mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{M}^\top - \mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}(\boldsymbol{\nu} - \mathbf{a})(\boldsymbol{\nu} - \mathbf{a})^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{M}^\top\right)^{1/2} \mathbf{z}_0,$$

*where* $\eta \sim \chi^2_{k_y-k-1}$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$, *and* $\boldsymbol{\nu}|\mathbf{y} \sim t_k\left(d_y, \mathbf{m}_y, \mathbf{S}_y/d_y\right)$. *Moreover,* $\eta$, $\mathbf{z}_0$ *and* $\boldsymbol{\nu}$ *are mutually independent.*

**Proof of Lemma 1.** Since $\boldsymbol{\Omega}^* \stackrel{d}{=} \boldsymbol{\Omega}|\boldsymbol{\nu} = \boldsymbol{\nu}^*, \mathbf{y} \sim \mathcal{IW}_k(k_y, \mathbf{S}_y^*(\boldsymbol{\nu}^*))$ and, consequently, $\boldsymbol{\Omega}^{*-1} \sim \mathcal{W}_k(k_y - k - 1, \mathbf{S}_y^*(\boldsymbol{\nu}^*)^{-1})$ (see, Theorem 3.4.1 in [30]), it holds that (see, e.g., Theorem 3.2.5 in [42])

$$\boldsymbol{\Xi}^* = \widetilde{\mathbf{M}}\boldsymbol{\Omega}^{*-1}\widetilde{\mathbf{M}}^\top \sim \mathcal{W}_k(k_y - k - 1, \mathbf{V}^*),$$

with $\widetilde{\mathbf{M}} = (\mathbf{M}^\top, \boldsymbol{\nu}^* - \mathbf{a})^\top$ and $\mathbf{V}^* = \tilde{\mathbf{M}}\mathbf{S}_y^*(\boldsymbol{\nu}^*)^{-1}\tilde{\mathbf{M}}^\top$. Next, we partition $\boldsymbol{\Xi}^*$ and $\mathbf{V}^*$ in the following way

$$\boldsymbol{\Xi}^* = \begin{pmatrix} \boldsymbol{\Xi}_{11}^* & \boldsymbol{\Xi}_{12}^* \\ \boldsymbol{\Xi}_{21}^* & \boldsymbol{\Xi}_{22}^* \end{pmatrix} = \begin{pmatrix} \mathbf{M}\boldsymbol{\Omega}^{*-1}\mathbf{M}^\top & (\boldsymbol{\nu}^* - \mathbf{a})^\top \boldsymbol{\Omega}^{*-1}\mathbf{M}^\top \\ \mathbf{M}\boldsymbol{\Omega}^{*-1}(\boldsymbol{\nu}^* - \mathbf{a}) & (\boldsymbol{\nu}^* - \mathbf{a})^\top \boldsymbol{\Omega}^{*-1}(\boldsymbol{\nu}^* - \mathbf{a}) \end{pmatrix}$$

and

$$\mathbf{V}^* = \begin{pmatrix} \mathbf{V}_{11}^* & \mathbf{V}_{12}^* \\ \mathbf{V}_{21}^* & \mathbf{V}_{22}^* \end{pmatrix} = \begin{pmatrix} \mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu}^*)^{-1}\mathbf{M}^\top & (\boldsymbol{\nu}^* - \mathbf{a})^\top \mathbf{S}_y^*(\boldsymbol{\nu}^*)^{-1}\mathbf{M}^\top \\ \mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu}^*)^{-1}(\boldsymbol{\nu}^* - \mathbf{a}) & (\boldsymbol{\nu}^* - \mathbf{a})^\top \mathbf{S}_y^*(\boldsymbol{\nu}^*)^{-1}(\boldsymbol{\nu}^* - \mathbf{a}) \end{pmatrix}.$$

The application of Theorem 3.2.10 in [42] yields

$$\boldsymbol{\Xi}_{12}^*|\boldsymbol{\Xi}_{22}^* \sim \mathcal{N}_p(\mathbf{V}_{12}V_{22}^{-1}\boldsymbol{\Xi}_{22}^*, \mathbf{V}_{11\cdot2}\boldsymbol{\Xi}_{22}^*) \text{ with } \mathbf{V}_{11\cdot2} = \mathbf{V}_{11} - \frac{\mathbf{V}_{12}\mathbf{V}_{21}}{V_{22}}.$$

Defining $\eta = \boldsymbol{\Xi}_{22}^*/V_{22}$ and using Theorem 3.2.8 of [42] we get that $\eta \sim \chi^2_{k_y-k-1}$. Since the $\chi^2_{k_y-k-1}$-distribution is independent of $\boldsymbol{\nu} = \boldsymbol{\nu}^*$ and $\mathbf{y}$ (the distribution of $\boldsymbol{\Xi}_{22}^*$ depends on $\mathbf{y}$ by definition of $\boldsymbol{\Xi}^*$), it is also the unconditional distribution of $\eta$ as well as $\eta$ is independent of both $\boldsymbol{\nu}$ and $\mathbf{y}$. Thus, the stochastic representation of $\mathbf{M}\boldsymbol{\Omega}^{-1}(\boldsymbol{\nu} - \mathbf{a})$ is given by

$$\mathbf{M}\boldsymbol{\Omega}^{-1}(\boldsymbol{\nu} - \mathbf{a}) \stackrel{d}{=} \eta \mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}(\boldsymbol{\nu} - \mathbf{a}) + \sqrt{\eta}\Bigg((\boldsymbol{\nu} - \mathbf{a})^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}(\boldsymbol{\nu} - \mathbf{a}) \cdot \mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{M}^\top$$
$$- \mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}(\boldsymbol{\nu} - \mathbf{a})(\boldsymbol{\nu} - \mathbf{a})^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{M}^\top\Bigg)^{1/2} \mathbf{z}_0,$$

where $\eta \sim \chi^2_{k_y-k-1}$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$, and $\boldsymbol{\nu}|\mathbf{y} \sim t_k\left(d_y, \mathbf{m}_y, \mathbf{S}_y/d_y\right)$; moreover, $\eta$, $\mathbf{z}_0$ and $\boldsymbol{\nu}$ are mutually independent. This completes the proof of the lemma. □

**Proof of Theorem 1.** The results of Theorem 1 follow from Lemma 1 with $\mathbf{M} = C_t\mathbf{L}$, $\boldsymbol{\Sigma} = \boldsymbol{\Omega}$, $\boldsymbol{v} = \boldsymbol{\mu}$, $\mathbf{a} = r_{f,t+1}\mathbf{1}$ and

  (a) $k_y = n + k + 1$, $d_y = n - k$, $v_y = n$, $\mathbf{m}_y = \bar{\mathbf{x}}_{t,d}$, $\mathbf{S}_y = \mathbf{S}_{t,d}/n$, and $\mathbf{S}_y^*(\boldsymbol{v}) = \mathbf{S}_{t,d}^*(\boldsymbol{\mu})$ in the case of the diffuse prior;
  (b) $k_y = n + d_0 + 1$, $d_y = n + d_0 - 2k$, $v_y = n + r_0$, $\mathbf{m}_y = \bar{\mathbf{x}}_{t,c}$, $\mathbf{S}_y = \mathbf{S}_{t,c}/(n+r_0)$, and $\mathbf{S}_y^*(\boldsymbol{v}) = \mathbf{S}_{t,c}^*(\boldsymbol{\mu})$ in the case of the conjugate prior.    □

**Lemma 2.** *Under the conditions of Lemma* 1, *we get the following stochastic representation of* $\mathbf{M}\boldsymbol{\Omega}^{-1}(\boldsymbol{v} - \mathbf{a})$ *expressed as*

$$\mathbf{M}\boldsymbol{\Omega}^{-1}(\boldsymbol{v} - \mathbf{a}) \stackrel{d}{=} v_y^{-1}\eta\mathbf{M}\boldsymbol{\zeta} + v_y^{-1}\sqrt{\eta}\left(\epsilon\mathbf{M}\boldsymbol{\Upsilon}\mathbf{M}^\top - \mathbf{M}\boldsymbol{\zeta}\boldsymbol{\zeta}^\top\mathbf{M}^\top\right)^{1/2}\mathbf{z}_0,$$

*with*

$$\epsilon = \epsilon(Q, \mathbf{u}) = (\mathbf{m}_y - \mathbf{a})^\top\mathbf{S}_y^{-1}(\mathbf{m}_y - \mathbf{a}) + 2\frac{\sqrt{kQ/d_y}}{1 + kQ/d_y}(\mathbf{m}_y - \mathbf{a})^\top\mathbf{S}_y^{-1/2}\mathbf{u} + \frac{kQ/d_y}{1 + kQ/d_y}\left(1 - \left((\mathbf{m}_y - \mathbf{a})^\top\mathbf{S}_y^{-1/2}\mathbf{u}\right)^2\right),$$

$$\boldsymbol{\zeta} = \boldsymbol{\zeta}(Q, \mathbf{u}) = \mathbf{S}_y^{-1}(\mathbf{m}_y - \mathbf{a}) + \frac{\sqrt{kQ/d_y}}{1 + kQ/d_y}\mathbf{S}_y^{-1/2}\mathbf{u} - \frac{kQ/d_y}{1 + kQ/d_y}\mathbf{S}_y^{-1/2}\mathbf{u}\mathbf{u}^\top\mathbf{S}_y^{-1/2}(\mathbf{m}_y - \mathbf{a}),$$

$$\boldsymbol{\Upsilon} = \boldsymbol{\Upsilon}(Q, \mathbf{u}) = \mathbf{S}_y^{-1} - \frac{kQ/d_y}{1 + kQ/d_y}\mathbf{S}_y^{-1/2}\mathbf{u}\mathbf{u}^\top\mathbf{S}_y^{-1/2},$$

*where* $\eta \sim \chi_{k_y-k-1}^2$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$, $Q \sim \mathcal{F}(k, d_y)$, *and* $\mathbf{u}$ *uniformly distributed on the unit sphere in* $\mathbb{R}^k$. *Moreover,* $\eta$, $\mathbf{z}_0$, $Q$, *and* $\mathbf{u}$ *are mutually independent.*

**Proof of Lemma 2.** The application of the Sherman–Morrison formula (see, e.g., [40, p. 124]) yields

$$(\mathbf{S}_y + (\boldsymbol{v} - \mathbf{m}_y)(\boldsymbol{v} - \mathbf{m}_y)^\top)^{-1} = \mathbf{S}_y^{-1} - \frac{\mathbf{S}_y^{-1}(\boldsymbol{v} - \mathbf{m}_y)(\boldsymbol{v} - \mathbf{m}_y)^\top\mathbf{S}_y^{-1}}{1 + (\boldsymbol{v} - \mathbf{m}_y)^\top\mathbf{S}_y^{-1}(\boldsymbol{v} - \mathbf{m}_y)}. \tag{10}$$

Let

$$\mathbf{u} = \frac{\mathbf{S}_y^{-1/2}(\boldsymbol{v} - \mathbf{m}_y)}{\sqrt{(\boldsymbol{v} - \mathbf{m}_y)^\top\mathbf{S}_y^{-1}(\boldsymbol{v} - \mathbf{m}_y)}} \quad \text{and} \quad Q = d_y(\boldsymbol{v} - \mathbf{m}_y)^\top\mathbf{S}_y^{-1}(\boldsymbol{v} - \mathbf{m}_y)/k. \tag{11}$$

Since $\boldsymbol{v}|\mathbf{y} \sim t_k(d_y, \mathbf{m}_y, \mathbf{S}_y/d_y)$ and that the multivariate $t$-distribution belongs to the class of the elliptically contoured distributions, we obtain that $\mathbf{u}$ and $Q$ are independent, and $\mathbf{u}$ is uniformly distributed on the unit sphere in $\mathbb{R}^k$ (see Theorem 2.15 of [31]). Moreover, from the properties of the multivariate $t$-distribution (see p. 19 of [34]), we get that $Q \sim \mathcal{F}(k, d_y)$, i.e., $Q$ has an $\mathcal{F}$-distribution with $k$ and $d_y$ degrees of freedom.

Hence, the application of (10) and (11) leads to

$$(\mathbf{S}_y + (\boldsymbol{v} - \mathbf{m}_y)(\boldsymbol{v} - \mathbf{m}_y)^\top)^{-1} = \mathbf{S}_y^{-1} - \frac{kQ/d_y}{1 + kQ/d_y}\mathbf{S}_y^{-1/2}\mathbf{u}\mathbf{u}^\top\mathbf{S}_y^{-1/2},$$

$$(\mathbf{S}_y + (\boldsymbol{v} - \mathbf{m}_y)(\boldsymbol{v} - \mathbf{m}_y)^\top)^{-1}(\boldsymbol{v} - \mathbf{a}) = \mathbf{S}_y^{-1}(\boldsymbol{v} - \mathbf{a}) - \frac{\mathbf{S}_y^{-1}(\boldsymbol{v} - \mathbf{m}_y)(\boldsymbol{v} - \mathbf{m}_y)^\top\mathbf{S}_y^{-1}(\boldsymbol{v} - \mathbf{m}_y + \mathbf{m}_y - \mathbf{a})}{1 + (\boldsymbol{v} - \mathbf{m}_y)^\top\mathbf{S}_y^{-1}(\boldsymbol{v} - \mathbf{m}_y)}$$

$$= \mathbf{S}_y^{-1}(\mathbf{m}_y - \mathbf{a}) + \frac{\mathbf{S}_y^{-1}(\boldsymbol{v} - \mathbf{m}_y)}{1 + (\boldsymbol{v} - \mathbf{m}_y)^\top\mathbf{S}_y^{-1}(\boldsymbol{v} - \mathbf{m}_y)}$$

$$- \frac{\mathbf{S}_y^{-1}(\boldsymbol{v} - \mathbf{m}_y)(\boldsymbol{v} - \mathbf{m}_y)^\top\mathbf{S}_y^{-1}(\mathbf{m}_y - \mathbf{a})}{1 + (\boldsymbol{v} - \mathbf{m}_y)^\top\mathbf{S}_y^{-1}(\boldsymbol{v} - \mathbf{m}_y)}$$

$$= \mathbf{S}_y^{-1}(\mathbf{m}_y - \mathbf{a}) + \frac{\sqrt{kQ/d_y}}{1 + kQ/d_y}\mathbf{S}_y^{-1/2}\mathbf{u} - \frac{kQ/d_y}{1 + kQ/d_y}\mathbf{S}_y^{-1/2}\mathbf{u}\mathbf{u}^\top\mathbf{S}_y^{-1/2}(\mathbf{m}_y - \mathbf{a}),$$

and

$$(\boldsymbol{v} - \mathbf{a})^\top(\mathbf{S}_y + (\boldsymbol{v} - \mathbf{m}_y)(\boldsymbol{v} - \mathbf{m}_y)^\top)^{-1}(\boldsymbol{v} - \mathbf{a}) = (\mathbf{m}_y - \mathbf{a})^\top\mathbf{S}_y^{-1}(\mathbf{m}_y - \mathbf{a}) + 2\frac{(\mathbf{m}_y - \mathbf{a})^\top\mathbf{S}_y^{-1/2}\mathbf{u}\sqrt{kQ/d_y}}{1 + kQ/d_y}$$

$$+ \frac{kQ/d_y}{1 + kQ/d_y} - \frac{kQ/d_y}{1 + kQ/d_y}\left((\mathbf{m}_y - \mathbf{a})^\top\mathbf{S}_y^{-1/2}\mathbf{u}\right)^2.$$

Putting the above results together we obtain the statement of the lemma.    □

**Proof of Theorem 2.** The results of Theorem 2 are obtained by using Lemma 2 with $\mathbf{M} = C_t\mathbf{L}$, $\boldsymbol{\Sigma} = \boldsymbol{\Omega}$, $\boldsymbol{v} = \boldsymbol{\mu}$, $\mathbf{a} = r_{f,t+1}\mathbf{1}$ and

(a) $k_y = n + k + 1$, $d_y = n - k$, $v_y = n$, $\mathbf{m}_y - \mathbf{a} = \bar{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1}$, $\mathbf{S}_y = \mathbf{S}_{t,d}/n$, and $\mathbf{S}_y^*(\boldsymbol{v}) = \mathbf{S}_{t,d}^*(\boldsymbol{\mu})$ for the diffuse prior;

(b) $k_y = n + d_0 + 1$, $d_y = n + d_0 - 2k$, $v_y = n + r_0$, $\mathbf{m}_y - \mathbf{a} = \bar{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1}$, $\mathbf{S}_y = \mathbf{S}_{t,c}/(n + r_0)$, and $\mathbf{S}_y^*(\boldsymbol{v}) = \mathbf{S}_{t,c}^*(\boldsymbol{\mu})$ in the case of the conjugate prior.  □

**Proof of Theorem 3.** The proof of the theorem is based on the stochastic representations obtained in Theorem 2. Let $\boldsymbol{\ell}$ be an arbitrary $k$-dimensional vector of constants.

(a) Using that $\eta$, $\mathbf{z}_0$ $Q$, and $\mathbf{u}$ are independent and that $\mathbb{E}(\mathbf{z}_0) = \mathbf{0}$, in the case of the diffuse prior we get

$$\mathbb{E}(\boldsymbol{\ell}^\top \mathbf{w}_t | \mathbf{x}_{t,n}) = C_t \mathbb{E}(\eta)\boldsymbol{\ell}^\top \mathbb{E}(\boldsymbol{\zeta}_d)$$

with $\mathbb{E}(\eta) = n$ and

$$\begin{aligned}
\mathbb{E}(\boldsymbol{\zeta}_d | \mathbf{x}_{t,n}) &= \mathbf{S}_{t,d}^{-1}(\bar{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1}) + \frac{1}{\sqrt{n}}\mathbb{E}\left(\frac{\sqrt{kQ/(n-k)}}{1 + kQ/(n-k)}\mathbf{S}_{t,d}^{-1/2}\right)\mathbb{E}(\mathbf{u}) \\
&\quad - \mathbb{E}\left(\frac{kQ/(n-k)}{1 + kQ/(n-k)}\mathbf{S}_{t,d}^{-1/2}\right)\mathbb{E}(\mathbf{u}\mathbf{u}^\top)\mathbf{S}_{t,d}^{-1/2}(\bar{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1}) \\
&= \mathbf{S}_{t,d}^{-1}(\bar{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1}) - \frac{k}{n}\frac{1}{k}\mathbf{S}_{t,d}^{-1}(\bar{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1}),
\end{aligned}$$

where we use that $E(\mathbf{u}) = \mathbf{0}$ and $E(\mathbf{u}\mathbf{u}^T) = \frac{1}{k}\mathbf{I_k}$ (see, e.g., [31]) as well as the fact that if $Q \sim \mathcal{F}(k, n - k)$, then $\frac{k}{n-k}Q/\left(1 + \frac{k}{n-k}Q\right) \sim Beta\left(\frac{k}{2}, \frac{n-k}{2}\right)$. Hence,

$$E\left(\frac{\frac{k}{(n-k)}Q}{1 + \frac{k}{(n-k)}Q}\right) = \frac{k}{n}$$

and, consequently, since $\boldsymbol{\ell}$ was an arbitrary vector, we get

$$\mathbb{E}(\mathbf{w}_t | \mathbf{x}_{t,n}) = C_t(n - 1)\mathbf{S}_{t,d}^{-1}(\bar{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1}).$$

(b) Similar computations as in part (a) lead to

$$\mathbb{E}(\mathbf{w}_t | \mathbf{x}_{t,n}) = C_t(n + d_0 - k - 1)\mathbf{S}_{t,c}^{-1}(\bar{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1})$$

under the conjugate prior.  □

**Lemma 3.** *Under the assumption of Lemma 2 with* $\mathbf{M} = \mathbf{b}^\top : 1 \times k$*, we get that*

$$\begin{aligned}
v_y^2\mathbb{E}((\mathbf{b}^\top \boldsymbol{\Omega}^{-1}(\boldsymbol{v} - \mathbf{a}))^2 | \mathbf{y}) &= (k_y - k - 1)(k_y - k)\left[\left(1 - \frac{2}{k + d_y} + \frac{2}{(k + d_y)(k + d_y + 2)}\right)c_{12}^2\right. \\
&\quad + \left.\left(\frac{d_y}{(k + d_y)(k + d_y + 2)} + \frac{1}{(k + d_y)(k + d_y + 2)}c_2\right)c_1\right] \\
&\quad + (k_y - k - 1)\left[\left(\frac{k - 1}{k + d_y} + \left(1 - \frac{1}{k} - \frac{1}{k + d_y} + \frac{1}{(k + d_y)(k + d_y + 2)}\right)c_2\right)c_1 + \frac{2}{(k + d_y)(k + d_y + 2)}c_{12}^2\right],
\end{aligned}$$

*where* $c_1 = \mathbf{b}^\top \mathbf{S}_y^{-1}\mathbf{b}$*,* $c_2 = (\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1}(\mathbf{m}_y - \mathbf{a})$*, and* $c_{12} = \mathbf{b}^\top \mathbf{S}_y^{-1}(\mathbf{m}_y - \mathbf{a})$*.*

**Proof of Lemma 3.** The proof of the lemma is based on the stochastic representations from Lemma 2. Since $\eta$, $\mathbf{z}_0$, $Q$, and $\mathbf{u}$ are independent as well as $\mathbb{E}(\mathbf{z}_0) = \mathbf{0}$ and $\mathbb{E}(\mathbf{z}_0\mathbf{z}_0^\top) = \mathbf{I}_p$, we obtain

$$\begin{aligned}
v_y^2\mathbb{E}((\mathbf{b}^\top \boldsymbol{\Omega}^{-1}(\boldsymbol{v} - \mathbf{a}))^2 | \mathbf{y}) &= \mathbb{E}(\eta^2)\mathbb{E}((\mathbf{b}^\top \boldsymbol{\zeta})^2 | \mathbf{y}) + \mathbb{E}(\eta)\left(\mathbb{E}(\epsilon\mathbf{b}^\top \boldsymbol{\Upsilon}\mathbf{b} | \mathbf{y}) - \mathbb{E}((\mathbf{b}^\top \boldsymbol{\zeta})^2 | \mathbf{y})\right) \\
&= (k_y - k - 1)(k_y - k)\mathbb{E}((\mathbf{b}^\top \boldsymbol{\zeta})^2 | \mathbf{y}) + (k_y - k - 1)\mathbb{E}(\epsilon\mathbf{b}^\top \boldsymbol{\Upsilon}\mathbf{b} | \mathbf{y})
\end{aligned}$$

with $\mathbb{E}(\eta) = k_y - k - 1$ and $\mathbb{E}(\eta^2) = (k_y - k - 1)(k_y - k + 1)$.

The application of $E(\mathbf{u}\mathbf{u}^T) = \frac{1}{k}\mathbf{I}_k$ and the fact that all odd mixed moments of $\mathbf{u}$ are zero yield

$$\begin{aligned}
\mathbb{E}((\mathbf{b}^\top \boldsymbol{\zeta})^2 | \mathbf{y}) &= (\mathbf{b}^\top \mathbf{S}_y^{-1}(\mathbf{m}_y - \mathbf{a}))^2 + \frac{1}{k}\mathbb{E}\left(\frac{kQ/d_y}{(1 + kQ/d_y)^2}\right)\mathbf{b}^\top \mathbf{S}_y^{-1}\mathbf{b} - \frac{2}{k}\mathbb{E}\left(\frac{kQ/d_y}{1 + kQ/d_y}\right)(\mathbf{b}^\top \mathbf{S}_y^{-1}(\mathbf{m}_y - \mathbf{a}))^2 \\
&\quad + E\left(\left(\frac{kQ/d_y}{1 + kQ/d_y}\right)^2\right)E\left((\mathbf{b}^\top \mathbf{S}_y^{-1/2}\mathbf{U})^2((\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1/2}\mathbf{U})^2 | \mathbf{y}\right)
\end{aligned}$$

and

$$\mathbb{E}\left(\epsilon\mathbf{b}^\top \boldsymbol{\Upsilon}\mathbf{b}|\mathbf{y}\right) = (\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1}(\mathbf{m}_y - \mathbf{a})\mathbf{b}^\top \mathbf{S}_y^{-1}\mathbf{b} + \mathbb{E}\left(\frac{kQ/d_y}{1+kQ/d_y}\right)\left(\mathbf{b}^\top \mathbf{S}_y^{-1}\mathbf{b} - \frac{1}{k}(\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1}(\mathbf{m}_y - \mathbf{a})\mathbf{b}^\top \mathbf{S}_y^{-1}\mathbf{b}\right)$$

$$- \frac{1}{k}(\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1}(\mathbf{m}_y - \mathbf{a})\mathbf{b}^\top \mathbf{S}_y^{-1}\mathbf{b} - \frac{1}{k}\mathbb{E}\left(\frac{kQ/d_y}{1+kQ/d_y}\right)\mathbf{b}^T \mathbf{S}_y^{-1}\mathbf{b}$$

$$+ E\left(\left(\frac{kQ/d_y}{1+kQ/d_y}\right)^2\right)\mathbb{E}\left((\mathbf{b}^\top \mathbf{S}_y^{-1/2}\mathbf{u})^2((\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1/2}\mathbf{u})^2|\mathbf{y}\right).$$

Since $\frac{kQ/d_y}{1+kQ/d_y}$ has a beta distribution with $k/2$ and $d_y/2$ degrees of freedom, we obtain

$$E\left(\frac{kQ/d_y}{1+kQ/d_y}\right) = \frac{k}{k+d_y},$$

$$E\left(\frac{kQ/d_y}{1+kQ/d_y}\right)^2 = \frac{2kd_y + k^2(k+d_y+2)}{(k+d_y)^2(k+d_y+2)} = \frac{k(k+2)}{(k+d_y)(k+d_y+2)}.$$

Furthermore, using $Q \sim \mathcal{F}(k, d_y)$, we get

$$E\left[\frac{kQ/d_y}{(1+kQ/d_y)^2}\right] = \frac{1}{n_0}\int_0^\infty \frac{kt/d_y}{(1+kt/d_y)^2}\frac{1}{B\left(\frac{k}{2},\frac{d_y}{2}\right)}\left(\frac{k}{d_y}\right)^{k/2}t^{k/2-1}\left(1+\frac{k}{d_y}t\right)^{-(k+d_y)/2}dt$$

$$= \frac{1}{B\left(\frac{k}{2},\frac{d_y}{2}\right)}\int_0^\infty \left(\frac{k}{d_y}\right)^{(k+2)/2}t^{(k+2)/2-1}\left(1+\frac{k}{d_y}t\right)^{-(k+d_y+4)/2}dt = \frac{B\left(\frac{k+2}{2},\frac{d_y+2}{2}\right)}{B\left(\frac{k}{2},\frac{d_y}{2}\right)}$$

$$= \frac{kd_y}{(k+d_y)(k+d_y+2)},$$

where $B(\cdot,\cdot)$ stands for the beta function (see, [38, p. 256]).

Next, we compute $\mathbb{E}\left((\mathbf{b}^\top \mathbf{S}_y^{-1/2}\mathbf{u})^2((\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1/2}\mathbf{u})^2|\mathbf{y}\right)$. Let $Q_N \sim \chi_k^2$ be independent of $\mathbf{u}$. Then $\sqrt{Q_N}\mathbf{u}$ has a multivariate standard normal distribution, i.e.

$$\begin{pmatrix}\mathbf{b}^\top \mathbf{S}_y^{-1/2}\\(\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1/2}\end{pmatrix}\sqrt{Q_N}\mathbf{u} \sim \mathcal{N}_2\left(\mathbf{0},\begin{pmatrix}\mathbf{b}^\top \mathbf{S}_y^{-1}\mathbf{b} & \mathbf{b}^\top \mathbf{S}_y^{-1}(\mathbf{m}_y - \mathbf{a})\\(\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1}\mathbf{b} & (\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1}(\mathbf{m}_y - \mathbf{a})\end{pmatrix}\right) = \mathcal{N}_2\left(\mathbf{0},\begin{pmatrix}c_1 & c_{12}\\c_{12} & c_2\end{pmatrix}\right),$$

where $c_1$, $c_2$, and $c_{12}$ are defined in the statement of Lemma 3. Hence,

$$\mathbb{E}\left((\mathbf{b}^\top \mathbf{S}_y^{-1/2}\mathbf{u})^2((\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1/2}\mathbf{u})^2|\mathbf{y}\right) = \mathbb{E}\left[\left(\mathbf{b}^\top \mathbf{S}_y^{-1/2}\mathbf{u}\right)^2\left((\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1/2}\mathbf{u}\right)^2|\mathbf{y}\right]\frac{\mathbb{E}(Q_N^2)}{\mathbb{E}(Q_N^2)}$$

$$= \frac{\mathbb{E}\left[\left(\mathbf{b}^\top \mathbf{S}_y^{-1/2}\sqrt{Q_N}\mathbf{u}\right)^2\left((\mathbf{m}_y - \mathbf{a})^\top \sqrt{Q_N}\mathbf{S}_y^{-1/2}\mathbf{u}\right)^2|\mathbf{y}\right]}{\mathbb{E}(Q_N^2)} = \frac{c_1 c_2 + 2c_{12}^2}{k(k+2)},$$

where the last equality follows from Isserlis' theorem (see e.g., [32]).

Hence,

$$E(\mathbf{b}^\top \boldsymbol{\zeta}\boldsymbol{\zeta}^\top \mathbf{b}) = c_{12}^2 + \frac{1}{k}\frac{kd_y}{(k+d_y)(k+d_y+2)}c_1 - \frac{2}{k}\frac{k}{k+d_y}c_{12}^2 + \frac{k(k+2)}{(k+d_y)(k+d_y+2)}\frac{c_1 c_2 + 2c_{12}^2}{k(k+2)}$$

$$= \left(1 - \frac{2}{k+d_y} + \frac{2}{(k+d_y)(k+d_y+2)}\right)c_{12}^2 + \left(\frac{d_y}{(k+d_y)(k+d_y+2)} + \frac{1}{(k+d_y)(k+d_y+2)}c_2\right)c_1$$

and

$$E\left(\epsilon\mathbf{b}^\top \boldsymbol{\Upsilon}\mathbf{b}\right) = c_1 c_2 + \frac{k}{k+d_y}c_1 - \frac{1}{k}\frac{k}{k+d_y}c_1 c_2 - \frac{1}{k}c_1 c_2 - \frac{1}{k}\frac{k}{k+d_y}c_1 + \frac{k(k+2)}{(k+d_y)(k+d_y+2)}\frac{c_1 c_2 + 2c_{12}^2}{k(k+2)}$$

$$= \frac{2}{(k+d_y)(k+d_y+2)}c_{12}^2 + \left(\frac{k-1}{(k+d_y)} + \left(1 - \frac{1}{k} - \frac{1}{k+d_y} + \frac{1}{(k+d_y)(k+d_y+2)}\right)c_2\right)c_1. \quad \square$$

**Proof of Theorem 4.** The results of Theorem 4 are obtained by using Lemma 3 with $\mathbf{b} = C_t\boldsymbol{\ell}$, $\boldsymbol{\Sigma} = \boldsymbol{\Omega}$, $\boldsymbol{\nu} = \boldsymbol{\mu}$, $\mathbf{a} = r_{f,t+1}\mathbf{1}$ and Theorem 3.

(a) In the case of the diffuse prior, using $k_y = n + k + 1$, $d_y = n - k$, $v_y = n$, $\mathbf{m}_y - \mathbf{a} = \bar{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1}$, $\mathbf{S}_y = \mathbf{S}_{t,d}/n$, $c_1 = nC_t^2 \boldsymbol{\ell}^\top \mathbf{S}_{t,d}^{-1} \boldsymbol{\ell}$, $c_2 = n(\bar{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})^\top \mathbf{S}_{t,d}^{-1}(\bar{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})$, and $c_{12} = nC_t \boldsymbol{\ell}^\top \mathbf{S}_{t,d}^{-1}(\bar{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})$ we get

$$\mathbb{V}ar(\boldsymbol{\ell}^\top \mathbf{w}_t | \mathbf{y}) = \frac{1}{n^2} \left\{ n(n+1) \left[ \left( 1 - \frac{2}{n} + \frac{2}{n(n+2)} \right) c_{12}^2 + \left( \frac{n-k}{n(n+2)} + \frac{1}{n(n+2)} c_2 \right) c_1 \right] \right.$$

$$+ n \left[ \left( \frac{k-1}{n} + \left( 1 - \frac{1}{k} - \frac{1}{n} + \frac{1}{n(n+2)} \right) c_2 \right) c_1 + \frac{2}{n(n+2)} c_{12}^2 \right] - \left. (n-1)^2 c_{12}^2 \right\}$$

$$= \frac{n-1}{n^2} c_{12}^2 + c_1 \frac{1}{n^2} \left( \frac{n^2 + k - 2}{n+2} + \frac{n(k-1)}{k} c_2 \right)$$

$$= \boldsymbol{\ell}^\top \left( C_t^2 \left( (n-1)\mathbf{S}_{t,d}^{-1}(\bar{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})(\bar{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})^\top \mathbf{S}_{t,d}^{-1} + \left( \frac{n^2 + k - 2}{n(n+2)} + \frac{k-1}{k} b_d \right) \mathbf{S}_{t,d}^{-1} \right) \right) \boldsymbol{\ell}$$

where $b_d = n(\bar{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})^\top \mathbf{S}_{t,d}^{-1}(\bar{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})$. Since $\boldsymbol{\ell}$ is an arbitrary vector, the results in part (a) follow.

(b) In the case of the conjugate prior, the application of $k_y = n + d_0 + 1$, $d_y = n + d_0 - 2k$, $v_y = n + r_0$, $\mathbf{m}_y - \mathbf{a} = \bar{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1}$, and $\mathbf{S}_y = \mathbf{S}_{t,c}/(n+r_0)$, $c_1 = (n+r_0)C_t^2 \boldsymbol{\ell}^\top \mathbf{S}_{t,c}^{-1} \boldsymbol{\ell}$, $c_2 = (n+r_0)(\bar{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})^\top \mathbf{S}_{t,c}^{-1}(\bar{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})$, and $c_{12} = (n+r_0)C_t \boldsymbol{\ell}^\top \mathbf{S}_{t,c}^{-1}(\bar{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1})$. leads to

$$\mathbb{V}ar(\boldsymbol{\ell}^\top \mathbf{w}_t | \mathbf{y}) = \frac{1}{(n+r_0)^2} \left\{ (n + d_0 - k)(n + d_0 - k + 1) \right.$$

$$\times \left[ \left( 1 - \frac{2}{n + d_0 - k} + \frac{2}{(n + d_0 - k)(n + d_0 - k + 2)} \right) c_{12}^2 \right.$$

$$+ \left. \left( \frac{n + d_0 - 2k}{(n + d_0 - k)(n + d_0 - k + 2)} + \frac{1}{(n + d_0 - k)(n + d_0 - k + 2)} c_2 \right) c_1 \right]$$

$$+ (n + d_0 - k) \left[ \left( \frac{k-1}{n + d_0 - k} + \left( 1 - \frac{1}{k} - \frac{1}{n + d_0 - k} + \frac{1}{(n + d_0 - k)(n + d_0 - k + 2)} \right) c_2 \right) c_1 \right.$$

$$+ \left. \frac{2}{(n + d_0 - k)(n + d_0 - k + 2)} c_{12}^2 \right] - \left. (n + d_0 - k - 1)^2 c_{12}^2 \right\}$$

$$= \frac{1}{(n+r_0)^2} \left[ \frac{n + d_0 - k - 1}{(n + d_0 - k)^2} c_{12}^2 + c_1 \left( \frac{(n + d_0 - k)^2 + k - 2}{n + d_0 - k + 2} + \frac{(n + d_0 - k)(k - 1)}{k} c_2 \right) \right]$$

$$= \boldsymbol{\ell}^\top \left\{ C_t^2 \left[ (n + d_0 - k - 1)\mathbf{S}_{t,c}^{-1}(\bar{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1})(\bar{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1})^\top \mathbf{S}_{t,c}^{-1} \right. \right.$$

$$+ \left. \left. \left( \frac{(n + d_0 - k)^2 + k - 2}{(n + r_0)(n + d_0 - k + 2)} + \frac{(n + d_0 - k)(k - 1)}{(n + r_0)k} b_c \right) \mathbf{S}_{t,c}^{-1} \right] \right\} \boldsymbol{\ell},$$

where $b_c = (n + r_0)(\bar{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1})^\top \mathbf{S}_{t,c}^{-1}(\bar{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1})$. Since $\boldsymbol{\ell}$ is an arbitrary vector, we get the statement of Theorem 4.(b). □

**Proof of Theorem 5.** Let $\boldsymbol{\ell}$ be an arbitrary $k$-dimensional vector. From Theorem 1 with $\mathbf{L} = \boldsymbol{\ell}^\top$, we get the following stochastic representations of $\mathbf{L}\mathbf{w}_t$ under the diffuse prior and the conjugate prior expressed as

$$\boldsymbol{\ell}^\top \mathbf{w}_t \stackrel{d}{=} C_t \eta \boldsymbol{\ell}^\top \mathbf{S}_{t,d}^*(\boldsymbol{\mu})^{-1}(\boldsymbol{\mu} - r_{f,t+1}) + C_t \sqrt{\eta} \left( (\boldsymbol{\mu} - r_{f,t+1})^\top \mathbf{S}_{t,d}^*(\boldsymbol{\mu})^{-1}(\boldsymbol{\mu} - r_{f,t+1}) \cdot \boldsymbol{\ell}^\top \mathbf{S}_{t,d}^*(\boldsymbol{\mu})^{-1} \boldsymbol{\ell} \right.$$

$$- \left. \boldsymbol{\ell}^\top \mathbf{S}_{t,d}^*(\boldsymbol{\mu})^{-1}(\boldsymbol{\mu} - r_{f,t+1})(\boldsymbol{\mu} - r_{f,t+1})^\top \mathbf{S}_{t,d}^*(\boldsymbol{\mu})^{-1} \boldsymbol{\ell} \right)^{1/2} z_0,$$

where $\eta \sim \chi_n^2$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$, and $\boldsymbol{\mu}|\mathbf{x} \sim t_k \left( n - k, \bar{\mathbf{x}}_{t,d}, \mathbf{S}_{t,d}/(n(n-k)) \right)$, and

$$\boldsymbol{\ell}^\top \mathbf{w}_t \stackrel{d}{=} C_t \eta \boldsymbol{\ell}^\top \mathbf{S}_{t,c}^*(\boldsymbol{\mu})^{-1}(\boldsymbol{\mu} - r_{f,t+1}) + C_t \sqrt{\eta} \left( (\boldsymbol{\mu} - r_{f,t+1})^\top \mathbf{S}_{t,c}^*(\boldsymbol{\mu})^{-1}(\boldsymbol{\mu} - r_{f,t+1}) \cdot \boldsymbol{\ell}^\top \mathbf{S}_{t,c}^*(\boldsymbol{\mu})^{-1} \boldsymbol{\ell} \right.$$

$$- \left. \boldsymbol{\ell}^\top \mathbf{S}_{t,c}^*(\boldsymbol{\mu})^{-1}(\boldsymbol{\mu} - r_{f,t+1})(\boldsymbol{\mu} - r_{f,t+1})^\top \mathbf{S}_{t,c}^*(\boldsymbol{\mu})^{-1} \boldsymbol{\ell} \right)^{1/2} z_0,$$

where $\eta \sim \chi_{n+d_0-k}^2$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$, and $\boldsymbol{\mu}|\mathbf{x} \sim t_k \left( n + d_0 - 2k, \bar{\mathbf{x}}_{t,c}, \mathbf{S}_{t,c}/((n+r_0)(n+d_0-2k)) \right)$.

Moreover, since

$$\sqrt{n}\left(\begin{pmatrix} \eta/n \\ \mathbf{z}_0/\sqrt{n} \\ \boldsymbol{\mu} \end{pmatrix} - \begin{pmatrix} 1 \\ \mathbf{0} \\ \overline{\mathbf{x}}_{t,d} \end{pmatrix}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \check{\mathbf{S}}_t \end{pmatrix}\right)$$

and

$$\sqrt{n}\left(\begin{pmatrix} \eta/n \\ \mathbf{z}_0/\sqrt{n} \\ \boldsymbol{\mu} \end{pmatrix} - \begin{pmatrix} 1 \\ \mathbf{0} \\ \overline{\mathbf{x}}_{t,c} \end{pmatrix}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \check{\mathbf{S}}_t \end{pmatrix}\right)$$

as $n \longrightarrow \infty$ as well as

$$\lim_{n \to \infty} \overline{\mathbf{x}}_{t,c} = \check{\mathbf{x}}_t = \lim_{n \to \infty} \overline{\mathbf{x}}_{t,d}$$

and

$$\lim_{n \to \infty} \frac{\mathbf{S}_{t,c}}{n + r_0} = \check{\mathbf{S}}_t = \lim_{n \to \infty} \frac{\mathbf{S}_{t,d}}{n - 1},$$

the application of the delta method (see, [24, Theorem 3.7]) proves that

$$\sqrt{n}(\boldsymbol{\ell}^{\top}\mathbf{w}_t - \boldsymbol{\ell}^{\top}\hat{\mathbf{w}}_t)|\mathbf{x}_{t,n} \xrightarrow{d.} \mathcal{N}_k(\mathbf{0}, f_d)$$

and

$$\sqrt{n}(\boldsymbol{\ell}^{\top}\mathbf{w}_t - \boldsymbol{\ell}^{\top}\hat{\mathbf{w}}_t)|\mathbf{x}_{t,n} \xrightarrow{d.} \mathcal{N}_k(\mathbf{0}, f_c),$$

as $n \longrightarrow \infty$ under the diffuse prior and the conjugate prior, respectively.

Finally, the results of Theorem 4 yield

$$f_d = \lim_{n \to \infty} \mathbb{V}ar(\sqrt{n}\boldsymbol{\ell}^{\top}\mathbf{w}_t) = \lim_{n \to \infty} \boldsymbol{\ell}^{\top}\left\{ C_t^2\left(n(n-1)\mathbf{S}_{t,d}^{-1}(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})^{\top}\mathbf{S}_{t,d}^{-1}\right.\right.$$

$$\left.\left. + \left(\frac{n^2 + k - 2}{n(n+2)} + \frac{k-1}{k}b_d\right)\mathbf{S}_{t,d}^{-1}\right)\right\}\boldsymbol{\ell}$$

$$= \boldsymbol{\ell}^{\top}\left\{ C_t^2\left[\check{\mathbf{S}}_t^{-1}(\check{\mathbf{x}}_t - r_{f,t+1}\mathbf{1})(\check{\mathbf{x}}_t - r_{f,t+1}\mathbf{1})^{\top}\check{\mathbf{S}}_t^{-1} + \left(1 + \frac{k-1}{k}(\check{\mathbf{x}}_t - r_{f,t+1}\mathbf{1})^{\top}\check{\mathbf{S}}_t^{-1}(\check{\mathbf{x}}_t - r_{f,t+1}\mathbf{1})\right)\check{\mathbf{S}}_t^{-1}\right]\right\}\boldsymbol{\ell}$$

and, similarly,

$$f_c = \boldsymbol{\ell}^{\top}\left\{ C_t^2\left[\check{\mathbf{S}}_t^{-1}(\check{\mathbf{x}}_t - r_{f,t+1}\mathbf{1})(\check{\mathbf{x}}_t - r_{f,t+1}\mathbf{1})^{\top}\check{\mathbf{S}}_t^{-1} + \left(1 + \frac{k-1}{k}(\check{\mathbf{x}}_t - r_{f,t+1}\mathbf{1})^{\top}\check{\mathbf{S}}_t^{-1}(\check{\mathbf{x}}_t - r_{f,t+1}\mathbf{1})\right)\check{\mathbf{S}}_t^{-1}\right]\right\}\boldsymbol{\ell} = f_d.$$

Since, for each $\boldsymbol{\ell}$ the linear combination $\boldsymbol{\ell}^{\top}\mathbf{w}_t$ is asymptotically normally distributed, then we also get that the vector of weights $\mathbf{w}_t$ is asymptotically normal. $\square$

**Proof of Theorem 6.** Since $\mathbf{x}_{t+1}|\boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and it is conditionally independent of $\mathbf{x}_{t,n}$, we get

$$\widehat{W}_{t+1}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{x}_{t,n} \sim \mathcal{N}(W_t(1 + r_{f,t+1} + \mathbf{v}_t^{\top}(\boldsymbol{\mu} - r_{f,t+1})), W_t^2 \mathbf{v}_t^{\top}\boldsymbol{\Sigma}\mathbf{v}_t).$$

(a) In the case of the diffuse prior, we observe that

$$\frac{\mathbf{v}_t^{\top}\boldsymbol{\Sigma}\mathbf{v}_t}{\mathbf{v}_t^{\top}\mathbf{S}_{t,d}(\boldsymbol{\mu})^*\mathbf{v}_t} \stackrel{d}{=} \frac{1}{\xi}, \tag{12}$$

where $\xi \sim \chi_{n-k+1}^2$ which is independent of $\boldsymbol{\mu}$ (see, e.g., Theorem 3.2.13 in [42]). Then the stochastic representation of $\widehat{W}_{t+1}$ is given by

$$\widehat{W}_{t+1} \stackrel{d}{=} W_t\left(1 + r_{f,t+1} + \mathbf{v}_t^{\top}(\boldsymbol{\mu} - r_{f,t+1}) + \frac{\sqrt{\mathbf{v}_t^{\top}\mathbf{S}_{t,d}(\boldsymbol{\mu})^*\mathbf{v}_t}}{\sqrt{n-k+1}}t_2\right),$$

where $t_2 \sim t_1(n - k + 1, 0, 1)$ is independent of $\boldsymbol{\mu}$. Finally, from the properties of the multivariate $t$-distribution, we obtain

$$\mathbf{v}_t^{\top}(\boldsymbol{\mu} - \overline{\mathbf{x}}_{t,d}) \sim t_1\left(n - k, 0, \frac{\mathbf{v}_t^{\top}\mathbf{S}_{t,d}\mathbf{v}_t}{n(n-k)}\right),$$

which leads to

$$\widehat{W}_{t+1} \stackrel{d}{=} W_t \left( 1 + r_{f,t+1} + \mathbf{v}_t^\top (\bar{\mathbf{x}}_{t,d} - r_{f,t+1}) + \sqrt{\mathbf{v}_t^\top \mathbf{S}_{t,d} \mathbf{v}_t} \left( \frac{t_1}{\sqrt{n(n-k)}} + \sqrt{1 + \frac{t_1^2}{n-k}} \frac{t_2}{\sqrt{n-k+1}} \right) \right),$$

where $t_1$ and $t_2$ are independent with $t_1 \sim t_{n-k}$ and $t_2 \sim t_{n-k+1}$.

(b) Similarly, for the conjugate prior, it holds that

$$\frac{\mathbf{v}_t^\top \mathbf{\Sigma} \mathbf{v}_t}{\mathbf{v}_t^\top \mathbf{S}_{t,c}(\boldsymbol{\mu})^* \mathbf{v}_t} \stackrel{d}{=} \frac{1}{\xi}, \tag{13}$$

where $\xi \sim \chi^2_{n+d_0-2k+1}$ which is independent of $\boldsymbol{\mu}$. Then the stochastic representation of $\widehat{W}_{t+1}$ is given by

$$\widehat{W}_{t+1} \stackrel{d}{=} W_t \left( 1 + r_{f,t+1} + \mathbf{v}_t^\top (\boldsymbol{\mu} - r_{f,t+1}) + \frac{\sqrt{\mathbf{v}_t^\top \mathbf{S}_{t,c}(\boldsymbol{\mu})^* \mathbf{v}_t}}{\sqrt{n+d_0-2k+1}} t_2 \right),$$

where $t_2 \sim t_{n+d_0-2k+1}$ is independent of $\boldsymbol{\mu}$. From the properties of the multivariate $t$-distribution, we get

$$\mathbf{v}_t^\top (\boldsymbol{\mu} - \bar{\mathbf{x}}_{t,c}) \sim t_1 \left( n + d_0 - 2k, 0, \frac{\mathbf{v}_t^\top \mathbf{S}_{t,c} \mathbf{v}_t}{(n+r_0)(n+d_0-2k)} \right),$$

which leads to

$$\widehat{W}_{t+1} \stackrel{d}{=} W_t \left( 1 + r_{f,t+1} + \mathbf{v}_t^\top (\bar{\mathbf{x}}_{t,c} - r_{f,t+1}) \right.$$

$$\left. + \sqrt{\mathbf{v}_t^\top \mathbf{S}_{t,c} \mathbf{v}_t} \left( \frac{t_1}{\sqrt{(n+r_0)(n+d_0-2k)}} + \sqrt{1 + \frac{t_1^2}{n+d_0-2k}} \frac{t_2}{\sqrt{n+d_0-2k+1}} \right) \right),$$

where $t_1$ and $t_2$ are independent with $t_1 \sim t_{n+d_0-2k}$ and $t_2 \sim t_{n+d_0-2k+1}$. □

## Appendix B. Empirical Bayes estimation of the hyperparameters in the conjugate prior

In this section, we derive the empirical Bayes estimates for the hyperparameters of the conjugate prior $\mathbf{m}_0$ and $\mathbf{S}_0$. Given the sample $\mathbf{x}_{\tau,n}$ the empirical Bayes estimates for $\mathbf{m}_0$ and $\mathbf{S}_0$ are obtained by maximizing (see, e.g., [23])

$$g(\mathbf{m}_0, \mathbf{S}_0) = \int_{\boldsymbol{\mu}} \int_{\mathbf{\Sigma}} L(\mathbf{x}_{t,n} | \boldsymbol{\mu}, \mathbf{\Sigma}) \pi(\boldsymbol{\mu}, \mathbf{\Sigma}) \mathrm{d}\mathbf{\Sigma} \mathrm{d}\boldsymbol{\mu} \tag{14}$$

with respect to $\mathbf{m}_0$ and $\mathbf{S}_0$.

First, we calculate the integral in (14), ignoring the terms which do not depend on $\mathbf{m}_0$ and $\mathbf{S}_0$, to get

$$g(\mathbf{m}_0, \mathbf{S}_0)$$

$$\propto \int_{\boldsymbol{\mu}} \int_{\mathbf{\Sigma}} L(\mathbf{x}_{t,n} | \boldsymbol{\mu}, \mathbf{\Sigma}) \pi(\boldsymbol{\mu}, \mathbf{\Sigma}) \mathrm{d}\mathbf{\Sigma} \mathrm{d}\boldsymbol{\mu} \propto \int_{\boldsymbol{\mu}} \int_{\mathbf{\Sigma}} |\mathbf{\Sigma}|^{-n/2} \exp\left\{ -\frac{n}{2} (\bar{\mathbf{x}}_\tau - \boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1} (\bar{\mathbf{x}}_\tau - \boldsymbol{\mu}) - \frac{n-1}{2} tr(\mathbf{S}_\tau \mathbf{\Sigma}^{-1}) \right\}$$

$$\times |\mathbf{\Sigma}|^{-1/2} \exp\left\{ -\frac{r_0}{2} (\boldsymbol{\mu} - \mathbf{m}_0)^\top \mathbf{\Sigma}^{-1} (\boldsymbol{\mu} - \mathbf{m}_0) - \right\} |\mathbf{\Sigma}|^{-d_0/2} |\mathbf{S}_0|^{(d_0-k-1)/2} \exp\left\{ -\frac{1}{2} tr(\mathbf{S}_0 \mathbf{\Sigma}^{-1}) \right\} \mathrm{d}\mathbf{\Sigma} \mathrm{d}\boldsymbol{\mu}$$

$$= |\mathbf{S}_0|^{(d_0-k-1)/2} \int_{\boldsymbol{\mu}} \int_{\mathbf{\Sigma}} |\mathbf{\Sigma}|^{-(n+d_0+1)/2} \exp\left\{ -\frac{1}{2} tr\left( \mathbf{\Sigma}^{-1} \mathbf{V}_\tau(\boldsymbol{\mu}; \mathbf{m}_0, \mathbf{S}_0) \right) \right\} \mathrm{d}\mathbf{\Sigma} \mathrm{d}\boldsymbol{\mu}$$

$$\propto |\mathbf{S}_0|^{(d_0-k-1)/2} \int_{\boldsymbol{\mu}} |\mathbf{V}_\tau(\boldsymbol{\mu}; \mathbf{m}_0, \mathbf{S}_0)|^{-(n+d_0-k)/2} \mathrm{d}\boldsymbol{\mu},$$

where the last identity is obtained by recognizing that under the integral with respect to $\mathbf{\Sigma}$ we have a kernel of the density function of $\mathcal{IW}_k(n + d_0 + 1, \mathbf{V}_\tau(\boldsymbol{\mu}; \mathbf{m}_0, \mathbf{S}_0))$ with $\bar{\mathbf{y}}_\tau(\mathbf{m}_0) = (n\bar{\mathbf{x}}_\tau + r_0\mathbf{m}_0)/(n+r_0)$ and

$$\mathbf{V}_\tau(\boldsymbol{\mu}; \mathbf{m}_0, \mathbf{S}_0) = \mathbf{S}_0 + (n-1)\mathbf{S}_\tau + r_0(\boldsymbol{\mu} - \mathbf{m}_0)(\boldsymbol{\mu} - \mathbf{m}_0)^\top + n(\bar{\mathbf{x}}_\tau - \boldsymbol{\mu})(\bar{\mathbf{x}}_\tau - \boldsymbol{\mu})^\top$$

$$= \mathbf{S}_0 + (n-1)\mathbf{S}_\tau + nr_0 \frac{(\mathbf{m}_0 - \bar{\mathbf{y}}_\tau(\mathbf{m}_0))(\mathbf{m}_0 - \bar{\mathbf{y}}_\tau(\mathbf{m}_0))^\top}{n+r_0} + (n+r_0)(\boldsymbol{\mu} - \bar{\mathbf{y}}_\tau((\mathbf{m}_0)))(\boldsymbol{\mu} - \bar{\mathbf{y}}_\tau(\mathbf{m}_0))^\top.$$

Let $\widetilde{\mathbf{V}}_\tau(\mathbf{m}_0, \mathbf{S}_0) = \mathbf{S}_0 + (n-1)\mathbf{S}_\tau + nr_0(\mathbf{m}_0 - \bar{\mathbf{y}}_\tau(\mathbf{m}_0))(\mathbf{m}_0 - \bar{\mathbf{y}}_\tau(\mathbf{m}_0))^\top / (n+r_0)$. The application of Sylvester's determinant theorem leads to

$$|\mathbf{V}_\tau(\boldsymbol{\mu}; \mathbf{m}_0, \mathbf{S}_0)| = |\widetilde{\mathbf{V}}_\tau(\mathbf{m}_0, \mathbf{S}_0)|(1 + (n+r_0)(\boldsymbol{\mu} - \bar{\mathbf{y}}_\tau(\mathbf{m}_0))^\top \widetilde{\mathbf{V}}_\tau(\mathbf{m}_0, \mathbf{S}_0)^{-1}(\boldsymbol{\mu} - \bar{\mathbf{y}}_\tau(\mathbf{m}_0))).$$

Hence,

$$
\begin{aligned}
g(\mathbf{m}_0, \mathbf{S}_0) & \\
& \propto |\mathbf{S}_0|^{(d_0-k-1)/2} \int_{\boldsymbol{\mu}} |\mathbf{V}_\tau(\boldsymbol{\mu}; \mathbf{m}_0, \mathbf{S}_0)|^{-(n+d_0-k)/2} \mathrm{d}\boldsymbol{\mu} \\
& \propto |\mathbf{S}_0|^{(d_0-k-1)/2} |\widetilde{\mathbf{V}}_\tau(\mathbf{m}_0, \mathbf{S}_0)|^{-(n+d_0-k)/2} \int_{\boldsymbol{\mu}} (1 + (n+r_0)(\boldsymbol{\mu} - \bar{\mathbf{y}}_\tau(\mathbf{m}_0))^\top \widetilde{\mathbf{V}}_\tau(\mathbf{m}_0, \mathbf{S}_0)^{-1} (\boldsymbol{\mu} - \bar{\mathbf{y}}_\tau(\mathbf{m}_0)))^{-(n+d_0-k)/2} \mathrm{d}\boldsymbol{\mu} \\
& \propto |\mathbf{S}_0|^{(d_0-k-1)/2} |\widetilde{\mathbf{V}}_\tau(\mathbf{m}_0, \mathbf{S}_0)|^{-(n+d_0-k-1)/2} = |\mathbf{S}_0|^{(d_0-k-1)/2} |\mathbf{S}_0 + (n-1)\mathbf{S}_\tau|^{-(n+d_0-k-1)/2} \\
& \quad \times \left(1 + nr_0(\mathbf{m}_0 - \bar{\mathbf{y}}_\tau(\mathbf{m}_0))^\top (\mathbf{S}_0 + (n-1)\mathbf{S}_\tau)^{-1}(\mathbf{m}_0 - \bar{\mathbf{y}}_\tau(\mathbf{m}_0))/(n+r_0)\right)^{-(n+d_0-k-1)/2},
\end{aligned}
$$

where we use Sylvester's determinant theorem for the second time. From the last line, we conclude that $g(\mathbf{m}_0, \mathbf{S}_0)$ is maximized with respect to $\mathbf{m}_0$ at $\hat{\mathbf{m}}_0$ satisfying $\mathbf{m}_0 = \bar{\mathbf{y}}_\tau(\mathbf{m}_0)$ independently of $\mathbf{S}_0$ leading to $\hat{\mathbf{m}}_0 = \bar{\mathbf{x}}_\tau$.

Taking the logarithms of $g(\mathbf{m}_0, \mathbf{S}_0)$, calculating the matrix derivative with respect to $\mathbf{S}_0$ which is then set to the zero matrix, and substituting $\mathbf{m}_0$ by $\hat{\mathbf{m}}_0$, we get the following matrix equation

$$
\frac{d_0 - k - 1}{2} \mathbf{S}_0^{-1} - \frac{n + d_0 - k - 1}{2} (\mathbf{S}_0 + (n-1)\mathbf{S}_\tau)^{-1} = \mathbf{O}
$$

with the solution given by

$$
\hat{\mathbf{S}}_0 = \frac{(d_0 - k - 1)(n - 1)}{n} \mathbf{S}_\tau.
$$

## References

[1] O. Aguilar, M. West, Bayesian dynamic factor models and portfolio allocation, J. Bus. Econom. Statist. 18 (2000) 338–357.
[2] T. Ando, Bayesian portfolio selection using a multifactor model, Int. J. Forecast. 25 (2009) 550–566.
[3] D. Avramov, G. Zhou, Bayesian portfolio analysis, Ann. Rev. Financial Econ. 2 (2010) 25–47.
[4] J. Bai, S. Shi, Estimating high dimensional covariance matrices and its applications, Ann. Econ. Finance 12 (2011) 199–215.
[5] C. Barry, Portfolio analysis under uncertain means, variances, and covariances, J. Finance 29 (1974) 515–522.
[6] S. Basak, G. Chabakauri, Dynamic mean-variance asset allocation, Rev. Financ. Stud. 23 (2010) 2970–3016.
[7] D. Bauder, R. Bodnar, T. Bodnar, W. Schmid, Bayesian estimation of the efficient frontier, Scand. J. Stat. 46 (2019) 802–830.
[8] J.M. Bernardo, A.F.M. Smith, Bayesian Theory, Wiley, 2000.
[9] T. Bodnar, H. Dette, N. Parolya, Testing for independence of large dimensional vectors, Ann. Statist. 47 (2019) 2977–3008.
[10] T. Bodnar, S. Dmytriv, N. Parolya, W. Schmid, Tests for the weights of the global minimum variance portfolio in a high-dimensional setting, IEEE Trans. Signal Process. 67 (2019) 4479–4493.
[11] T. Bodnar, S. Mazur, Y. Okhrin, Bayesian estimation of the global minimum variance portfolio, European J. Oper. Res. 256 (2017) 292–307.
[12] T. Bodnar, S. Mazur, K. Podgórski, Singular inverse wishart distribution and its application to portfolio theory, J. Multivariate Anal. 143 (2016) 314–326.
[13] T. Bodnar, Y. Okhrin, On the product of inverse Wishart and normal distributions with applications to discriminant analysis and portfolio theory, Scand. J. Statist. 38 (2011) 311–331.
[14] T. Bodnar, N. Parolya, W. Schmid, A closed-form solution of the multi-period portfolio choice problem for a quadratic utility function, Ann. Oper. Res. 229 (2015) 121–158.
[15] T. Bodnar, N. Parolya, W. Schmid, On the exact solution of the multi-period portfolio choice problem for an exponential utility under return predictability, European J. Oper. Res. 246 (2015) 528–542.
[16] T. Bodnar, W. Schmid, A test for the weights of the global minimum variance portfolio in an elliptical model, Metrika 67 (2) (2008) 127–143.
[17] T. Bodnar, W. Schmid, Econometrical analysis of the sample efficient frontier, Eur. J. Finance 15 (2009) 317–335.
[18] T. Bodnar, W. Schmid, On the exact distribution of the estimated expected utility portfolio weights: Theory and applications, Statist. Risk Model. 28 (2011) 319–342.
[19] M. Brandt, Portfolio choice problems, in: Y. Ait-Sahalia, L. Hansen (Eds.), Handbook of Financial Econometrics, volume 1, Tools and Techniques, North Holland, 2010, pp. 269–336.
[20] M. Brandt, Santa-Clara, Dynamic portfolio selection by augmenting the asset space, J. Finance 61 (2006) 2187–2217.
[21] S. Brown, Optimal Portfolio Choice under Uncertainty: A Bayesian Approach (Ph.D. thesis), University of Chicago, 1976.
[22] E. Çanakoğlu, S. Özekici, Portfolio selection in stochastic markets with exponential utility functions, Ann. Oper. Res. 166 (2009) 281–297.
[23] B.P. Carlin, T.A. Louis, Bayes and Empirical Bayes Methods for Data Analysis, Chapman & Hall/CRC, 2000.
[24] A. DasGupta, Asymptotic Theory of Statistics and Probability, in: Springer Texts in Statistics, Springer, 2008.
[25] E.J. Elton, M.J. Gruber, On the optimality of some multiperiod portfolio selection criteria, J. Bus. 47 (1974) 231–243.
[26] P. Frost, J. Savarino, An empirical Bayes approach to efficient portfolio selection, J. Financ. Quant. Anal. 21 (1986) 293–305.
[27] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, Bayesian Data Analysis, volume 2, Chapman & Hall/CRC Boca Raton, FL, USA, 2014.
[28] M.R. Gibbons, S.A. Ross, J. Shanken, A test of the efficiency of a given portfolio, Econometrica 57 (1989) 1121–1152.
[29] G.H. Givens, J.A. Hoeting, Computational Statistics, John Wiley & Sons, 2012.
[30] A. Gupta, D. Nagar, Matrix Variate Distributions, Chapman and Hall/CRC, Boca Raton, 2000.
[31] A. Gupta, T. Varga, T. Bodnar, Elliptically Contoured Models in Statistics and Portfolio Theory, second ed., Springer, 2013.
[32] L. Isserlis, On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables, Biometrika 12 (1918) 134–139.
[33] R. Klein, V. Bawa, The effect of estimation risk on optimal portfolio choice, J. Financ. Econ. 3 (1976) 215–231.
[34] S. Kotz, S. Nadarajah, Multivariate *t* Distributions and their Applications, Cambridge University Press, Cambridge, United Kingdom, 2004.
[35] D. Li, W.-L. Ng, Optimal dynamic portfolio selection: Multiperiod mean-variance formulation, Math. Finance 10 (2000) 387–406.
[36] H. Markowitz, Portfolio selection, J. Finance 7 (1952) 77–91.
[37] H. Markowitz, Portfolio Selection: Efficient Diversification of Investments, John Wiley, New York, 1959.
[38] A.M. Mathai, S.B. Provost, Quadratic Forms an Random Variables, Marcel Dekker, New York, 1992.

[39] R.C. Merton, Lifetime portfolio selection under uncertainty: the continuous time case, Rev. Econ. Stat. 50 (1969) 247–257.
[40] C.D. Meyer, Matrix Analysis and Applied Linear Algebra, SIAM, 2000.
[41] J. Mossin, Optimal multiperiod portfolio policies, J. Bus. 41 (1968) 215–229.
[42] R.J. Muirhead, Aspects of Multivariate Statistical Theory, Wiley, New York, 1982.
[43] Y. Okhrin, W. Schmid, Distributional properties of portfolio weights, J. Econometrics 134 (2006) 235–256.
[44] L. Pástor, Portfolio selection and asset pricing models, J. Finance 55 (2000) 179–223.
[45] L. Pástor, R.F. Stambaugh, Comparing asset pricing models: an investment perspective, J. Financ. Econ. 56 (2000) 335–381.
[46] G.G. Pennacchi, Theory of Asset Pricing, Pearson/Addison-Wesley Boston, 2008.
[47] N.G. Polson, B.V. Tew, Bayesian portfolio selection: An empirical analysis of the S&P 500 index 1970–1996, J. Bus. Econom. Statist. 18 (2000) 164–173.
[48] S.T. Rachev, J.S.J. Hsu, B.S. Bagasheva, F.J. Fabozzi, Bayesian Methods in Finance, Wiley, New Jersey, 2008.
[49] P.A. Samuelson, Lifetime portfolio selection by dynamic stochastic programming, Rev. Econ. Stat. 51 (1969) 239–246.
[50] M. Sekerke, Bayesian Risk Management: A Guide to Model Risk and Sequential Learning in Financial Markets, Wiley, New Jersey, 2015.
[51] J. Shanken, On the estimation of beta-pricing models, Rev. Financ. Stud. 5 (1992) 1–33.
[52] J. Shanken, G. Zhou, Estimating and testing beta pricing models: Alternative methods and their performance in simulations, J. Financ. Econ. 84 (2007) 40–86.
[53] D. Sun, J. Berger, Objective Bayesian analysis for the multivariate normal model, in: J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith, M. West (Eds.), Bayesian Statistics, volume 8, Oxford: University Press, 2007, pp. 525–547.
[54] R.L. Winkler, Bayesian models for forecasting future security prices, J. Financ. Quant. Anal. 8 (1973) 387–405.
[55] R.L. Winkler, C.B. Barry, A Bayesian model for portfolio selection and revision, J. Finance 30 (1975) 179–192.
[56] A. Zellner, An Introduction to Bayesian Inference in Econometrics, John Wiley, New York, 1971.
[57] A. Zellner, T. Ando, A direct Monte Carlo approach for Bayesian analysis of the seemingly unrelated regression model, J. Econometrics 159 (2010) 33–45.