ÖREBRO UNIVERSITY
SCHOOL OF BUSINESS

WORKING PAPER

2/2018

Bayesian inference for the tangent portfolio

David Bauder, Taras Bodnar, Stepan Mazur, Yarema Okhrin
Statistics

ISSN 1403-0586

http://www.oru.se/Institutioner/Handelshogskolan-vid-Orebro-universitet/Forskning/Publikationer/Working-papers/
Örebro University School of Business
701 82 Örebro
SWEDEN

Bayesian inference for the tangent portfolio

David Bauder[a], Taras Bodnar[b], Stepan Mazur[c], Yarema Okhrin[d] [1]

[a] *Department of Mathematics, Humboldt-University of Berlin, D-10099 Berlin, Germany*

[b] *Department of Mathematics, Stockholm University, SE-10691 Stockholm, Sweden*

[c] *Department of Statistics, School of Business, Örebro University, SE-70281 Örebro, Sweden*

[d] *Department of Statistics, University of Augsburg, D-86159 Augsburg, Germany*

Abstract

In this paper we consider the estimation of the weights of tangent portfolios from the Bayesian point of view assuming normal conditional distributions of the logarithmic returns. For diffuse and conjugate priors for the mean vector and the covariance matrix, we derive stochastic representations for the posterior distributions of the weights of tangent portfolio and their linear combinations. Separately we provide the mean and variance of the posterior distributions, which are of key importance for portfolio selection. The analytic results are evaluated within a simulation study, where the precision of coverage intervals is assessed.

Keywords: asset allocation, tangent portfolio, Bayesian analysis

JEL Classification: C10, C44

[1]Corresponding author: Yarema Okhrin. E-mail address: yarema.okhrin@wiwi.uni-augsburg.de. This research was partly supported by the German Science Foundation (DFG) via the projects BO 3521/3-1 and SCHM 859/13-1 "Bayesian Estimation of the Multi-Period Optimal Portfolio Weights and Risk Measures" and by the Swedish Research Council (VR) via the project "Bayesian Analysis of Optimal Portfolios and Their Risk Measures". The third author appreciates the financial support of Swedish Research Council: Dynamical stochastic models for efficient spatial analysis of linkages in financial markets, Statistics in the Empirical Sciences

1 Introduction

The seminal paper of Markowitz (1952) suggests a simple and intuitive approach for determining the optimal portfolios of risky assets. It allows us to determine the optimal portfolio weights which lead to the lowest risk for a given expected portfolio return. If the asset returns are assumed to follow normal distribution, then this task is equivalent to minimizing the expected quadratic utility of the future wealth. Depending on the level of the risk aversion or on the expected targeted portfolio return, all the resulting portfolios will lie on a hyperbolic efficient frontier in the μ-σ-space. Taking the risk-free asset into account changes the paradigm of the classical Markowitz approach. In this case the efficient portfolios lie on straight line which crosses the vertical axis at the level of the risk-free rate and is tangent to the mean-variance efficient frontier of Markowitz. The line is usually referred to as the capital market line and the tangent point is the tangent portfolio. Every investor holds a portfolio which consists of the tangent portfolio and the risk-free asset, while the proportions are determined by the risk aversion.

In practice, however, the tagency and other portfolios frequently lead to investment strategies with modest profits and high risk. Several approaches were developed to improve the performance. The first strand of research analyses the estimation risk in portfolio weights, which arises if we replace the unknown parameters of the distribution of asset returns with their sample counterparts. If the estimation risk is properly quantified it can be taken into account when constructing estimation-risk-adjusted portfolios. Alternatively one can shrink the optimal portfolio weights to constant target weights. Typically one takes equally weighted portfolio for this purpose. The objective is to minimize an appropriate objective function, usually the utility function of the investor.

The second strand of research uses the Bayesian framework. The Bayesian setting resembles the human way of information utilization. The investors use the past experiences or additional information for decisions at a given time point. These subjective beliefs are reflected in a Bayesian setup using specific prior distributions. The first applications of Bayesian statistics in portfolio analysis were completely based on uninformative or data-based priors, see Winkler (1973), Winkler and Barry (1975). Bawa et al. (1979) provided an excellent review on early examples of Bayesian studies on portfolio choice. These contributions stimulated a steady growth of interest in Bayesian tools for asset allocation problems. Jorion (1986), Kandel and Stambaugh (1996), Barberis (2000), Pastor (2000) used the Bayesian framework to analyze the impact of the underlying asset pricing or predictive model for asset returns on the optimal portfolio choice. Wang (2005), Kan and Zhou (2007), Golosnoy and Okhrin (2007), Golosnoy and Okhrin (2008), Bodnar et al. (2015) concentrated on shrinkage estimation, which allows to shift the portfolio weights to prespecified values, which reflect the prior beliefs of investors. Brandt (2010) gives a state of the art review of the modern portfolio selection techniques, paying a particular attention to Bayesian approaches.

In this paper we consider diffuse and conjugate priors for the parameters of asset returns. In both cases we derive stochastic representation of the posterior distribution of the tangent portfolio and the corresponding first two moments. These results simplify numerical computation of the optimal portfolios and their analysis, since random sampling is required only for simple and standard distribution such as t, N and F. Additionally we provide the asymptotic distribution, which is Gaussian with a simple expression for the covariance matrix. The established results are evaluated within a simulation study, which assesses the coverage probabilities of credible intervals.

The rest of the paper is structured as follows. Bayesian estimation of the tangent portfolio and main theoretical results are summarized in Section 2. The results of numerical study are given in Section 3, while Section 4 summarizes the paper. The appendix (Section 5) contains proofs and additional technical results.

2 Bayesian estimation of tangent portfolio

We consider a portfolio consisting of k assets. The k-dimensional vector of the asset (logarithmic) returns taken at time point t is denoted by $\mathbf{x}_t$. Let w_i be the i-th weight in the portfolio and let $\mathbf{w} = (w_1, \ldots, w_k)^\top$ be the vector of weights. Throughout the paper it is assumed that data drawn from the random vector of asset returns consist of conditionally independent observations which are conditionally normally distributed. That is we assume that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are independent given $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ with $\mathbf{x}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu}$ is a mean vector, $\boldsymbol{\Sigma}$ is a positive definite covariance matrix, and $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are independent given $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. It is remarkable that only the conditional distribution of the asset returns is assumed to be normal, while the unconditional distribution depends on the priors assigned to $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ and it is usually a heavy-tailed distribution. Moreover, the observation vectors are unconditionally dependent. These two features, namely heavy tails and time dependence, are, usually, observed in the stochastic behavior of the asset returns.

The aim of this section is to provide a Bayesian analysis of the tangent portfolio (TP) which is an optimal portfolio when the investment into a risk-free asset with return r_f is possible. In the mean-variance space this type of optimal portfolios is determined by a tangent line drawn from the portfolio which consists of the risk-free asset only to the set of optimal portfolios, the so-called efficient frontier, constructed in the case without a risk-free asset. The tangent portfolio weights are calculated by

$$\mathbf{w}_{TP} = \alpha^{-1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f\mathbf{1}_k), \tag{1}$$

where α is the coefficient of risk aversion which describes the investor's attitude towards risk and $\mathbf{1}_k$ stands for the k-dimensional vector of ones. If the sum of the weights in (1) is not equal to one, what is usually observed in practice, then the rest of the investor's wealth is invested into the risk-free rate whose weight is $w_0 = 1 - \mathbf{1}_k^\top \mathbf{w}_{TP}$. Otherwise, if it is normalized so that $\mathbf{1}_k^\top \mathbf{w}_{TP} = 1$, then the TP portfolio coincides with the optimal portfolio that maximizes the Sharpe ratio and it is also known as the market portfolio. This portfolio lies on the intersection of the mean-variance efficient frontier and the capital market line constructed with a risk-free asset.

Obviously the TP weights cannot be calculated since both the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ of the asset return distribution are unknown quantities. They have to be replaced by corresponding estimators using the historical data on asset returns $\mathbf{x} = (\mathbf{x}_1, ..., \mathbf{x}_n)$ observed at time points $1, ..., n$. Using these data, the sample estimators for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, namely the sample mean vector and the sample covariance matrix, are constructed and they are given by

$$\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \quad \text{and} \quad \mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^\top, \tag{2}$$

respectively. Replacing $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ by $\overline{\mathbf{x}}$ and $\mathbf{S}$ in (1) we obtain the sample estimator for the TP weights expressed as

$$\hat{\mathbf{w}}_{TP} = \alpha^{-1}\mathbf{S}^{-1}(\overline{\mathbf{x}} - r_f \mathbf{1}_k). \tag{3}$$

In this section we deal with a more general problem. Namely, the aim is to estimate arbitrary linear combinations of the TP weights. Let $\mathbf{L}$ be a $p \times k$ matrix of constants such that $\text{rank}(\mathbf{L}) = p < k$, and define

$$\boldsymbol{\theta} = \mathbf{L}\mathbf{w}_{TP} = \alpha^{-1}\mathbf{L}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f \mathbf{1}_k) \tag{4}$$

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

$$\hat{\boldsymbol{\theta}} = \mathbf{L}\hat{\mathbf{w}}_{TP} = \alpha^{-1}\mathbf{L}\mathbf{S}^{-1}(\overline{\mathbf{x}} - r_f \mathbf{1}_k). \tag{5}$$

The frequentist distribution of the TP weight and of $\hat{\boldsymbol{\theta}}$ as well as test theory on the TP weights were derived by Bodnar and Okhrin (2011) for $n > k$, whereas Bodnar et al. (2016b) extended these results to the case $n < k$.

Here, we deal with the problem of estimating the TP portfolio from the viewpoint of Bayesian statistics. The distributional properties of the TP weights and/or their linear combinations will be presented in terms of the posterior distribution. Thus we obtain not only the point estimator of the weights but the whole distribution. Using the posterior distribution the Bayesian estimate of the TP weights are derived as the posterior mean vector along with their uncertainties which are characterized by the posterior covariance matrix.

The starting point of the Bayesian analysis is the Bayes theorem which relates the posterior distribution of the parameter to the prior distribution and the likelihood function. The latter contains the knowledge about the parameter before the sample is taken. Since the distribution of the asset returns does not directly depend on the TP weights, the posterior distribution of $\mathbf{w}_{TP}$ as well as of $\boldsymbol{\theta}$ is derived from the posterior obtained for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ expressed as

$$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{x}) \propto L(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{6}$$

with the likelihood function given by

$$\begin{aligned} L(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= L(\mathbf{x}_1, \ldots, \mathbf{x}_n|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\right\} \\ &\propto |\boldsymbol{\Sigma}|^{-n/2} \exp\left\{-\frac{n}{2}(\overline{\mathbf{x}} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu}) - \frac{n-1}{2}\mathrm{tr}[\mathbf{S}\boldsymbol{\Sigma}^{-1}]\right\}. \end{aligned} \tag{7}$$

There are several approaches how the prior for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ could be chosen with the diffuse prior and the conjugate prior being the most widely used priors. The diffuse prior belongs to non-informative priors, i.e., it does not incorporate any information for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The diffuse prior is also known as Jeffreys' prior and it is given by

$$\pi_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{k+1}{2}}. \tag{8}$$

The second considered prior is the conjugate prior which is an informative one with a normal prior for $\boldsymbol{\mu}$ (conditional on $\boldsymbol{\Sigma}$) and an inverse Wishart prior for $\boldsymbol{\Sigma}$. It is expressed as

$$\begin{aligned} \pi_c(\boldsymbol{\mu}|\boldsymbol{\Sigma}) &\propto |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{\kappa_c}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_c)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_c)\right\}, \\ \pi_c(\boldsymbol{\Sigma}) &\propto |\boldsymbol{\Sigma}|^{-\nu_c/2} \exp\left\{-\frac{1}{2}\mathrm{tr}[\mathbf{V}_c\boldsymbol{\Sigma}^{-1}]\right\}, \end{aligned}$$

where $\boldsymbol{\mu}_c$ is the prior mean, κ_c is the parameter reflecting the prior precision of $\boldsymbol{\mu}_c$, ν_c is a prior precision on $\boldsymbol{\Sigma}$, and $\mathbf{V}_c$ is a known prior matrix of $\boldsymbol{\Sigma}$. The joint prior for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is then given by

$$\pi_c(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(\nu_c+1)/2} \exp\left\{-\frac{\kappa_c}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_c)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_c) - \frac{1}{2}\mathrm{tr}[\mathbf{V}_c\boldsymbol{\Sigma}^{-1}]\right\}. \tag{9}$$

Both the diffuse and the conjugate priors are successfully applied in finance by Barry (1974), Brown (1976), Klein and Bawa (1976), Frost and Savarino (1986), Rachev et al. (2008), Avramov and Zhou (2010), Sekerke (2015), Bodnar et al. (2016a) among others. The diffuse prior mimics the situation, when the investor has no additional information about the model parameters. The conjugate prior, however, reflects the prior beliefs trough the additional information with the expectations $\boldsymbol{\mu}_c$ and $\boldsymbol{\Sigma}_c$.

In Theorem 1, we derive the stochastic representations of the posterior distributions for $\boldsymbol{\theta}$ under the diffuse prior and the conjugate prior. The stochastic representation is a very powerful tool in multivariate statistics. It plays an important role in the theory of elliptically contoured distributions (c.f., Gupta et al. (2013)) and in Bayesian statistics (see, e.g., Bodnar et al. (2016a)) as well as it is widely used in Monte Carlo studies. In particular, the simulation of the values of the weights is considerably simplified if we use the stochastic representation.

Theorem 1. *Let* $\mathbf{x}_1, \ldots, \mathbf{x}_n|\boldsymbol{\mu}, \boldsymbol{\Sigma}$ *be conditionally independently and identically distributed with* $\mathbf{x}_t|\boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. *Let* $\mathbf{L}$ *be a* $p \times k$ *matrix of constants of rank* $p < k$, *and* $\mathbf{1}_k$ *denotes the vector of ones. We define*

$$\mathbf{a}_1 = \overline{\mathbf{x}} - r_f \mathbf{1}_k, \quad \mathbf{a}_2 = \boldsymbol{\mu}_c - r_f \mathbf{1}_k, \quad \text{and} \quad \mathbf{a}_{12} = \frac{1}{n+\kappa_c}\left(n\mathbf{a}_1 + \kappa_c \mathbf{a}_2\right).$$

Then the stochastic representation of the posterior distribution for $\boldsymbol{\theta} = \mathbf{L}\mathbf{w}_{TP}$

(a) under the diffuse prior $\pi_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *is given by*

$$\boldsymbol{\theta}_d \stackrel{d}{=} \frac{\eta_d}{\alpha} \cdot \mathbf{L}\mathbf{S}_d^{-1}\breve{\boldsymbol{\mu}}_d + \frac{\sqrt{\eta_d}}{\alpha}\left(\breve{\boldsymbol{\mu}}_d^T \mathbf{S}_d^{-1}\breve{\boldsymbol{\mu}}_d \cdot \mathbf{L}\mathbf{S}_d^{-1}\mathbf{L}^\top - \mathbf{L}\mathbf{S}_d^{-1}\breve{\boldsymbol{\mu}}_d\breve{\boldsymbol{\mu}}_d^T\mathbf{S}_d^{-1}\mathbf{L}^\top\right)^{1/2}\mathbf{z}_0, \quad (10)$$

with

$$\mathbf{S}_d = \mathbf{S}_d(\breve{\boldsymbol{\mu}}_d) = (n-1)\mathbf{S} + n(\breve{\boldsymbol{\mu}}_d - \mathbf{a}_1)(\breve{\boldsymbol{\mu}}_d - \mathbf{a}_1)^\top,$$

where $\eta_d \sim \chi^2_n$, $\breve{\boldsymbol{\mu}}_d|\mathbf{x} \sim t_k\left(n-k, \mathbf{a}_1, \frac{n-1}{n(n-k)}\mathbf{S}\right)$, *and* $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$; *moreover,* η_d, $\breve{\boldsymbol{\mu}}_d$, *and* $\mathbf{z}_0$ *are mutually independent.*

(b) under the conjugate prior $\pi_c(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *is given by*

$$\boldsymbol{\theta}_c \stackrel{d}{=} \frac{\eta_c}{\alpha} \cdot \mathbf{L}\mathbf{S}_c^{-1}\breve{\boldsymbol{\mu}}_c + \frac{\sqrt{\eta_c}}{\alpha}\left(\breve{\boldsymbol{\mu}}_c^T \mathbf{S}_c^{-1}\breve{\boldsymbol{\mu}}_c \cdot \mathbf{L}\mathbf{S}_c^{-1}\mathbf{L}^\top - \mathbf{L}\mathbf{S}_c^{-1}\breve{\boldsymbol{\mu}}_c\breve{\boldsymbol{\mu}}_c^T\mathbf{S}_c^{-1}\mathbf{L}^\top\right)^{1/2}\mathbf{z}_0, \quad (11)$$

with

$$\begin{aligned}\mathbf{S}_c &= \mathbf{S}_c(\breve{\boldsymbol{\mu}}_c) = \tilde{\mathbf{S}} + (n+\kappa_c)\left(\breve{\boldsymbol{\mu}}_c - \mathbf{a}_{12}\right)\left(\breve{\boldsymbol{\mu}}_c - \mathbf{a}_{12}\right)^\top,\\ \tilde{\mathbf{S}} &= (n-1)\mathbf{S} + \mathbf{V}_c - (n+\kappa_c)\mathbf{a}_{12}\mathbf{a}_{12}^\top + (n\mathbf{a}_1\mathbf{a}_1^\top + \kappa_c\mathbf{a}_2\mathbf{a}_2^\top),\end{aligned}$$

where $\eta_c \sim \chi^2_{\nu_c+n-k}$, $\breve{\boldsymbol{\mu}}_c|\mathbf{x} \sim t_k\left(\nu_c + n - 2k, \mathbf{a}_{12}, \frac{1}{(n+\kappa_c)(\nu_c+n-2k)}\tilde{\mathbf{S}}\right)$, *and* $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$; *moreover,* η_c, $\breve{\boldsymbol{\mu}}_c$, *and* $\mathbf{z}_0$ *are mutually independent.*

The proof of the theorem is given in the appendix. It is noted that the distribution of the weights is given in terms of a χ^2 random variable, a t-distributed random vector and a standard multivariate normal random vector which are independently distributed.

Moreover, only the distribution of $\breve{\boldsymbol{\mu}}_d$ (and $\breve{\boldsymbol{\mu}}_c$) in both stochastic representations depends on data.

To enhance computational efficiency in applications, we rewrite $\mathbf{S}_d^{-1}$ using the Sherman-Morrison formula (see, for example, Meyer (2000, p. 125))

$$\mathbf{S}_d^{-1} = \mathbf{S}_d^{-1}(\breve{\boldsymbol{\mu}}_d) = \frac{1}{n-1}\mathbf{S}^{-1} - \frac{n}{(n-1)^2}\frac{\mathbf{S}^{-1}(\breve{\boldsymbol{\mu}}_d-\mathbf{a}_1)(\breve{\boldsymbol{\mu}}_d-\mathbf{a}_1)^\top\mathbf{S}^{-1}}{1+\frac{n}{n-1}(\breve{\boldsymbol{\mu}}_d-\mathbf{a}_1)^\top\mathbf{S}^{-1}(\breve{\boldsymbol{\mu}}_d-\mathbf{a}_1)}.$$

Similarly, we obtain that

$$\mathbf{S}_c^{-1} = \mathbf{S}_c^{-1}(\breve{\boldsymbol{\mu}}_c) = \tilde{\mathbf{S}}_c^{-1} - (n+\kappa_c)\frac{\tilde{\mathbf{S}}_c^{-1}(\breve{\boldsymbol{\mu}}_c-\mathbf{a}_{12})(\breve{\boldsymbol{\mu}}_c-\mathbf{a}_{12})^\top\tilde{\mathbf{S}}_c^{-1}}{1+(n+\kappa_c)(\breve{\boldsymbol{\mu}}_c-\mathbf{a}_{12})^\top\tilde{\mathbf{S}}_c^{-1}(\breve{\boldsymbol{\mu}}_c-\mathbf{a}_{12})}.$$

The application of these equalities leads to more computationally efficient stochastic representations of $\boldsymbol{\theta}$ which are stated in Corollary 1.

Corollary 1. *Under the assumptions of Theorem 1, the stochastic representation of* $\boldsymbol{\theta}$

(a) under the diffuse prior $\pi_d(\boldsymbol{\mu},\boldsymbol{\Sigma})$ *is given by*

$$\boldsymbol{\theta}_d \overset{d}{=} \frac{\eta_d}{\alpha}\cdot\mathbf{L}\boldsymbol{\zeta}_d + \frac{\sqrt{\eta_d}}{\alpha}\left(\epsilon_d\cdot\mathbf{L}\boldsymbol{\Upsilon}_d\mathbf{L}^\top - \mathbf{L}\boldsymbol{\zeta}_d\boldsymbol{\zeta}_d^\top\mathbf{L}^\top\right)^{1/2}\mathbf{z}_0, \tag{12}$$

with

$$\begin{aligned}
\epsilon_d &= \epsilon_d(Q_d,\mathbf{U}) = \frac{1}{n-1}\mathbf{a}_1^\top\mathbf{S}^{-1}\mathbf{a}_1 + \frac{2}{\sqrt{n-1}}\frac{\sqrt{\frac{k}{n(n-k)}Q_d}}{1+\frac{k}{n-k}Q_d}\mathbf{a}_1^\top\mathbf{S}^{-1/2}\mathbf{U}\\
&\quad+\frac{\frac{k}{n(n-k)}Q_d}{1+\frac{k}{n-k}Q_d} - \frac{\frac{k}{n-k}Q_d}{1+\frac{k}{n-k}Q_d}\frac{1}{n-1}\left(\mathbf{a}_1^\top\mathbf{S}^{-1/2}\mathbf{U}\right)^2,\\
\boldsymbol{\zeta}_d &= \boldsymbol{\zeta}_d(Q_d,\mathbf{U}) = \frac{1}{n-1}\mathbf{S}^{-1}\mathbf{a}_1 + \frac{\sqrt{\frac{k}{n(n-k)}Q_d}}{1+\frac{k}{n-k}Q_d}\frac{1}{\sqrt{n-1}}\mathbf{S}^{-1/2}\mathbf{U}\\
&\quad-\frac{\frac{k}{n-k}Q_d}{1+\frac{k}{n-k}Q_d}\frac{1}{n-1}\mathbf{S}^{-1/2}\mathbf{U}\mathbf{U}^\top\mathbf{S}^{-1/2}\mathbf{a}_1,\\
\boldsymbol{\Upsilon}_d &= \boldsymbol{\Upsilon}_d(Q_d,\mathbf{U}) = \frac{1}{n-1}\mathbf{S}^{-1} - \frac{\frac{k}{n-k}Q_d}{1+\frac{k}{n-k}Q_d}\frac{1}{n-1}\mathbf{S}^{-1/2}\mathbf{U}\mathbf{U}^\top\mathbf{S}^{-1/2},
\end{aligned}$$

where $\eta_d\sim\chi^2_n$, $\mathbf{z}_0\sim\mathcal{N}_p(\mathbf{0},\mathbf{I}_p)$, $Q_d\sim\mathcal{F}(k,n-k)$, *and* $\mathbf{U}$ *is uniformly distributed on the unit sphere in* $\mathbb{R}^k$*; moreover,* η_d, $\mathbf{z}_0$, Q_d, *and* $\mathbf{U}$ *are mutually independent.*

(b) under the conjugate prior $\pi_c(\boldsymbol{\mu},\boldsymbol{\Sigma})$ *is given by*

$$\boldsymbol{\theta}_c \overset{d}{=} \frac{\eta_c}{\alpha}\cdot\mathbf{L}\boldsymbol{\zeta}_c + \frac{\sqrt{\eta_c}}{\alpha}\left(\epsilon_c\cdot\mathbf{L}\boldsymbol{\Upsilon}_c\mathbf{L}^\top - \mathbf{L}\boldsymbol{\zeta}_c\boldsymbol{\zeta}_c^\top\mathbf{L}^\top\right)^{1/2}\mathbf{z}_0, \tag{13}$$

with

$$\begin{aligned}
\epsilon_c &= \epsilon_c(Q_c,\mathbf{U}) = \mathbf{a}_{12}^\top\tilde{\mathbf{S}}^{-1}\mathbf{a}_{12} + 2\frac{\sqrt{\frac{k}{(\kappa_c+n)(\nu_c+n-2k)}Q_c}}{1+\frac{k}{\nu_c+n-2k}Q_c}\mathbf{a}_{12}^\top\tilde{\mathbf{S}}^{-1/2}\mathbf{U}\\
&+\frac{\frac{k}{(\kappa_c+n)(\nu_c+n-2k)}Q_c}{1+\frac{k}{\nu_c+n-2k}Q_c} - \frac{\frac{k}{\nu_c+n-2k}Q_c}{1+\frac{k}{\nu_c+n-2k}Q_c}\left(\mathbf{a}_{12}^\top\tilde{\mathbf{S}}^{-1/2}\mathbf{U}\right)^2,\\
\boldsymbol{\zeta}_c &= \boldsymbol{\zeta}_c(Q_c,\mathbf{U}) = \tilde{\mathbf{S}}^{-1}\mathbf{a}_{12} + \frac{\sqrt{\frac{k}{(\kappa_c+n)(\nu_c+n-2k)}Q_c}}{1+\frac{k}{\nu_c+n-2k}Q_c}\tilde{\mathbf{S}}^{-1/2}\mathbf{U}\\
&-\frac{\frac{k}{\nu_c+n-2k}Q_c}{1+\frac{k}{\nu_c+n-2k}Q_c}\tilde{\mathbf{S}}^{-1/2}\mathbf{U}\mathbf{U}^\top\tilde{\mathbf{S}}^{-1/2}\mathbf{a}_{12},\\
\boldsymbol{\Upsilon}_c &= \boldsymbol{\Upsilon}_c(Q_c,\mathbf{U}) = \tilde{\mathbf{S}}^{-1} - \frac{\frac{k}{\nu_c+n-2k}Q_c}{1+\frac{k}{\nu_c+n-2k}Q_c}\tilde{\mathbf{S}}^{-1/2}\mathbf{U}\mathbf{U}^\top\tilde{\mathbf{S}}^{-1/2},
\end{aligned}$$

where $\eta_c \sim \chi^2_{\nu_c+n-k}$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0},\mathbf{I}_p)$, $Q_c \sim \mathcal{F}(k,\nu_c+n-2k)$, *and* $\mathbf{U}$ *is uniformly distributed on the unit sphere in* $\mathbb{R}^k$*; moreover,* η_c, $\mathbf{z}_0$, Q_c, *and* $\mathbf{U}$ *are mutually independent.*

In contrast to the stochastic representations given in Theorem 1, the random variables in (12) and (13) do not depend on data. Consequently, the inverses of the matrices $\mathbf{S}$ and $\tilde{\mathbf{S}}$ have to be calculated only once within the simulation study for a given draw. This would surely speed up the generation of realizations of $\boldsymbol{\theta}$. To this end, it has to be noted that the uniform distribution on a unit sphere in $\mathbb{R}^k$ is not a standard distribution in many statistical packages. However, realizations of $\mathbf{U}$ can easily be obtained from the k-dimensional standard normal vector $\mathbf{Z}$ by using $\mathbf{U} = \mathbf{Z}/\sqrt{\mathbf{Z}^\top\mathbf{Z}}$.

In Theorem 2 we derive the analytical expressions of the Bayesian estimates for $\boldsymbol{\theta}$ calculated under the diffuse prior and the conjugate prior. These expressions are derived as posterior means of $\boldsymbol{\theta}$ which can be calculated by using the results of Corollary 1.

Theorem 2. *Under the assumptions of Theorem 1 the Bayesian estimate for* $\mathbf{w}_{TP}$

(a) under the diffuse prior (8) *is given by*

$$\hat{\mathbf{w}}_{TP;d} = E(\mathbf{w}_{TP}|\mathbf{x}) \;=\; \frac{1}{\alpha}\mathbf{S}^{-1}(\overline{\mathbf{x}} - r_f\mathbf{1}_k);$$

(b) under the conjugate prior (9) *is given by*

$$\hat{\mathbf{w}}_{TP;c} = E(\mathbf{w}_{TP}|\mathbf{x}) = \frac{\nu_c+n-k-1}{\alpha}\tilde{\mathbf{S}}^{-1}\mathbf{a}_{12},$$

where $\tilde{\mathbf{S}}$ *and* $\mathbf{a}_{12}$ *are given in Theorem 1.*

The proof is given in the appendix. From Theorem 2 we observe that the point estimator based on the diffuse prior coincides with the classical estimator. This is consistent with our expectations, since the diffuse prior adds no information, but merely reflect the uncertainty. The conjugate prior is an informative prior. This leads to a new point estimator that is obtained in Theorem 2, which reflects the additional information. The structure of the estimator is of shrinkage-type. The mean vector of excess returns is replaced by the weighted sum of the sample mean return and the prior mean. The weights reflect the precision of both sources of information. The covariance matrix is similarly a weighted sum of the sample and prior information. Thus we shrink the sample parameters towards the priors.

The formulas for the covariance matrix of $\mathbf{w}_{TP}$ under both priors are summarized in Theorem 3 whose proof is given in the appendix.

Theorem 3. *Under the assumptions of Theorem 1 the covariance matrix for* $\mathbf{w}_{TP}$

(a) under the diffuse prior (8) *is given by*

$$
\begin{aligned}
Var(\mathbf{w}_{TP}|\mathbf{x}) &= \frac{1}{n-1}\alpha^{-2}\mathbf{S}^{-1}(\overline{\mathbf{x}} - r_f\mathbf{1}_k)(\overline{\mathbf{x}} - r_f\mathbf{1}_k)^\top\mathbf{S}^{-1} \\
+ \quad & \alpha^{-2}\left[\frac{n^2+k-2}{(n-1)n(n+2)} + \frac{n}{(n-1)^2}\frac{k-1}{k}b_d\right]\mathbf{S}^{-1}
\end{aligned}
$$

with $b_d = (\overline{\mathbf{x}} - r_f\mathbf{1}_k)^\top\mathbf{S}^{-1}(\overline{\mathbf{x}} - r_f\mathbf{1}_k)$*;*

(b) under the conjugate prior (9) *is given by*

$$
\begin{aligned}
Var(\mathbf{w}_{TP}|\mathbf{x}) &= (\nu_c + n - k - 1)\alpha^{-2}\tilde{\mathbf{S}}^{-1}\mathbf{a}_{12}\mathbf{a}_{12}^\top\tilde{\mathbf{S}}^{-1} \\
+ \quad & \alpha^{-2}\left[\frac{(\nu_c+n-k)^2+k-2}{(n+\kappa_c)(\nu_c+n-k+2)} + (\nu_c+n-k)\frac{k-1}{k}b_c\right]\tilde{\mathbf{S}}^{-1}
\end{aligned}
$$

with $b_c = \mathbf{a}_{12}^\top\tilde{\mathbf{S}}^{-1}\mathbf{a}_{12}$ *where* $\tilde{\mathbf{S}}$ *and* $\mathbf{a}_{12}$ *are given in Theorem 1.*

Finally, in Theorem 4 we proof that both posterior distributions converge to the same normal distribution as the sample size increases. This results is not surprising and is in line with Bernstein-von-Mises theorem.

Theorem 4. *Under the assumptions of Theorem 1 it holds that*

$$
\sqrt{n}(\mathbf{w}_{TP} - \alpha^{-1}\mathbf{S}^{-1}(\overline{\mathbf{x}} - r_f\mathbf{1}_k))|\mathbf{x} \xrightarrow{d} \mathcal{N}_k(\mathbf{0}, \mathbf{F}) \tag{14}
$$

as $n \to \infty$ *under both the diffuse prior* (8) *and the conjugate prior* (9) *where*

$$
\mathbf{F} = \alpha^{-2}\breve{\mathbf{S}}^{-1}(\breve{\mathbf{x}} - r_f\mathbf{1}_k)(\breve{\mathbf{x}} - r_f\mathbf{1}_k)^\top\breve{\mathbf{S}}^{-1} + \alpha^{-2}\left[1 + \frac{k-1}{k}(\breve{\mathbf{x}} - r_f\mathbf{1}_k)^\top\breve{\mathbf{S}}^{-1}(\breve{\mathbf{x}} - r_f\mathbf{1}_k)\right]\breve{\mathbf{S}}^{-1}
$$

where

$$\breve{\mathbf{x}} = \lim_{n \longrightarrow \infty} \overline{\mathbf{x}} \quad and \quad \breve{\mathbf{S}} = \lim_{n \longrightarrow \infty} \mathbf{S}.$$

The proof of the theorem is given in the appendix. In practice, the asymptotic covariance matrix of $\mathbf{w}_{TP}$ is computed by using $\overline{\mathbf{x}}$ and $\mathbf{S}$ instead of $\breve{\mathbf{x}}$ and $\breve{\mathbf{S}}$.

3 Simulation Study

In this section we assess the performance of the suggested within a simulation study. We compute the coverage probabilities of credible intervals for the portfolio weights based on the diffuse and conjugate priors suggested in the previous section and compare it to the coverage probability stemming from the asymptotic distribution. Since the posterior distribution cannot be determined explicitly, the quantiles are computed via simulations using the respective stochastic representation. The number of repetitions is set to 10000. To speed up the computations we use the representation in Corollary 1.

The setup of the simulation study is as follows. Without loss of generality we restrict the discussion to the first portfolio weight, i.e. $p = 1$, $\mathbf{L} = \mathbf{e}_1^T$. The riskless rate of return is 0.001. The true expected returns $\boldsymbol{\mu}$ are taken as a unform grid of length k between $-0.01 + r_f$ and $0.01 + r_f$. For the covariance matrix we opt for the AR(1)-type structure $\boldsymbol{\Sigma} = (\rho^{|i-j|})_{i,j=1,..,k}$, where ρ takes values between -1 and 1. Since the dimension of the portfolio is of particular interest we consider $k \in \{5, 10, 20, 30\}$. The sample size n is set to 60, which is a typical value in financial literature and corresponds to roughly two months of daily data or a year of weekly data, respectively. In all considered cases we take the following parameters for the conjugate prior $\nu_c = \kappa_c = n/2$. $\boldsymbol{\mu}_c$ is set equal a uniform grid between 0 and 0.003. $\mathbf{S}_c$ is an identity matrix of a corresponding size. Specifically, the boundaries of the credible intervals are computed using the following procedure:

1. Generate independently
   - Diffuse: $\eta_d \sim \chi^2_n$, conjugate: $\eta_c \sim \chi^2_{\nu_c+n-k}$
   - $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$
   - Diffuse: $Q_d \sim \mathcal{F}(k, n-k)$, conjugate: $Q_c \sim \mathcal{F}(k, \nu_c + n - 2k)$
   - $\mathbf{Z} \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I}_k) \to \mathbf{u} = \mathbf{Z}/\sqrt{\mathbf{Z}'\mathbf{Z}}$
2. Compute the vector of weights using (12) for the diffuse prior and (13) for the conjugate prior and using true parameters.
3. Repeat steps (1) and (2) $B = 10000$ times.
4. Compute the credible intervals bounded by the sample quantiles.

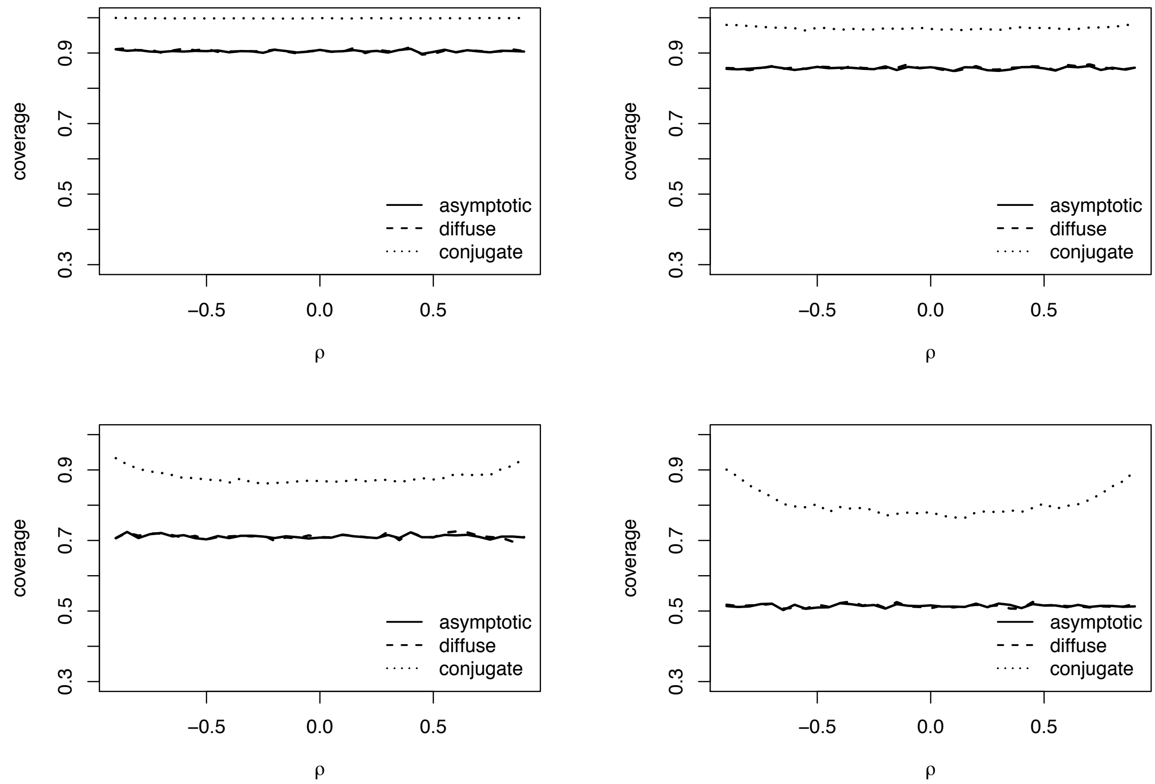

Figure 1: Coverage probabilities for k =5, 10 (top) and k =20, 30 (bottom) and 95% level of significance.

To determine the coverage probabilities we sample the asset returns from the original distribution, estimate the portfolio weights and count the fraction of times the weights are covered by the credible intervals. The results for different dimensions and values of ρ are illustrated in Figure 1. We conclude that the diffuse prior leads to the coverage probabilities almost identical to the ones based on the asymptotic distribution and does not depend on the strength of correlation between the assets. The conjugate prior, however, show much higher coverage probabilities especially at the boundaries of ρ. This is reasonable, since higher ρ values induce covariance matrices which are close to singularity. This leads to wider intervals and higher coverage probabilities. Other forms of the correlation structure or other parameters of the conjugate prior might obviously deteriorate these conclusions.

4 Summary

In this paper we analyze the tangent portfolio within a Bayesian framework. The suggested approach allows us to incorporate uncertainty about the model parameters quantified as prior beliefs of the investors and to incorporate these into the portfolio decisions.

Assuming different priors for the asset returns, we derive the stochastic representation of the posterior distributions of linear combinations of tangent portfolio weights. In particular, we consider non-informative diffuse and informative conjugate priors. Additionally we derive the mean and the variance of the posterior distribution. The results are evaluated within a numerical study, where we assess the coverage probabilities of credible intervals.

5 Appendix

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

Lemma 1. *Assume*

$$\begin{aligned}\Xi|\boldsymbol{\nu},\mathbf{x} &\sim \mathcal{IW}_k(\tau_0,\mathbf{V}_0),\\ \boldsymbol{\nu}|\mathbf{x} &\sim f(\cdot|\mathbf{x}),\end{aligned}$$

where $\mathbf{V}_0=\mathbf{V}_0(\boldsymbol{\nu})$ *and the symbol* $f(\cdot|\mathbf{x})$ *stands for the posterior distribution of* $\boldsymbol{\nu}$*. Let* $\mathbf{M}$ *be a* $p\times k$ *matrix of constants such that rank*$(\mathbf{M})=p\leq k$*. Then the stochastic representation of* $\mathbf{M}\Xi^{-1}\boldsymbol{\nu}$ *is given by*

$$\mathbf{M}\Xi^{-1}\boldsymbol{\nu}\overset{d}{=}\eta\cdot\mathbf{M}\mathbf{V}_0^{-1}\boldsymbol{\nu}+\sqrt{\eta}\left(\boldsymbol{\nu}^\top\mathbf{V}_0^{-1}\boldsymbol{\nu}\cdot\mathbf{M}\mathbf{V}_0^{-1}\mathbf{M}^\top-\mathbf{M}\mathbf{V}_0^{-1}\boldsymbol{\nu}\boldsymbol{\nu}^\top\mathbf{V}_0^{-1}\mathbf{M}^\top\right)^{1/2}\mathbf{z}_0,$$

where $\eta\sim\chi^2_{\tau_0-k-1}$*,* $\mathbf{z}_0\sim\mathcal{N}_p(\mathbf{0},\mathbf{I}_p)$*, and* $\boldsymbol{\nu}|\mathbf{x}\sim f(\cdot|\mathbf{x})$*; moreover,* $\eta,\mathbf{z}_0$ *and* $\boldsymbol{\nu}$ *are mutually independent.*

Proof. From Theorem 3.4.1 of Gupta and Nagar (2000) we obtain that

$$\Xi^{-1}|\boldsymbol{\nu},\mathbf{x}\sim\mathcal{W}_k(\tau_0-k-1,\mathbf{V}_0^{-1}).$$

Next, we fix $\boldsymbol{\nu}=\boldsymbol{\nu}^*$ and define $\tilde{\mathbf{M}}=(\mathbf{M}^\top,\boldsymbol{\nu}^*)^\top$, $\hat{\mathbf{H}}=\tilde{\mathbf{M}}\Xi^{-1}\tilde{\mathbf{M}}^\top=\left\{\hat{\mathbf{H}}_{ij}\right\}_{i,j=1,2}$ with $\hat{\mathbf{H}}_{11}=\mathbf{M}\Xi^{-1}\mathbf{M}^\top$, $\hat{\mathbf{H}}_{12}=\mathbf{M}\Xi^{-1}\boldsymbol{\nu}^*$, $\hat{\mathbf{H}}_{21}=\boldsymbol{\nu}^{*T}\Xi^{-1}\mathbf{M}^\top$, and $\hat{H}_{22}=\boldsymbol{\nu}^{*T}\Xi^{-1}\boldsymbol{\nu}^*$ as well as $\mathbf{H}=\tilde{\mathbf{M}}\tilde{\mathbf{V}}_0^{-1}\tilde{\mathbf{M}}^\top=\{\mathbf{H}_{ij}\}_{i,j=1,2}$ with $\tilde{\mathbf{V}}_0=\mathbf{V}_0(\boldsymbol{\nu}^*)$, $\mathbf{H}_{11}=\mathbf{M}\tilde{\mathbf{V}}^{-1}\mathbf{M}^\top$, $\mathbf{H}_{12}=\mathbf{M}\tilde{\mathbf{V}}^{-1}\boldsymbol{\nu}^*$, $\mathbf{H}_{21}=\boldsymbol{\nu}^{*T}\tilde{\mathbf{V}}^{-1}\mathbf{M}^\top$, and $H_{22}=\boldsymbol{\nu}^{*T}\tilde{\mathbf{V}}^{-1}\boldsymbol{\nu}^*$.

Since

$$\Xi^{-1}|\boldsymbol{\nu}=\boldsymbol{\nu}^*,\mathbf{x}\sim\mathcal{W}_k(\tau_0-k-1,\tilde{\mathbf{V}}^{-1})$$

and rank$(\tilde{\mathbf{M}})=p+1\leq k$, we get from Theorem 3.2.5 in Muirhead (1982) that

$$\hat{\mathbf{H}}|\boldsymbol{\nu}=\boldsymbol{\nu}^*,\mathbf{x}\sim\mathcal{W}_{p+1}(\tau_0-k-1,\mathbf{H}).$$

Moreover, from Theorem 3.2.10 of Muirhead (1982) we obtain that

$$\hat{\mathbf{H}}_{12}|\hat{H}_{22},\boldsymbol{\nu}=\boldsymbol{\nu}^*,\mathbf{x}\sim\mathcal{N}_p(\mathbf{H}_{12}H_{22}^{-1}\hat{H}_{22},\mathbf{H}_{11\cdot2}\hat{H}_{22}),$$

where $\mathbf{H}_{11\cdot 2} = \mathbf{H}_{11} - \mathbf{H}_{12}H_{22}^{-1}\mathbf{H}_{21}$ is the Schur complement.

Let $\eta = \hat{H}_{22}/H_{22}$. Then the application of Theorem 3.2.8 of Muirhead (1982) leads to

$$\eta|\boldsymbol{\nu} = \boldsymbol{\nu}^*, \mathbf{x} \sim \chi^2_{\tau_0-k-1}.$$

Moreover, since the conditional distribution of η given $\boldsymbol{\nu} = \boldsymbol{\nu}^*$ and $\mathbf{x}$ does not depend on $\boldsymbol{\nu}^*$ and $\mathbf{x}$, it is also the unconditional one as well as η and $\boldsymbol{\nu}$ are independent, i.e. $\eta \sim \chi^2_{\tau_0-k-1}$. Thus, the stochastic representation of $\mathbf{M}\boldsymbol{\Xi}^{-1}\boldsymbol{\nu}$ is given by

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

where $\eta \sim \chi^2_{\tau_0-k-1}$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$, and $\boldsymbol{\nu}|\mathbf{x} \sim f(\cdot|\mathbf{x})$; moreover, $\eta, \mathbf{z}_0$ and $\boldsymbol{\nu}$ are mutually independent. This completes the proof of the lemma. □

Proof of Theorem 1. a) Using the expression of the likelihood function (7) and the diffuse prior $\pi_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ as in (8), the posterior distribution of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is given by

$$\begin{aligned} \pi_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{x}) &\propto L(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\pi_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &\propto |\boldsymbol{\Sigma}|^{-(n+k+1)/2} \exp\left\{-\frac{n}{2}(\overline{\mathbf{x}} - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu}) - \frac{n-1}{2}\mathrm{tr}[\mathbf{S}\boldsymbol{\Sigma}^{-1}]\right\}. \end{aligned} \quad (15)$$

Integrating out $\boldsymbol{\Sigma}$ we obtain the marginal posterior for $\boldsymbol{\mu}$ expressed as

$$\begin{aligned} \pi_d(\boldsymbol{\mu}|\mathbf{x}) &\propto \int_{\boldsymbol{\Sigma}>0} |\boldsymbol{\Sigma}|^{-(n+k+1)/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[(n(\overline{\mathbf{x}} - \boldsymbol{\mu})(\overline{\mathbf{x}} - \boldsymbol{\mu})^\top + (n-1)\mathbf{S})\boldsymbol{\Sigma}^{-1}\right]\right\} d\boldsymbol{\Sigma} \\ &\propto |n(\overline{\mathbf{x}} - \boldsymbol{\mu})(\overline{\mathbf{x}} - \boldsymbol{\mu})^\top + (n-1)\mathbf{S}|^{-\frac{n}{2}}, \end{aligned}$$

where the last equality follows by observing that the function under the integral is the density function of the inverse Wishart distribution with $n+k+1$ degrees of freedom and parameter matrix $n(\overline{\mathbf{x}}-\boldsymbol{\mu})(\overline{\mathbf{x}}-\boldsymbol{\mu})^\top + (n-1)\mathbf{S}$. The application of Silvester's determinant theorem leads to

$$\pi_d(\boldsymbol{\mu}|\mathbf{x}) \propto \left(1 + \frac{n}{n-1}(\overline{\mathbf{x}} - \boldsymbol{\mu})^\top\mathbf{S}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu})\right)^{-\frac{n}{2}}$$

which shows that $\boldsymbol{\mu}|\mathbf{x} \sim t_k\left(n-k, \overline{\mathbf{x}}, \frac{n-1}{n(n-k)}\mathbf{S}\right)$. Using the properties of the multivariate t-distribution we then get with $\breve{\boldsymbol{\mu}}_d = (\boldsymbol{\mu} - r_f\mathbf{1}_k)$ that

$$\breve{\boldsymbol{\mu}}_d|\mathbf{x} \sim t_k\left(n-k, \mathbf{a}_1, \frac{n-1}{n(n-k)}\mathbf{S}\right) \quad \text{with} \quad \mathbf{a}_1 = \overline{\mathbf{x}} - r_f\mathbf{1}_k.$$

Furthermore, from (15) we obtained that $\boldsymbol{\Sigma}|\breve{\boldsymbol{\mu}}_d, \mathbf{x} \sim \mathcal{IW}_k(n+k+1, \mathbf{S}_d)$ with $\mathbf{S}_d = \mathbf{S}_d(\breve{\boldsymbol{\mu}}) = (n-1)\mathbf{S} + n(\breve{\boldsymbol{\mu}}_d - \mathbf{a}_1)(\breve{\boldsymbol{\mu}}_d - \mathbf{a}_1)^\top$.

Finally, the application of Lemma 1 with $\tau_0 = n + k + 1$ and $\mathbf{V}_0 = \mathbf{S}_d$ leads to

$$\boldsymbol{\theta} \stackrel{d}{=} \eta_d \cdot \mathbf{L}\mathbf{S}_d^{-1}\breve{\boldsymbol{\mu}}_d + \sqrt{\eta}\left(\breve{\boldsymbol{\mu}}_d^\top \mathbf{S}_d^{-1}\breve{\boldsymbol{\mu}}_d \cdot \mathbf{L}\mathbf{S}_d^{-1}\mathbf{L}^\top - \mathbf{L}\mathbf{S}_d^{-1}\breve{\boldsymbol{\mu}}_d\breve{\boldsymbol{\mu}}_d^\top \mathbf{S}_d^{-1}\mathbf{L}^\top\right)^{1/2}\mathbf{z}_0,$$

where $\eta_d \sim \chi^2_n$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$, and $\breve{\boldsymbol{\mu}}_d|\mathbf{x} \sim t_k\left(n-k, \mathbf{a}_1, \frac{n-1}{n(n-k)}\mathbf{S}\right)$; moreover, $\eta, \mathbf{z}_0$ and $\breve{\boldsymbol{\mu}}_d$ are mutually independent.

b) The joint posterior for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ under the conjugate prior (9) is given by

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

where $\breve{\boldsymbol{\mu}}_c = \boldsymbol{\mu} - r_f\mathbf{1}_k$, $\mathbf{a}_1$ and $\mathbf{a}_2$ as in the statement of the theorem and

$$\begin{aligned}
\mathbf{S}_c &= \mathbf{S}_c(\breve{\boldsymbol{\mu}}_c) = (n-1)\mathbf{S} + \mathbf{V}_c + n(\breve{\boldsymbol{\mu}}_c - \mathbf{a}_1)(\breve{\boldsymbol{\mu}}_c - \mathbf{a}_1)^\top + \kappa_c(\breve{\boldsymbol{\mu}}_c - \mathbf{a}_2)(\breve{\boldsymbol{\mu}}_c - \mathbf{a}_2)^\top,\\
&= \tilde{\mathbf{S}} + (n+\kappa_c)\left[\breve{\boldsymbol{\mu}}_c - \mathbf{a}_{12}\right]\left[\breve{\boldsymbol{\mu}}_c - \mathbf{a}_{12}\right]^\top,
\end{aligned}$$

with $\mathbf{a}_{12}$ given in the statement of the theorem and

$$\tilde{\mathbf{S}} = (n-1)\mathbf{S} + \mathbf{V}_c - (n+\kappa_c)\mathbf{a}_{12}\mathbf{a}_{12}^\top + (n\mathbf{a}_1\mathbf{a}_1^\top + \kappa_c\mathbf{a}_2\mathbf{a}_2^\top).$$

Following the proof of part a) of the theorem we get

$$\begin{aligned}
\boldsymbol{\Sigma}|\breve{\boldsymbol{\mu}}_c, \mathbf{x} &\sim \mathcal{IW}_k(\nu_c + n + 1, \mathbf{S}_c),\\
\breve{\boldsymbol{\mu}}_c &\sim t_k\left(\nu_c + n - 2k, \mathbf{a}_{12}, \frac{1}{(n+\kappa_c)(\nu_c+n-2k)}\tilde{\mathbf{S}}\right).
\end{aligned}$$

Finally, the application of Lemma 1 with $\tau_0 = \nu_c + n + 1$ and $\mathbf{V}_0 = \mathbf{S}_c$ leads to the statement of the theorem. □

In the proof of Corollary 1 we use the following lemma.

Lemma 2. *Assume that the stochastic representation of* $\mathbf{M}\boldsymbol{\Xi}^{-1}\boldsymbol{\nu}$ *is given by*

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

with $\mathbf{V}_0 = \mathbf{V}_0(\boldsymbol{\nu}) = \mathbf{S}_0 + n_0(\boldsymbol{\nu} - \mathbf{b}_0)(\boldsymbol{\nu} - \mathbf{b}_0)^\top$ *and* $\mathbf{M}$ *a* $p \times k$ *matrix of constants such that* rank$(\mathbf{M}) = p \leq k$ *where* $\eta \sim \chi^2_{\tau_0}$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$, *and* $\boldsymbol{\nu}|\mathbf{x} \sim t_k(d_0, \mathbf{b}_0, \lambda_0\mathbf{S}_0)$*; moreover,* $\eta, \mathbf{z}_0$ *and* $\boldsymbol{\nu}$ *are mutually independent. Then*

$$\mathbf{M}\boldsymbol{\Xi}^{-1}\boldsymbol{\nu} \quad \stackrel{d}{=} \quad \eta \cdot \mathbf{M}\boldsymbol{\zeta} + \sqrt{\eta}\left(\epsilon \cdot \mathbf{M}\boldsymbol{\Upsilon}\mathbf{M}^\top - \mathbf{M}\boldsymbol{\zeta}\boldsymbol{\zeta}^\top\mathbf{M}^\top\right)^{1/2}\mathbf{z}_0$$

with

$$\begin{aligned}
\epsilon &= \epsilon(Q,\mathbf{U}) = \mathbf{b}_0^\top \mathbf{S}_0^{-1}\mathbf{b}_0 + 2\frac{\sqrt{\lambda_0 kQ}}{1+n_0\lambda_0 kQ}\mathbf{b}_0^\top \mathbf{S}_0^{-1/2}\mathbf{U} + \frac{\lambda_0 kQ}{1+n_0\lambda_0 kQ} - \frac{n_0\lambda_0 kQ}{1+n_0\lambda_0 kQ}\left(\mathbf{b}_0^\top \mathbf{S}_0^{-1/2}\mathbf{U}\right)^2,\\
\boldsymbol{\zeta} &= \boldsymbol{\zeta}(Q,\mathbf{U}) = \mathbf{S}_0^{-1}\mathbf{b}_0 + \frac{\sqrt{\lambda_0 kQ}}{1+n_0\lambda_0 kQ}\mathbf{S}_0^{-1/2}\mathbf{U} - \frac{n_0\lambda_0 kQ}{1+n_0\lambda_0 kQ}\mathbf{S}_0^{-1/2}\mathbf{U}\mathbf{U}^\top \mathbf{S}_0^{-1/2}\mathbf{b}_0,\\
\boldsymbol{\Upsilon} &= \boldsymbol{\Upsilon}(Q,\mathbf{U}) = \mathbf{S}_0^{-1} - \frac{n_0\lambda_0 kQ}{1+n_0\lambda_0 kQ}\mathbf{S}_0^{-1/2}\mathbf{U}\mathbf{U}^\top \mathbf{S}_0^{-1/2},
\end{aligned}$$

where $\eta \sim \chi^2_{\tau_0}$*,* $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0},\mathbf{I}_p)$*,* $Q \sim \mathcal{F}(k,d_0)$*, and* $\mathbf{U}$ *uniformly distributed on the unit sphere in* $\mathbb{R}^k$*; moreover,* η*,* $\mathbf{z}_0$*,* Q*, and* $\mathbf{U}$ *are mutually independent.*

Proof. Using the Sherman-Morrison formula (see p.125 of Meyer (2000)) we obtain

$$\mathbf{V}_0^{-1} = \mathbf{S}_0^{-1} - n_0\frac{\mathbf{S}_0^{-1}(\boldsymbol{\nu}-\mathbf{b}_0)(\boldsymbol{\nu}-\mathbf{b}_0)^\top \mathbf{S}_0^{-1}}{1+n_0(\boldsymbol{\nu}-\mathbf{b}_0)^\top \mathbf{S}_0^{-1}(\boldsymbol{\nu}-\mathbf{b}_0)} \tag{16}$$

Let

$$\mathbf{U} = \frac{\mathbf{S}_0^{-1/2}(\boldsymbol{\nu}-\mathbf{b}_0)}{\sqrt{(\boldsymbol{\nu}-\mathbf{b}_0)^\top \mathbf{S}_0^{-1}(\boldsymbol{\nu}-\mathbf{b}_0)}} \quad \text{and} \quad Q = \lambda_0^{-1}(\boldsymbol{\nu}-\mathbf{b}_0)^\top \mathbf{S}_0^{-1}(\boldsymbol{\nu}-\mathbf{b}_0)/k. \tag{17}$$

Using the facts that $\boldsymbol{\nu}|\mathbf{x} \sim t_k(d_0,\mathbf{b}_0,\lambda_0\mathbf{S}_0)$ 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 Gupta et al. (2013)). Moreover, from the properties of the multivariate t-distribution (see p. 19 of Kotz and Nadarajah (2004)), we get that $Q \sim \mathcal{F}(k,d_0)$, i.e., Q has an $\mathcal{F}$-distribution with k and d_0 degrees of freedom.

Hence, the application of the (16) and (17) leads to

$$\begin{aligned}
\mathbf{V}_0^{-1} &= \mathbf{S}_0^{-1} - \frac{n_0\lambda_0 kQ}{1+n_0\lambda_0 kQ}\mathbf{S}_0^{-1/2}\mathbf{U}\mathbf{U}^\top \mathbf{S}_0^{-1/2},\\
\mathbf{V}_0^{-1}\boldsymbol{\nu} &= \mathbf{S}_0^{-1}\boldsymbol{\nu} - n_0\frac{\mathbf{S}_0^{-1}(\boldsymbol{\nu}-\mathbf{b}_0)(\boldsymbol{\nu}-\mathbf{b}_0)^\top \mathbf{S}_0^{-1}(\boldsymbol{\nu}-\mathbf{b}_0+\mathbf{b}_0)}{1+n_0(\boldsymbol{\nu}-\mathbf{b}_0)^\top \mathbf{S}_0^{-1}(\boldsymbol{\nu}-\mathbf{b}_0)}\\
&= \mathbf{S}_0^{-1}\mathbf{b}_0 + \frac{\mathbf{S}_0^{-1}(\boldsymbol{\nu}-\mathbf{b}_0)}{1+n_0(\boldsymbol{\nu}-\mathbf{b}_0)^\top \mathbf{S}_0^{-1}(\boldsymbol{\nu}-\mathbf{b}_0)} - n_0\frac{\mathbf{S}_0^{-1}(\boldsymbol{\nu}-\mathbf{b}_0)(\boldsymbol{\nu}-\mathbf{b}_0)^\top \mathbf{S}_0^{-1}\mathbf{b}_0}{1+n_0(\boldsymbol{\nu}-\mathbf{b}_0)^\top \mathbf{S}_0^{-1}(\boldsymbol{\nu}-\mathbf{b}_0)}\\
&= \mathbf{S}_0^{-1}\mathbf{b}_0 + \frac{\sqrt{\lambda_0 kQ}}{1+n_0\lambda_0 kQ}\mathbf{S}_0^{-1/2}\mathbf{U} - \frac{n_0\lambda_0 kQ}{1+n_0\lambda_0 kQ}\mathbf{S}_0^{-1/2}\mathbf{U}\mathbf{U}^\top \mathbf{S}_0^{-1/2}\mathbf{b}_0,
\end{aligned}$$

and

$$\boldsymbol{\nu}^\top \mathbf{V}_0^{-1}\boldsymbol{\nu} = \mathbf{b}_0^\top \mathbf{S}_0^{-1}\mathbf{b}_0 + 2\frac{\mathbf{b}_0^\top \mathbf{S}_0^{-1/2}\mathbf{U}\sqrt{\lambda_0 kQ}}{1+n_0\lambda_0 kQ} + \frac{\lambda_0 kQ}{1+n_0\lambda_0 kQ} - \frac{n_0\lambda_0 kQ}{1+n_0\lambda_0 kQ}\left(\mathbf{b}_0^\top \mathbf{S}_0^{-1/2}\mathbf{U}\right)^2.$$

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

Proof of Corollary 1. The statement of the corollary follows directly from Lemma 2 with

$\tau_0 = n$, $n_0 = n$, $d_0 = n-k$, $\lambda_0 = \frac{1}{n(n-k)}$, $\mathbf{b}_0 = \mathbf{a}_1$, $\mathbf{S}_0 = (n-1)\mathbf{S}$, $\mathbf{M} = \frac{1}{\alpha}\mathbf{L}$ in the case of the diffuse prior and with $\tau_0 = \nu_c + n - k$, $n_0 = n + \kappa_c$, $d_0 = \nu_c + n - 2k$, $\lambda_0 = \frac{1}{(n+\kappa_c)(\nu_c+n-2k)}$, $\mathbf{b}_0 = \mathbf{a}_{12}$, $\mathbf{S}_0 = \tilde{\mathbf{S}}$, $\mathbf{M} = \frac{1}{\alpha}\mathbf{L}$ for the conjugate prior. □

Lemma 3. *Under the assumption of Lemma 2 with $n_0\lambda_0 = 1/d_0$ we get that*

$$E(\mathbf{M}\mathbf{\Xi}^{-1}\boldsymbol{\nu}|\mathbf{x}) = \tau_0\left(1 - \frac{1}{k+d_0}\right)\mathbf{M}^\top\mathbf{S}_0^{-1}\mathbf{b}_0.$$

Proof. Since η, $\mathbf{z}_0$, Q, and $\mathbf{U}$ are independent with $E(\mathbf{z}_0) = \mathbf{0}$, $E(\mathbf{U}) = \mathbf{0}$ and $E(\mathbf{U}\mathbf{U}^\top) = \frac{1}{k}\mathbf{I_k}$, we get

$$E(\mathbf{M}\mathbf{\Xi}^{-1}\boldsymbol{\nu}|\mathbf{x}) = E(\eta\mathbf{M}\boldsymbol{\zeta}|\mathbf{x}) = \tau_0\left(\mathbf{M}\mathbf{S}_0^{-1}\mathbf{b}_0 - E\left(\frac{n_0\lambda_0 kQ}{1+n_0\lambda_0 kQ}\right)\frac{1}{k}\mathbf{M}\mathbf{S}_0^{-1}\mathbf{b}_0\right).$$

Since $Q \sim \mathcal{F}(k, d_0)$, then from the properties of the $\mathcal{F}$-distribution, we obtain that

$$\frac{\frac{k}{d_0}Q}{1+\frac{k}{d_0}Q} \sim Beta\left(\frac{k}{2}, \frac{d_0}{2}\right).$$

Hence, $E\left(\frac{\frac{k}{d_0}Q}{1+\frac{k}{d_0}Q}\right) = \frac{k}{k+d_0}$ which leads to

$$E(\mathbf{M}\mathbf{\Xi}^{-1}\boldsymbol{\nu}|\mathbf{x}) = \tau_0\left(1 - \frac{1}{k+d_0}\right)\mathbf{M}\mathbf{S}_0^{-1}\mathbf{b}_0.$$

□

Proof of Theorem 2. The application of Lemma 3 with $\tau_0 = n$, $d_0 = n-k$, $\mathbf{b}_0 = (\overline{\mathbf{x}} - r_f\mathbf{1}_k)$, $\mathbf{S}_0 = (n-1)\mathbf{S}$, $\mathbf{M} = \frac{1}{\alpha}\mathbf{l}^\top$ for the diffuse prior and with $\tau_0 = \nu_c + n - k$, $d_0 = \nu_c + n - 2k$, $\mathbf{b}_0 = \mathbf{a}_{12}$, $\mathbf{S}_0 = \tilde{\mathbf{S}}$, $\mathbf{M} = \frac{1}{\alpha}\mathbf{l}^\top$ in the case of the conjugate prior for an arbitrary vector $\mathbf{l}$ leads to

$$E(\mathbf{l}^\top\mathbf{w}_{TP}|\mathbf{x}) \;=\; \frac{1}{\alpha}\mathbf{l}^\top\mathbf{S}^{-1}(\overline{\mathbf{x}} - r_f\mathbf{1}_k)$$

and

$$E(\mathbf{l}^\top\mathbf{w}_{TP}|\mathbf{x}) = \frac{\nu_c + n - k - 1}{\alpha}\mathbf{l}^\top\tilde{\mathbf{S}}_c^{-1}\mathbf{a}_{12},$$

respectively. Since the vector $\mathbf{l}$ is arbitrary chosen, we get the statement of the theorem. □

In the proof of Theorem 3 we use the following lemma.

Lemma 4. *Under the assumption of Lemma 2 with* $n_0\lambda_0 = 1/d_0$ *and* $\mathbf{M} = \mathbf{m}^\top : 1 \times k$, *we get that*

$$\begin{aligned}
Var(\mathbf{m}^\top \boldsymbol{\Xi}^{-1}\boldsymbol{\nu}|\mathbf{x}) &= \tau_0(1+\tau_0)\left[\left(1 - \frac{2}{k+d_0} + \frac{2}{(k+d_0)(k+d_0+2)}\right)c_{12}^2\right.\\
&+ \left.\left(\frac{d_0}{n_0(k+d_0)(k+d_0+2)} + \frac{1}{(k+d_0)(k+d_0+2)}c_2\right)c_1\right]\\
&+ \tau_0\left[\left(\frac{k-1}{n_0(k+d_0)} + \left(1 - \frac{1}{k} - \frac{1}{k+d_0} + \frac{1}{(k+d_0)(k+d_0+2)}\right)c_2\right)c_1\right.\\
&+ \left.\frac{2}{(k+d_0)(k+d_0+2)}c_{12}^2\right] - \tau_0^2\left(1 - \frac{1}{k+d_0}\right)^2 c_{12}^2,
\end{aligned}$$

where $c_1 = \mathbf{m}^\top \mathbf{S}_0^{-1}\mathbf{m}$, $c_2 = \mathbf{b}_0^\top \mathbf{S}_0^{-1}\mathbf{b}_0$, *and* $c_{12} = \mathbf{m}^\top \mathbf{S}_0^{-1}\mathbf{b}_0$.

Proof. It holds that

$$Var(\mathbf{m}\boldsymbol{\Xi}^{-1}\boldsymbol{\nu}|\mathbf{x}) = E(\mathbf{m}^\top \boldsymbol{\Xi}^{-1}\boldsymbol{\nu}\boldsymbol{\nu}^\top \boldsymbol{\Xi}^{-1}\mathbf{m}|\mathbf{x}) - E(\mathbf{m}^\top \boldsymbol{\Xi}^{-1}\boldsymbol{\nu}|\mathbf{x})^2,$$

where $E(\mathbf{m}\boldsymbol{\Xi}^{-1}\boldsymbol{\nu}|\mathbf{x})$ is given in Lemma 3.

The application of Lemma 2 together with $E(\mathbf{z}_0) = \mathbf{0}$, $E(\mathbf{z}_0\mathbf{z}_0^\top) = \mathbf{I}_p$ and the independence of η, $\mathbf{z}_0$, Q, and $\mathbf{U}$ leads to

$$\begin{aligned}
E(\mathbf{m}^\top \boldsymbol{\Xi}^{-1}\boldsymbol{\nu}\boldsymbol{\nu}^\top \boldsymbol{\Xi}^{-1}\mathbf{m}|\mathbf{x}) &= E(\eta^2)E(\mathbf{m}^\top \boldsymbol{\zeta}\boldsymbol{\zeta}^\top \mathbf{m}|\mathbf{x}) + E(\eta)E\left(\epsilon \mathbf{m}^\top \boldsymbol{\Upsilon}\mathbf{m} - \mathbf{m}^\top \boldsymbol{\zeta}\boldsymbol{\zeta}^\top \mathbf{m}|\mathbf{x}\right)\\
&= \tau_0(1+\tau_0)E(\mathbf{m}^\top \boldsymbol{\zeta}\boldsymbol{\zeta}^\top \mathbf{m}|\mathbf{x}) + \tau_0 E\left(\epsilon \mathbf{m}^\top \boldsymbol{\Upsilon}\mathbf{m}|\mathbf{x}\right),
\end{aligned}$$

where we use that $E(\eta) = \tau_0$ and $E(\eta^2) - E(\eta) = \tau_0(1+\tau_0)$.

Using that $E(\mathbf{U}\mathbf{U}^\top) = \frac{1}{k}\mathbf{I}_k$ and all odd mixed moments of the elements of $\mathbf{U}$ are zero, we get

$$\begin{aligned}
E(\mathbf{m}^\top \boldsymbol{\zeta}\boldsymbol{\zeta}^\top \mathbf{m}|\mathbf{x}) &= (\mathbf{m}^\top \mathbf{S}_0^{-1}\mathbf{b}_0)^2 + \frac{1}{k}E\left(\frac{\lambda_0 kQ}{(1+n_0\lambda_0 kQ)^2}\right)\mathbf{m}^\top \mathbf{S}_0^{-1}\mathbf{m}\\
&- \frac{2}{k}E\left(\frac{n_0\lambda_0 kQ}{1+n_0\lambda_0 kQ}\right)(\mathbf{m}\mathbf{S}_0^{-1}\mathbf{b}_0)^2\\
&+ E\left(\left(\frac{n_0\lambda_0 kQ}{1+n_0\lambda_0 kQ}\right)^2\right)E\left((\mathbf{m}\mathbf{S}_0^{-1/2}\mathbf{U})^2(\mathbf{b}_0^\top \mathbf{S}_0^{-1/2}\mathbf{U})^2|\mathbf{x}\right)
\end{aligned}$$

and

$$
\begin{aligned}
E\left(\epsilon \mathbf{m}^\top \boldsymbol{\Upsilon} \mathbf{m}|\mathbf{x}\right) &= \mathbf{b}_0^\top \mathbf{S}_0^{-1}\mathbf{b}_0 \mathbf{m}^\top \mathbf{S}_0^{-1}\mathbf{m} + E\left(\frac{\lambda_0 k Q}{1+n_0\lambda_0 k Q}\right)\mathbf{m}^\top \mathbf{S}_0^{-1}\mathbf{m} \\
&- \frac{1}{k}E\left(\frac{n_0\lambda_0 k Q}{1+n_0\lambda_0 k Q}\right)\mathbf{b}_0^\top \mathbf{S}_0^{-1}\mathbf{b}_0 \mathbf{m}^\top \mathbf{S}_0^{-1}\mathbf{m} \\
&- \frac{1}{k}\mathbf{b}_0^\top \mathbf{S}_0^{-1}\mathbf{b}_0 \mathbf{m}^\top \mathbf{S}_0^{-1}\mathbf{m} - \frac{1}{k}E\left(\frac{\lambda_0 k Q}{1+n_0\lambda_0 k Q}\right)\mathbf{m}^\top \mathbf{S}_0^{-1}\mathbf{m} \\
&+ E\left(\left(\frac{n_0\lambda_0 k Q}{1+n_0\lambda_0 k Q}\right)^2\right) E\left((\mathbf{m}\mathbf{S}_0^{-1/2}\mathbf{U})^2(\mathbf{b}_0^\top \mathbf{S}_0^{-1/2}\mathbf{U})^2|\mathbf{x}\right).
\end{aligned}
$$

Since $\frac{n_0\lambda_0 kQ}{1+n_0\lambda_0 kQ}$ has a beta distribution with $k/2$ and $d_0/2$ degrees of freedom (see the end of the proof of Lemma 3), we obtain

$$
\begin{aligned}
E\left(\frac{n_0\lambda_0 k Q}{1+n_0\lambda_0 k Q}\right) &= \frac{k}{k+d_0}, \\
E\left(\frac{n_0\lambda_0 k Q}{1+n_0\lambda_0 k Q}\right)^2 &= \frac{2kd_0 + k^2(k+d_0+2)}{(k+d_0)^2(k+d_0+2)} = \frac{k(k+2)}{(k+d_0)(k+d_0+2)}.
\end{aligned}
$$

Furthermore, using $Q \sim \mathcal{F}(k, d_0)$, we get

$$
\begin{aligned}
E\left[\frac{\lambda_0 k Q}{(1+n_0\lambda_0 k Q)^2}\right] &= \frac{1}{n_0}\int_0^\infty \frac{kt/d_0}{(1+kt/d_0)^2}\frac{1}{B\left(\frac{k}{2},\frac{d_0}{2}\right)}\left(\frac{k}{d_0}\right)^{k/2} t^{k/2-1}\left(1+\frac{k}{d_0}t\right)^{-(k+d_0)/2} dt \\
&= \frac{1}{n_0}\frac{1}{B\left(\frac{k}{2},\frac{d_0}{2}\right)}\int_0^\infty \left(\frac{k}{d_0}\right)^{(k+2)/2} t^{(k+2)/2-1}\left(1+\frac{k}{d_0}t\right)^{-(k+d_0+4)/2} dt \\
&= \frac{1}{n_0}\frac{B\left(\frac{k+2}{2},\frac{d_0+2}{2}\right)}{B\left(\frac{k}{2},\frac{d_0}{2}\right)} = \frac{kd_0}{n_0(k+d_0)(k+d_0+2)},
\end{aligned}
$$

where $B(\cdot,\cdot)$ stands for the beta function (see, Mathai and Provost (1992, p. 256)).

Let $Q_N \sim \chi^2_k$ be independent of $\mathbf{U}$. Then $\sqrt{Q_N}\mathbf{U}$ has a multivariate standard normal distribution, i.e.

$$
\begin{pmatrix} \mathbf{m}^\top \mathbf{S}_0^{-1/2} \\ \mathbf{b}_0^\top \mathbf{S}_0^{-1/2} \end{pmatrix} \sqrt{Q_N}\mathbf{U}|\mathbf{x} \sim \mathcal{N}_2\left(\mathbf{0}, \begin{pmatrix} \mathbf{m}^\top \mathbf{S}_0^{-1}\mathbf{m} & \mathbf{m}^\top \mathbf{S}_0^{-1}\mathbf{b}_0 \\ \mathbf{b}_0^\top \mathbf{S}_0^{-1}\mathbf{m} & \mathbf{b}_0^\top \mathbf{S}_0^{-1}\mathbf{b}_0 \end{pmatrix}\right) = \mathcal{N}_2\left(\mathbf{0}, \begin{pmatrix} c_1 & c_{12} \\ c_{12} & c_2 \end{pmatrix}\right),
$$

where c_1, c_2, and c_{12} are defined in the statement of Lemma 4. Hence,

$$
\begin{aligned}
& E\left((\mathbf{m}\mathbf{S}_0^{-1/2}\mathbf{U})^2(\mathbf{b}_0^\top \mathbf{S}_0^{-1/2}\mathbf{U})^2|\mathbf{x}\right) = E\left[\left(\mathbf{m}^\top \mathbf{S}_0^{-1/2}\mathbf{U}\right)^2\left(\mathbf{b}_0^\top \mathbf{S}_0^{-1/2}\mathbf{U}\right)^2|\mathbf{x}\right]\frac{E(Q_N^2)}{E(Q_N^2)} \\
= \; & \frac{E\left[\left(\mathbf{m}^\top \mathbf{S}_0^{-1/2}\sqrt{Q_N}\mathbf{U}\right)^2\left(\mathbf{b}_0^\top \sqrt{Q_N}\mathbf{S}_0^{-1/2}\mathbf{U}\right)^2|\mathbf{x}\right]}{E(Q_N^2)} = \frac{c_1c_2+2c_{12}^2}{k(k+2)},
\end{aligned}
$$

where the last equality follows from the Isserlis' theorem (c.f., Isserlis (1918)).

Thus, we get

$$\begin{aligned} E(\mathbf{m}^\top \boldsymbol{\zeta}\boldsymbol{\zeta}^\top \mathbf{m}|\mathbf{x}) &= c_{12}^2 + \frac{1}{k}\frac{kd_0}{n_0(k+d_0)(k+d_0+2)}c_1 \\ &- \frac{2}{k}\frac{k}{k+d_0}c_{12}^2 + \frac{k(k+2)}{(k+d_0)(k+d_0+2)}\frac{c_1c_2+2c_{12}^2}{k(k+2)} \\ &= \left(1-\frac{2}{k+d_0}+\frac{2}{(k+d_0)(k+d_0+2)}\right)c_{12}^2 \\ &+ \left(\frac{d_0}{n_0(k+d_0)(k+d_0+2)}+\frac{1}{(k+d_0)(k+d_0+2)}c_2\right)c_1 \end{aligned}$$

and

$$\begin{aligned} &E\left(\epsilon \mathbf{m}^\top \boldsymbol{\Upsilon}\mathbf{m}|\mathbf{x}\right) = c_1c_2 + \frac{1}{n_0}\frac{k}{k+d_0}c_1 - \frac{1}{k}\frac{k}{k+d_0}c_1c_2 \\ -& \frac{1}{k}c_1c_2 - \frac{1}{k}\frac{1}{n_0}\frac{k}{k+d_0}c_1 + \frac{k(k+2)}{(k+d_0)(k+d_0+2)}\frac{c_1c_2+2c_{12}^2}{k(k+2)} \\ =& \frac{2}{(k+d_0)(k+d_0+2)}c_{12}^2 + \left(\frac{k-1}{n_0(k+d_0)}+\left(1-\frac{1}{k}-\frac{1}{k+d_0}+\frac{1}{(k+d_0)(k+d_0+2)}\right)c_2\right)c_1 \,. \end{aligned}$$

□

Proof of Theorem 3. For the fixed arbitrary chosen vector $\mathbf{l}$ we apply the results of Lemma 4 with $\tau_0 = n$, $n_0 = n$, $d_0 = n-k$, $\mathbf{b}_0 = (\overline{\mathbf{x}} - r_f \mathbf{1}_k)$, $\mathbf{S}_0 = (n-1)\mathbf{S}$, $\mathbf{m} = \frac{1}{\alpha}\mathbf{l}$ for the diffuse prior and with $\tau_0 = \nu_c + n - k$,$n_0 = n + \kappa_c$, $d_0 = \nu_c + n - 2k$, $\mathbf{b}_0 = \mathbf{a}_{12}$, $\mathbf{S}_0 = \tilde{\mathbf{S}}$, $\mathbf{m} = \frac{1}{\alpha}\mathbf{l}$ in the case of the conjugate prior. This leads to

$$\begin{aligned} &Var(\mathbf{l}^\top \mathbf{w}_{TP}|\mathbf{x}) = \frac{n(n+1)}{(n-1)^2}\left[\left(1-\frac{2}{n}+\frac{2}{n(n+2)}\right)c_{12}^2 + \left(\frac{(n-k)(n-1)}{n^2(n+2)}+\frac{1}{n(n+2)}c_2\right)c_1\right] \\ +& \frac{n}{(n-1)^2}\left[\left(\frac{(k-1)(n-1)}{n^2}+\left(1-\frac{1}{k}-\frac{1}{n}+\frac{1}{n(n+2)}\right)c_2\right)c_1 + \frac{2}{n(n+2)}c_{12}^2\right] - c_{12}^2 \\ =& \frac{1}{n-1}c_{12}^2 + \left[\frac{n^2+k-2}{(n-1)n(n+2)}+\frac{n}{(n-1)^2}\frac{k-1}{k}c_2\right]c_1 \end{aligned}$$

with $c_1 = \mathbf{l}^\top \mathbf{S}^{-1}\mathbf{l}/\alpha^2$, $c_2 = (\overline{\mathbf{x}} - r_f\mathbf{1}_k)^\top \mathbf{S}^{-1}(\overline{\mathbf{x}} - r_f\mathbf{1}_k)$, $c_{12} = \mathbf{l}^\top \mathbf{S}^{-1}(\overline{\mathbf{x}} - r_f\mathbf{1}_k)/\alpha$ and, similarly,

$$\begin{aligned} &Var(\mathbf{l}^\top \mathbf{w}_{TP}|\mathbf{x}) = (\nu_c+n-k-1)c_{12}^2 \\ +& \left[\frac{(\nu_c+n-k)^2+k-2}{(n+\kappa_c)(\nu_c+n-k+2)}+(\nu_c+n-k)\frac{k-1}{k}c_2\right]c_1 \end{aligned}$$

with $c_1 = \mathbf{l}^\top \tilde{\mathbf{S}}^{-1}\mathbf{l}/\alpha^2$, $c_2 = \mathbf{a}_{12}^\top \tilde{\mathbf{S}}^{-1}\mathbf{a}_{12}$, and $c_{12} = \mathbf{l}^\top \tilde{\mathbf{S}}^{-1}\mathbf{a}_{12}/\alpha$.

Using the structure of both the variances and the fact that $\mathbf{l}$ is arbitrary chosen, we get the statement of the theorem. $\square$

Proof of Theorem 4. The application of Theorem 1 leads to

$$\boldsymbol{\theta} \stackrel{d}{=} \frac{\eta_d}{\alpha} \cdot \mathbf{L}\mathbf{S}_d^{-1}\breve{\boldsymbol{\mu}}_d + \frac{\sqrt{\eta_d}}{\alpha}\left(\breve{\boldsymbol{\mu}}_d^T\mathbf{S}_d^{-1}\breve{\boldsymbol{\mu}}_d \cdot \mathbf{L}\mathbf{S}_d^{-1}\mathbf{L}^\top - \mathbf{L}\mathbf{S}_d^{-1}\breve{\boldsymbol{\mu}}_d\breve{\boldsymbol{\mu}}_d^T\mathbf{S}_d^{-1}\mathbf{L}^\top\right)^{1/2}\mathbf{z}_0,$$

under the diffuse prior, where $\eta_d \sim \chi^2_n$, $\breve{\boldsymbol{\mu}}_d|\mathbf{x} \sim t_k\left(n-k, \mathbf{a}_1, \frac{n-1}{n(n-k)}\mathbf{S}\right)$, and $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$ which are mutually independent, and

$$\boldsymbol{\theta} \stackrel{d}{=} \frac{\eta_c}{\alpha} \cdot \mathbf{L}\mathbf{S}_c^{-1}\breve{\boldsymbol{\mu}}_c + \frac{\sqrt{\eta_c}}{\alpha}\left(\breve{\boldsymbol{\mu}}_c^T\mathbf{S}_c^{-1}\breve{\boldsymbol{\mu}}_c \cdot \mathbf{L}\mathbf{S}_c^{-1}\mathbf{L}^\top - \mathbf{L}\mathbf{S}_c^{-1}\breve{\boldsymbol{\mu}}_c\breve{\boldsymbol{\mu}}_c^T\mathbf{S}_c^{-1}\mathbf{L}^\top\right)^{1/2}\mathbf{z}_0,$$

where $\eta_c \sim \chi^2_{\nu_c+n-k}$, $\breve{\boldsymbol{\mu}}_c|\mathbf{x} \sim t_k\left(\nu_c+n-2k, \mathbf{a}_{12}, \frac{1}{(n+\kappa_c)(\nu_c+n-2k)}\tilde{\mathbf{S}}\right)$, and $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$ which are mutually independent.

Consequently, we get that

$$\sqrt{n}\left(\begin{pmatrix}\eta_d/n\\ \mathbf{z}_0/\sqrt{n}\\ \breve{\boldsymbol{\mu}}_d\end{pmatrix} - \begin{pmatrix}1\\ \mathbf{0}\\ \mathbf{a}_1\end{pmatrix}\right) \stackrel{d}{\longrightarrow} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix}2 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_p & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{S}\end{pmatrix}\right)$$

and

$$\sqrt{n}\left(\begin{pmatrix}\eta_c/n\\ \mathbf{z}_0/\sqrt{n}\\ \breve{\boldsymbol{\mu}}_c\end{pmatrix} - \begin{pmatrix}1\\ \mathbf{0}\\ \mathbf{a}_1\end{pmatrix}\right) \stackrel{d}{\longrightarrow} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix}2 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_p & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{S}\end{pmatrix}\right)$$

as $n \longrightarrow \infty$.

The application of the delta method (c.f., (DasGupta, 2008, Theorem 3.7)) proves that

$$\sqrt{n}(\mathbf{w}_{TP} - \alpha^{-1}\mathbf{S}^{-1}(\overline{\mathbf{x}} - r_f\mathbf{1}_k))|\mathbf{x} \stackrel{d.}{\longrightarrow} \mathcal{N}_k(\mathbf{0}, \mathbf{F}_d)$$

and

$$\sqrt{n}(\mathbf{w}_{TP} - \alpha^{-1}\mathbf{S}^{-1}(\overline{\mathbf{x}} - r_f\mathbf{1}_k))|\mathbf{x} \stackrel{d.}{\longrightarrow} \mathcal{N}_k(\mathbf{0}, \mathbf{F}_c),$$

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

Finally, using the results of Theorem 3 we get

$$\begin{aligned}
\mathbf{F}_d &= \lim_{n\longrightarrow\infty} Var(\sqrt{n}\mathbf{w}_{TP}) = \lim_{n\longrightarrow\infty}\left\{n\frac{1}{n-1}\alpha^{-2}\mathbf{S}^{-1}(\overline{\mathbf{x}} - r_f\mathbf{1}_k)(\overline{\mathbf{x}} - r_f\mathbf{1}_k)^\top\mathbf{S}^{-1}\right.\\
&+ \left.\alpha^{-2}n\left[\frac{n^2+k-2}{(n-1)n(n+2)} + \left(\frac{1}{(n-1)(n+2)} + \frac{n}{(n-1)^2}\frac{k-1}{k}\right)b_d\right]\mathbf{S}^{-1}\right\}\\
&= \alpha^{-2}\breve{\mathbf{S}}^{-1}(\breve{\mathbf{x}} - r_f\mathbf{1}_k)(\breve{\mathbf{x}} - r_f\mathbf{1}_k)^\top\breve{\mathbf{S}}^{-1} + \alpha^{-2}\left[1 + \frac{k-1}{k}\breve{b}_d\right]\breve{\mathbf{S}}^{-1}
\end{aligned}$$

where $\breve{b}_d = (\breve{\mathbf{x}} - r_f \mathbf{1}_k)^\top \breve{\mathbf{S}}^{-1} (\breve{\mathbf{x}} - r_f \mathbf{1}_k)$ with $\breve{\mathbf{x}}$ and $\breve{\mathbf{S}}$ defined in the statement of the theorem.

Similarly,

$$\mathbf{F}_c = \alpha^{-2} \breve{\mathbf{S}}^{-1} (\breve{\mathbf{x}} - r_f \mathbf{1}_k)(\breve{\mathbf{x}} - r_f \mathbf{1}_k)^\top \breve{\mathbf{S}}^{-1} + \alpha^{-2} \left[1 + \frac{k-1}{k} \breve{b}_d \right] \breve{\mathbf{S}}^{-1} = \mathbf{F}_d,$$

which completes the proof of the theorem.

□

References

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

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

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

Bawa, V. S., Brown, S., and Klein, R. (1979). Estimation Risk and Optimal Portfolio Choice. North-Holland, Amsterdam.

Bodnar, T., Mazur, S., and Okhrin, Y. (2016a). Bayesian estimation of the global minimum variance portfolio. European Journal of Operational Research, page To appear.

Bodnar, T., Mazur, S., Podgórski, K., and Tyrcha, J. (2016b). Tangency portfolio weights for small sample and singular covariance matrix: estimation and test theory. Technical report, Lund University.

Bodnar, T. and Okhrin, Y. (2011). On the product of inverse Wishart and normal distributions with applications to discriminant analysis and portfolio theory. Scandinavian Jornal of Statistics, 38:311–331.

Bodnar, T., Parolya, N., and Schmid, W. (2015). Estimation of the global minimum variance portfolio in high dimensions. Technical report.

Brandt, M. (2010). Handbook of Financial Econometrics, chapter Portfolio choice problems, pages 269–336. North Holland.

Brown, S. (1976). Optimal portfolio choice under uncertainty: A bayesian approach. PhD thesis, University of Chicago.

DasGupta, A. (2008). Asymptotic Theory of Statistics and Probability. Springer Texts in Statistics. Springer.

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

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

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

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

Gupta, A., Varga, T., and Bodnar, T. (2013). Elliptically contoured models in statistics and portfolio theory. Springer, second edition.

Isserlis, L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. Biometrika, 12:134–139.

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

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

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

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

Kotz, S. and Nadarajah, S. (2004). Multivariate t distributions and their applications. Cambridge University Press, The Pitt Building, Trumpington Street, Cambridge, United Kingdom.

Markowitz, H. M. (1952). Mean-variance analysis in portfolio choice and capital markets. Journal of Finance, 7:77–91.

Mathai, A. M. and Provost, S. B. (1992). Quadratic forms an random variables. Marcel Dekker, New York.

Meyer, C. D. (2000). Matrix analysis and applied linear algebra. SIAM.

Muirhead, R. J. (1982). Aspects of Multivariate Statistical Theory. Wiley, New York.

Pastor, L. . (2000). Portfolio selection and asset pricing models. Journal of Finance, 55:179–223.

Rachev, S. T., Hsu, J. S. J., Bagasheva, B. S., and Fabozzi, F. J. (2008). Bayesian methods in finance. The Frank J. Fabozzi. Wiley, New Jersey.

Sekerke, M. (2015). Bayesian risk management: a guide to model risk and sequential learning in financial markets. Wiley, New Jersey.

Wang, Z. (2005). A shrinkage approach to model uncertainty and asset allocation. Review of Financial Studies, 18:673–705.

Winkler, R. (1973). Bayesian models for forecasting future security prices. Journal of Financial and Quantitative Analysis, 8:387–405.

Winkler, R. L. and Barry, C. B. (1975). A bayesian model for portfolio selection and revision. Journal of Finance, 30:179–192.