Higher order moments of the estimated tangency portfolio weights

Farrukh Javed, Stepan Mazur & Edward Ngailo

View supplementary material

Published online: 05 Mar 2020.

Submit your article to this journal

Article views: 486

View related articles

View Crossmark data

Higher order moments of the estimated tangency portfolio weights

Farrukh Javed[a], Stepan Mazur[a] and Edward Ngailo[b]

[a]Örebro University School of Business, Örebro, Sweden; [b]Department of Mathematics, Linköping University, Linköping, Sweden

ABSTRACT

In this paper, we consider the estimated weights of the tangency portfolio. We derive analytical expressions for the higher order non-central and central moments of these weights when the returns are assumed to be independently and multivariate normally distributed. Moreover, the expressions for mean, variance, skewness and kurtosis of the estimated weights are obtained in closed forms. Later, we complement our results with a simulation study where data from the multivariate normal and t-distributions are simulated, and the first four moments of estimated weights are computed by using the Monte Carlo experiment. It is noteworthy to mention that the distributional assumption of returns is found to be important, especially for the first two moments. Finally, through an empirical illustration utilizing returns of four financial indices listed in NASDAQ stock exchange, we observe the presence of time dynamics in higher moments.

1. Introduction

The fundamental goal of the portfolio theory, as devised by Markowitz [39], is to determine an efficient way of portfolio allocation. The mean–variance optimization technique plays a central role in allocating investments among different assets. According to it, the investor allocates the wealth among risky assets by maximizing the expected return based on a given level of risk or by minimizing the risk for a given level of expected returns. The trade-off between the risk and return of the portfolio is at the heart of portfolio theory, which seeks to find optimal allocations of the investor's initial wealth to the available assets. The tangency portfolio (TP) is one such portfolio which consists of both risky and risk-free assets. In order to have entire understanding of the conditions and processes in a portfolio, the study on its statistical properties is crucial and unavoidable. Therefore, in this paper, we derive analytical results for the higher moments of TP estimated weights, which also include the expressions for skewness and kurtosis.

Statistical properties of the estimated TP weights are intensively discussed in the existing literature. For example, Britten-Jones [20] developed an F-test for the TP weights, while Bodnar [6] delivered sequential monitoring procedures for the TP weights. The univariate

density of the TP weights as well as its asymptotic distribution under the assumption of independently and multivariate normally distributed returns are obtained by Okhrin and Schmid [44]. Later on, Bodnar and Okhrin [16] derived the explicit density of the linear transformation of the estimated weights and suggested several exact tests of general linear hypothesis about the elements of the portfolio weights. Kotsiuba and Mazur [34] derived the approximate density function formula for the weights, which is based on the Gaussian integral and the third-order Taylor expansion. A test on the location of the TP on the set of feasible portfolios is developed by Muhinyuza *et al.* [42]. Bodnar *et al.* [15] extended the results by Bodnar and Okhrin [16] in the setting when both the population and the sample covariance matrices are singular. Moreover, they established the high-dimensional asymptotic distribution of the estimated weights of the TP when both the portfolio dimension and the sample size increase to infinity. In [46], the authors delivered new theory-based portfolio strategies which are the combinations of the naive $1/N$ rule with the sophisticated theory-based strategies. Shrinkage estimators for the optimal portfolio weights that allow us to shrink the estimated classical Markowitz weights to the deterministic target portfolio weights are proposed by, for example, Wang [47]. More recently, Bauder *et al.* [5] studied the distributional properties of the weights of the TP from the Bayesian point of view.

To contribute to the existing literature on TP weights, in this paper, we aim to derive the higher order moments of the sample weights of the TP in closed forms when the returns are assumed to be independently and multivariate normally distributed. The results presented here are further derived from [32], where the idea was discussed in a more compact form. This article, however, can be seen as a detailed extension of the mentioned working paper. Let us note that there is a reasonable amount of the literature available (see, e.g. [30,33]) discussing portfolio selection based on higher moments of asset returns, but not much has been done from the perspective of the distribution of portfolio weights. This article is a step further in this direction. Higher order moments can be used for the approximation of the density function of the estimated weights (see, e.g. [37]). As argued by Okhrin and Schmid [44], the knowledge of portfolio weights leads to information about the expected portfolio return and the variance of the portfolio return. Since the expected portfolio returns play a crucial role in most financial theories, the knowledge of the first two moments of the estimated portfolio weights can be helpful in learning about the expected portfolio return as well as the portfolio finance. Similarly, via deriving expressions for moments greater than 2, such as the skewness and kurtosis of estimated weights, we would be able to understand the tail and asymmetric behavior of the fraction of wealth allocated to assets in the portfolio. It will help us in indicating how much the estimated weights deviate from normality. More specifically, the measures of skewness and kurtosis can account for asymmetry and tail risk. In [44], the authors show that the moments of the optimal portfolio weights are very sensitive to changes in the moments of stock returns. The obtained expressions for the higher moments of the estimated portfolio weights can, therefore, be very informative for practitioners to account for tail risks in making portfolio strategies. Following the lines of [18], we obtain explicit relations of minimum VaR and minimum CVaR portfolio weights in terms of estimated tangency portfolio weights, where these higher moments come into play a significant role in accounting for portfolio risk. These measures, for our chosen portfolios, will help to better understand the driving forces of the market's portfolio risk since the distributional properties of weights are the consequential inputs for investment and

asset allocation decisions, pricing derivatives and hedging against portfolio risk. In this particular article, we would obtain explicit expressions for partial cases such as the mean, variance, skewness, and kurtosis. More specifically, we will take a look at the skewness and kurtosis for measuring the deviation from the normal distribution. It would be interesting to see how moments of estimated weights behave when the assumption of normally distributed data is violated. We are going to analyze it numerically by simulating data from the multivariate t-distribution with 5 and 10 degrees of freedom, and by computing moments of the TP weights using the Monte Carlo experiment.

This paper is organized as follows. In Section 2, we present our main results where we deliver explicit formulas for the higher order non-central and central moments of the estimated TP weights. Moreover, we derive the mean, variance, skewness and kurtosis in closed forms. Section 3 is devoted to stress possible application implications of the main results. In Section 4, we establish auxiliary results which we use in proving the main results of Section 2. The results of simulation studies and applications are given in Section 5, while Section 6 summarizes the paper. All the proofs of the main results are collected in the appendix. Proofs of auxiliary results and some tables are collected in the online supplementary materials.

2. Main results

We consider a portfolio that consists of k assets. Let $\mathbf{x}_t = (x_{1t}, \ldots, x_{kt})^{\mathrm{T}}$ be the k-dimensional vector of log-returns of these assets at time $t = 1, \ldots, n$. The fraction of the wealth allocated to the ith asset in the portfolio is denoted by w_i, and let $\mathbf{w} = (w_1, \ldots, w_k)^{\mathrm{T}}$ be the vector of weights. Let the mean vector of the asset returns be denoted by $\boldsymbol{\mu}$ and the covariance matrix by $\boldsymbol{\Sigma}$ which is assumed to be positive definite.

In [36,38], the authors showed both theoretically and empirically that the mean–variance optimal portfolio problem is equivalent to maximizing the expected quadratic utility. Since the risk is usually measured by the variance of the portfolios return, the optimal portfolio without a risk-free asset is obtained by minimizing the portfolio variance for a given level of the expected return under the constraint $\mathbf{w}^{\mathrm{T}}\mathbf{1}_k = 1$ where $\mathbf{1}_k$ denotes the vector of ones. However, if short selling is allowed and a risk-free asset, with return r_f, is available, then part of investor's wealth is invested into the risk-free asset, whereas the rest of the wealth is invested into the portfolio from the efficient frontier. The return of risky assets is given as $\mu_p = \mathbf{w}^{\mathrm{T}}(\boldsymbol{\mu} - r_f\mathbf{1}_k) + r_f$ with the variance $\sigma_p^2 = \mathbf{w}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{w}$.

In this paper, we consider the weights of the TP that are obtained as the solution to the following optimization problem:

$$\max_{\mathbf{w}} \left[\mu_p - \frac{\alpha}{2}\sigma_p^2 \right], \tag{1}$$

where $\alpha > 0$ denotes the investor's attitude towards risk and is called risk aversion. The higher number of α representing lesser tolerance to risk. This level of aversion to risk can be characterized by defining the investor's indifference curve which represents the investor's preferences for risk and return. There is a large amount of literature on measuring risk aversion, and a different approach has been suggested to estimate this coefficient. The most common choice of the risk aversion lies between 1 and 3, but one can find a wide range of α in the literature – from 0.2 to 10 and even higher (see, e.g. [27] and references therein).

Let us note that in order to obtain an explicit solution to the investor's problem we do allow short sales, i.e. no restrictions are placed on the portfolio weights. Since $\mathbf{\Sigma} > 0$, the TP weights are given by

$$\mathbf{w}_{TP} = \alpha^{-1}\mathbf{\Sigma}^{-1}(\mathbf{\mu} - r_f\mathbf{1}_k). \tag{2}$$

The vector of the tangency portfolio weights $\mathbf{w}_{TP}$ that is defined in (2) determines the structure of the portfolio that corresponds to risky assets only, while $1 - \mathbf{w}_{TP}^{\mathrm{T}}\mathbf{1}_k$ determines the part of the wealth that should be invested in the risk-free asset. According to Ingersoll [31], the TP lies on the intersection of the mean–variance frontier and the tangency line drawn from the portfolio consisting of the risk-free asset.

Since $\mathbf{\mu}$ and $\mathbf{\Sigma}$ are unknown parameters, the investor cannot determine $\mathbf{w}_{TP}$. Consequently, both $\mathbf{\mu}$ and $\mathbf{\Sigma}$ need to be estimated. There are numerous estimation techniques for the mean vector (see [17,25]), covariance matrix and its inverse (see [7,22,26]). In this paper, we consider the classical unbiased sample estimators for $\mathbf{\mu}$ and $\mathbf{\Sigma}$ which are expressed as

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_j \quad \text{and} \quad \mathbf{S} = \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^{\mathrm{T}}.$$

Throughout the paper, it is assumed that the asset returns $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are independent and identically distributed (iid) such that $\mathbf{x}_t \sim \mathcal{N}_k(\mathbf{\mu}, \mathbf{\Sigma})$, $t = 1, \ldots, n$. Replacing $\mathbf{\mu}$ and $\mathbf{\Sigma}$ with $\bar{\mathbf{x}}$ and $\mathbf{S}$ in (2), we obtain the sample estimator $\widehat{\mathbf{w}}_{TP}$ of TP weights $\mathbf{w}_{TP}$, i.e.

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

In this paper, we focus on the linear combination of the TP weights. In particular, we are interested in

$$\theta = \mathbf{l}^{\mathrm{T}}\mathbf{w}_{TP} = \alpha^{-1}\mathbf{l}^{\mathrm{T}}\mathbf{\Sigma}^{-1}(\mathbf{\mu} - r_f\mathbf{1}_k),$$

where $\mathbf{l}$ is a k-dimensional vector of constants. From the investment point of view, the choice of vector $\mathbf{l}$ can be made in different ways. For example, if $\mathbf{l}^{\mathrm{T}} = (1, 0, \ldots, 0)$ then an investor will know the behavior of the TP weight of the first asset in the portfolio. Similarly, if $\mathbf{l}^{\mathrm{T}} = (1, 1, 0, \ldots, 0)$, the behavior of the sum of the first two assets in the portfolio can be analyzed and so on. If, on the other hand, $\mathbf{l} = \mathbf{1}_k$, it means an investor is only interested in knowing about how much will be invested into the *risky assets*.

In a more general setting, the sample estimator of θ is given by

$$\hat{\theta} = \mathbf{l}^{\mathrm{T}}\widehat{\mathbf{w}}_{TP} = \alpha^{-1}\mathbf{l}^{\mathrm{T}}\mathbf{S}^{-1}(\bar{\mathbf{x}} - r_f\mathbf{1}_k).$$

In Theorem 2.1, we deliver explicit expressions for the higher order non-central and central moments of $\hat{\theta}$. Both expressions depend on a confluent hypergeometric function $_1F_1(a; b; x)$, which is defined as

$$_1F_1(a; b; x) = 1 + \frac{a}{b}x + \frac{a(a+1)}{b(b+1)}\frac{x^2}{2!} + \cdots = \sum_{k=0}^{\infty}\frac{(a)_k}{(b)_k}\frac{x^k}{k!},$$

where $(a)_k$ and $(b)_k$ are Pochhammer symbols [1]. Note that the computation of a confluent hypergeometric function is a standard routine within many mathematical software

packages, such as in R. Let us also note that the non-central and central moments of the estimated TP weights exist only up to the order $(n - k)/2$, while the moments of the order higher than $(n - k)/2$ do not exist at all.

Theorem 2.1: *Let* $\mathbf{x}_1, \ldots, \mathbf{x}_n$ *be iid random vectors with* $\mathbf{x}_1 \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $k < n-1$ *and* $\boldsymbol{\Sigma} > 0$. *Also, let* $\mathbf{l}$ *be a k-dimensional vector of constants,* $\check{s} = n\check{\boldsymbol{\mu}}^{\mathrm{T}}\mathbf{R}_{\mathbf{l}}\check{\boldsymbol{\mu}}$ *with* $\check{\boldsymbol{\mu}} = \boldsymbol{\mu} - r_f\mathbf{1}_k$ *and* $\mathbf{R}_{\mathbf{l}} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{l}\mathbf{l}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}/\mathbf{l}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{l}$. *Then*

(a) *the rth order moment of* $\hat{\theta}$ *is given by*

$$\mu_r := \mathrm{E}[\hat{\theta}^r] = \frac{(n-1)^r}{\alpha^r(n-k-2)\cdots(n-k-2r)}$$

$$\times \left[(\mathbf{l}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\check{\boldsymbol{\mu}})^r + \sum_{j=1}^{\lfloor r/2 \rfloor} \binom{r}{2j} \frac{(2j)!}{(2n)^j j!} \left(\mathbf{l}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\check{\boldsymbol{\mu}} \right)^{r-2j} \left(\mathbf{l}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{l} \right)^j \right.$$

$$\left. \times \left(1 + \sum_{m=1}^{j} \binom{j}{m} \left(\frac{k-1}{n-k+1} \right)^m \check{c}_m \right) \right], \quad n-k > 2r,$$

where

$$\check{c}_m = \frac{(k-1+2(m-1))\cdots(k-1)}{(n-k-2(m-1)-1)\cdots(n-k-1)} \mathrm{e}^{-\check{s}/2}{}_1F_1\left(m + \frac{k-1}{2}; \frac{k-1}{2}; \frac{\check{s}}{2} \right);$$

(b) *the rth order central moment of* $\hat{\theta}$ *is given by*

$$\bar{\mu}_r := \mathrm{E}[(\hat{\theta} - \mu_1)^r] = (-\mu_1)^r + \sum_{i=1}^{r} \binom{r}{i} \frac{(-\mu_1)^{r-i}(n-1)^i}{\alpha^i(n-k-2)\cdots(n-k-2i)}$$

$$\times \left[(\mathbf{l}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\check{\boldsymbol{\mu}})^i + \sum_{j=1}^{\lfloor i/2 \rfloor} \binom{i}{2j} \frac{(2j)!}{(2n)^j j!} \left(\mathbf{l}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\check{\boldsymbol{\mu}} \right)^{i-2j} \left(\mathbf{l}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{l} \right)^j \right.$$

$$\left. \times \left(1 + \sum_{m=1}^{j} \binom{j}{m} \left(\frac{k-1}{n-k+1} \right)^m \check{c}_m \right) \right], \quad n-k > 2r.$$

Remark 2.1: From (3), we can observe that the sample estimator of TP weights $\widehat{\mathbf{w}}_{TP}$ depends on the inverse of the sample covariance matrix $\mathbf{S}$. In Theorem 2.1, we assume that $k < n - 1$, and it ensures that the distribution of $\mathbf{S}$ to be non-singular, which makes it invertible. If $k > n-1$, then $\mathbf{S}$ is singular and a regular inverse cannot be taken. This problem has been addressed in the portfolio context by employing Moore–Penrose inverse (see [13–15]). Alternatively, various regularization techniques can be used. For example, one can employ the ridge-type approach that is based on adding a diagonal matrix to the covariance matrix [45], Landweber–Fridman iterative algorithm [35], the spectral cut-off that is based on a singular value decomposition [24], a form of Lasso where the l_1 norm of the optimal portfolio weights is penalized [21], or an iterative method that is based on second-order damped dynamical systems [28].

The following corollary delivers the expressions of the mean and the variance for $\widehat{\mathbf{w}}_{TP}$.

Corollary 2.2: *Let* $\mathbf{x}_1, \ldots, \mathbf{x}_n$ *be iid random vectors with* $\mathbf{x}_1 \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $k < n-1$, *and* $\boldsymbol{\Sigma} > 0$. *Also let* $\breve{\boldsymbol{\mu}} = \boldsymbol{\mu} - r_f \mathbf{1}_k$ *and* $\delta = n\breve{\boldsymbol{\mu}}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \breve{\boldsymbol{\mu}}$. *Then the mean and the variance of* $\widehat{\mathbf{w}}_{TP}$ *are given by*

$$\mathrm{E}[\widehat{\mathbf{w}}_{TP}] = \frac{n-1}{n-k-2}\mathbf{w}_{TP} \quad \text{and} \quad \mathrm{Var}[\widehat{\mathbf{w}}_{TP}] = \breve{d}_1^{(0)}\mathbf{w}_{TP}\mathbf{w}_{TP}^{\mathrm{T}} + \breve{d}_2^{(0)}\alpha^{-2}\boldsymbol{\Sigma}^{-1}$$

with

$$\breve{d}_1^{(0)} = \frac{(n-k)(n-1)^2}{(n-k-1)(n-k-2)^2(n-k-4)},$$

$$\breve{d}_2^{(0)} = \frac{(n-1)^2(n-2+\delta)}{n(n-k-1)(n-k-2)(n-k-4)}.$$

From Corollary 2.2, we can see that the sample estimator of the weights is biased, meaning that the $\mathrm{E}[\widehat{\mathbf{w}}_{TP}] \neq \mathbf{w}_{TP}$. However, note that for large sample size, asymptotically, the estimator is unbiased since $\lim_{n\to\infty} \mathrm{E}[\widehat{\mathbf{w}}_{TP}] = \mathbf{w}_{TP}$. Consequently, the sample estimator of the TP weights is consistent, i.e. $\mathrm{plim}_{n\to\infty} \widehat{\mathbf{w}}_{TP} = \mathbf{w}_{TP}$, where plim denotes convergence in probability.

In the next corollary, we derive the expressions for skewness and the kurtosis of $\hat{\theta} = \mathbf{l}^{\mathrm{T}}\widehat{\mathbf{w}}_{TP}$.

Corollary 2.3: *Let* $\mathbf{x}_1, \ldots, \mathbf{x}_n$ *be iid random vectors with* $\mathbf{x}_1 \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $k < n-1$ *and* $\boldsymbol{\Sigma} > 0$. *Also, let* $\mathbf{l}$ *be a k-dimensional vector of constants,* $\check{\theta} = \mathbf{l}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\breve{\boldsymbol{\mu}}$ *with* $\breve{\boldsymbol{\mu}} = \boldsymbol{\mu} - r_f \mathbf{1}_k$, *and* $\check{s} = n\breve{\boldsymbol{\mu}}^{\mathrm{T}}\mathbf{R_l}\breve{\boldsymbol{\mu}}$ *with* $\mathbf{R_l} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{l}\mathbf{l}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}/\mathbf{l}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{l}$. *Then the skewness* $\hat{\theta}$ *is given by*

$$\mathrm{Skewness}[\hat{\theta}] = \left(\breve{d}_1^{(1)}\check{\theta}^3 + \breve{d}_2^{(1)}\check{\theta}\mathbf{l}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{l}\right)\left(\breve{d}_1^{(0)}\check{\theta}^2 + \breve{d}_2^{(0)}\mathbf{l}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{l}\right)^{-3/2}$$

with $\breve{d}_1^{(0)}$ *and* $\breve{d}_2^{(0)}$, *which are defined in Corollary 2.2, and*

$$\breve{d}_1^{(1)} = \frac{16(n-1)^3}{(n-k-2)^3(n-k-4)(n-k-6)},$$

$$\breve{d}_2^{(1)} = \frac{12(n-1)^3}{n(n-k-2)^2(n-k-4)(n-k-6)}\left(1 + \frac{\check{s}+k-1}{n-k-1}\right),$$

while the kurtosis of $\hat{\theta}$ *is expressed as*

$$\mathrm{Kurtosis}[\hat{\theta}] = \left(\breve{d}_1^{(2)}\check{\theta}^4 + \breve{d}_2^{(2)}\check{\theta}^2\mathbf{l}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{l} + \breve{d}_3^{(2)}(\mathbf{l}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{l})^2\right)\left(\breve{d}_1^{(0)}\check{\theta}^2 + \breve{d}_2^{(0)}\mathbf{l}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{l}\right)^{-2},$$

where

$$\breve{d}_1^{(2)} = \frac{3(n-1)^4[(n-k)(n-k-6)(n-k-8) - (n-k-2)^2(n-k-10)]}{(n-k-2)^4(n-k-4)(n-k-6)(n-k-8)},$$

$$\breve{d}_2^{(2)} = \frac{6(1+\check{c}_1)(n-1)^4[(n-k-2)^2 - (n-k+2)(n-k-8)]}{n(n-k-2)^3(n-k-4)(n-k-6)(n-k-8)},$$

$$\breve{d}_3^{(2)} = \frac{3(1 + 2\breve{c}_1 + \breve{c}_2)(n-1)^4}{n^2(n-k-2)(n-k-4)(n-k-6)(n-k-8)},$$

with

$$\breve{c}_1 = \frac{\breve{s} + k - 1}{n - k - 1} \quad \text{and} \quad \breve{c}_2 = \frac{\breve{s}^2 + (2\breve{s} + k - 1)(k + 1)}{(n - k - 1)(n - k - 3)}.$$

One of the important factors to consider when selecting the optimal portfolio for a particular investor is the degree of risk aversion coefficient α, where the higher the number is, the lesser the tolerance to risk becomes. We observe that the skewness and kurtosis of estimated portfolio weights are found to be not depending on α, and the level of risk aversion does not influence these higher moments. This finding is consistent with the existing literature (see, e.g. [23]). It indicates that the magnitude of the investor's tolerance level to risk does not affect the higher moments (skewness and kurtosis) of estimated weights.

The proofs of the main results are provided in the appendix.

3. Application implications of main results

The results obtained in Section 2 can be used in many different ways. Below we summarize few applications which are of immediate interest both for theoreticians and practitioners.

It is well known that the cumulants and moments can be used to define the probability distribution of a random variable under study. For example, for the Gaussian case, all cumulants of order greater than two are zero; therefore, higher order cumulants can be used for testing of Gaussianity as well as for proving classical central limit theorems.

Let us consider the characteristic function of $\hat{\theta} = \mathbf{1}^{\mathrm{T}}\widehat{\mathbf{w}}_{TP}$ denoted by $\varphi_{\hat{\theta}}(t)$, $t \in \mathbb{R}$. It can be expressed using series expansion that is given by

$$\varphi_{\hat{\theta}}(t) = \mathrm{E}\left[e^{it\hat{\theta}}\right] = 1 + \sum_{j=1}^{\infty} \mu_j \frac{(it)^j}{j!}, \quad t \in \mathbb{R},$$

where $\mu_j = \mathrm{E}[\hat{\theta}^j]$. It also holds that $(-i)^j(\mathrm{d}^j\varphi_{\hat{\theta}}(t)/\mathrm{d}t^j)|_{t=0} = \mu_j$. Hence, we can observe the connection between moments of $\hat{\theta}$ and its characteristic function that completely defines the probability distribution. Having the characteristic function of $\hat{\theta}$, the cumulant generating function can be defined as (see [37])

$$\psi_{\hat{\theta}}(t) = \ln\left[\varphi_{\hat{\theta}}(t)\right] = \sum_{j=1}^{\infty} \kappa_j \frac{(it)^j}{j!}, \quad t \in \mathbb{R},$$

where κ_j denotes the jth cumulant of $\hat{\theta}$ that can be obtained in terms of moments. For example, $\kappa_1 = \mu_1$, $\kappa_2 = \mu_2 - \mu_1^2$, etc. (see [48]).

From Theorem 2.1 we know that the non-central and central moments of $\hat{\theta}$ exist only up to the order $(n - k)/2$, while the moments of the order higher than $(n - k)/2$ do not exist at all. Consequently, we can deliver the approximations of the characteristic and cumulant generating functions of $\hat{\theta}$ that are based on the higher order moments and cumulants

expressed as

$$\varphi_{\hat{\theta}}(t) \approx 1 + \sum_{j=1}^{\lfloor (n-k)/2 \rfloor} \mu_j \frac{(it)^j}{j!} \quad \text{and} \quad \psi_{\hat{\theta}}(t) \approx \sum_{j=1}^{\lfloor (n-k)/2 \rfloor} \kappa_j \frac{(it)^j}{j!}.$$

Let us recall that the skewness is a measure for the degree of symmetry in the distribution and deviation from zero to the left or right side indicates the presence of asymmetry. Negatively skewed distributions lead to a long left tail which, from an investor's perspective, can mean a greater chance of extremely negative outcomes. While the positive skew implies a long right tail, it can result in a greater chance of extremely positive outcomes. On the other hand, the kurtosis is a measure of the fatness in the tails and deviation from 3 for a Gaussian distributed variable, is an indicator of the presence of tails fatter than Gaussian, and therefore, increases the likelihood of extreme events. The closed-form expressions presented in Corollaries 2.2 and 2.3, can be used as a measure of asymmetry and tail behavior in the fraction of weights allocated to different assets in the portfolio. Moreover, with the help of standard deviation, one can observe how dramatically estimated portfolio weights oscillate over a period of time.

The quantification of the risk of a portfolio has been of immense interest both for theoreticians and practitioners. Usually, the variance of the portfolio is considered as a measure of the portfolio risk. However, it is not always an appropriate risk measure since it takes into account a two-sided risk. A recent development in this direction highlights that the quantile-based measures are well-suited functions to quantify risk. Among these, the most popular are the Value-at-Risk (VaR) and Conditional VaR (CVaR), where the latter is also known as the expected shortfall (see, e.g. [2]). In contrast to the variance, the VaR and the CVaR are one-sided risk measures. The Basle Committee on Banking Supervision allows banks to use VaR when determining their capital-adequacy requirements arising from their exposure to market risk. The portfolio selection problems based on minimizing the portfolio VaR (CVaR) have been considered in a number of literature studies. For example, Alexander and Baptista [3,4] suggested the application of the VaR and the CVaR as measures of the risk in Markowitz's optimization problem instead of the variance and examined the economic implications of a mean-VaR model for portfolio selection. A brief connection between the TP and the minimum VaR portfolio has been drawn by Bodnar and Zabolotskyy [19] while focusing mainly on the riskiness of an optimal portfolio which maximizes the Sharpe ratio. Following the lines of Bodnar *et al.* [18], we obtain explicit relations of minimum VaR and minimum CVaR portfolio weights in terms of estimated tangency portfolio weights, where these higher moments come into play their role. Since the main concern for the minimum VaR and minimum CVaR measure is a tail risk, the knowledge of higher moments of estimated weights can, therefore, be helpful for practitioners in better understanding the driving forces of markets portfolio risk and making asset allocation decisions against the portfolio risk.

4. Auxiliary results

In this section, we present the auxiliary results, which are used in proving our main results of Section 2 and can be applied in the discriminant analysis (see [9]). Let us note that our findings are complementing the existing results obtained in [8,10–12,16,34].

The assumption of normally distributed data is a standard in different fields of applied and theoretical statistics. Consequently, we can find many expressions involving the estimated mean and the estimated covariance matrix of a k-dimensional normal distribution, i.e. $\mathbf{x}_t \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for $t = 1, \ldots, n$ and $n > k$, where n is a sample size. Considering the sample estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ that are defined in (3) and assuming normality, we obtain that

$$\bar{\mathbf{x}} \sim \mathcal{N}_k\left(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}\right) \quad \text{and} \quad (n-1)\mathbf{S} \sim \mathcal{W}_k(n-1, \boldsymbol{\Sigma}),$$

where $\mathcal{W}_k(n-1, \boldsymbol{\Sigma})$ stands for a k-dimensional Wishart distribution with $n-1$ degrees of freedom and a positive definite covariance matrix $\boldsymbol{\Sigma}$; moreover, $\bar{\mathbf{x}}$ and $\mathbf{S}$ are independently distributed (see [43, Chapter 3]). Hence, we can observe that the sample estimator $\widehat{\mathbf{w}}_{TP}$ of the TP weights $\mathbf{w}_{TP}$ given in (3) is expressed as a product of an inverse Wishart random matrix and a Gaussian random vector. The same structure appears in the discriminant analysis, where the coefficients of a discriminant function that maximizes the discrepancy between two datasets are expressed as a product of an inverse Wishart random matrix and a Gaussian random vector (see, e.g. [9]).

Both objects in the portfolio theory and discriminant analysis can be generalized in the expression $\mathbf{l}^T\mathbf{A}^{-1}\mathbf{z}$, where $\mathbf{l}$ is a k-dimensional vector of constants, $\mathbf{A} \sim \mathcal{W}_k(n, \boldsymbol{\Sigma})$, and $\mathbf{z} \sim \mathcal{N}_k(\boldsymbol{\mu}, \lambda\boldsymbol{\Sigma})$ which is independent of $\mathbf{A}$. We assume that $n > k$, implying that the matrix $\mathbf{A}$ is non-singular. We also assume that $\lambda > 0$ is a constant and $\boldsymbol{\Sigma}$ is a positive definite matrix.

In the next theorem, we consider the higher order moments of the generalized expression $\mathbf{l}^T\mathbf{A}^{-1}\mathbf{z}$.

Theorem 4.1: *Let $\mathbf{A} \sim \mathcal{W}_k(n, \boldsymbol{\Sigma})$, $n > k$ and $\mathbf{z} \sim \mathcal{N}_k(\boldsymbol{\mu}, \lambda\boldsymbol{\Sigma})$ with $\lambda > 0$ and positive definite $\boldsymbol{\Sigma}$. Furthermore, let $\mathbf{A}$ and $\mathbf{z}$ be independent and $\mathbf{l}$ be a k-dimensional vector of constants. Then the rth order moment of $\mathbf{l}^T\mathbf{A}^{-1}\mathbf{z}$ is given by*

$$\mathrm{E}\left[(\mathbf{l}^T\mathbf{A}^{-1}\mathbf{z})^r\right] = \frac{1}{(n-k-1)\cdots(n-k-2r+1)}$$

$$\times \left[(\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^r + \sum_{j=1}^{\lfloor r/2 \rfloor}\binom{r}{2j}\frac{(2j)!}{2^j j!}\left(\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right)^{r-2j}\right.$$

$$\left.\times\left(\lambda\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\mathbf{l}\right)^j\left(1 + \sum_{m=1}^{j}\binom{j}{m}c_m\right)\right]$$

for $n-k+1 > 2r$ with

$$c_m = \frac{(k-1+2(m-1))\cdots(k-1)}{(n-k-2(m-1))\cdots(n-k)}e^{-s/2}{}_1F_1\left(m+\frac{k-1}{2}; \frac{k-1}{2}; \frac{s}{2}\right),$$

where $s = \boldsymbol{\mu}^T\mathbf{R}_l\boldsymbol{\mu}/\lambda$ and $\mathbf{R}_l = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{l}\mathbf{l}^T\boldsymbol{\Sigma}^{-1}/\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\mathbf{l}$.

The proof of Theorem 4.1 is given in the online supplementary materials. From Theorem 4.1, we can observe that the non-central and central moments of the estimated

TP weights exist only up to the order $(n - k)/2$, while the moments of the order higher than $(n - k)/2$ do not exist at all. It is also noticed that the formula for the higher order moments of $\mathbf{l}^T\mathbf{A}^{-1}\mathbf{z}$ depends on the confluent hypergeometric function.

Now we consider an explicit formula for the higher order central moments of $\mathbf{l}^T\mathbf{A}^{-1}\mathbf{z}$ which is given in the next corollary, while its proof is given in the online supplementary materials.

Corollary 4.2: *Let $\mathbf{A} \sim \mathcal{W}_k(n, \boldsymbol{\Sigma})$, $n > k$ and $\mathbf{z} \sim \mathcal{N}_k(\boldsymbol{\mu}, \lambda\boldsymbol{\Sigma})$ with $\lambda > 0$ and positive definite $\boldsymbol{\Sigma}$. Furthermore, let $\mathbf{A}$ and $\mathbf{z}$ be independent and $\mathbf{l}$ be a k-dimensional vector of constants. Then the rth order central moment of $\mathbf{l}^T\mathbf{A}^{-1}\mathbf{z}$ is given by*

$$
\mathrm{E}\left[\left(\mathbf{l}^T\mathbf{A}^{-1}\mathbf{z} - \mathrm{E}[\mathbf{l}^T\mathbf{A}^{-1}\mathbf{z}]\right)^r\right] = (-\kappa_1)^r + \sum_{i=1}^{r}\binom{r}{i}\frac{(-\kappa_1)^{r-i}}{(n-k-1)\cdots(n-k-2i+1)}
$$

$$
\times \left[\left(\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right)^i + \sum_{j=1}^{\lfloor i/2 \rfloor}\binom{i}{2j}\frac{(2j)!}{2^j j!}\left(\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right)^{i-2j}\right.
$$

$$
\left. \times \left(\lambda\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\mathbf{l}\right)^j\left(1 + \sum_{m=1}^{j}\binom{j}{m}c_m\right)\right]
$$

for $n - k + 1 > 2r$ with $\kappa_1 = (1/(n-k-1))\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ and

$$
c_m = \frac{(k-1+2(m-1))\cdots(k-1)}{(n-k-2(m-1))\cdots(n-k)}e^{-s/2}\,_1F_1\left(m+\frac{k-1}{2};\frac{k-1}{2};\frac{s}{2}\right),
$$

where $s = \boldsymbol{\mu}^T\mathbf{R}_{\mathbf{l}}\boldsymbol{\mu}/\lambda$ and $\mathbf{R}_{\mathbf{l}} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{l}\mathbf{l}^T\boldsymbol{\Sigma}^{-1}/\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\mathbf{l}$.

In the following corollary, we deliver the expressions of the second-order central moment, the third-order central moment, and the fourth-order central moment for $\mathbf{l}^T\mathbf{A}^{-1}\mathbf{z}$ in closed forms without using the confluent hypergeometric function. These results play a fundamental role in the understanding of the variation, asymmetry, and tail behavior of the estimated weights. The proof of the corollary can be found in the online supplementary materials.

Corollary 4.3: *Let $\mathbf{A} \sim \mathcal{W}_k(n, \boldsymbol{\Sigma})$, $n > k$ and $\mathbf{z} \sim \mathcal{N}_k(\boldsymbol{\mu}, \lambda\boldsymbol{\Sigma})$ with $\lambda > 0$ and positive definite $\boldsymbol{\Sigma}$. Furthermore, let $\mathbf{A}$ and $\mathbf{z}$ be independent and $\mathbf{l}$ be a k-dimensional vector of constants. Also, let $s = \boldsymbol{\mu}^T\mathbf{R}_{\mathbf{l}}\boldsymbol{\mu}/\lambda$ with $\mathbf{R}_{\mathbf{l}} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{l}\mathbf{l}^T\boldsymbol{\Sigma}^{-1}/\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\mathbf{l}$. Then*

(a) *the second-order central moment of $\mathbf{l}^T\mathbf{A}^{-1}\mathbf{z}$ is given by*

$$
\mathrm{E}[(\mathbf{l}^T\mathbf{A}^{-1}\mathbf{z} - \mathrm{E}[\mathbf{l}^T\mathbf{A}^{-1}\mathbf{z}])^2] = d_1^{(0)}(\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^2 + d_2^{(0)}\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\mathbf{l},
$$

for $n - k > 3$ with

$$
d_1^{(0)} = \frac{n-k+1}{(n-k)(n-k-1)^2(n-k-3)}, \quad d_2^{(0)} = \frac{\lambda(n-1) + \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{(n-k)(n-k-1)(n-k-3)};
$$

(b) *the third-order central moment of* $\mathbf{l}^T\mathbf{A}^{-1}\mathbf{z}$ *is given by*

$$E[(\mathbf{l}^T\mathbf{A}^{-1}\mathbf{z} - E[\mathbf{l}^T\mathbf{A}^{-1}\mathbf{z}])^3] = d_1^{(1)}(\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^3 + d_2^{(1)}\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \cdot \mathbf{l}^T\boldsymbol{\Sigma}^{-1}\mathbf{l}$$

for $n-k > 5$ *with*

$$d_1^{(1)} = \frac{16}{(n-k-1)^3(n-k-3)(n-k-5)},$$

$$d_2^{(1)} = \frac{12\lambda}{(n-k-1)^2(n-k-3)(n-k-5)}\left(1 + \frac{s+k-1}{n-k}\right);$$

(c) *the fourth-order central moment of* $\mathbf{l}^T\mathbf{A}^{-1}\mathbf{z}$ *is given by*

$$E[(\mathbf{l}^T\mathbf{A}^{-1}\mathbf{z} - E[\mathbf{l}^T\mathbf{A}^{-1}\mathbf{z}])^4] = d_1^{(3)}(\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^4 + d_2^{(3)}(\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^2\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\mathbf{l}$$
$$+ d_3^{(3)}(\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\mathbf{l})^2$$

for $n-k > 7$ *with*

$$d_1^{(2)} = \frac{3[(n-k+1)(n-k-5)(n-k-7) - (n-k-1)^2(n-k-9)]}{(n-k-1)^4(n-k-3)(n-k-5)(n-k-7)},$$

$$d_2^{(2)} = \frac{6\lambda(1+c_1)[(n-k-1)^2 - (n-k+3)(n-k-7)]}{(n-k-1)^3(n-k-3)(n-k-5)(n-k-7)},$$

$$d_3^{(2)} = \frac{3\lambda^2(1+2c_1+c_2)}{(n-k-1)(n-k-3)(n-k-5)(n-k-7)},$$

with

$$c_1 = \frac{s+k-1}{n-k} \quad \text{and} \quad c_2 = \frac{s^2 + (2s+k-1)(k+1)}{(n-k)(n-k-2)}.$$

5. Simulation studies and application

5.1. Simulation studies

The theoretical results of the paper are obtained under the assumption that the returns are independently and multivariate normally distributed. In this section, we also discuss what happens when the assumption of normality is violated. In particular, it is done numerically by simulating data from the multivariate t-distribution with 5 and 10 degrees of freedom. In what follows, the symbol $t_k(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for the k-variate t-distribution with ν degrees of freedom, the location parameter $\boldsymbol{\mu}$ and the dispersion matrix $\boldsymbol{\Sigma}$ as defined in [29, Section 2.7.2.4].

In our simulations, we put $k \in \{5, 10, 15\}$, $r_f = 0.001$ and $\mathbf{l} = \mathbf{1}_k$. The results for $k \in \{10, 15\}$ are available in the online supplementary materials. Each element of the mean vector $\boldsymbol{\mu}$ is uniformly distributed on $[-1, 1]$, and the covariance matrix $\boldsymbol{\Sigma}$ is taken to be diagonal, where each diagonal element is uniformly distributed on $[0, 1]$. For the Gaussian data, the mean and variance estimates depend on α, while it is not the case for skewness and kurtosis. In order to see how they behave for the non-Gaussian data, we consider several

values of $\alpha \in \{3, 5, 10, 50, 100\}$ and study its effect on the moments of estimated weights together with the sample size $n \in \{30, 60, 120\}$.

The simulated data consist of $N = 10^5$ independent realizations which are used to fit the corresponding moment estimators with Epanechnikov kernel. The bandwidth parameters are determined via cross-validation for every sample. Below we summarize the corresponding algorithm:

(i) Generate independently $\mathbf{x}_1, \ldots, \mathbf{x}_n$ from $t_k(\nu_i, \boldsymbol{\mu}, ((\nu_i - 2)/\nu_i)\boldsymbol{\Sigma})$, $i \in \{1, 2\}$, with $\nu_1 = 5$ and $\nu_2 = 10$;

(ii) Generate $\hat{\theta} = \mathbf{l}^T \widehat{\mathbf{w}}_{TP}$ by using

$$\hat{\theta} = \alpha^{-1} \mathbf{l}^T \mathbf{S}^{-1} (\bar{\mathbf{x}} - r_f \mathbf{1}_k),$$

where $\bar{\mathbf{x}} = (n^{-1}) \sum_{i=1}^{n} \mathbf{x}_i$ and $\mathbf{S} = (n-1)^{-1} \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$;

(iii) Repeat (i)–(ii) N times.

The results of this simulation study are presented in Table 1. It is interesting to notice that the mean and variance estimates vary for different values of risk aversion coefficient α. In general, lesser tolerance to risk leads to reduce the magnitude of expected value and variance of estimated weights. It has been observed that the distributional assumption of returns is important, especially, for the first two moments, which help to construct portfolio strategies, such as the efficient frontier curve. For all the cases, a relatively large reduction in the magnitude of the first two moment estimates has been noticed. In particular, smaller values of mean and variance can be seen for non-Gaussian returns. However, for skewness and kurtosis, the result is otherwise, and as pointed out earlier, they do not show any dependence on α, even for the non-Gaussian data. We further notice that, with the increase in sample size n and the degrees of freedom ν for the t-distributed data, the estimated moments converge to nominal values provided by the Gaussian distribution, and this finding is in accordance with the existing theory. Now, we can observe interesting behavior for the first two moments. They are very similar to the normal distribution for larger sample sizes and higher degrees of freedom in t-distribution. And it should be so according to the classical theory, i.e. where for larger degrees of freedom in t distribution, the resulting behavior should be closer to the normal distribution. The overall picture does not change much with the increase in k, see Tables 1–2 in the online supplementary materials. Furthermore, we provide the bias and MSE measures, and the 95% CIs of the estimated TP weights. The results are reported for the Gaussian and t-distributed cases in Tables 3–8 in the online supplementary materials. As can be seen, the estimator shows some biases for small n, but with the increase in sample size n and number of assets k, it starts reducing. It is interesting to point out here that relatively large bias and MSE, and wider CIs are observed for small n and large k, which further reduces with the increase in α.

5.2. Application

In this section, we present the results of the empirical study, where we show how theoretical results obtained in Section 2 can be applied to real data. We consider weekly data of $k = 4$ financial indexes[1] which are listed in NASDAQ stock exchange. Their abbreviated symbolic names are `IXTC`, `IXCO`, `TRAN`, `INDS`. The data are taken for the period from August

Table 1. Mean, variance, skewness and kurtosis of the estimated TP weights.

Risk aversion	Moments	n = 30			n = 60			n = 120		
		$\mathcal{N}_5(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$t_5(5, \boldsymbol{\mu}, 0.6\boldsymbol{\Sigma})$	$t_5(10, \boldsymbol{\mu}, 0.8\boldsymbol{\Sigma})$	$\mathcal{N}_5(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$t_5(5, \boldsymbol{\mu}, 0.6\boldsymbol{\Sigma})$	$t_5(10, \boldsymbol{\mu}, 0.8\boldsymbol{\Sigma})$	$\mathcal{N}_5(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$t_5(5, \boldsymbol{\mu}, 0.6\boldsymbol{\Sigma})$	$t_5(10, \boldsymbol{\mu}, 0.8\boldsymbol{\Sigma})$
$\alpha = 3$	Mean	0.693058	0.823531	0.739655	0.611893	0.677623	0.633911	0.578853	0.617318	0.590747
	Variance	0.442182	0.800563	0.561839	0.146369	0.260536	0.184290	0.061006	0.111257	0.076105
	Skewness	0.635559	0.705868	0.668208	0.378587	0.435739	0.402370	0.249499	0.251552	0.269590
	Kurtosis	5.248779	5.304775	5.289139	3.773318	3.902418	3.772289	3.333380	3.374525	3.335348
$\alpha = 5$	Mean	0.415835	0.492226	0.441431	0.367136	0.408328	0.380652	0.347312	0.369663	0.352586
	Variance	0.159185	0.291613	0.199524	0.052693	0.094338	0.066394	0.021962	0.040233	0.027381
	Skewness	0.635559	0.726802	0.675312	0.378587	0.445028	0.440394	0.249499	0.240287	0.255678
	Kurtosis	5.248779	5.517219	5.321095	3.773318	3.969844	3.879399	3.333380	3.358223	3.329926
$\alpha = 10$	Mean	0.207917	0.247109	0.222150	0.183568	0.204042	0.189601	0.173656	0.184632	0.176978
	Variance $\times$ 10	0.397963	0.732816	0.504198	0.131732	0.235984	0.165355	0.054905	0.100686	0.068726
	Skewness	0.635559	0.759917	0.700800	0.378587	0.438122	0.406656	0.249499	0.257948	0.274644
	Kurtosis	5.248779	5.420355	5.401659	3.773318	3.898568	3.884597	3.333380	3.407303	3.351953
$\alpha = 50$	Mean $\times$ 10	0.415835	0.492298	0.440564	0.367136	0.408150	0.380350	0.347312	0.369670	0.353839
	Variance $\times$ 10^2	0.159185	0.289685	0.200941	0.052693	0.095361	0.065758	0.021962	0.040143	0.027416
	Skewness	0.635559	0.756199	0.682990	0.378587	0.445572	0.404225	0.249499	0.235320	0.272103
	Kurtosis	5.248779	5.789996	5.379279	3.773318	3.915872	3.795098	3.333380	3.372282	3.313795
$\alpha = 100$	Mean $\times$ 10	0.207917	0.246221	0.221592	0.183568	0.203620	0.189536	0.173650	0.184987	0.176462
	Variance $\times$ 10^3	0.397963	0.726510	0.507236	0.131732	0.237021	0.164691	0.054905	0.100726	0.068846
	Skewness	0.635559	0.740576	0.715550	0.378587	0.416021	0.413437	0.249499	0.262654	0.265099
	Kurtosis	5.248779	5.391738	5.712833	3.773318	3.830333	3.841699	3.333380	3.381111	3.327772

Note: The returns are assumed to be independently multivariate normally and t-distributed. k is taken to be 5, and $\mathbf{l} = \mathbf{1}_k$.
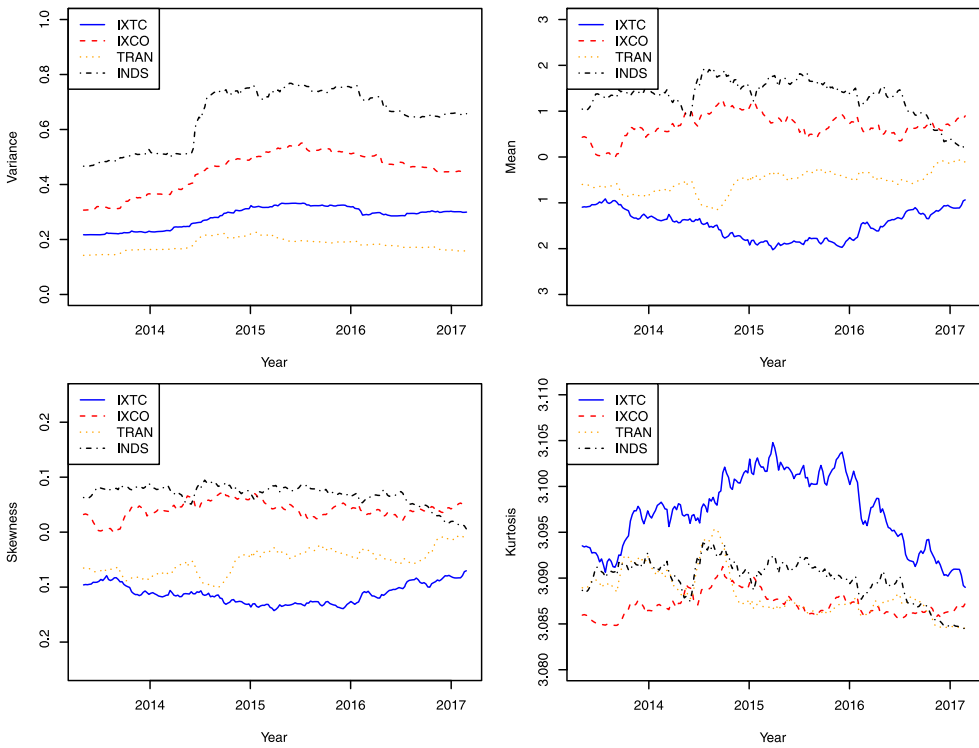
Figure 1. The rolling estimators for the mean (top-left), variance (top-right), skewness (bottom-left) and kurtosis (bottom-right) of four financial indexes with the estimation window of 300 weeks and $\alpha = 3$.

2007 to April 2017. Weekly log returns on each index have been considered, due to the fact that they usually follow the Gaussian distribution. The weekly log returns on the three-month US treasury bill are used as the risk-free rate. The risk aversion coefficient α is taken as 3.

Figure 1 presents the dynamic behavior of mean, variance, skewness, and kurtosis for the estimated TP weights by using the rolling estimator with the estimation window of 300 weeks, i.e. $n = 300$. We observe that mean for all the indices shows a noticeable dynamic behavior. More specifically, for IXTC and TRAN we observe that expected values are negative throughout the sample period, which indicates short selling for these indices. While for TRAN, INDS indices, positive expected values can be seen for almost all the sample period.

From the top-right plot of variance, a time-varying structure can be seen for all stock indices. A shift in variance can be observed after mid-2014 which seems to settle down to its original position by the end of the sample period.

The bottom-left plot of skewness displays almost similar behavior as for the mean plot for all the stock markets and do not deviate significantly from zero (the nominal skewness for the normally distributed data). However, a minimal negative skewness is observed throughout the sample period for IXTC and TRAN indices. Finally, the bottom-right plot of kurtosis shows decent variation for most of the stocks, except for IXTC which shows a

relatively high variation. However, the overall dynamic revolves around the nominal kurtosis for a normal distribution. Through this empirical exercise, we confirm that there exists a reasonable time-varying behavior in the moments of estimated TP weights, but eventually, they can be nicely approximated by the normal distribution for a large sample size.

6. Conclusions

In this paper, we studied higher order moments of the estimated TP weights obtained under the assumption of normally and independently distributed returns. In particular, we derived the higher order non-central and central moments of estimated weights that depend on the confluent hypergeometric function. Moreover, we provided the expressions for the mean, variance, skewness and kurtosis in closed forms without using the confluent hypergeometric function. The results are supported by a simulation study where data from normal and the multivariate t-distributions have been simulated and moments of the estimated TP weights have been evaluated by using a Monte Carlo experiment. The investor's attitude towards risk influences the portfolio strategy and can be displayed through, such as the efficient frontier .[2] Through this simulation study, we noticed a sharp decline in the mean and variance of estimated weights with high-risk aversion parameter. The skewness and kurtosis, however, remain almost unchanged with respect to the varying nature of the risk aversion parameter, which indicates that tolerance to risk does not derive the tail risk of estimated weights. For a small sample, the values of skewness and kurtosis of estimated weights show some deviation from the normal distribution. However, it is observed that the estimated weights can be well approximated by a normal distribution for a large sample size. Additionally, we studied the behavior of Bias, MSE and CIs of the sample estimator of TP weights. Bias and MSE are found to be relatively large for small sample size and small risk aversion, while they are significantly reduced for large sample size. These results are available in the online supplementary materials.

Through the empirical study for four financial indexes listed in the NASDAQ stock exchange, we obtained first four moments' expressions aiming to observe the presence of time-varying behavior. For some stocks, we observed that expected values are negative throughout the sample period, which indicates short selling for these indices. A reasonable time-varying structure is observed for a variance with a few relatively high values under the sample period. While the skewness and kurtosis revolve around their average values since the sample size is taken to be relatively large.

Notes

1. Note that the number of indexes $k = 4$ is used here for illustration purposes, and similar results can easily be obtained for any k such that $k < n$.
2. It also holds for a continuous-time Merton's portfolio (see [40,41])

Acknowledgments

Disclosure statement

Funding

References

[1] M. Abramowitz and I.A. Stegun, *Pocketbook of Mathematical Functions*, Verlag Harri Deutsch, Frankfurt (Main), 1984.

[2] C. Acerbi and D. Tasche, *Expected shortfall: A natural coherent alternative to value at risk*, Econ. Notes 31 (2002), pp. 379–388.

[3] G.J. Alexander and A.M. Baptista, *Economic implications of using a mean-VaR model for portfolio selection: A comparison with mean-variance analysis*, J. Econ. Dyn. Control 26 (2002), pp. 1159–1193.

[4] G.J. Alexander and A.M. Baptista, *A comparison of VaR and CVaR constraints on portfolio selection with the mean-variance model*, Manage. Sci. 50 (2004), pp. 1261–1273.

[5] D. Bauder, T. Bodnar, S. Mazur, and Y. Okhrin, *Bayesian inference for the tangent portfolio*, Int. J. Theor. Appl. Finance 21 (2018), pp. 1–27.

[6] O. Bodnar, *Sequential surveillance of the tangency portfolio weights*, Int. J. Theor. Appl. Finance 12 (2009), pp. 797–810.

[7] T. Bodnar, A.K. Gupta, and N. Parolya, *Direct shrinkage estimation of large dimensional precision matrix*, J. Multivar. Anal. 146 (2016), pp. 223–236.

[8] T. Bodnar, S. Mazur, S. Muhinyuza, and N. Parolya, *On the product of a singular Wishart matrix and a singular Gaussian vector in high dimensions*, Theory Probab. Math. Statist. 99 (2018), pp. 37–50.

[9] T. Bodnar, S. Mazur, E. Ngailo, and N. Parolya, *Discriminant analysis in small and large dimensions*, Theory Probab. Math. Statist. 100 (2019), pp. 24–42.

[10] T. Bodnar, S. Mazur, and Y. Okhrin, *On the exact and approximate distributions of the product of a Wishart matrix with a normal vector*, J. Multivar. Anal. 122 (2013), pp. 70–81.

[11] T. Bodnar, S. Mazur, and Y. Okhrin, *Distribution of the product of a singular Wishart matrix and a normal vector*, Theory Probab. Math. Statist. 91 (2014), pp. 1–15.

[12] T. Bodnar, S. Mazur, and N. Parolya, *Central limit theorems for functionals of large sample covariance matrix and mean vector in matrix-variate location mixture of normal distributions*, Scand. J. Statist. 46 (2019), pp. 636–660.

[13] T. Bodnar, S. Mazur, and K. Podgórski, *Singular inverse Wishart distribution and its application to portfolio theory*, J. Multivar. Anal. 143 (2016), pp. 314–326.

[14] T. Bodnar, S. Mazur, and K. Podgórski, *A test for the global minimum variance portfolio for small sample and singular covariance*, AStA Adv. Stat. Anal. 101 (2017), pp. 253–265.

[15] T. Bodnar, S. Mazur, K. Podgórski, and J. Tyrcha, *Tangency portfolio weights for singular covariance matrix in small and large dimensions: Estimation and test theory*, J. Stat. Plan. Inference 201 (2019), pp. 40–57.

[16] T. Bodnar and Y. Okhrin, *On the product of inverse Wishart and normal distributions with applications to discriminant analysis and portfolio theory*, Scand. J. Statist. 38 (2011), pp. 311–331.

[17] T. Bodnar, O. Okhrin, and N. Parolya, *Optimal shrinkage estimator for high-dimensional mean vector*, J. Multivar. Anal. 170 (2019), pp. 63–79.

[18] T. Bodnar, W. Schmid, and T. Zabolotskyy, *Asymptotic behavior of the estimated weights and of the estimated performance measures of the minimum VaR and the minimum CVaR optimal portfolios for dependent data*, Metrika 76 (2013), pp. 1105–1134.

[19] T. Bodnar and T. Zabolotskyy, *How risky is the optimal portfolio which maximizes the Sharpe ratio?*, AStA Adv. Stat. Anal. 101 (2017), pp. 1–28.

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

[21] J. Brodie, I. Daubechies, C. DeMol, D. Giannone, and I. Loris, *Sparse and stable Markowitz portfolios*, Proc. Natl. Acad. Sci. USA 106 (2009), pp. 12267–12272.

[22] J. Bun, J.P. Bouchaud, and M. Potters, *Cleaning large correlation matrices: Tools from random matrix theory*, Phys. Rep. 666 (2017), pp. 1–109.

[23] P. Carr and L. Wu, *Variance risk premiums*, Rev. Financ. Stud. 22 (2009), pp. 1311–1341.

[24] E. Chernousova and Yu. Golubev, *Spectral cut-off regularizations for ill-posed linear models*, Math. Methods Statist. 23 (2014), pp. 116–131.

[25] B. Efron, *Minimum volume confidence regions for a multivariate normal mean vector*, J. R. Stat. Soc. Ser. B 68 (2006), pp. 655–670.

[26] J. Fan, Y. Liao, and H. Liu, *An overview of the estimation of large covariance and precision matrices*, Econom. J. 19 (2016), pp. C1–C32.

[27] N. Gandelman and R. Hernandez-Murillo, *Risk aversion at the country level*, Review 97 (2015), pp. 53–66.

[28] M. Gulliksson and S. Mazur, *An iterative approach to ill-conditioned Optimal Portfolio Selection*, Comput. Econ. (2019). doi:10.1007/s10614-019-09943-6.

[29] A.K. Gupta, D.K. Nagar, and T. Bodnar, *Elliptically Contoured Models in Statistics and Portfolio Theory*, Springer, New York, 2013.

[30] C.R. Harvey, J.C. Liechty, M.W. Liechty, and P. Müller, *Portfolio selection with higher moments*, Quant. Finance 10 (2010), pp. 469–485.

[31] J.E. Ingersoll, *Theory of Financial Decision Making*, Rowman & Littlefield Publishers, Savage, 1987.

[32] F. Javed, S. Mazur, and E. Ngailo, *Higher order moments of the estimated tangency portfolio weights*, Working Paper, Örebro University, Örebro, 2017, pp. 1–18.

[33] E. Jurczenko and B. Maillet, *Theoretical foundations of asset allocation and pricing models with higher-order moments*, in *Multi-moment Asset Allocation and Pricing Models*, E. Jurczenko and B. Maillet, eds., Wiley & Sons, 2015.

[34] I. Kotsiuba and S. Mazur, *On the asymptotic and approximate distributions of the product of an inverse Wishart matrix and a Gaussian random vector*, Theory Probab. Math. Statist. 93 (2015), pp. 95–104.

[35] R. Kress, *Linear Integral Equations*, Springer, New York, 1999.

[36] Y. Kroll, H. Levy, and H. Markowitz, *Mean-variance versus direct utility maximization*, J. Finance 39 (1984), pp. 47–61.

[37] V.P. Leonov and A.N. Shiryaev, *On a method of calculation semi-invariants*, Theory Probab. Appl. 4 (1959), pp. 319–329.

[38] H. Levy and H. Markowitz, *Approximating expected utility by a function of mean and variance*, Am. Econ. Rev. 69 (1979), pp. 308–317.

[39] H. Markowitz, *Portfolio selection*, J. Finance 7 (1952), pp. 77–91.

[40] R. Merton, *Lifetime portfolio selection under uncertainty: The continuous-time case*, Rev. Econ. Stat. 51 (1969), pp. 247–257.

[41] R. Merton, *Optimum consumption and portfolio rules in a continuous-time model*, J. Econ. Theory 3 (1971), pp. 373–413.

[42] S. Muhinyuza, T. Bodnar, and M. Lindholm, *A test on the location of the tangency portfolio on the set of feasible portfolios*, Research Report 26, Stockholm University, 2017.

[43] R.J. Muirhead, *Aspects of Multivariate Statistical Theory*, Wiley, New York, 1982.

[44] Y. Okhrin and W. Schmid, *Distributional properties of portfolio weights*, J. Econom. 134 (2006), pp. 235–256.

[45] A.N. Tikhonov and V.Y. Arsenin, *Solutions of Ill-Posed Problems*, Winston, New York, 1977.

[46] J. Tu and G. Zhou, *Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies*, J. Financ. Econ. 99 (2011), pp. 204–215.

[47] Z. Wang, *A shrinkage approach to model uncertainty and asset allocation*, Rev. Financ. Stud. 18 (2005), pp. 673–705.
[48] C.S. Withers and S. Nadarajah, *Moments from cumulants and vice versa*, Int. J. Math. Educ. Sci. Technol. 40 (2009), pp. 842–845.

Appendix

Here we collect all the proofs of our main results obtained in Section 2.

Proof of Theorem 2.1.**:** From [43, Chapter 3], we have that

$$\bar{\mathbf{x}} \sim \mathcal{N}_k \left(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma} \right) \quad \text{and} \quad \mathbf{V} := (n-1)\mathbf{S} \sim \mathcal{W}_k(n-1, \boldsymbol{\Sigma});$$

moreover, $\bar{\mathbf{x}}$ and $\mathbf{V}$ are independently distributed. Since

$$\hat{\theta} = \mathbf{l}^{\mathrm{T}}\widehat{\mathbf{w}}_{TP} = \alpha^{-1}\mathbf{l}^{\mathrm{T}}\mathbf{S}^{-1}(\bar{\mathbf{x}} - r_f\mathbf{1}_k) = \frac{n-1}{\alpha}\mathbf{l}^{\mathrm{T}}\mathbf{V}^{-1}(\bar{\mathbf{x}} - r_f\mathbf{1}_k),$$

the rest of the proof follows from Theorem 4.1 and Corollary 4.3. ■

Proof of Corollary 2.2.**:** From Corollary 4.3, we get the first two moments of $\hat{\theta}$ which are given by

$$\mathrm{E}[\hat{\theta}] = \frac{n-1}{n-k-2}\theta \quad \text{and} \quad \mathrm{Var}[\hat{\theta}] = \breve{d}_1^{(0)}\theta^2 + \breve{d}_2^{(0)}\alpha^{-2}\mathbf{l}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{l}$$

with $\breve{d}_1^{(0)}$ and $\breve{d}_2^{(0)}$ which are the same as in the formulation of the corollary.
 Moreover, since $\mathbf{l}$ is an arbitrary vector of constants, we get the statement of the corollary. ■

Proof of Corollary 2.3.**:** The skewness of $\hat{\theta}$ is given by

$$\mathrm{Skewness}[\hat{\theta}] = \frac{\overline{\mu}_3}{\left[\mathrm{Var}(\hat{\theta})\right]^{3/2}} \frac{\overline{\mu}_3}{\left[\mathrm{Var}(\hat{\theta})\right]^{3/2}} = \overline{\mu}_3 \left(\breve{d}_1^{(0)}\theta^2 + \breve{d}_2^{(0)}\alpha^{-2}\mathbf{l}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{l} \right)^{-3/2},$$

where $\mathrm{Var}(\hat{\theta})$ is obtained from Corollary 2.3. From Corollary 4.3, it follows that

$$\overline{\mu}_3 = \tilde{d}_1^{(1)}(\mathbf{l}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\breve{\boldsymbol{\mu}})^3 + \tilde{d}_2^{(1)}\mathbf{l}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\breve{\boldsymbol{\mu}} \cdot \mathbf{l}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{l},$$

where

$$\tilde{d}_1^{(1)} = \frac{16(n-1)^3}{\alpha^3(n-k-2)^3(n-k-4)(n-k-6)}$$

$$\tilde{d}_2^{(1)} = \frac{12(n-1)^3}{\alpha^3 n(n-k-2)^2(n-k-4)(n-k-6)} \left(1 + \frac{\breve{s}+k-1}{n-k-1} \right)$$

with $\breve{s} = n\breve{\boldsymbol{\mu}}^{\mathrm{T}}\mathbf{R}_{\mathbf{l}}\breve{\boldsymbol{\mu}}$ and $\mathbf{R}_{\mathbf{l}} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{l}\mathbf{l}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}/\mathbf{l}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{l}$. Furthermore, $\overline{\mu}_3$ can be rewritten in the next form

$$\overline{\mu}_3 = \alpha^{-3} \left(\breve{d}_1^{(1)}\breve{\theta}^3 + \breve{d}_2^{(1)}\breve{\theta}\mathbf{l}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{l} \right),$$

where $\breve{d}_1^{(1)}$ and $\breve{d}_2^{(1)}$ are the same as in the formulation of the corollary. Putting all above together we get the skewness of $\hat{\theta}$.

We later move on and derive the explicit formula for the kurtosis of $\hat{\theta}$. It holds that

$$\text{Kurtosis}[\hat{\theta}] = \frac{\overline{\mu}_4}{\left[\text{Var}(\hat{\theta})\right]^2} = \overline{\mu}_4 \left(\breve{d}_1^{(0)}\theta^2 + \breve{d}_2^{(0)}\alpha^{-2}\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\mathbf{l}\right)^{-2}.$$

Using Corollary 4.3, we obtain

$$\overline{\mu}_4 = \tilde{d}_1^{(2)}(\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\breve{\boldsymbol{\mu}})^4 + \tilde{d}_2^{(2)}(\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\breve{\boldsymbol{\mu}})^2\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\mathbf{l} + \tilde{d}_3^{(2)}(\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\mathbf{l})^2,$$

where

$$\tilde{d}_1^{(2)} = \frac{3(n-1)^4[(n-k)(n-k-6)(n-k-8) - (n-k-2)^2(n-k-10)]}{\alpha^4(n-k-2)^4(n-k-4)(n-k-6)(n-k-8)},$$

$$\tilde{d}_2^{(2)} = \frac{6(1+\breve{c}_1)(n-1)^4[(n-k-2)^2 - (n-k+2)(n-k-8)]}{\alpha^4 n(n-k-2)^3(n-k-4)(n-k-6)(n-k-8)},$$

$$\tilde{d}_3^{(2)} = \frac{3(1+2\breve{c}_1+\breve{c}_2)(n-1)^4}{\alpha^4 n^2(n-k-2)(n-k-4)(n-k-6)(n-k-8)},$$

with

$$\breve{c}_1 = \frac{\breve{s}+k-1}{n-k-1} \quad \text{and} \quad \breve{c}_2 = \frac{\breve{s}^2 + (2\breve{s}+k-1)(k+1)}{(n-k-1)(n-k-3)}.$$

Moreover, $\overline{\mu}_4$ can be rewritten as

$$\overline{\mu}_4 = \alpha^{-4}\left(\breve{d}_1^{(2)}\breve{\theta}^4 + \breve{d}_2^{(2)}\breve{\theta}^2\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\mathbf{l} + \breve{d}_3^{(2)}(\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\mathbf{l})^2\right),$$

where $\breve{d}_1^{(3)}$, $\breve{d}_2^{(3)}$, and $\breve{d}_3^{(3)}$ are the same as in the formulation of the corollary. It completes the proof of the corollary. ∎