WORKING PAPER 13/2021 (STATISTICS)

Tangency portfolio weights under a skew-normal model in small and large dimensions

Farrukh Javed, Stepan Mazur and Erik Thorsén

ISSN 1403-0586

Örebro University School of Business
SE-701 82 Örebro, Sweden

Tangency portfolio weights under a skew-normal model in small and large dimensions

Farrukh Javed[(a)], Stepan Mazur[(a)] and Erik Thorsén[(b)]
[(a)]Unit of Statistics, Örebro University School of Business, SE-70182 Örebro, Sweden
[(b)]Department of Mathematics, Stockholm University, SE-10691 Stockholm, Sweden

ARTICLE HISTORY
Compiled June, 2021

ABSTRACT
In this paper, we investigate the distributional properties of the estimated tangency portfolio (TP) weights assuming that the asset returns follow a matrix variate closed skew-normal distribution. We establish a stochastic representation of the linear combination of the estimated TP weights that fully characterize its distribution. Using the stochastic representation we derive the mean and variance of the estimated weights of TP which are of key importance in portfolio analysis. Furthermore, we provide the asymptotic distribution of the linear combination of the estimated TP weights under the high-dimensional asymptotic regime, i.e. the dimension of the portfolio p and the sample size n tend to infinity such that $p/n \to c \in (0, 1)$. A good performance of the theoretical findings is documented in the simulation study. In the empirical study, we apply the theoretical results to real data of the stocks included in the S&P 500 index.

KEYWORDS
Asset allocation, high-dimensional asymptotics, matrix variate skew-normal distribution, stochastic representation, tangency portfolio

1. Introduction

Modern portfolio theory, introduced by Markowitz (1952), has been a pinnacle of investment theory since its introduction in the 1950s. The problems posed by Markowitz aim to find a portfolio that is characterized by the investor's belief in risk and return. In an optimal fashion, the obtained portfolio is then used by the investor to allocate her/his wealth. Tangency portfolio (TP) is one of these optimal portfolios which determines how an investor should allocate the wealth between the risk-free rate and some risky assets. When an investor tries to construct a portfolio, either there is a need to specify all the parameters in the portfolio allocation procedure or to make use of data to estimate them. We believe that the latter is more common than the former and by doing so we introduce estimation uncertainty into the allocation process. This uncertainty is paramount to quantify for the investors since their expectations might not match with what the portfolio can deliver. It is also essential to communicate to stakeholders and assert compliance to regulatory frameworks which makes analytic results more compelling as these are easy to reason about.

The implications of estimation uncertainty in modern portfolio theory, in general, and in TP, in particular, have been extensively researched in the literature. The research dates back as far as late 90s, see e.g. Britten-Jones (1999), where the statistical test for the TP weights is derived. Okhrin and Schmid (2006) continued on this path and derived the asymptotic distribution for the portfolio weights. The moments of the TP weights were then later characterized by Kan and Zhou (2007) under the assumption of normally distributed returns. Bodnar and Okhrin (2011) derived statistical tests for the composite hypothesis of the TP weights and Kotsiuba and Mazur (2016) approximated the density using Taylor expansion. Characterising uncertainty fits well into the Bayesian framework and along this line of research, in Bauder et al. (2018) distributional properties of the TP weights are studied.[1]

Recent literature has continued to build upon the previously mentioned works. Bodnar et al. (2019) extended the works of Okhrin and Schmid (2006) and Bodnar and Okhrin (2011) to the scenario when both the population and the sample covariance matrices are singular. This can be seen as a high-dimensional setting where the sample size is comparable, or smaller, than the portfolio size. Recently this topic has gained a lot of attention and a large number of different approaches have been taken to construct statistical tests, characterizing the distribution of the TP weights and functions thereof. Bodnar et al. (2021) derived the distribution in small and large dimension for a large class of portfolios, including the normalised TP weights. In Muhinyuza et al. (2020) and Muhinyuza (2020) the statistical test for the TP in small and large dimension is derived to deduce whether the portfolio is efficient or not. Karlsson et al. (2020) delivered the high-dimensional asymptotic distribution of the estimated TP weights and high-dimensional asymptotic test on the linear combination of the elements of TP weights. Javed et al. (2021) obtained analytical expressions for the higher order moments of the estimated TP weights. In Alfelt and Mazur (2020), the mean and variance of the estimated TP weights are studied when the sample covariance matrix is singular.

When the influence of uncertainty is to be understood in a finite-sample setting, there is usually a need for a statistical model to account for it. The chosen model should take into account the characteristics of asset returns, which are usually known as the *stylized facts* (see e.g. Cont (2001)). One such characteristic is skewness, which is quite often present in low-frequency data, such as weekly, monthly or quarterly,

[1]In Bayesian framework, the posterior distribution of the TP weights is proportional to the product of (singular) Wishart matrix and (singular) normal vector under the assumption of normally distributed data. The distributional properties of these products are well studied by Bodnar et al. (2013, 2014), Bodnar et al. (2018), Bodnar et al. (2019).

and has been documented in the literature (see e.g. Kraus and Litzenberger (1976), Alles and Kling (1994) or Peiro (1999)). Following this line of research, Bodnar and Gupta (2015) incorporated skewness into a portfolio allocation problem through the Closed Skew-Normal (CSN) model. This model can incorporate many different aspects of asset returns due to its flexible parameter structure.

Along this direction, in this article, we consider the CSN model for assets returns with a focus on TP. Our contribution to the existing literature is as follows. First, we derive a stochastic representation of the linear combination of the TP weights that is a computationally effective tool for studying distributional properties. Second, we deliver closed-form expressions for the mean and variance of the TP weights. The moments are vital for quickly understanding the implications of estimation uncertainty in a portfolio. A plug-in or sample version of the TP is one realisation from its distribution. The moments would help quantify the overall uncertainty in the point estimate. Third, we obtain the asymptotic distribution of the linear combination of the TP weights under a high-dimensional asymptotic regime, i.e. both portfolio size p and sample size n tend to infinity such that $p/n \to c \in (0, 1)$. There have been a large number of portfolios derived to constrain higher order moments of the portfolio. For the interested reader, we recommend Harvey et al. (2010) and the references therein. However, one can see this as an entirely different problem since it aims to constrain the portfolio choice allocation problem. Our contribution helps investors to understand the influence of skewness, and not constrain it.

This paper is organized as follows. In Section 2, we briefly introduce the CSN model. Section 3 provides a framework for TP within the domain of the CSN model together with numerous results for its sample counterpart in small and high dimension. In Section 4, we study the performance of the theoretical results through simulation and empirical studies. We finish the paper with discussion in Section 5.

2. Skew-normal model

In this section, we will briefly present the matrix-variate closed skew-normal (CSN) distribution and discuss its properties, especially in connection to asset returns. For this, we need some notation, with which we start first. Let $\mathbf{1}_k$ denotes a k-dimensional vector of ones, while $\mathbf{0}_k$ stands for a k-dimensional vector of zeros. We would note that the vectors in this work are identified with single-column matrices. Let $\mathbf{I}_k$ stands for the identity matrix of size k. The symbol $\otimes$ stands for the Kronecker product and $\text{vec}(\boldsymbol{A})$ denotes the `vec` operator, which vectorizes a matrix through stacking its columns, e.g. if $\boldsymbol{A}$ is a $k \times p$ matrix then $\text{vec}(\boldsymbol{A}) = (a_{11}, \ldots, a_{k1}, a_{12}, \ldots, a_{k2}, \ldots, a_{1p}, \ldots, a_{kp})^\top$. Furthermore, let $\stackrel{d}{=}$ denotes the equality in distributions. Finally, let $\boldsymbol{A} \succ 0$ implies that $\boldsymbol{A}$ is symmetric and positive definite.

Let

$$\mathbf{X} = \begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{np} \end{pmatrix} = (\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_p) = \begin{pmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix}$$

be the $n \times p$ observation matrix of the asset returns, where each observation $\tilde{\mathbf{x}}_i$ is a n-dimensional vector of returns for the i:th asset, $i = 1, \dots, p$, while $\mathbf{x}_t$ is a p-dimensional vector of the asset returns at time point t, $t = 1, \dots, n$.

Throughout the paper we assume that the matrix $\mathbf{X}$ is random and follows a matrix-variate CSN distribution (see e.g. Domínguez-Molina et al. (2007)), denoted by $\mathbf{X} \sim \mathcal{CSN}_{n,p;1,1}(\mathbf{1}_n \otimes \boldsymbol{\mu}^\top, \mathbf{I}_n \otimes \boldsymbol{\Sigma}, \mathbf{1}_n^\top \otimes \mathbf{e}^\top, 0, v)$, where $\boldsymbol{\mu}$ is a p-dimensional vector, $\boldsymbol{\Sigma}$ is a $p \times p$ symmetric positive definite matrix, $\mathbf{1}_n$ is a n-dimensional vector, $\mathbf{e}$ is a p-dimensional vector and v is a strictly positive number.

To discuss the parameters in detail, the density function of $\mathbf{X}$ is expressed in terms of $\mathrm{vec}(\mathbf{X}^\top)$ and is given by

$$f_{\mathrm{vec}(\mathbf{X}^\top)}(\mathbf{y}) = 2\phi_{np}(\mathbf{y}; \mathrm{vec}(\mathbf{1}_n^\top \otimes \boldsymbol{\mu}), \boldsymbol{\Sigma} \otimes \mathbf{I}_n)\Phi((\mathbf{e}^\top \otimes \mathbf{1}_n^\top)(\mathbf{y} - \mathrm{vec}(\mathbf{1}_n^\top \otimes \boldsymbol{\mu})); 0, v), \tag{2.1}$$

where $\phi_k(\mathbf{y}; \mathbf{m}, \boldsymbol{A})$ and $\Phi_k(\mathbf{y}; \mathbf{m}, \boldsymbol{A})$ stand for the density function and cumulative distribution function, respectively, of the k-dimensional normal distribution with mean $\mathbf{m}$ and covariance matrix $\boldsymbol{A}$. The density function of $\mathrm{vec}(\mathbf{X}^\top)$ presented in (2.1) corresponds to the density function of multivariate skew-normal distribution considered by Arellano-Valle and Azzalini (2006). If there is no skewness, i.e. if $\mathbf{e} = \mathbf{0}_p$, then it leads us to the matrix-variate normal distribution with mean matrix $\mathbf{1}_n \otimes \boldsymbol{\mu}^\top$ and covariance matrix $\mathbf{I}_n \otimes \boldsymbol{\Sigma}$ denoted by $\mathcal{N}_{n,p}(\mathbf{1}_n \otimes \boldsymbol{\mu}^\top, \mathbf{I}_n \otimes \boldsymbol{\Sigma})$. The parameter v is harder to reason about from the density since it is part of the CDF of the normal distribution. Due to Proposition 2.1 Domínguez-Molina et al. (2007) the CSN distribution can also be written by a stochastic representation as

$$\mathrm{vec}(\mathbf{X}^\top) \overset{d}{=} \mathrm{vec}(\mathbf{1}_n^\top \otimes \boldsymbol{\mu}) + \left((\mathbf{I}_n \otimes \boldsymbol{\Sigma})^{-1} + \frac{(\mathbf{1}_n \otimes \mathbf{e})(\mathbf{1}_n^\top \otimes \mathbf{e}^\top)}{v} \right)^{-1/2} \mathbf{z} + \frac{\mathbf{1}_n \otimes \boldsymbol{\Sigma}\mathbf{e}}{\sqrt{v + n\mathbf{e}^\top \boldsymbol{\Sigma}\mathbf{e}}} |z_0|, \tag{2.2}$$

where $\mathbf{z} \sim \mathcal{N}_{np}(\mathbf{0}_{np}, \mathbf{I}_{np})$ and $z_0 \sim \mathcal{N}(0, 1)$; moreover, $\mathbf{z}$ and z_0 are independent. The stochastic representation makes it easier to reason about v. The random variable z_0 is latent and represents the distortions that the considered data gets to experience. From the stochastic representation, we can actually see that the parameter v represents the absence of the distortions. The larger it becomes, the smaller the skewness will be.

As mentioned earlier, the CSN distribution is known to capture some of the dynamics and stylized facts that asset returns are known to exhibit. Specifically, the parameter vector $\mathbf{e}$ takes care of the skewness of the asset returns and its influence is present in both the mean and the variance. The larger it becomes the more dispersed the values of $\mathbf{X}$ will be. Depending on the sign of $\mathbf{e}$ the mean will also change accordingly.

The flexibility of the matrix-variate CSN distribution can be seen from these parameters. Given that our asset return distribution follows the stochastic representation (2.2) we have from Bodnar and Gupta (2015, Section 2), that the covariance for n different observations of ith asset class $\tilde{\mathbf{x}}_i$ is equal to

$$\mathrm{Cov}[\tilde{\mathbf{x}}_i] = \sigma_{ii}\mathbf{I}_n - \frac{2}{\pi}\frac{(\mathbf{e}^\top\mathbf{\Sigma}\mathbf{a}_{p;i})^2}{\sigma_{ii}^2(v + n\mathbf{e}^\top\mathbf{\Sigma}\mathbf{e})}\mathbf{1}_n\mathbf{1}_n^\top,$$

where σ_{ii} is the ith diagonal element of $\mathbf{\Sigma}$ and $\mathbf{a}_{p;i} = (0, \ldots, 0, 1, 0, \ldots, 0)^\top$. The non-diagonal elements of the matrix $\mathrm{Cov}[\tilde{\mathbf{x}}_i]$ are non-zero and, therefore, the elements of $\tilde{\mathbf{x}}_i$ are dependent. For our special case of the matrix-variate CSN distribution, the model is stationary, since

$$\begin{aligned}
\mathrm{E}[\mathbf{x}_t] &= \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}(v + n\mathbf{e}^\top\mathbf{\Sigma}\mathbf{e})^{-1/2}\mathbf{e}, \\
\mathrm{Var}[\mathbf{x}_t] &= \mathbf{\Sigma} - \frac{2}{\pi}\frac{1}{v + n\mathbf{e}^\top\mathbf{\Sigma}\mathbf{e}}\mathbf{e}\mathbf{e}^\top,
\end{aligned}$$

are independent of time.

3. Tangency portfolio under the skew normality

In this section, we will construct the TP as well as derive its properties using the statistical model described in Section 2. Let $\mathbf{x}$ denotes the vector of asset returns with mean vector $\boldsymbol{\eta}$ and covariance matrix $\boldsymbol{\Psi}$, while r_f stands for the return on the risk-free asset, which can be the interest rate of a risk-free bond or any other risk-free contract. Let $\mathbf{w} = (w_1, w_2, \ldots, w_p)$ denotes any vector of portfolio weights where each element w_i represents the amount allocated in the i:th asset.

Mean-variance portfolios have been extensively studied since their introduction in Markowitz (1952). The solution to these classical portfolio selection problems is optimal in the sense that we can not expect to receive more return without accepting more risk or vice versa. Many portfolio allocation problems give solutions which are optimal in the same sense (see e.g. Bodnar et al. (2013)). However, one specific portfolio is able to attain all mean-variance efficient portfolios, namely the solution to maximizing the expected quadratic utility. Assuming that the investor wants to use the expected quadratic utility to optimize the portfolio and include the risk-free asset, then the investor will end up with what is known as the TP. The precense of a risk-free asset r_f, implies that any portfolio can be obtained by borrowing or placing a large enough portion on the risk-free asset. To see this, consider the common portfolio constraint $w_0 + \mathbf{w}^\top \mathbf{1}_p = 1$, where w_0 is the amount allocated in the risk-free rate. The constraint can be removed by considering $w_0 = 1 - \mathbf{w}^\top \mathbf{1}_p$ which implies that the portfolio return distribution will be given by $x_p = w_0 r_f + \mathbf{w}^\top \mathbf{x} = \mathbf{w}^\top(\mathbf{x} - r_f \mathbf{1}_p) + r_f$. Investing nothing in the market, e.g. $\mathbf{w} = \mathbf{0}$, results in the investor receiving the risk-free rate as the portfolio return. Consider now the expected payoff for such a portfolio, which is equal to $\mu_p = \mathbf{w}^\top(\boldsymbol{\eta} - r_f \mathbf{1}_p) + r_f$. The mean of the assets $\boldsymbol{\eta}$ are discounted according to the risk-free rate. Risk in this scenario is then measured by the portfolio variance, $\sigma_p^2 = \mathbf{w}^\top \boldsymbol{\Psi} \mathbf{w}$.

Using the expected quadratic utility, the portfolio can be obtained through the following unconstrained optimization problem

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

where $\alpha > 0$ is the risk aversion coefficient, representing investor's risk profile. A large risk aversion represents a risk averse investor. The solution is given by

$$\mathbf{w}_{TP} = \alpha^{-1} \boldsymbol{\Psi}^{-1} (\boldsymbol{\eta} - r_f \mathbf{1}_p).$$

In this paper, we focus on the linear combination of the TP weights expressed as

$$\theta := \mathbf{l}^\top \mathbf{w}_{TP} = \alpha^{-1} \mathbf{l}^\top \boldsymbol{\Psi}^{-1} (\boldsymbol{\eta} - r_f \mathbf{1}_p) \tag{3.1}$$

where $\mathbf{l}$ is a p-dimensional vector of constants. Since both parameters $\boldsymbol{\eta}$ and $\boldsymbol{\Psi}$ are unknown in practice, they need to be estimated from historical data. The most common estimators are sample mean vector and sample covariance matrix that are given by

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{t=1}^{n} \mathbf{x}_t = \frac{1}{n} \mathbf{X}^\top \mathbf{1}_n, \tag{3.2}$$

$$\mathbf{S} = \frac{1}{n-1} \sum_{t=1}^{n} (\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_t - \bar{\mathbf{x}})^\top = \frac{1}{n-1} \mathbf{X}^\top \mathbf{V} \mathbf{X}, \tag{3.3}$$

where $\mathbf{V} = \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top$ is a symmetric idempotent matrix, i.e. $\mathbf{V} = \mathbf{V}^\top$ and $\mathbf{V} = \mathbf{V}^2$. The sample estimator of θ can then be expressed as

$$\hat{\theta} := \mathbf{l}^\top \hat{\mathbf{w}}_{TP} = \alpha^{-1} \mathbf{l}^\top \mathbf{S}^{-1} (\bar{\mathbf{x}} - r_f \mathbf{1}_p). \tag{3.4}$$

The choice of $\mathbf{l}$ can represent the investors' preference and interest in the portfolio. It can be used to understand what kind of contribution a certain asset has to the portfolio performance or what performance one might achieve by excluding (or including) a certain asset. It can also be used to examine future portfolio performance when $\mathbf{l}$ is the log-returns at time $n+1$. This can include stress testing the portfolio when $\mathbf{l}$ follows a distribution which might be different from the current. This scenario requires some additional attention as the vector $\mathbf{l}$ is now stochastic.

In the following proposition, we derive the distribution of the sample mean vector and sample covariance matrix when asset returns follow a matrix-variate CSN distribution.

Proposition 3.1. *Let* $\mathbf{X} \sim \mathcal{CSN}_{n,p;1,1}(\mathbf{1}_n \otimes \boldsymbol{\mu}^\top, \mathbf{I}_n \otimes \boldsymbol{\Sigma}, \mathbf{1}_n^\top \otimes \mathbf{e}^\top, 0, v)$ *with* $\boldsymbol{\Sigma} \succ 0$. *Then it holds that*

(i) $\bar{\mathbf{x}} \sim \mathcal{CSN}_{p,1}(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}, n\mathbf{e}^\top, 0, v)$*;*

(ii) $(n-1)\mathbf{S} \sim \mathcal{W}_p(n-1, \boldsymbol{\Sigma})$ *(p-dimensional Wishart distribution when* $p \le n-1$ *and p-dimensional singular Wishart distribution when* $p > n-1$ *with* $n-1$ *degrees of freedom and the parameter matrix* $\boldsymbol{\Sigma}$*);*

(iii) $\bar{\mathbf{x}}$ *and* $\mathbf{S}$ *are independently distributed.*

Proof. Since $(\mathbf{I}_p \otimes \mathbf{1}_n^\top)$ is of full row rank, from Proposition 3.1 of Domínguez-Molina et al. (2007) we have that

$$\bar{\mathbf{x}} \sim \mathcal{CSN}_{p,1}\left(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}, n\mathbf{e}^\top, 0, v\right) \tag{3.5}$$

which shows the first part of the statement. To show the second and third parts of the statement, we make use of the Sherman-Morrison formula (Harville, 1997, Corollary 18.2.10) on the matrix square root in (2.2) and get that

$$\left((\mathbf{I}_n \otimes \boldsymbol{\Sigma})^{-1} + \frac{(\mathbf{1}_n \otimes \mathbf{e})(\mathbf{1}_n^\top \otimes \mathbf{e}^\top)}{v}\right)^{-1/2} = \left(\mathbf{I}_n \otimes \boldsymbol{\Sigma} - \frac{(\mathbf{1}_n\mathbf{1}_n^\top) \otimes (\boldsymbol{\Sigma}\mathbf{e}\mathbf{e}^\top\boldsymbol{\Sigma})}{v + n\mathbf{e}^\top\boldsymbol{\Sigma}\mathbf{e}}\right)^{1/2}.$$

Hence, using Proposition 2.1 in Domínguez-Molina et al. (2007), it holds that

$$(n-1)\mathbf{S} \overset{d}{=} \mathbf{X}^\top\mathbf{V}\mathbf{X} \overset{d}{=} \mathbf{Y}^\top\mathbf{V}\mathbf{Y},$$

where $\mathbf{Y} \sim \mathcal{N}_{n,p}\left(\mathbf{1}_n \otimes \boldsymbol{\mu}, \mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)$. Therefore, we get that $(n-1)\mathbf{S} \sim \mathcal{W}_p(n-1, \boldsymbol{\Sigma})$. Next, let us note that the stochastic representation of $\bar{\mathbf{x}}$ has the following form

$$\bar{\mathbf{x}} \overset{d}{=} \boldsymbol{\mu} + \left(n\boldsymbol{\Sigma}^{-1} + \frac{n^2}{v}\mathbf{e}\mathbf{e}^\top\right)^{-1/2}\tilde{\mathbf{z}} + \frac{\boldsymbol{\Sigma}\mathbf{e}}{\sqrt{v + n\mathbf{e}^\top\boldsymbol{\Sigma}\mathbf{e}}}|\tilde{z}_0|,$$

where $\tilde{\mathbf{z}} \sim \mathcal{N}_p(\mathbf{0}_p, \mathbf{I}_p)$ and $\tilde{z}_0 \sim \mathcal{N}(0,1)$; moreover, $\tilde{\mathbf{z}}$ and $\tilde{z}_0$ are independent. Since the distribution of $\mathbf{S}$ doesn't depend on $\tilde{z}_0$, we get that $\bar{\mathbf{x}}$ and $\mathbf{S}$ are independently distributed. The proposition is proved. □

From Proposition 3.1 we get that the sample mean vector follows multivariate CSN distribution, while the sample covariance matrix follows regular Wishart distribution when $p \le n-1$ and singular Wishart distribution when $p > n-1$. It also holds that the sample mean vector and sample covariance matrix are independent. In what follows, we focus on the case when $p \le n-1$ since it guarantees us that the sample covariance matrix is not singular and its regular inverse can be taken. For the case when $p > n-1$, there is a need for deriving distributional properties of the generalized inverse Wishart matrix and it is not a trivial task.[2]

From Proposition 3.1 we can also see that the introduction of skewness in the data-generating process will affect the mean of the estimators but not the sample covariance matrix. That is, for investors using this type of model for investment, there is a need to adjust their expectations since currently the sample mean is not centred around the true mean and will experience shocks, modelled by the parameter $\mathbf{e}$ and v. The influence of the latent variable scales with n, so the larger the sample size is, the smaller the skewness parameter is expected to be.

3.1. Finite sample results

The sampling distribution of the TP can be derived in many ways. Here, we derive the stochastic representation of the linear combination of the TP weights that fully describes the distribution. This result is delivered in the next theorem.

Theorem 3.2. *Let $\mathbf{X} \sim \mathcal{CSN}_{n,p;1,1}(\mathbf{1}_n \otimes \boldsymbol{\mu}^\top, \mathbf{I}_n \otimes \boldsymbol{\Sigma}, \mathbf{1}_n^\top \otimes \mathbf{e}^\top, 0, v)$ with $n > p$ and $\boldsymbol{\Sigma} \succ 0$. Also, let $\mathbf{l}$ be a p-dimensional vector of constants and $\mathbf{R}_\mathbf{l} := \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{l}\mathbf{l}^\top\boldsymbol{\Sigma}^{-1}/\mathbf{l}^\top\boldsymbol{\Sigma}^{-1}\mathbf{l}$. Then the stochastic representation of $\hat{\theta} = \mathbf{l}^\top\hat{\mathbf{w}}_{TP}$ is given by*

$$\hat{\theta} \stackrel{d}{=} \alpha^{-1}\frac{n-1}{\xi}\left(\mathbf{l}^\top\boldsymbol{\Sigma}^{-1}\bar{\mathbf{z}} + t_0\sqrt{\frac{\mathbf{l}^\top\boldsymbol{\Sigma}^{-1}\mathbf{l}\cdot\bar{\mathbf{z}}^\top\mathbf{R}_\mathbf{l}\bar{\mathbf{z}}}{n-p+1}}\right) \tag{3.6}$$

where $\xi \sim \chi^2_{n-p}$, $t_0 \sim t(n-p+1,0,1)$, and $\bar{\mathbf{z}} \sim \mathcal{CSN}_{p,1}(\boldsymbol{\mu} - r_f\mathbf{1}_p, \frac{1}{n}\boldsymbol{\Sigma}, n\mathbf{e}^\top, 0, v)$ and ξ, t_0, $\bar{\mathbf{z}}$ are mutually independent.

Proof. From Proposition 3.1 we know that $\mathbf{S}$ and $\bar{\mathbf{x}}$ are independently distributed. Consequently, it follows that the conditional distribution of $\hat{\theta}|\bar{\mathbf{x}} = \bar{\mathbf{x}}^*$ is equal to the distribution of $\breve{\theta} := \alpha^{-1}\mathbf{l}^\top\mathbf{S}^{-1}\bar{\mathbf{z}}^*$ with $\bar{\mathbf{z}}^* := (\bar{\mathbf{x}}^* - r_f\mathbf{1}_p)$. Moreover, $\breve{\theta}$ can be rewritten

[2] Additionally assuming that rank($\boldsymbol{\Sigma}$)=$r \le n-1$, Bodnar et al. (2016, 2017) and Bodnar et al. (2019) employed the Moore-Penrose inverse in the portfolio context. One can also make use of different regularization methods such as the ridge-type approach (Tikhonov and Arsenin, 1977), the Landweber-Fridman algorithm (Kress, 1999), the spectral cut-off approach (Chernousova and Golubev, 2014), the Lasso-type method (Brodie et al., 2009), and an iterative method based on a second order damped dynamical systems (Gulliksson and Mazur, 2020; Gulliksson et al., 2021).

as

$$\breve{\theta} \stackrel{d}{=} \alpha^{-1} \bar{\mathbf{z}}^{*\top} \boldsymbol{\Sigma}^{-1} \bar{\mathbf{z}}^{*} \frac{\mathbf{l}^\top \mathbf{S}^{-1} \bar{\mathbf{z}}^{*}}{\bar{\mathbf{z}}^{*\top} \mathbf{S}^{-1} \bar{\mathbf{z}}^{*}} \frac{\bar{\mathbf{z}}^{*\top} \mathbf{S}^{-1} \bar{\mathbf{z}}^{*}}{\bar{\mathbf{z}}^{*\top} \boldsymbol{\Sigma}^{-1} \bar{\mathbf{z}}^{*}}. \tag{3.7}$$

Now, we shall show that $\bar{\mathbf{z}}^{*\top} \boldsymbol{\Sigma}^{-1} \bar{\mathbf{z}}^{*} \cdot \mathbf{l}^\top \mathbf{S}^{-1} \bar{\mathbf{z}}^{*} / \bar{\mathbf{z}}^{*\top} \mathbf{S}^{-1} \bar{\mathbf{z}}^{*}$ and $\bar{\mathbf{z}}^{*\top} \mathbf{S}^{-1} \bar{\mathbf{z}}^{*} / \bar{\mathbf{z}}^{*\top} \boldsymbol{\Sigma}^{-1} \bar{\mathbf{z}}^{*}$ are independently distributed and derive their distributions.

Let $\mathbf{M} = (\mathbf{l}, \bar{\mathbf{z}}^{*})^\top$ such that $\mathbf{l} \neq \bar{\mathbf{z}}^{*}$. Through the application of Theorem 3.2.11 in Muirhead (1990), we obtain that

$$(n-1)(\mathbf{M}\mathbf{S}^{-1}\mathbf{M}^\top)^{-1} \sim \mathcal{W}_2\left(n-p+1, (\mathbf{M}\boldsymbol{\Sigma}^{-1}\mathbf{M}^\top)^{-1}\right),$$

and, through Theorem 3.4.1 in Gupta and Nagar (2018), we receive

$$(n-1)^{-1}\mathbf{M}\mathbf{S}^{-1}\mathbf{M}^\top \sim \mathcal{IW}_2\left(n-p+4, \mathbf{M}\boldsymbol{\Sigma}^{-1}\mathbf{M}^\top\right),$$

i.e. $(n-1)^{-1}\mathbf{M}\mathbf{S}^{-1}\mathbf{M}^\top$ has a 2-dimensional inverse Wishart distribution with $n-p+4$ degrees of freedom and the parameter matrix $\mathbf{M}\boldsymbol{\Sigma}^{-1}\mathbf{M}^\top$. It also holds that

$$\mathbf{M}\mathbf{S}^{-1}\mathbf{M}^\top = \begin{pmatrix} \mathbf{l}^\top \mathbf{S}^{-1} \mathbf{l} & \mathbf{l}^\top \mathbf{S}^{-1} \bar{\mathbf{z}}^{*} \\ \bar{\mathbf{z}}^{*\top} \mathbf{S}^{-1} \mathbf{l} & \bar{\mathbf{z}}^{*\top} \mathbf{S}^{-1} \bar{\mathbf{z}}^{*} \end{pmatrix}. \tag{3.8}$$

Applying Theorem 3(d) of Bodnar and Okhrin (2008), we have that $\bar{\mathbf{z}}^{*\top} \mathbf{S}^{-1} \bar{\mathbf{z}}^{*}$ is independent of $\mathbf{l}^\top \mathbf{S}^{-1} \bar{\mathbf{z}}^{*} / \bar{\mathbf{z}}^{*\top} \mathbf{S}^{-1} \bar{\mathbf{z}}^{*}$. Therefore, $\bar{\mathbf{z}}^{*\top} \mathbf{S}^{-1} \bar{\mathbf{z}}^{*} / \bar{\mathbf{z}}^{*\top} \boldsymbol{\Sigma}^{-1} \bar{\mathbf{z}}^{*}$ is independent of $\bar{\mathbf{z}}^{*\top} \boldsymbol{\Sigma}^{-1} \bar{\mathbf{z}}^{*} \cdot \mathbf{l}^\top \mathbf{S}^{-1} \bar{\mathbf{z}}^{*} / \bar{\mathbf{z}}^{*\top} \mathbf{S}^{-1} \bar{\mathbf{z}}^{*}$. Moreover, from Theorem 3.2.12 of Muirhead (1990), we get that

$$(n-1)\frac{\bar{\mathbf{z}}^{*\top} \boldsymbol{\Sigma}^{-1} \bar{\mathbf{z}}^{*}}{\bar{\mathbf{z}}^{*\top} \mathbf{S}^{-1} \bar{\mathbf{z}}^{*}} \sim \chi^2_{n-p}$$

that is also independent of $\bar{\mathbf{z}}^{*}$. This implies that $\bar{\mathbf{z}}^\top \mathbf{S}^{-1} \bar{\mathbf{z}} / \bar{\mathbf{z}}^\top \boldsymbol{\Sigma}^{-1} \bar{\mathbf{z}}$ is independent of $\bar{\mathbf{z}}^\top \boldsymbol{\Sigma}^{-1} \bar{\mathbf{z}} \cdot \mathbf{l}^\top \mathbf{S}^{-1} \bar{\mathbf{z}} / \bar{\mathbf{z}}^\top \mathbf{S}^{-1} \bar{\mathbf{z}}$, where $\bar{\mathbf{z}} := \bar{\mathbf{x}} - r_f \mathbf{1}_p$.

From the proof in Theorem 1 of Bodnar and Schmid (2008) we obtain

$$\bar{\mathbf{z}}^{*\top} \boldsymbol{\Sigma}^{-1} \bar{\mathbf{z}}^{*} \frac{\mathbf{l}^\top \mathbf{S}^{-1} \bar{\mathbf{z}}^{*}}{\bar{\mathbf{z}}^{*\top} \mathbf{S}^{-1} \bar{\mathbf{z}}^{*}} \sim t\left(n-p+1, \mathbf{l}^\top \boldsymbol{\Sigma}^{-1} \bar{\mathbf{z}}^{*}, \frac{\mathbf{l}^\top \boldsymbol{\Sigma}^{-1} \mathbf{l} \cdot \bar{\mathbf{z}}^{*\top} \mathbf{R}_{\mathbf{l}} \bar{\mathbf{z}}^{*}}{n-p+1}\right)$$

with $\mathbf{R}_{\mathbf{l}} := \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{l}\mathbf{l}^\top\boldsymbol{\Sigma}^{-1}/\mathbf{l}^\top\boldsymbol{\Sigma}^{-1}\mathbf{l}$. Hence, the stochastic representation of $\breve{\theta}$ can be further simplified to

$$\breve{\theta} \stackrel{d}{=} \alpha^{-1}\frac{n-1}{\xi}\left(\mathbf{l}^\top \boldsymbol{\Sigma}^{-1} \bar{\mathbf{z}}^{*} + t_0\sqrt{\frac{\mathbf{l}^\top \boldsymbol{\Sigma}^{-1} \mathbf{l} \cdot \bar{\mathbf{z}}^{*\top} \mathbf{R}_{\mathbf{l}} \bar{\mathbf{z}}^{*}}{n-p+1}}\right) \tag{3.9}$$

where $\xi \sim \chi^2_{n-p}$ and $t_0 \sim t(n-p+1, 0, 1)$ which are independently distributed. Finally, since $\bar{\mathbf{z}} \sim \mathcal{CSN}_p(\boldsymbol{\mu} - r_f \mathbf{1}_p, \frac{1}{n}\boldsymbol{\Sigma}, n\mathbf{e}^\top, 0, \tilde{v})$ (see Genton (2004, Chapter 2.3)), the stochastic representation of $\hat{\theta}$ follows straightforward. The theorem is proved. $\square$

From Theorem 3.2, we can observe that the stochastic representation of $\hat{\theta}$ is expressed as a function of independent univariate random variables that follow χ^2 and t distributions and random vector that follows multivariate CSN distribution. This result helps us to speed up the simulation of $\hat{\theta}$ as we shouldn't simulate the inverse of the sample covariance matrix $\mathbf{S}^{-1}$ that is a computationally heavy task, especially in high dimensions. Let us note that the obtained stochastic representation plays a fundamental role in the derivations of the mean and covariance of $\hat{\mathbf{w}}_{TP}$, and of the asymptotic distribution of $\hat{\theta}$ under a high-dimensional asymptotic regime. We can also use equation (3.6) to compute what influence a certain asset has on the portfolio and what would happen if it was to be excluded. If we let $\mathbf{l}_j = \mathbf{1}_p - \mathbf{b}_j$, where $\mathbf{b}_j$ is the canonical basis in $\mathbb{R}^p$ then we would be investigating how the portfolio size is affected by the exclusion of the j:th asset.

To this end, we further simplify the portfolio diagnostics for the investor by deriving the moments of the portfolio weights distribution. By doing so, the investor can compare several assets in a portfolio and their corresponding returns through a small number of quantities.

Theorem 3.3. *Let* $\mathbf{X} \sim \mathcal{CSN}_{n,p;1,1}(\mathbf{1}_n \otimes \boldsymbol{\mu}^\top, \mathbf{I}_n \otimes \boldsymbol{\Sigma}, \mathbf{1}_n^\top \otimes \mathbf{e}^\top, 0, v)$ *with* $n > p$ *and* $\boldsymbol{\Sigma} \succ 0$. *Also, let* $\mathbf{l}$ *be a p-dimensional vector of constants,* $\tilde{\boldsymbol{\mu}} := \boldsymbol{\mu} - r_f \mathbf{1}_p$ *and* $\tilde{\mathbf{e}} := \alpha^{-1}\sqrt{2/\pi}(v + n\mathbf{e}^\top \boldsymbol{\Sigma} \mathbf{e})^{-1/2}\mathbf{e}$. *Then it holds that*

$$\mathrm{E}[\hat{\mathbf{w}}_{TP}] = \frac{n-1}{n-p-2}\left(\mathbf{w}_{TP} + \tilde{\mathbf{e}}\right)$$

and

$$\begin{aligned}\mathrm{Var}[\hat{\mathbf{w}}_{TP}] &= c_1\left(\mathbf{w}_{TP} + \tilde{\mathbf{e}}\right)\left(\mathbf{w}_{TP} + \tilde{\mathbf{e}}\right)^\top - c_2 \tilde{\mathbf{e}}\tilde{\mathbf{e}}^\top \\ &\quad + c_3\left(1 - \frac{2}{n} + \tilde{\boldsymbol{\mu}}^\top \boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\mu}} + 2\alpha\tilde{\mathbf{e}}^\top \tilde{\boldsymbol{\mu}}\right)\boldsymbol{\Sigma}^{-1}\end{aligned}$$

with

$$\begin{aligned} c_1 &= \frac{(n-1)^2(n-p)}{(n-p-1)(n-p-2)^2(n-p-4)}, \\ c_2 &= \frac{(n-1)^2}{(n-p-1)(n-p-4)}, \\ c_3 &= \frac{c_1(n-p-2)}{\alpha^2(n-p)}. \end{aligned}$$

Proof of Theorem 3.3. First, we shall evaluate $\mathrm{E}[\hat{\theta}]$. Application of Theorem 3.2 leads us to

$$\begin{aligned}\mathrm{E}[\hat{\theta}] &= \mathrm{E}\left[\alpha^{-1}\frac{n-1}{\xi}\left(\mathbf{l}^\top\mathbf{\Sigma}^{-1}\bar{\mathbf{z}} + t_0\sqrt{\frac{\mathbf{l}^\top\mathbf{\Sigma}^{-1}\mathbf{l}\cdot\bar{\mathbf{z}}^\top\mathbf{R_l}\bar{\mathbf{z}}}{n-p+1}}\right)\right]\\ &= \frac{n-1}{\alpha}\,\mathrm{E}\left[\frac{1}{\xi}\right]\mathbf{l}^\top\mathbf{\Sigma}^{-1}\,\mathrm{E}[\bar{\mathbf{z}}], \qquad (3.10)\end{aligned}$$

where the last equality follows from the fact that ξ, t_0, and $\bar{\mathbf{z}}$ are mutually independent, and $\mathrm{E}[t_0]=0$. Since $\xi \sim \chi^2_{n-p}$, it holds that $1/\xi \sim Inv-\chi^2_{n-p}$ (inverse-chi-squared distribution with $n-p$ degrees of freedom). From Gelman et al. (2013, p. 575) it follows that

$$\mathrm{E}\left[\frac{1}{\xi}\right] = \frac{1}{n-p-2}. \qquad (3.11)$$

Next, we shall evaluate $\mathrm{E}(\bar{\mathbf{z}})$ using the moment generating function of $\bar{\mathbf{z}}$ which is given by

$$m_{\bar{\mathbf{z}}}(\mathbf{t}) = 2\Phi_1\left(\mathbf{e}^\top\mathbf{\Sigma}\mathbf{t}; 0, v+n\mathbf{e}^\top\mathbf{\Sigma}\mathbf{e}\right)\exp\left(\tilde{\boldsymbol{\mu}}^\top\mathbf{t} + \frac{1}{2n}\mathbf{t}^\top\mathbf{\Sigma}\mathbf{t}\right)$$

for $\mathbf{t}\in\mathbb{R}^p$ (see Genton (2004, Lemma 2.2.2)). Therefore, $\mathrm{E}[\bar{\mathbf{z}}]$ can be evaluated as

$$\begin{aligned}\mathrm{E}[\bar{\mathbf{z}}] &= \left.\frac{\partial m_{\bar{\mathbf{z}}}(\mathbf{t})}{\partial\mathbf{t}}\right|_{\mathbf{t}=\mathbf{0}}\\ &= 2\Bigg[\phi_1\left(\mathbf{e}^\top\mathbf{\Sigma}\mathbf{t}; 0, \tilde{v}+n\mathbf{e}^\top\mathbf{\Sigma}\mathbf{e}\right)\mathbf{\Sigma}\mathbf{e}\\ &\quad +\Phi_1\left(\mathbf{e}^\top\mathbf{\Sigma}\mathbf{t}; 0, v+n\mathbf{e}^\top\mathbf{\Sigma}\mathbf{e}\right)\left(\tilde{\boldsymbol{\mu}}+\frac{1}{n}\mathbf{\Sigma}\mathbf{t}\right)\Bigg]\\ &\quad \times\exp\left(\tilde{\boldsymbol{\mu}}^\top\mathbf{t}+\frac{1}{2n}\mathbf{t}^\top\mathbf{\Sigma}\mathbf{t}\right)\Bigg|_{\mathbf{t}=\mathbf{0}}\\ &= \sqrt{\frac{2}{\pi}}(v+n\mathbf{e}^\top\mathbf{\Sigma}\mathbf{e})^{-1/2}\mathbf{\Sigma}\mathbf{e}+\tilde{\boldsymbol{\mu}}. \qquad (3.12)\end{aligned}$$

Substituting (3.11) and (3.12) in (3.10) and using the fact that $\mathbf{l}$ is an arbitrary vector show the first part of the theorem.

Next, we shall evaluate $\mathrm{Var}[\hat{\theta}]$. Let us recall that

$$\mathrm{Var}[\hat{\theta}] = \mathrm{E}[\hat{\theta}^2] - (\mathrm{E}[\hat{\theta}])^2,$$

where $\mathrm{E}[\hat{\theta}]$ is known from above, while $\mathrm{E}[\hat{\theta}^2]$ should be evaluated. From Theorem 3.2 we get that

$$\begin{aligned}
\mathrm{E}[\hat{\theta}^2] &= \mathrm{E}\left[\alpha^{-2}\frac{(n-1)^2}{\xi^2}\left(\mathbf{l}^\top\mathbf{\Sigma}^{-1}\bar{\mathbf{z}} + t_0\sqrt{\frac{\mathbf{l}^\top\mathbf{\Sigma}^{-1}\mathbf{l}\cdot\bar{\mathbf{z}}^\top\mathbf{R}_\mathbf{l}\bar{\mathbf{z}}}{n-p+1}}\right)^2\right] \\
&= \frac{(n-1)^2}{\alpha^2}\mathrm{E}\left[\frac{1}{\xi^2}\right]\left(\mathrm{E}\left[(\mathbf{l}^\top\mathbf{\Sigma}^{-1}\bar{\mathbf{z}})^2\right] + \frac{\mathbf{l}^\top\mathbf{\Sigma}^{-1}\mathbf{l}}{n-p-1}\mathrm{E}\left[\bar{\mathbf{z}}^\top\mathbf{R}_\mathbf{l}\bar{\mathbf{z}}\right]\right), \qquad (3.13)
\end{aligned}$$

where the last equality follows from the fact that ξ, t_0, and $\bar{\mathbf{z}}$ are mutually independent, and $\mathrm{E}[t_0]=0$, $\mathrm{E}[t_0^2]=\frac{n-p+1}{n-p-1}$. From Gelman et al. (2013, p. 575), we obtain that

$$\mathrm{E}\left[\frac{1}{\xi^2}\right] = \frac{1}{(n-p-2)(n-p-4)} \qquad (3.14)$$

while through the application of Lemma 5.1 it holds that

$$\begin{aligned}
\mathrm{E}\left[(\mathbf{l}^\top\mathbf{\Sigma}^{-1}\bar{\mathbf{z}})^2\right] &= \mathrm{E}\left[\bar{\mathbf{z}}^\top\mathbf{\Sigma}^{-1}\mathbf{l}\mathbf{l}^\top\mathbf{\Sigma}^{-1}\bar{\mathbf{z}}\right] \\
&= (\tilde{\boldsymbol{\mu}}^\top\mathbf{\Sigma}^{-1}\mathbf{l})^2 + \frac{\mathbf{l}^\top\mathbf{\Sigma}^{-1}\mathbf{l}}{n} + 2\sqrt{\frac{2}{\pi}}\frac{\mathbf{e}^\top\mathbf{l}\mathbf{l}^\top\mathbf{\Sigma}^{-1}\tilde{\boldsymbol{\mu}}}{(v+n\mathbf{e}^\top\mathbf{\Sigma}\mathbf{e})^{1/2}} \\
&= (\alpha\theta)^2 + \frac{\mathbf{l}^\top\mathbf{\Sigma}^{-1}\mathbf{l}}{n} + 2\sqrt{\frac{2}{\pi}}\frac{\alpha\theta\mathbf{e}^\top\mathbf{l}}{(v+n\mathbf{e}^\top\mathbf{\Sigma}\mathbf{e})^{1/2}} \qquad (3.15)
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}\left[\bar{\mathbf{z}}^\top\mathbf{R}_\mathbf{l}\bar{\mathbf{z}}\right] &= \tilde{\boldsymbol{\mu}}^\top\mathbf{R}_\mathbf{l}\tilde{\boldsymbol{\mu}} + \frac{p-1}{n} + 2\sqrt{\frac{2}{\pi}}\frac{\mathbf{e}^\top\mathbf{\Sigma}\mathbf{R}_\mathbf{l}\tilde{\boldsymbol{\mu}}}{(v+n\mathbf{e}^\top\mathbf{\Sigma}\mathbf{e})^{1/2}} \\
&= \tilde{\boldsymbol{\mu}}^\top\mathbf{\Sigma}^{-1}\tilde{\boldsymbol{\mu}} - \frac{(\alpha\theta)^2}{\mathbf{l}^\top\mathbf{\Sigma}^{-1}\mathbf{l}} + \frac{p-1}{n} + 2\sqrt{\frac{2}{\pi}}\frac{\mathbf{e}^\top\mathbf{\Sigma}\mathbf{R}_\mathbf{l}\tilde{\boldsymbol{\mu}}}{(v+n\mathbf{e}^\top\mathbf{\Sigma}\mathbf{e})^{1/2}}. \qquad (3.16)
\end{aligned}$$

Hence, through (3.14), (3.15) and (3.16), we receive

$$
\begin{aligned}
\mathrm{E}[\hat{\theta}^2] &= \frac{(n-1)^2}{\alpha^2(n-p-2)(n-p-4)}\left[(\alpha\theta)^2+\frac{\mathbf{l}^\top\mathbf{\Sigma}^{-1}\mathbf{l}}{n}+2\sqrt{\frac{2}{\pi}}\frac{\alpha\theta\mathbf{e}^\top\mathbf{l}}{(v+n\mathbf{e}^\top\mathbf{\Sigma}\mathbf{e})^{1/2}}\right.\\
&\quad\left.+\frac{\mathbf{l}^\top\mathbf{\Sigma}^{-1}\mathbf{l}}{n-p-1}\left(\tilde{\boldsymbol{\mu}}^\top\mathbf{\Sigma}^{-1}\tilde{\boldsymbol{\mu}}-\frac{(\alpha\theta)^2}{\mathbf{l}^\top\mathbf{\Sigma}^{-1}\mathbf{l}}+\frac{p-1}{n}+2\sqrt{\frac{2}{\pi}}\frac{\mathbf{e}^\top\mathbf{\Sigma}\mathbf{R}_\mathbf{l}\tilde{\boldsymbol{\mu}}}{(v+n\mathbf{e}^\top\mathbf{\Sigma}\mathbf{e})^{1/2}}\right)\right]\\
&= \frac{(n-1)^2}{\alpha^2(n-p-1)(n-p-2)(n-p-4)}\\
&\quad\times\left[(n-p-2)\alpha^2\left(\theta+\frac{1}{\alpha}\sqrt{\frac{2}{\pi}}\frac{\mathbf{e}^\top\mathbf{l}}{(v+n\mathbf{e}^\top\mathbf{\Sigma}\mathbf{e})^{1/2}}\right)^2-\frac{2(n-p-2)}{\pi}\frac{(\mathbf{e}^\top\mathbf{l})^2}{v+n\mathbf{e}^\top\mathbf{\Sigma}\mathbf{e}}\right.\\
&\quad\left.+\left(n-2+n\tilde{\boldsymbol{\mu}}^\top\mathbf{\Sigma}^{-1}\tilde{\boldsymbol{\mu}}+2\sqrt{\frac{2}{\pi}}\frac{n\mathbf{e}^\top\tilde{\boldsymbol{\mu}}}{(v+n\mathbf{e}^\top\mathbf{\Sigma}\mathbf{e})^{1/2}}\right)\frac{\mathbf{l}^\top\mathbf{\Sigma}^{-1}\mathbf{l}}{n}\right].
\end{aligned}
$$

Therefore, we get that

$$
\begin{aligned}
\mathrm{Var}[\hat{\theta}] &= \mathrm{E}[\hat{\theta}^2]-\left(\mathrm{E}[\hat{\theta}]\right)^2\\
&= c_1\left(\theta+\frac{1}{\alpha}\sqrt{\frac{2}{\pi}}\frac{\mathbf{e}^\top\mathbf{l}}{(v+n\mathbf{e}^\top\mathbf{\Sigma}\mathbf{e})^{1/2}}\right)^2-c_2\frac{2}{\alpha^2\pi}\frac{(\mathbf{e}^\top\mathbf{l})^2}{v+n\mathbf{e}^\top\mathbf{\Sigma}\mathbf{e}}\\
&\quad+c_3\left(1-\frac{2}{n}+\tilde{\boldsymbol{\mu}}^\top\mathbf{\Sigma}^{-1}\tilde{\boldsymbol{\mu}}+2\sqrt{\frac{2}{\pi}}\frac{\mathbf{e}^\top\tilde{\boldsymbol{\mu}}}{(v+n\mathbf{e}^\top\mathbf{\Sigma}\mathbf{e})^{1/2}}\right)\mathbf{l}^\top\mathbf{\Sigma}^{-1}\mathbf{l}
\end{aligned}
$$

with c_1, c_2 and c_3 which are the same as in the formulation of the theorem. Finally, using the fact that $\mathbf{l}$ is an arbitrary vector we arrive the statement of the theorem. □

From Theorem 3.3 we can clearly see that our estimates are biased when returns follow the CSN distribution. Any increase in the elements of $\mathbf{e}$ will increase that bias and an increase in the sample size n or v will decrease that bias. We can also see that it is quite hard to disentangle the parameters $\mathbf{e}$ and v. It is not surprising since the definition of the distribution is in terms of shocks based of the matrix $\mathbf{1}^\top \otimes \mathbf{e}^\top$ and v.

3.2. Asymptotic distribution under double asymptotic regime

Today investors need to traverse large asset universes, trying to convey which assets to be included in the portfolio and which not. This often implies that their portfolio grows in relation to the number of samples they can obtain for the specific asset classes. When the asset universe is large, the number of parameters that need to be estimated grow respectively, especially in comparison to the sample size. It is therefore of interest to study the influence of the number of assets included in the portfolio and what happens in the limit.

In the high-dimensional setting, both n and p grow towards infinity such that $p/n = c \in (0,1)$. The following are some necessary assumptions for the existence of the asymptotic distribution.

(A1) There exists a constant M_1 such that $\max_i |\mu_i| \le M_1$ uniformly in p.
(A2) Let λ_1 and λ_p denote the largest and smallest eigenvalue of $\mathbf{\Sigma}$. There exist constants M_2 and M_3 such that $M_2 < \lambda_1,\ \lambda_p < M_3$ uniformly in p.
(A3) There exists a constant M_4 such that $\max_i |e_i| \le M_4$.

The assumptions (A1) and (A2) concern the mean vector and covariance matrix in high dimensions (see, Ledoit and Wolf (2017) or Bodnar et al. (2021)). As for (A1), it is empirically and intuitively justified, there does not exist an asset return with infinite mean, if so, the whole market would converge to that asset. If the mean would be negative enough then, in a limiting case, the asset will not survive in the market as no investor would invest in it. As for (A2), it covers the fluctuations of classes of asset returns along the axis of their eigenvectors. The eigenvectors are in turn a rotation to make the assets uncorrelated. The assumption (A3) is due to the fact that CSN model depends on several parameters and can be interpreted as the skewness for each individual asset in the portfolio needs to be bounded, it can not grow with n or p. It is merely a technicality since infinite skewness has little economical interpretation.

Let $X \overset{a.s.}{\to} x$ denotes almost sure convergence of a random variable X to a quantity x and $X \overset{d}{\to} F$ denotes convergence in distribution. Define $\mathcal{L} = \{\mathbf{z} : z_i < \infty \ \forall\ i = 1,2,...,p,\ \mathbf{z}^\top \mathbf{1}_p < \infty\}$ as a feasible set of linear combinations. Assuming that $\mathbf{l} \in \mathcal{L}$, we limit the investor to choose the combinations to make inference of. This is a technicality since in most practical applications especially in higher dimensions, the vector $\mathbf{l}$ will be sparse. In the following theorem, we present results for $\hat{\theta}$ in the high-dimensional setting.

Theorem 3.4. *Let* $\mathbf{X} \sim \mathcal{CSN}_{n,p;1,1}(\mathbf{1}_n \otimes \boldsymbol{\mu}^\top, \mathbf{I}_n \otimes \mathbf{\Sigma}, \mathbf{1}_n^\top \otimes \mathbf{e}^\top, 0, v)$ *with* $n > p$ *and* $\mathbf{\Sigma} \succ 0$. *Also, let* $\mathbf{l} \in \mathcal{L}$. *Then, under the assumptions (A1)-(A3), it holds that*

$$\hat{\theta} \overset{a.s.}{\to} \frac{1}{1-c}\theta$$

and

$$\frac{\sqrt{n-p}}{\tilde{\sigma}} \left(\hat{\theta} - \frac{1}{1-p/n}\theta \right) \overset{d}{\to} \mathcal{N}(0,1)$$

with

$$\tilde{\sigma}^2 = \frac{\alpha^{-2}}{(1-c)^2} \left[2\,(\alpha\theta)^2 + \mathbf{l}^\top \mathbf{\Sigma}^{-1} \mathbf{l} \left((1-c) + (\boldsymbol{\mu} - r_f \mathbf{1}_p)^\top \mathbf{R}_{\mathbf{l}} (\boldsymbol{\mu} - r_f \mathbf{1}_p) \right) \right]$$

for $p/n \to c \in (0,1)$ as $p \to \infty$ and $n \to \infty$.

Proof. From Theorem 3.2 we know that

$$\hat{\theta} \stackrel{d}{=} \alpha^{-1}\frac{n-1}{\xi}\left(\mathbf{l}^\top\mathbf{\Sigma}^{-1}\bar{\mathbf{z}} + t_0\sqrt{\frac{\mathbf{l}^\top\mathbf{\Sigma}^{-1}\mathbf{l}\cdot\bar{\mathbf{z}}^\top\mathbf{R}_\mathbf{l}\bar{\mathbf{z}}}{n-p+1}}\right) \tag{3.17}$$

where $\xi \sim \chi^2_{n-p}$, $t_0 \sim t(n-p+1, 0, 1)$, and $\bar{\mathbf{z}} \sim \mathcal{CSN}_{p,1}(\boldsymbol{\mu} - r_f\mathbf{1}_p, \frac{1}{n}\mathbf{\Sigma}, n\mathbf{e}^\top, 0, v)$; moreover, ξ, t_0 and $\bar{\mathbf{z}}$ are mutually independent.

Through the properties of the χ^2-distribution we have that

$$\frac{\xi}{n-p} \stackrel{a.s.}{\to} 1 \tag{3.18}$$

for $p/n \to c \in (0,1)$ as $p \to \infty$ and $n \to \infty$ (Bodnar and Reiß, 2016, Lemma 3). From Proposition 2.1 of Domínguez-Molina et al. (2007) we have that $\bar{\mathbf{z}}$ permits the following stochastic representation

$$\begin{aligned}
\bar{\mathbf{z}} &\stackrel{d}{=} \boldsymbol{\mu} - r_f\mathbf{1}_p + \left(n\mathbf{\Sigma}^{-1} + \frac{n^2}{v}\mathbf{e}\mathbf{e}^\top\right)^{-1/2}\tilde{\mathbf{z}} + \frac{\mathbf{\Sigma}\mathbf{e}}{\sqrt{v + n\mathbf{e}^\top\mathbf{\Sigma}\mathbf{e}}}|\tilde{z}_0| \\
&= \boldsymbol{\mu} - r_f\mathbf{1}_p + \frac{1}{\sqrt{n}}\left(\mathbf{\Sigma} - \frac{\mathbf{\Sigma}\mathbf{e}\mathbf{e}^\top\mathbf{\Sigma}}{\frac{v}{n} + \mathbf{e}^\top\mathbf{\Sigma}\mathbf{e}}\right)^{1/2}\tilde{\mathbf{z}} + \frac{\mathbf{\Sigma}\mathbf{e}}{\sqrt{v + n\mathbf{e}^\top\mathbf{\Sigma}\mathbf{e}}}|\tilde{z}_0|,
\end{aligned} \tag{3.19}$$

where $\tilde{\mathbf{z}} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$, $\tilde{z}_0 \sim \mathcal{N}(0,1)$ and they are independently distributed. We would note that in the last equality we used the Sherman-Morrison inversion formula (Sherman and Morrison (1950)). From (3.19), under the assumptions (A1)-(A3), we obtain that

$$\mathbf{l}^\top\bar{\mathbf{z}} \stackrel{a.s.}{\to} \mathbf{l}^\top(\boldsymbol{\mu} - r_f\mathbf{1}_p) \tag{3.20}$$

for $p/n \to c \in (0,1)$ as $p \to \infty$ and $n \to \infty$.

Therefore, through (3.18), (3.20) and assumptions (A1)-(A3), we get

$$\begin{aligned}
\hat{\theta} &\stackrel{d}{=} \alpha^{-1}\frac{n-1}{n-p}\frac{1}{\xi/(n-p)}\left(\mathbf{l}^\top\mathbf{\Sigma}^{-1}\bar{\mathbf{z}} + t_0\sqrt{\frac{\mathbf{l}^\top\mathbf{\Sigma}^{-1}\mathbf{l}\cdot\bar{\mathbf{z}}^\top\mathbf{R}_\mathbf{l}\bar{\mathbf{z}}}{n-p+1}}\right) \\
&\stackrel{d}{\to} \frac{1}{1-c}\theta.
\end{aligned}$$

Since convergence in distribution to a point implies convergence almost surely, we receive, through the continuous mapping theorem (Billingsley (2013)), that the one dimensional objects in (3.17) converge to their desired components. The first part of the theorem is shown.

Next, we derive the high-dimensional asymptotic distribution of $\hat{\theta}$. For sufficiently large n and p, through the stochastic representation $\hat{\theta}$ given in Theorem 3.2, we have that

$$\begin{aligned}
&\sqrt{n-p}\left(\hat{\theta} - \frac{1}{1-p/n}\theta\right) \\
\stackrel{d}{=}\;& \sqrt{n-p}\left[\alpha^{-1}\frac{n-1}{\xi}\left(\mathbf{l}^\top\mathbf{\Sigma}^{-1}\bar{\mathbf{z}} + t_0\sqrt{\frac{\mathbf{l}^\top\mathbf{\Sigma}^{-1}\mathbf{l}\cdot\bar{\mathbf{z}}^\top\mathbf{R}_{\mathbf{l}}\bar{\mathbf{z}}}{n-p+1}}\right) - \frac{\alpha^{-1}}{1-p/n}\mathbf{l}^\top\mathbf{\Sigma}^{-1}(\boldsymbol{\mu}-r_f\mathbf{1})\right] \\
\approx\;& \frac{\alpha^{-1}}{\xi/(n-p)}\frac{\sqrt{n-p}}{1-p/n}\left[\mathbf{l}^\top\mathbf{\Sigma}^{-1}\left(\bar{\mathbf{z}} - \frac{\xi}{n-p}(\boldsymbol{\mu}-r_f\mathbf{1}_p)\right) + t_0\sqrt{\frac{\mathbf{l}^\top\mathbf{\Sigma}^{-1}\mathbf{l}\cdot\bar{\mathbf{z}}^\top\mathbf{R}_{\mathbf{l}}\bar{\mathbf{z}}}{n-p+1}}\right]. \quad (3.21)
\end{aligned}$$

Furthermore, by (3.19) we have that

$$\begin{aligned}
\bar{\mathbf{z}} - \frac{\xi}{n-p}(\boldsymbol{\mu}-r_f\mathbf{1}_p) &= \left(\frac{\xi}{n-p}-1\right)(r_f\mathbf{1}_p-\boldsymbol{\mu}) + \frac{1}{\sqrt{n}}\left(\mathbf{\Sigma} - \frac{\mathbf{\Sigma}\mathbf{e}\mathbf{e}^\top\mathbf{\Sigma}}{\frac{v}{(n}+\mathbf{e}^\top\mathbf{\Sigma}\mathbf{e}}\right)^{1/2}\tilde{\mathbf{z}} \\
&\quad + \frac{\mathbf{\Sigma}\mathbf{e}}{\sqrt{v+n\mathbf{e}^\top\mathbf{\Sigma}\mathbf{e}}}|\tilde{z}_0|. \quad (3.22)
\end{aligned}$$

From Bodnar and Reiß (2016, Lemma 3) we have that

$$\sqrt{n-p}\left(\frac{\xi}{n-p}-1\right)\xrightarrow{d}\mathcal{N}(0,2)$$

and, therefore, it holds that

$$\sqrt{n-p}\mathbf{l}^\top\mathbf{\Sigma}^{-1}\left(\bar{\mathbf{z}} - \frac{\xi}{n-p}(\boldsymbol{\mu}-r_f\mathbf{1}_p)\right)\xrightarrow{d}\sqrt{2}\mathbf{l}^\top\mathbf{\Sigma}^{-1}(\boldsymbol{\mu}-r_f\mathbf{1})\tilde{z}_1 + \sqrt{1-c}\mathbf{l}^\top\mathbf{\Sigma}^{-1/2}\tilde{\mathbf{z}},$$

where we used the fact that $-\tilde{z}_1 \stackrel{d}{=} \tilde{z}_1$ if $\tilde{z}_1 \sim \mathcal{N}(0,1)$ and that the dependence on $\tilde{z}_0$ vanishes asymptotically. By assumptions (A1)-(A3) together with (3.22) it holds that

$$\bar{\mathbf{z}}^\top\mathbf{R}_{\mathbf{l}}\bar{\mathbf{z}}\xrightarrow{\mathbb{P}}(\boldsymbol{\mu}-r_f\mathbf{1}_p)^\top\mathbf{R}_{\mathbf{l}}(\boldsymbol{\mu}-r_f\mathbf{1}_p).$$

Thus, we have that (3.21) converges in distribution towards

$$\begin{aligned}
\frac{\alpha^{-1}}{1-c}\Bigg\{&\sqrt{2}\mathbf{l}^\top\mathbf{\Sigma}^{-1}(\boldsymbol{\mu}-r_f\mathbf{1})\tilde{z}_1 + \sqrt{1-c}\mathbf{l}^\top\mathbf{\Sigma}^{-1/2}\tilde{\mathbf{z}} \\
&+ \sqrt{\mathbf{l}^\top\mathbf{\Sigma}^{-1}\mathbf{l}}\sqrt{(\boldsymbol{\mu}-r_f\mathbf{1}_p)^\top\mathbf{R}_{\mathbf{l}}(\boldsymbol{\mu}-r_f\mathbf{1}_p)}\tilde{z}_2\Bigg\}, \quad (3.23)
\end{aligned}$$

where $\tilde{\mathbf{z}} \sim \mathcal{N}_p(\mathbf{0}_p,\mathbf{I}_p)$, $\tilde{z}_1 \sim \mathcal{N}(0,1)$, and $\tilde{z}_2 \sim \mathcal{N}(0,1)$; moreover, $\tilde{\mathbf{z}}$, $\tilde{z}_1$ and $\tilde{z}_2$ are mutually independently distributed. Evaluating the variance of (3.23) we receive the desired statement. The theorem is proved. □

4. Simulation and empirical studies

In this section, we will investigate different properties of the estimators through a simulation study and finish the section with an empirical application.

4.1. Simulation study

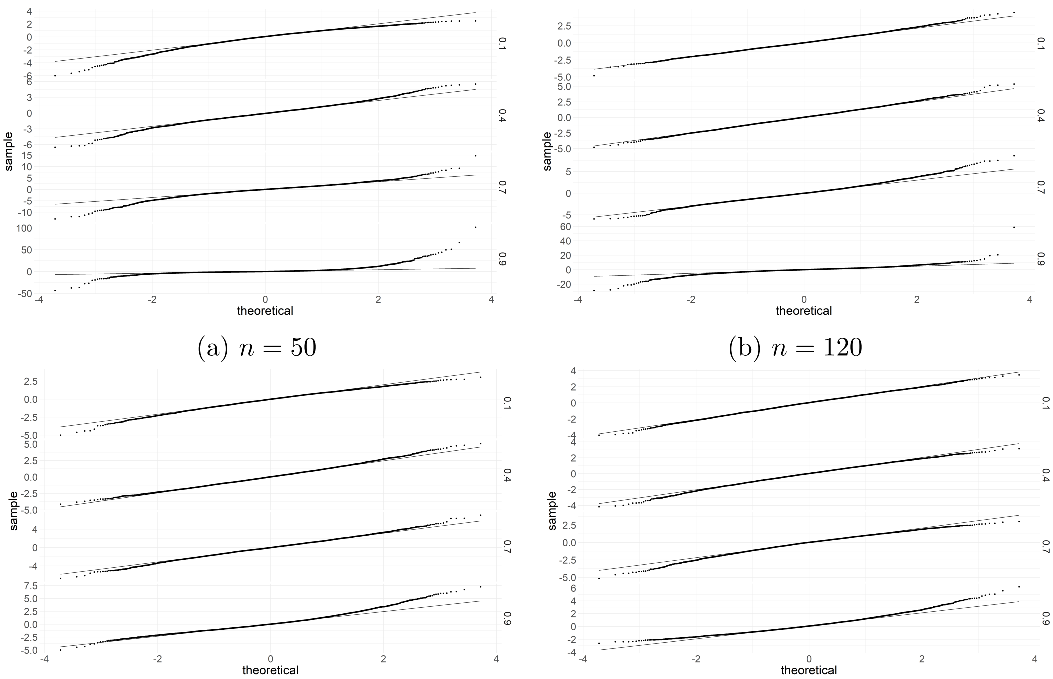
We will investigate how well the high-dimensional asymptotic distribution approximates the finite-sample distribution given by Theorem 3.4 and 3.2, respectively. A number of parameters are simulated to analyze it. The following setup for the simulation study is applied. We set $\alpha = 2$, $r_f = 0.01$ and $v = 3$. For each combination of $n \in \{50, 120, 250, 500\}$ and $c \in \{0.1, 0.3, 0.7, 0.9\}$, we

(1) simulate a random matrix $\mathbf{Y}$ of size $p \times n$ with entries following a centered normal distribution with standard deviation 0.2 and fix $\mathbf{\Sigma} = \mathbf{Y}\mathbf{Y}^\top$;
(2) simulate the elements of the mean vector $\boldsymbol{\mu}$ according to $\mu_j \sim U(-0.1, 0.1)$, $j = 1, 2, \ldots, p$;
(3) simulate the elements of the skewness parameter $\mathbf{e}$ from a standard t-distribution with 5 degrees of freedom;
(4) simulate 10^4 observations from the sampling distribution of $\hat{\theta}$ given by Theorem 3.2.

In Figure 1 we compare the quantiles of the empirical sampling distribution and its high-dimensional asymptotic distribution for the different cases of $n \in \{50, 120, 250, 500\}$. We can see that for small values of c (the two first rows of each subfigure), the approximation works well. The only occasion where the approximation does not seem sufficient, is for $n = 50$. However, for larger values of n it does not seem to be an issue. For c closer to 1 the asymptotic distribution fails to account for the tails and there are a number of explanations for this. The effective sample size is very small since the number of parameters we need are estimating is large. In each scenario we are estimating p quantities in the mean vector and $p(p-1)/2$ elements of the covariance matrix to construct the tangency portfolio.

4.2. Empirical study

This section presents an application of the theoretical results obtained in Section 3. Weekly log-returns of a randomly selected number of stocks from the S&P 500 index for the period from the 19th of February, 2011, to the 6th of May, 2020, are used in this study. No special preference is given to the selection of stocks as it is not in our



(a) $n = 50$ (b) $n = 120$

(c) $n = 250$ (d) $n = 500$

Figure 1.: QQ-plot of realisations of $\hat{\theta}$. In this Figure, we compare the empirical distribution to its theoretical high dimensional asymptotic distribution. On the right hand, we display the quantity $c_n = p/n$ for each simulated scenario.

interest to select a good (or bad) collection of stocks.

To estimate the parameters of the CSN model, we employ the method of weighted moments developed by Bodnar and Gupta (2015). Due to the complex relationship between v and $\mathbf{e}$, estimation of the skewness vector $\mathbf{e}$ is hard. However, estimation of a skewness vector that is proportional to the skewness parameter $\mathbf{e}$ is much easier. That is, we are able to estimate

$$\breve{\mathbf{e}} = \frac{1}{(v + n\mathbf{e}^\top \boldsymbol{\Sigma} \mathbf{e})^{1/2}} \mathbf{e}.$$

If we estimate $\breve{\mathbf{e}}$ instead of $\mathbf{e}$ the parameter v is implicitly determined by the algorithm. However, to use the results in Section 3 we still need to specify it, ex ante. To estimate the parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and $\breve{\mathbf{e}}$, we employ the algorithm listed below. For a more detailed exposition on the approach, we suggest the mentioned reference and the references therein.

(1) From a data matrix $\mathbf{X}$ of size $n \times p$ construct the following entities

$$\begin{aligned}
\bar{\mathbf{x}} &= \frac{1}{n}\sum_{t=1}^{n}\mathbf{x}_t, & (4.1)\\
\hat{\mathbf{\Sigma}} &= \frac{1}{n-1}\sum_{t=1}^{n}(\mathbf{x}_t-\bar{\mathbf{x}})(\mathbf{x}_t-\bar{\mathbf{x}})^\top, & (4.2)\\
\bar{\phi}(\gamma) &= \frac{\gamma^{-p/2}}{n}\sum_{t=1}^{n}\exp\left(\frac{\gamma-1}{2\gamma}\mathbf{x}_t^\top\hat{\mathbf{\Sigma}}^{-1}\mathbf{x}_t\right), & (4.3)\\
\bar{\mathbf{x}}_\phi(\gamma) &= \frac{\gamma^{-p/2}}{n}\sum_{t=1}^{n}\mathbf{x}_t\exp\left(\frac{\gamma-1}{2\gamma}\mathbf{x}_t^\top\hat{\mathbf{\Sigma}}^{-1}\mathbf{x}_t\right), & (4.4)
\end{aligned}$$

where $\gamma > 1$.

(2) Solve (numerically) the following root problem

$$\left(\gamma\bar{\mathbf{x}}-\frac{\bar{\mathbf{x}}_\phi(\gamma)}{\bar{\phi}(\gamma)}\right)^\top\hat{\mathbf{\Sigma}}\left(\gamma\bar{\mathbf{x}}-\frac{\bar{\mathbf{x}}_\phi(\gamma)}{\bar{\phi}(\gamma)}\right)+\frac{2}{(\gamma-1)\pi n}(1-f(\gamma))^2\left(1-\frac{1}{f(\gamma)^2}\right)=0,$$

where

$$\begin{aligned}
f(\gamma) = & \frac{1}{\bar{\mathbf{x}}^\top\hat{\mathbf{\Sigma}}^{-1}\bar{\mathbf{x}}-\frac{2}{\gamma-1}\log\bar{\phi}(\gamma)}\left[\frac{\bar{\mathbf{x}}^\top\hat{\mathbf{\Sigma}}^{-1}\bar{\mathbf{x}}_\phi(\gamma)}{\bar{\phi}(\gamma)}-\frac{2\gamma}{\gamma-1}\log\bar{\phi}(\gamma)\right.\\
&\pm\left(\left[\frac{\bar{\mathbf{x}}^\top\hat{\mathbf{\Sigma}}^{-1}\bar{\mathbf{x}}_\phi(\gamma)}{\bar{\phi}(\gamma)}-\frac{2\gamma}{\gamma-1}\log\bar{\phi}(\gamma)\right]^2-\left(\bar{\mathbf{x}}^\top\hat{\mathbf{\Sigma}}^{-1}\bar{\mathbf{x}}-\frac{2}{\gamma-1}\log\bar{\phi}(\gamma)\right)\right.\\
&\left.\left.\times\left[\frac{\bar{\mathbf{x}}_\phi(\gamma)^\top\hat{\mathbf{\Sigma}}^{-1}\bar{\mathbf{x}}_\phi(\gamma)}{\bar{\phi}(\gamma)^2}-\frac{2\gamma^2}{\gamma-1}\log\bar{\phi}(\gamma)\right]\right)^{1/2}\right] \qquad (4.5)
\end{aligned}$$

in terms of γ. We first consider the positive part in (4.5) and if no solution is found there, we try with the negative one. Let $\hat{\gamma}$ denotes its solution.

(3) The estimator for $\boldsymbol{\mu}, \mathbf{\Sigma}$ and $\breve{\mathbf{e}}$ are then determined by

$$\begin{aligned}
\hat{\mathbf{\Sigma}} &= \mathbf{S}, & (4.6)\\
\hat{\boldsymbol{\mu}} &= \frac{1}{f(\hat{\gamma})-\hat{\gamma}}\left(f(\hat{\gamma})\bar{\mathbf{x}}-\frac{\bar{\mathbf{x}}_\phi(\hat{\gamma})}{\bar{\phi}(\hat{\gamma})}\right), & (4.7)\\
\hat{\breve{\mathbf{e}}} &= \sqrt{\frac{\pi}{2}}\frac{1}{1-f(\hat{\gamma})}\left(\hat{\gamma}\bar{\mathbf{x}}-\frac{\bar{\mathbf{x}}_\phi(\hat{\gamma})}{\bar{\phi}(\hat{\gamma})}\right). & (4.8)
\end{aligned}$$

We take these estimates and use them in a plug-in approach for the moments described in Theorem 3.3 and 3.4. In Figures 2 and 3, we show the comparison between the moments obtained in the finite- and high-dimensional settings. The first row shows the estimates of mean while the second row shows estimates of variance. The (estimated) moments are based on a moving window sample of size 300. The number of stocks are chosen to be 150 and 200. The parameter α was chosen to 5 and $r_f = 0.01$. From the moments given by Theorem 3.3 and 3.4, we can see that α will only change the scale of the mean and variance. Hence, we have chose one value of α in this application. In this illustration, we set $\mathbf{l}_1 = (1, 0, 0, \ldots, 0)$ and $\mathbf{l}_2 = (0, 1, 0, \ldots, 0)$ in Theorem 3.4. We will showcase the mean and variance (diagonal element) of the two first elements from Theorem 3.3. These weights correspond to the M&T Bank Corporation (MTB) and Cisco Systems, Inc. (CSCO) stocks. The results can be seen in Figures 2 and 3.

In Figure 2, we can see the mean and variance for the first set of portfolios, where $c = 0.5$. There is a sharp decline in the mean of the MTB weight. It goes from a long position to quite a heavy short over night. This change is then recovered throughout the rest of the investment period. In the same Figure, in the second column we can see that there is a steady decrease in the mean of CSCO (second column, CSCO) from 2017 to 2019. After that the weights are reasonably stable with the exception for COVID, which impacts both means. Looking at the variance, second row in the same figure, we can see a slightly bigger difference between the methods. For MTB, the decrease in mean around 2019 causes an increase in the variance. Thereafter there is a steady decrease and then COVID impacts the variance, as should be expected. The variance of the second weight, for the asset CSCO, is almost twice as high. One possible explanation to it is the constant change in the weight, seen in the mean. Although not noticed in the mean for the CSCO weight, there is a sharp increase in the variance for the same period as the MTB mean drops. This is due to the fact that the variance of the portfolio weights is influenced by all assets, it contains the inverse of $\mathbf{\Sigma}$. We also expect this weight to vary much more when COVID hits.

If we now increase the portfolio size, whose results are seen in Figure 3, we can see that the mean and variance for these weights are quite different. We should now hold very little of MTB and if any position is to be taken then it seems that on average, it should be a short position. The mean for the CSCO weight is on a different scale from before which is not necessarily surprising. Portfolios in general depend on a number of assets and their relation to each other. The decrease is not the same though there is a large drop in the mean of CSCO where we previously saw a drop in in MTB around 2019. The variance, on the second row of the same figure, decreases throughout the whole period with the exception of an increase in 2019. The difference between the two methods seem to be larger in the beginning of the investment period and decrease as we go further into the period. When COVID-19 hits, the effect of using a large portfolio, the diversification effect, can be noticed in almost no change in the mean of these specific portfolio weights. However, it can be seen in the variance.

Figure 2.: Window based sample moments for the TP weights when assuming that asset returns are distributed according to the CSN distribution and its high-dimensional asymptotic counterpart. The window size is equal to 300, α is taken to be 5 while p is equal to 150.

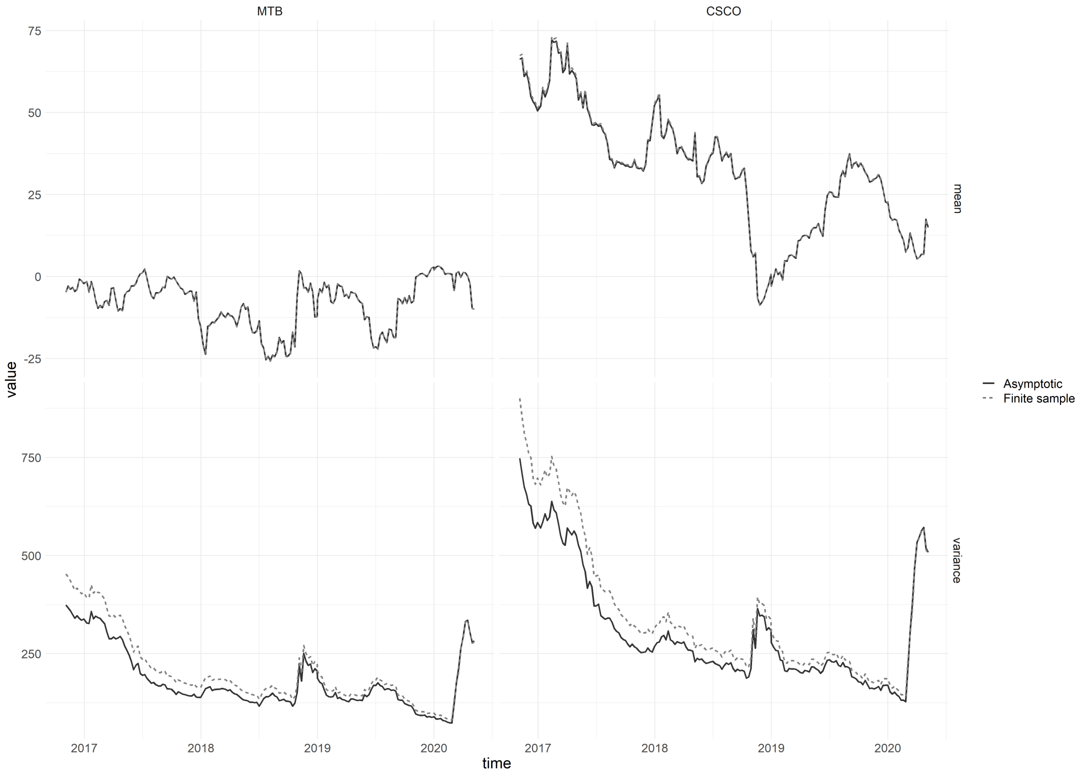


Figure 3.: Window based sample moments for the TP weights when assuming that asset returns are distributed according to the CSN distribution and its high-dimensional asymptotic counterpart. The window size is equal to 300, α is taken to be 5 while p is equal to 200.

5. Conclusions

In this paper, we investigated the implications of skewness on the sample TP weights within the context of the CSN distribution. Using the finite-sample distribution we computed the first two moments of the TP weights. Through these, we can see that skewness has implications on the estimated TP. They are biased, which means that the investor doesn't hold the correct portfolio, on average. If returns are CSN distributed then holding the sample TP implies that the investor holds the wrong portfolio. In the high-dimensional asymptotic setting, the TP is especially sensitive to the concentration ratio, the ratio between the number of assets and the number of observations used to estimate the parameters. If the concentration ratio is close to one, then our portfolio weights are extremely biased. This is common in high dimensional asymptotics, see e.g. Karlsson et al. (2020).

In the simulation and empirical study, we investigated the application of our results. The simulations showed that the finite-sample distribution can be well approximated by the high-dimensional distribution in most cases. However, for increasing values the concentration ratio the approximation becomes worse and most often in that is seen in the tails. In the empirical study, we employed a set of generalised method of moments estimators for the parameters in the CSN distribution and, thereafter, showed the difference between the derived moments of the finite-sample and the high-dimensional distribution. The mean between the two does not differ much between the two methods. The variance, on the other hand, can differ quite a lot. The difference also seems to increase with higher concentration ratios.

Throughout this paper we assumed that skewness parameter in the matrix-variate CSN distribution takes the form of $\mathbf{d}^\top \otimes \mathbf{e}^\top$ with $\mathbf{d} = \mathbf{1}_n$. This assumption on $\mathbf{d}$ means that observations are independent in time and it is known to be a limiting factor in higher-frequency data. This parameter can introduce a lot of flexibility and practical relevance in the use of the CSN distribution for modelling asset returns. This restriction is a simplification to derive the analytic results. A tradeoff between flexibility and interpretability. We aim to investigate how to approach modelling matrix-variate distributions with dependent components in future research.

Acknowledgment

The authors acknowledge financial support from the project "Models for macro and financial economics after the financial crisis" (Dnr: P18-0201) funded by the Jan Wallander and Tom Hedelius Foundation. Farrukh Javed and Stepan Mazur also acknowledge financial support from the internal research grants of Örebro University.

References

Alfelt, G. and S. Mazur (2020). On the mean and variance of the estimated tangency portfolio weights for small samples. Technical Report 8, Örebro University School of Business.

Alles, L. A. and J. L. Kling (1994). Regularities in the variation of skewness in asset returns. *Journal of financial Research 17*(3), 427–438.

Arellano-Valle, R. B. and A. Azzalini (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics 33*(3), 561–574.

Bauder, D., T. Bodnar, S. Mazur, and Y. Okhrin (2018). Bayesian inference for the tangent portfolio. *International Journal of Theoretical and Applied Finance 21*(08), 1850054.

Billingsley, P. (2013). *Convergence of probability measures*. John Wiley & Sons.

Bodnar, T., H. Dette, N. Parolya, and E. Thorsén (2021). Sampling distributions of optimal portfolio weights and characteristics in low and large dimensions. *Random Matrices: Theory and Applications*.

Bodnar, T. and A. K. Gupta (2015). Robustness of the inference procedures for the global minimum variance portfolio weights in a skew-normal model. *The European Journal of Finance 21*(13-14), 1176–1194.

Bodnar, T., S. Mazur, S. Muhinyuza, and N. Parolya (2018). On the product of a singular Wishart matrix and a singular Gaussian vector in high dimension. *Theory of Probability and Mathematical Statistics 99*(2), 37–50.

Bodnar, T., S. Mazur, and Y. Okhrin (2013). On the exact and approximate distributions of the product of a Wishart matrix with a normal vector. *Journal of Multivariate Analysis 122*, 70–81.

Bodnar, T., S. Mazur, and Y. Okhrin (2014). Distribution of the product of singular Wishart matrix and normal vector. *Theory of Probability and Mathematical Statistics 91*, 1–15.

Bodnar, T., S. Mazur, and N. Parolya (2019). Central limit theorems for functionals of large sample covariance matrix and mean vector in matrix-variate location mixture of normal distributions. *Scandinavian Journal of Statistics 46*(2), 636–660.

Bodnar, T., S. Mazur, and K. Podgórski (2016). Singular inverse Wishart distribution and its application to portfolio theory. *Journal of Multivariate Analysis 143*, 314–326.

Bodnar, T., S. Mazur, and K. Podgórski (2017). A test for the global minimum variance portfolio for small sample and singular covariance. *AStA Advances in Statistical Analysis 101*(3), 253–265.

Bodnar, T., S. Mazur, K. Podgórski, and J. Tyrcha (2019). Tangency portfolio weights for singular covariance matrix in small and large dimensions: estimation and test theory. *Journal of Statistical Planning and Inference 201*, 40–57.

Bodnar, T. and Y. Okhrin (2008). Properties of the singular, inverse and generalized inverse partitioned wishart distributions. *Journal of Multivariate Analysis 99*(10), 2389–2405.

Bodnar, T. and Y. Okhrin (2011). On the product of inverse wishart and normal distributions with applications to discriminant analysis and portfolio theory. *Scandinavian Journal of Statistics 38*(2), 311–331.

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

Bodnar, T. and M. Reiß (2016). Exact and asymptotic tests on a factor model in low and large dimensions with applications. *Journal of Multivariate Analysis 150*, 125–151.

Bodnar, T. and W. Schmid (2008). A test for the weights of the global minimum variance portfolio in an elliptical model. *Metrika 67*(2), 127–143.

Britten-Jones, M. (1999). The sampling error in estimates of mean-variance efficient portfolio weights. *The Journal of Finance 54*(2), 655–671.

Brodie, J., I. Daubechies, C. De Mol, D. Giannone, and I. Loris (2009). Sparse and stable markowitz portfolios. *Proceedings of the National Academy of Sciences of the USA 106*, 12267–12272.

Chernousova, E. and Y. Golubev (2014). Spectral cut-off regularizations for ill-posed linear models. *Mathematical Methods of Statistics 23*(2), 116–131.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues.

Domínguez-Molina, J. A., G. González-Farías, R. Ramos-Quiroga, and A. K. Gupta (2007). A matrix variate closed skew-normal distribution with applications to stochastic frontier analysis. *Communications in Statistics—Theory and Methods 36*(9), 1691–1703.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013). *Bayesian data analysis.* CRC press.

Genton, M. G. (2004). *Skew-elliptical distributions and their applications: a journey beyond normality.* CRC Press.

Gulliksson, M. and S. Mazur (2020). An iterative approach to ill-conditioned optimal portfolio selection. *Computational Economics 56*, 773–794.

Gulliksson, M., A. Oleynik, and S. Mazur (2021). Portfolio selection with a rank-deficient covariance matrix. Technical Report 12, Örebro University School of Business.

Gupta, A. K. and D. K. Nagar (2018). *Matrix variate distributions.* Chapman and Hall/CRC.

Harvey, C. R., J. C. Liechty, M. W. Liechty, and P. Müller (2010). Portfolio selection with higher moments. *Quantitative Finance 10*(5), 469–485.

Harville, D. A. (1997). *Matrix algebra from a statistician's perspective.* New York: Springer-Verlag.

Javed, F., S. Mazur, and E. Ngailo (2021). Higher order moments of the estimated tangency portfolio weights. *Journal of Applied Statistics 48*(3), 517–535.

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

Karlsson, S., S. Mazur, and S. Muhinyuza (2020). Statistical inference for the tangency portfolio in high dimension. Technical Report 10, Örebro University School of Business.

Kotsiuba, I. and S. Mazur (2016). On the asymptotic and approximate distributions of the product of an inverse wishart matrix and a gaussian vector. *Theory of Probability and Mathematical Statistics 93*, 103–112.

Kraus, A. and R. H. Litzenberger (1976). Skewness preference and the valuation of risk assets. *The Journal of finance 31*(4), 1085–1100.

Kress, R. (1999). *Linear Integral Equations.* Springer.

Ledoit, O. and M. Wolf (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *The Review of Financial Studies 30*(12), 4349–4388.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance 7*(1).

Muhinyuza, S. (2020). A test on mean-variance efficiency of the tangency portfolio in high-dimensional setting. *Theory of Probability and Mathematical Statistics*, In press.

Muhinyuza, S., T. Bodnar, and M. Lindholm (2020). A test on the location of the tangency portfolio on the set of feasible portfolios. *Applied Mathematics and Computation 386*, 125519.

Muirhead, R. (1990). *Aspects of Multivariate Statistical Theory.* New York: Wiley.

Okhrin, Y. and W. Schmid (2006). Distributional properties of portfolio weights. *Journal of econometrics 134*(1), 235–256.

Peiro, A. (1999). Skewness in financial returns. *Journal of Banking & Finance 23*(6), 847–862.

Sherman, J. and W. J. Morrison (1950). Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *The Annals of Mathematical Statistics 21*(1), 124–127.

Tikhonov, A. and V. Arsenin (1977). *Solutions of Ill-Posed Problems.* New York: Winston.

Appendix

In this section we state some results necessary for deriving the results in this paper.

Lemma 5.1. *Let* $\mathbf{X} \sim \mathcal{CSN}_{n,p;1,1}(\mathbf{1}_n \otimes \boldsymbol{\mu}^\top, \mathbf{I}_n \otimes \boldsymbol{\Sigma}, \mathbf{1}^\top \otimes \mathbf{e}^\top, 0, v)$ *with* $n > p$ *and* $\boldsymbol{\Sigma} \succ 0$*. Also, let* $\bar{\mathbf{z}} := \bar{\mathbf{x}} - r_f \mathbf{1}_p$*, where* $\bar{\mathbf{x}} = \mathbf{X}^\top \mathbf{1}_p / n$*. Furthermore, let* $\tilde{\boldsymbol{\mu}} := \boldsymbol{\mu} - r_f \mathbf{1}_p$

and $\mathbf{B}$ *be a* $p \times p$ *symmetric matrix. It then holds that*

$$\mathrm{E}\left[\bar{\mathbf{z}}^\top \mathbf{B}\bar{\mathbf{z}}\right] = \tilde{\boldsymbol{\mu}}^\top \mathbf{B}\tilde{\boldsymbol{\mu}} + \frac{\mathrm{tr}(\boldsymbol{\Sigma}\mathbf{B})}{n} + 2\sqrt{\frac{2}{\pi}}\frac{\mathbf{e}^\top \boldsymbol{\Sigma}\mathbf{B}\tilde{\boldsymbol{\mu}}}{(v + n\mathbf{e}^\top \boldsymbol{\Sigma}\mathbf{e})^{1/2}}. \tag{5.1}$$

Proof of Lemma 5.1. From the properties of the trace we have that

$$\begin{aligned}
\mathrm{E}\left[\bar{\mathbf{z}}^\top \mathbf{B}\bar{\mathbf{z}}\right] &= \mathrm{E}\left[\mathrm{tr}\left(\bar{\mathbf{z}}^\top \mathbf{B}\bar{\mathbf{z}}\right)\right] \\
&= \mathrm{E}\left[\mathrm{tr}\left(\mathbf{B}\bar{\mathbf{z}}\bar{\mathbf{z}}^\top\right)\right] \\
&= \mathrm{tr}\left(\mathbf{B}\,\mathrm{E}\left[\bar{\mathbf{z}}\bar{\mathbf{z}}^\top\right]\right) \\
&= \mathrm{tr}\left(\mathbf{B}\,\mathrm{Var}(\bar{\mathbf{z}})\right) + \mathrm{tr}\left(\mathbf{B}\,\mathrm{E}\left[\bar{\mathbf{z}}\right]\mathrm{E}\left[\bar{\mathbf{z}}\right]^\top\right).
\end{aligned}$$

Using the moments from (3.1), we receive the desired results. □