



Mathematical Statistics
Stockholm University

**A test on mean-variance efficiency of the
tangency portfolio in high dimensional setting**

Stanislas Muhinyuza

Research Report 2019:13

ISSN 1650-0377

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

<http://www.math.su.se>



A test on mean-variance efficiency of the tangency portfolio in high dimensional setting

Stanislas Muhinyuza

September 2019

Abstract

In this paper we derive the asymptotic distribution of the test of the efficiency of the tangency portfolio in high dimensional settings, namely when both the portfolio dimension and the sample size grow to infinity. Moreover, we propose a new test based on the estimator for the slope parameter of the efficient frontier in the mean-variance space when there is a possibility in investing into the riskless asset, and derive the asymptotic distribution of that test statistic under both the null and alternative hypotheses. Additionally, we study the finite sample performance of the derived theoretical results via simulations.

Keywords Tangency Portfolio, Mean-variance portfolio, High dimension settings.

1 Introduction

Since the introduction of mean-variance theory by Markowitz (1952), a large number of papers devoted to the optimal portfolio selection have been published and brought remarkable contributions in different avenues of finance, be it in research or in practice. Since then, the problem of testing the efficiency of a given portfolio has gained a lot of attention (Bodnar and Schmid (2008); Bodnar et al. (2019a); Gibbons et al. (1989); Glombek (2014); Britten-Jones (1999); Muhinyuza et al. (2017)) to just name a few. In this regard, the aim of the investor is to find an optimal portfolio that minimizes the risk, i.e, the variance of the portfolio for a given level of the expected return. In the absence of a risk-free asset, the risk aversion strategy leads to minimal variance portfolio. Merton (1972) showed that all Markowitz' optimal portfolios lie on the upper part of the parabola in the mean-variance space (known as Efficient Frontier(EF)) and its equation is given by

$$(R - R_{GMV})^2 = s(V - V_{GMV}) \quad (1)$$

where

$$R_{GMV} = \frac{\mathbf{1}\Sigma^{-1}\boldsymbol{\mu}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}, V_{GMV} = \frac{1}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}, \text{ and } s = \boldsymbol{\mu}'\mathbf{R}\boldsymbol{\mu} \text{ with } \mathbf{R} = \Sigma^{-1} - \frac{\Sigma^{-1}\mathbf{1}\mathbf{1}'\Sigma^{-1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}. \quad (2)$$

Whereas, in the presence of a risk-free asset, the tangency portfolio (TP), i.e, a linear combination of risky assets and a risk-free asset needs to be considered and the equation of the efficiency frontier in case of the presence of risk-free asset is given by

$$(R - r_f)^2 = \lambda V \quad (3)$$

where

$$\lambda = (\boldsymbol{\mu} - r_f\mathbf{1})'\Sigma^{-1}(\boldsymbol{\mu} - r_f\mathbf{1}). \quad (4)$$

In practice, this investment theory appears to be challenging because of the presence of the sampling error while estimating the unknown theoretical quantities. A big number of the literature in this area of research treats the case of classical asymptotics (the sample size n increases while the size of the portfolio k remains constant). In this situation, the plug-in estimator (sample estimator) of the optimal portfolio turns to be a good estimator due to its attractive properties, namely the consistency and asymptotic normality.

Nowadays, in several applications the number of assets k in a portfolio are comparable to the sample size n , i.e., the portfolio size k and the sample size n grows to infinity at the same order, that means $\frac{k}{n} \rightarrow c \in (0, \infty)$. In this situation the traditional asymptotic theory cannot be applied because of the failure in delivering consistent estimators of the unknown parameters of the assets returns, namely, the mean vector and the covariance

matrix. A number of papers treat the high dimensional asymptotics in portfolio theory by the help of the results from random matrix theory (see, e.g., Frahm and Jaekel (2008); Glombek (2014); Bodnar et al. (2019b, 2016b, 2018)). Recently, Bodnar et al. (2019a) studied the distributional properties of the estimated TP weights and suggested inference procedures in small and high-dimensions. Furthermore, they delivered the high-dimensional asymptotic distribution of the estimated TP weights and they proposed a test statistic when both the population and sample covariance matrices are singular. This paper complements the existing literature in different ways. It provides the asymptotics of the test statistic for testing the existence of the EF and for testing the efficiency of the TP under high-dimensional regime.

The rest of the paper is structured as follow: Section 2 discusses the main results of the two provided tests including the stochastic representations of their test statistics. These stochastic representations are later used to obtain the high dimensional asymptotic distributions of the test statistics under both the null and the alternative hypotheses. It also provides the power functions of the suggested tests for different values of c . Section 3 presents the results of the simulation study, in which we compare the performance of the two proposed tests, while the concluding remarks are given in the Section 4.

2 Test theory on the location of the TP in high dimension

Through out this section, we assume $\mathbf{x}_1, \dots, \mathbf{x}_n$ to be independent k -dimensional sample of the asset returns, with $\mathbf{E}(\mathbf{x}_t) = \boldsymbol{\mu}$ and $\mathbf{cov}(\mathbf{x}_t) = \boldsymbol{\Sigma}$, for $t = 1, \dots, n$, where $\boldsymbol{\Sigma}$ is assumed to be positive definite.

2.1 A test on the existence of the EF based on the slope parameter

The slope parameter plays an important role in the construction of the efficient frontier. It shows how the market is profitable, i.e. how large is the increase in the portfolio profit in relation to the unit increase of the portfolio variance. If the slope parameters is zero, the population efficient frontier reduces to a straight line. In this case, the GMV portfolio is the only available investment. If there is a possibility to invest in the risk-free asset with return r_f , a part of the investor wealth may be invested into the riskless asset and it may reduce the variance, whereas the rest of the wealth can be invested into the risky assets. In this case, a test for the existence of efficient frontier would be of importance and its hypotheses are given by

$$H_0 : \lambda = 0 \text{ against } H_1 : \lambda > 0, \quad (5)$$

where λ is defined in (4). The rejection of the null hypothesis ensures the existence of the efficient frontier, i.e. it confirms that the positiveness of the slope parameter ($\lambda > 0$), and the investor has a number of investment options to choose from including the TP. On the other hand, the non rejection of the null hypothesis means that the slope coefficient of the efficient frontier is equal to zero. In this case the GMV portfolio is the only available portfolio for investment, and also the allocation of the whole wealth into the riskless asset could be considered as a suitable alternative.

The test statistic for testing (5) is based on the derivation in Bodnar and Schmid (2009) and is given by

$$T_\lambda = \frac{n(n-k)}{(n-1)k} (\hat{\boldsymbol{\mu}} - r_f \mathbf{1})' \hat{\boldsymbol{\Sigma}}^{-1} (\hat{\boldsymbol{\mu}} - r_f \mathbf{1}), \quad (6)$$

with

$$\hat{\boldsymbol{\mu}} = \frac{1}{n} \sum_{t=1}^n \mathbf{x}_t \text{ and } \hat{\boldsymbol{\Sigma}} = \frac{1}{n-1} \sum_{t=1}^n (\mathbf{x}_t - \hat{\boldsymbol{\mu}})(\mathbf{x}_t - \hat{\boldsymbol{\mu}})' \quad (7)$$

are the sample mean vector and the sample covariance matrix, respectively. The distribution of the test statistic T_λ in the equation (6) is given in the following proposition

Proposition 1. *Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be i.i.d random vectors with $\mathbf{x}_1 \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $k < n$. $\boldsymbol{\Sigma}$ is assumed to be positive definite. Then it holds that*

- a) $T_\lambda \sim \mathcal{F}_{k, n-k, \delta_T}$ under H_1 , where $\delta_T = n\lambda$ with λ as defined in (4);
- b) $T_\lambda \sim \mathcal{F}_{k, n-k}$ under H_0 .

Proof of proposition 1. From Theorem 3.1.2 of Muirhead (1982), it follows that

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

where $W_k(n-1, \boldsymbol{\Sigma})$ denotes a k -dimensional Wishart distribution with $n-1$ degree of freedom and the parameter matrix $\boldsymbol{\Sigma}$. On top of that, $(\hat{\boldsymbol{\mu}} - r_f \mathbf{1})$ and $(n-1)\mathbf{S}$ are independent. Applying the results of Theorem 6.7a.1 in Mathai and Provost (1992), we get the statement of the proposition. The statement of Theorem 1(b) follows by setting $\lambda = 0$ under the null hypothesis. \square

Alternatively, one can represent the distribution of T_λ using the stochastic represen-

tation given below

$$T_\lambda = \frac{n - k (\sqrt{n\lambda} + z_1)^2 + \zeta_1}{k \zeta_2} \quad (8)$$

where $z_1 \sim \mathcal{N}(0, 1)$, $\zeta_1 \sim \chi_{k-1}^2$ and $\zeta_2 \sim \chi_{n-k}^2$. Moreover, z_1, ζ_1 and ζ_2 are independent.

From Proposition 1, it is remarkable that the density function of the statistic T_λ depends on the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ only over the non-centrality parameter δ_T . Thus, the exact power function of the test can be easily computed using any mathematical software package. However, some numerical difficulties may be encountered when the power function of the test is computed for large values of k and n . To address this problem, we derive the asymptotic distribution of T_λ for high-dimensional setting. This result is given in Theorem 1. We note that for finite case, this result has been used to test the efficiency of any portfolio from the efficient frontier with respect to the GMV portfolio (Bodnar and Schmid (2009); Bodnar and Bodnar (2010)).

We also observe that, because of positive definiteness of the covariance matrix, the situation under the null hypothesis occurs only if $\boldsymbol{\mu} = r_f \mathbf{1}$. Moreover, the same test is used to test the hypothesis $H_0 : \mu_1 = \dots = \mu_k$ (see, e.g., Rencher and Christensen (2012)).

Theorem 1. *Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be i.i.d random vectors with $\mathbf{x}_1 \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $k < n$. $\boldsymbol{\Sigma}$ is assumed to be positive definite. Then it holds that*

$$\sqrt{k} \left(\frac{T_\lambda - 1 - \frac{n}{k} \lambda}{\sigma_{T_\lambda}} \right) \xrightarrow{d} \mathcal{N}(0, 1)$$

where

$$\sigma_{T_\lambda}^2 = 2 + 4 \frac{\lambda}{c} + 2 \frac{c}{1-c} \left(1 + \frac{\lambda}{c} \right)^2$$

for $k/n \rightarrow c \in (0, 1)$ as $n \rightarrow \infty$. Under the null hypothesis, $\sqrt{k} (T_\lambda - 1) \xrightarrow{d} \mathcal{N}(0, 2/(1-c))$ for $k/n \rightarrow c \in (0, 1)$ as $n \rightarrow \infty$.

Proof of Theorem 1. Using the stochastic representation given in equation (8), it holds that

$$\begin{aligned}
T_\lambda - 1 - \frac{n}{k}\lambda &= \frac{n-k}{k} \frac{(\sqrt{n\lambda} + z_1)^2 + \zeta_1}{\zeta_2} - 1 - \frac{n}{k}\lambda \\
&= \frac{n-k}{\zeta_2} \left(\frac{\zeta_1 + n\lambda + 2\sqrt{n\lambda}z_1 + z_1^2}{k} - \left(1 + \frac{n}{k}\lambda\right) \frac{\zeta_2}{n-k} \right) \\
&= \frac{n-k}{\zeta_2} \left(\frac{\zeta_1 + n\lambda + 2\sqrt{n\lambda}z_1 + z_1^2}{k} - \left(1 + \frac{n}{k}\lambda\right) + \left(1 + \frac{n}{k}\lambda\right) - \left(1 + \frac{n}{k}\lambda\right) \frac{\zeta_2}{n-k} \right) \\
&= \frac{n-k}{\zeta_2} \left(\frac{\zeta_1}{k} + \frac{n}{k}\lambda + 2\frac{\sqrt{n\lambda}}{k}z_1 + \frac{z_1^2}{k} - 1 - \frac{n}{k}\lambda - \left(1 + \frac{n}{k}\lambda\right) \left(\frac{\zeta_2}{n-k} - 1\right) \right) \\
&= \frac{n-k}{\zeta_2} \left(\left(\frac{\zeta_1}{k} - 1\right) - \left(1 + \frac{n}{k}\lambda\right) \left(\frac{\zeta_2}{n-k} - 1\right) + 2\frac{\sqrt{n\lambda}}{k}z_1 + \frac{z_1^2}{k} \right)
\end{aligned}$$

We then have

$$\sqrt{k} \left(T_\lambda - 1 - \frac{n}{k}\lambda \right) = \frac{n-k}{\zeta_2} \left(\sqrt{k} \left(\frac{\zeta_1}{k} - 1 \right) - \left(1 + \frac{n}{k}\lambda\right) \sqrt{k} \left(\frac{\zeta_2}{n-k} - 1 \right) + 2\frac{\sqrt{n\lambda}}{\sqrt{k}}z_1 + \frac{z_1^2}{\sqrt{k}} \right)$$

Using Lemma 3 in Bodnar and Reiß (2016) and the proof of Lemma 4 in Bodnar et al. (2016b), we obtain the following results:

$$\frac{\zeta_2}{n-k} \xrightarrow{a.s.} 1, 1 + \frac{n}{k}\lambda \xrightarrow{a.s.} 1 + \frac{\lambda}{c}, \sqrt{k} \left(\frac{\zeta_1}{k} - 1 \right) \xrightarrow{d} \mathcal{N} \left(0, 2\frac{c}{1-c} \right) \text{ and } 2\frac{\sqrt{n\lambda}}{\sqrt{k}}z_1 \xrightarrow{d} \mathcal{N} \left(0, 4\frac{\lambda}{c} \right).$$

The fact that z_1 , ζ_1 and ζ_2 are independent, the application of Slutsky's lemma (see, e.g., Theorem 2.8 in Van der Vaart (2000)) gives us

$$\sqrt{k} \left(\frac{T_\lambda - 1 - \frac{n}{k}\lambda}{\sigma_{T_\lambda}} \right) \xrightarrow{d} \mathcal{N}(0, 1)$$

where

$$\sigma_{T_\lambda}^2 = 2 + 4\frac{\lambda}{c} + 2\frac{c}{1-c} \left(1 + \frac{\lambda}{c} \right)^2$$

□

The application of Theorem 1 leads to an asymptotic expression of the power function given by

$$P\left(\frac{\sqrt{k}(T_\lambda - 1)}{\sqrt{2/(1-c)}} > z_{1-\alpha}\right) = 1 - P\left(\frac{\sqrt{k}T_\lambda - 1 - \frac{n}{k}\lambda}{\sigma_{T_\lambda}} \leq \frac{\sqrt{2/(1-c)}z_{1-\alpha} - \sqrt{k}\frac{n}{k}\lambda}{\sigma_{T_\lambda}}\right) \quad (9)$$

$$\approx 1 - \Phi\left(\frac{\sqrt{2/(1-c)}z_{1-\alpha} - \sqrt{k}\frac{\lambda}{c}}{\sigma_{T_\lambda}}\right) \quad (10)$$

where $z_{1-\alpha}$ is the $(1 - \alpha)$ -quantile of the standard normal distribution.

In Figures 1 and 2 we plot the power function (10) as a function of λ for different values of c and n as a solid line. We additionally plot the empirical power of the test for the same values of c and n as dashed line, and it is interpreted as the number of rejections of the null hypothesis obtained via a simulation study. As it can be seen from the proof of Theorem 1, instead of generating the huge random matrix of order $k \times n$ of asset returns in each simulation run, we instead simulate three independent random variables from standard univariate distributions and compute the statistic T_λ for given values of λ in the stochastic representation given in equation (8). The following algorithm can be used to compute the asymptotic power:

(i) generate independently $z_1^{(b)} \sim \mathcal{N}(0, 1)$, $\zeta_1^{(b)} \sim \chi_{k-1}^2$ and $\zeta_2^{(b)} \sim \chi_{n-k}^2$;

(ii) For fixed λ , compute

$$T_s^{(b)} = \frac{n - k}{k} \frac{(\sqrt{n\lambda} + z_1^{(b)})^2 + \zeta_1^{(b)}}{\zeta_2^{(b)}}$$

(iii) repeat steps (i)-(ii) for $b = 1, \dots, B$ and approximate P by

$$\hat{P} = \frac{1}{B} \sum_{b=1}^B \mathbf{1}_{\{z_{1-\alpha, +\infty}\}} \left(\sqrt{k} \frac{(T_\lambda^{(b)} - 1)}{\sqrt{2/(1-c)}} \right)$$

where $\mathbf{1}_{\{\mathcal{A}\}}$ is the indicator function of set \mathcal{A}

In Figures 1 and 2 we observe that the high dimensional expression of the power function provides a reliable approximation of the true power function. We also note that for small values of the concentration constant the two power functions are indistinguishable while a moderate discrepancy is present for large values of the concentration coefficient c .

2.2 Test on the location of the TP on EF in high dimensions

The location of the TP on the EF depends crucially on the relation between the risk-free rate r_f and the expected return of the global minimum variance portfolio (GMVP). At each time point, the investor wants to check whether holding TP is mean-variance

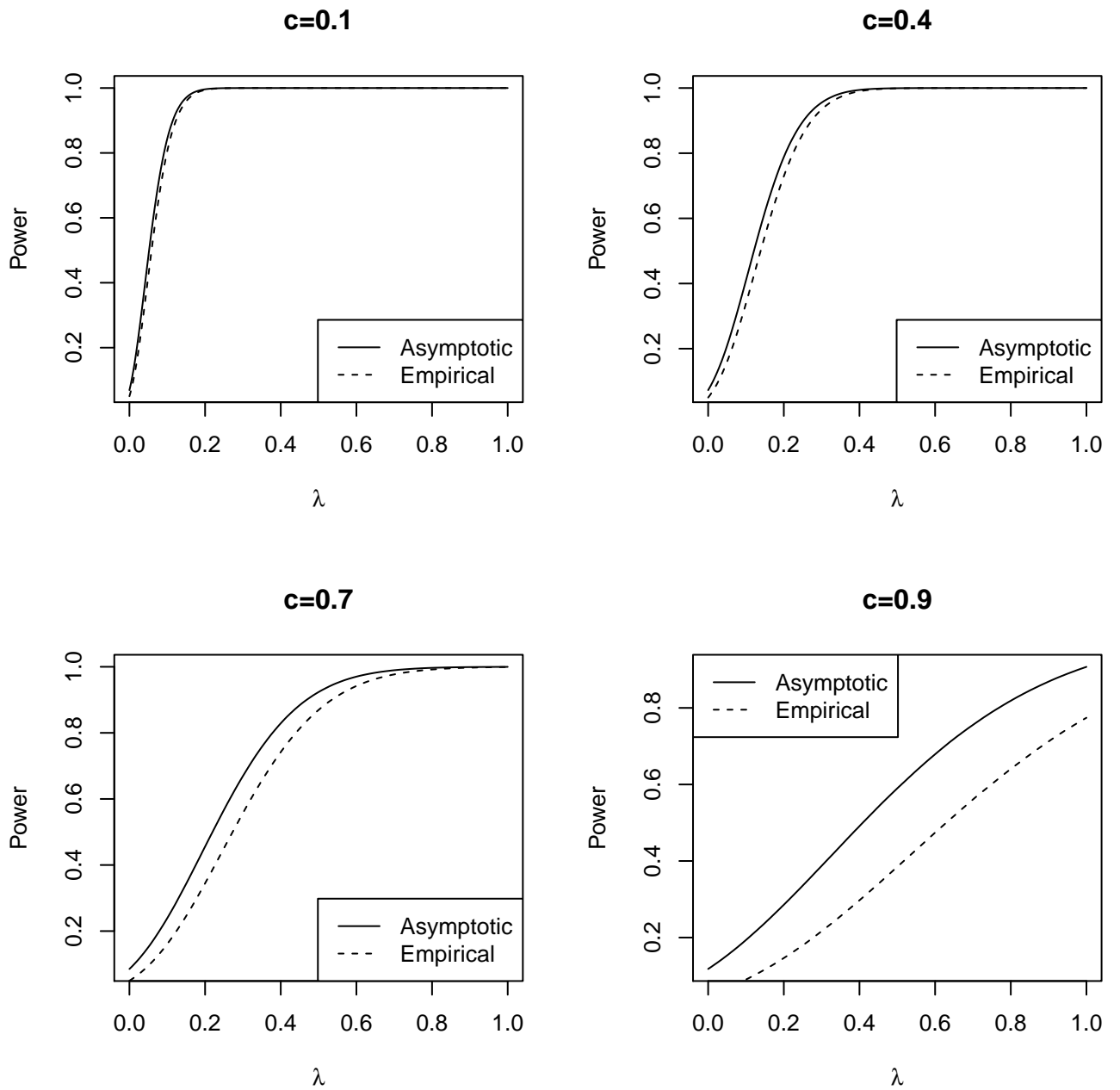


Figure 1: Asymptotic power versus empirical power for different values of c as a function of λ , significance level 5% and $n = 250$.

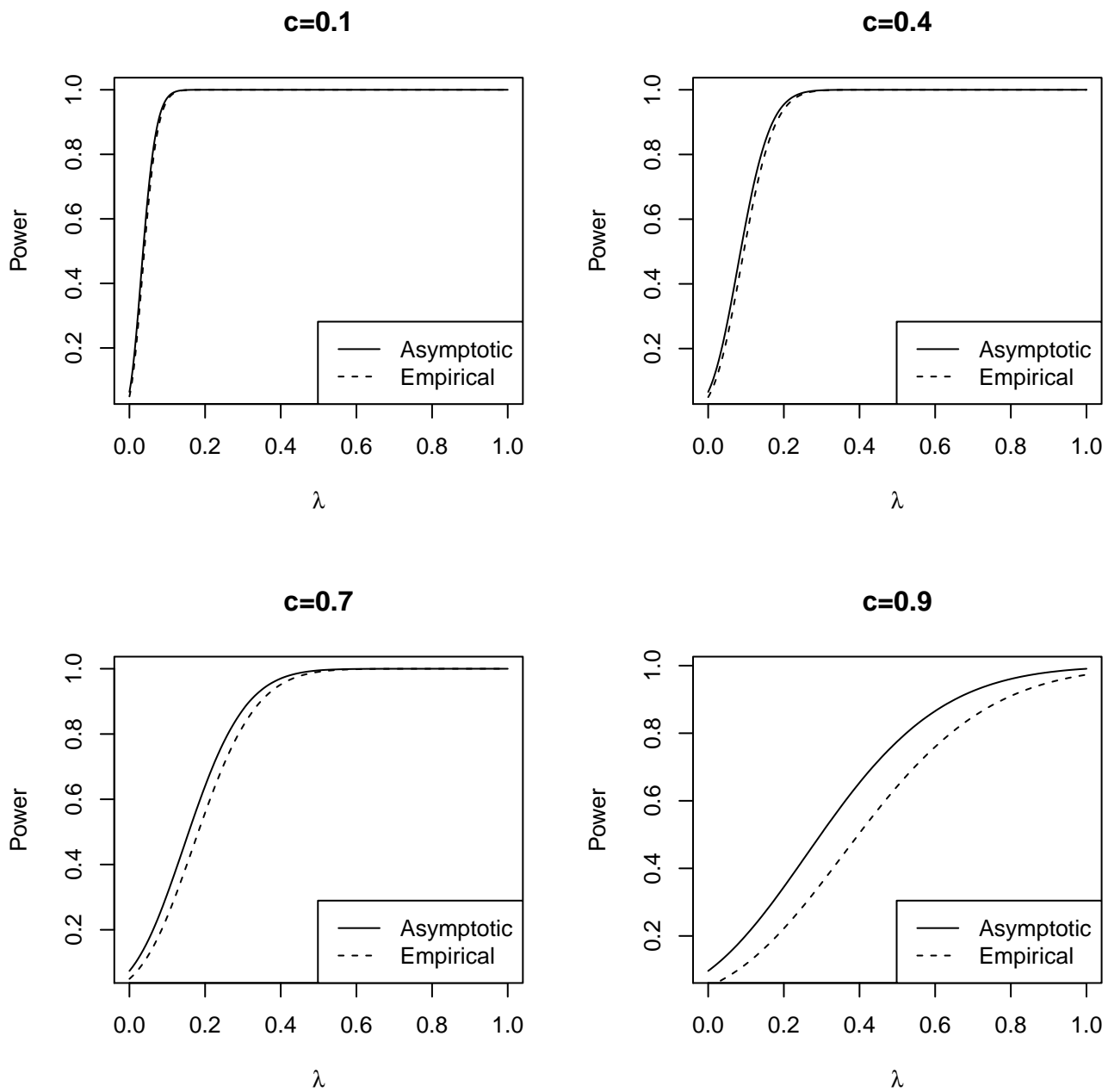


Figure 2: Asymptotic power versus empirical power for different values of c as a function of λ , significance level 5% and $n = 500$.

efficient or it has to be reconstructed. For that reason, this problem can be formalised as a statistical test problem, with the following hypotheses

$$H_0 : R_{GMV} \leq r_f \text{ versus } H_1 : R_{GMV} > r_f \quad (11)$$

Rejecting H_0 in (11), means that the TP is mean-variance efficient while the non rejection of H_0 in (5) does not guarantee the efficiency of the TP. In the case of non rejection of H_0 , the investor cannot be sure of the optimality of the TP and the investment in the risk-free rate could be considered as a suitable alternative.

Muhinyuza et al. (2017) proposed the following test statistics for (11)

$$T = \frac{\sqrt{n-k}}{\sqrt{n-1}} \frac{\hat{R}_{GMV} - r_f}{\sqrt{1 + \frac{n}{n-1} \hat{s} \sqrt{\frac{\hat{V}_{GMV}}{n}}}} \quad (12)$$

where \hat{R}_{GMV} , \hat{V}_{GMV} and \hat{s} are the sample estimators for R_{GMV} , V_{GMV} and s given by

$$\hat{R}_{GMV} = \frac{\mathbf{1}' \hat{\Sigma}^{-1} \hat{\boldsymbol{\mu}}}{\mathbf{1}' \hat{\Sigma}^{-1} \mathbf{1}} \text{ and } \hat{V}_{GMV} = \frac{1}{\mathbf{1}' \hat{\Sigma}^{-1} \mathbf{1}} \quad (13)$$

and

$$\hat{s} = \hat{\boldsymbol{\mu}}' \hat{\mathbf{R}} \hat{\boldsymbol{\mu}} \text{ with } \hat{\mathbf{R}} = \hat{\Sigma}^{-1} - \frac{\hat{\Sigma}^{-1} \mathbf{1} \mathbf{1}' \hat{\Sigma}^{-1}}{\mathbf{1}' \hat{\Sigma}^{-1} \mathbf{1}}. \quad (14)$$

Moreover, they provided the distribution of T in form of a density function and a stochastic representation under both the null and alternative hypotheses (see, Muhinyuza et al. (2017, Proposition 1, Theorem 1)). The following proposition summarizes the distribution of T .

Proposition 2. *Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be i.i.d random vectors with $\mathbf{x}_1 \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $k < n$. $\boldsymbol{\Sigma}$ is assumed to be positive definite. Then*

(a) *the density of T is given by*

$$f_T(x) = \frac{n(n-k+1)}{(k-1)(n-1)} \int_0^\infty f_{t_{n-k, \delta(y)}}(x) f_{F_{k-1, n-k+1, ns}}\left(\frac{n(n-k+1)}{(k-1)(n-1)}y\right) dy \quad (15)$$

where

$$\delta(y) = \frac{\sqrt{n}}{\sqrt{1 + \frac{n}{n-1}y}} S_{GMV} \text{ with } S_{GMV} = \frac{R_{GMV} - r_f}{\sqrt{V_{GMV}}}$$

is the Sharpe ratio of the GMVP. The slope parameter $s = \boldsymbol{\mu}' \mathbf{R} \boldsymbol{\mu}$ with $\mathbf{R} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{1}' \boldsymbol{\Sigma}^{-1} / \mathbf{1}' \boldsymbol{\Sigma}^{-1} \mathbf{1}$.

(b) the stochastic representation of T is given by

$$T \stackrel{d}{=} \frac{\sqrt{n-k}}{\sqrt{\xi}} \frac{1}{\sqrt{1 + \frac{\xi_3 + (\sqrt{ns} + z_1)^2}{\xi_2}}} \left(\sqrt{n} S_{GMV} + z_2 + \sqrt{\frac{\xi_3 + (\sqrt{ns} + z_1)^2}{\xi_2}} z_3 \right) \quad (16)$$

where $z_1, z_2, z_3 \sim \mathcal{N}(0, 1)$, $\xi \sim \chi_{n-k}^2$, $\xi_2 \sim \chi_{n-k+1}^2$, $\xi_3 \sim \chi_{k-2}^2$ are mutually independent.

It is seen that the distribution of test statistic in equation (12) can be represented as a mixture of a non-central t distribution with $n - k$ degrees of freedom and a non-centrality parameter $\delta(y)$, or can be given in the form of a stochastic representation as in the equation (16).

Furthermore, using the results from Proposition 2 the power function of the test can be easily obtained and it is easy to see that it depends on the parameter $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ through s and S_{GMV} . However, this approach may encounter some difficulties for the large values of k and n , since $\boldsymbol{\Sigma}$ becomes unstable for large values of k . To deal with this problem, we derive the asymptotic distribution of T in a high-dimensional environment. We assume that $k/n \rightarrow c \in (0, 1)$ as $n \rightarrow \infty$. No further relation is imposed between k and n . Note that, under high dimensional setting, the usual estimators for the precision matrix (Inverse of the covariance matrix) performs poorly and are not consistent anymore (Bodnar et al. (2016a)). Therefore, it is worth to study the behaviour of the test statistic developed by Muhinyuza et al. (2017) under high dimensional environment and propose an alternative test which takes into account the correction of the estimated precision matrix.

The next theorem presents the asymptotic distribution of T under double asymptotic regime.

Theorem 2. Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be i.i.d random vectors with $\mathbf{x}_1 \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $k < n$. $\boldsymbol{\Sigma}$ is assumed to be positive definite. Let $\frac{k}{n} \rightarrow c \in (0, 1)$ as $n \rightarrow \infty$. Then, it holds that the asymptotic distribution of T is given by

(a)

$$\sigma_T^{-1} \left(T - \sqrt{n} \frac{S_{GMV}}{\sqrt{1 + \frac{k-1}{n-k+1} \left(1 + \frac{n}{k-1} s\right)}} \right) \xrightarrow{d} \mathcal{N}(0, 1) \quad (17)$$

where

$$\sigma_T^2 = 1 + \frac{S_{GMV}^2}{1+s} \left(\frac{1}{2} + \frac{1}{2} \frac{s^2 + 2s + c}{(1+s)^2} \right) \quad (18)$$

for $\frac{k}{n} \rightarrow c \in (0, 1)$ as $n \rightarrow \infty$.

(b) Under the null hypothesis it holds that $T \sim N(0, 1)$

Proof of Theorem 2. From the stochastic representation given in (16) and using the properties of a normally distributed random variable, we have that

$$T \stackrel{d}{=} \sqrt{\frac{n-k}{\xi}} \frac{\sqrt{n} S_{GMV}}{\sqrt{1 + \frac{\zeta}{\xi_2}}} + z_0 \sqrt{\frac{n-k}{\xi}} \quad (19)$$

where $z_0 = \frac{z_5}{\sqrt{1 + \frac{\zeta}{\xi_2}}} \sim \mathcal{N}(0, 1)$, $\xi \sim \chi_{n-k}^2$, $\xi_2 \sim \chi_{n-k+1}^2$, $\zeta \sim \chi_{k-1, ns}^2$, and on top of that z_0, ζ, ξ and ξ_2 are mutually independent.

We then have

$$T - \sqrt{n} \frac{S_{GMV}}{\sqrt{1 + \frac{k-1}{n-k+1} \left(1 + \frac{n}{k-1} s\right)}} = \sqrt{\frac{n-k}{\xi}} z_0 + \sqrt{n} \sqrt{\frac{n-k}{\xi}} \frac{S_{GMV}}{\sqrt{1 + \frac{\zeta}{\xi_2}}} - \sqrt{n} \frac{S_{GMV}}{\sqrt{1 + \frac{k-1}{n-k+1} \left(1 + \frac{n}{k-1} s\right)}}$$

Adding and subtracting $\sqrt{n} \frac{S_{GMV}}{\sqrt{1 + \frac{\zeta}{\xi_2}}}$ and factoring out $\frac{S_{GMV}}{\sqrt{1 + \frac{\zeta}{\xi_2}}}$ and rearranging we get

$$\begin{aligned} & T - \sqrt{n} \frac{S_{GMV}}{\sqrt{1 + \frac{k-1}{n-k+1} \left(1 + \frac{n}{k-1} s\right)}} \\ &= \sqrt{\frac{n-k}{\xi}} z_0 + \frac{S_{GMV}}{\sqrt{1 + \frac{\zeta}{\xi_2}}} \left(\sqrt{n} \left(\sqrt{\frac{n-k}{\xi}} - 1 \right) + \sqrt{n} \left(1 - \frac{\sqrt{1 + \frac{\zeta}{\xi_2}}}{\sqrt{1 + \frac{k-1}{n-k+1} \left(1 + \frac{n}{k-1} s\right)}} \right) \right) \end{aligned}$$

Putting the last expression on common denominator and multiplying the numerator by its conjugate, we obtain

$$\begin{aligned} & \sqrt{1 + \frac{k-1}{n-k+1} \left(1 + \frac{n}{k-1} s\right)} - \sqrt{1 + \frac{\zeta}{\xi_2}} \\ &= \frac{\frac{k-1}{n-k+1} \left(1 + \frac{n}{k-1} s\right) - \frac{\zeta}{\xi_2}}{\sqrt{1 + \frac{k-1}{n-k+1} \left(1 + \frac{n}{k-1} s\right)} + \sqrt{1 + \frac{\zeta}{\xi_2}}} \end{aligned}$$

We then have

$$\frac{k-1}{n-k+1} \left(1 + \frac{n}{k-1} s\right) - \frac{\zeta}{\xi_2} = \frac{k-1}{n-k+1} \left(1 + \frac{n}{k-1} s - \frac{\zeta/(k-1)}{\xi_2/(n-k+1)}\right) \quad (20)$$

Putting (20) on common denominator and rearranging it, we get

$$1 + \frac{n}{k-1} s - \frac{\zeta/(k-1)}{\xi_2/(n-k+1)} = \frac{n-k+1}{\xi_2} \left[\left(1 + \frac{n}{k-1} s\right) \left(\frac{\xi_2}{n-k+1} - 1\right) - \left(\frac{\zeta}{k-1} - 1 - \frac{n}{k-1} s\right) \right] (2)$$

We also have that

$$\sqrt{\frac{n-k}{\xi}} - 1 = \left(\sqrt{\frac{n-k}{\xi}} - 1 \right) \frac{\sqrt{\frac{n-k}{\xi}} + 1}{\sqrt{\frac{n-k}{\xi}} + 1} = \frac{\frac{n-k}{\xi} - 1}{\sqrt{\frac{n-k}{\xi}} + 1} = \frac{\left(1 - \frac{\xi}{n-k}\right) \frac{n-k}{\xi}}{\sqrt{\frac{n-k}{\xi}} + 1} \quad (22)$$

Considering (21) and (22), we then get

$$\begin{aligned} T &= \sqrt{n} \frac{S_{GMV}}{\sqrt{1 + \frac{k-1}{n-k+1} \left(1 + \frac{n}{k-1}s\right)}} = \sqrt{\frac{n-k}{\xi}} z_0 + \frac{S_{GMV}}{\sqrt{1 + \frac{\zeta}{\xi_2}}} \left(\frac{\sqrt{n} \left(1 - \frac{\xi}{n-k}\right) \frac{n-k}{\xi}}{\sqrt{\frac{n-k}{\xi}} + 1} \right) \\ &+ \frac{S_{GMV}}{\sqrt{1 + \frac{\zeta}{\xi_2}}} \left(\frac{\frac{k-1}{n-k+1} \left[\frac{n-k+1}{\xi_2} \left(\left(1 + \frac{n}{k-1}s\right) \sqrt{n} \left(\frac{\xi_2}{n-k+1} - 1 \right) - \sqrt{n} \left(\frac{\zeta}{k-1} - 1 - \frac{n}{k-1}s \right) \right) \right]}{\sqrt{1 + \frac{k-1}{n-k+1} \left(1 + \frac{n}{k-1}s\right)} \left(\sqrt{1 + \frac{k-1}{n-k+1} \left(1 + \frac{n}{k-1}s\right)} + \sqrt{1 + \frac{\zeta}{\xi_2}} \right)} \right) \end{aligned}$$

Using Lemma 3(a) in Bodnar and Reiß (2016), we obtain

$$\frac{\xi}{n-k} \xrightarrow{a.s.} 1, \frac{\xi_2}{n-k+1} \xrightarrow{a.s.} 1 \text{ and } \frac{\zeta}{k-1} - 1 - \frac{n}{k-1}s \xrightarrow{a.s.} 0 \Rightarrow \frac{\zeta}{k-1} \xrightarrow{a.s.} 1 + \frac{s}{c} \quad (23)$$

Consequently,

$$\sqrt{1 + \frac{\zeta}{\xi_2}} = \sqrt{1 + \frac{\zeta}{\xi_2} \frac{k-1}{k-1} \frac{n-k+1}{n-k+1}} = \sqrt{1 + \frac{\zeta}{k-1} \frac{n-k+1}{\xi_2} \frac{k-1}{n-k+1}} \xrightarrow{a.s.} \sqrt{\frac{1+s}{1-c}} \quad (24)$$

and

$$\sqrt{1 + \frac{k-1}{n-k+1} \left(1 + \frac{n}{k-1}s\right)} \xrightarrow{a.s.} \sqrt{\frac{1+s}{1-c}} \quad (25)$$

By using Lemma 3(b) in Bodnar and Reiß (2016) and the proof of Lemma 4 in Bodnar et al. (2016b), we get

$$\sqrt{n} \left(\frac{\xi}{n-k} - 1 \right) \xrightarrow{d} \mathcal{N}(0, 2/(1-c)), \quad (26)$$

$$\left(1 + \frac{n}{k-1}s\right) \sqrt{n} \left(\frac{\xi_2}{n-k+1} - 1 \right) \xrightarrow{d} \mathcal{N}\left(0, \frac{2}{1-c} \left(1 + \frac{s}{c}\right)^2\right), \quad (27)$$

and

$$\sqrt{n} \left(\frac{\zeta}{k-1} - 1 - \frac{n}{k-1}s \right) \xrightarrow{d} \mathcal{N}\left(0, \frac{2}{c} \left(1 + 2\frac{s}{c}\right)\right) \quad (28)$$

for $k/n \rightarrow c \in (0, 1)$ as $n \rightarrow \infty$.

We also know that

$$\sqrt{n} \left(z_0 / \sqrt{n} \right) \xrightarrow{d} \mathcal{N}(0, 1) \quad (29)$$

Since z_0, ξ, ξ_2 and ζ are independent and taking into account (23),(24),(25), (26),(27),(28) and (29), we obtain the following asymptotic result

$$T - \sqrt{n} \frac{S_{GMV}}{\sqrt{1 + \frac{k-1}{n-k+1} \left(1 + \frac{n}{k-1}s\right)}} \xrightarrow{d} \bar{z}_0 + \frac{S_{GMV}}{\sqrt{\frac{1+s}{1-c}}} \frac{\bar{z}_1}{2} + \frac{S_{GMV}}{\sqrt{\frac{1+s}{1-c}}} \left(\frac{c/(1-c)(\bar{z}_2 + \bar{z}_3)}{2\frac{1+s}{1-c}} \right)$$

where $\bar{z}_0 \sim \mathcal{N}(0, 1)$, $\bar{z}_1 \sim \mathcal{N}\left(0, \frac{2}{1-c}\right)$, $\bar{z}_2 \sim \mathcal{N}\left(0, \frac{2}{1-c} \left(1 + \frac{s}{c}\right)^2\right)$ and $\bar{z}_3 \sim \mathcal{N}\left(0, \frac{2}{c} \left(1 + 2\frac{s}{c}\right)\right)$. Moreover, $\bar{z}_0, \bar{z}_1, \bar{z}_2$ and \bar{z}_3 are independently distributed.

Finally, the application of Slutsky's lemma (see, e.g., Theorem 2.8 in Van der Vaart (2000)) leads to

$$\sigma_T^{-1} \left(T - \sqrt{n} \frac{S_{GMV}}{\sqrt{1 + \frac{k-1}{n-k+1} \left(1 + \frac{n}{k-1}s\right)}} \right) \xrightarrow{d} \mathcal{N}(0, 1)$$

with

$$\begin{aligned} \sigma_T^2 &= 1 + \frac{S_{GMV}^2}{1+s} (1-c) \left(\frac{1}{2(1-c)} + \frac{(1-c)^2}{4(1+s)^2} \left(\frac{c^2}{(1-c)^2} \left(\frac{2}{1-c} \left(1 + \frac{s}{c}\right)^2 + \frac{2}{c} \left(1 + 2\frac{s}{c}\right) \right) \right) \right) \\ &= 1 + \frac{S_{GMV}^2}{1+s} (1-c) \left(\frac{1}{2(1-c)} + \frac{s^2 + 2s + c}{2(1+s)^2(1-c)} \right) \\ &= 1 + \frac{S_{GMV}^2}{1+s} \left(\frac{1}{2} + \frac{s^2 + 2s + c}{2(1+s)^2} \right) \end{aligned}$$

The statement of Theorem 1(b) follows by setting $S_{GMV} = 0$ under the null hypothesis. \square

In Figures 3 and 4, we deliver the results of the power functions of the exact test and of the high-dimensional asymptotic test that was obtained in Theorem 2 for different values of $c \in \{0.1, 0.4, 0.7, 0.9\}$, with $s = 1$ and $\alpha = 5\%$. The dashed black line represents the power function of the exact test, while the power function of the high-dimensional test is indicated by a solid black line. A good performance of the asymptotic power is observed for all considered values of c .

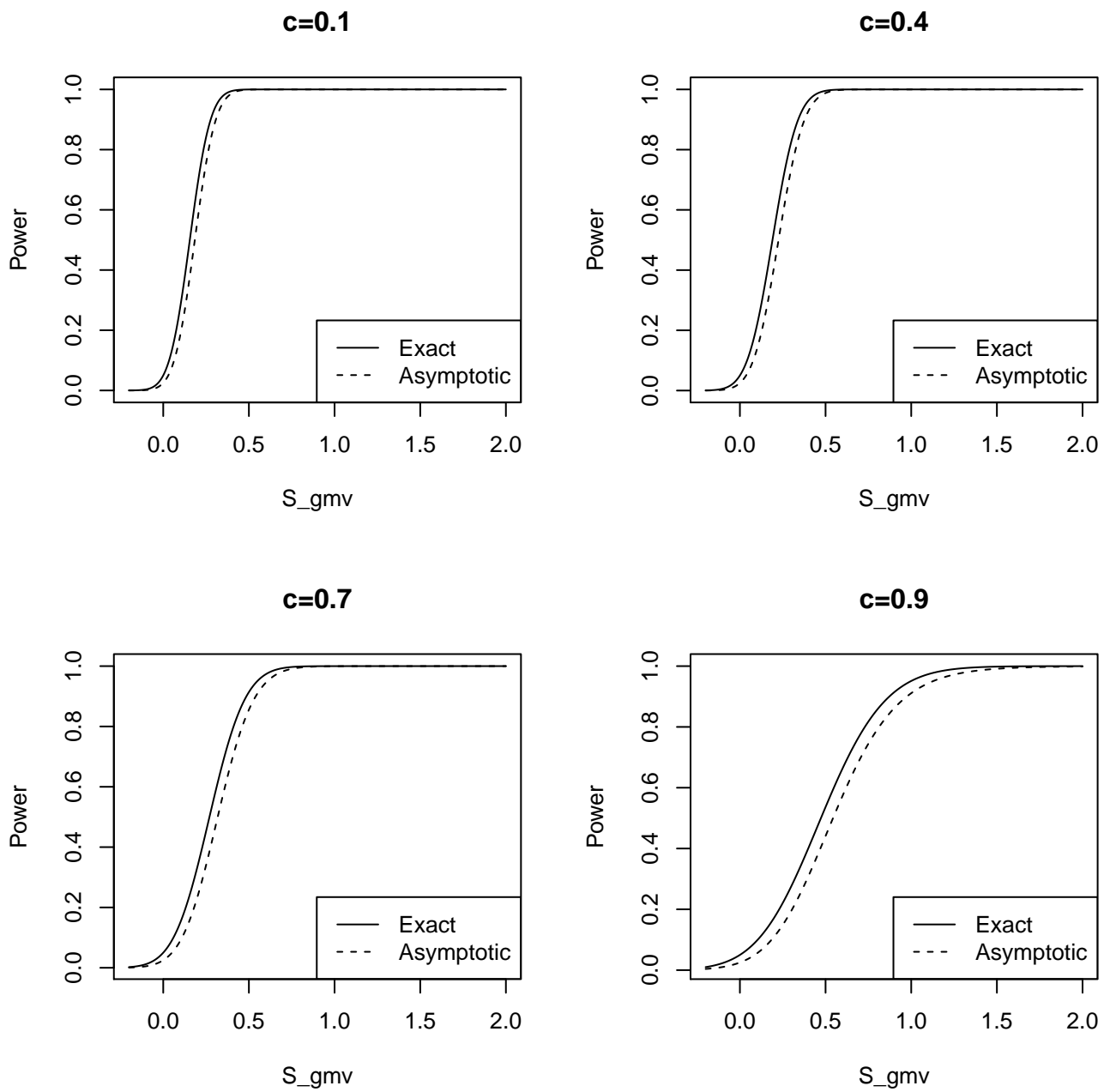


Figure 3: Asymptotic power versus Exact power for different values of c as a function of S_{GMV} , significance level 5% and $n = 250$. We set $s = 1$.

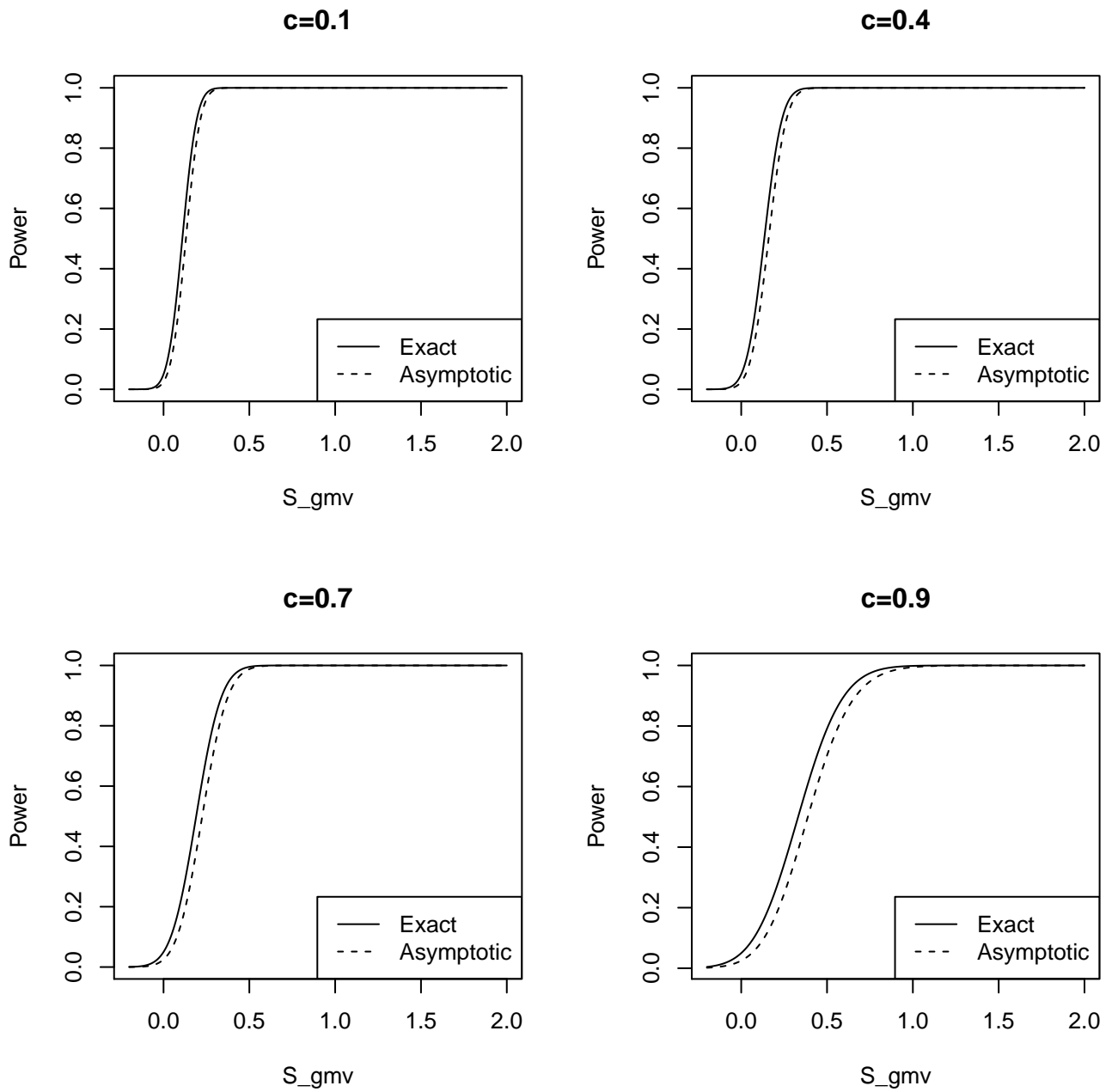


Figure 4: Asymptotic power versus Exact power for different values of c as a function of S_{GMV} , significance level 5% and $n = 500$. We set $s = 1$.

3 Comparison of the tests

In this section we compare the performance of the power function of the test derived in Section 2.1 and the power function of the statistical test developed in Muhinyuza et al. (2017) for $k/n \rightarrow c \in (0, 1)$ as $n \rightarrow \infty$. Results of both Theorem 2 and Muhinyuza et al. (2017, Proposition 1) show that the power functions of the tests depend on the mean vector and the covariance matrix only through the slope parameter s of the efficient frontier and the Sharpe ratio S_{GMV} of the GMVP. For that reason, we put $\Sigma = \mathbf{I}_k$, an identity matrix of dimension k in the simulation study, and consider several values of $\boldsymbol{\mu}$ chosen as follows:

- for $\boldsymbol{\mu}_1$, 15% of its first values are 0.1 and the remainder are set to zero;
- for $\boldsymbol{\mu}_2$, 20% of its first values are 0.1 and the remainder are set to zero;
- for $\boldsymbol{\mu}_3$, 25% of its first values are 0.1 and the remainder are set to zero;
- for $\boldsymbol{\mu}_4$, 30% of its first values are 0.1 and the remainder are set to zero;
- for $\boldsymbol{\mu}_5$, 35% of its first values are 0.1 and the remainder are set to zero.

Table 1 contains several values of s and S_{GMV} obtained using the aforementioned values of $\boldsymbol{\mu}$ and Σ . We note that the values with $S_{GMV} = 0$ corresponds to the null hypothesis in (11) and expect the empirical significance level of the test obtained through out the simulation to be the nominal significance level 0.05. Whereas the other values designate the alternative hypothesis. In addition we set the risk-free rate to be 0.01 and the portfolio size is $k \in \{50, 200, 350, 450\}$. We observe that the slope parameter s becomes larger as k increases, while the Sharpe ratio S_{GMV} increases when the number of non-zero elements in the mean vector becomes larger, the increase is also noted when the portfolio size get larger. On the other hand, we notice that for the values with $\lambda = 0$ are the equivalent to the null hypothesis in (5) while the other values favours the alternative hypothesis. We also note that λ grows as either k or the number of non-zero elements in the mean vector become larger.

In Table 2, we present the results from the simulation study of the powers of the tests for the hypotheses (5) and (11) for different values of $k \in \{50, 200, 350, 450\}$ and $n = 500$. Each value of the power function given in the table was obtained by generating 10^6 independent sample from the corresponding model. All obtained results show a good performance even for data generated from heavy tailed t -distribution. Moreover, we observe an increase of the power functions a the number of non-zero elements of the mean vector becomes larger. To this end, we note that the asymptotic power function of T_λ grows faster compared to the asymptotic power function of T . We also note that for both hypotheses, our tests are moderately conservative when data are generated from a

k	s & S_{GMV}	μ_1	μ_2	μ_3	μ_4	μ_5
50	s	0.0672	0.08	0.0962	0.105	0.1152
	λ	0.069	0.085	0.109	0.125	0.149
	S_{GMV}	0.0424	0.0707	0.1131	0.1414	0.1838
200	s	0.255	0.32	0.375	0.42	0.455
	λ	0.26	0.34	0.42	0.5	0.58
	S_{GMV}	0.0707	0.1414	0.2121	0.2828	0.3536
350	s	0.4497	0.56	0.6587	0.735	0.7977
	λ	0.459	0.595	0.739	0.875	1.019
	S_{GMV}	0.0962	0.1871	0.2833	0.3742	0.4704
450	s	0.5772	0.72	0.8462	0.945	1.0252
	λ	0.589	0.765	0.949	1.125	1.309
	S_{GMV}	0.1084	0.2121	0.3206	0.4243	0.5327

Table 1: Slope parameters s , λ and Sharpe ratio S_{GMV} for the portfolio dimension $k \in \{50, 200, 350, 450\}$ and several values of μ .

	Power	T	T_λ	T	T_λ	T	T_λ	T	T_λ	T	T_λ
k	Distribution	μ_1		μ_2		μ_3		μ_4		μ_5	
50	Normal	0.226	0.823	0.424	0.919	0.768	0.981	0.876	0.994	0.983	1
	t_5	0.148	0.746	0.339	0.863	0.606	0.957	0.786	0.983	0.951	0.994
	t_{10}	0.177	0.804	0.39	0.906	0.702	0.969	0.863	0.988	0.971	0.994
200	Normal	0.302	0.992	0.676	1	0.931	1	0.995	1	1	1
	t_5	0.127	0.973	0.473	0.997	0.777	1	0.952	1	0.992	1
	t_{10}	0.221	0.987	0.57	0.997	0.881	1	0.98	1	0.998	1
350	Normal	0.271	0.981	0.585	1	0.85	1	0.956	1	0.993	1
	t_5	0.109	0.954	0.282	0.991	0.587	0.998	0.805	1	0.945	1
	t_{10}	0.169	0.964	0.437	0.996	0.746	1	0.914	1	0.984	1
450	Normal	0.128	0.735	0.29	0.902	0.493	0.961	0.67	0.984	0.823	0.998
	t_5	0.074	0.65	0.146	0.767	0.285	0.875	0.44	0.937	0.631	0.977
	t_{10}	0.114	0.679	0.231	0.839	0.365	0.912	0.562	0.969	0.701	0.991

Table 2: Power function for the portfolio dimension $k \in \{50, 200, 350, 450\}$ and the sample size $n = 500$. The nominal significance level of the test is $\alpha = 0.05$.

heavy-tailed distribution, since the powers obtained under the t -distribution are in all cases smaller than the one obtained for the normal distribution.

4 Conclusion

In this paper we focus on the property of the TP in high dimension. Especially, we provide the high dimensional asymptotic distribution of the test statistic for testing the existence of the EF and for testing the efficiency of the TP under high-dimensional regime. In either test the asymptotic distribution is obtained under the null and alternative hypotheses. With an extensive simulation study, we have shown that both tests are robust to the violation of the normality assumption and perform well for heavy-tailed t -distribution.

References

- Bodnar, O. and Bodnar, T. (2010). On the unbiased estimator of the efficient frontier. *International Journal of Theoretical and Applied Finance*, 13(07):1065–1073.
- Bodnar, T., Gupta, A. K., and Parolya, N. (2016a). Direct shrinkage estimation of large dimensional precision matrix. *Journal of Multivariate Analysis*, 146:223–236.
- Bodnar, T., Hautsch, N., and Parolya, N. (2016b). Consistent estimation of the high dimensional efficient frontier. Technical report, Technical report.
- Bodnar, T., Mazur, S., Podgórski, K., and Tyrcha, J. (2019a). 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., Okhrin, Y., and Parolya, N. (2019b). Optimal shrinkage-based portfolio selection in high dimensions. *arXiv preprint arXiv:1611.01958*.
- Bodnar, T., Parolya, N., and Schmid, W. (2018). Estimation of the global minimum variance portfolio in high dimensions. *European Journal of Operational Research*, 266(1):371–390.
- Bodnar, T. and Reiß, M. (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 Schmid, W. (2008). A test for the weights of the global minimum variance portfolio in an elliptical model. *Metrika*, 67(2):127–143.
- Bodnar, T. and Schmid, W. (2009). Estimation of optimal portfolio compositions for gaussian returns. *Statistics & Decisions*, 26(3):179–201.

- Britten-Jones, M. (1999). The sampling error in estimates of mean-variance efficient portfolio weights. *The Journal of Finance*, 54(2):655–671.
- Frahm, G. and Jaekel, U. (2008). Tyler’s m-estimator, random matrix theory, and generalized elliptical distributions with applications to finance. *Random Matrix Theory, and Generalized Elliptical Distributions with Applications to Finance (October 21, 2008)*.
- Gibbons, M. R., Ross, S. A., and Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, pages 1121–1152.
- Glombek, K. (2014). Statistical inference for high-dimensional global minimum variance portfolios. *Scandinavian Journal of Statistics*, 41(4):845–865.
- Markowitz, H. (1952). Portfolio selection. *The journal of finance*, 7(1):77–91.
- Mathai, A. M. and Provost, S. B. (1992). *Quadratic Forms in Random Variables*. Marcel Dekker.
- Merton, R. C. (1972). An analytic derivation of the efficient portfolio frontier. *Journal of financial and quantitative analysis*, 7(4):1851–1872.
- Muhinyuza, S., Bodnar, T., and Lindholm, M. (2017). A test on the location of the tangency portfolio on the set of feasible portfolios. *Research Report 26, Stockholm University*.
- Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.
- Rencher, A. C. and Christensen, W. F. (2012). *Methods of multivariate analysis*. John Wiley & Sons.
- Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.