



Delft University of Technology

Statistical Inference for the Expected Utility Portfolio in High Dimensions

Bodnar, Taras; Dmytriv, Solomiia; Okhrin, Yarema ; Parolya, Nestor; Schmid, Wolfgang

DOI

[10.1109/TSP.2020.3037369](https://doi.org/10.1109/TSP.2020.3037369)

Publication date

2021

Document Version

Final published version

Published in

IEEE Transactions on Signal Processing

Citation (APA)

Bodnar, T., Dmytriv, S., Okhrin, Y., Parolya, N., & Schmid, W. (2021). Statistical Inference for the Expected Utility Portfolio in High Dimensions. *IEEE Transactions on Signal Processing*, 69, 1-14. Article 9258421.
<https://doi.org/10.1109/TSP.2020.3037369>

Important note

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

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' - Taverne project

<https://www.openaccess.nl/en/you-share-we-take-care>

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

Statistical Inference for the Expected Utility Portfolio in High Dimensions

Taras Bodnar , Solomiia Dmytriv , Yarema Okhrin , Nestor Parolya , and Wolfgang Schmid

Abstract—In this paper, using the shrinkage-based approach for portfolio weights and modern results from random matrix theory we construct an effective procedure for testing the efficiency of the expected utility (EU) portfolio and discuss the asymptotic behavior of the proposed test statistic under the high-dimensional asymptotic regime, namely when the number of assets p increases at the same rate as the sample size n such that their ratio p/n approaches a positive constant $c \in (0, 1)$ as $n \rightarrow \infty$. We provide an extensive simulation study where the power function and receiver operating characteristic curves of the test are analyzed. In the empirical study, the methodology is applied to the returns of S&P 500 constituents.

Index Terms—Finance, portfolio analysis, mean-variance optimal portfolio, statistical test, shrinkage estimator, random matrix theory.

I. INTRODUCTION

FOLLOWING the mean-variance approach of [1], which is considered to be one of the most popular portfolio choice strategies, the weights of an optimal portfolio are obtained by the mean-variance tradeoff aiming to minimize the portfolio variance while maximizing the portfolio return. This set of optimal portfolios determines the efficient frontier in the mean-variance space. The Markowitz approach formalizes the advantages of portfolio diversification and has become a benchmark for both researchers and practitioners in portfolio management.

Markowitz optimal portfolios, also known as mean-variance optimal portfolios, can also be obtained as solutions of other optimization problems (e.g., [2]), like by maximizing the expected quadratic utility function (see, [3]) expressed as

$$\mathbf{w}'\boldsymbol{\mu} - \frac{\gamma}{2}\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} \rightarrow \max \quad \text{subject to } \mathbf{w}'\mathbf{1}_p = 1, \quad (1)$$

Manuscript received May 22, 2020; revised October 14, 2020; accepted October 28, 2020. Date of publication November 16, 2020; date of current version December 23, 2020. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Monica F. Bugallo. This work was supported in part by the Swedish Research Council (VR) via the project “Bayesian Analysis of Optimal Portfolios and Their Risk Measures,” and in part by the DFG project OK103/5-2 “Vine copula base modelling and forecasting of multivariate realized volatility time-series.” (Corresponding author: Yarema Okhrin.)

Taras Bodnar is with the Department of Mathematics, Stockholm University, Stockholm SE-10691, Sweden (e-mail: taras.bodnar@math.su.se).

Solomiia Dmytriv is with the Department of Statistics and Operations Research, University of Vienna, 1090 Vienna, Austria (e-mail: solomiia.dmytriv@univie.ac.at).

Yarema Okhrin is with the Department of Statistics, University of Augsburg, D-86159 Augsburg, Germany (e-mail: yarema.okhrin@wiwi.uni-augsburg.de).

Nestor Parolya is with the Delft Institute of Applied Mathematics, Delft University of Technology, 2628XE, The Netherlands (e-mail: n.parolya@tudelft.nl.).

Wolfgang Schmid is with the Department of Statistics, European University Viadrina, Frankfurt (Oder) D-15232, Germany (e-mail: schmid@europa-unide).

Digital Object Identifier 10.1109/TSP.2020.3037369

where $\mathbf{w} = (w_1, \dots, w_p)'$ is the vector of portfolio weights, $\mathbf{1}_p$ is the p -dimensional vector of ones, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean vector and the covariance matrix of the random vector of asset returns $\mathbf{x} = (x_1, \dots, x_p)'$. The quantity $\gamma > 0$ measures the investors attitude towards risk. If $\gamma = \infty$, then the investor is fully risk averse and determines the investment strategy by minimizing the portfolio variance without paying attention to the expected portfolio return, i.e., she constructs the so-called global minimum variance (GMV) portfolio. Under the assumption that the asset returns are normally distributed, the problem of maximization the mean-variance objective function (1) is equivalent to the maximization of the expected exponential utility, which implies constant absolute risk aversion (CARA). In this case, γ is equal to the investors absolute risk aversion coefficient (see, e.g., [3]).

We denote the solution of (1) by \mathbf{w}_{EU} and it is given by

$$\mathbf{w}_{EU} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}_p}{\mathbf{1}'_p\boldsymbol{\Sigma}^{-1}\mathbf{1}_p} + \gamma^{-1}\mathbf{Q}\boldsymbol{\mu}, \quad (2)$$

where

$$\mathbf{Q} = \boldsymbol{\Sigma}^{-1} - \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}_p\mathbf{1}'_p\boldsymbol{\Sigma}^{-1}}{\mathbf{1}'_p\boldsymbol{\Sigma}^{-1}\mathbf{1}_p}. \quad (3)$$

The case of fully risk averse investor, i.e., $\gamma = \infty$, leads to the weights of the GMV portfolio expressed as

$$\mathbf{w}_{GMV} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}_p}{\mathbf{1}'_p\boldsymbol{\Sigma}^{-1}\mathbf{1}_p}. \quad (4)$$

The derived formulas of optimal portfolio weights (2) and (4) cannot directly be used in practice, since they both depend on unknown parameters of the data generating process. The mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ are not observable in practice and have to be estimated by using historical data for asset returns. This, however, introduces further sources of risk into the investment process, namely the estimation risk which has been ignored for a long time in finance.

The most commonly used approach to estimate the weights of optimal portfolios is based on simple replacing the unknown first two moments of the asset returns by their sample counterparts. As a result, we obtain a “plug-in” estimator for the optimal portfolio weights also known as its sample estimator, which is a traditional way to construct a portfolio in practice. Assuming that the asset returns are independent and normally distributed [4] obtain the asymptotic distribution of the sample estimator of the EU portfolio weights, while the corresponding exact distributional results can be found in [5]. Further theoretical and practically relevant findings related to the characterization of

the distribution of the sample estimator of the optimal portfolio weights and their characteristics can be found in [6], [7], [8], [9], among others.

The use of the “plug-in” estimators in practice has been widely criticized in statistical and financial literature. One of the main drawbacks of the sample estimators is the investor overoptimism about the optimality of the constructed portfolio. Several studies (see, e.g., [10], [11], [12]) show with theoretical and empirical arguments that the plug-in estimator of the efficient frontier overestimates the location of the true efficient frontier in the mean-variance space. This leads to too optimistic trading strategies which perform in practice typically much worse than expected.

In recent years, other types of estimators for the optimal portfolio weights have been introduced in the literature. Some estimators attempt to improve the estimators for the parameters of the asset returns. Relying on the idea of [13] we can use a shrinkage estimator for the mean vector and for the covariance matrix or its inverse (see, e.g., [14] and [15]). Alternatively, one can apply the shrinkage method directly to portfolio weights as suggested by [16], [17], [18], etc. The goal of the approach is to reduce the estimation uncertainty and to decrease the variance in the estimated portfolio weights.

The problem of assessing the estimation risk, when an optimal portfolio is constructed, becomes very challenging from the high-dimensional perspectives, i.e., when both the number of included assets p and the sample size n tend to infinity simultaneously such that p/n tends to the concentration ratio $c > 0$ as $n \rightarrow \infty$ (see, [19], [20]). Under the classical asymptotic regime, when the number of assets p is fixed and substantially smaller than the sample size n , the traditional “plug-in” estimator of optimal portfolio weights is consistent (see, [4], [5]). On the other hand, the sample estimators of the mean vector and of the covariance matrix are not longer feasible under the high-dimensional asymptotics ([21], [20], [22]), which has a negative impact on the performance of the asset allocation strategy. Moreover, the inverse covariance matrix does not exist anymore for $c > 1$ and the optimal portfolios cannot be constructed in a traditional way.

The high-dimensional treatment of the portfolio problem is not only of theoretical interest, but also of great importance in practice. Nowadays, the technological advances and the availability of financial information make the whole universe of assets easily accessible for private and institutional investors (see, [23]). As a result, the investors have a possibility to allocate their wealth into thousands of stocks around the world. However, the sample size of asset returns at selected frequency cannot be enlarged without increasing the time span of the collected data. Typically, one restricts the estimation period to a few years to mitigate the impact of time-varying characteristics of the asset returns. This leads to portfolios consisting of hundreds of assets estimated over a relatively short period of time. Therefore, new results on constructing optimal portfolios in this high-dimensional setting are in great demand. Similarly as in the low-dimensional case, the first line of the research deals with deriving improved estimators for the mean vector and the covariance matrix of asset returns. These are used to obtain

improved plug-in estimators of the optimal portfolio weights (see, [24], [25]). The second possibility is to improve the estimators of the optimal portfolio weights directly. This can be achieved by taking their functional dependence on the mean vector and of covariance matrix into account. Following this approach [26] suggest the optimal shrinkage estimator for the GMV portfolio weights, while [27] propose the optimal shrinkage estimator for the EU portfolio weights. Both estimators are derived by using recent results in random matrix theory and appear to be feasible even in the case of $c > 1$. Other optimal portfolio choice strategies under the high-dimensional regime were established by [28], [29], [9].

It is important to note that the statistical methods developed for estimating optimal portfolio weights can be linked to the classical methods used in statistical signal processing. For example, the Capon or minimum variance spatial filter is equivalent to the GMV portfolio in signal processing literature (see, [30] and [31]). The estimation risk of the high-dimensional minimum variance beamformer is studied in [28] and [32], while its constrained versions are discussed in [33]. Moreover, [34] discuss the finite-sample size effect on minimum variance filter and [35] present an improved calibration of the precision matrix. Further literature on the applications of random matrix theory to signal processing and portfolio optimization can be found in [36] and references therein.

We contribute to the recent literature in portfolio theory and signal processing theory by developing new statistical tests on the weights of the EU portfolio in a high-dimensional setting. From practical point of view an investor will have an opportunity to test if the current large portfolio coincides with a prespecified benchmark portfolio or there are significant deviations. From the theoretical perspective we contribute by deriving confidence intervals and test theory for expressions including functions of both the mean vector and the covariance matrix. This directly extends the existent results on testing the structure of the covariance matrix in high-dimensional settings (see, e.g., [37], [38], [39]). The new approach is based on the shrinkage estimator of the EU portfolio weights and extends the one derived for the weights of the GMV portfolio in [40] by taking the uncertainty about the estimated mean vector into account when the high-dimensional asymptotic distribution of the test statistic is derived. One of the main advantages of the approach is that the whole high-dimensional vector of portfolio weights can be tested in a single step. Moreover, the investor can make a decision about the efficiency of the holding portfolio based on the result of the testing procedure.

The rest of paper is organized as follows. In Section II, we describe the existent approaches in testing the finite number of the EU portfolio weights in both low and high dimensions. New test based on the shrinkage approach is suggested in Section III. Here, the asymptotic distribution of the test statistic is derived under both the null and the alternative hypotheses under high-dimensional settings. In Section IV-A, we compare the new test with the existent approaches in terms of size and power properties, while an empirical illustration is provided in Section IV-B. Concluding remarks are presented in Section V.

II. SAMPLE ESTIMATOR OF THE EU PORTFOLIO AND TEST THEORY

We consider a financial market consisting of p risky assets. Let \mathbf{x}_t denote the p -dimensional vector of the returns on risky assets at time t . Suppose that $E(\mathbf{x}_t) = \boldsymbol{\mu}$ and $Cov(\mathbf{x}_t) = \boldsymbol{\Sigma}$ where $\boldsymbol{\Sigma}$ is assumed to be positive definite. Let $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ be a sample of asset return vectors consisting of their n independent realizations and let $\mathbf{X}_n = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ stand for the $p \times n$ data matrix. Throughout the paper we assume that the asset returns are independent and identically normally distributed, i.e., $\mathbf{x}_i \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $i = 1, \dots, n$.

The sample estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are given by

$$\bar{\mathbf{x}}_n = \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j \text{ and } \hat{\boldsymbol{\Sigma}}_n = \frac{1}{n-1} \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}}_n)(\mathbf{x}_j - \bar{\mathbf{x}}_n)'. \quad (5)$$

Replacing $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in (2) by their sample estimators from (5), we obtain the sample estimator of the EU portfolio weights expressed as

$$\hat{\mathbf{w}}_{EU} = \frac{\hat{\boldsymbol{\Sigma}}_n^{-1} \mathbf{1}_p}{\mathbf{1}'_p \hat{\boldsymbol{\Sigma}}_n^{-1} \mathbf{1}_p} + \gamma^{-1} \hat{\mathbf{Q}}_n \hat{\boldsymbol{\mu}}_n,$$

where

$$\hat{\mathbf{Q}}_n = \hat{\boldsymbol{\Sigma}}_n^{-1} - \frac{\hat{\boldsymbol{\Sigma}}_n^{-1} \mathbf{1}_p \mathbf{1}'_p \hat{\boldsymbol{\Sigma}}_n^{-1}}{\mathbf{1}'_p \hat{\boldsymbol{\Sigma}}_n^{-1} \mathbf{1}_p}. \quad (6)$$

In [4], Y. Okhrin and W. Schmid derive the analytical expression for the expectation and the covariance matrix of $\hat{\mathbf{w}}_{EU}$, and obtain its asymptotic distribution assuming that the portfolio size is considerably smaller than the sample size. These results are extended in [5] who derive the finite-sample distribution of the estimated EU portfolio weights and use these results in the derivation of an asymptotic test on the weights which we present in the next subsection.

A. Tests Based on Mahalanobis Distance

At each time point an investor has to decide whether the holding portfolio is efficient or it has to be adjusted (see, [41], [5]). This problem can be presented as a special case of the general linear hypotheses formulated for the portfolio weights. Let \mathbf{L} denote the $k \times p$ dimensional matrix of constants with $k < p - 1$ and let \mathbf{r} be the k -dimensional vector of constants. [5] consider the following hypotheses for linear combinations of the EU portfolio weights

$$H_0 : \mathbf{L}\mathbf{w}_{EU} = \mathbf{r} \quad \text{against} \quad H_1 : \mathbf{L}\mathbf{w}_{EU} \neq \mathbf{r}, \quad (7)$$

If one sets $\mathbf{L} = [\mathbf{I}_k \ \mathbf{O}_{k,p-k}]$ in (7) where \mathbf{I}_k is the k -dimensional identity matrix and $\mathbf{O}_{k,p-k}$ is the $k \times (p - k)$ matrix with zeros, then the null hypothesis states that the first k weights in \mathbf{w}_{EU} are equal to the corresponding components defined by \mathbf{r} . It also has to be noted that the whole structure of the EU portfolio cannot be tested by using (7) because of the restriction imposed on the number of linear combinations which should be smaller than $p - 1$. Thus, the test on the whole vector of the EU portfolio weights

should be performed by testing at least two null hypotheses of the form (7) by selecting matrices \mathbf{L} in each of the null hypotheses such that all elements in \mathbf{w}_{EU} are tested. This leads to a multiple testing problem also discussed below.

In order to test (7) for a given matrix \mathbf{L} and a vector \mathbf{r} , [5] suggest the following test statistic:

$$T_{\mathbf{L}} = (n - p + 1) (\hat{\mathbf{w}}_{\mathbf{L}} - \mathbf{r})' \left(\frac{\mathbf{L}\hat{\mathbf{Q}}_n\mathbf{L}'}{\mathbf{1}'_p \hat{\boldsymbol{\Sigma}}_n^{-1} \mathbf{1}_p} + \gamma^{-1} \frac{\mathbf{L}\hat{\mathbf{Q}}_n\mathbf{L}'}{\bar{\mathbf{x}}'_n \hat{\mathbf{Q}}_n \bar{\mathbf{x}}_n} \right)^{-1} (\hat{\mathbf{w}}_{\mathbf{L}} - \mathbf{r}), \quad (8)$$

where

$$\hat{\mathbf{w}}_{\mathbf{L}} = \mathbf{L}\hat{\mathbf{w}}_{EU} = \frac{\mathbf{L}\hat{\boldsymbol{\Sigma}}_n^{-1} \mathbf{1}_p}{\mathbf{1}'_p \hat{\boldsymbol{\Sigma}}_n^{-1} \mathbf{1}_p} + \gamma^{-1} \mathbf{L}\hat{\mathbf{Q}}_n \bar{\mathbf{x}}_n. \quad (9)$$

In [5], T. Bodnar and W. Schmid show that the test statistic $T_{\mathbf{L}}$ is asymptotically non-central χ^2 -distributed with k degrees of freedom and the non-centrality parameter

$$\begin{aligned} \lambda = & n (\mathbf{w}_{\mathbf{L}} - \mathbf{r})' \\ & \times \left(\frac{\mathbf{L}\mathbf{Q}\mathbf{L}'}{\mathbf{1}'_p \hat{\boldsymbol{\Sigma}}_n^{-1} \mathbf{1}_p} + \frac{1}{\gamma} \frac{\mathbf{L}\mathbf{Q}\mathbf{L}'}{\boldsymbol{\mu}' \mathbf{Q} \boldsymbol{\mu}} + \frac{1}{\gamma^2} (\mathbf{L}\mathbf{Q}\mathbf{L}') \boldsymbol{\mu}' \mathbf{Q} \boldsymbol{\mu} \right. \\ & \left. - \mathbf{L}\mathbf{Q}\boldsymbol{\mu}\boldsymbol{\mu}'\mathbf{Q}\mathbf{L}' \right)^{-1} (\mathbf{w}_{\mathbf{L}} - \mathbf{r}) \end{aligned} \quad (10)$$

with

$$\mathbf{w}_{\mathbf{L}} = \mathbf{L}\mathbf{w}_{EU} = \frac{\mathbf{L}\boldsymbol{\Sigma}^{-1} \mathbf{1}_p}{\mathbf{1}'_p \boldsymbol{\Sigma}^{-1} \mathbf{1}_p} + \gamma^{-1} \mathbf{L}\mathbf{Q}\boldsymbol{\mu}, \quad (11)$$

when both p and k are relatively small with respect to the sample size n . As a special case, we obtain the asymptotic distribution of $T_{\mathbf{L}}$ under the null hypothesis in (7) given by $T_{\mathbf{L}} \xrightarrow{d} \chi_k^2$ where the symbol \xrightarrow{d} denotes the convergence in distribution.

Since the asymptotic distribution of the test statistic $T_{\mathbf{L}}$ is obtained under classical asymptotic regime, this test, in general, is not applicable when the portfolio size is comparable to the sample size. We illustrate this point in Figure 1. Here we plot the kernel density estimator (KDE) of the distribution of the test statistic $T_{\mathbf{L}}$ under the null hypothesis together with the asymptotic χ^2 -distribution (green and red curves, respectively). For this purpose we generate samples from a multivariate normal distribution with mean vector and covariance matrix as specified in the numerical study of Section IV-A. The vector \mathbf{r} consists of the first k components of the true EU portfolio weights and we set $\mathbf{L} = [\mathbf{I}_k \ \mathbf{O}_{k,p-k}]$. For each sample we compute the value of the test statistic $T_{\mathbf{L}}$ and then plot the KDE. To robustify the conclusions we set $\gamma = 5$, $p = 300$, $c_n = p/n \in \{0.3, 0.8\}$ and $k \in \{10, 30, 100\}$. We observe that already for $k = 10$ the difference between the KDE and the asymptotic distribution is very large and this evidence becomes stronger if k increases. For $k = 100$ the KDE shifts strongly to the right and is not shown to retain the same scaling on the x -axis. Table I gives the realized sizes (estimated probabilities of type I error) of

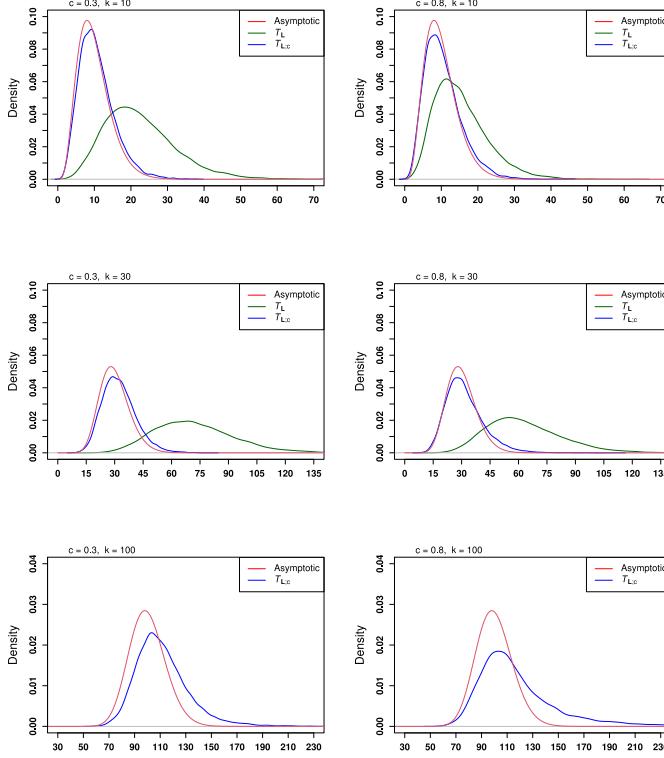


Fig. 1. The high-dimensional asymptotic χ^2 approximation of the densities of T_L and $T_{L;c}$ together with their kernel density estimators for $\gamma = 5$, $p = 300$, $c_n = p/n \in \{0.3, 0.8\}$ and $k \in \{10, 30, 100\}$.

TABLE I
EMPIRICAL SIZES OF THE TESTS BASED ON T_L AND $T_{L;c}$ USING 10^4
INDEPENDENT REPLICATIONS

$c = 0.3$			
	$k = 10$	$k = 30$	$k = 100$
T_L	0.526	0.890	1
$T_{L;c}$	0.059	0.071	0.172
$c = 0.8$			
	$k = 10$	$k = 30$	$k = 100$
T_L	0.216	0.762	1
$T_{L;c}$	0.069	0.105	0.220

the considered test based on the 10^4 independent replications and with the nominal level $\alpha = 0.05$. For different values of $k \in \{10, 30, 100\}$, it can be seen that T_L is highly inconsistent and has a much higher size than the nominal value α . We conclude that the test is highly unreliable if we wish to test many or all weights simultaneously.

B. Improvement of the Test Based on Mahalanobis Distance for Large-Dimensional Portfolios

In [42], T. Bodnar *et al.* show that the sample estimator of the EU portfolio weights is not consistent under the high-dimensional asymptotic regime, i.e., when $p/n \rightarrow c \in [0, 1)$ as $n \rightarrow \infty$. Moreover, they derive a consistent estimator for

the elements of \mathbf{w}_L and use these findings to construct a high-dimensional asymptotic test on the finite number of linear combinations of the EU portfolio weights.

Let \mathbf{L} be a $k \times p$ matrix of constants as defined in Section II-A and let

$$\hat{\mathbf{w}}_{GMV;L} = \mathbf{L}\hat{\mathbf{w}}_{GMV} = \frac{\mathbf{L}\hat{\Sigma}_n^{-1}\mathbf{1}_p}{\mathbf{1}'_p\hat{\Sigma}_n^{-1}\mathbf{1}_p}, \quad \hat{s} = \bar{\mathbf{x}}'_n \hat{\mathbf{Q}}_n \bar{\mathbf{x}}_n$$

$$\text{and } \hat{\boldsymbol{\eta}}_L = \frac{\mathbf{L}\hat{\mathbf{Q}}_n \bar{\mathbf{x}}_n}{\bar{\mathbf{x}}'_n \hat{\mathbf{Q}}_n \bar{\mathbf{x}}_n}. \quad (12)$$

Assuming that k is finite, i.e., considerably smaller than both p and n , [42] prove that

$$\hat{\mathbf{w}}_{GMV;L} \xrightarrow{a.s.} \mathbf{L}\mathbf{w}_{GMV}, \quad \hat{s}_c = (1 - c_n)\hat{s} - c_n \xrightarrow{a.s.} s \quad (13)$$

$$\text{and } \hat{\boldsymbol{\eta}}_{L;c} = \frac{\hat{s}_c + c_n}{\hat{s}_c} \hat{\boldsymbol{\eta}}_L \xrightarrow{a.s.} \boldsymbol{\eta}_L$$

for $c_n = p/n \rightarrow c \in [0, 1)$ as $n \rightarrow \infty$ with

$$s = \boldsymbol{\mu}' \mathbf{Q} \boldsymbol{\mu} \quad \text{and} \quad \boldsymbol{\eta}_L = \frac{\mathbf{L} \mathbf{Q} \boldsymbol{\mu}}{\boldsymbol{\mu}' \mathbf{Q} \boldsymbol{\mu}}. \quad (14)$$

The symbol $\xrightarrow{a.s.}$ denotes the almost surely convergence.

Using (13), [42] propose a high-dimensional asymptotic test on the hypotheses (7) with the test statistic given by

$$T_{L;c} = (n - p)(\hat{\mathbf{w}}_{L;c} - \mathbf{r})' \hat{\Omega}_{L;c}^{-1} (\hat{\mathbf{w}}_{L;c} - \mathbf{r}), \quad (15)$$

where

$$\hat{\mathbf{w}}_{L;c} = \hat{\mathbf{w}}_{GMV;L} + \gamma^{-1} \hat{s}_c \hat{\boldsymbol{\eta}}_{L;c} \quad (16)$$

and

$$\begin{aligned} \hat{\Omega}_{L;c} = & \left(\left(\frac{1 - c_n}{\hat{s}_c + c_n} + (\hat{s}_c + c_n)\gamma^{-1} \right) \gamma^{-1} + \hat{V}_c \right) (1 - c_n) \mathbf{L} \hat{\mathbf{Q}}_n \mathbf{L}^\top \\ & + \gamma^{-2} \left\{ \frac{2(1 - c_n)c_n^3}{(\hat{s}_c + c_n)^2} + 4(1 - c_n)c_n \frac{\hat{s}_c(\hat{s}_c + 2c_n)}{(\hat{s}_c + c_n)^2} \right. \\ & \left. + \frac{2(1 - c_n)c_n^2(\hat{s}_c + c_n)^2}{\hat{s}_c^2} - \hat{s}_c^2 \right\} \hat{\boldsymbol{\eta}}_{L;c} \hat{\boldsymbol{\eta}}'_{L;c}, \end{aligned} \quad (17)$$

where

$$\hat{V}_c = \frac{\hat{V}_{GMV}}{1 - c_n} \quad \text{with} \quad \hat{V}_{GMV} = \frac{1}{\mathbf{1}'_p \hat{\Sigma}_n^{-1} \mathbf{1}_p} \quad (18)$$

are the consistent and the sample estimators of the variance of the GMV portfolio (4), that is (see, e.g., [26, p.387])

$$\hat{V}_c \xrightarrow{a.s.} V_{GMV} = \frac{1}{\mathbf{1}'_p \Sigma^{-1} \mathbf{1}_p}$$

for $c_n = p/n \rightarrow c \in [0, 1)$ as $n \rightarrow \infty$.

The application of the results of Theorem 4.4 in [42] leads to the high-dimensional asymptotic distribution of $T_{L;c}$ under both hypotheses in (7). Namely, it holds that the asymptotic distribution of $T_{L;c}$ under H_1 is a non-central χ^2 -distribution

with k degrees of freedom and non-centrality parameter given by

$$\lambda_c = (n - p)(\mathbf{w}_L - \mathbf{r})' \boldsymbol{\Omega}_{L;c}^{-1} (\mathbf{w}_L - \mathbf{r}), \quad (19)$$

where

$$\begin{aligned} \boldsymbol{\Omega}_{L;c} = & \left(\left(\frac{1-c}{s+c} + (s+c)\gamma^{-1} \right) \gamma^{-1} + V_{GMV} \right) (1-c) \mathbf{L} \mathbf{Q} \mathbf{L}^\top \\ & + \gamma^{-2} \left\{ \frac{2(1-c)c^3}{(s+c)^2} + 4(1-c)c \frac{s(s+2c)}{(s+c)^2} \right. \\ & \left. + \frac{2(1-c)c^2(s+c)^2}{s^2} - s^2 \right\} \boldsymbol{\eta}_L \boldsymbol{\eta}_L' . \end{aligned} \quad (20)$$

Moreover, $T_{L;c} \xrightarrow{d} \chi_k^2$ under H_0 .

In Figure 1, we present the KDE of the distribution of $T_{L;c}$ (blue curve) and compare it to its high-dimensional asymptotic distribution (red curve). The kernel density estimator as well as the realized sizes of the test are obtained under the same simulation setup as one used at the end of Section II-A. The approximation works well and much better than in the case of T_L for smaller values of k , but discrepancy becomes large if k increases. The same conclusion can be drawn from Table I. Here the method proposed by [42] has a much better realized size which still increases dramatically with growing k .

III. TEST BASED ON THE SHRINKAGE APPROACH

Both tests based on the Mahalanobis distance are designed to test a finite number of linear restrictions imposed on the EU portfolio weights. Although the high-dimensional test shows a considerable improvement in terms of the size (see, Figures 1 and Table I), this test, similarly to the test based on the statistic T_L , cannot be applied to test the structure of the whole EU portfolio. In practice, one has to fix the number k of the EU portfolio weights (or their linear restrictions) and apply the test $T_{L;c}$ several times in order to cover the whole vector \mathbf{w}_{EU} . This approach is a single-step multiple test (see, [43]) with the number of marginal hypotheses to be tested equal to $[p/k] + 1$. Since the dependence structure between the marginal tests is very complicated, one has to monitor the overall type I error rate by using the so-called Bonferroni correction (see, [43]). This would worse the power properties of each individual test, especially when the number of tests is relatively large.

As a solution to this challenging problem, we suggest a new approach for testing the structure of the EU portfolio by a single test. The new procedure is based on the shrinkage estimator of the EU portfolio weights as suggested by [27] and extend our previous results obtained for the GMV portfolio in [40], which is a very special case of the EU portfolio. In contrast to the EU portfolio, the weights of the GMV portfolio do not depend on the mean vector. As a result, the derivation of the test for the EU portfolio becomes a very challenging task and completely new results in random matrix theory have to be derived to handle it.

A. Optimal Shrinkage Estimator of the EU Portfolio Weights

The shrinkage estimator for the EU portfolio weights is a convex combination of the sample estimator and a fixed well behaved target portfolio $\mathbf{b} \in \mathbb{R}^p$ with bounded expected return and variance, i.e., $R_b = \mathbf{b}'\boldsymbol{\mu} < \infty$ and $V_b = \mathbf{b}'\boldsymbol{\Sigma}^{-1}\mathbf{b} < \infty$ uniformly in p . Thus, the shrinkage estimator is expressed as

$$\hat{\mathbf{w}}_{GSE} = \alpha_n \hat{\mathbf{w}}_{EU} + (1 - \alpha_n) \mathbf{b} \quad \text{with } \mathbf{b}'\mathbf{1}_p = 1, \quad (21)$$

where α_n is the shrinkage intensity. One of the main ideas behind the shrinkage estimator (21) is to reduce the large variability present in the sample estimator $\hat{\mathbf{w}}_{EU}$ by shrinking it to a vector of constants. This approach might introduce a bias in the estimator, but on the other side it reduces the variability of the sample estimator considerably.

In [27], T. Bodnar, Y. Okhrin, and N. Parolya determine the optimal shrinkage intensity α_n^* as the solution of the maximization problem based on the mean-variance objective function. It is given by

$$\alpha_n^* = \frac{(\hat{\mathbf{w}}_{EU} - \mathbf{b})'(\boldsymbol{\mu} - \gamma\boldsymbol{\Sigma}\mathbf{b})}{(\hat{\mathbf{w}}_{EU} - \mathbf{b})'\boldsymbol{\Sigma}(\hat{\mathbf{w}}_{EU} - \mathbf{b})}. \quad (22)$$

Since the expression of α_n^* depends on both the population mean vector and covariance matrix and on their sample counterparts, it cannot directly be applied in practice. As such, [27] propose a two-stage procedure. First, the deterministic quantity α^* which is asymptotically equivalent to α_n^* is found. Second, it is consistently estimated under the high-dimensional asymptotic regime.

It holds that (see, [27, Theorem 2.1])

$$\alpha^* =$$

$$\frac{\gamma^{-1} \left((R_{GMV} - R_b) \left(1 + \frac{1}{1-c} \right) + \gamma(V_b - V_{GMV}) + \frac{\gamma^{-1}}{1-c} s \right)}{\frac{V_{GMV}}{1-c} - 2 \left(V_{GMV} + \frac{\gamma^{-1}}{1-c} (R_b - R_{GMV}) \right) + \gamma^{-2} \left(\frac{s+c}{(1-c)^3} \right) + V_b}, \quad (23)$$

where $R_{GMV} = \frac{\mathbf{1}_p' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{\mathbf{1}_p' \boldsymbol{\Sigma}^{-1} \mathbf{1}_p}$ is the expected return of the GMV portfolio. Following [27] we assume throughout the paper that uniformly in p the quadratic form $\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}_p$ is bounded away from zero and $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ is bounded from above by some positive constant. These conditions guarantee among others the boundedness of R_{GMV} , V_{GMV} and s as $p \rightarrow \infty$, thus, keeping the limiting expressions coming further well defined asymptotically. Consistent estimators for the variance of the GMV portfolio V_{GMV} and for the slope parameter of the efficient frontier s are given in (18) and (13), respectively. [27] show that the sample estimators of R_{GMV} , R_b , and V_b are consistent, that is

$$\begin{aligned} \hat{R}_{GMV} &= \frac{\mathbf{1}_p' \hat{\boldsymbol{\Sigma}}_n^{-1} \bar{\mathbf{x}}_n}{\mathbf{1}_p' \hat{\boldsymbol{\Sigma}}_n^{-1} \mathbf{1}_p} \xrightarrow{a.s} R_{GMV}, \\ \hat{R}_b &= \mathbf{b}' \bar{\mathbf{x}}_n \xrightarrow{a.s} R_b, \\ \hat{V}_b &= \mathbf{b}' \hat{\boldsymbol{\Sigma}}_n \mathbf{b} \xrightarrow{a.s} V_b, \end{aligned} \quad (24)$$

for $p/n \rightarrow c \in [0, 1]$ as $n \rightarrow \infty$.

Hence, a consistent estimator for α^* is constructed as

$$\hat{\alpha}_c^* = \frac{\gamma^{-1} \left((\hat{R}_{GMV} - \hat{R}_b) \left(1 + \frac{1}{1-c_n} \right) + \gamma(\hat{V}_b - \hat{V}_c) + \frac{\gamma^{-1}}{1-c_n} \hat{s}_c \right)}{\frac{\hat{V}_c}{1-c_n} - 2 \left(\hat{V}_c + \frac{\gamma^{-1}}{1-c_n} (\hat{R}_b - \hat{R}_{GMV}) \right) + \gamma^{-2} \left(\frac{\hat{s}_c + c_n}{(1-c_n)^3} \right) + \hat{V}_b}, \quad (25)$$

while the bona fide shrinkage estimator for the weights of the EU portfolio is expressed as

$$\hat{\mathbf{w}}_{BFGSE} = \hat{\alpha}_c^* \hat{\mathbf{w}}_{EU} + (1 - \hat{\alpha}_c^*) \mathbf{b}. \quad (26)$$

Next, we prove that $\hat{\alpha}_c^*$ is asymptotically normally distributed. This result will then be used to derive a test for the structure of the EU portfolio in Section III-B. Let $\alpha^* = \frac{A}{B}$ and $\hat{\alpha}_c^* = \frac{\hat{A}_n}{\hat{B}_n}$. Then, we get

$$\begin{aligned} \sqrt{n}(\hat{\alpha}_c^* - \alpha^*) &= \sqrt{n} \left(\frac{\hat{A}_n - A}{\hat{B}_n} - \frac{A(\hat{B}_n - B)}{B\hat{B}_n} \right) \\ &= \frac{1}{\hat{B}_n} \left(\sqrt{n}(\hat{A}_n - A) - \frac{A}{B} \sqrt{n}(\hat{B}_n - B) \right) \\ &= \frac{\mathbf{d}'}{\hat{B}_n} \sqrt{n}\mathbf{t} + o_P(1) \end{aligned} \quad (27)$$

for $p/n = c + o(n^{-1/2})$ as $n \rightarrow \infty$ with

$$\mathbf{t} = \begin{pmatrix} \hat{R}_{GMV} - R_{GMV} \\ \hat{V}_c - V_{GMV} \\ \hat{s}_c - s \\ \hat{R}_b - R_b \\ \hat{V}_b - V_b \end{pmatrix}, \mathbf{d} = \begin{pmatrix} \gamma^{-1} + \frac{\gamma^{-1}}{1-c_n} (1 - 2\frac{A}{B}) \\ -1 - \frac{A}{B} \left(\frac{1}{1-c_n} - 2 \right) \\ \frac{\gamma^{-2}}{1-c_n} \left(1 - \frac{1}{(1-c_n)^2} \frac{A}{B} \right) \\ -\gamma^{-1} - \frac{\gamma^{-1}}{1-c_n} (1 - 2\frac{A}{B}) \\ 1 - \frac{A}{B} \end{pmatrix}, \quad (28)$$

where the symbol $o_P(1)$ denotes a sequence which tends almost surely to zero. In Theorem 1, we derive the asymptotic distribution of \mathbf{t} .

Theorem 1: Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be independent and identically distributed with $\mathbf{x}_i \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for $i = 1, \dots, n$ with $\boldsymbol{\Sigma}$ positive definite. Then it holds that

$$\sqrt{n}\mathbf{t} \xrightarrow{d} \mathcal{N}_5(\mathbf{0}, \boldsymbol{\Omega}_\alpha) \quad (29)$$

for $p/n \rightarrow c \in [0, 1]$ as $n \rightarrow \infty$ where $\boldsymbol{\Omega}_\alpha$ is given in Eq. (30).

Since

$$\hat{B}_n \xrightarrow{a.s} B \quad \text{for } \frac{p}{n} \rightarrow c \in [0, 1] \quad \text{as } n \rightarrow \infty,$$

the application of Slutsky's lemma (c.f., [44, Theorem 1.5]) leads to the asymptotic distribution of $\hat{\alpha}_c^*$ as given in Theorem 2.

Theorem 2: Under the assumptions of Theorem 1, it holds that

$$\sqrt{n}(\hat{\alpha}_c^* - \alpha^*) \xrightarrow{d} \mathcal{N}(0, C_\alpha), \quad (31)$$

for $p/n \rightarrow c \in [0, 1]$ as $n \rightarrow \infty$ where

$$C_\alpha = \frac{1}{B^2} \mathbf{d}' \boldsymbol{\Omega}_\alpha \mathbf{d}. \quad (32)$$

Finally, using (13), (18), and (24) a consistent estimator for C_α is given by

$$\hat{C}_\alpha = \frac{1}{\hat{B}_n^2} \mathbf{d}' \hat{\boldsymbol{\Omega}}_{\alpha;c} \mathbf{d}, \quad (33)$$

where $\hat{\boldsymbol{\Omega}}_{\alpha;c}$ is a consistent estimator for $\boldsymbol{\Omega}_\alpha$ expressed as Eq. (34) shown at the bottom of this page.

Remark 1: In the case of the investor who invests into the GMV portfolio ($\gamma = \infty$), the formulas (23) and (25) simplify to

$$\alpha^* = \frac{(1-c)(V_b - V_{GMV})}{cV_{GMV} + (1-c)(V_b - V_{GMV})}$$

and

$$\hat{\alpha}_c^* = \frac{(1-c)(\hat{V}_b - \hat{V}_c)}{c\hat{V}_c + (1-c)(\hat{V}_b - \hat{V}_c)}.$$

Moreover, the application of Theorem 1 leads to

$$\sqrt{n}(\hat{\alpha}_c^* - \alpha^*) \xrightarrow{d} \mathcal{N} \left(0, \frac{2(1-c)c^2(L_b + 1)}{((1-c)L_b + c)^4} ((2-c)L_b + c) \right) \quad (35)$$

for $p/n \rightarrow c \in (0, 1)$ as $n \rightarrow \infty$ with $L_b = V_b/V_{GMV} - 1$, which coincides with the results obtained in Theorem 2 of [40].

B. Test Based on a Shrinkage Estimator

We use the properties of the shrinkage intensity α^* and of its consistent estimator $\hat{\alpha}_c^*$ to derive an asymptotic test on the

$$\boldsymbol{\Omega}_\alpha = \begin{pmatrix} \frac{V_{GMV}(s+1)}{1-c} & 0 & 0 & V_{GMV} & -2V_{GMV}(R_b - R_{GMV}) \\ 0 & 2\frac{V_{GMV}^2}{1-c} & 0 & 0 & 2V_{GMV}^2 \\ 0 & 0 & \frac{2((s+1)^2+c-1)}{1-c} & 2(R_b - R_{GMV}) & -2(R_b - R_{GMV})^2 \\ V_{GMV} & 0 & 2(R_b - R_{GMV}) & V_b & 0 \\ -2V_{GMV}(R_b - R_{GMV}) & 2V_{GMV}^2 & -2(R_b - R_{GMV})^2 & 0 & 2V_b^2 \end{pmatrix} \quad (30)$$

$$\hat{\boldsymbol{\Omega}}_{\alpha;c} = \begin{pmatrix} \frac{\hat{V}_c(\hat{s}_c+1)}{1-c} & 0 & 0 & \hat{V}_c & -2\hat{V}_c(\hat{R}_b - \hat{R}_{GMV}) \\ 0 & 2\frac{\hat{V}_c^2}{1-c} & 0 & 0 & 2\hat{V}_c^2 \\ 0 & 0 & \frac{2((\hat{s}_c+1)^2+c-1)}{1-c} & 2(\hat{R}_b - \hat{R}_{GMV}) & -2(\hat{R}_b - \hat{R}_{GMV})^2 \\ \hat{V}_c & 0 & 2(\hat{R}_b - \hat{R}_{GMV}) & \hat{V}_b & 0 \\ -2\hat{V}_c(\hat{R}_b - \hat{R}_{GMV}) & 2\hat{V}_c^2 & -2(\hat{R}_b - \hat{R}_{GMV})^2 & 0 & 2\hat{V}_b^2 \end{pmatrix} \quad (34)$$

structure of the EU portfolio. The testing hypotheses are given by

$$H_0 : \mathbf{w}_{EU} = \mathbf{w}_0 \quad \text{against} \quad H_1 : \mathbf{w}_{EU} \neq \mathbf{w}_0, \quad (36)$$

which, in contrast to the hypotheses considered in Section II, allow to test the structure of the whole vector of the EU portfolio weights by using a single test avoiding the problem of multiplicity.

Following [40], the idea behind a statistical test based on the shrinkage approach is the usage of \mathbf{w}_0 as a fixed target portfolio, i.e., to set $\mathbf{b} = \mathbf{w}_0$ in (21). Since \mathbf{w}_0 is the EU optimal portfolio under the null hypothesis in (36), its expected return and variance should satisfy

$$R_{\mathbf{w}_0} = R_{GMV} + \gamma^{-1}s \quad \text{and} \quad V_{\mathbf{w}_0} = V_{GMV} + \gamma^{-2}s. \quad (37)$$

As a result, the numerator in (23) becomes

$$\begin{aligned} A(\mathbf{w}_0) &= (R_{GMV} - R_b) \left(\gamma^{-1} + \frac{\gamma^{-1}}{1-c} \right) \\ &\quad + (V_b - V_{GMV}) + \frac{\gamma^{-2}s}{1-c} \\ &= -\gamma^{-2}s \left(1 + \frac{1}{1-c} \right) + \gamma^{-2}s + \frac{\gamma^{-2}}{1-c}s = 0, \end{aligned} \quad (38)$$

proving that

$$\alpha^* = 0 \quad \text{under } H_0. \quad (39)$$

Hence, for testing (36), one can derive a test on the hypotheses

$$H_0 : \alpha^*(\mathbf{w}_0) = 0 \quad \text{against} \quad H_1 : \alpha^*(\mathbf{w}_0) \neq 0, \quad (40)$$

where the notation $\alpha^*(\mathbf{w}_0)$ denotes the optimal shrinkage intensity as in (23) computed with target portfolio \mathbf{w}_0 . It has to be noted that the hypotheses (36) and (40) are not equivalent. Namely, it is shown in (38) that the null hypothesis in (36) implies (40), but the implication in the other direction is not obviously true. Nevertheless, if either the mean or the variance of the target portfolio coincide with those of the EU portfolio then the hypotheses (36) and (40) are equivalent. Moreover, the rejection of the null hypothesis in (40) ensures the rejection of the null hypothesis in (36) meaning that \mathbf{w}_0 is not the EU optimal portfolio.

Let $\hat{\alpha}_c^*(\mathbf{w}_0)$ be the consistent estimator of $\alpha^*(\mathbf{w}_0)$ as constructed in (25) when the shrinkage target is $\mathbf{b} = \mathbf{w}_0$. Then the application of Theorem 2 shows that

$$\hat{\alpha}_c^*(\mathbf{w}_0) \xrightarrow{a.s.} 0 \text{ for } \frac{p}{n} \rightarrow c \in [0, 1] \text{ as } n \rightarrow \infty,$$

when the null hypothesis in (36) is true.

Moreover, since the numerator in the expression of $\alpha^*(\mathbf{w}_0)$ in (23) under the null hypothesis in (40) is equal to zero, i.e. $A = 0$ where A is defined before (27), we get the following stochastic representation of $\sqrt{n}\hat{\alpha}_c^*(\mathbf{w}_0)$ expressed as

$$\sqrt{n}\hat{\alpha}_c^*(\mathbf{w}_0) = \frac{1}{\hat{B}_n} \mathbf{d}'_0 \sqrt{n} \mathbf{t}, \quad \mathbf{d}_0 = \begin{pmatrix} \gamma^{-1} + \frac{\gamma^{-1}}{1-c_n} \\ -1 \\ \frac{\gamma^{-2}}{1-c_n} \\ -\gamma^{-1} - \frac{\gamma^{-1}}{1-c_n} \\ 1 \end{pmatrix} \quad (41)$$

and \mathbf{t} is defined in (28). The application of Theorem 1 then leads to the following result

Theorem 3: Assume that the conditions of Theorem 1 are fulfilled. Then, under the null hypothesis in (40), it holds that

$$\sqrt{n}\hat{\alpha}_c^*(\mathbf{w}_0) \xrightarrow{d} \mathcal{N}(0, C_{\alpha;0}), \quad (42)$$

for $p/n \rightarrow c \in [0, 1]$ as $n \rightarrow \infty$ with $C_{\alpha;0} = \frac{1}{B^2} \mathbf{d}'_0 \Omega_\alpha \mathbf{d}_0$ where Ω_α is given in (30) and B is defined before (27).

Replacing B^2 and Ω_α by their consistent estimators \hat{B}_n^2 and $\hat{\Omega}_{\alpha;c}$, we get a consistent estimator for $C_{\alpha;0}$ expressed as

$$\hat{C}_{\alpha;0} = \frac{1}{\hat{B}_n^2} \mathbf{d}'_0 \hat{\Omega}_{\alpha;c} \mathbf{d}_0. \quad (43)$$

Then for testing hypotheses (40), we obtain the following test statistic

$$T_\alpha = \sqrt{n} \frac{\hat{\alpha}_c^*(\mathbf{w}_0)}{\sqrt{\hat{C}_{\alpha;0}}} = \sqrt{n} \frac{\hat{\alpha}_c^*(\mathbf{w}_0) \hat{B}_n}{\sqrt{\mathbf{d}'_0 \hat{\Omega}_{\alpha;c} \mathbf{d}_0}}, \quad (44)$$

where $\hat{\alpha}_c^*(\mathbf{w}_0)$ with $\mathbf{b} = \mathbf{w}_0$ and $\hat{\Omega}_{\alpha;c}$ are given in (25) and (34), respectively. Under the null hypothesis in (40) we get that

$$T_\alpha \xrightarrow{d} \mathcal{N}(0, 1)$$

for $p/n \rightarrow c \in [0, 1]$ as $n \rightarrow \infty$ and, hence, the hypothesis that \mathbf{w}_0 are the weights of the EU portfolio is rejected as soon as $|T_\alpha| > z_{1-\beta/2}$ where $z_{1-\beta/2}$ is the $(1 - \beta/2)$ quantile of the standard normal distribution. Under the alternative hypothesis in (40), it does not hold that $\hat{\alpha}_c^*(\mathbf{w}_0) \xrightarrow{a.s.} 0$ and consequently, the test based on T_α can detect the deviation in the null hypotheses of both (36) and (40).

Remark 2: Using that $s = \gamma(R_{\mathbf{w}_0} - R_{GMV})$ (see, (37)) and $\hat{R}_{\mathbf{w}_0}$ and \hat{R}_{GMV} are consistent estimators of $R_{\mathbf{w}_0}$ and R_{GMV} , respectively (see, (24)), another consistent estimator of Ω_α under H_0 in (40) is given by Eq. (46).

$$\tilde{\Omega}_{\alpha;c} = \begin{pmatrix} \frac{\hat{V}_c(\gamma(\hat{R}_{\mathbf{w}_0} - \hat{R}_{GMV})+1)}{1-c} & 0 & 0 & \hat{V}_c & -2\hat{V}_c(\hat{R}_{\mathbf{w}_0} - \hat{R}_{GMV}) \\ 0 & 2\frac{\hat{V}_c^2}{1-c} & 0 & 0 & 2\hat{V}_c^2 \\ 0 & 0 & 2\frac{((\gamma(\hat{R}_{\mathbf{w}_0} - \hat{R}_{GMV})+1)^2+c-1)}{1-c} & 2(\hat{R}_{\mathbf{w}_0} - \hat{R}_{GMV}) & -2(\hat{R}_{\mathbf{w}_0} - \hat{R}_{GMV})^2 \\ \hat{V}_c & 0 & 2(\hat{R}_{\mathbf{w}_0} - \hat{R}_{GMV}) & \hat{V}_{\mathbf{w}_0} & 0 \\ -2\hat{V}_c(\hat{R}_{\mathbf{w}_0} - \hat{R}_{GMV}) & 2\hat{V}_c^2 & -2(\hat{R}_{\mathbf{w}_0} - \hat{R}_{GMV})^2 & 0 & 2\hat{V}_{\mathbf{w}_0}^2 \end{pmatrix} \quad (46)$$

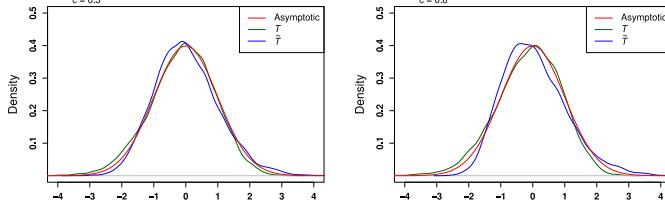


Fig. 2. The high-dimensional asymptotic normal approximation of the densities of T_α and \tilde{T}_α together with their kernel density estimators for $\gamma = 5$, $p = 300$ and $c_n = p/n \in \{0.3, 0.8\}$.

TABLE II
EMPIRICAL SIZES OF THE TWO TESTS BASED ON T_α AND \tilde{T}_α USING 10^4 INDEPENDENT REPLICATIONS.

$c = 0.3$	$c = 0.8$
T_α	0.052
\tilde{T}_α	0.051

Then, the hypotheses in (40) can also be tested by using the following test statistic

$$\tilde{T}_\alpha = \sqrt{n} \frac{\hat{\alpha}_c^*(\mathbf{w}_0) \hat{B}_n}{\sqrt{\mathbf{d}'_0 \tilde{\Omega}_{\alpha;c} \mathbf{d}_0}} \quad (45)$$

which is asymptotically standard normally distributed under H_0 in (40).

Remark 3: Using the duality between the test theory and confidence interval (see, [45]), the null hypothesis in (40) and consequently in (36) are rejected at significance level β as soon as the $(1 - \beta)$ confidence interval constructed for $\alpha^*(\mathbf{w}_0)$ does not include zero. This confidence interval in the case of the test T_α has the boundaries

$$\hat{\alpha}_c^*(\mathbf{w}_0) \pm \frac{z_{1-\beta/2}}{\sqrt{n}} \sqrt{\frac{\mathbf{d}'_0 \hat{\Omega}_{\alpha;c} \mathbf{d}_0}{\hat{B}_n}}, \quad (47)$$

while for the test based on \tilde{T}_α we get

$$\hat{\alpha}_c^*(\mathbf{w}_0) \pm \frac{z_{1-\beta/2}}{\sqrt{n}} \sqrt{\frac{\mathbf{d}'_0 \tilde{\Omega}_{\alpha;c} \mathbf{d}_0}{\hat{B}_n}}. \quad (48)$$

To assess the precision of the asymptotic distribution we use a similar setting as in the last section. In Figure 2, we show the KDEs of the distribution of the test statistics T_α and \tilde{T}_α under the null hypothesis together with their high-dimensional asymptotic distribution. The latter approximates the simulated exact distributions very precisely, although the fit appears to be slightly better for T_α . The empirical size on both cases is close to the nominal size of 5% as it is shown in Table II. Summarizing, we conclude that the derived high-dimensional asymptotic distributions provide good approximations for the distributions of the proposed test statistics for different values of c .

IV. SIMULATION AND EMPIRICAL STUDY

The performance of the derived tests is investigated throughout an extensive simulation study. In particular, we explore the behavior of the tests with respect to their power functions and receiver operative characteristic curves. Additionally, we apply in this section the derived inference procedures to real data.

A. Simulation Study

The sample of asset returns $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ are generated independently from $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. To mimic the behavior of real data we generate the eigenvalues of population covariance matrix $\boldsymbol{\Sigma}$ according to the law $\lambda_i = 0.1e^{\delta c(i-1)/p}$, $i = 1, \dots, p$ (see, [27]) and take its eigenvectors from the spectral decomposition of the standard Wishart random matrix. Then, the covariance matrix is given as follows

$$\boldsymbol{\Sigma} = \boldsymbol{\Theta} \boldsymbol{\Lambda} \boldsymbol{\Theta}', \quad (49)$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix of the predefined eigenvalues and $\boldsymbol{\Theta}$ is a $p \times p$ matrix of eigenvectors. By changing the value of δ , we can control the conditional index of the covariance matrix for different values of c . We set condition index equals to 450. This setting reflects the parametrisation we observed in the empirical study in the next section. The elements of the mean vector are randomly generated from $U(-0.2, 0.2)$, which also corresponds to the natural behavior of daily asset returns.

We assume that the portfolio weights and thus the shrinkage intensity change due to a change in the mean of asset returns. Under the alternative hypothesis, there is an additive shift to the mean vector of the asset returns defined as

$$\boldsymbol{\mu}_1 = \boldsymbol{\mu} + \boldsymbol{\epsilon}, \quad (50)$$

where

$$\boldsymbol{\epsilon} = -a \cdot (\underbrace{1, \dots, 1}_m, \underbrace{0, \dots, 0}_m),$$

where $a = 0.01\kappa$, $\kappa \in \{0, 1, 2, \dots, 35\}$, $m = 0.5p$. Thus, we assume that the expected return on the first m assets decreases.

We conduct the test at the 5% level of significance. We put $p = 300$ and $c \in \{0.3, 0.8\}$. The number of repetitions is 10^4 and $\gamma = 5$. For the ROC curves we fix a at 0.08. The results are illustrated in Figure 3. It can be seen that both tests T_α and \tilde{T}_α display an overall consistency and a good performance in terms of power functions and ROC curves. The behavior is better for smaller values of c and not substantially worse in case of $c = 0.8$. The test based on the test statistic given in (45) outperforms the test given in (44) and for this reason it is used in the empirical illustration in Section IV-B.

B. Empirical Study

In this section, we apply the derived theoretical results to real data. The objective is to determine the periods where the shrinkage intensity is significantly different from zero and thus the EU optimal portfolio is significantly different from the target or the benchmark portfolio \mathbf{b} . This study is based on daily return data of all companies listed in the S&P 500 index for the period

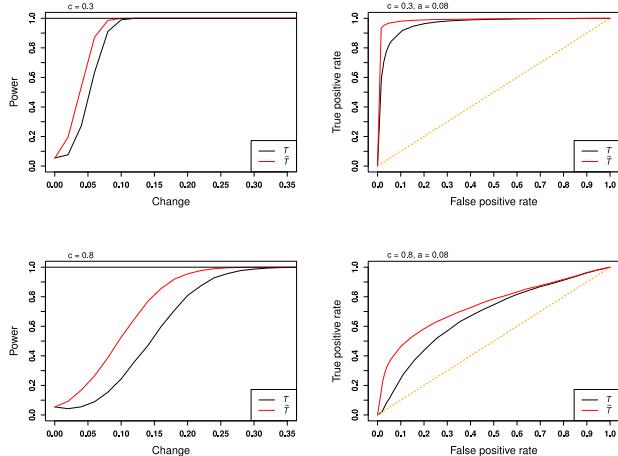


Fig. 3. Empirical power functions of the proposed tests as a function of the change a (left) and ROC curves of two tests for $a = 0.08$ (right) for different values of c according to the scenario given in (50) and $p = 300$.

from April 1999 to March 2020. We assume that the investor allocates her wealth to portfolios of size $p \in \{100, 300\}$ with daily reallocation. She selects the first p assets in alphabetic order from the available data. Note that due permanent changes in the index composition, only 369 out of 500 assets were included in the index over the whole time span. For this reason and to guarantee comparability of the results we cannot consider all 500 assets. The sample size n is chosen to attain $c \in \{0.3, 0.5, 0.8\}$, i.e. $n = p/c$. We put $\gamma = 5$ which is a common value for the risk aversion coefficient in financial literature. As the target portfolio we consider the equally weighted portfolio with all weights equal $1/p$. Despite of its simplicity this portfolio appears to show a superior long-run performance and dominates many more sophisticated trading strategies (see, [46]).

Figure 4 shows the time series of estimated shrinkage intensities together with 95% confidence intervals as defined in (48). If $c = 0.3$, then the shrinkage intensity is close to one indicating that the EU portfolio clearly dominates benchmark in the convex combination. This is due to the fact, that the investor has more historical data to estimate the unknown parameters and the estimation risk is relatively low. If c increases, the sample available for a portfolio of a fixed size gets smaller and the shrinkage intensity shifts towards zero. The benchmark portfolio gets higher weight and for $c = 0.8$ it even becomes dominant. The same reasoning applies if we analyse the impact of increase in p from 100 to 300. Fixed c and larger p increase the sample size n and has a stabilizing impact on the shrinkage intensity.

We cannot reject the null hypothesis of the test based on \hat{T}_α in (45) that the shrinkage intensity is zero if the confidence intervals cover the zero value (see, Remark 3 above). The figures reveal that we never opt for H_0 if $c = 0.3$ or 0.5. Thus, for this parameter constellation the portfolio weights of the EU portfolio are always significantly different from the weights of the equally weighted portfolio. The situation changes for $c = 0.8$ where we do have periods with not rejected H_0 in (40). Similar behavior is observed for $p = 300$ too, however, here the intensities and their variances are more stable leading to less periods with not rejected H_0 .

Recall that a non-rejection of H_0 in (40) does not guarantee that the weights of the EU portfolio coincide with the weights of the target portfolio. To elaborate on the difference between the two portfolios and to get more economic insight into the dynamics of the intensities we consider Figure 5. Here we plot the differences between the means and variances of the GMV and the equally weighted benchmark. These quantities determine the behavior of the empirical shrinkage intensity in (23). On the one hand, we observe in Figure 4 that the shrinkage intensity increases during a crisis period, e.g., 2002-2003 and 2008-2010. This seems to be surprising since the volatility of returns is high in this period and the equally weighted portfolio is believed to reduce the risk. However, Figure 5 shows that the variance of the benchmark portfolio is much higher (i.e., $\hat{V}_b > \hat{V}_c$) and its return is much lower (i.e., $\hat{R}_b < \hat{R}_{GMV}$) compared to the GMV portfolio in the crisis period leading to a higher relative precision and efficiency of the EU portfolio. On another hand, the mean returns and the variances are almost indistinguishable in calm periods leading to shrinkage intensities closer to zero and even insignificant for larger c 's. Thus, we conclude that non-rejecting H_0 is driven by high similarity between the mean and the variance of the target and GMV portfolios.

V. SUMMARY

This paper is dedicated to portfolio selection problems driven by high-dimensional financial data sets. In particular, we deal with optimal asset allocation in a high-dimensional asymptotic regime, namely when the number of assets and the sample size tend to infinity at the same rate. Due to the curse of dimensionality in the parameter estimation process, asset allocation for such portfolios becomes a challenging task. Using the techniques from the theory of random matrices, new inferential procedures based on the optimal shrinkage intensity for testing the efficiency of the high-dimensional EU portfolio are developed and the asymptotic distributions of the proposed test statistics are derived. In extensive simulations, we show that the suggested tests have excellent performance characteristics for various values of c . The practical advantage of the proposed procedures are demonstrated in an empirical study based on stocks included into the S&P 500 index. Namely, we found that depending on the value of p/n there are periods of time where one can clearly reject the null hypothesis of equality between the EU portfolio and the equally weighted portfolio. Moreover, the EU portfolio is dominating the benchmark in these periods. In particular, we show that one can undoubtedly beat the naive portfolio during volatile time periods, which brings some new insights into the understanding of the expected utility portfolios. This fact is surprising because if the volatility of returns is high the equally weighted portfolio is widely believed to reduce the risk. On the other hand, in calm periods the mean returns and variances of assets are almost indistinguishable leading to a similarity between the mean and the variance of the target and GMV portfolios.

Another issue, which is of central importance is the assumption of normality that we impose on the distribution of the asset

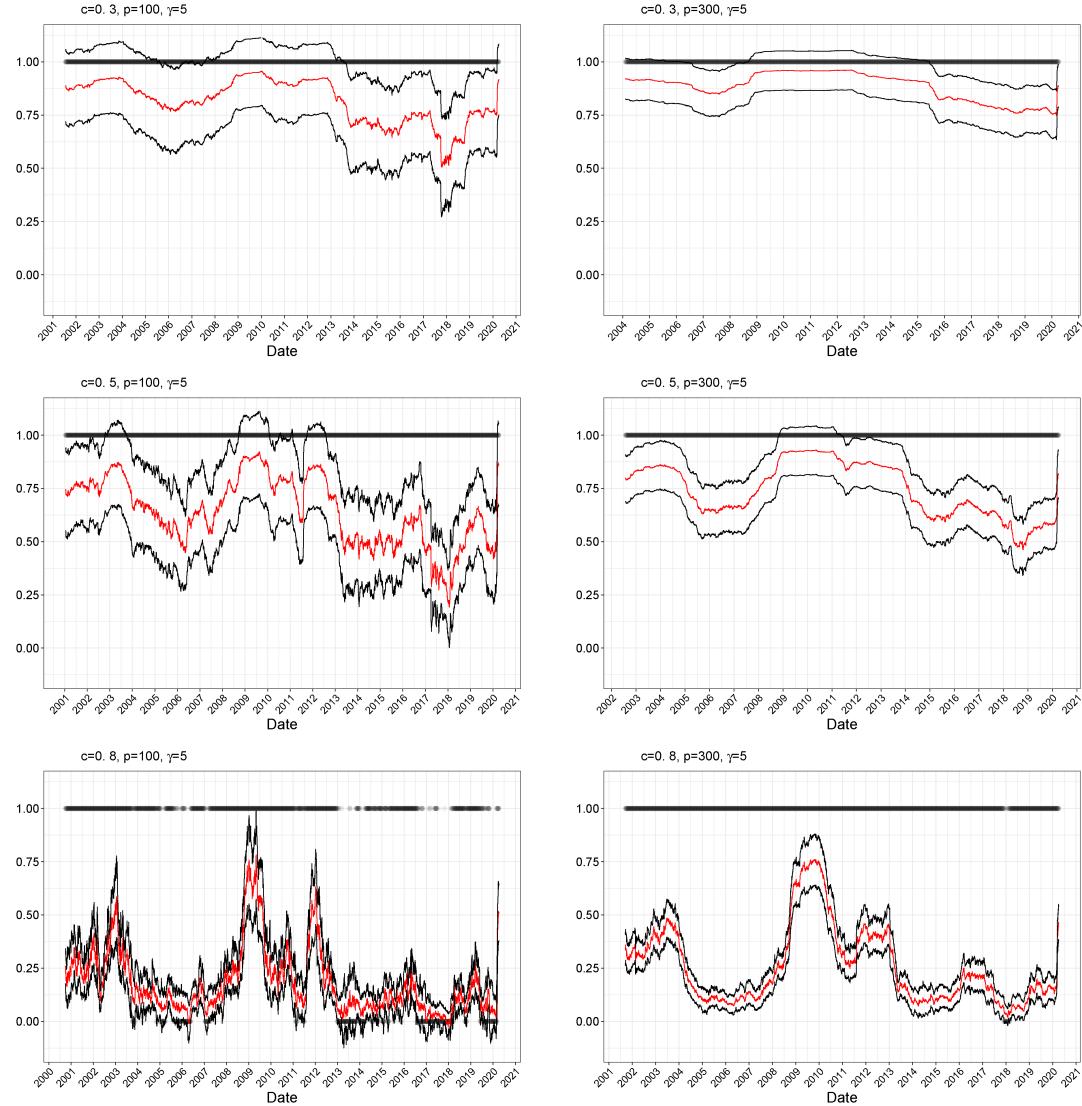


Fig. 4. Estimated shrinkage intensities for the equally weighted portfolio as the target portfolio ($p = 100$ on the right and $p = 300$ on the left) with 95% pointwise confidence intervals. The black dots indicate the periods with rejected H_0 (1-values) and not rejected H_0 (0-values).

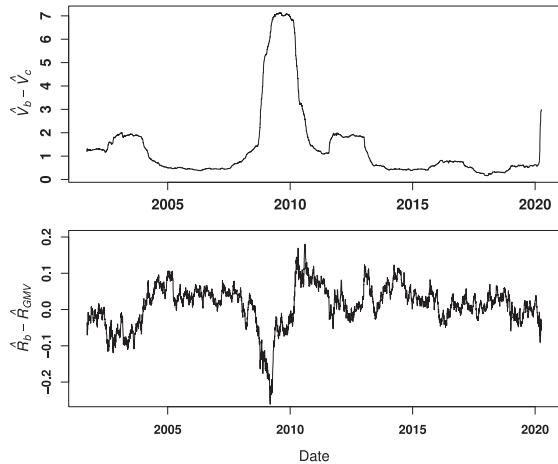


Fig. 5. Components of the estimated shrinkage intensity given in (25) using equally weighted target for $c = 0.8$, $p = 300$ and $\gamma = 5$.

returns. Although the asset returns are clearly not normally distributed our procedure gives an effective method how to assess the difference between high-dimensional benchmark and optimal portfolio in a simplistic way and heavy-tailed returns might make our conclusions even more extreme. In general, we expect the asymptotic distribution of optimal shrinkage intensity to depend of the fourth moment (kurtosis). This interesting topic will be investigated in a separate paper and is left for future research. Furthermore, an extension to a multivariate shrinkage intensity can introduce more flexibility into the model by letting each asset to be shrunk individually. The asset-specific intensity mimics the risk of each asset and allows us to evaluate their individual proximity to the benchmark. Such setting would allow for more general tests and, particularly, for building tracking portfolios. This will be also considered in future research projects.

APPENDIX

In this section the proofs of the theoretical results are given. The proof of Theorem 1 is based on Lemmas 1-2.

Lemma 1: Let $\mathbf{z}_1, \dots, \mathbf{z}_n$ be an independent sample from the p -dimensional standard normal distribution and let

$$\mathbf{S}_n = \frac{1}{n-1} \sum_{j=1}^n (\mathbf{z}_j - \bar{\mathbf{z}})(\mathbf{z}_j - \bar{\mathbf{z}})' \quad (51)$$

be the corresponding sample covariance matrix. Let $\mathbf{m}_1, \mathbf{m}_2$, and \mathbf{m}_3 be the p -dimensional vector of constants with the Euclidean norms equal to one. Then

$$\begin{aligned} & \sqrt{n} \begin{pmatrix} \mathbf{m}'_1 \mathbf{S}_n \mathbf{m}_1 - 1 \\ \mathbf{m}'_2 \mathbf{S}_n^{-1} \mathbf{m}_2 - \frac{1}{1-c_n} \\ \mathbf{m}'_2 \mathbf{S}_n^{-1} \mathbf{m}_3 - \frac{1}{1-c_n} \mathbf{m}'_2 \mathbf{m}_3 \\ \mathbf{m}'_3 \mathbf{S}_n^{-1} \mathbf{m}_3 - \frac{1}{1-c_n} \end{pmatrix} \\ & \xrightarrow{d} \mathcal{N}_4 \left(\mathbf{0}, \frac{2}{c} \Theta(\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3) \circ \Lambda \right), \end{aligned} \quad (52)$$

with

$$\Theta(\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3) = \begin{pmatrix} 1 & \theta_{12}^2 & \theta_{12}\theta_{13} & \theta_{13}^2 \\ \theta_{12}^2 & 1 & \theta_{23} & \theta_{23}^2 \\ \theta_{12}\theta_{13} & \theta_{23} & 0.5 + 0.5\theta_{23}^2 & \theta_{23} \\ \theta_{13}^2 & \theta_{23}^2 & \theta_{23} & 1 \end{pmatrix}, \quad (53)$$

where

$$\theta_{12} = \lim_{n \rightarrow \infty} (\mathbf{m}'_1 \mathbf{m}_2),$$

$$\theta_{13} = \lim_{n \rightarrow \infty} (\mathbf{m}'_1 \mathbf{m}_3),$$

$$\theta_{23} = \lim_{n \rightarrow \infty} (\mathbf{m}'_2 \mathbf{m}_3)$$

and

$$\Lambda = \begin{pmatrix} c & -\frac{c}{1-c} & -\frac{c}{1-c} & -\frac{c}{1-c} \\ -\frac{c}{1-c} & \frac{c}{(1-c)^3} & \frac{c}{(1-c)^3} & \frac{c}{(1-c)^3} \\ -\frac{c}{1-c} & \frac{c}{(1-c)^3} & \frac{c}{(1-c)^3} & \frac{c}{(1-c)^3} \\ -\frac{c}{1-c} & \frac{c}{(1-c)^3} & \frac{c}{(1-c)^3} & \frac{c}{(1-c)^3} \end{pmatrix}, \quad (54)$$

where the symbol \circ denotes the Hadamard (elementwise) product of matrices.

Proof of Lemma 1 Since $(n-1)\mathbf{S}_n$ has a p -dimensional Wishart distribution with the identity covariance matrix, we get that there exists a $p \times (n-1)$ matrix $\tilde{\mathbf{Z}}$ whose entries are independent and standard normally distributed such that $(n-1)\mathbf{S}_n = \tilde{\mathbf{Z}}\tilde{\mathbf{Z}}'$. The application of Theorem 2 in [47] leads to (52) with Θ as in (53) and Λ given by

$$\Lambda = \begin{pmatrix} \lambda_1 & \lambda_2 & \lambda_2 & \lambda_2 \\ \lambda_2 & \lambda_3 & \lambda_3 & \lambda_3 \\ \lambda_2 & \lambda_3 & \lambda_3 & \lambda_3 \\ \lambda_2 & \lambda_3 & \lambda_3 & \lambda_3 \end{pmatrix}$$

with

$$\lambda_1 = \int_{a_-}^{a_+} z^2 dF_c(z) - \left(\int_{a_-}^{a_+} zdF_c(z) \right)^2,$$

$$\lambda_2 = 1 - \int_{a_-}^{a_+} zdF_c(z) \int_{a_-}^{a_+} \frac{1}{z} dF_c(z),$$

$$\lambda_3 = \int_{a_-}^{a_+} \frac{1}{z^2} dF_c(z) - \left(\int_{a_-}^{a_+} \frac{1}{z} dF_c(z) \right)^2$$

where the function $F_c(z)$ denotes the cumulative distribution function of the Marchenko-Pastur law (see, [21]) for $c < 1$ expressed as

$$dF_c(z) = \frac{1}{2\pi z c} \sqrt{(a_+ - z)(z - a_-)} \mathbb{1}_{[a_-, a_+]}(z) dz,$$

where $a_\pm = (1 \pm \sqrt{c})^2$. The moments of $F_c(z)$ present in Λ can be found in [48, Lemma 14]. This completes the proof of the lemma. ■

Lemma 2: Under the conditions of Theorem 1 it holds that

$$\begin{aligned} \sqrt{n} \mathbf{h} &= \sqrt{n} \begin{pmatrix} \mathbf{1}'_p \hat{\Sigma}_n^{-1} \bar{\mathbf{x}}_n - \frac{1}{1-c_n} \mathbf{1}'_p \Sigma^{-1} \mu \\ \mathbf{1}'_p \hat{\Sigma}_n^{-1} \mathbf{1}_p - \frac{1}{1-c_n} \mathbf{1}'_p \Sigma^{-1} \mathbf{1}_p \\ \bar{\mathbf{x}}'_n \hat{\Sigma}_n^{-1} \bar{\mathbf{x}}_n - \frac{1}{1-c_n} \mu' \Sigma^{-1} \mu - \frac{c_n}{1-c_n} \\ \mathbf{b}' \bar{\mathbf{x}}_n - \mathbf{b}' \mu \\ \mathbf{b}' \hat{\Sigma}_n \mathbf{b} - \mathbf{b}' \Sigma \mathbf{b} \end{pmatrix} \\ &\xrightarrow{d} \mathcal{N}_5(\mathbf{0}, \Xi) \end{aligned} \quad (55)$$

for $c_n = p/n \rightarrow c \in [0, 1]$ as $n \rightarrow \infty$ with

$$\Xi = \begin{pmatrix} \xi_{11} & \xi_{12} & \xi_{13} & \xi_{14} & \xi_{15} \\ \xi_{12} & \xi_{22} & \xi_{23} & \xi_{24} & \xi_{25} \\ \xi_{13} & \xi_{23} & \xi_{33} & \xi_{34} & \xi_{35} \\ \xi_{14} & \xi_{24} & \xi_{34} & \xi_{44} & \xi_{45} \\ \xi_{15} & \xi_{25} & \xi_{35} & \xi_{45} & \xi_{55} \end{pmatrix}, \quad (56)$$

where $\xi_{11} = \frac{1}{(1-c)^3} \frac{1}{V_{GMV}} (s^* + \frac{R_{GMV}^2}{V_{GMV}})$,
 $\xi_{22} = \frac{2}{(1-c)^3} \frac{1}{V_{GMV}^2}$, $\xi_{33} = \frac{2}{(1-c)^3} ((s^*)^2 + c - 1)$,
 $\xi_{44} = V_b$, $\xi_{55} = 2V_b^2$, $\xi_{12} = \frac{2}{(1-c)^3} \frac{R_{GMV}}{V_{GMV}^2}$,
 $\xi_{13} = \frac{2}{(1-c)^3} \frac{R_{GMV}s^*}{V_{GMV}^2}$, $\xi_{14} = \frac{1}{1-c}$, $\xi_{15} = -\frac{2}{1-c} R_b$,
 $\xi_{23} = \frac{2}{(1-c)^3} \frac{R_{GMV}^2}{V_{GMV}^2}$, $\xi_{24} = 0$, $\xi_{25} = -\frac{2}{1-c}$, $\xi_{34} = \frac{2R_b}{1-c}$,
 $\xi_{35} = -\frac{2}{1-c} R_b^2$, $\xi_{45} = 0$ and $s^* = s + \frac{R_{GMV}^2}{V_{GMV}} + 1$.

Proof of Lemma 2 Let $\mathbf{a}' = (a_1, a_2, a_3, a_4, a_5)$ be an arbitrary vector of constants. Next, we show that $\sqrt{n}\mathbf{a}'\mathbf{h} \xrightarrow{d} \mathcal{N}(0, \mathbf{a}'\Xi\mathbf{a})$, which will prove the statement of the lemma.

Since $\mathbf{x}_1, \dots, \mathbf{x}_n$ are independent and identically distributed with $\mathbf{x}_i \sim \mathcal{N}_p(\mu, \Sigma)$, we get that $\mathbf{x}_i = \mu + \Sigma \mathbf{z}_i$ where $\mathbf{z}_1, \dots, \mathbf{z}_n$ are independent standard normally distributed and $\Sigma^{1/2}$ is the symmetric square root of Σ . Moreover, it holds that

$$\bar{\mathbf{x}}_n = \mu + \Sigma \bar{\mathbf{z}}_n \quad \text{and} \quad \hat{\Sigma}_n = \Sigma^{1/2} \mathbf{S}_n \Sigma^{1/2},$$

where

$$\bar{\mathbf{z}}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i \quad \text{and} \quad \mathbf{S}_n = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{z}_i - \bar{\mathbf{z}}_n)(\mathbf{z}_i - \bar{\mathbf{z}}_n)'.$$

To this end, we have that $\bar{\mathbf{z}}_n$ and \mathbf{S}_n are independent with $\sqrt{n}\bar{\mathbf{z}}_n$ standard normally distributed and $(n-1)\mathbf{S}_n$ standard Wishart distributed.

Let $\nu = \Sigma^{-1/2}\mu$. We get

$$\sqrt{n}\mathbf{a}'\mathbf{h} = H_1(\bar{\mathbf{z}}_n, \mathbf{S}_n) + H_2(\bar{\mathbf{z}}_n),$$

with

$$\begin{aligned} H_1(\bar{\mathbf{z}}_n, \mathbf{S}_n) &= a_1 \sqrt{\mathbf{1}'_p \Sigma^{-1} \mathbf{1}_p} \sqrt{\bar{\mathbf{x}}'_n \Sigma^{-1} \bar{\mathbf{x}}_n} \sqrt{n} \\ &\times \left(\frac{\mathbf{1}'_p \hat{\Sigma}_n^{-1} \bar{\mathbf{x}}_n}{\sqrt{\mathbf{1}'_p \Sigma^{-1} \mathbf{1}_p} \sqrt{\bar{\mathbf{x}}'_n \Sigma^{-1} \bar{\mathbf{x}}_n}} - \frac{\frac{1}{1-c_n} \mathbf{1}'_p \Sigma^{-1} \bar{\mathbf{x}}_n}{\sqrt{\mathbf{1}'_p \Sigma^{-1} \mathbf{1}_p} \sqrt{\bar{\mathbf{x}}'_n \Sigma^{-1} \bar{\mathbf{x}}_n}} \right) \\ &+ a_2 \mathbf{1}'_p \Sigma^{-1} \mathbf{1}_p \sqrt{n} \left(\frac{\mathbf{1}'_p \hat{\Sigma}_n^{-1} \mathbf{1}_p}{\mathbf{1}'_p \Sigma^{-1} \mathbf{1}_p} - \frac{1}{1-c_n} \right) \\ &+ a_3 \bar{\mathbf{x}}'_n \Sigma^{-1} \bar{\mathbf{x}}_n \sqrt{n} \left(\frac{\bar{\mathbf{x}}'_n \hat{\Sigma}_n^{-1} \bar{\mathbf{x}}_n}{\bar{\mathbf{x}}'_n \Sigma^{-1} \bar{\mathbf{x}}_n} - \frac{1}{1-c_n} \right) \\ &+ a_5 \mathbf{b}' \Sigma \mathbf{b} \sqrt{n} \left(\frac{\mathbf{b}' \hat{\Sigma}_n \mathbf{b}}{\mathbf{b}' \Sigma \mathbf{b}} - 1 \right) \\ &= \mathbf{d}'_1(\bar{\mathbf{z}}_n) \sqrt{n} \mathbf{h}_1(\bar{\mathbf{z}}_n, \mathbf{S}_n) \end{aligned}$$

and

$$\begin{aligned} H_2(\bar{\mathbf{z}}_n) &= a_1 \frac{1}{1-c_n} \sqrt{n} (\mathbf{1}'_p \Sigma^{-1} \bar{\mathbf{x}}_n - \mathbf{1}'_p \Sigma^{-1} \mu) \\ &+ a_3 \frac{1}{1-c_n} \sqrt{n} (\bar{\mathbf{x}}'_n \Sigma^{-1} \bar{\mathbf{x}}_n - \mu' \Sigma^{-1} \mu - c_n) \\ &+ a_4 \sqrt{n} (\mathbf{b}' \bar{\mathbf{x}}_n - \mathbf{b}' \mu) \\ &= \frac{a_3}{1-c_n} \sqrt{n} ((\bar{\mathbf{z}}_n + \mathbf{d}_2)' (\bar{\mathbf{z}}_n + \mathbf{d}_2) - \mathbf{d}'_2 \mathbf{d}_2 - c_n) \end{aligned}$$

with

$$\begin{aligned} \mathbf{d}_1(\bar{\mathbf{z}}_n) &= \begin{pmatrix} a_5 \mathbf{b}' \Sigma \mathbf{b} \\ a_2 \mathbf{1}'_p \Sigma^{-1} \mathbf{1}_p \\ a_1 \sqrt{\mathbf{1}'_p \Sigma^{-1} \mathbf{1}_p} \sqrt{(\nu + \bar{\mathbf{z}}_n)' (\nu + \bar{\mathbf{z}}_n)} \\ a_3 (\nu + \bar{\mathbf{z}}_n)' (\nu + \bar{\mathbf{z}}_n) \end{pmatrix}, \\ \mathbf{d}_2 &= \frac{1-c_n}{a_3} \\ &\times \left(\frac{a_3}{1-c_n} \Sigma^{-1/2} \mu + \frac{a_1}{2(1-c_n)} \Sigma^{-1/2} \mathbf{1}_p + \frac{a_4}{2} \Sigma^{1/2} \mathbf{b} \right), \end{aligned}$$

and

$$\begin{aligned} \mathbf{h}_1(\bar{\mathbf{z}}_n, \mathbf{S}_n) &= \begin{pmatrix} \mathbf{b}' \Sigma^{1/2} \mathbf{S}_n \Sigma^{1/2} \mathbf{b} \\ \mathbf{1}'_p \Sigma^{-1/2} \mathbf{S}_n^{-1} \Sigma^{-1/2} \mathbf{1}_p \\ \mathbf{1}'_p \Sigma^{-1} \mathbf{1}_p \\ \frac{\mathbf{1}'_p \Sigma^{-1/2} \mathbf{S}_n^{-1} (\nu + \bar{\mathbf{z}}_n)}{\sqrt{\mathbf{1}'_p \Sigma^{-1} \mathbf{1}_p} \sqrt{(\nu + \bar{\mathbf{z}}_n)' (\nu + \bar{\mathbf{z}}_n)}} \\ \frac{(\nu + \bar{\mathbf{z}}_n)' \mathbf{S}_n^{-1} (\nu + \bar{\mathbf{z}}_n)}{(\nu + \bar{\mathbf{z}}_n)' (\nu + \bar{\mathbf{z}}_n)} \end{pmatrix} \\ &- \begin{pmatrix} 1 \\ \frac{1}{1-c_n} \\ \frac{\mathbf{1}'_p \Sigma^{-1} (\nu + \bar{\mathbf{z}}_n)}{\sqrt{\mathbf{1}'_p \Sigma^{-1} \mathbf{1}_p} \sqrt{(\nu + \bar{\mathbf{z}}_n)' (\nu + \bar{\mathbf{z}}_n)}} \\ \frac{1}{1-c_n} \end{pmatrix}. \end{aligned}$$

Since \mathbf{S}_n and $\bar{\mathbf{z}}_n$ are independent the conditional distribution of $H_1(\bar{\mathbf{z}}_n, \mathbf{S}_n)$ given $\bar{\mathbf{z}}_n = \mathbf{v}$ coincides with $H_1(\mathbf{v}, \mathbf{S}_n)$. Furthermore, the application of Lemma 1 to $\sqrt{n}\mathbf{h}_1(\mathbf{v}, \mathbf{S}_n)$ proves that it is asymptotically normally distributed and, thus, the asymptotic stochastic representation of $H_1(\bar{\mathbf{z}}_n, \mathbf{S}_n)$ is given by Eq. (57) shown at the bottom of this page, where $\omega_1 \xrightarrow{d} \mathcal{N}(0, 1)$ and is independent of $\bar{\mathbf{z}}_n$ and hence of $H_2(\bar{\mathbf{z}}_n)$. Finally, we have that $n(\bar{\mathbf{z}}_n + \mathbf{d}_2)'(\bar{\mathbf{z}}_n + \mathbf{d}_2)$ has a non-central χ^2 distribution with p degrees of freedom and noncentrality parameter $n\mathbf{d}'_2 \mathbf{d}_2$. The application of [49] leads to

$$\begin{aligned} &\sqrt{p} \left(\frac{n(\bar{\mathbf{z}}_n + \mathbf{d}_2)'(\bar{\mathbf{z}}_n + \mathbf{d}_2)}{p} - \frac{n\mathbf{d}'_2 \mathbf{d}_2}{p} - 1 \right) \\ &\xrightarrow{d} \mathcal{N} \left(0, 2 + 4 \frac{\mathbf{d}'_2 \mathbf{d}_2}{c} \right) \end{aligned}$$

and, consequently,

$$H_2(\bar{\mathbf{z}}_n) \stackrel{d}{=} \frac{\sqrt{c_n}}{1-c_n} a_3 \sqrt{2 + 4 \frac{\mathbf{d}'_2 \mathbf{d}_2}{c_n}} \omega_2. \quad (58)$$

where $\omega_2 \xrightarrow{d} \mathcal{N}(0, 1)$.

Using that $\nu' \bar{\mathbf{z}}_n \xrightarrow{a.s.} 0$ and $\bar{\mathbf{z}}'_n \bar{\mathbf{z}}_n \xrightarrow{a.s.} c$, the application of Slutsky's lemma (c.f., [44, Theorem 1.5]) leads to

$$\sqrt{n}\mathbf{a}'\mathbf{h} \xrightarrow{d} \mathcal{N}(0, \mathbf{a}'\boldsymbol{\Xi}\mathbf{a})$$

for $p/n = c + o(n^{-1/2})$ as $n \rightarrow \infty$ where $\boldsymbol{\Xi}$ is given in (56). Since \mathbf{a} is an arbitrary vector, the statement of Lemma 2 is proved. \blacksquare

Proof of Theorem 1: It holds that

$$\begin{aligned} \hat{R}_{GMV} - R_{GMV} &= \hat{V}_{GMV} \left(\left(\mathbf{1}'_p \hat{\Sigma}_n^{-1} \bar{\mathbf{x}}_n - \frac{1}{1-c_n} \mathbf{1}'_p \Sigma^{-1} \mu \right) \right. \\ &\quad \left. - R_{GMV} \left(\mathbf{1}'_p \hat{\Sigma}_n^{-1} \mathbf{1}_p - \frac{1}{1-c_n} \mathbf{1}'_p \Sigma^{-1} \mathbf{1}_p \right) \right), \\ \hat{V}_c - V_{GMV} &= -V_{GMV} \hat{V}_{GMV} \left(\mathbf{1}'_p \hat{\Sigma}_n^{-1} \mathbf{1}_p - \frac{1}{1-c_n} \mathbf{1}'_p \Sigma^{-1} \mathbf{1}_p \right) \end{aligned}$$

$$H_1(\bar{\mathbf{z}}_n \mathbf{v}, \mathbf{S}_n) \stackrel{d}{=} \sqrt{\frac{2}{c}} \omega_1 \sqrt{\mathbf{d}'_1 \left(\Theta \left(\frac{\Sigma^{1/2} \mathbf{b}}{\sqrt{\mathbf{b}' \Sigma \mathbf{b}}}, \frac{\Sigma^{-1/2} \mathbf{1}_p}{\sqrt{\mathbf{1}'_p \Sigma^{-1} \mathbf{1}_p}}, \frac{(\nu + \bar{\mathbf{z}}_n)}{\sqrt{(\nu + \bar{\mathbf{z}}_n)' (\nu + \bar{\mathbf{z}}_n)}} \right) \circ \Lambda \right) \mathbf{d}_1} \quad (57)$$

$$\mathbf{D} = \begin{pmatrix} (1 - c_n)\hat{V}_c & -(1 - c_n)\hat{V}_c R_{GMV} & 0 & 0 & 0 \\ 0 & -(1 - c_n)\hat{V}_c V_{GMV} & 0 & 0 & 0 \\ (1 - c_n)\hat{V}_c \left(\frac{R_{GMV}}{V_{GMV}} - \frac{\hat{R}_{GMV}}{\hat{V}_c} \right) & (1 - c_n)\hat{V}_c \frac{R_{GMV}^2}{V_{GMV}} & (1 - c_n) & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (59)$$

$$\begin{aligned} \hat{s}_c - s_{GMV} &= (1 - c_n) \left(\bar{\mathbf{x}}_n' \hat{\Sigma}_n^{-1} \bar{\mathbf{x}}_n - \frac{1}{1 - c_n} \boldsymbol{\mu}' \Sigma^{-1} \boldsymbol{\mu} - \frac{c_n}{1 - c_n} \right) \\ &\quad - (1 - c_n) \left(\frac{(\mathbf{1}_p' \hat{\Sigma}_n^{-1} \bar{\mathbf{x}}_n)^2}{\mathbf{1}_p' \hat{\Sigma}_n^{-1} \mathbf{1}_p} - \frac{1}{1 - c_n} \frac{(\mathbf{1}_p' \Sigma^{-1} \boldsymbol{\mu})^2}{\mathbf{1}_p' \Sigma^{-1} \mathbf{1}_p} \right) \\ &= (1 - c_n) \left(\bar{\mathbf{x}}_n' \hat{\Sigma}_n^{-1} \bar{\mathbf{x}}_n - \frac{1}{1 - c_n} \boldsymbol{\mu}' \Sigma^{-1} \boldsymbol{\mu} - \frac{c_n}{1 - c_n} \right) \\ &\quad - (1 - c_n) \hat{V}_{GMV} \left(\left(\mathbf{1}_p' \hat{\Sigma}_n^{-1} \bar{\mathbf{x}}_n + \frac{1}{1 - c_n} \frac{R_{GMV}}{V_{GMV}} \right) \right. \\ &\quad \times \left(\mathbf{1}_p' \hat{\Sigma}_n^{-1} \bar{\mathbf{x}}_n - \frac{1}{1 - c_n} \mathbf{1}_p' \Sigma^{-1} \boldsymbol{\mu} \right) \\ &\quad \left. - \frac{1}{1 - c_n} \frac{R_{GMV}^2}{V_{GMV}} \left(\mathbf{1}_p' \hat{\Sigma}_n^{-1} \mathbf{1}_p - \frac{1}{1 - c_n} \mathbf{1}_p' \Sigma^{-1} \mathbf{1}_p \right) \right). \end{aligned}$$

Hence,

$$\sqrt{n} \begin{pmatrix} \hat{R}_{GMV} - R_{GMV} \\ \hat{V}_c - V_{GMV} \\ \hat{s}_c - s \\ \hat{R}_b - R_b \\ \hat{V}_b - V_b \end{pmatrix} = \mathbf{D} \sqrt{n} \mathbf{h},$$

with \mathbf{h} defined in (55) and \mathbf{D} in Eq. (59) shown at the top of this page.

The application of $\hat{R}_{GMV} \xrightarrow{a.s.} R_{GMV}$ and $\hat{V}_c \xrightarrow{a.s.} V_{GMV}$ for $p/n \rightarrow c \in [0, 1]$ as $n \rightarrow \infty$, the results of Lemma 2, and Slutsky's lemma (c.f., [44, Theorem 1.5]) complete the proof of the theorem. ■

ACKNOWLEDGMENT

The authors would like to thank Professor Pier Luigi Dragotti, the Associate Editor, and three anonymous Reviewers for their helpful suggestions.

REFERENCES

- [1] H. Markowitz, "Portfolio selection," *J. Finance*, vol. 7, no. 1, pp. 77–91, Mar. 1952.
- [2] T. Bodnar, N. Parolya, and W. Schmid, "On the equivalence of quadratic optimization problems commonly used in portfolio theory," *Eur. J. Oper. Res.*, vol. 229, no. 3, pp. 637–644, Sep. 2013.
- [3] J. Ingersoll, *Theory of Financial Decision Making*, ser. G - Reference, Information and Interdisciplinary Subjects Series. Rowman & Littlefield, 1987.
- [4] Y. Okhrin and W. Schmid, "Distributional properties of portfolio weights," *J. Econometrics*, vol. 134, no. 1, pp. 235–256, Feb. 2006.
- [5] T. Bodnar and W. Schmid, "On the exact distribution of the estimated expected utility portfolio weights: Theory and applications," *Statist. Risk Model.*, vol. 28, no. 4, pp. 319–342, Dec. 2011.
- [6] L. Yang, R. Couillet, and M. R. McKay, "A robust statistics approach to minimum variance portfolio optimization," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6684–6697, Dec. 2015.
- [7] A. Woodgate and A. F. Siegel, "How much error is in the tracking error? The impact of estimation risk on fund tracking error," *J. Portfolio Manag.*, vol. 41, no. 2, pp. 84–99, 2015.
- [8] M. Simaan, Y. Simaan, and Y. Tang, "Estimation error in mean returns and the mean-variance efficient frontier," *Int. Rev. Econ. Finance*, vol. 56, pp. 109–124, Jul. 2018.
- [9] Z. Zhao, R. Zhou, and D. P. Palomar, "Optimal mean-reverting portfolio with leverage constraint for statistical arbitrage in finance," *IEEE Trans. Signal Process.*, vol. 67, no. 7, pp. 1681–1695, Apr. 2019.
- [10] A. F. Siegel and A. Woodgate, "Performance of portfolios optimized with estimation error," *Manag. Sci.*, vol. 53, no. 6, pp. 1005–1015, Jun. 2007.
- [11] R. Kan and D. R. Smith, "The distribution of the sample minimum-variance frontier," *Manag. Sci.*, vol. 54, no. 7, pp. 1364–1380, May 2008.
- [12] O. Bodnar and T. Bodnar, "On the unbiased estimator of the efficient frontier," *Int. J. Theor. Appl. Finance*, vol. 13, no. 07, pp. 1065–1073, Nov. 2010.
- [13] C. Stein, "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," in *Proc. 3rd Berkeley Symp. Math. Statist. Probability, Volume 1: Contributions Theory Statist.*, Berkeley, California: 197–206: University of California Press, 1956.
- [14] T. Bodnar, A. K. Gupta, and N. Parolya, "On the strong convergence of the optimal linear shrinkage estimator for large dimensional covariance matrix," *J. Multivariate Anal.*, vol. 132, pp. 215–228, Nov. 2014.
- [15] T. Bodnar, A. K. Gupta, and N. Parolya, "Direct shrinkage estimation of large dimensional precision matrix," *J. Multivariate Anal.*, vol. 146, pp. 223–236, Apr. 2016.
- [16] V. Golosnoy and Y. Okhrin, "Multivariate shrinkage for optimal portfolio weights," *Eur. J. Finance*, vol. 13, no. 5, pp. 441–458, Jul. 2007.
- [17] Y. Okhrin and W. Schmid, *Int. J. Theor. Appl. Finance*, vol. 11, no. 3, pp. 249–276, May 2008.
- [18] G. Frahm and C. Memmel, "Dominating estimators for minimum-variance portfolios," *J. Econometrics*, vol. 159, no. 2, pp. 289–302, Dec. 2010.
- [19] Z. Bai, H. Liu, and W.-K. Wong, "Enhancement of the applicability of Markowitz's portfolio optimization by utilizing random matrix theory," *Math. Finance*, vol. 19, no. 4, pp. 639–667, Oct. 2009.
- [20] J. Bai and S. Shi, "Estimating high dimensional covariance matrices and its applications," *Annals Econom. Finance*, vol. 12, no. 2, pp. 199–215, Nov. 2011.
- [21] Z. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*. Berlin, Germany: Springer, 2010, vol. 20.
- [22] T. Bodnar, O. Okhrin, and N. Parolya, "Optimal shrinkage estimator for high-dimensional mean vector," *J. Multivariate Anal.*, vol. 170, pp. 63–79, Mar. 2019.
- [23] N. Hautsch and S. Voigt, "Large-scale portfolio allocation under transaction costs and model uncertainty," *J. Econometrics*, vol. 212, no. 1, pp. 221–240, Sep. 2019.
- [24] O. Ledoit and M. Wolf, "Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks," *Rev. Fin. Stud.*, vol. 30, no. 12, pp. 4349–4388, Jun. 2017.
- [25] T. Holgersson, P. Karlsson, and A. Stephan, "A risk perspective of estimating portfolio weights of the global minimum-variance portfolio," *ASTA Adv. Statist. Anal.*, vol. 104, no. 1, pp. 59–80, Mar. 2020.
- [26] T. Bodnar, N. Parolya, and W. Schmid, "Estimation of the global minimum variance portfolio in high dimensions," *Eur. J. Oper. Res.*, vol. 266, no. 1, pp. 371–390, Apr. 2018.
- [27] T. Bodnar, Y. Okhrin, and N. Parolya, "Optimal shrinkage-based portfolio selection in high dimensions," *J. Bus. Econ. Statist.*, no. under revision, 2020.
- [28] F. Rubio, X. Mestre, and D. P. Palomar, "Performance analysis and optimal selection of large minimum variance portfolios under estimation risk," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 4, pp. 337–350, Aug. 2012.

- [29] K. Benidis, Y. Feng, and D. P. Palomar, "Sparse portfolios for high-dimensional financial index tracking," *IEEE Trans. Signal Process.*, vol. 66, no. 1, pp. 155–170, Jan. 2018.
- [30] S. Verdú, *Multiuser Detection*. New York, NY, USA: Cambridge Univ. Press, 1998.
- [31] H. L. Van Trees, *Optimum Array Processing*. New York, NY, USA: Wiley, 2002.
- [32] L. Yang, M. R. McKay, and R. Couillet, "High-dimensional MVDR beamforming: Optimized solutions based on spiked random matrix models," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1933–1947, Apr. 2018.
- [33] J. Li, P. Stoica, and Z. Wang, "Doubly constrained robust capon beamformer," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2407–2423, Sep. 2004.
- [34] X. Mestre and M. Lagunas, "Finite sample size effect on MV beamformers: optimum diagonal loading factor for large arrays," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 69–82, Jan. 2006.
- [35] M. Zhang, F. Rubio, X. Mestre, and D. Palomar, "Improved calibration of high-dimensional precision matrices," *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1509–1519, Mar. 2013.
- [36] Y. Feng and D. P. Palomar, *A Signal Processing Perspective on Financial Engineering*. Boston, MA, USA: now Publishers, 2016, vol. 9.
- [37] Z. Bai, D. Jiang, J.-F. Yao, and S. Zheng, "Corrections to LRT on large-dimensional covariance matrix by RMT," *Annals Statist.*, vol. 37, no. 6B, pp. 3822–3840, 2009.
- [38] J. Yao, S. Zheng, and Z. Bai, *Sample Covariance Matrices and High-Dimensional Data Analysis*. New York, NY, USA: Cambridge Univ. Press, 2015.
- [39] T. Bodnar, H. Dette, and N. Parolya, "Testing for independence of large dimensional vectors," *Annals Statist.*, vol. 47, no. 5, pp. 2977–3008, 2019.
- [40] T. Bodnar, S. Dmytryiv, N. Parolya, and W. Schmid, "Tests for the weights of the global minimum variance portfolio in a high-dimensional setting," *IEEE Trans. Signal Process.*, vol. 67, no. 17, pp. 4479–4493, Sep. 2019.
- [41] T. Bodnar and W. Schmid, "A test for the weights of the global minimum variance portfolio in an elliptical model," *Metrika*, vol. 67, no. 2, pp. 127–143, Feb. 2008.
- [42] T. Bodnar, H. Dette, N. Parolya, and E. Thorsén, "Sampling distributions of optimal portfolio weights and characteristics in low and large dimensions," *Random Matrices: Theory Appl.*, no. under revision, 2020.
- [43] T. Dickhaus, *Simultaneous Statistical Inference*. Berlin, Germany: Springer, 2014.
- [44] A. DasGupta, *Asymptotic Theory of Statistics and Probability*. Berlin, Germany: Springer, 2008.
- [45] J. Aitchison, "Confidence-region tests," *J. Royal Statist. Soc.: Series B (Methodological)*, vol. 26, no. 3, pp. 462–476, 1964.
- [46] V. DeMiguel, L. Garlappi, and R. Uppal, "Optimal versus naive diversification: How inefficient is the $1/n$ portfolio strategy?," *Rev. Financial Stud.*, vol. 22, no. 5, pp. 1915–1953, 2009.
- [47] Z. Bai *et al.*, "Asymptotic properties of eigenmatrices of a large sample covariance matrix," *Annals Appl. Probability*, vol. 21, no. 5, pp. 1994–2015, Dec. 2011.
- [48] K. Glombeck, "Statistical inference for high-dimensional global minimum variance portfolios," *Scandinavian J. Statist.*, vol. 41, no. 4, pp. 845–865, Dec. 2014.
- [49] T. Bodnar and M. Reiß, "Exact and asymptotic tests on a factor model in low and large dimensions with applications," *J. Multivariate Analysis*, vol. 150, pp. 125–151, Sep. 2016.



Taras Bodnar received the M.S. degree in mathematics (Hons.) from the Ivan Franko National University of Lviv, Lviv, Ukraine, in 2001, the Ph.D. degree in economics and statistics (summa cum laude) and mathematics and physics from the European University Viadrina, Frankfurt (Oder), Frankfurt, Germany and the Taras Shevchenko National University of Kyiv, Kyiv, Ukraine, in 2004 and 2009, respectively. He is Professor of mathematical statistics with the Department of Mathematics, the Stockholm University, Stockholm, Sweden. From March 2005 to August 2012, he was a Research Associate with the Department of Statistics, European University Viadrina, Frankfurt (Oder), Germany. From September 2012 to November 2014, he was a Research Associate with the Department of Mathematics, Humboldt-University of Berlin, Berlin, Germany. He is a Member of the Editorial Boards of *Journal of Multivariate Analysis* and *Theory of Probability and Mathematical Statistics*. His research interests include high-dimensional statistics, random matrix theory, Bayesian statistics, statistical methods in finance among others. In November 2008, he was the recipient of the Young Scientists Award of the Brandenburg state for an outstanding work in the Humanities or Social Sciences.

August 2012, he was a Research Associate with the Department of Statistics, European University Viadrina, Frankfurt (Oder), Germany. From September 2012 to November 2014, he was a Research Associate with the Department of Mathematics, Humboldt-University of Berlin, Berlin, Germany. He is a Member of the Editorial Boards of *Journal of Multivariate Analysis* and *Theory of Probability and Mathematical Statistics*. His research interests include high-dimensional statistics, random matrix theory, Bayesian statistics, statistical methods in finance among others. In November 2008, he was the recipient of the Young Scientists Award of the Brandenburg state for an outstanding work in the Humanities or Social Sciences.



Solomiia Dmytryiv received the B.Sc. degree in mathematics and the M.Sc. degree in statistics from the Ivan Franko National University of Lviv, Lviv, Ukraine, in 2013 and 2014 respectively, and the Ph.D. degree in economics and statistics from the European University Viadrina, Frankfurt (Oder), Germany in 2020. Since August 2020, she has been a Postdoctoral Researcher with the Department of Statistics and Operations Research, the University of Vienna, Vienna, Austria. Her research interests include the intersection of high-dimensional statistics and random matrix theory with mathematical and statistical finance.



Yarema Okhrin received the M.S. degree in mathematics (with Hons.) from the Ivan Franko National University of Lviv, Ukraine, in 1999, and the Ph.D. degree in economics and statistics (summa cum laude) from the European University Viadrina, Frankfurt (Oder), Germany. He is Professor of Statistics with the Faculty of Business and Economics, the University of Augsburg, Germany. From April 2008 to September 2009, he was an Assistant Professor of Econometrics with the University of Bern, Switzerland. From March 2000 to March 2008, he was a Research Associate with the Department of Statistics, European University Viadrina, Frankfurt (Oder), Germany. His current research interests cover financial econometrics, dependence modelling, analysis of image, network and environmental data. He is an Editor-in-Chief of *Advances in Statistical Analysis* and was the Vice-President of the German Statistical Society from 2012 to 2020.



Nestor Parolya received the B.S. and M.S. degrees in mathematics and statistics (both with Hons.) from the Ivan Franko National University of Lviv, Ukraine, in 2009 and 2010, respectively, and the Ph.D. degree in economics/statistics (summa cum laude) from the European University Viadrina, Frankfurt (Oder), Germany. From September 2013 to April 2014, he was doing his Postdoctoral research with the Department of Statistics and Econometrics, Ruhr-University Bochum, and thereafter during 2014–2019 was appointed as an Assistant Professor (non-tenured) for financial econometrics with Institute of Statistics, Leibniz University Hannover. Since April 2019, he has been an Assistant Professor of statistics (tenure track) with the Delft Institute of Applied Mathematics, Delft University of Technology, the Netherlands. His research interests include high-dimensional statistics, large dimensional random matrix theory, statistical/mathematical finance, financial engineering and operations research.



Wolfgang Schmid received the M.S., Ph.D. and Habilitation degrees in mathematics from the University of Ulm, Germany, in 1982, 1984, and 1991, respectively. Since 1995, he has been a Full Professor in Statistics with the European University Viadrina in Frankfurt (Oder), Germany. At the moment, his main research activities are in statistical analysis of financial markets, statistical process control, and statistical analysis of environmental processes. He is the author of more than 150 publications, Joint Editor of the books *Frontiers in Statistical Quality Control Vol. 9* and *Vol. 12* and primary supervisor of 25 dissertations. Dr. Schmid is a Member of the Editorial Board of *ASTA Advances in Statistical Analysis* and of the *Journal of Multivariate Analysis* and an Associate Editor of the journal *Sequential Analysis*. He is elected as a Member of the International Statistical Institute and was the President of the German Statistical Society from 2012–2020.