



Delft University of Technology

Document Version

Final published version

Citation (APA)

Bodnar, T., Hautsch, N., Okhrin, Y., & Parolya, N. (2025). Consistent estimation of the high-dimensional efficient frontier. *European Journal of Finance*, 32(4-6), 482-509. <https://doi.org/10.1080/1351847X.2025.2505043>

Important note

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

Copyright

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership. Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse

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.

This work is downloaded from Delft University of Technology.



Consistent estimation of the high-dimensional efficient frontier

Taras Bodnar, Nikolaus Hautsch, Yarema Okhrin & Nestor Parolya

To cite this article: Taras Bodnar, Nikolaus Hautsch, Yarema Okhrin & Nestor Parolya (22 May 2025): Consistent estimation of the high-dimensional efficient frontier, The European Journal of Finance, DOI: [10.1080/1351847X.2025.2505043](https://doi.org/10.1080/1351847X.2025.2505043)

To link to this article: <https://doi.org/10.1080/1351847X.2025.2505043>



© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 22 May 2025.



Submit your article to this journal [↗](#)



Article views: 104



View related articles [↗](#)



View Crossmark data [↗](#)

Consistent estimation of the high-dimensional efficient frontier

Taras Bodnar ^a, Nikolaus Hautsch ^b, Yarema Okhrin ^c and Nestor Parolya ^d

^aDepartment of Management and Engineering, Linköping University, Linköping, Sweden; ^bDepartment of Statistics and Operations Research, University of Vienna, Vienna, Austria; ^cChair of Statistics and Data Science, Augsburg University, Augsburg, Germany; ^dDelft Institute of Applied Mathematics, Delft University of Technology, Delft, The Netherlands

ABSTRACT

In this paper, we analyze the asymptotic behavior of the main characteristics of the mean-variance efficient frontier employing random matrix theory. Our particular interest covers the case when the dimension p and the sample size n tend to infinity simultaneously and their ratio p/n tends to a positive constant $c \in (0, 1)$. We neither impose any distributional nor structural assumptions on the asset returns. For the developed theoretical framework, some regularity conditions, like the existence of the 4th moments, are needed. It is shown that two out of three quantities of interest are biased and overestimated by their sample counterparts under the high-dimensional asymptotic regime. This becomes evident based on the asymptotic deterministic equivalents of the sample plug-in estimators. Using them we construct consistent estimators of the three characteristics of the efficient frontier. Furthermore, the asymptotic normality of the considered estimators of the parameters of the efficient frontier is proved. Verifying the theoretical results based on an extensive simulation study we show that the proposed estimator for the efficient frontier is a valuable alternative to the sample estimator for high dimensional data. Finally, we present an empirical application, where we estimate the efficient frontier based on stocks from the S&P 500 index.

ARTICLE HISTORY

Received 30 September 2024
Accepted 29 April 2025

KEYWORDS

Efficient frontier;
large-dimensional
asymptotics; random matrix
theory; high-frequency
financial data

JEL CLASSIFICATIONS

G11; C13; C14; C58; C65

1. Introduction

The efficient frontier is a key object of modern portfolio theory as derived by Markowitz (1952). Merton (1972) shows that the efficient frontier is the upper part of the parabola in the mean-variance space defined by three characteristics. In particular, using Merton's notation, the efficient frontier is given by (cf. Merton 1972)

$$V = \frac{a - 2bR + cR^2}{ac - b^2}, \quad (1)$$

where the constants $a = \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, $b = \mathbf{1}'_p\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, and $c = \mathbf{1}'_p\boldsymbol{\Sigma}^{-1}\mathbf{1}_p$ fully determine the location of the vertex and the slope coefficient of the parabola in the mean-variance space. The vector $\boldsymbol{\mu}$ denotes the p -dimensional mean vector of asset returns and the $p \times p$ matrix $\boldsymbol{\Sigma}$ is the corresponding positive definite covariance matrix. The vector $\mathbf{1}_p$ is the p -dimensional unity vector.

The three parameters a , b , and c do not possess an appropriate interpretation in the financial literature. For that reason, we rewrite (1) as

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

where

$$R_{GMV} = \frac{\mathbf{1}'_p\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{1}'_p\boldsymbol{\Sigma}^{-1}\mathbf{1}_p} \quad (3)$$

CONTACT Nestor Parolya  N.Parolya@tudelft.nl

is the expected return and

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

is the variance of the global minimum variance (GMV) portfolio. The parameter

$$s = \boldsymbol{\mu}' \mathbf{Q} \boldsymbol{\mu} \quad \text{with } \mathbf{Q} = \boldsymbol{\Sigma}^{-1} - \frac{\boldsymbol{\Sigma}^{-1} \mathbf{1}_p \mathbf{1}'_p \boldsymbol{\Sigma}^{-1}}{\mathbf{1}'_p \boldsymbol{\Sigma}^{-1} \mathbf{1}_p}, \quad (5)$$

denotes the slope parameter of the parabola in the mean-variance space. The parameters $\{R_{GMV}, V_{GMV}, s\}$ are the objects of main interest in this paper.

Although the calculation of the set of parameters R_{GMV}, V_{GMV}, s is straightforward when $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are known, their computation becomes a challenging task otherwise. Since the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ are unknown in practice, they must be estimated from return data before the efficient frontier is constructed. The usual estimation technique is based on sample estimators or so-called plug-in estimators (see, e.g. Kan and Smith 2008; Okhrin and Schmid 2006; T. Bodnar and Schmid 2008, 2009) which replace $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in (3)–(5) by their sample counterparts. This estimation method yields the sample efficient frontier, specified by the sample estimators $\hat{R}_{GMV}, \hat{V}_{GMV}$, and \hat{s} of R_{GMV}, V_{GMV} , and s , respectively.

The asymptotic behavior of the sample efficient frontier for finite dimension p under the assumption of normality is investigated in Jobson and Korkie (1980) and Jobson (1991), while T. Bodnar and Schmid (2008) and Kan and Smith (2008) study the finite sample distributional properties of the sample parameters $\{\hat{R}_{GMV}, \hat{V}_{GMV}, \hat{s}\}$ of the efficient frontier. However, Basak, Jagannathan, and Ma (2005) and Siegel and Woodgate (2007) show that the sample efficient frontier overestimates the true location of the efficient frontier in the mean-variance space. To correct this overestimation, Kan and Smith (2008) derives improved estimates for the parameters of the efficient frontier. At the same time, O. Bodnar and Bodnar (2010) constructs an unbiased estimator of the efficient frontier. Finally, Bayesian estimators of the efficient frontier were derived in Bauder et al. (2019, 2021), and T. Bodnar, Lindholm, et al. (2022).

The finite-sample techniques work well only under the assumption of normally distributed asset returns, which is not supported by in most applications. Typically, asset returns are skewed and heavy-tailed (e.g. C. Adcock 2014; C. Adcock, Eling, and Loperfido 2015; Fama 1965; Markowitz 1991). T. Bodnar and Gupta (2009) and Gupta, Varga, and Bodnar (2013) construct the efficient frontier under elliptical models and provide inference procedures for its main parameters.

More challenging problems arise by considering the asymptotic properties of the sample efficient frontier when the dimension of the portfolio p increases together with the sample size n . It is noted that the sample estimators work well only if the number of assets p is fixed and significantly smaller than the number of observations n . This case is often used in statistics and is referred to the case of standard asymptotics (see, Le Cam and Lo Yang 2000). In this regime, the traditional sample estimators are consistent for the main parameters of the efficient frontier. As a result, for a small fixed dimension $p \in \{2, 3, 5\}$ with a large n the sample estimator can be used. However, it is unclear what to do if the number of assets in the portfolio are (very) large comparable to n . Here, we are in a situation when both the number of assets p and the sample size n tend to infinity. This double asymptotic has an interpretation when p and n are of comparable size, i.e. when p/n tends to a concentration ratio $c > 0$.¹ This regime of asymptotics is known as a high-dimensional asymptotic or ‘Kolmogorov’ asymptotic (see, e.g. Bühlmann and van de Geer 2011; T. Bodnar, Dette, and Parolya 2019; T. Cai and Shen 2011). In this setting, the sample estimators behave unpredictably and are far from the optimal ones. In general, the larger the concentration ratio c becomes, the worse is the performance of the sample estimators. This is a well-known problem in statistics called ‘the curse of dimensionality’ (see, e.g. Bellman 1961).

Under the presence of financial high-frequency (intraday) data, the curse of dimensionality can be reduced by employing in-fill asymptotics, however, it requires to handle the impact of market microstructure noise (and potentially ill-conditioned estimates). One of the first applications of high-frequency data to estimate huge-dimensional covariance matrices is Hautsch, Kyj, and Oomen (2012) who propose a blocking and regularization

approach. Hautsch, Kyj, and Malec (2015) further extend this idea and show its statistical and economic usefulness in a global minimum variance framework. Recently, De Nard et al. (2022) generalize the multivariate GARCH process by incorporating intraday data in the model specification, while Golosnoy, Gribisch, and Seifert (2022) and Golosnoy and Gribisch (2022) apply the realized covariance matrix to construct realized global minimum variance portfolio.

To handle the curse of dimensionality for the efficient frontier, we employ results from asymptotic random matrix theory. Random matrix theory is a fast growing branch of probability theory with many applications in statistics and finance. It studies the asymptotic behavior of the eigenvalues of random matrices under general asymptotics (see, e.g. Bai and Silverstein 2010; G. Anderson, Guionnet, and Zeitouni 2010). The asymptotic behavior of the functionals of the sample covariance matrices is analyzed in Marčenko and Pastur (1967), Yin (1986), Girko and Gupta (1994, 1996a, 1996b), Silverstein (1995), Bai, Miao, and Pan (2007), Bai and Silverstein (2010), Rubio and Mestre (2011), T. Bodnar, Gupta, and Parolya (2014), and T. Bodnar, Gupta, and Parolya (2016), among others. Marčenko and Pastur (1967) formally derive the behavior of the limiting spectral measure of the sample covariance matrix. It depends on the corresponding spectral measure of the population covariance matrix, which is unknown. Silverstein (1995) proves the validity of the Marčenko and Pastur (MP) equation under more general assumptions and show the strong convergence of the spectral measure of the sample covariance matrix. Recently, high-dimensional optimal portfolio theory has attracted the attention of researchers and practitioners of the financial sector (see, e.g. Ao, Yingying, and Zheng 2019; Ding, Li, and Zheng 2021; El Karoui 2010; Glombek 2014; Jagannathan and Ma 2003; Kan and Wang 2024; Lassance, Vanderveken, and Vrins 2024; T. Bodnar et al. 2019, 2021, 2025; T. Bodnar, Dette, et al. 2022; T. Bodnar, Okhrin, and Parolya 2023; T. Bodnar, Parolya, and Schmid 2018; T. Bodnar, Parolya, and Thorsén 2024; T. T. Cai et al. 2020). We use the above-mentioned theoretical results to find the asymptotic behavior of the sample efficient frontier $\{\hat{R}_{GMV}, \hat{V}_{GMV}, \hat{s}\}$ and to prove that \hat{R}_{GMV} is indeed a consistent estimator for R_{GMV} , while \hat{V}_{GMV} and \hat{s} are highly biased and inconsistent in high dimensions. We show that the additive and multiplicative biases are solely functions of the concentration ratio c , thus we can construct consistent estimators handling these biases. No distributional neither structural assumptions are imposed on the asset returns. Theoretically, we only need the existence of the 4th moments. Moreover, under additional assumptions imposed on the distribution of the asset returns we prove that the consistent estimators of the three parameters of the efficient frontier are mutually independent and asymptotically normally distributed under the high-dimensional asymptotic regime. Our findings confirm that the Bayesian estimator of the efficient frontier, as derived in Bauder et al. (2021), is indeed a consistent estimator from a frequentist perspective.

The rest of the paper is organized as follows. In Section 2, we provide the main assumptions and notations used in the paper and formulate the main theoretical result on the asymptotic behavior of the efficient frontier. At the end of Section 2, consistent estimators of the three parameters of the efficient frontier are presented, whereas the results on the asymptotic normality are summarized in Section 2.1. In Section 3, we provide an extensive simulation study to verify the rate of convergence as well as the performance of the obtained estimators under high-dimensional asymptotics. Here, the performance of the derived estimator is compared with existent estimation techniques for high-dimensional data. The results of the empirical study are presented in Section 4. Section 5 provides concluding remarks. The proofs of the theoretical results are presented in the appendix (Appendix).

2. Consistent estimation of the efficient frontier under large dimensional asymptotics

In this section, we present our main result on the asymptotics of the main characteristics of the efficient frontier.

Let the $p \times n$ matrix \mathbf{Y}_n be the observation matrix of the p -dimensional vector of the asset returns taken at time points $1, \dots, n$. The vector of the expected asset returns is denoted by $\boldsymbol{\mu}_n$, while $\boldsymbol{\Sigma}_n$ denotes the covariance matrix.² We assume that the observation matrix is equal in distribution to

$$\mathbf{Y}_n \stackrel{d.}{=} \boldsymbol{\Sigma}_n^{\frac{1}{2}} \mathbf{X}_n + \boldsymbol{\mu}_n \mathbf{1}'_n, \quad (6)$$

where the symbol ' $\stackrel{d.}{=}$ ' denotes the equality in distribution and the $p \times n$ matrix \mathbf{X}_n contains independent and identically distributed (i.i.d.) real random variables with zero mean and unit variance. Note that only the matrix \mathbf{Y}_n is observable. We know neither \mathbf{X}_n , $\boldsymbol{\Sigma}_n$ nor $\boldsymbol{\mu}_n$.

Note that the observation matrix \mathbf{Y}_n has dependent rows, which is assured by the covariance matrix $\boldsymbol{\Sigma}_n$, but independent columns. The assumption of the independence of the samples can be weakened to allow for dependent elements of \mathbf{X}_n by controlling the growth of the number of dependent entries but not their joint distribution (see Friesen, Löwe, and Stolz 2013). Here, we stick to the assumption of independence of the random samples to simplify the proof of the theoretical results.

The sample mean vector of the asset returns is given by

$$\bar{\mathbf{y}}_n = \frac{1}{n} \mathbf{Y}_n \mathbf{1}_n, \quad (7)$$

whereas the sample covariance matrix is defined as

$$\mathbf{S}_n = \frac{1}{n} (\mathbf{Y}_n - \bar{\mathbf{y}}_n \mathbf{1}'_n) (\mathbf{Y}_n - \bar{\mathbf{y}}_n \mathbf{1}'_n)' = \frac{1}{n} \mathbf{Y}_n \mathbf{Y}'_n - \bar{\mathbf{y}}_n \bar{\mathbf{y}}'_n. \quad (8)$$

Using (7) and (8) we obtain the sample estimators for the three parameters of the efficient frontier given by

$$\hat{R}_{GMV} = \frac{\mathbf{1}'_p \mathbf{S}_n^{-1} \bar{\mathbf{y}}_n}{\mathbf{1}'_p \mathbf{S}_n^{-1} \mathbf{1}_p}, \quad \hat{V}_{GMV} = \frac{1}{\mathbf{1}'_p \mathbf{S}_n^{-1} \mathbf{1}_p}, \quad (9)$$

and

$$\hat{s} = \bar{\mathbf{y}}'_n \hat{\mathbf{Q}} \bar{\mathbf{y}}_n \quad \text{with} \quad \hat{\mathbf{Q}} = \mathbf{S}_n^{-1} - \frac{\mathbf{S}_n^{-1} \mathbf{1}_p \mathbf{1}'_p \mathbf{S}_n^{-1}}{\mathbf{1}'_p \mathbf{S}_n^{-1} \mathbf{1}_p}. \quad (10)$$

Our theoretical findings are based on the following assumptions:

- (A1) The population covariance matrix $\boldsymbol{\Sigma}_n$ is a nonrandom p -dimensional positive definite matrix for all dimensions p .
- (A2) The elements of the matrix \mathbf{X}_n have uniformly bounded $4 + \varepsilon$, $\varepsilon > 0$ moments.
- (A3) It exists $M_l, M_u \in (0, +\infty)$ and $q \in [0, +\infty)$ such that

$$M_l p^q \leq \mathbf{1}'_p \boldsymbol{\Sigma}_n^{-1} \mathbf{1}_p, \boldsymbol{\mu}'_n \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_n \leq M_u p^q.$$

All of these regularity assumptions are very general and fit many practical situations. Assumption (A1) together with (6) is typical for financial and statistical problems and does not impose strong restrictions. Assumption (A2) is a technical one. Our simulation study shows that this assumption can be relaxed for practical purposes. Assumption (A3) requires that the quantities used in the calculation of R_{GMV} , V_{GMV} , and s are of the same order. This assumption is very general and imposes no additional constraints neither on the mean vector $\boldsymbol{\mu}_n$ (such that its Euclidean norm is bounded) nor on the covariance matrix $\boldsymbol{\Sigma}_n$ (such that its eigenvalues lie in the compact interval). The last point allows us to assume a factor model for the data matrix \mathbf{Y}_n which implies that the largest eigenvalue of $\boldsymbol{\Sigma}_n$ is of order p (c.f. Fan, Fan, and Lv 2008; Fan, Liao, and Mincheva 2013; Fan, Zhang, and Yu 2012). Finally, assumption (A3) ensures that $\mathbf{1}'_p \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_n$ is at most of order p^q which follows directly from the Cauchy-Schwarz inequality.

The main result about the strong convergence of the sample estimators for the three parameters of the efficient frontier is presented in Theorem 2.1.

Theorem 2.1: Let $q \geq 0$. Then, under the assumptions (A1)–(A3) it holds that

$$\left| \hat{R}_{GMV} - R_{GMV} \right| \xrightarrow{a.s.} 0 \quad \text{for } p/n \rightarrow c \in (0, 1) \text{ as } n \rightarrow \infty, \quad (11)$$

$$p^q \left| \hat{V}_{GMV} - (1 - c)V_{GMV} \right| \xrightarrow{a.s.} 0 \quad \text{for } p/n \rightarrow c \in (0, 1) \text{ as } n \rightarrow \infty, \quad (12)$$

$$p^{-q} \left| \hat{s} - \frac{1}{1-c}s - \frac{p}{n} \mathbb{1}_{\mathcal{K}_{\{0\}}}(q) \right| \xrightarrow{a.s.} 0 \quad \text{for } p/n \rightarrow c \in (0, 1) \text{ as } n \rightarrow \infty, \quad (13)$$

where $\mathbb{1}_{\mathcal{A}}(\cdot)$ denotes the indicator function of set \mathcal{A} .

The proof of Theorem 2.1 is given in the appendix (Appendix). Theorem 2.1 establishes the asymptotic behavior of the main characteristics of the efficient frontier as both the dimension and the sample size grow to infinity simultaneously. It claims that the sample estimator of the expected return of the GMV portfolio \hat{R}_{GMV} is indeed a consistent estimator under the high-dimensional asymptotic regime. In contrast, the sample estimators of the other efficient frontier parameters are inconsistent. The estimator for the sample variance of the GMV portfolio, \hat{V}_{GMV} , possesses a multiplicative bias $\frac{1}{1-c}$, whereas the sample estimator of the slope parameter s has a multiplicative bias $(1 - c)$. Consequently, if the concentration ratio c is close to one, both the estimators produce a high bias implying inconsistency.

Fortunately, these biases can be easily handled using Theorem 2.1. Hence, consistent estimators \hat{R}_c , \hat{V}_c and \hat{s}_c for the three parameters of the efficient frontier are presented in Corollary 2.1. This result follows directly from Theorem 2.1.

Corollary 2.1: Under assumptions (A1)–(A3), consistent estimators of R_{GMV} , V_{GMV} and s are given by

$$\hat{R}_c = \hat{R}_{GMV}, \quad (14)$$

$$\hat{V}_c = \frac{1}{1 - p/n} \hat{V}_{GMV}, \quad (15)$$

$$\hat{s}_c = (1 - p/n)\hat{s} - \frac{p}{n} \mathbb{1}_{\mathcal{K}_{\{0\}}}(q). \quad (16)$$

Assuming that the asset returns are normally distributed and the number of assets is fixed, i.e. $c = 0$, O. Bodnar and Bodnar (2010) derive unbiased estimators for the three parameters of the efficient frontier given by

$$\hat{R}_u = \hat{R}_{GMV}, \quad \hat{V}_u = \frac{n-1}{n-p} \hat{V}_{GMV}, \quad \text{and} \quad \hat{s}_u = \frac{n-p-1}{n-1} \hat{s} - \frac{p-1}{n}, \quad (17)$$

which are asymptotically equivalent to \hat{R}_c , \hat{V}_c , and \hat{s}_c for $q = 0$.

Returning to the equivalent Merton notation of the efficient frontier, we present in Corollary 2.2 the corresponding consistent estimators of the parameters \mathbf{a} , \mathbf{b} , and \mathbf{c} given in (1).

Corollary 2.2: Under the assumptions (A1)–(A3) the consistent estimators of the Merton's constants \mathbf{a} , \mathbf{b} and \mathbf{c} are given by

$$\hat{\mathbf{a}}_c = (1 - p/n)\hat{\mathbf{a}}, \quad (18)$$

$$\hat{\mathbf{b}}_c = (1 - p/n)\hat{\mathbf{b}}, \quad (19)$$

$$\hat{\mathbf{c}}_c = (1 - p/n)\hat{\mathbf{c}}, \quad (20)$$

where $\hat{\mathbf{a}} = \bar{\mathbf{y}}_n' \mathbf{S}_n^{-1} \bar{\mathbf{y}}_n$, $\hat{\mathbf{b}} = \mathbf{1}_p' \mathbf{S}_n^{-1} \bar{\mathbf{y}}_n$, and $\hat{\mathbf{c}} = \mathbf{1}_p' \mathbf{S}_n^{-1} \mathbf{1}_p$ are the corresponding sample counterparts.

The proof of Corollary 2.2 follows directly from the proof of Theorem 2.1 and Corollary 2.1. Both corollaries provide equivalent results. Nevertheless, the parametrization discussed in Corollary 2.1 is more tractable. It is easy to deduce that under the standard asymptotics $p/n \rightarrow 0$ the consistent estimators of the parameters of the efficient frontier coincide with their sample counterparts.

The obtained results make the practical implementation of the Markowitz portfolio analysis feasible. They ensure the consistency of the estimators in the case of high-dimensional portfolios if the concentration ratio $c = p/n$ does not exceed 1. In the case $c > 1$, the sample covariance matrix is singular. Thus, it is more complicated to find consistent estimators for the parameters of the efficient frontier. In this case, the naive choice would be to replace the inverse sample covariance matrix by its pseudo-inverse, but the consistency of such replacement is not obvious. We leave this question for future research.

2.1. Asymptotic normality

Below we prove that the consistent estimators of the three parameters of the efficient frontier are asymptotically normally distributed under the high-dimensional asymptotic regime. These results are formulated as Theorem 2.2. Note that only for the proof of this theorem we impose an additional assumption on the distribution of the entries of \mathbf{X}_n , which are assumed to be standard normally distributed.

Theorem 2.2: *Let the assumptions (A1)–(A3) are fulfilled. If $q = 0$ and $x_{ij} \sim \mathcal{N}(0, 1)$, then it holds that*

$$\sqrt{n} \begin{pmatrix} \hat{R}_c - R_{GMV} \\ \hat{V}_c - V_{GMV} \\ \hat{s}_c - s \end{pmatrix} \xrightarrow{d} \mathcal{N} \left(\mathbf{0}, \begin{pmatrix} \left(1 + \frac{s+c}{1-c}\right) V_{GMV} & 0 & 0 \\ 0 & \frac{2V_{GMV}^2}{1-c} & 0 \\ 0 & 0 & \sigma_s^2 \end{pmatrix} \right),$$

where

$$\sigma_s^2 = 2(c + 2s) + 2 \frac{(c+s)^2}{1-c} \quad (21)$$

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

The findings of Theorem 2.2 show that the consistent estimators for the three parameters of the efficient frontier are independent under high-dimensional asymptotics. We observe that their asymptotic variances increase as c approaches 1. On the other side, interesting results are obtained in the special case $c = 0$. Here, it holds that

$$\sqrt{n} \begin{pmatrix} \hat{R}_c - R_{GMV} \\ \hat{V}_c - V_{GMV} \\ \hat{s}_c - s \end{pmatrix} \xrightarrow{d} \mathcal{N} \left(\mathbf{0}, \begin{pmatrix} (1+s) V_{GMV} & 0 & 0 \\ 0 & 2V_{GMV}^2 & 0 \\ 0 & 0 & 4s + 2s^2 \end{pmatrix} \right),$$

which coincides with the findings obtained for finite p under the standard asymptotic regime in T. Bodnar and Schmid (2009). Comparing the above asymptotic distribution with those presented in Theorem 2.2, we conclude that ignoring the high-dimensional effect in the calculation of the asymptotic distribution yields a considerable underestimation of the uncertainties of \hat{V}_c and \hat{s}_c , especially for values of c closed to 1. Finally, we also observe that an additional additive bias is present in the sample estimator for s when the quadratic forms $\mathbf{1}_p' \Sigma_n^{-1} \mathbf{1}_p$ and $\mu_n' \Sigma_n^{-1} \mu_n$ are bounded, i.e. in case of $q = 0$.

Finally, application of Slutsky's Lemma (cf., Theorem 1.5 in DasGupta 2008) leads to the marginal confidence intervals for R_{GMV} , V_{GMV} , and s given by

$$I_{1-\alpha}(R_{GMV}) = \left[\hat{R}_c - \frac{z_{1-\alpha/2}}{\sqrt{n}} \sqrt{1 + \frac{\hat{\delta}_c + c}{1-c}} \sqrt{\hat{V}_c}, \hat{R}_c + \frac{z_{1-\alpha/2}}{\sqrt{n}} \sqrt{1 + \frac{\hat{\delta}_c + c}{1-c}} \sqrt{\hat{V}_c} \right],$$

$$I_{1-\alpha}(V_{GMV}) = \left[\hat{V}_c - \frac{z_{1-\alpha/2}}{\sqrt{n}} \sqrt{\frac{2}{1-c}} \hat{V}_c, \hat{V}_c + \frac{z_{1-\alpha/2}}{\sqrt{n}} \sqrt{\frac{2}{1-c}} \hat{V}_c \right],$$

$$I_{1-\alpha}(s) = \left[\hat{\delta}_c - \frac{z_{1-\alpha/2}}{\sqrt{n}} \sqrt{2(c + 2\hat{\delta}_c) + 2\frac{(c + \hat{\delta}_c)^2}{1-c}}, \hat{\delta}_c + \frac{z_{1-\alpha/2}}{\sqrt{n}} \sqrt{2(c + 2\hat{\delta}_c) + 2\frac{(c + \hat{\delta}_c)^2}{1-c}} \right],$$

where z_β denotes the β -quantile of the standard normal distribution.

3. Finite sample performance

3.1. Estimation of R_{GMV} , V_{GMV} , and s

In this section, we examine the rate of convergence and the performance of the derived consistent estimators given in Corollary 2.1. In order to analyze the large sample performance of the estimators, we use a quadratic loss function. Without loss of generality, in our simulation study we take the covariance matrix Σ_n as given with the same proportion of the eigenvalues over all dimensions, namely 20% of eigenvalues are equal to 0.5, 40% to 1 and 40% to 5, respectively. This guarantees to keep the spectrum of the covariance matrix unchanged for all p . The elements of the mean vector μ_n are generated from the uniform distribution, i.e. $\mu_n^i \sim \text{Unif}[-0.2, 0.2]$ for $i \in \{1, \dots, n\}$.

Three scenarios are considered in the simulation study, namely

Scenario 1 : The entries of \mathbf{X}_n are generated independently from $\mathcal{N}(0, 1)$.

Scenario 2 : The entries of \mathbf{X}_n are generated independently from $t_3(0, 1/3)$ (t -distribution with 3 degrees of freedom and scale parameter 1/3). This choice of the scale parameter ensures that the covariance matrix of each column of \mathbf{Y}_n is equal to Σ_n .

Scenario 3 : The columns of \mathbf{Y}_n are assumed to follow a CCC-GARCH(1,1) process with correlation matrix Ω_n computed from Σ_n and conditional variance equations expressed as

$$h_{i,t} = \alpha_{0,i} + \alpha_{1,i}(Y_n^{i,t-1} - \mu_n^i)^2 + \beta_{1,i}h_{i,t-1} \quad \text{for } i = 1, \dots, p \quad \text{and } t = 1, \dots, n$$

with $\alpha_{0,i} > 0$ and $\alpha_{1,i}, \beta_{1,i} \in [0, 1)$ such that $\alpha_{1,i} + \beta_{1,i} < 1$ for all $i = 1, \dots, p$. The parameters $\alpha_{1,i}$ are generated from $\text{Unif}[0, 0.1]$, while the parameters $\beta_{1,i}$ are simulated from $\text{Unif}[0.8, 0.89]$. Finally, the values $\alpha_{0,i}$ are calculated using Ω_n , $\alpha_{1,i}$ and $\beta_{1,i}$ ensuring that unconditional covariance matrix of the simulated CCC-GARCH(1,1) process is equal to Σ_n .

The model in Scenario 1 satisfies the conditions (A1)–(A3) and is used to study the speed of the convergence of the obtained theoretical results. The stochastic models from Scenarios 2 and 3 are designed in such a way, that at least one of the assumptions (A1)–(A3) is violated. In particular, in case of Scenario 2, the moments of order $(4 + \varepsilon)$ do not exist. For the model in Scenario 3, we additionally assume dependence between the columns of matrix \mathbf{Y}_n . These two scenarios are used in the simulation study in order to investigate the robustness of the obtained theoretical results to the violation of the imposed assumptions. The results of the simulation study are based on 1000 independent repetitions performed for $c \in \{0.5, 0.9\}$.

In Figure 1, we present the results for the estimator \hat{R}_c . The solid line denotes the average quadratic loss while the dotted lines present the 0.95 and 0.05 quantiles of the quadratic loss. In the top of the figure, the results for the normal distribution are presented, whereas findings for the t -distribution and the CCC-GARCH model are

given in the middle and in the bottom, correspondingly. In case of the normal distribution, we observe a fast convergence speed of the estimator for moderate values of $c = 0.5$. The variance of the quadratic loss is small and vanishes as the dimension increases. In the case of very noisy data, namely $c = 0.9$, the estimator \hat{R}_c also reveals a fast convergence rate in the average quadratic loss. However, as it can be seen from the figure, this estimator is quite noisy. Similar results are observed for the t -distribution and the CCC-GARCH process, where only the convergence rate is little slower than for the normal distribution.

Figure 2 contains the simulation results for the estimator \hat{V}_c , which is a consistent estimator for V_{GMV} . We observe a similar behavior independently of the scenario used to simulate the data. As expected, better results are presented in the case of the normal distribution, whereas the worst ones correspond to the CCC-GARCH process, although the difference is not large. The convergence rate is relatively small for $p < 200$ in the case of $c = 0.5$ and for $p < 500$ in the case of $c = 0.9$. If p increases the quadratic loss tends to zero.

The results obtained for \hat{s}_c in Figure 3 are stronger as those obtained for \hat{R}_c and \hat{V}_c . In the case of $c = 0.5$, the quadratic loss function tends to zero already for small values of p , while for $c = 0.9$ larger values of p are needed. Interestingly, the best results are present in the case of the CCC-GARCH model, where the convergence rate is even larger than in case of the normal distribution.

In Figures 4–6 we plot the histograms of $\sqrt{n}(\hat{R}_c - R_{GMV})$, $\sqrt{n}(\hat{V}_c - V_{GMV})$, and $\sqrt{n}(\hat{s}_c - s)$ as well as their asymptotic densities presented in Theorem 2.2. Although the results of Theorem 2.2 are derived under the assumption of normality, in Figures 4–6 we also include the results for the t -distribution and the CCC-GARCH process. In Figure 4, we observe that the densities of the normal distribution provide very good approximations of the corresponding histograms independently of the considered scenario. Furthermore, as expected, the variance in the case of $c = 0.9$ is much larger than in the case of $c = 0.5$. In contrast to the findings of Figure 4, a good approximation by the normal distribution is present only in the case of the normal distribution. In the case of $\sqrt{n}(\hat{V}_c - V_{GMV})$, the histograms are shifted slightly to the left for the t -distribution and to the right for the CCC-GARCH model (see Figure 6). Similar results are also present in Figure 6 for $\sqrt{n}(\hat{s}_c - s)$. In particular, a good approximation is observed in the case of a normal distribution, whereas the histogram is shifted to the right in the case of the t -distribution and to the left in the case of the CCC-GARCH process.

To summarize the simulation results, we observe fast convergence of the proposed estimators for the three parameters of the efficient frontier in the average quadratic loss. The convergence speed is controlled by the existence of the moments of the distribution, i.e. the heavier its tails the slower the convergence rate. It is noted that the estimators \hat{R}_c and \hat{s}_c can still be improved by applying a shrinkage estimator for the mean vector. A suitable shrinkage can significantly reduce the noise of the considered estimators for the parameters of the efficient frontier and optimal portfolios (see, e.g. Golosnoy and Okhrin 2007; Kan and Smith 2008; O. Bodnar, Bodnar, and Parolya 2022; T. Bodnar, Parolya, and Thorsén 2023).

3.2. Estimation of the efficient frontier

To further investigate the performance of the obtained estimators, we compare them with state-of-art estimators used for the estimation of the inverse covariance matrix, namely the nonlinear shrinkage estimator of Ledoit and Wolf (2017, 2020), the empirical Bayes estimator of Efron and Morris (1976), and the ridge-type estimator suggested by Kubokawa and Srivastava (2008).

The empirical Bayes estimator is obtained by

$$\hat{\Pi}_{EBE} = \frac{n-p-2}{n-1} \mathbf{S}_n^{-1} + \frac{p^2+p-2}{(n-1)\text{tr}(\mathbf{S}_n)} \mathbf{I}_p, \quad (22)$$

while the ridge-type of Kubokawa and Srivastava (2008) is computed as

$$\hat{\Pi}_{RTE} = p \left((n-1)\mathbf{S}_n + \text{tr}(\mathbf{S}_n)\mathbf{I}_p \right)^{-1}, \quad (23)$$

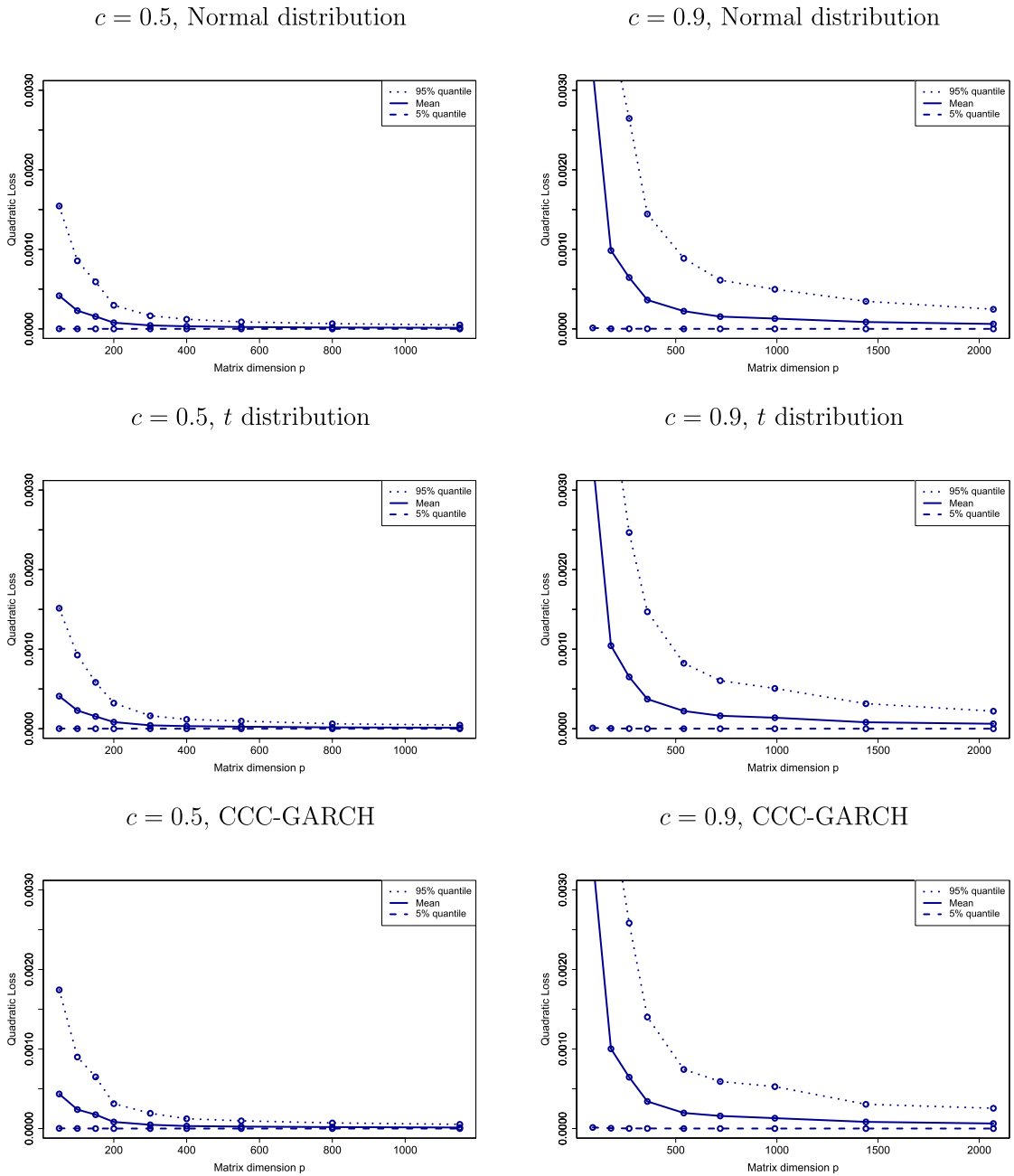
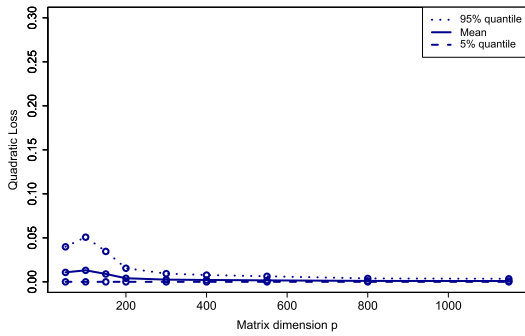


Figure 1. The quadratic loss of \hat{R}_c for the normal distribution (above), for the t -distribution with 3 degrees of freedom (in the middle), and for the CCC-GARCH process (below).

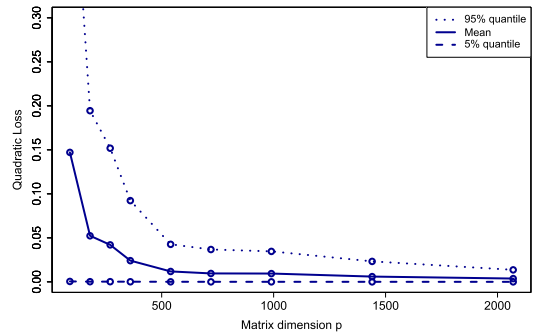
where \mathbf{I}_p is the p -dimensional identity matrix. Finally, the nonlinear shrinkage estimator for the covariance matrix of Ledoit and Wolf (2020) is given by

$$\hat{\mathbf{\Pi}}_{LWE} = \mathbf{H} \left(\text{diag}(d_1^{or}, \dots, d_p^{or}) \right)^{-1} \mathbf{H}^\top,$$

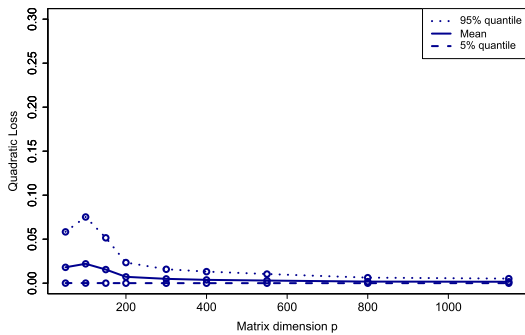
$c = 0.5$, Normal distribution



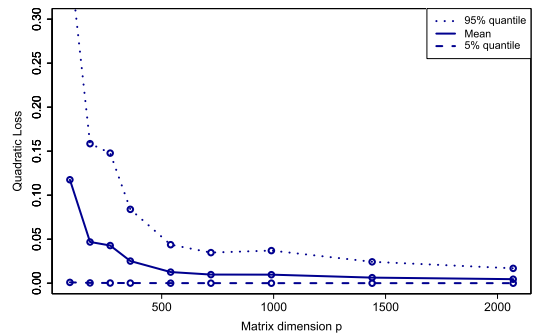
$c = 0.9$, Normal distribution



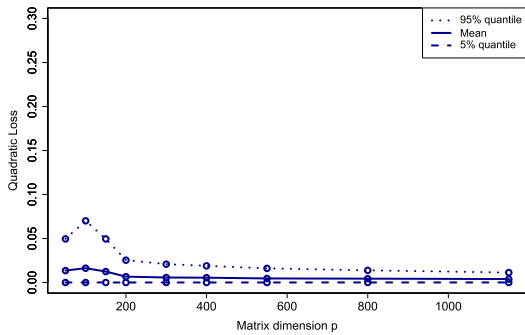
$c = 0.5$, t distribution



$c = 0.9$, t distribution



$c = 0.5$, CCC-GARCH



$c = 0.9$, CCC-GARCH

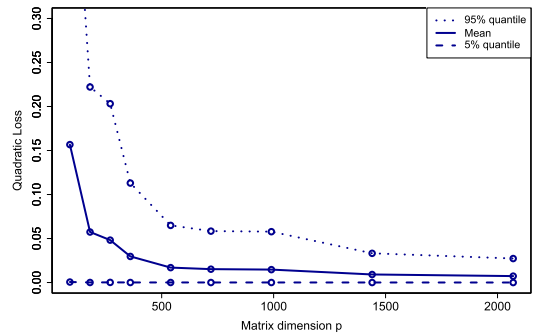


Figure 2. The quadratic loss of \hat{V}_c for the normal distribution (above), for the t -distribution with 3 degrees of freedom (in the middle), and for the CCC-GARCH process (below).

$$d_i^{or} = \begin{cases} \frac{d_i}{|1 - p/n - p/nd_i \check{m}_F(d_i)|^2}, & \text{if } d_i > 0, \\ \frac{1}{(p/n - 1) \check{m}_F(0)}, & \text{if } d_i = 0, \end{cases} \quad \text{for } i \in \{1, \dots, p\}, \quad (24)$$

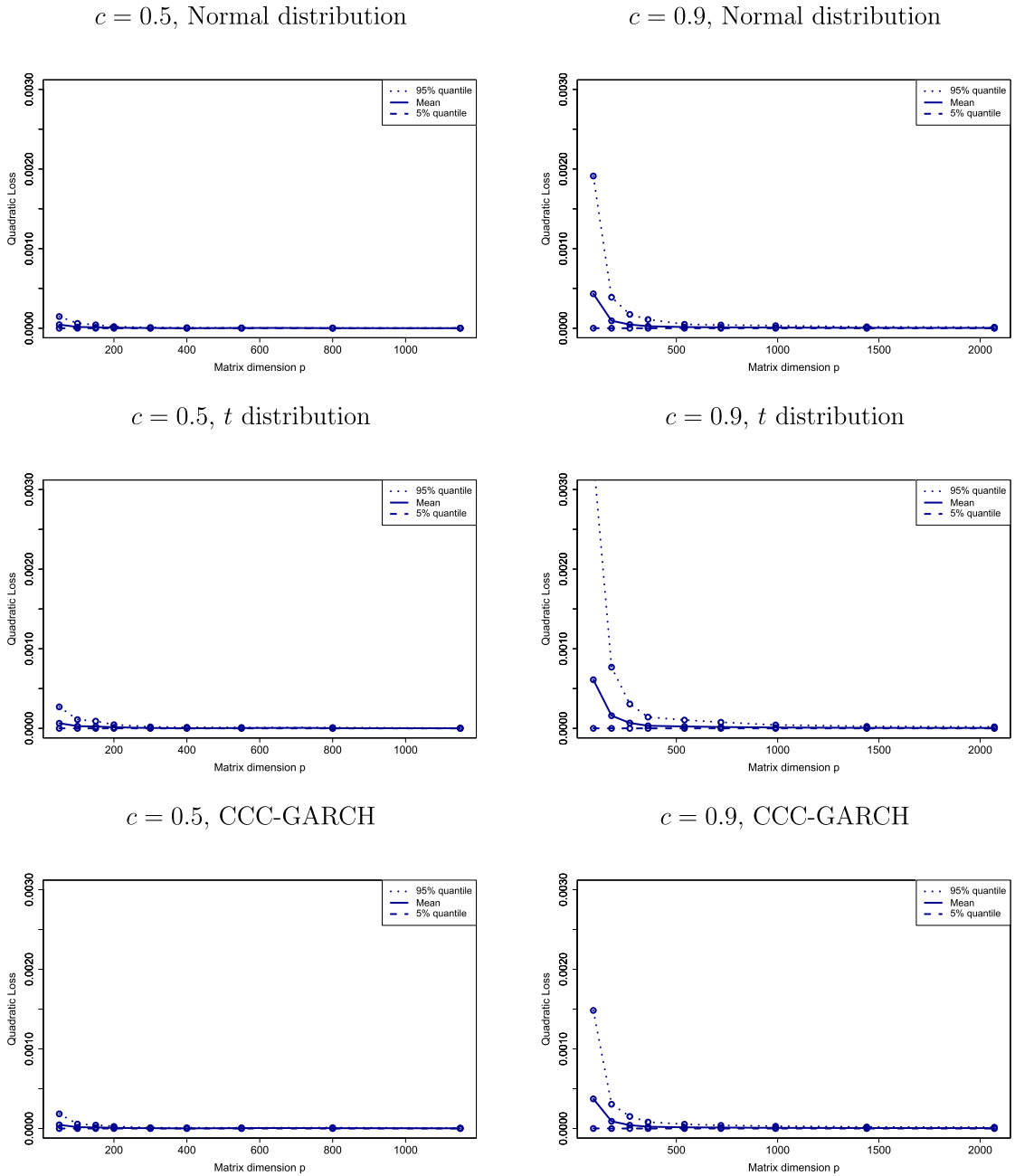


Figure 3. The quadratic loss of \hat{s}_c for the normal distribution (above), for the t -distribution with 3 degrees of freedom (in the middle), and for the CCC-GARCH process (below).

where $m_F(z)$ is the limiting Stieltjes transform of the sample covariance matrix \mathbf{S}_n which for a distribution function $G : \mathbb{R} \rightarrow \mathbb{R}$ is defined by

$$m_G(z) = \int_{-\infty}^{+\infty} \frac{1}{\lambda - z} dG(\lambda); \quad z \in \mathbb{C}^+ \equiv \{z \in \mathbb{C} : \Im z > 0\}.$$

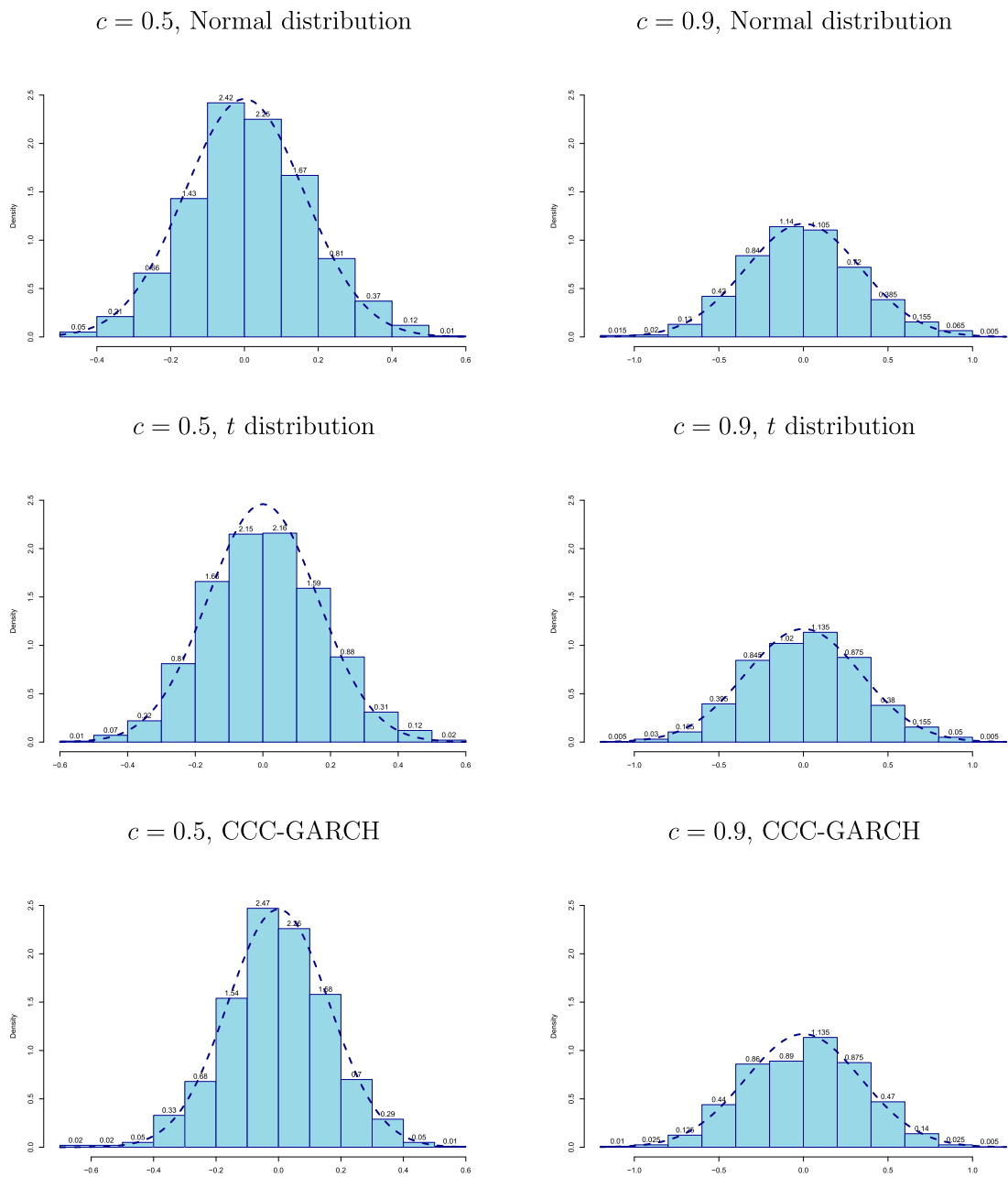


Figure 4. Histogram and the asymptotic density (see Theorem 2.2) of $\sqrt{n}(\hat{R}_c - R_{GMV})$ for the normal distribution (above), for the t -distribution with 3 degrees of freedom (in the middle), and for the CCC-GARCH process (below). We put $n = 1000$.

The estimator of the mean vector is left unchanged and corresponds to the sample. Using the estimators $\hat{\Pi}_{LWE}$ ($\hat{\Pi}_{EBE}$ or $\hat{\Pi}_{RTE}$) and \hat{y}_n in (3)–(5) instead of Σ_n^{-1} and μ_n , respectively, we obtain the nonlinear shrinkage, the empirical Bayes, and the ridge-type estimators of the parameters of the efficient frontier.

In Figure 7, we consider the behavior of the whole efficient frontier by comparing the proposed estimator with the three benchmarks presented above. Moreover, we also include the population efficient frontier and

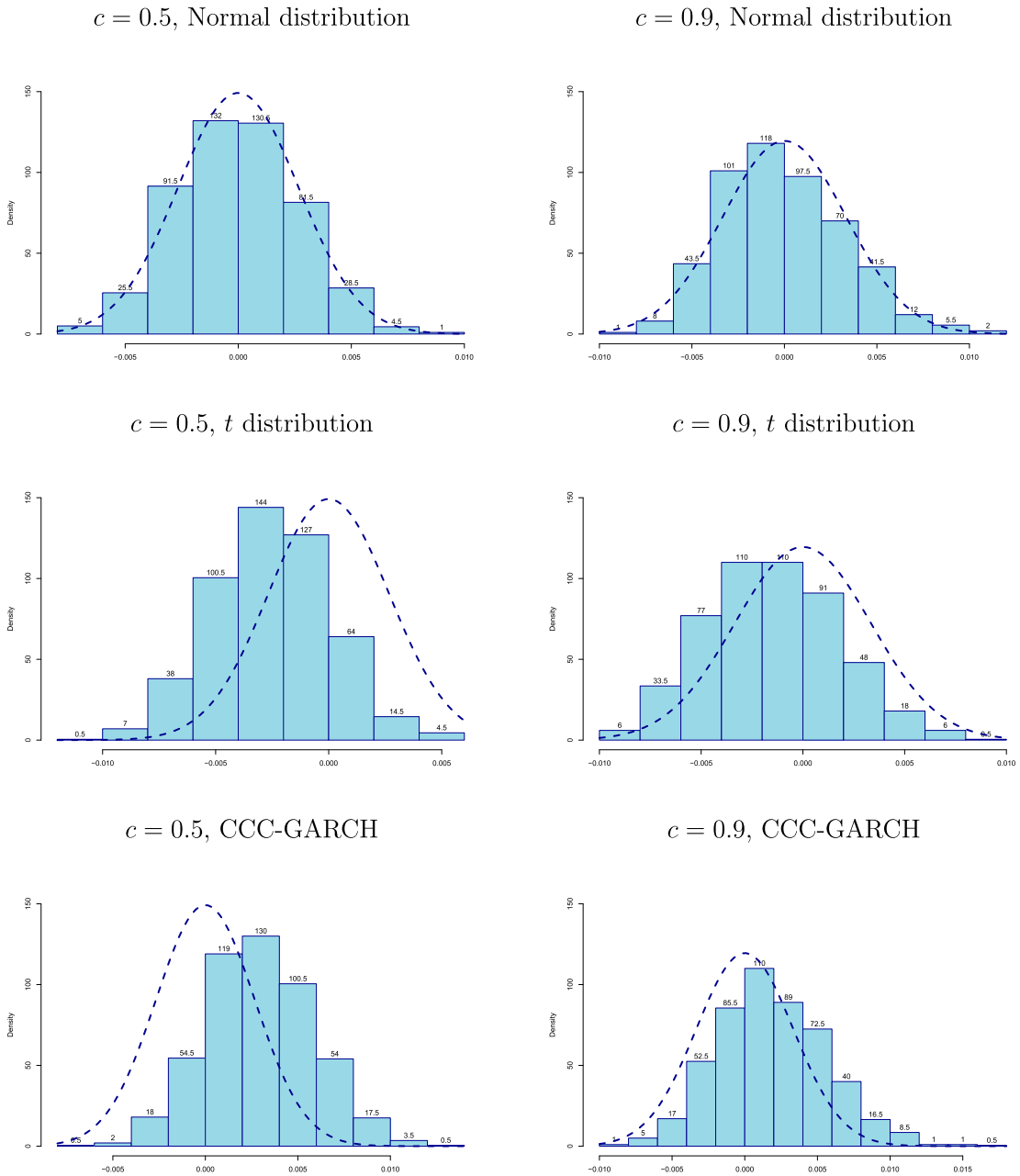


Figure 5. Histogram and the asymptotic density (see Theorem 2.2) of $\sqrt{n}(\hat{V}_c - V_{GMV})$ for the normal distribution (above), for the t -distribution with 3 degrees of freedom (in the middle), and for the CCC-GARCH process (below). We put $n = 1000$.

its sample estimator. The results are presented for all three scenarios considered in Section 3.1 in the case of $c \in \{0.5, 0.9\}$ and $n = 100$.

As expected, we observe that the sample estimator performs very poorly in all of the considered cases. It considerably overestimates the location of the true efficient frontier in the mean-variance space. In contrast, the consistent estimator of the efficient frontier shows a very good performance in almost all of the considered cases.

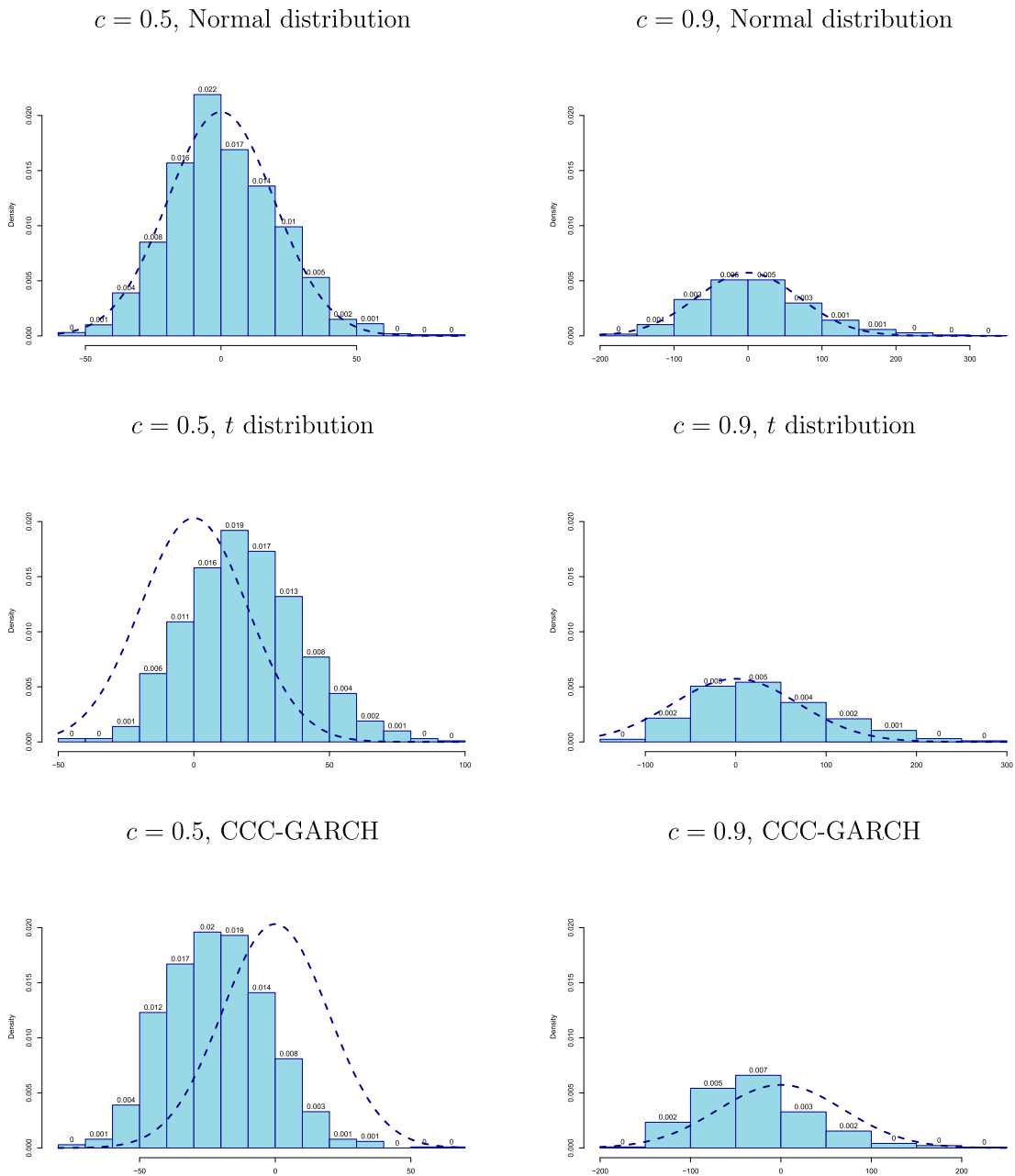
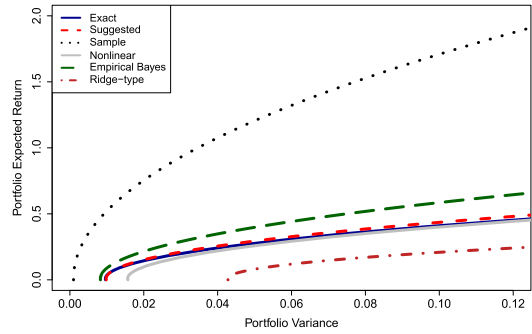
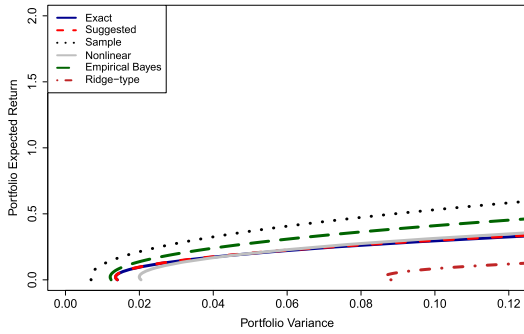


Figure 6. Histogram and the asymptotic density (see Theorem 2.2) of $\sqrt{n}(\hat{s}_c - s)$ for the normal distribution (above), for the t -distribution with 3 degrees of freedom (in the middle), and for the CCC-GARCH process (below). We put $n = 1000$.

It nearly coincides with the population efficient frontier in the case of normally distributed data for both values of c . In the case of the t -distribution, it slightly overestimates the location of the efficient frontier. Remarkably, the deviations from the population efficient frontier are very small in all of the considered cases even for $c = 0.9$. The nonlinear shrinkage estimator performs similarly to the consistent estimator proposed in the paper. However, it significantly overestimates the variance of the GMV portfolio in all of the considered cases leading to

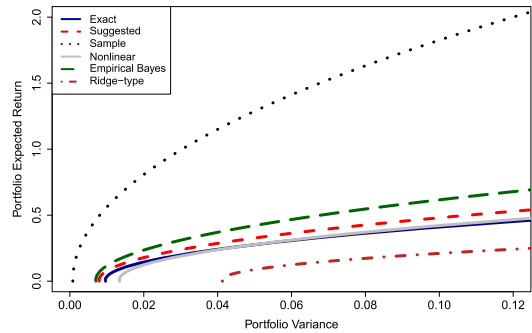
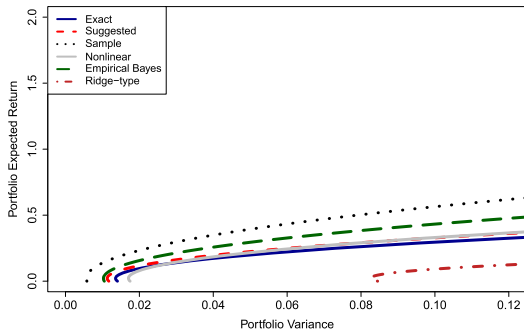
$c = 0.5$, Normal distribution

$c = 0.9$, Normal distribution



$c = 0.5$, t distribution

$c = 0.9$, t distribution



$c = 0.5$, CCC-GARCH

$c = 0.9$, CCC-GARCH

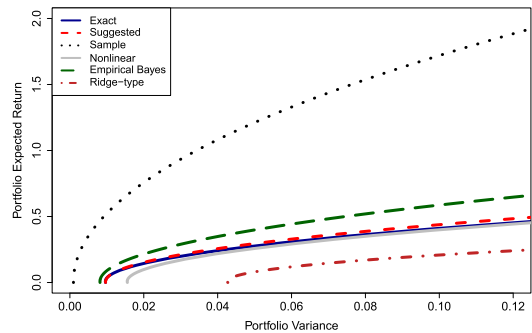
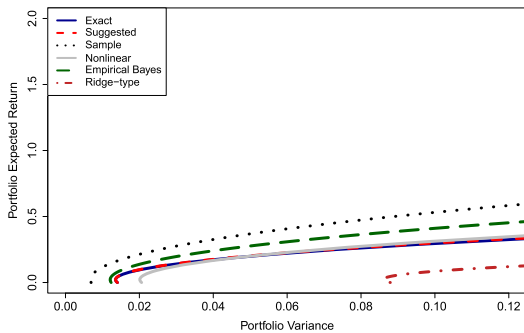


Figure 7. Estimated efficient frontier for the normal distribution (above), for the t -distribution with 3 degrees of freedom (in the middle), and for the CCC-GARCH process (below). We put $n = 100$.

considerable overestimation of the location of the vertex of the efficient frontier in the mean-variance space. The consistent and nonlinear shrinkage estimators are followed by the empirical Bayes approach. Finally, the ridge-type estimator underestimates the location of the population efficient frontier in the mean-variance space for all considered cases.

To summarize, we conclude that the proposed estimator for the efficient frontier shows significant improvement over the sample estimator for large dimensional data. In the case of the normal distribution, it is a great alternative to all of the considered benchmarks. Under heavy-tailed data, the proposed estimator is still much better in comparison to the estimators based on the empirical Bayes estimator and the ridge-type estimator of the inverse covariance matrix.

4. Empirical application

In this section, we apply the theoretical results of the paper to real data. Our goal is to estimate efficient frontiers based on a large cross-section of stocks without relying on data from overly extended periods. Instead, we seek to provide insights into the time-varying nature of efficient frontiers over relatively short intervals. This approach excludes the use of daily data, where estimating efficient frontiers for hundreds of assets would necessitate sample periods spanning several months or years. Consequently, we use one-minute intraday returns of a broad cross-section of S&P 500 constituents over the period from March 3, 2017, to June 6, 2022. This period is particularly interesting as it encompasses the COVID-19 pandemic. The choice of this time interval highlights the importance of short estimation periods while maintaining a large number of assets. Using daily data at the end of the period would inevitably include observations from before the pandemic, leading to deteriorated estimators of the actual frontier. This further underscores the relevance of high-frequency data and short estimation windows for portfolio selection.

We use a $p = 200$ -dimensional portfolio constructed based on the most frequently traded assets each day of the sample window. We set $n = 375$, corresponding to the number of 1-minute intervals on a single trading day from 9:45 to 16:00 (leaving out the first and last 15 minutes to avoid market opening and closure effects). To weaken the influence of very large and very small values of asset returns on the estimation procedure, we perform a proper truncation of large (in absolute terms) returns in the considered data sets that is commonly referred to as winsorization. This method is widely used by quantitative portfolio managers (Chincarini and Kim 2006, 180).

Intraday data at high-frequency, for example, 1-minute data, induce market-microstructure noise. A common practice is to deploy 5-min observations (see Barndorff-Nielsen and Shephard 2002; T. Anderson et al. 2001). Using returns at higher frequency causes the difficulty that the underlying data is clearly not i.i.d. due to serial correlation and persistent variances.

Therefore, we alternatively consider 5-, 10-, 30-, and 60-minute returns, which are shown to be significantly less autocorrelated but are still subject to volatility clustering. The latter effect does not align with the assumptions underlying the proposed estimators. However, our simulation results in Section 3 provide promising evidence that our estimator is sufficiently robust to violations from the underlying assumptions, particularly GARCH-type effects. Moreover, high-frequency data are subject to intraday periodicities challenging the assumption of an underlying deterministic covariance matrix. Bibinger et al. (2014) show that intraday seasonalities in covariances are clearly less pronounced than seasonalities in variances. Since in our high-dimensional portfolio application, the contribution of covariances clearly dominates the role of variances, we suggest that the effects of seasonalities on our estimates should be of minor importance.

To make the estimates based on alternative sampling schemes comparable, we fix the ratio p/n with $p = 200$ and $n = 375$. This, however, requires utilizing data from neighboring days. In this case, we estimate the parameters using rolling windows, which are moved forward on a day-to-day basis. Hence, efficient frontiers based on 5 minutes utilize data from rolling 5-day windows before the corresponding day. The resulting estimated efficient frontier can be seen as the 5-min frontier averaged over the past 5 trading days. Accordingly, using 10-, 30-, and 60-minute returns result in estimates averaged over periods of 10-, 30-, and 60 trading days. To make the resulting efficient frontiers comparable, the underlying returns are scaled to 60-min holding periods corresponding to the lowest sampling frequency in the given context. The efficient frontier parameters R_{GMV} , V_{GMV} , and s are alternatively estimated using the sample estimator, the proposed consistent estimator, the empirical Bayes estimator, and the ridge-type estimator.

Figures 8–10 display the estimates based on our proposed estimator of the expected return, the variance, and the slope of the global minimum variance portfolio for different frequencies. In both figures, we observe more

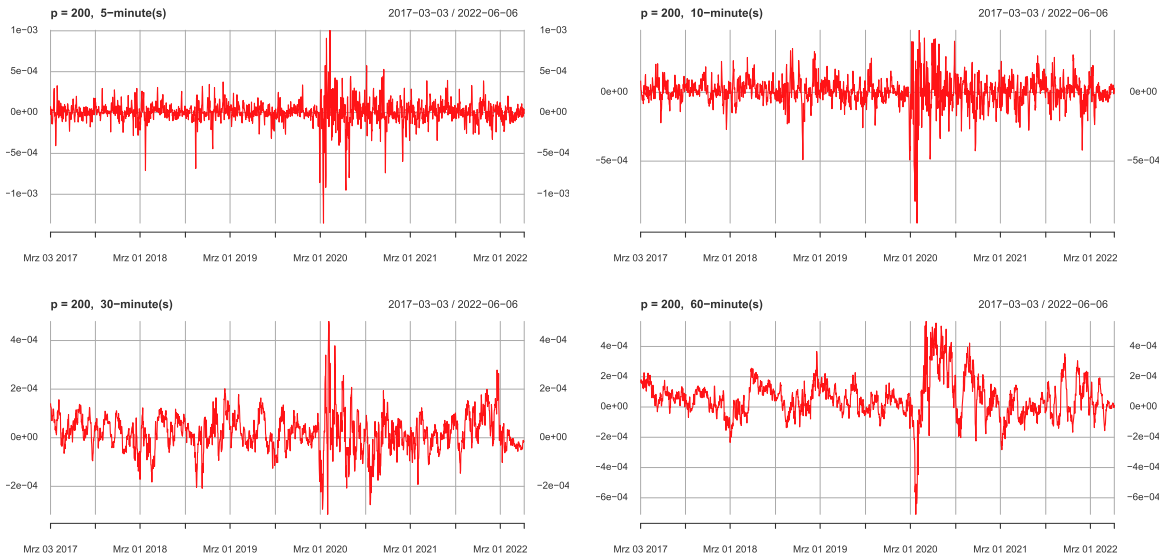


Figure 8. Estimated expected return of the global minimum variance portfolio based on 5-, 10-, 30-, and 60-minute returns on first 200 stocks from S&P500 from 03 March 2017 to 6 June 2022.

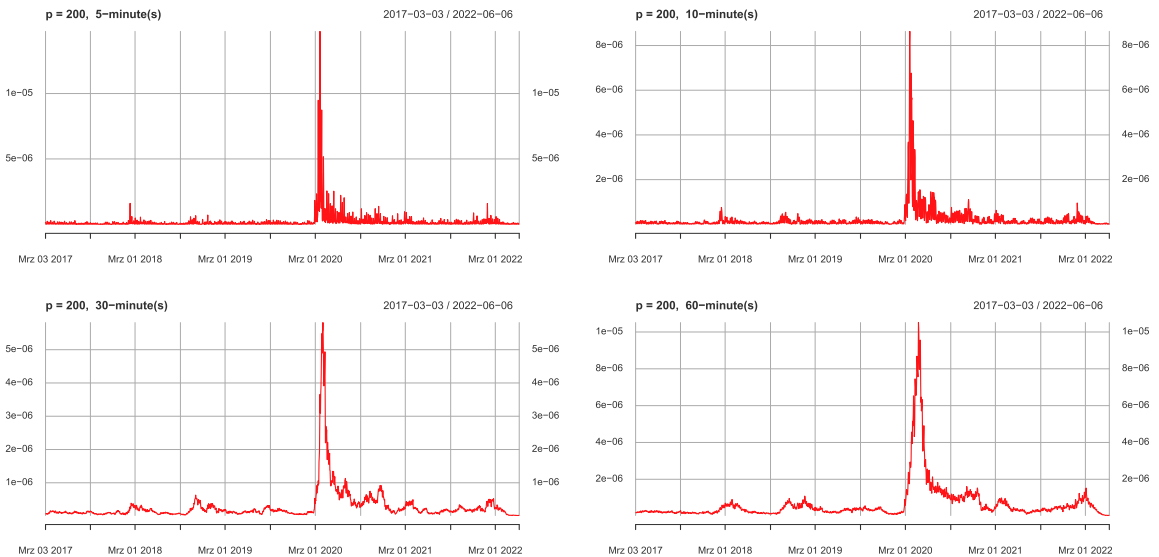


Figure 9. Estimated variance of the global minimum variance portfolio based on 5-, 10-, 30-, and 60-minute returns on first 200 stocks from S&P500 from 03 March 2017 to 6 June 2022.

volatile behavior when the returns of higher frequency are used. Remarkably, the parameters of the efficient frontier react to the financial crisis in different ways. While a large increase in the variance of the GMV portfolio is observed, the changes in the other two parameters are not large, especially in the case of the slope parameter of the efficient frontier whose behavior shows only minor changes during this period. Important is the fact that the slope parameter decreases during the crisis. In general, however, the level and pattern of all characteristics are similar for different data frequencies but with more variability for more frequent returns.

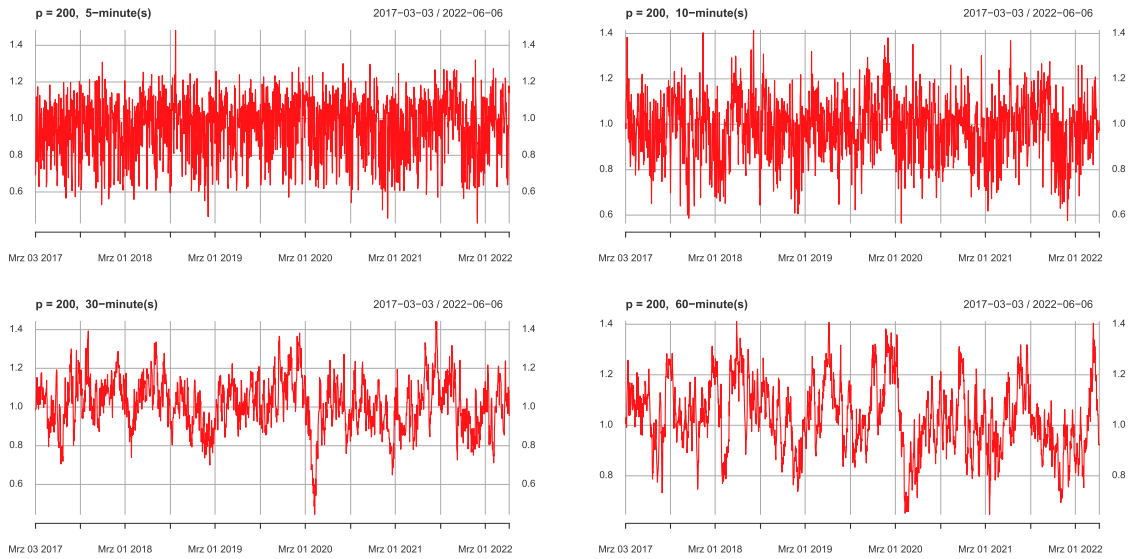


Figure 10. Estimated slope parameter of the efficient frontier based on 5-, 10-, 30-, and 60-minute returns on first 200 stocks from S&P500 from 03 March 2017 to 6 June 2022.

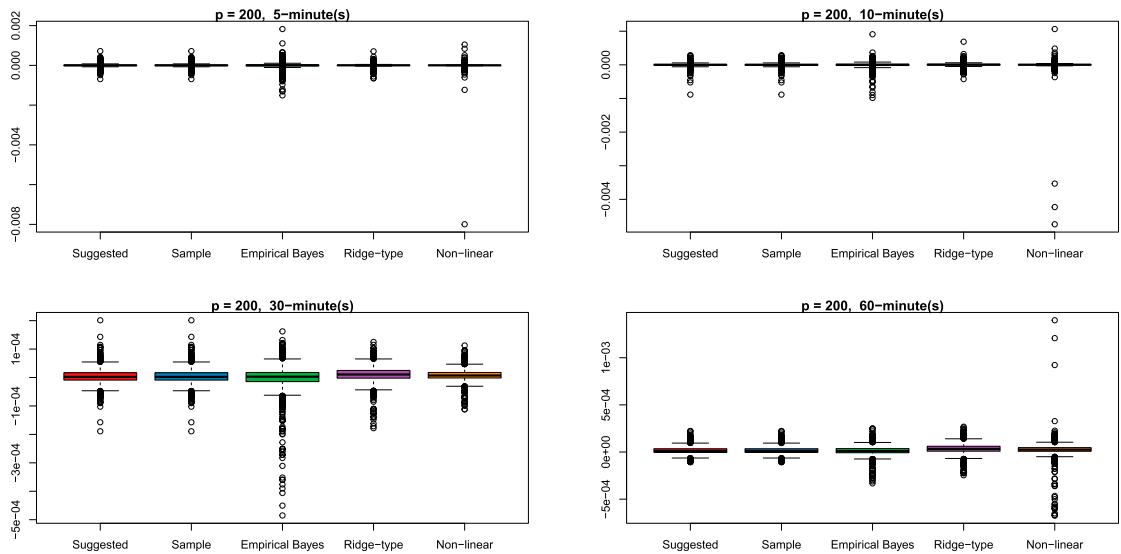


Figure 11. Box plot of the estimated expected return of the global minimum variance portfolio based on 5-, 10-, 30-, and 60-minute returns on first 200 stocks from S&P500 from 03 March 2017 to 6 June 2022.

In Figures 11–13, we present the box plots of the considered estimators for the three parameters of the efficient frontier. In the case of the expected return of the global minimum variance portfolio (cf. Figure 11), the box plots are almost identical with the exception of the empirical Bayes estimator with more volatile behavior for all of the considered frequencies. In contrast, the box plots in the case of the other two parameters of the efficient frontier are very different (cf. Figures 12 and 13). Furthermore, we observe that the values of the estimated expected return and the estimated variance of the global minimum variance portfolios get closer to zero with increasing

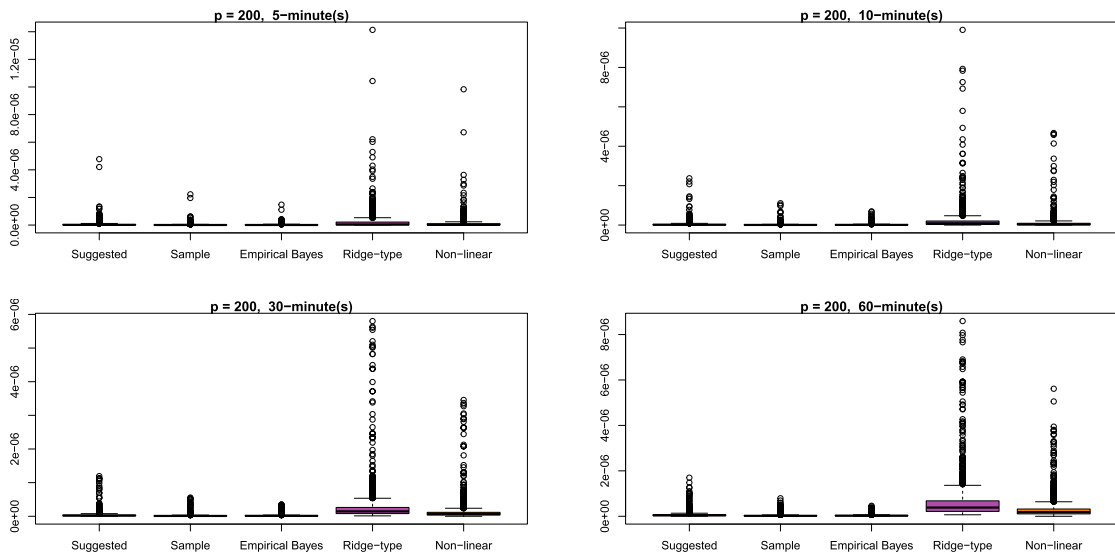


Figure 12. Box plot of the estimated variance of the global minimum variance portfolio based on 5-, 10-, 30-, and 60-minute returns on first 200 stocks from S&P500 from 03 March 2017 to 6 June 2022.

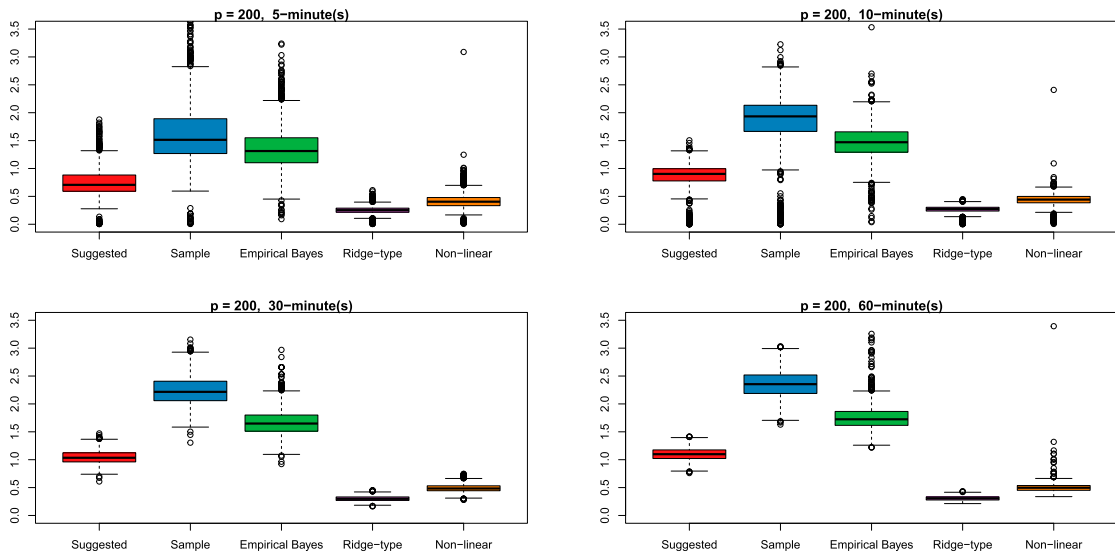


Figure 13. Box plot of the estimated slope coefficient of the efficient frontier based on 5-, 10-, 30-, and 60-minute returns on first 200 stocks from S&P500 from 03 March 2017 to 6 June 2022.

frequency. Surprisingly, the ridge-type estimators are the most volatile for the variance and the least volatile for the slope. This effect is not present in the figure for the efficient frontier. Finally, as expected, the sample estimator underestimates the GMV portfolio’s variance and overestimates the efficient frontier’s slope parameter.

The estimators for the efficient frontier are presented for several frequencies in Figure 14 in the case of June 3, 2022. The results of this figure are in line with the finding of the simulation study. Namely, the sample and the empirical Bayes efficient frontiers are overoptimistic and deviate considerably from the consistent estimator of

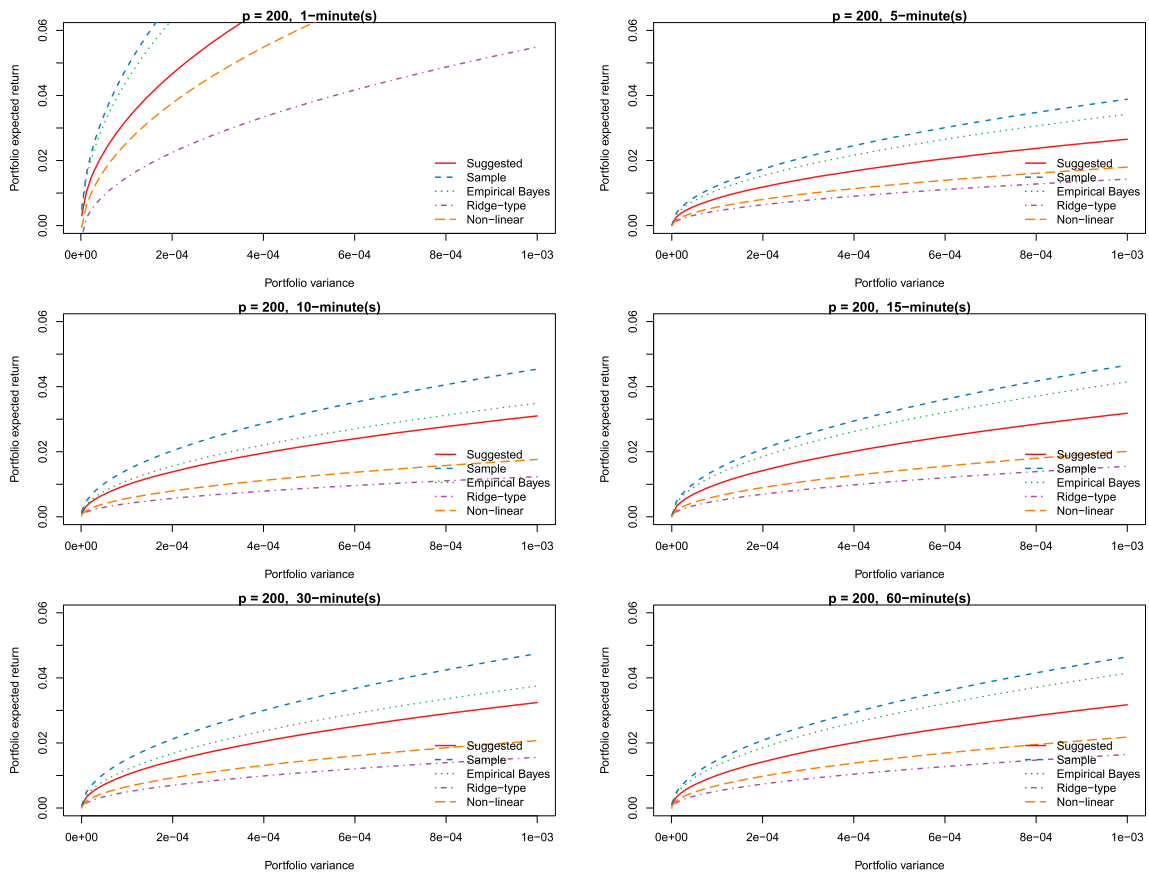


Figure 14. The sample, the consistent, the empirical Bayes, the ridge-type, and the non-linear shrinkage estimators for the efficient frontier based on 5-, 10-, 30-, and 60-minute returns on 200 stocks from S&P500 calculated on June 3, 2022.

the efficient frontier. On the contrary, the ridge-type estimator and the non-linear shrinkage are too conservative and reflect the specific behavior of the variance and the slope in the figures above.

5. Conclusions

The efficient frontier derived by Markowitz (1952) describes a set of portfolios that indicates the optimal asset allocation for the investor. Under uncertainty, the efficient frontier must be estimated from a given data set. In the case of a fixed number of assets p and a large sample size n , the sample estimator is a good intuitive choice. However, when p grows together with n such that their ratio approaches a constant $c \in (0, 1)$, the sample estimator becomes a poor alternative.

To handle ‘the curse of dimensionality,’ we investigate the asymptotic behavior of the sample estimator for the efficient frontier under large-dimensional asymptotics. We identify the asymptotic biases and correct them to develop consistent estimators for the parameters of the efficient frontier. We prove that two of the three parameters of the efficient frontier are highly biased under the large-dimensional asymptotic regime and that their biases are pure functions of the concentration ratio $c \in (0, 1)$.

Finally, we provide a simulation study to investigate the convergence rate and the performance of the proposed estimator for the efficient frontier. We take state-of-the-art estimators for high-dimensional data and compare them with the suggested one. The proposed estimator, due to its simple form, appears to be a strong alternative in the case of large-dimensional data.

One possible direction of extension of the results derived in the paper is to incorporate high-order moments in the construction of an efficient frontier. The importance of higher-order moments of the asset returns in portfolio theory has recently been highlighted in the number of papers, see Patton (2004), C. Adcock (2005), Jondeau and Rockinger (2006), Mencía and Sentana (2009), Harvey et al. (2010), and C. Adcock, Eling, and Loperfido (2015), among others. While Patton (2004) studies the empirical importance of the skewness for portfolio selection, C. Adcock (2005) presents several methods for including skewness in the decision process. Several theoretical findings for portfolio selection with higher moments are provided in De Athayde and Flôres Jr (2004), Mencía and Sentana (2009), C. J. Adcock (2010), and C. Adcock (2014). In particular, Mencía and Sentana (2009) and C. Adcock (2014) discuss the expression of the mean-variance-skewness efficient frontier. The derivation of a consistent estimator of the high-dimensional mean-variance-skewness efficient frontier can be one of the potential topics, left for future research.

Another direction of possible research may deal with the estimation of the efficient frontier under structural constraints imposed on portfolio weights. In particular, Jagannathan and Ma (2003) argue that imposing linear constraints on optimal portfolio weights may reduce the estimation risk, while the performance of optimal portfolios under the transaction costs is investigated in Mei, DeMiguel, and Nogales (2016), DeMiguel et al. (2020), DeMiguel, Martin-Utrera, and Uppal (2024), and O. Bodnar, Bodnar, and Niklasson (2025), among others. It is expected that constraints imposed on portfolio weights will shorten the feasible set of portfolios and will change the location of the efficient frontier in the mean-variance space by increasing the variance of the global minimum variance portfolio and reducing the expected return of the global minimum variance portfolio and the slope coefficient of the efficient frontier. The statistical significance of these changes is a challenging topic that can be considered in future research.

Notes

1. Note that this is not Merton's constant c , which is written in calligraphic font.
2. Under large-dimensional asymptotics, both the dimension p and the sample size n tend to infinity. Therefore, it is natural to assume, without loss of generality, that the dimension $p \equiv p(n)$ is the function of the sample size n . Consequently, the mean vector and the covariance matrix are indexed with n as well.

Acknowledgments

The authors would like to thank Professor Chris Adcock and two anonymous Reviewers for their constructive comments that improved the quality of this paper. The authors express their gratitude to the participants of the CMS Statistics 2015 conference in London and the World Congress of the Econometric Society 2015 in Montreal for their valuable discussion.

Data availability statement

The data were obtained from Thomson Reuters Eikon. These data are not publicly available.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Taras Bodnar is Professor of Mathematical Finance at the Department of Management and Engineering, Linköping University, Sweden. He received his M.S. degree in mathematics (with honors) from Ivan Franko National University of Lviv in 2001. He earned two Ph.D. degrees: one in economics and statistics (*summa cum laude*) from the European University Viadrina, Frankfurt (Oder), in 2004, and another in mathematics and physics from Taras Shevchenko National University of Kyiv in 2009. From 2005 to 2012, he served as a Research Associate at the European University Viadrina, and from 2012 to 2014 at the Department of Mathematics, Humboldt University of Berlin. He held from 2015 to 2019 an associate professorship and from 2019 to 2024 a professorship in mathematical statistics at Stockholm University. His research spans high-dimensional statistics, random matrix theory, Bayesian inference, and statistical methods in finance. He has published widely in these areas and serves on the editorial boards of the Journal of Multivariate Analysis and Theory of Probability and Mathematical Statistics. In 2008, he received the Young Scientists Award of the Brandenburg state for outstanding work in the humanities and social sciences. His contributions bridge theoretical statistics and practical applications in finance and data science.

Nikolaus Hautsch is Professor of Finance and Statistics at the University of Vienna. He earned his Ph.D. in econometrics in 2003 from the University of Konstanz. From 2004 to 2007 he joined the Department of Economics of the University of Copenhagen as Assistant Professor and Associate Professor. Until 2013 he held the Chair of Econometrics at Humboldt University Berlin and served as director of the Berlin Doctoral Program in Economics and Management Science. He is elected fellow of the Society for Financial Econometrics, research fellow of the Center for Financial Studies (CFS) Frankfurt and a staff member of the Vienna Graduate School of Finance. Hautsch had visiting positions at the University of Technology, Sydney, the University of Melbourne, the Université Catholique de Louvain, the University of Cambridge and Duke University. His research focuses on financial econometrics, modelling of large-scale high-frequency data, financial risk management, market microstructure, analysis of systemic risk and economic networks, blockchain technology in finance, machine learning in finance as well as statistics in sports. Currently, he serves as associate editor of the *Journal of Business and Economic Statistics*, the *Journal of Financial Econometrics* and the *International Journal of Forecasting*, among others.

Yarema Okhrin is Professor of Statistics and Data Science at the Faculty of Business and Economics, University of Augsburg, Germany. He received his M.S. degree in from the Ivan Franko National University of Lviv in 1999, and his Ph.D. in economics and statistics from the European University Viadrina, Frankfurt (Oder), Germany. From 2000 to 2008, he was a Research Associate at the Department of Statistics at the European University Viadrina. He subsequently held the position of Assistant Professor of Econometrics at the University of Bern, Switzerland, from 2008 to 2009, before joining the University of Augsburg. His research focuses on financial econometrics, machine learning in finance, and the statistical analysis of image, network, and environmental data. He has published widely in these areas and is Editor-in-Chief of *Advances in Statistical Analysis*. Professor Okhrin has been active in the German Statistical Society, serving as Vice-President from 2012 to 2022. As of 2024, he is the elected President of the Society. His work combines methodological rigor with interdisciplinary applications, contributing to both theoretical and applied developments in modern statistics.

Nestor Parolya is Assistant Professor of Statistics at the Delft Institute of Applied Mathematics, Delft University of Technology, the Netherlands. He earned his B.S. and M.S. degrees in mathematics and statistics (both with honors) from the Ivan Franko National University of Lviv in 2009 and 2010, respectively, and received his Ph.D. in economics and statistics (*summa cum laude*) from the European University Viadrina, Frankfurt (Oder), in 2014. He conducted postdoctoral research in statistics at Ruhr-University Bochum, and from 2014 to 2019 served as Assistant Professor (non-tenured) of Financial Econometrics at Leibniz University Hannover. During 2017–2018, he held visiting professorships at the universities of Heidelberg and Mannheim. His research focuses on high-dimensional statistics, random matrix theory, statistical learning, and applications in financial econometrics, mathematical finance, and operations research. He has published in leading journals in statistics and finance and is recognized for contributions that bridge theory and practice in complex data environments. In 2019, Dr. Parolya was awarded the Wolfgang-Wetzlar-Preis by the German Statistical Society for outstanding contributions to statistical methodology and its applications.

ORCID

Taras Bodnar  <http://orcid.org/0000-0001-7855-8221>

Nikolaus Hautsch  <https://orcid.org/0000-0001-9702-3871>

Yarema Okhrin  <http://orcid.org/0000-0003-4704-5233>

Nestor Parolya  <http://orcid.org/0000-0003-2147-2288>

References

- Adcock, C. 2005. "Exploiting Skewness to Build An Optimal Hedge Fund with a Currency Overlay." *The European Journal of Finance* 11 (5): 445–462. <https://doi.org/10.1080/13518470500039527>.
- Adcock, C. J. 2010. "Asset Pricing and Portfolio Selection Based on the Multivariate Extended Skew-Student-t Distribution." *Annals of Operations Research* 176 (1): 221–234. <https://doi.org/10.1007/s10479-009-0586-4>.
- Adcock, C. 2014. "Mean-variance-skewness Efficient Surfaces, Stein's Lemma and the Multivariate Extended Skew-Student Distribution." *European Journal of Operational Research* 234 (2): 392–401. <https://doi.org/10.1016/j.ejor.2013.07.011>.
- Adcock, C., M. Eling, and N. Loperfido. 2015. "Skewed Distributions in Finance and Actuarial Science: A Review." *The European Journal of Finance* 21 (13–14): 1253–1281. <https://doi.org/10.1080/1351847X.2012.720269>.
- Anderson, T., T. Bollerslev, F. Diebold, and P. Labys. 2001. "Modeling and Forecasting Realized Volatility." *Econometrica* 71 (2): 579–625. <https://doi.org/10.1111/ecta.2003.71.issue-2>.
- Anderson, G., A. Guionnet, and O. Zeitouni. 2010. *An Introduction to Random Matrices*. New York, USA: Cambridge University Press.
- Ao, M., L. Yingying, and X. Zheng. 2019. "Approaching Mean-Variance Efficiency for Large Portfolios." *The Review of Financial Studies* 32 (7): 2890–2919. <https://doi.org/10.1093/rfs/hhy105>.
- Bai, Z., B. Miao, and G. Pan. 2007. "On Asymptotics of Eigenvectors of Large Sample Covariance Matrix." *Annals of Probability* 35 (4): 1532–1572. <https://doi.org/10.1214/009117906000001079>.

- Bai, Z., and J. Silverstein. 2010. *Spectral Analysis of Large Dimensional Random Matrices*. New York, Dordrecht, Heidelberg, London: Springer.
- Barndorff-Nielsen, O. E., and N. Shephard. 2002. "Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models." *Journal of the Royal Statistical Society Series B* 64 (2): 253–280. <https://doi.org/10.1111/1467-9868.00336>.
- Basak, G. K., R. Jagannathan, and T. Ma. 2005. "Estimation the Risk in Sample Efficient Portfolios." Working Paper, Northwestern University.
- Bauder, D., R. Bodnar, T. Bodnar, and W. Schmid. 2019. "Bayesian Estimation of the Efficient Frontier." *Scandinavian Journal of Statistics* 46 (3): 802–830. <https://doi.org/10.1111/sjos.v46.3>.
- Bauder, D., T. Bodnar, N. Parolya, and W. Schmid. 2021. "Bayesian Mean-Variance Analysis: Optimal Portfolio Selection under Parameter Uncertainty." *Quantitative Finance* 21 (2): 221–242. <https://doi.org/10.1080/14697688.2020.1748214>.
- Bellman, R. 1961. *Adaptive Control Processes: A Guided Tour*. Princeton: Princeton University Press.
- Bibinger, M., N. Hautsch, P. Malec, and M. Reiss. 2014. "Estimating the Quadratic Covariation Matrix from Noisy Observations: Local Method of Moments and Efficiency." *The Annals of Statistics* 42 (4): 1312–1346. <https://doi.org/10.1214/14-AOS1224>.
- Bodnar, O., and T. Bodnar. 2010. "On the Unbiased Estimator of the Efficient Frontier." *International Journal of Theoretical and Applied Finance* 13 (7): 1065–1073. <https://doi.org/10.1142/S021902491000611X>.
- Bodnar, O., T. Bodnar, and V. Niklasson. 2025. "Incorporating Different Sources of Information for Bayesian Optimal Portfolio Selection." *Journal of Business & Economic Statistics* 43 (2): 365–377. <https://doi.org/10.1080/07350015.2024.2379361>.
- Bodnar, O., T. Bodnar, and N. Parolya. 2022. "Recent Advances in Shrinkage-Based High-Dimensional Inference." *Journal of Multivariate Analysis* 188:104826. <https://doi.org/10.1016/j.jmva.2021.104826>.
- Bodnar, T., H. Dette, and N. Parolya. 2019. "Testing for Independence of Large Dimensional Vectors." *The Annals of Statistics* 47 (5): 2977–3008. <https://doi.org/10.1214/18-AOS1771>.
- Bodnar, T., H. Dette, N. Parolya, and E. Thorsén. 2022. "Sampling Distributions of Optimal Portfolio Weights and Characteristics in Low and Large Dimensions." *Random Matrices: Theory and Applications* 11 (1): 2250008. <https://doi.org/10.1142/S2010326322500083>.
- Bodnar, T., S. Dmytriv, Y. Okhrin, D. Otryakhin, and N. Parolya. 2025. "High-Dimensional Portfolio Selection with HDShOP Package." *The European Journal of Finance*. <https://doi.org/10.1080/1351847X.2025.2501637>.
- Bodnar, T., S. Dmytriv, Y. Okhrin, N. Parolya, and W. Schmid. 2021. "Statistical Inference for the Expected Utility Portfolio in High Dimensions." *IEEE Transactions on Signal Processing* 69:1–14. <https://doi.org/10.1109/TSP.78>.
- Bodnar, T., S. Dmytriv, N. Parolya, and W. Schmid. 2019. "Tests for the Weights of the Global Minimum Variance Portfolio in a High-Dimensional Setting." *IEEE Transactions on Signal Processing* 67 (17): 4479–4493. <https://doi.org/10.1109/TSP.78>.
- Bodnar, T., and A. K. Gupta. 2009. "Construction and Inferences of the Efficient Frontier in Elliptical Models." *Journal of the Japan Statistical Society* 39 (2): 193–207. <https://doi.org/10.14490/jjss.39.193>.
- Bodnar, T., A. Gupta, and N. Parolya. 2014. "On the Strong Convergence of the Optimal Linear Shrinkage Estimator for Large Dimensional Covariance Matrix." *Journal of Multivariate Analysis* 132:215–228. <https://doi.org/10.1016/j.jmva.2014.08.006>.
- Bodnar, T., A. Gupta, and N. Parolya. 2016. "Direct Shrinkage Estimation of Large Dimensional Precision Matrix." *Journal of Multivariate Analysis* 146: 223–236. <http://dx.doi.org/10.1016/j.jmva.2015.09.010>.
- Bodnar, T., M. Lindholm, V. Niklasson, and E. Thorsén. 2022. "Bayesian Portfolio Selection Using VaR and CVaR." *Applied Mathematics and Computation* 427:127120. <https://doi.org/10.1016/j.amc.2022.127120>.
- Bodnar, T., Y. Okhrin, and N. Parolya. 2023. "Optimal Shrinkage-Based Portfolio Selection in High Dimensions." *Journal of Business & Economic Statistics* 41 (1): 140–156. <https://doi.org/10.1080/07350015.2021.2004897>.
- Bodnar, T., N. Parolya, and W. Schmid. 2018. "Estimation of the Global Minimum Variance Portfolio in High Dimensions." *European Journal of Operational Research* 266 (1): 371–390. <https://doi.org/10.1016/j.ejor.2017.09.028>.
- Bodnar, T., N. Parolya, and E. Thorsén. 2023. "Dynamic Shrinkage Estimation of the High-Dimensional Minimum-Variance Portfolio." *IEEE Transactions on Signal Processing* 71:1334–1349. <https://doi.org/10.1109/TSP.2023.3263950>.
- Bodnar, T., N. Parolya, and E. Thorsén. 2024. "Two is Better than One: Regularized Shrinkage of Large Minimum Variance Portfolios." *Journal of Machine Learning Research* 25 (173): 1–32.
- Bodnar, T., and M. Reiß. 2016. "Exact and Asymptotic Tests on a Factor Model in Low and Large Dimensions with Applications." *Journal of Multivariate Analysis* 150:125–151. <https://doi.org/10.1016/j.jmva.2016.05.011>.
- Bodnar, T., and W. Schmid. 2008. "Estimation of Optimal Portfolio Compositions for Gaussian Returns." *Statistics & Decisions* 26:179–201. <https://doi.org/10.1524/stnd.2008.0918>.
- Bodnar, T., and W. Schmid. 2009. "Econometrical Analysis of the Sample Efficient Frontier." *The European Journal of Finance* 15 (3): 317–335. <https://doi.org/10.1080/13518470802423478>.
- Bühlmann, P., and S. van de Geer. 2011. *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Heidelberg, New York: Springer.
- Cai, T. T., J. Hu, Y. Li, and X. Zheng. 2020. "High-Dimensional Minimum Variance Portfolio Estimation Based on High-Frequency Data." *Journal of Econometrics* 214 (2): 482–494. <https://doi.org/10.1016/j.jeconom.2019.04.039>.
- Cai, T., and X. Shen. 2011. *High-Dimensional Data Analysis. Vol. 2 of Frontiers of Statistics*. Singapore: World Scientific.
- Chincarini, L. B., and D. Kim. 2006. *Quantitative Equity Portfolio Management: An Active Approach to Portfolio Construction and Management*. New York: McGraw-Hill.
- DasGupta, A. 2008. *Asymptotic Theory of Statistics and Probability*. Vol. 180. New York: Springer.

- De Athayde, G. M., and R. G. Jr. Flóres . 2004. "Finding a Maximum Skewness Portfolio - a General Solution to Three-Moments Portfolio Choice." *Journal of Economic Dynamics and Control* 28 (7): 1335–1352. [https://doi.org/10.1016/S0165-1889\(02\)00084-2](https://doi.org/10.1016/S0165-1889(02)00084-2).
- DeMiguel, V., A. Martin-Utrera, F. J. Nogales, and R. Uppal. 2020. "A Transaction-Cost Perspective on the Multitude of Firm Characteristics." *The Review of Financial Studies* 33 (5): 2180–2222. <https://doi.org/10.1093/rfs/hhz085>.
- DeMiguel, V., A. Martin-Utrera, and R. Uppal. 2024. "A Multifactor Perspective on Volatility-Managed Portfolios." *The Journal of Finance* 79 (6): 3859–3891.
- De Nard, G., R. F. Engle, O. Ledoit, and M. Wolf. 2022. "Large Dynamic Covariance Matrices: Enhancements Based on Intraday Data." *Journal of Banking & Finance* 138:106426. <https://doi.org/10.1016/j.jbankfin.2022.106426>.
- Ding, Y., Y. Li, and X. Zheng. 2021. "High Dimensional Minimum Variance Portfolio Estimation under Statistical Factor Models." *Journal of Econometrics* 222 (1): 502–515. <https://doi.org/10.1016/j.jeconom.2020.07.013>.
- Efron, B., and C. Morris. 1976. "Multivariate Empirical Bayes and Estimation of Covariance Matrices." *Annals of Statistics* 4:22–32.
- El Karoui, N. 2010. "High-Dimensionality Effects in the Markowitz Problem and Other Quadratic Programs with Linear Constraints: Risk Underestimation." *The Annals of Statistics* 38 (6): 3487–3566. <https://doi.org/10.1214/10-AOS795>.
- Fama, E. F. 1965. "The Behavior of Stock Market Prices." *The Journal of Business* 38 (1): 34–105. <https://doi.org/10.1086/jb.1965.38.issue-1>.
- Fan, J., Y. Fan, and J. Lv. 2008. "High Dimensional Covariance Matrix Estimation Using a Factor Model." *Journal of Econometrics* 147 (1): 186–197. <https://doi.org/10.1016/j.jeconom.2008.09.017>.
- Fan, J., Y. Liao, and M. Mincheva. 2013. "Large Covariance Estimation by Thresholding Principal Orthogonal Complements." *Journal of the Royal Statistical Society: Series B* 75 (4): 603–680. <https://doi.org/10.1111/rssb.12016>.
- Fan, J., J. Zhang, and K. Yu. 2012. "Vast Portfolio Selection with Gross-Exposure Constraints." *Journal of American Statistical Association* 107 (498): 592–606. <https://doi.org/10.1080/01621459.2012.682825>.
- Friesen, O., M. Löwe, and M. Stolz. 2013. "Gaussian Fluctuations for Sample Covariance Matrices with Dependent Data." *Journal of Multivariate Analysis* 114:270–287. <https://doi.org/10.1016/j.jmva.2012.08.004>.
- Girko, V., and A. K. Gupta. 1994. "Asymptotic Behaviour of Spectral Function of Empirical Covariance Matrices." *Random Operators and Stochastic Equations* 2:43–60.
- Girko, V., and A. K. Gupta. 1996a. "Canonical Equation for the Resolvent of Empirical Covariance Matrices Pencil." In *Multidimensional Statistical Analysis and Theory of Random Matrices*, edited by A. K. Gupta and V. L. Girko. Netherlands: VSP Publishers.
- Girko, V., and A. K. Gupta. 1996b. "Multivariate Elliptically Contoured Linear Models and Some Aspects of the Theory of Random Matrices." In *Multidimensional Statistical Analysis and Theory of Random Matrices*, edited by A. K. Gupta and V. L. Girko. Netherlands: VSP Publishers.
- Glombek, K. 2014. "Statistical Inference for High-Dimensional Global Minimum Variance Portfolios." *Scandinavian Journal of Statistics* 41 (4): 845–865. <https://doi.org/10.1111/sjos.v41.4>.
- Golosnoy, V., and B. Gribisch. 2022. "Modeling and Forecasting Realized Portfolio Weights." *Journal of Banking & Finance* 138:106404. <https://doi.org/10.1016/j.jbankfin.2022.106404>.
- Golosnoy, V., B. Gribisch, and M. I. Seifert. 2022. "Sample and Realized Minimum Variance Portfolios: Estimation, Statistical Inference, and Tests." *Wiley Interdisciplinary Reviews: Computational Statistics* 14 (5): e1556. <https://doi.org/10.1002/wics.v14.5>.
- Golosnoy, V., and Y. Okhrin. 2007. "Multivariate Shrinkage for Optimal Portfolio Weights." *The European Journal of Finance* 13 (5): 441–458. <https://doi.org/10.1080/13518470601137592>.
- Gupta, A. K., T. Varga, and T. Bodnar. 2013. *Elliptically Contoured Models in Statistics and Portfolio Theory*. New York: Springer.
- Harvey, C. R., J. C. Liechty, M. W. Liechty, and P. Müller. 2010. "Portfolio Selection with Higher Moments." *Quantitative Finance* 10 (5): 469–485. <https://doi.org/10.1080/14697681003756877>.
- Hautsch, N., L. Kyj, and P. Malec. 2015. "Do High-Frequency Data Improve High-Dimensional Portfolio Allocation?" *Journal of Applied Econometrics* 30 (2): 263–290. <https://doi.org/10.1002/jae.2361>.
- Hautsch, N., L. Kyj, and R. Oomen. 2012. "A Blocking and Regularization Approach to High Dimensional Realized Covariance Estimation." *Journal of Applied Econometrics* 27 (4): 625–645. <https://doi.org/10.1002/jae.v27.4>.
- Jagannathan, R., and T. Ma. 2003. "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps." *The Journal of Finance* 58 (4): 1651–1683. <https://doi.org/10.1111/jofi.2003.58.issue-4>.
- Jobson, J. D. 1991. "Confidence Regions for the Mean-Variance Efficient Set: An Alternative Approach to Estimation Risk." *Review of Quantitative Finance and Accounting* 1 (3): 235–257. <https://doi.org/10.1007/BF02408379>.
- Jobson, J. D., and B. Korkie. 1980. "Estimation for Markowitz Efficient Portfolios." *Journal of the American Statistical Association* 75 (371): 544–554. <https://doi.org/10.1080/01621459.1980.10477507>.
- Jondeau, E., and M. Rockinger. 2006. "Optimal Portfolio Allocation under Higher Moments." *European Financial Management* 12 (1): 29–55. <https://doi.org/10.1111/eufm.2006.12.issue-1>.
- Kan, R., and D. R. Smith. 2008. "The Distribution of the Sample Minimum-Variance Frontier." *Management Science* 54 (7): 1364–1380. <https://doi.org/10.1287/mnsc.1070.0852>.
- Kan, R., and X. Wang. 2024. "Optimal Portfolio Choice with Unknown Benchmark Efficiency." *Management Science* 70 (9): 6117–6138. <https://doi.org/10.1287/mnsc.2021.01767>.
- Kubokawa, T., and M. Srivastava. 2008. "Estimation of the Precision Matrix of a Singular Wishart Distribution and Its Application in High-Dimensional Data." *Journal of Multivariate Analysis* 99 (9): 1906–1928. <https://doi.org/10.1016/j.jmva.2008.01.016>.

Lassance, N., R. Vanderveken, and F. Vrins. 2024. “On the Combination of Naive and Mean-Variance Portfolio Strategies.” *Journal of Business & Economic Statistics* 42 (3): 875–889. <https://doi.org/10.1080/07350015.2023.2256801>.

Le Cam, L., and G. Lo Yang. 2000. *Asymptotics in Statistics: Some Basic Concepts*. New York: Springer.

Ledito, O., and M. Wolf. 2017. “Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks.” *The Review of Financial Studies* 30 (12): 4349–4388. <https://doi.org/10.1093/rfs/hhx052>.

Ledito, O., and M. Wolf. 2020. “Analytical Nonlinear Shrinkage of Large-Dimensional Covariance Matrices.” *The Annals of Statistics* 48 (5): 3043–3065. <https://doi.org/10.1214/19-AOS1921>.

Marčenko, V. A., and L. A. Pastur. 1967. “Distribution of Eigenvalues for Some Sets of Random Matrices.” *Sbornik: Mathematics* 1 (4): 457–483. <https://doi.org/10.1070/SM1967v001n04ABEH001994>.

Markowitz, H. 1952. “Portfolio Selection.” *Journal of Finance* 7:77–91.

Markowitz, H. 1991. “Foundations of Portfolio Theory.” *Journal of Finance* 7:469–477. <https://doi.org/10.1111/jofi.1991.46.issue-2>.

Mei, X., V. DeMiguel, and F. J. Nogales. 2016. “Multiperiod Portfolio Optimization with Multiple Risky Assets and General Transaction Costs.” *Journal of Banking & Finance* 69:108–120. <https://doi.org/10.1016/j.jbankfin.2016.04.002>.

Mencía, J., and E. Sentana. 2009. “Multivariate Location–scale Mixtures of Normals and Mean–variance–skewness Portfolio Allocation.” *Journal of Econometrics* 153 (2): 105–121. <https://doi.org/10.1016/j.jeconom.2009.05.001>.

Merton, R. 1972. “An Analytical Derivation of the Efficient Frontier.” *Journal of Financial and Quantitative Analysis* 7 (4): 1851–1872. <https://doi.org/10.2307/2329621>.

Okhrin, Y., and W. Schmid. 2006. “Distributional Properties of Portfolio Weights.” *Journal of Econometrics* 134 (1): 235–256. <https://doi.org/10.1016/j.jeconom.2005.06.022>.

Patton, A. J. 2004. “On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation.” *Journal of Financial Econometrics* 2 (1): 130–168. <https://doi.org/10.1093/jjfinec/nbh006>.

Rubio, F., and X. Mestre. 2011. “Spectral Convergence for a General Class of Random Matrices.” *Statistics and Probability Letters* 81 (5): 592–602. <https://doi.org/10.1016/j.spl.2011.01.004>.

Siegel, A. F., and A. Woodgate. 2007. “Performance of Portfolios Optimized with Estimation Error.” *Management Science* 53 (6): 1005–1015. <https://doi.org/10.1287/mnsc.1060.0664>.

Silverstein, J. W. 1995. “Strong Convergence of the Empirical Distribution of Eigenvalues of Large-Dimensional Random Matrices.” *Journal of Multivariate Analysis* 55 (2): 331–339. <https://doi.org/10.1006/jmva.1995.1083>.

Yin, Y. 1986. “Limiting Spectral Distribution for a Class of Random Matrices.” *Journal of Multivariate Analysis* 20 (1): 50–68. [https://doi.org/10.1016/0047-259X\(86\)90019-9](https://doi.org/10.1016/0047-259X(86)90019-9).

Appendix

We start with three important lemmas investigating the asymptotic behavior of quadratic forms in \mathbf{X}_n under the high-dimensional asymptotic regime.

Lemma A.1: Assume (A1)–(A3). Let a nonrandom $p \times p$ -dimensional matrix Θ_p and a nonrandom $n \times n$ -dimensional matrix Θ_n possess a uniformly bounded trace norms (sum of singular values) and let $\Sigma_n = \mathbf{I}_p$. Then it holds that

$$\left| \text{tr} \left(\Theta_p (1/n \mathbf{X}_n \mathbf{X}'_n - z \mathbf{I}_p)^{-1} \right) - m(z) \text{tr} \left(\Theta_p \right) \right| \xrightarrow{a.s.} 0, \tag{A1}$$

$$\left| \text{tr} \left(\Theta_n (1/n \mathbf{X}'_n \mathbf{X}_n - z \mathbf{I}_n)^{-1} \right) - \underline{m}(z) \text{tr} \left(\Theta_n \right) \right| \xrightarrow{a.s.} 0 \tag{A2}$$

for $p/n \rightarrow c \in (0, +\infty)$ as $n \rightarrow \infty$ where

$$m(z) = (x(z) - z)^{-1} \quad \text{and} \quad \underline{m}(z) = -\frac{1-c}{z} + cm(z) \tag{A3}$$

with

$$x(z) = \frac{1}{2} \left(1 - c + z + \sqrt{(1-c+z)^2 - 4z} \right). \tag{A4}$$

Proof of Lemma A.1: See, the proof of Lemma 1.1 in supplementary material of T. Bodnar, Okhrin, and Parolya (2023). ■

Lemma A.2: Assume (A1)–(A3). Let $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ be the universal nonrandom vectors from the set $\mathcal{V} = \{p^{-q/2} \Sigma_n^{-1/2} \mathbf{1}, p^{-q/2} \Sigma_n^{-1/2} \boldsymbol{\mu}_n\}$. Define $\bar{\mathbf{x}}_n = \frac{1}{n} \mathbf{X}_n \mathbf{1}$ and $\tilde{\mathbf{S}}_n = \frac{1}{n} \mathbf{X}_n \mathbf{X}'_n$. Then it holds that

$$\left| \boldsymbol{\xi}' \tilde{\mathbf{S}}_n^{-1} \boldsymbol{\theta} - (1-c)^{-1} \boldsymbol{\xi}' \boldsymbol{\theta} \right| \xrightarrow{a.s.} 0, \tag{A5}$$

$$\bar{\mathbf{x}}'_n \tilde{\mathbf{S}}_n^{-1} \bar{\mathbf{x}}_n \xrightarrow{a.s.} c, \tag{A6}$$

$$\bar{\mathbf{x}}'_n \tilde{\mathbf{S}}_n^{-1} \boldsymbol{\theta} \xrightarrow{a.s.} 0 \tag{A7}$$

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

Proof of Lemma A.2: See, Lemma 1.2 in the supplementary material of T. Bodnar, Okhrin, and Parolya (2023). ■

Lemma A.3: Assume (A1)–(A3). Let $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ be the universal nonrandom vectors from the set $\mathcal{V} = \{\mathbf{1}, \boldsymbol{\mu}_n\}$. Let $q \geq 0$. Then it holds that

$$p^{-q} \left| \boldsymbol{\xi}' \mathbf{S}_n^{-1} \boldsymbol{\theta} - (1-c)^{-1} \boldsymbol{\xi}' \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\theta} \right| \xrightarrow{a.s.} 0, \quad (\text{A8})$$

$$\bar{\mathbf{x}}_n' \boldsymbol{\Sigma}_n^{1/2} \mathbf{S}_n^{-1} \boldsymbol{\Sigma}_n^{1/2} \bar{\mathbf{x}}_n \xrightarrow{a.s.} \frac{c}{1-c}, \quad (\text{A9})$$

$$p^{-q} \bar{\mathbf{x}}_n' \boldsymbol{\Sigma}_n^{1/2} \mathbf{S}_n^{-1} \boldsymbol{\theta} \xrightarrow{a.s.} 0 \quad (\text{A10})$$

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

Proof of Lemma A.3: From the Sherman-Morrison formula we get

$$\mathbf{S}_n^{-1} = (\boldsymbol{\Sigma}_n^{1/2} \tilde{\mathbf{S}}_n \boldsymbol{\Sigma}_n^{1/2} - \boldsymbol{\Sigma}_n^{1/2} \bar{\mathbf{x}}_n \bar{\mathbf{x}}_n' \boldsymbol{\Sigma}_n^{1/2})^{-1} = \boldsymbol{\Sigma}_n^{-1/2} \tilde{\mathbf{S}}_n^{-1} \boldsymbol{\Sigma}_n^{-1/2} + \frac{\boldsymbol{\Sigma}_n^{-1/2} \tilde{\mathbf{S}}_n^{-1} \bar{\mathbf{x}}_n \bar{\mathbf{x}}_n' \tilde{\mathbf{S}}_n^{-1} \boldsymbol{\Sigma}_n^{-1/2}}{1 - \bar{\mathbf{x}}_n' \tilde{\mathbf{S}}_n^{-1} \bar{\mathbf{x}}_n}.$$

Hence, using the results of Lemma A.2 we get

$$\begin{aligned} p^{-q} \left| \boldsymbol{\xi}' \mathbf{S}_n^{-1} \boldsymbol{\theta} - (1-c)^{-1} \boldsymbol{\xi}' \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\theta} \right| &\leq p^{-q} \left| \boldsymbol{\xi}' \boldsymbol{\Sigma}_n^{-1/2} \tilde{\mathbf{S}}_n^{-1} \boldsymbol{\Sigma}_n^{-1/2} \boldsymbol{\theta} - (1-c)^{-1} \boldsymbol{\xi}' \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\theta} \right| \\ &\quad + \left| \frac{p^{-q} \boldsymbol{\xi}' \boldsymbol{\Sigma}_n^{-1/2} \tilde{\mathbf{S}}_n^{-1} \bar{\mathbf{x}}_n \bar{\mathbf{x}}_n' \tilde{\mathbf{S}}_n^{-1} \boldsymbol{\Sigma}_n^{-1/2} \boldsymbol{\theta}}{1 - \bar{\mathbf{x}}_n' \tilde{\mathbf{S}}_n^{-1} \bar{\mathbf{x}}_n} \right| \xrightarrow{a.s.} 0, \\ \bar{\mathbf{x}}_n' \boldsymbol{\Sigma}_n^{1/2} \mathbf{S}_n^{-1} \boldsymbol{\Sigma}_n^{1/2} \bar{\mathbf{x}}_n &= \bar{\mathbf{x}}_n' \tilde{\mathbf{S}}_n^{-1} \bar{\mathbf{x}}_n + \frac{(\bar{\mathbf{x}}_n' \tilde{\mathbf{S}}_n^{-1} \bar{\mathbf{x}}_n)^2}{1 - \bar{\mathbf{x}}_n' \tilde{\mathbf{S}}_n^{-1} \bar{\mathbf{x}}_n} \xrightarrow{a.s.} c + \frac{c^2}{1-c} = \frac{c}{1-c}, \\ p^{-q} \bar{\mathbf{x}}_n' \boldsymbol{\Sigma}_n^{1/2} \mathbf{S}_n^{-1} \boldsymbol{\theta} &= p^{-q} \bar{\mathbf{x}}_n' \tilde{\mathbf{S}}_n^{-1} \boldsymbol{\Sigma}_n^{-1/2} \boldsymbol{\theta} \\ &\quad + p^{-q} \frac{\bar{\mathbf{x}}_n' \tilde{\mathbf{S}}_n^{-1} \bar{\mathbf{x}}_n \bar{\mathbf{x}}_n' \tilde{\mathbf{S}}_n^{-1} \boldsymbol{\Sigma}_n^{-1/2} \boldsymbol{\theta}}{1 - \bar{\mathbf{x}}_n' \tilde{\mathbf{S}}_n^{-1} \bar{\mathbf{x}}_n} \xrightarrow{a.s.} 0 \end{aligned}$$

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

Proof of Theorem 2.1: The statement (12) follows immediately from (A8) with $\boldsymbol{\xi} = \boldsymbol{\theta} = \mathbf{1}_p$.

In order to prove (11), we consider

$$\hat{R}_{GMV} = \frac{p^{-q} \mathbf{1}_p \mathbf{S}_n^{-1} \bar{\mathbf{y}}_n}{p^{-q} \mathbf{1}_p \mathbf{S}_n^{-1} \mathbf{1}_n} = \frac{p^{-q} \mathbf{1}_p \mathbf{S}_n^{-1} \boldsymbol{\mu}_n}{p^{-q} \mathbf{1}_p \mathbf{S}_n^{-1} \mathbf{1}_p} + \frac{p^{-q} \mathbf{1}_p \mathbf{S}_n^{-1} \boldsymbol{\Sigma}_n^{1/2} \bar{\mathbf{x}}_n}{p^{-q} \mathbf{1}_p \mathbf{S}_n^{-1} \mathbf{1}_p}.$$

The application of (A8) and (A10) leads to

$$\frac{p^{-q} \mathbf{1}_p \mathbf{S}_n^{-1} \boldsymbol{\mu}_n}{p^{-q} \mathbf{1}_p \mathbf{S}_n^{-1} \mathbf{1}_p} \xrightarrow{a.s.} \frac{p^{-q} (1-c)^{-1} \mathbf{1}_p \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_n}{p^{-q} (1-c)^{-1} \mathbf{1}_p \boldsymbol{\Sigma}_n^{-1} \mathbf{1}_p} = \frac{\mathbf{1}_p \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_n}{\mathbf{1}_p \boldsymbol{\Sigma}_n^{-1} \mathbf{1}_p} = R_{GMV}$$

and

$$\frac{p^{-q} \mathbf{1}_p \mathbf{S}_n^{-1} \boldsymbol{\Sigma}_n^{1/2} \bar{\mathbf{x}}_n}{p^{-q} \mathbf{1}_p \mathbf{S}_n^{-1} \mathbf{1}_p} \xrightarrow{a.s.} 0,$$

for $p/n \rightarrow c \in (0, +\infty)$ as $n \rightarrow \infty$, from which (11) follows.

Finally, we get

$$\begin{aligned} \hat{s} &= \bar{\mathbf{y}}_n' \hat{\mathbf{Q}} \bar{\mathbf{y}}_n = (\boldsymbol{\mu}_n + \boldsymbol{\Sigma}_n^{1/2} \bar{\mathbf{x}}_n)' \mathbf{S}_n^{-1} (\boldsymbol{\mu}_n + \boldsymbol{\Sigma}_n^{1/2} \bar{\mathbf{x}}_n) - \frac{((\boldsymbol{\mu}_n + \boldsymbol{\Sigma}_n^{1/2} \bar{\mathbf{x}}_n)' \mathbf{S}_n^{-1} \mathbf{1}_p)^2}{\mathbf{1}_p' \mathbf{S}_n^{-1} \mathbf{1}_p} \\ &= \boldsymbol{\mu}_n' \mathbf{S}_n^{-1} \boldsymbol{\mu}_n + 2 \boldsymbol{\mu}_n' \mathbf{S}_n^{-1} \boldsymbol{\Sigma}_n^{1/2} \bar{\mathbf{x}}_n + \bar{\mathbf{x}}_n' \boldsymbol{\Sigma}_n^{1/2} \mathbf{S}_n^{-1} \boldsymbol{\Sigma}_n^{1/2} \bar{\mathbf{x}}_n \\ &\quad - \frac{(\boldsymbol{\mu}_n' \mathbf{S}_n^{-1} \mathbf{1}_p)^2}{\mathbf{1}_p' \mathbf{S}_n^{-1} \mathbf{1}_p} - \frac{(\bar{\mathbf{x}}_n' \boldsymbol{\Sigma}_n^{1/2} \mathbf{S}_n^{-1} \mathbf{1}_p)^2}{\mathbf{1}_p' \mathbf{S}_n^{-1} \mathbf{1}_p} - 2 \frac{(\boldsymbol{\mu}_n' \mathbf{S}_n^{-1} \mathbf{1}_p) (\bar{\mathbf{x}}_n' \boldsymbol{\Sigma}_n^{1/2} \mathbf{S}_n^{-1} \mathbf{1}_p)}{\mathbf{1}_p' \mathbf{S}_n^{-1} \mathbf{1}_p}. \end{aligned}$$

The application of the results of Lemma A.3 leads to (13). The theorem is proved. ■

Lemma A.4: (a) Let $Z_{p,n} \sim \chi_{n-p}^2$ (χ^2 -distribution with $n-p$ degrees of freedom). Then

$$\sqrt{n} \left(\frac{Z_{p,n}}{n-p} - 1 \right) \xrightarrow{d.} \mathcal{N} \left(0, \frac{2}{1-c} \right) \quad \text{for } \frac{p}{n} \rightarrow c \in (0, 1) \text{ as } n \rightarrow \infty. \quad (\text{A11})$$

(b) Let $F_{p,n} \sim F_{p,n-p,n\lambda}$ (non-central F -distribution with p and $n-p$ degrees of freedom and non-centrality parameter $n\lambda$) with $\lambda \in (0, \infty)$. Then

$$\sqrt{n} \left(\frac{n-p}{p} F_{p,n} - 1 - \frac{n}{p} \lambda \right) \xrightarrow{d.} \mathcal{N} \left(0, \frac{2}{c} \left(1 + 2 \frac{\lambda}{c} \right) + \frac{2}{1-c} \left(1 + \frac{\lambda}{c} \right)^2 \right) \quad (\text{A12})$$

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

Proof of Lemma A.4: (a) The statement follows directly from the properties of χ^2 -distribution and the fact that $c < 1$ which ensures that $n-p \rightarrow \infty$ for $\frac{p}{n} \rightarrow c \in (0, 1)$ as $n \rightarrow \infty$.

(b) From the properties of the non-central F -distribution, we get

$$\frac{n-p}{p} F_{p,n} \stackrel{d.}{=} \frac{f_1/p}{f_2/(n-p)},$$

where $f_1 \sim \chi_{p;n\lambda}^2$ (non-central χ^2 -distribution with p degrees of freedom and non-centrality parameter $n\lambda$), $f_2 \sim \chi_{n-p}^2$, and f_1, f_2 are independently distributed.

It holds that

$$\begin{aligned} \sqrt{n} \left(\frac{n-p}{p} F_{p,n} - 1 - \frac{n}{p} \lambda \right) &= \frac{1}{f_2/(n-p)} \left(\sqrt{n} \left(\frac{f_1}{p} - 1 - \frac{n}{p} \lambda \right) \right. \\ &\quad \left. - \sqrt{n} \left(\frac{f_2}{n-p} - 1 \right) \left(1 + \frac{n}{p} \lambda \right) \right). \end{aligned}$$

From Lemma A.4(a), we get

$$\sqrt{n} \left(\frac{f_2}{n-p} - 1 \right) \xrightarrow{d.} \mathcal{N} \left(0, \frac{2}{1-c} \right) \quad \text{for } \frac{p}{n} \rightarrow c \in (0, 1) \text{ as } n \rightarrow \infty.$$

Furthermore, the application of Lemma 3 in T. Bodnar and Reiß (2016) leads to $f_2/(n-p) \xrightarrow{a.s.} 1$ and

$$\sqrt{n} \left(\frac{f_1}{p} - 1 - \frac{n}{p} \lambda \right) \xrightarrow{d.} \mathcal{N} \left(0, \frac{2}{c} \left(1 + 2 \frac{\lambda}{c} \right) \right)$$

for $\frac{p}{n} \rightarrow c \in (0, 1)$ as $n \rightarrow \infty$. Putting these results together and applying Slutsky's lemma (see, e.g. Theorem 1.5 in DasGupta 2008), we obtain the statement of Lemma A.4(b).

The following result is Lemma A1 in T. Bodnar and Schmid (2009) ■

Lemma A.5: Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be a random sample of independent vectors such that $\mathbf{x}_i \sim \mathcal{N}_p(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$ for $i = 1, \dots, n$. Then for any p and n with $n > p$ it holds that

- (a) \hat{V}_{GMV} is independent of $(\hat{R}_{GMV}, \hat{\delta})$.
- (b) $(n-1)\hat{V}_{GMV}/V_{GMV} \sim \chi_{n-p}^2$.
- (c) $\frac{n(n-p+1)}{(n-1)(p-1)}\hat{\delta} \sim F_{p-1, n-p+1, n\delta}$.
- (d) $\hat{R}_{GMV}|\hat{\delta} = y \sim \mathcal{N}(R_{GMV}, \frac{1+\frac{n}{n-1}y}{n}V_{GMV})$.

Proof of Theorem 2.2: From Lemma A.5(a), we get that \hat{V}_{GMV} is independent of $(\hat{R}_{GMV}, \hat{\delta})$ for any p and n and, hence, \hat{V}_{GMV} is independent of $(\hat{R}_{GMV}, \hat{\delta})$ also under high-dimensional asymptotics. Moreover, from Lemma A.4(a) we get

$$\sqrt{n}(\hat{V}_c - V_{GMV}) = \sqrt{n} \left(\frac{1}{1-p/n} \hat{V}_{GMV} - V_{GMV} \right) = \sqrt{n} \left(\frac{n}{n-p} \frac{\hat{V}_{GMV}}{V_{GMV}} - 1 \right) V_{GMV}.$$

The application of Lemma A.5(b) leads to $\sqrt{n}(\hat{V}_c - V_{GMV}) \xrightarrow{d.} \mathcal{N}(0, \frac{2}{1-c} V_{GMV}^2)$.

Similarly, we obtain

$$\sqrt{n}(\hat{s}_c - s) = \sqrt{n}\left(\left(1 - \frac{p}{n}\right)\hat{s} - \frac{p}{n} - s\right) = \sqrt{n}\left(\frac{n-p}{p}\hat{s} - 1 - \frac{n}{p}s\right)\frac{p}{n}.$$

The application of Lemmas A.4(b) and A.5(c) leads to

$$\sqrt{n}(\hat{s}_c - s) \xrightarrow{d.} \mathcal{N}\left(0, 2(c + 2s) + \frac{2}{1-c}(c + s)^2\right).$$

Finally, the application of Lemma A.5(d) leads to

$$\sqrt{n}\begin{pmatrix} \hat{R}_c - R_{GMV} \\ \hat{s}_c - s \end{pmatrix} \stackrel{d.}{=} \begin{pmatrix} \sqrt{1 + \frac{n}{n-1} \frac{\hat{s}_c + p/n}{1-p/n}} \sqrt{V_{GMV}} \sqrt{n}z_0 \\ \sqrt{n}(\hat{s}_c - s) \end{pmatrix},$$

where

$$\sqrt{n}\begin{pmatrix} z_0 \\ \hat{s}_c - s \end{pmatrix} \xrightarrow{d.} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \sigma_s^2 \end{pmatrix}\right),$$

with σ_s^2 as in (21). Now, the application of the δ -method (see, e.g. Theorem 3.7 in DasGupta 2008) and the fact that \hat{V}_{GMV} is independent of (\hat{R}_{GMV}, \hat{s}) leads to the statement of the theorem. ■