



Delft University of Technology

Is the empirical out-of-sample variance an informative risk measure for the high-dimensional portfolios?

Bodnar, Taras; Parolya, Nestor; Thorsén, Erik

DOI

[10.1016/j.frl.2023.103807](https://doi.org/10.1016/j.frl.2023.103807)

Publication date

2023

Document Version

Final published version

Published in

Finance Research Letters

Citation (APA)

Bodnar, T., Parolya, N., & Thorsén, E. (2023). Is the empirical out-of-sample variance an informative risk measure for the high-dimensional portfolios? *Finance Research Letters*, 54, Article 103807. <https://doi.org/10.1016/j.frl.2023.103807>

Important note

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

Copyright

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

Takedown policy

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



Is the empirical out-of-sample variance an informative risk measure for the high-dimensional portfolios?

Taras Bodnar^a, Nestor Parolya^{b,*}, Erik Thorsén^a

^a Department of Mathematics, Stockholm University, Stockholm, SE-10691, Sweden

^b Department of Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands

ARTICLE INFO

Keywords:

Shrinkage estimator
High-dimensional covariance matrix
Random matrix theory
Minimum variance portfolio
Parameter uncertainty

ABSTRACT

The main contribution of this paper is the derivation of the asymptotic behavior of the out-of-sample variance, the out-of-sample relative loss, and of their empirical counterparts in the high-dimensional setting, i.e., when both ratios p/n and p/m tend to some positive constants as $m \rightarrow \infty$ and $n \rightarrow \infty$, where p is the portfolio dimension, while n and m are the sample sizes from the in-sample and out-of-sample periods, respectively. The results are obtained for the traditional estimator of the global minimum variance (GMV) portfolio and for the two shrinkage estimators introduced by Frahm and Memmel (2010) and Bodnar et al. (2018). We show that the behavior of the empirical out-of-sample variance may be misleading in many practical situations, leading, for example, to a comparison of zeros. On the other hand, this will never happen with the empirical out-of-sample relative loss, which seems to provide a natural normalization of the out-of-sample variance in the high-dimensional setup. As a result, an important question arises if the out-of-sample variance can safely be used in practice for portfolios constructed from a large asset universe.

1. Introduction

Mean–variance analysis of Markowitz is a well established tool for optimal portfolio selection which is one of the most popular approaches today in financial literature (see, e.g., Markowitz (1952, 1959), Britten-Jones (1999), Ao et al. (2019), Bodnar et al. (2021b) and Ding et al. (2021)). The idea behind the approach is to invest in the portfolio which has the smallest variance for a given level of the expected return. In the limiting case of the fully risk-averse investor, the so-called global minimum variance (GMV) portfolio is selected. The latter portfolio possesses the smallest variance among all mean–variance optimal portfolios and lies on the vertex of the efficient frontier which is a parabola in the mean–variance space (see, Merton (1972), Kan and Smith (2008) and Bodnar and Schmid (2009)).

One of the important challenges, which arise when the Markowitz theory is implemented in practice, is related to the estimation error which appears when unknown parameters of the data-generating process are replaced by their sample counterparts in the expressions of the optimal portfolio weights (see, Okhrin and Schmid (2006), El Karoui (2010), Cai et al. (2020) and Bodnar et al. (2022a, 2023)). The impact of the parameter uncertainty on the performance of optimal portfolios is usually comparable to or even larger than the one described by the model uncertainty which is determined by using the covariance matrix in Markowitz optimization problem. Moreover, the estimation error present in an estimator of the mean vector has even a larger influence on the performance of optimal portfolios than the error related to the estimation of the covariance matrix (see, e.g., Merton (1980), Best and Grauer (1991) and Chopra and Ziemba (1993)). This is usually used in financial literature as an argument to hold the GMV

* Corresponding author.

E-mail address: n.parolya@tudelft.nl (N. Parolya).

portfolio whose weights only depend on the covariance matrix (see, Chan et al. (1999), Jagannathan and Ma (2003), Frahm and Memmel (2010), Bodnar et al. (2022b)).

Let \mathbf{y} denote the k -dimensional vector of the asset returns and let $\boldsymbol{\mu} = \mathbb{E}(\mathbf{y})$ and $\boldsymbol{\Sigma} = \text{Var}(\mathbf{y})$ be its mean vector and covariance matrix. Then the expected return and the variance of the portfolio with the weights \mathbf{w} are given by

$$R_p = \mathbf{w}^\top \boldsymbol{\mu} \quad \text{and} \quad V_p = \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w},$$

respectively. The weights of the GMV portfolio are found by minimizing V_p given that the whole investor wealth is invested in the selected assets, i.e., under the constraint $\mathbf{w}^\top \mathbf{1} = 1$ where $\mathbf{1}$ denotes the p -dimensional vector of ones. They are given by

$$\mathbf{w}_{GMV} = \frac{\boldsymbol{\Sigma}^{-1} \mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}}, \tag{1.1}$$

while the variance of the GMV portfolio is expressed as

$$V_{GMV} = \mathbf{w}_{GMV}^\top \boldsymbol{\Sigma} \mathbf{w}_{GMV} = \frac{1}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}}. \tag{1.2}$$

We refer to \mathbf{w}_{GMV} and V_{GMV} as the population weights and the population variance of the GMV portfolio, since they both depend on the unknown parameter $\boldsymbol{\Sigma}$ of the data-generating model. It has to be noted that V_{GMV} is also called the in-sample variance in financial literature (see, Frahm and Memmel (2010)).

In practical applications, the population GMV portfolio cannot be constructed since its weights \mathbf{w}_{GMV} depend on the unobservable quantity $\boldsymbol{\Sigma}$. Given historical realizations of the asset returns, $\mathbf{y}_1, \dots, \mathbf{y}_n$, the population covariance matrix is estimated by its sample counterpart expressed as

$$\mathbf{S}_n = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}}_n)(\mathbf{y}_i - \bar{\mathbf{y}}_n)^\top \quad \text{with} \quad \bar{\mathbf{y}}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i. \tag{1.3}$$

Then, the traditional GMV portfolio is determined as the sample estimator of \mathbf{w}_{GMV} where the unknown $\boldsymbol{\Sigma}$ is replaced by \mathbf{S}_n , i.e.,

$$\hat{\mathbf{w}}_{n:S} = \frac{\mathbf{S}_n^{-1} \mathbf{1}}{\mathbf{1}^\top \mathbf{S}_n^{-1} \mathbf{1}}. \tag{1.4}$$

If the portfolio dimension p is considerably smaller than the sample size n , then \mathbf{S}_n consistently estimates $\boldsymbol{\Sigma}$ under weak conditions imposed on the data-generating model of the asset returns and, consequently, the traditional GMV portfolio provides a good approximation of the population GMV portfolio.

The situation is completely different in the high-dimensional setting when the portfolio dimension is comparable to the sample size such that $p/n \rightarrow c \in [0, 1)$ as $n \rightarrow \infty$ where the constant c is called the concentration ratio (see, Bai and Silverstein (2010) and Bodnar et al. (2019a)). In this case the sample covariance matrix \mathbf{S}_n is not longer a consistent estimator for $\boldsymbol{\Sigma}$. As a result, the traditional GMV portfolio might deviate considerably from the population GMV portfolio. In order to ensure a good performance of the holding portfolio, the weights of the traditional GMV portfolio have to be adjusted by taking the parameter uncertainty into account (see, e.g., Jagannathan and Ma (2003), Glombek (2014), Ao et al. (2019), Bodnar et al. (2019b), Cai et al. (2020) and Ding et al. (2021)).

In order to define an improved estimator of the high-dimensional GMV portfolio in terms of out-of-sample variance,¹ the optimization problem has to be formulated. As a performance measure, the out-of-sample variance is usually used which is given by

$$V_{\hat{\mathbf{w}}_n} = \hat{\mathbf{w}}_n^\top \boldsymbol{\Sigma} \hat{\mathbf{w}}_n, \tag{1.5}$$

where $\hat{\mathbf{w}}_n$ is an estimator of \mathbf{w}_{GMV} based on the asset returns $\mathbf{y}_1, \dots, \mathbf{y}_n$. Alternatively, one can use the out-of-sample relative loss

$$L_{\hat{\mathbf{w}}_n} = \frac{V_{\hat{\mathbf{w}}_n} - V_{GMV}}{V_{GMV}} = \mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1} \hat{\mathbf{w}}_n^\top \boldsymbol{\Sigma} \hat{\mathbf{w}}_n - 1, \tag{1.6}$$

as a performance measure. By definitions of $V_{\hat{\mathbf{w}}_n}$ and $L_{\hat{\mathbf{w}}_n}$, one directly gets that the portfolio which minimizes the out-of-sample variance also minimizes the out-of-sample relative loss and vice versa.

Unfortunately, due to the presence of $\boldsymbol{\Sigma}$ in (1.5) and in (1.6), both the performance measures can only be used in theoretical derivations or in the comparison study based on the simulated data where the covariance matrix $\boldsymbol{\Sigma}$ is known. In practice, $\boldsymbol{\Sigma}$ is usually replaced by its estimator $\mathbf{S}_{n+1:m}$ constructed by using the asset returns $\mathbf{y}_{n+1}, \dots, \mathbf{y}_{n+m}$ from time $n+1$ to $n+m$ and defined by

$$\mathbf{S}_{n+1:n+m} = \frac{1}{m-1} \sum_{i=n+1}^{n+m} (\mathbf{y}_i - \bar{\mathbf{y}}_{n+1:n+m})(\mathbf{y}_i - \bar{\mathbf{y}}_{n+1:n+m})^\top \quad \text{with} \quad \bar{\mathbf{y}}_{n+1:n+m} = \frac{1}{m} \sum_{i=n+1}^{n+m} \mathbf{y}_i. \tag{1.7}$$

Consequently, the out-of-sample variance and the out-of-sample relative loss are replaced by the sample counterparts, the so-called empirical out-of-sample variance and the empirical out-of-sample relative loss expressed as

$$\hat{V}_{\hat{\mathbf{w}}_n:m} = \hat{\mathbf{w}}_n^\top \mathbf{S}_{n+1:n+m} \hat{\mathbf{w}}_n, \tag{1.8}$$

¹ Actually, what the literature typically uses as a measure to compare different estimators/strategies is the Sharpe ratio of the portfolio. But this case is much more complex and should be considered separately because of the influence of the sample mean. This important topic will be treated in a separate paper.

and

$$\hat{L}_{\hat{\mathbf{w}}_n;m} = \frac{\hat{V}_{\hat{\mathbf{w}}_n} - (1 - \tilde{c})^{-1} \hat{V}_{n+1:n+m;GMV}}{(1 - \tilde{c})^{-1} \hat{V}_{n+1:n+m;GMV}} = (1 - \tilde{c}) \mathbf{1}^\top \mathbf{S}_{n+1:m+n}^{-1} \mathbf{1} \hat{\mathbf{w}}_n^\top \mathbf{S}_{n+1:m+n} \hat{\mathbf{w}}_n - 1, \tag{1.9}$$

respectively, with $p/m \rightarrow \tilde{c}$ as $m \rightarrow \infty$. In (1.9), $(1 - \tilde{c}) \hat{V}_{n+1:n+m;GMV}$ is a consistent estimator for V_{GMV} in the high-dimensional setting (see, Lemma 1.3 in Bodnar et al. (2023)).

We contribute in this paper by deriving the asymptotic behavior of the out-of-sample variance, of the out-of-sample relative loss, and of their empirical counterparts in the high-dimensional setting, i.e., when $p/n \rightarrow c$ as $n \rightarrow \infty$ and $p/m \rightarrow \tilde{c}$ as $m \rightarrow \infty$. The results are obtained for the sample estimator (1.4) of the GMV portfolio (1.1) and for two shrinkage estimators introduced by Frahm and Memmel (2010) and Bodnar et al. (2018). We show that the empirical out-of-sample variance might tend to zero independently of chosen estimator of the GMV portfolio, which makes hard to distinguish between the estimators in practice. In contrast, the empirical out-of-sample losses of the considered estimators of the GMV portfolio tend to deterministic finite quantities. As such, a decision about the ranking of the estimators can be drawn. Moreover, one needs milder conditions for the derivation of the asymptotic properties of the empirical out-of-sample relative loss in comparison to the empirical out-of-sample variance, which is an additional advantage for the application of the former in practice.

The rest of the paper is structured as follows. In Section 2, the asymptotic behavior of the out-of-sample variance and of the out-of-sample relative loss is established for the traditional sample estimator and for the two shrinkage approaches. Section 3 presents the corresponding results in the case of the empirical performance measures. The results of a comprehensive simulation study are provided in Section 4, while the theoretical findings are implemented to real data in Section 5. Concluding remarks are drawn in Section 6. The proofs of the theoretical results are postponed to the supplementary material.

2. Out-of-sample variance and relative loss

Let the vector of asset returns, $\mathbf{y}_1, \dots, \mathbf{y}_n, \mathbf{y}_{n+1}, \dots, \mathbf{y}_{n+m}$ be independent and identically distributed with the following stochastic representation

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2} \mathbf{x}_t, \tag{2.1}$$

where the components of \mathbf{x}_t are independent and identically distributed with zero mean, unit variance, and finite $4 + \epsilon$ moments for some $\epsilon > 0$. No specific distributional assumptions are imposed on the components of \mathbf{x}_t . The symbol $\boldsymbol{\Sigma}^{1/2}$ denotes the square root of a positive definite matrix $\boldsymbol{\Sigma}$, i.e., $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2} (\boldsymbol{\Sigma}^{1/2})^\top$. Finally, we note that only $\mathbf{y}_t, t = 1, \dots, n + m$, are observable, while $\boldsymbol{\mu}, \boldsymbol{\Sigma}$, and $\mathbf{x}_t, t = 1, \dots, n + m$, are all unknown.

Depending on the performance measure different assumptions on the covariance matrix $\boldsymbol{\Sigma}$ and on the weights \mathbf{b} of the target portfolio are imposed. They are summarized as follows:

(A1) The variance of the GMV portfolio V_{GMV} as given in (1.2) is uniformly bounded in p .

(A2) The variance of the target portfolio $V_{\mathbf{b}} = \mathbf{b}^\top \boldsymbol{\Sigma} \mathbf{b}$ is uniformly bounded in p .

(A3) The relative loss of the target portfolio

$$L_{\mathbf{b}} = \frac{V_{\mathbf{b}} - V_{GMV}}{V_{GMV}} = \mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{b}^\top \boldsymbol{\Sigma} \mathbf{b} - 1,$$

is uniformly bounded in p .

The considered assumptions are very general and are fulfilled in many applications. For instance, all three assumptions are fulfilled when the eigenvalues of $\boldsymbol{\Sigma}$ are uniformly bounded in p and the Euclidean norm of the target vector \mathbf{b} is uniformly bounded in p . Assumptions (A1) and (A2) will be needed when the out-of-sample variance (1.5) and its empirical counterpart (1.8) are analyzed, while Assumption (A3) is only required in the case of the out-of-sample relative loss (1.6) and its empirical counterpart (1.9). This is not surprising, since the relative loss functions are already normalized and for that reason less restrictive assumptions are needed to study their asymptotic behavior. Furthermore, the normalization constant does not depend on an estimator of the GMV portfolio weights and thus, the normalization has no impact on the selected estimator.

Two shrinkage estimators for the GMV portfolio weights were derived in Frahm and Memmel (2010) and Bodnar et al. (2018), and they are given by

$$\hat{\mathbf{w}}_{n;FM} = \hat{\alpha}_{n;FM} \hat{\mathbf{w}}_{n;S} + (1 - \hat{\alpha}_{n;FM}) \mathbf{b} \tag{2.2}$$

with

$$\hat{\alpha}_{n;FM} = 1 - \frac{p - 3}{n - p + 2} (\mathbf{1}^\top \mathbf{S}_n^{-1} \mathbf{1} \mathbf{b}^\top \mathbf{S}_n \mathbf{b} - 1)^{-1}. \tag{2.3}$$

and

$$\hat{\mathbf{w}}_{n;BPS} = \hat{\alpha}_{n;BPS} \hat{\mathbf{w}}_{n;S} + (1 - \hat{\alpha}_{n;BPS}) \mathbf{b} \tag{2.4}$$

with

$$\hat{\alpha}_{n,BPS} = \frac{(1 - p/n) \left((1 - p/n) \mathbf{1}^\top \mathbf{S}_n^{-1} \mathbf{1} \mathbf{b}^\top \mathbf{S}_n \mathbf{b} - 1 \right)}{p/n + (1 - p/n) \left((1 - p/n) \mathbf{1}^\top \mathbf{S}_n^{-1} \mathbf{1} \mathbf{b}^\top \mathbf{S}_n \mathbf{b} - 1 \right)}, \tag{2.5}$$

respectively.

Next, we present the asymptotic behavior of the out-of-sample variance (Theorem 2.1) and of the out-of-sample relative loss (Theorem 2.2) calculated for the sample estimator $\hat{\mathbf{w}}_{n,S}$ of the GMV portfolio weights and for two shrinkage estimators $\hat{\mathbf{w}}_{n,FM}$ and $\hat{\mathbf{w}}_{n,BPS}$ in the high-dimensional setting. The proofs of the theorems are given in the supplementary material. To this end, we note that the out-of-sample variance and the out-of-sample loss of the target portfolio \mathbf{b} are, by definition, expressed as

$$V_{\mathbf{b}} = \mathbf{b}^\top \Sigma \mathbf{b} \tag{2.6}$$

and

$$L_{\mathbf{b}} = \frac{V_{\mathbf{b}}}{V_{GMV}} - 1 = \mathbf{1}^\top \Sigma^{-1} \mathbf{1} \mathbf{b}^\top \Sigma \mathbf{b} - 1, \tag{2.7}$$

respectively.

Theorem 2.1. Let $y_t, t = 1, \dots, n$ follow model (2.1). Then,

(i) under Assumption (A1), for the out-of-sample variance of the sample GMV portfolio $\hat{\mathbf{w}}_{n,S}$ it holds that

$$\left| V_{\hat{\mathbf{w}}_{n,S}} - (1 - c)^{-1} V_{GMV} \right| \xrightarrow{a.s.} 0, \tag{2.8}$$

(ii) under Assumptions (A1) and (A2), for the out-of-sample variance of the shrinkage GMV portfolio $\hat{\mathbf{w}}_{n,BPS}$ it holds that

$$\left| V_{\hat{\mathbf{w}}_{n,BPS}} - \left(V_{GMV} + \alpha_{BPS}^2 \frac{c}{1 - c} V_{GMV} + (1 - \alpha_{BPS})^2 (V_{\mathbf{b}} - V_{GMV}) \right) \right| \xrightarrow{a.s.} 0 \tag{2.9}$$

with

$$\alpha_{BPS} = \frac{(1 - c) L_{\mathbf{b}}}{c + (1 - c) L_{\mathbf{b}}}, \tag{2.10}$$

(iii) under Assumptions (A1) and (A2), for the out-of-sample variance of the shrinkage GMV portfolio $\hat{\mathbf{w}}_{n,FM}$ it holds that

$$\left| V_{\hat{\mathbf{w}}_{n,FM}} - \left(V_{GMV} + \alpha_{FM}^2 \frac{c}{1 - c} V_{GMV} + (1 - \alpha_{FM})^2 (V_{\mathbf{b}} - V_{GMV}) \right) \right| \xrightarrow{a.s.} 0 \tag{2.11}$$

with

$$\alpha_{FM} = 1 - \frac{c}{1 - c} \left((1 - c)^{-1} (L_{\mathbf{b}} + 1) - 1 \right)^{-1} = \frac{L_{\mathbf{b}}}{L_{\mathbf{b}} + c}, \tag{2.12}$$

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

Theorem 2.1 reveals an interesting fact that the asymptotic variance of the traditional estimator as well as other considered estimators of GMV portfolio are the linear combinations of V_{GMV} and $V_{\mathbf{b}} - V_{GMV}$. These quantities, however, could be very small if the number of assets is large. Indeed, taking for example V_{GMV} we can get the following upper bound using the trace inequality

$$V_{GMV} = (\mathbf{1}^\top \Sigma^{-1} \mathbf{1})^{-1} \leq (p \cdot \lambda_{\min}(\Sigma^{-1}))^{-1} = (p \cdot \lambda_{\max}^{-1}(\Sigma))^{-1} = \frac{\lambda_{\max}(\Sigma)}{p},$$

which obviously is vanishing for $p \rightarrow \infty$ in case the maximum eigenvalue of Σ , namely $\lambda_{\max}(\Sigma)$, is of smaller order than p . Similarly, this may happen for $V_{\mathbf{b}}$ by taking, for example, the naive portfolio $\mathbf{b} = \mathbf{1}/p$. Nevertheless, as the next theorem shows, it is not the case if the relative loss is taken instead.

Theorem 2.2. Let $y_t, t = 1, \dots, n$ follow model (2.1). Then,

(i) for the out-of-sample relative loss of the sample GMV portfolio $\hat{\mathbf{w}}_{n,S}$ it holds that

$$\left| L_{\hat{\mathbf{w}}_{n,S}} - \frac{c}{1 - c} \right| \xrightarrow{a.s.} 0, \tag{2.13}$$

(ii) under Assumption (A3), for the out-of-sample relative loss of the shrinkage GMV portfolio $\hat{\mathbf{w}}_{n,BPS}$ it holds that

$$\left| L_{\hat{\mathbf{w}}_{n,BPS}} - \left(\alpha_{BPS}^2 \frac{c}{1 - c} + (1 - \alpha_{BPS})^2 L_{\mathbf{b}} \right) \right| \xrightarrow{a.s.} 0, \tag{2.14}$$

(iii) under Assumption (A3), for the out-of-sample relative loss of the shrinkage GMV portfolio $\hat{\mathbf{w}}_{n,FM}$ it holds that

$$\left| L_{\hat{\mathbf{w}}_{n,FM}} - \left(\alpha_{FM}^2 \frac{c}{1 - c} + (1 - \alpha_{FM})^2 L_{\mathbf{b}} \right) \right| \xrightarrow{a.s.} 0, \tag{2.15}$$

for $p/n \rightarrow c \in (0, 1)$ as $n \rightarrow \infty$ where α_{BPS} and α_{FM} are given in (2.10) and (2.12), respectively.

The findings of [Theorem 2.2](#) show that the relative loss of shrinkage portfolios is present as a linear combination of the relative loss of the corresponding target portfolio and of the limiting relative loss of the traditional GMV portfolio. The relative loss of the traditional GMV portfolio $\hat{w}_{n,S}$ tends to a constant $c/(1 - c)$ that does not depend on the covariance matrix of the asset returns. Moreover, if c tends to 1, then the relative loss of the traditional GMV portfolio tends to infinity showing that the impact of the estimation error could be drastically large in the high-dimensional setting. Furthermore, using [\(2.10\)](#) and [\(2.12\)](#) the limiting values of relative loss computed for two shrinkage estimators can be rewritten as

$$\alpha_{FM}^2 \frac{c}{1 - c} + (1 - \alpha_{FM})^2 L_b = \frac{L_b^2}{(c + L_b)^2} \frac{c}{1 - c} + \frac{c^2}{(c + L_b)^2} L_b \tag{2.16}$$

for the shrinkage estimator of [Frahm and Memmel \(2010\)](#) and

$$\alpha_{BPS}^2 \frac{c}{1 - c} + (1 - \alpha_{BPS})^2 L_b = \frac{(1 - c)L_b^2}{(c + (1 - c)L_b)^2} c + \frac{c^2}{(c + (1 - c)L_b)^2} L_b \tag{2.17}$$

for the shrinkage estimator of [Bodnar et al. \(2018\)](#). As a result, expressions [\(2.16\)](#) and [\(2.17\)](#) show that the out-of-sample relative loss of the shrinkage estimator [\(2.2\)](#) tends to infinity as c approaches one, similarly to the traditional estimator $\hat{w}_{n,S}$, while the out-of-sample relative loss of the shrinkage estimator [\(2.4\)](#) tends to the relative loss of the target portfolio when c tends to one.

The results of [Theorem 2.2](#) lead also to some dominance statements presented in [Corollary 2.3](#) in terms of the out-of-sample relative loss. Due to the relationship between the out-of-sample variance and the out-of-sample loss the same statements also hold for the out-of-sample variance by using the findings of [Theorem 2.1](#).

Corollary 2.3. *Let $y_t, t = 1, \dots, n$ follow model [\(2.1\)](#). Then, under Assumption [\(A3\)](#) it holds that*

(i)

$$L_{\hat{w}_{n,S}} - L_{\hat{w}_{n,FM}} \xrightarrow{a.s.} \frac{c^2(c + L_b + cL_b)}{(1 - c)(c + L_b)^2} \geq 0, \text{ for } \frac{p}{n} \rightarrow c \in (0, 1) \text{ as } n \rightarrow \infty,$$

with equality if and only if $c = 0$ or $L_b = \infty$, i.e., when the sample size is considerably larger than the portfolio dimension or the target portfolio deviates too strong from the true GMV portfolio;

(ii)

$$L_{\hat{w}_{n,S}} - L_{\hat{w}_{n,BPS}} \xrightarrow{a.s.} \frac{c^2}{(1 - c)(c + (1 - c)L_b)} \geq 0 \text{ for } \frac{p}{n} \rightarrow c \in (0, 1) \text{ as } n \rightarrow \infty,$$

with equality if and only if $c = 0$ or $L_b = \infty$, i.e., when the sample size is considerably larger than the portfolio dimension or the target portfolio deviates too strong from the true GMV portfolio;

(iii)

$$L_{\hat{w}_{n,FM}} - L_{\hat{w}_{n,BPS}} \xrightarrow{a.s.} \frac{c^4 L_b^2}{(1 - c)(c + L_b)^2(c + (1 - c)L_b)} \geq 0 \text{ for } \frac{p}{n} \rightarrow c \in (0, 1) \text{ as } n \rightarrow \infty,$$

with equality if and only if $c = 0$ or $c > 0, L_b = 0$ or $c > 0, L_b = \infty$, i.e., when the target portfolio coincides with the true GMV portfolio or the target portfolio deviates too strong from the true GMV portfolio when the concentration ratio is positive.

The findings of [Corollary 2.3](#) show that the shrinkage estimator of [Bodnar et al. \(2018\)](#) outperforms the other two estimators, while the shrinkage estimator of [Frahm and Memmel \(2010\)](#) is always better than the sample estimator $\hat{w}_{n,S}$. The exception is present when the sample size n is considerably larger than the portfolio dimension p such that the concentration ratio is equal to zero or when the target portfolio is very poorly chosen such that its relative loss is infinity. In the latter situation, the investor might consider a different target portfolio in order to get the advantage of the shrinkage approaches over the sample estimator. Interestingly, when the target portfolio coincides with the population GMV portfolio, then both shrinkage estimators perform similarly.

3. Empirical out-of-sample variance and relative loss

The out-of-sample variance is one of the performance measures which are widely used to compare trading strategies between each other. However, it cannot directly be applied in practice, since its definition depends on the unknown population covariance matrix Σ . As a result, different portfolio strategies are compared between each other based on the empirical counterparts of the out-of-sample performance measures as presented in [\(1.8\)](#) and [\(1.9\)](#), respectively, where the sample of the asset returns y_{n+1}, \dots, y_{n+m} is used to construct an estimator of the covariance matrix denoted by $S_{n+1:n+m}$ as in [\(1.7\)](#). In this section we show that the empirical out-of-sample variance can be an improper performance measure in the high-dimensional setting, since it may tend to zero independently of the used trading strategy due to the dimensionality effect. As a solution, the empirical out-of-sample relative loss is suggested and asymptotic properties are derived.

In [Theorems 3.1](#) and [3.2](#) we derive the asymptotic properties of the empirical out-of-sample variance and of the empirical out-of-sample relative loss computed for the four portfolios discussed in Section 2. The proofs of the theorems are presented in the supplement. It is remarkable that the results of [Theorems 3.1](#) and [3.2](#) are deduced under the same conditions as given in the statements of [Theorems 2.1](#) and [2.2](#), even though additional randomness is taken into account in the derivations of the results. Moreover, both the empirical out-of-sample variances and the out-of-sample relative losses converge to the same limiting values as given in [Theorems 2.1](#) and [2.2](#).

Theorem 3.1. Let $y_t, t = 1, \dots, n + m$ follow model (2.1). Then,

(i) under Assumption (A1), for the empirical out-of-sample variance of the sample GMV portfolio $\hat{w}_{n,S}$ it holds that

$$\left| \hat{V}_{\hat{w}_{n,S};m} - (1 - c)^{-1} V_{GMV} \right| \xrightarrow{a.s.} 0, \tag{3.1}$$

(ii) under Assumption (A2), for the empirical out-of-sample variance of the target portfolio \mathbf{b} it holds that

$$\left| \hat{V}_{\mathbf{b};m} - V_{\mathbf{b}} \right| \xrightarrow{a.s.} 0, \tag{3.2}$$

(iii) under Assumptions (A1) and (A2), for the empirical out-of-sample variance of the shrinkage GMV portfolio $\hat{w}_{n,BPS}$ it holds that

$$\left| \hat{V}_{\hat{w}_{n,BPS};m} - \left(V_{GMV} + \alpha_{BPS}^2 \frac{c}{1 - c} V_{GMV} + (1 - \alpha_{BPS})^2 (V_{\mathbf{b}} - V_{GMV}) \right) \right| \xrightarrow{a.s.} 0, \tag{3.3}$$

with α_{BPS} as in (2.10),

(iv) under Assumptions (A1) and (A2), for the empirical out-of-sample variance of the shrinkage GMV portfolio $\hat{w}_{n,FM}$ it holds that

$$\left| \hat{V}_{\hat{w}_{n,FM};m} - \left(V_{GMV} + \alpha_{FM}^2 \frac{c}{1 - c} V_{GMV} + (1 - \alpha_{FM})^2 (V_{\mathbf{b}} - V_{GMV}) \right) \right| \xrightarrow{a.s.} 0, \tag{3.4}$$

with α_{FM} as in (2.12),

for $p/n \rightarrow c \in (0, 1)$ and $p/m \rightarrow \tilde{c} \in (0, \infty)$ as $n \rightarrow \infty$.

Theorem 3.2. Let $y_t, t = 1, \dots, n + m$ follow model (2.1). Then,

(i) under Assumption (A3), for the empirical out-of-sample relative loss of the sample GMV portfolio $\hat{w}_{n,S}$ it holds that

$$\left| \hat{L}_{\hat{w}_{n,S};m} - \frac{c}{1 - c} \right| \xrightarrow{a.s.} 0, \tag{3.5}$$

(ii) under Assumption (A3), for the empirical out-of-sample relative loss of the target portfolio \mathbf{b} it holds that

$$\left| \hat{L}_{\mathbf{b};m} - L_{\mathbf{b}} \right| \xrightarrow{a.s.} 0, \tag{3.6}$$

(iii) under Assumptions (A3), for the empirical out-of-sample relative loss of the shrinkage GMV portfolio $\hat{w}_{n,BPS}$ it holds that

$$\left| \hat{L}_{\hat{w}_{n,BPS};m} - \left(\alpha_{BPS}^2 \frac{c}{1 - c} + (1 - \alpha_{BPS})^2 L_{\mathbf{b}} \right) \right| \xrightarrow{a.s.} 0, \tag{3.7}$$

with α_{BPS} as in (2.10),

(iv) under Assumptions (A3), for the empirical out-of-sample relative loss of the shrinkage GMV portfolio $\hat{w}_{n,FM}$ it holds that

$$\left| \hat{L}_{\hat{w}_{n,FM};m} - \left(\alpha_{FM}^2 \frac{c}{1 - c} + (1 - \alpha_{FM})^2 L_{\mathbf{b}} \right) \right| \xrightarrow{a.s.} 0, \tag{3.8}$$

with α_{FM} as in (2.12),

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

Since the empirical out-of-sample losses $\hat{L}_{\hat{w}_{n,S};m}$, $\hat{L}_{\hat{w}_{n,BPS};m}$, and $\hat{L}_{\hat{w}_{n,FM};m}$ possess the same high-dimensional asymptotic behavior as the corresponding out-of-sample losses $L_{\hat{w}_{n,S}}$, $L_{\hat{w}_{n,BPS}}$, and $L_{\hat{w}_{n,FM}}$ in Theorem 2.2, the results of Corollary 2.3 remain also valid. Namely, we get

Corollary 3.3. Let $y_t, t = 1, \dots, n + m$ follow model (2.1). Then, under Assumption (A3) it holds that

(i)

$$\hat{L}_{\hat{w}_{n,S};m} - \hat{L}_{\hat{w}_{n,FM};m} \xrightarrow{a.s.} \frac{c^2(c + L_{\mathbf{b}} + cL_{\mathbf{b}})}{(1 - c)(c + L_{\mathbf{b}})^2} \geq 0$$

for $p/n \rightarrow c \in (0, 1)$, $p/m \rightarrow \tilde{c} \in (0, 1)$ as $n, m \rightarrow \infty$, with equality if and only if $c = 0$ or $L_{\mathbf{b}} = \infty$, i.e., when the sample size is considerably larger than the portfolio dimension or the target portfolio deviates too strong from the true GMV portfolio;

(ii)

$$\hat{L}_{\hat{w}_{n,S};m} - \hat{L}_{\hat{w}_{n,BPS};m} \xrightarrow{a.s.} \frac{c^2}{(1 - c)(c + (1 - c)L_{\mathbf{b}})} \geq 0$$

for $p/n \rightarrow c \in (0, 1)$, $p/m \rightarrow \tilde{c} \in (0, 1)$ as $n, m \rightarrow \infty$, with equality if and only if $c = 0$ or $L_{\mathbf{b}} = \infty$, i.e., when the sample size is considerably larger than the portfolio dimension or the target portfolio deviates too strong from the true GMV portfolio;

(iii)

$$\hat{L}_{\hat{w}_{n,FM};m} - \hat{L}_{\hat{w}_{n,BPS};m} \xrightarrow{a.s.} \frac{c^4 L_{\mathbf{b}}^2}{(1 - c)(c + L_{\mathbf{b}})^2(c + (1 - c)L_{\mathbf{b}})} \geq 0$$

for $p/n \rightarrow c \in (0, 1)$, $p/m \rightarrow \tilde{c} \in (0, 1)$ as $n, m \rightarrow \infty$, with equality if and only if $c = 0$ or $c > 0$, $L_b = 0$ or $c > 0$, $L_b = \infty$, i.e., when the target portfolio coincides with the true GMV portfolio or the target portfolio deviates too strong from the true GMV portfolio when the concentration ratio is positive.

Corollary 3.3 provides the limiting behavior of the differences of the empirical out-of-sample losses and, consequently, the same ranking between the three estimators of the GMV portfolio weights as previously obtained in Corollary 2.3. Furthermore, the difference between the asymptotic behavior of the three estimators is negligible only when the concentration ratio is zero, i.e., the portfolio size is considerably smaller than the sample size, or when the target portfolio is poorly chosen such that its relative loss becomes infinity.

4. Simulation study

In this section we will investigate the finite sample behavior of the high-dimensional asymptotic results presented in Corollary 3.3 via an extensive Monte Carlo study. The aim of the study is twofold: (i) first, we investigate how fast the difference of the empirical out-of-sample relative loss functions tend to the corresponding limiting value provided in the statement of Corollary 3.3; (ii) second, we study the impact of the presence of non-linear time dependence in the data-generating model on the performance of the three considered trading strategies.

For each fixed value of the portfolio size p we first simulated the elements of the mean vector μ as $\mu_i \sim U(-0.1, 0.1)$, $i = 1, 2, \dots, p$ and the elements of the covariance matrix Σ using the RandCovMtrx function from the HDSHOP package (Bodnar et al., 2021a). Then these values were used in simulating samples of the asset returns from the following data-generating models:

Scenario 1: t -distribution The elements of x_t are drawn independently from the t -distribution with 5 degrees of freedom, that is $x_{tj} \sim t(5)$ for $j = 1, \dots, p$, while y_t is constructed according to (2.1). Moreover since the variance of the t -distribution with 5 degrees of freedom is equal to $5/3$ we, additionally multiply the vector x_t in (2.1) by $\sqrt{3/5}$. As such, all $\sqrt{3/5}x_{tj}$ have mean zero and variance one.

Scenario 2: CAPM The vector of asset returns y_t is generated according to the CAPM (Capital Asset Pricing Model), i.e.,

$$y_t = \mu + \beta z_t + \Sigma^{1/2} x_t,$$

with independently distributed $z_t \sim N(0, 1)$ and $x_t \sim N_p(\mathbf{0}, \mathbf{I})$. The elements of vector β are drawn from the uniform distribution, that is $\beta_i \sim U(-1, 1)$ for $i = 1, \dots, p$.

Scenario 3: CCC-GARCH model of Bollerslev (1990) The asset returns are simulated according to

$$y_t | \Sigma_t \sim N_p(\mu, \Sigma_t)$$

where the conditional covariance matrix is specified by

$$\Sigma_t = \mathbf{D}_t^{1/2} \mathbf{C} \mathbf{D}_t^{1/2} \quad \text{with} \quad \mathbf{D}_t = \text{diag}(h_{1,t}, h_{2,t}, \dots, h_{p,t}),$$

with

$$h_{j,t} = \alpha_{j,0} + \alpha_{j,1}(y_{j,t-1} - \mu_j)^2 + \beta_{j,1}h_{j,t-1}, \quad \text{for } j = 1, 2, \dots, p, \quad \text{and } t = 1, 2, \dots, n + m.$$

The coefficients of the CCC-GARCH model are generated by $\alpha_{j,1} \sim U(0, 0.1)$ and $\beta_{j,1} \sim U(0.6, 0.7)$ which implies that the stationarity conditions, $\alpha_{j,1} + \beta_{j,1} < 1$, are always fulfilled. The intercepts $\alpha_{j,0}$, $j = 1, \dots, p$ is thereafter chosen such that the unconditional covariance matrix is equal to Σ .

The model under scenario 1 fulfills the assumptions imposed in Section 2 by drawing the vector x_t independently each of other. Scenario 2 considers a common model in asset pricing, namely the CAPM model. In this scenario, the tails are less fat than Scenario 1, but the largest eigenvalue of the covariance matrix is of order p . Furthermore, CAPM has the potential to capture many different market regimes and structures, such as cyclical patterns through its conditional distribution. In contrast to these two, scenario 3 possesses some time dependence structure, thus violating the assumption imposed on the data-generating model in Section 2. Here a complicated non-linear time dependence structure is assumed in scenario 3 which is accompanied with conditionally time-dependent covariance matrix Σ_t . This model captures cyclical patterns seen in market data, such as volatility clustering. Finally, the equally weighted portfolio is used as a target portfolio in all scenarios.

In Figs. 1 to 3 we present the relative differences of empirical out-of-sample losses as considered in Corollary 3.3 divided by the corresponding asymptotic limit determined for each difference in the statement of the corollary in the right hand-side of each inequality. For each scenario we set $n = \{100, 250, 500, 750, 1000\}$, $c = \{0.5, 0.9\}$ and $\tilde{c} = \{0.5, 0.9\}$. The portfolio size p and the sample size m are thereafter determined by $p = nc$ and in turn $m = p/\tilde{c}$. If necessary we round to the closest integer. The results in the figures are based on the 1000 independent repetitions and present the corresponding average values.

Fig. 1 depicts the results of the simulation study obtained under scenario 1. The relative differences in the empirical out-of-sample losses converge quickly to one, indicating that the results of Corollary 3.3 may also be used when samples of asset returns of moderate size are used. As expected, the fastest convergence is observed in the case $c = \tilde{c} = 0.5$, while the largest deviations from one is present in the case of $c = 0.5$ and $\tilde{c} = 0.9$, when the sample size is small. Finally, we note that all computed values in the plots

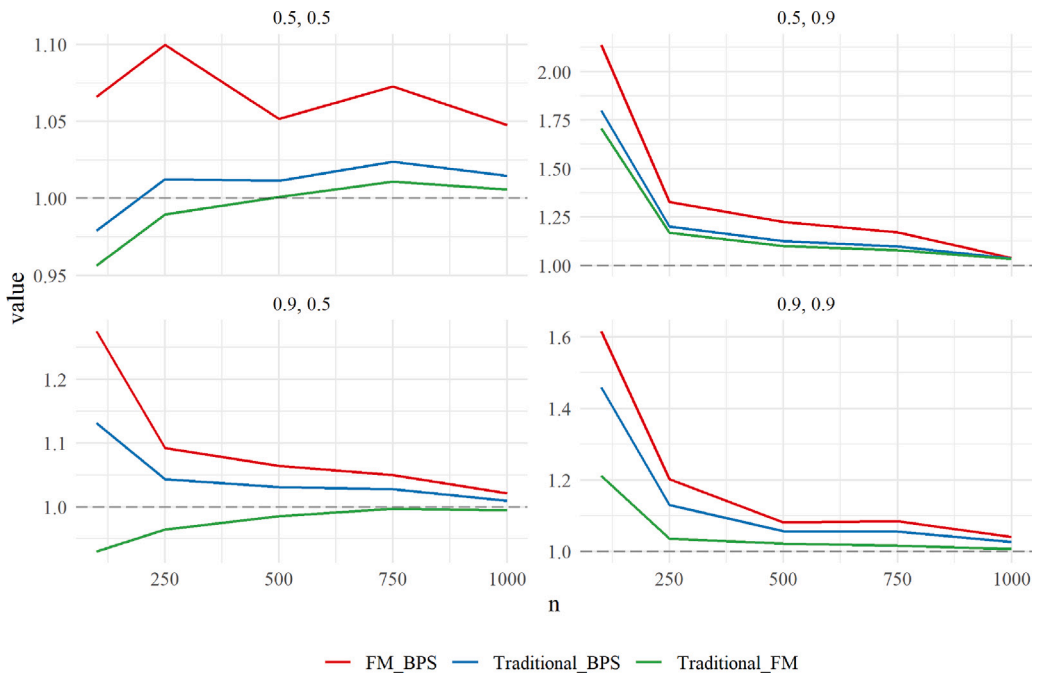


Fig. 1. Relative differences in the empirical out-of-sample losses divided by the corresponding asymptotic limit as given in Corollary 3.3 for $n = \{100, 250, 500, 750, 1000\}$, $c = \{0.5, 0.9\}$ and $\bar{c} = \{0.5, 0.9\}$. The samples of asset returns are drawn following scenario 1.

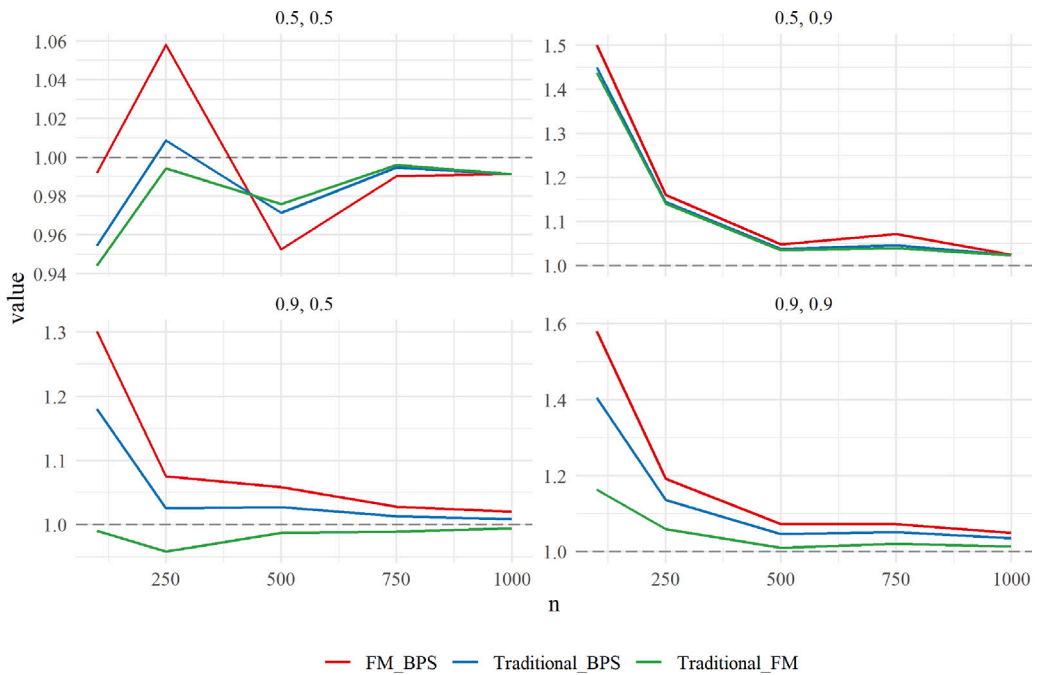


Fig. 2. Relative differences in the empirical out-of-sample losses divided by the corresponding asymptotic limit as given in Corollary 3.3 for $n = \{100, 250, 500, 750, 1000\}$, $c = \{0.5, 0.9\}$ and $\bar{c} = \{0.5, 0.9\}$. The samples of asset returns are drawn following scenario 2.

are positive and, as such, the shrinkage estimator of Bodnar et al. (2018) outperforms the other two trading strategies followed by the shrinkage approach of Frahm and Memmel (2010) in all of the considered cases.

In Fig. 2 the results of the simulation study obtained under scenario 2 are presented. All values are positive, which is in line with the ordering presented in the previous section. At first glance, the results seem slightly more volatile for the upper left hand

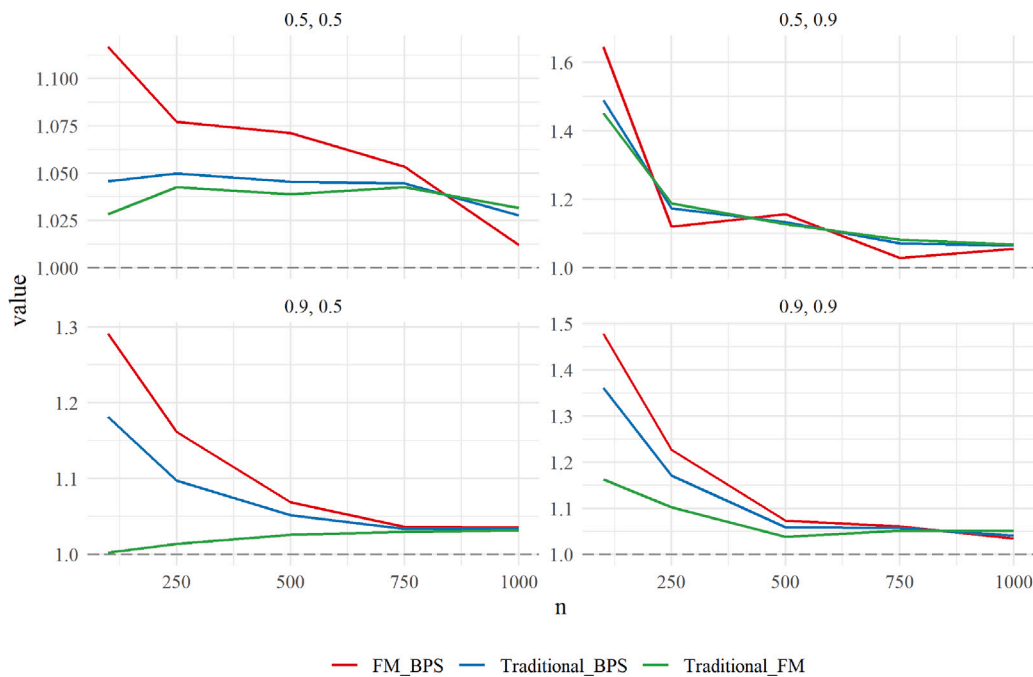


Fig. 3. Relative differences in the empirical out-of-sample losses divided by the corresponding asymptotic limit as given in Corollary 3.3 for $n = \{100, 250, 500, 750, 1000\}$, $c = \{0.5, 0.9\}$ and $\tilde{c} = \{0.5, 0.9\}$. The samples of asset returns are drawn following scenario 3.

figure, i.e. $c = \tilde{c} = 0.5$, though that is simply a matter of scale. As c and \tilde{c} grow, all simulations illustrate the pattern as in scenario 1.

Fig. 3 illustrates the results of the simulation study under the third scenario. In this setting, the returns are simulated from a CCC-GARCH model which captures volatility clustering and also introduces a non-linear time dependence structure in the vectors of the asset returns. Here the relative differences do not converge to one, although the departure from one is not considerably large. Moreover, the relative losses converge to the values which are larger one and the computed values are all positive. As such, the ranking between the three trading strategies is preserved and one can also use the expression of the limiting values of Corollary 3.3 as the corresponding lower bounds for the differences under the assumption of the CCC-GARCH model.

5. Empirical illustration

In the empirical application we use 10 years of daily data for 100 and 190 stocks included in the S&P500 index from the first of June 2011 to the seventh of January 2021. During the considered period of time, 380 stocks were continuously included in the S&P500 index from which we randomly choose 100 and 190 stocks to build the GMV portfolio. The first $n = 200$ observations were used to estimate the weights of the GMV portfolio by employing the traditional estimator and the two shrinkage estimators introduced in Section 2, while the next $m = 200$ observations were used to compute the values of the empirical out-of-sample variances and the empirical out-of-sample relative losses for each trading strategy. Then, using the rolling window approach the same computations, related to the short-term prediction of risk measures, are subsequently performed over the time period from the fourteenth of February, 2013 to the seventh of January 2021. As a target portfolio in the construction of the two shrinkage estimator, the equally weighted portfolio was used.

Fig. 4 depicts the values of the empirical out-of-sample variances and of the empirical out-of-sample relative losses computed for three estimators of the GMV portfolio considered in the paper. The result are presented for two portfolio sizes which correspond to $c = \tilde{c} = 0.5$ and $c = \tilde{c} = 0.95$. A considerable increase in both the empirical out-of sample variances and losses of each estimator is observed in March 2020 which corresponds to the crisis on international financial market caused by the beginning of COVID-19 spread over the world. The rapid increase of volatility is more pronounced in the case of the smaller dimensional portfolio, i.e., when $p = 100$. In the case of the portfolio which is based on $p = 190$ stocks the jump in the values of the two considered performance measures is smoothed due to higher variability of these two measures presented during the whole period of observation. Another rapid increase in the loss functions for $p = 100$ occurs in late December 2020. This date can be related to the second wave of the COVID-19 spread. Similar increases in the behavior of the relative loss function are also present for the portfolio consisting of $p = 190$ stocks, although they are somehow hidden by the more volatile behavior of the loss function in the latter case.

In general, the results in Fig. 4 confirm the ordering of the three trading strategies which is deduced in Corollary 3.3 and confirmed in the finite-sample case in the simulation study of Section 4. Namely, the shrinkage estimator of Bodnar et al. (2018)

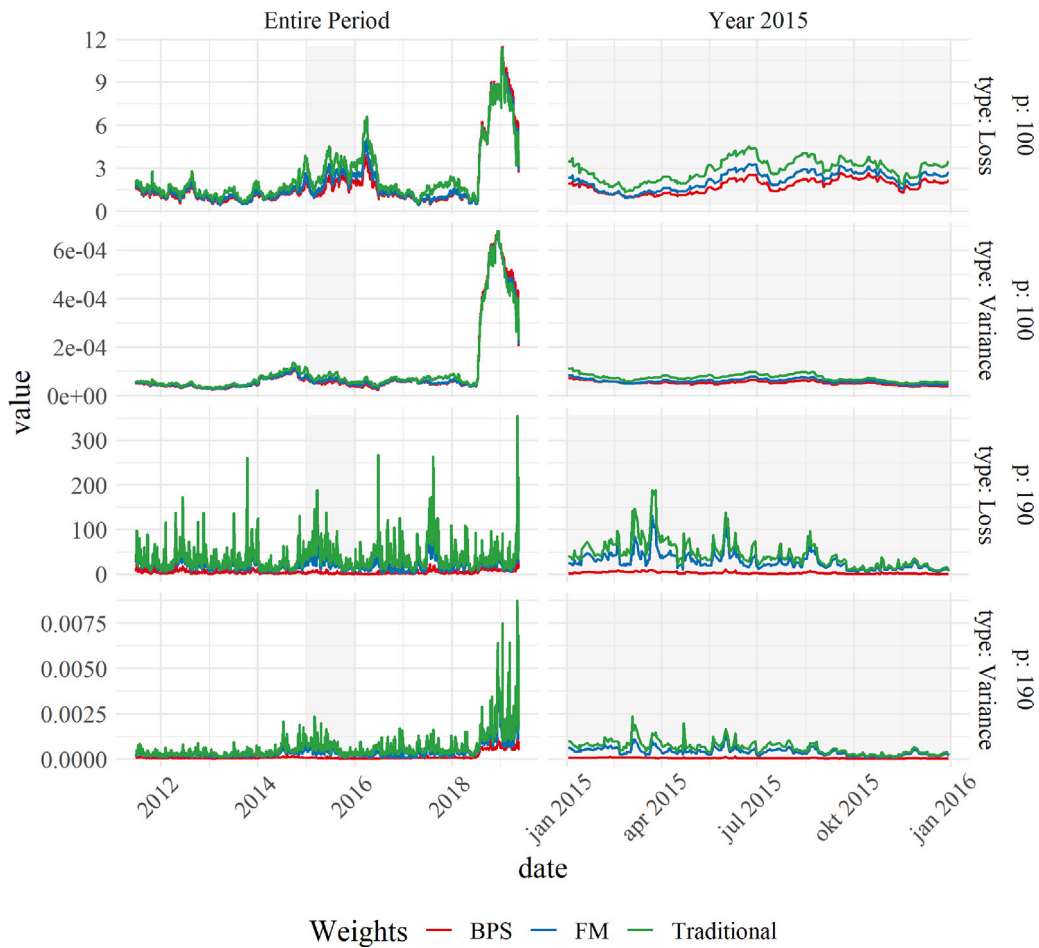


Fig. 4. Empirical out-of-sample variance and out-of-sample relative loss of the traditional GMV portfolio and the two shrinkage estimators based on the rolling window approach with window size equal 200 and computed for two portfolios which consist of 100 and 190 stocks traded in the S&P 500 index.

shows the smallest values of both the empirical out-of-sample variance and the empirical out-of-sample relative loss, while the shrinkage estimator of [Frahm and Memmel \(2010\)](#) is ranked on the second place. On the other side, when the empirical out-of-sample variance is used as a performance measure, the distinction between the strategies become visually negligible in almost all cases presented for $p = 100$ and in majority of cases when the portfolio with $p = 190$ is constructed. This empirical finding can be explained by noting that most of the values of the empirical out-of-sample variance were computed during the stable period on the capital market and as such, the true value of the global minimum variance was very small at that time. In contrast, the usage of the empirical out-of-sample loss can lead to the obvious conclusion about the performance of each of the considered three trading strategies. Finally, the impact of portfolio dimensionality which is accompanied with a huge amount of estimation error becomes more pronounced when the empirical relative loss is used, especially during the turbulent period on the capital market.

6. Summary

The sample variance of the GMV portfolio is known to be biased and to significantly underestimate the true population variance of this portfolio, especially when the portfolio size is comparable to the sample size. In many practical situations it is not a good measure for the portfolio performance and the out-of-sample variance is usually used instead. However, its value could be very close to zero in a high-dimensional asset universe. That is why many researchers use its annualized version by multiplying the out-of-sample variance by the square root of periods (see, e.g., [Zhao et al., 2021](#)). Still, it is not clear what is the natural normalization of risk, when big number of assets is involved. The only way to find this out is to consistently estimate the risk of the true GMV portfolio and construct a unit-free measure of the out-of-sample variance, i.e., its relative loss with respect to the true risk.

For this purpose we derive the asymptotic properties of the out-of-sample variance and of the out-of-sample relative loss as well as of their empirical counterparts. Under weak conditions imposed on the data-generating model it is shown that the out-of-sample variance and the empirical out-of-sample variance might tend to zero independently of chosen estimator of the GMV portfolio

weights, which can make the comparison between the trading strategies intractable. This is not, however, an issue when the out-of-sample relative loss and the empirical out-of-sample relative loss are used instead. In the latter case a clear ordering between the estimators of the three considered estimators can be made.

As a by-product of the derived theoretical findings, we also prove by way of example that the shrinkage estimator of Bodnar et al. (2018) outperforms the shrinkage estimator of Frahm and Memmel (2010) and the traditional estimator of the GMV portfolio. Moreover, we quantify the difference in the performance of the three trading strategies by deducing the asymptotic difference of their empirical out-of-sample relative loss functions. Within a comprehensive numerical study it is shown that the derived asymptotic limits can still be used when the sample of moderate size is present and when the asset returns possess a non-linear time dependence structure. As a result a new measure of risk, i.e., the out-of-sample relative loss, could be used by practitioners, which is always bounded away from zero and infinity, and provides a natural ordering between investment strategies in terms of their out-of-sample performance.

CRedit authorship contribution statement

Taras Bodnar: Conceptualization, Methodology. **Nestor Parolya:** Conceptualization, Methodology. **Erik Thorsén:** Methodology, Software, Formal analysis.

Data availability

The data is free available on the Yahoo-Finance.

Acknowledgment

The authors thank Professor Samuel Vigne and the anonymous Reviewer for their comments and suggestions which have improved the presentation of the paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.frl.2023.103807>.

References

- Ao, M., Yingying, L., Zheng, X., 2019. Approaching mean-variance efficiency for large portfolios. *Rev. Financ. Stud.* 32 (7), 2890–2919.
- Bai, Z., Silverstein, J.W., 2010. *Spectral Analysis of Large Dimensional Random Matrices*. Springer.
- Best, M.J., Grauer, R.R., 1991. On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results. *Rev. Financ. Stud.* 4 (2), 315–342.
- Bodnar, O., Bodnar, T., Parolya, N., 2022a. Recent advances in shrinkage-based high-dimensional inference. *J. Multivariate Anal.* 188, 104826.
- Bodnar, T., Dette, H., Parolya, N., 2019a. Testing for independence of large dimensional vectors. *Ann. Statist.* 47 (5), 2977–3008.
- Bodnar, T., Dette, H., Parolya, N., Thorsén, E., 2022b. Sampling distributions of optimal portfolio weights and characteristics in low and large dimensions. *Random Matrices: Theory Appl.* 11, 2250008.
- Bodnar, T., Dmytriv, S., Okhrin, Y., Otryakhin, D., Parolya, N., 2021a. HDSHOP: High-dimensional shrinkage optimal portfolios. R package version 0.1.1.
- Bodnar, T., Dmytriv, S., Okhrin, Y., Parolya, N., Schmid, W., 2021b. Statistical inference for the expected utility portfolio in high dimensions. *IEEE Trans. Signal Process.* 69, 1–14.
- Bodnar, T., Dmytriv, S., Parolya, N., Schmid, W., 2019b. Tests for the weights of the global minimum variance portfolio in a high-dimensional setting. *IEEE Trans. Signal Process.* 67 (17), 4479–4493.
- Bodnar, T., Okhrin, Y., Parolya, N., 2023. Optimal shrinkage-based portfolio selection in high dimensions. *J. Bus. Econom. Statist.* 41, 140–156.
- Bodnar, T., Parolya, N., Schmid, W., 2018. Estimation of the global minimum variance portfolio in high dimensions. *European J. Oper. Res.* 266 (1), 371–390.
- Bodnar, T., Schmid, W., 2009. Econometrical analysis of the sample efficient frontier. *Eur. J. Finance* 15 (3), 317–335.
- Bollerslev, T., 1990. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *Rev. Econ. Stat.* 72 (3), 498–505.
- Britten-Jones, M., 1999. The sampling error in estimates of mean-variance efficient portfolio weights. *J. Finance* 54 (2), 655–671.
- Cai, T.T., Hu, J., Li, Y., Zheng, X., 2020. High-dimensional minimum variance portfolio estimation based on high-frequency data. *J. Econometrics* 214 (2), 482–494.
- Chan, L.K., Karceski, J., Lakonishok, J., 1999. On portfolio optimization: Forecasting covariances and choosing the risk model. *Rev. Financ. Stud.* 12 (5), 937–974.
- Chopra, V.K., Ziemba, W.T., 1993. The effect of errors in means, variances, and covariances on optimal portfolio choice. *J. Portf. Manag.* 19 (2), 6–11.
- Ding, Y., Li, Y., Zheng, X., 2021. High dimensional minimum variance portfolio estimation under statistical factor models. *J. Econometrics* 222 (1), 502–515.
- El Karoui, N., 2010. High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: Risk underestimation. *Ann. Statist.* 38 (6), 3487–3566.
- Frahm, G., Memmel, C., 2010. Dominating estimators for minimum-variance portfolios. *J. Econometrics* 159, 289–302.
- Glombek, K., 2014. Statistical inference for high-dimensional global minimum variance portfolios. *Scand. J. Stat.* 41 (4), 845–865.
- Jagannathan, R., Ma, T., 2003. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *J. Finance* 58 (4), 1651–1683.
- Kan, R., Smith, D.R., 2008. The distribution of the sample minimum-variance frontier. *Manage. Sci.* 54 (7), 1364–1380.
- Markowitz, H., 1952. Portfolio selection. *J. Finance* 7 (1), 77–91.
- Markowitz, H., 1959. *Portfolio Selection: Efficient Diversification of Investments*. John Wiley & Sons, Inc and Chapman & Hall, Ltd.
- Merton, R.C., 1972. An analytic derivation of the efficient Portfolio frontier. *J. Final. Quant. Anal.* 7 (4), 1851–1872.
- Merton, R.C., 1980. On estimating the expected return on the market: An exploratory investigation. *J. Financ. Econ.* 8 (4), 323–361.
- Okhrin, Y., Schmid, W., 2006. Distributional properties of portfolio weights. *J. Econometrics* 134 (1), 235–256.
- Zhao, Z., Ledoit, O., Jiang, H., 2021. Risk reduction and efficiency increase in large portfolios: Cross-exposure constraints and shrinkage of the covariance matrix*. *J. Financ. Econ.* 21 (1), 73–105.