The European Journal of Finance

ISSN: 1351-847X (Print) 1466-4364 (Online) Journal homepage: www.tandfonline.com/journals/rejf20

High-Dimensional portfolio selection with HDShOP package

Taras Bodnar, Solomiia Dmytriv, Yarema Okhrin, Dmitry Otryakhin & Nestor Parolya

To cite this article: Taras Bodnar, Solomiia Dmytriv, Yarema Okhrin, Dmitry Otryakhin & Nestor Parolya (12 May 2025): High-Dimensional portfolio selection with HDShOP package, The European Journal of Finance, DOI: 10.1080/1351847X.2025.2501637

To link to this article: https://doi.org/10.1080/1351847X.2025.2501637

Published online: 12 May 2025.

THE EUROPEAN JOURNAL OF FINANCE
https://doi.org/10.1080/1351847X.2025.2501637

High-Dimensional portfolio selection with HDShOP package

Taras Bodnar [a], Solomiia Dmytriv [b], Yarema Okhrin [c], Dmitry Otryakhin [d] and Nestor Parolya [e]

[a]Department of Management and Engineering, Linköping University, Linköping, Sweden; [b]Department of Statistics and Operations Research, University of Vienna, Vienna, Austria; [c]Chair of Statistics and Data Science, Augsburg University, Augsburg, Germany; [d]Department of Mathematics, Stockholm University, Stockholm, Sweden; [e]Delft Insitute of Applied Mathematics, Delft University of Technology, Delft, The Netherlands

ABSTRACT
This paper discusses the practical aspects of working with high-dimensional shrinkage portfolios. It presents the R package `HDShOP` which provides a comprehensive framework for such work. In particular, we cover the construction of portfolios using shrinkage-based estimators for the mean vector, covariance matrix, and precision matrix of asset returns, as well as the shrinkage estimators derived directly for the weights of optimal portfolios. Moreover, shrinkage-based tests on the mean-variance efficiency of a given portfolio are discussed. Aspects related to programming, such as classes and methods used in the construction of optimal portfolios, are described. The description of the software is preceded by underlying theory and it is accompanied by several empirical illustrations based on the data consisting of returns on stocks from the S&P 500 index.

ARTICLE HISTORY
Received 29 July 2024
Accepted 29 April 2025

KEYWORDS
High-dimensional portfolio theory; shrinkage estimation; shrinkage-based test; high-dimensional statistical inference; quantitative finance

1. Introduction

The mean-variance approach of Markowitz is a well-established and very popular tool for optimal portfolio selection in financial literature (see, e.g. Adcock 2014; Adcock, Eling, and Loperfido 2015; Bodnar et al. 2021; Bodnar, Parolya, and Thorsén 2024; Britten-Jones 1999; Ding, Li, and Zheng 2021; Lassance, Vanderveken, and Vrins 2024; Markowitz 1952). The idea behind the approach is to invest in the portfolio that has the smallest variance for a given level of the expected return. In case of no restriction on the expected portfolio return, 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 Bodnar, Hautsch, et al. 2025; Bodnar and Schmid 2009; Kan and Smith 2008; Merton 1972).

However, when the Markowitz theory is implemented in practice, one of the important challenges arises when unknown parameters of the data-generating process are replaced by their sample counterparts in the expressions of the optimal portfolio weights (see Adcock 2015; Bodnar, Bodnar, and Parolya 2022; Bodnar, Okhrin, and Parolya 2023; Cai et al. 2020; El Karoui 2010; Okhrin and Schmid 2006). The impact of the estimation error on the performance of optimal portfolio weights is sometimes even larger than the uncertainty that is present in the data-generating model of the asset returns. Moreover, it seems that the errors in the sample mean vector have even a larger influence on the performance of optimal portfolios than the errors related to the estimation of the covariance matrix (see, e.g. Best and Grauer 1991; Chopra and Ziemba 1993; Merton 1980). These two types of uncertainty should usually be taken into account simultaneously when an optimal portfolio is constructed in practice, see, e.g. Bodnar, Bodnar, and Parolya (2022) and Bodnar, Dette, et al. (2022). Other important aspects of

CONTACT Taras Bodnar taras.bodnar@liu.se

modern portfolio theory are discussed in Alexander and Baptista (2002), Alexander and Baptista (2004), Bodnar, Parolya, and Schmid (2015), Garcia et al. (2019), Bodnar, Lindholm, et al. (2022), Huang and Ma (2023), X. Wang et al. (2023), Bodnar, Bodnar, and Niklasson (2025) among others.

The estimation problem is even more severe in the case of high-dimensionality is present. If the portfolio dimension is considerably small compared to the sample size, then the traditional sample covariance matrix consistently estimates the population counterpart under weak conditions imposed on the data-generating model of the asset returns and, consequently, the traditional portfolio provides a good approximation of the population one. The situation is completely different in the high-dimensional setting when the portfolio dimension is comparable to the sample size (see Bai and Silverstein 2010; Bodnar, Dette, and Parolya 2019). In this case, the sample covariance matrix is no longer a consistent estimator for the true covariance matrix. As a result, the traditional portfolio weights might deviate considerably from the population ones. To ensure a good performance of the holding portfolio, the weights of the traditional portfolio have to be adjusted by taking the parameter uncertainty into account (see, e.g. Ao, Yingying, and Zheng 2019; Bodnar, Dmytriv, et al. 2019; Bodnar, Parolya, and Thorsén 2024; Cai et al. 2020; Ding, Li, and Zheng 2021; Glombek 2014; Jagannathan and Ma 2003).

In the `HDShOP` package (see Bodnar, Dmytriv, et al. 2024), we present a framework that is based on two strategies for the construction of shrinkage portfolios in the high-dimensional setting: (i) direct shrinkage estimation of the optimal portfolio weights; (ii) shrinkage estimation of the mean vector and the covariance/precision matrix. Also, we introduce numerical implementations of the recent results related to those techniques. In particular, we include the methods for the shrinkage estimation of optimal portfolio weights developed in Bodnar, Parolya, and Schmid (2018) and Bodnar, Okhrin, and Parolya (2023), as well as the routines for the shrinkage estimation of the mean vector based on Jorion (1986), Bodnar, Okhrin, and Parolya (2019) and the covariance matrix presented in Bodnar, Gupta, and Parolya (2014), Bodnar, Gupta, and Parolya (2016), Ledoit and Wolf (2017) and Ledoit and Wolf (2020). Also, the package provides the mechanisms for analyzing and testing the efficiency of a given portfolio in a high-dimensional setting. The suggested workflow is based on the means of object-oriented programming: Optimal portfolios are represented by S3 classes for which methods are defined.

Although `HDShOP` has a unique functionality, packages with related tools exist. There are packages implementing shrinkage estimators for the covariance matrices. For example, `ShrinkCovMat` contains methods developed in Touloumis (2015), while `CovTools` (Lee and You 2021) provides the functions based on the approaches developed in Ledoit and Wolf (2003, 2004) as well as the oracle approximating shrinkage estimator and the Rao-Blackwell Ledoit-Wolf estimator (Chen et al. 2010). Moreover, `nlshrink` (Ramprasad 2016) provides the methods from Ledoit and Wolf (2004, 2015). There are several packages devoted to portfolio management. Among these, `fPortfolio` (Wuertz et al. 2023) seems to be one of the most used R packages. An important task of this package is the construction of optimal portfolios by minimizing the portfolio variance while constraining the portfolio return and asset allocation. To achieve this, relatively simple estimators of the mean vector and the covariance matrix of the asset returns are used, and then a quadratic solver performs optimization to find the portfolio weights. Numerical solvers allow for different kinds of constraints such as the prohibition of short sells or group and box constraints. Another approach included in the package fixes the portfolio variance and maximizes the portfolio return leading to the so-called minimum risk efficient portfolio. Apart from the construction of optimal portfolios, `fPortfolio` also provides means for backtesting and various plotting functions. The other R packages dealing with the optimal portfolio construction are `parma` (Galanos and Pfaff 2022), `MarkowitzR` (Pav 2023), `PortfolioOptim` (Palczewski 2019), `epo` (Reckziegel 2023). `MarkowitzR` presents a collection of methods to analyze the significance of optimal portfolios, while `PortfolioOptim` determines portfolios that have smallest distance to the benchmark portfolio. The contribution of `HDShOP` package to the existing package ecosystem is twofold: (i) It implements the popular shrinkage approach for constructing optimal portfolios following the recently published papers of Bodnar, Gupta, and Parolya (2014), Bodnar, Gupta, and Parolya (2016), Bodnar, Parolya, and Schmid (2018), Bodnar, Dmytriv, et al. (2019), Bodnar, Okhrin, and Parolya (2019), Bodnar et al. (2021), Bodnar, Bodnar, and Parolya (2022), Bodnar, Dette, et al. (2022) and Bodnar, Okhrin, and Parolya (2023) written by the authors of the package; (ii) It is the only R package dealing with the construction of optimal portfolios in the high-dimensional setting. In this paper, we contribute to the literature by

- presenting a detailed review of shinkage estimators for the mean vector, covariance matrix, and precision matrix derived in the high-dimensional settings;
- providing the analytical formulae of the direct shrinkage estimators for both the global minimum variance and the mean-variance optimal portfolios;
- describing a high-dimensional test on the mean-variance efficiency of a given portfolio;
- highlighting the main features of the R package *HDShOP* designed for the practical implementation of high-dimensional portfolio theory;
- applying the presented approaches to the data consisting of returns on stocks from the S&P 500 index for the period from 2005 to 2025;
- documenting a good performance of the shrinkage approaches, especially during the COVID period.

The rest of the paper is structured as follows. In the next section, we review the classical results of portfolio theory together with the sample estimators of the optimal portfolios and the efficient frontier, the set of optimal portfolios in the mean-variance space. Section 3 presents the improved shrinkage estimators of optimal portfolio weights and discusses the shrinkage method for creating optimal portfolios. The package architecture is described and examples of its usage are provided in Section 4. This is followed by the results of the empirical illustration in Section 5 where the functionality of the package is demonstrated in the construction of optimal portfolios based on real data of returns on the stocks included in the S&P 500 index. Concluding remarks are summarized in Section 6.

2. Classical mean-variance analysis

In 1952 Harry Markowitz (Markowitz 1952) established a new approach for constructing optimal portfolios which is still known in the literature as modern portfolio theory. In the derivation of an optimal portfolio, the tradeoff between the expected portfolio return and portfolio variance is taken into account and the optimal portfolios are obtained by minimizing the portfolio variance given that the expected portfolio return is equal to a certain value.

Let $\mathbf{x} = (x_1, \ldots, x_p)^\top$ be the p-dimensional vector of the asset returns and let $\mathbf{w} = (w_1, \ldots, w_p)^\top$ be the p-dimensional vector of portfolio weights. Then the expected return and the variance of the portfolio with weights $\mathbf{w}$ are given by

$$\mu_p = \mathbf{w}^\top \boldsymbol{\mu} \quad \text{and} \quad V_p = \mathbf{w}^\top \Sigma \mathbf{w}, \tag{1}$$

respectively, where $\boldsymbol{\mu} = \mathbb{E}(\mathbf{x})$ is the mean vector and $\Sigma = \mathbb{C}ov(\mathbf{x})$ is the covariance matrix of the asset returns. The mean-variance (MV) optimal portfolios obtained by following the Markowitz approach can also be deduced as the solutions of the expected utility optimization problem (e.g. Bodnar, Parolya, and Schmid 2013) expressed as

$$\mathbf{w}^\top \boldsymbol{\mu} - \frac{\gamma}{2} \mathbf{w}^\top \Sigma \mathbf{w} \to \max \quad \text{subject to } \mathbf{w}^\top \mathbf{1}_p = 1, \tag{2}$$

where $\mathbf{1}_p$ is the p-dimensional vector of size p and $\gamma > 0$ is the coefficient of risk aversion that measures the investor's attitude towards risk. The solution of (2) is given by

$$\mathbf{w}_{MV} = \frac{\Sigma^{-1}\mathbf{1}_p}{\mathbf{1}_p^\top \Sigma^{-1}\mathbf{1}_p} + \gamma^{-1}\mathbf{Q}\boldsymbol{\mu} \quad \text{with } \mathbf{Q} = \Sigma^{-1} - \frac{\Sigma^{-1}\mathbf{1}_p\mathbf{1}_p^\top\Sigma^{-1}}{\mathbf{1}_p^\top \Sigma^{-1}\mathbf{1}_p}. \tag{3}$$

In the special case of a fully risk-averse investor, i.e. $\gamma = \infty$, the optimization problem (2) becomes

$$\mathbf{w}^\top \Sigma \mathbf{w} \to \min \quad \text{subject to } \mathbf{w}^\top \mathbf{1}_p = 1, \tag{4}$$

whose solution is expressed as

$$\mathbf{w}_{GMV} = \frac{\Sigma^{-1}\mathbf{1}_p}{\mathbf{1}_p^\top \Sigma^{-1}\mathbf{1}_p}. \tag{5}$$

The portfolio with the weights (5) is the optimal portfolio with the smallest variance and thus it is known as the global minimum variance (GMV) portfolio.

Varying the coefficient of risk aversion γ from 0 to ∞, we get all mean-variance optimal portfolios that lie on the parabola in the mean-variance space, known as the efficient frontier. The equation of the efficient frontier was first derived by Merton (1972). Following Kan and Smith (2008) and Bodnar and Schmid (2009), it is given by

$$(R - R_{GMV})^2 = s(V - V_{GMV}), \tag{6}$$

where

$$R_{GMV} = \frac{\boldsymbol{\mu}^\top \Sigma^{-1} \mathbf{1}_p}{\mathbf{1}_p^\top \Sigma^{-1} \mathbf{1}_p}, \quad V_{GMV} = \frac{1}{\mathbf{1}_p^\top \Sigma^{-1} \mathbf{1}_p}, \quad s = \boldsymbol{\mu}^\top \mathbf{Q} \boldsymbol{\mu} \tag{7}$$

are the expected return of the GMV portfolio, the variance of the GMV portfolio, and the slope parameter of the efficient frontier, respectively.

2.1. Traditional sample estimators of optimal portfolios

In practice, however, the formulas (3) and (5) cannot be directly implemented due to the presence of unobservable quantities $\boldsymbol{\mu}$ and Σ. The traditional approach suggests replacing $\boldsymbol{\mu}$ and Σ with their sample counterparts computed by using the historical data of asset returns. Let $\mathbf{X}_n = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ be the $p \times n$ observation matrix. Then, the traditional sample estimators of $\boldsymbol{\mu}$ and Σ are given by

$$\begin{aligned} \bar{\mathbf{x}}_n &= \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i = \frac{1}{n} \mathbf{X}_n \mathbf{1}_n \quad \text{and} \\ \mathbf{S}_n &= \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}}_n)(\mathbf{x}_i - \bar{\mathbf{x}}_n)^\top = \frac{1}{n-1} \mathbf{X}_n \left(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \mathbf{X}_n^\top, \end{aligned} \tag{8}$$

where $\mathbf{I}_n$ denotes the identity matrix of size n.

Replacing $\boldsymbol{\mu}$ and Σ by $\bar{\mathbf{x}}_n$ and $\mathbf{S}_n$, we get the traditional (sample) MV and GMV portfolios expressed as

$$\hat{\mathbf{w}}_{MV;S} = \hat{\mathbf{w}}_{GMV;S} + \gamma^{-1} \hat{\mathbf{Q}}_n \bar{\mathbf{x}}_n \tag{9}$$

with

$$\hat{\mathbf{w}}_{GMV;S} = \begin{cases} \dfrac{\mathbf{S}_n^{-1} \mathbf{1}_p}{\mathbf{1}_p^\top \mathbf{S}_n^{-1} \mathbf{1}_p} & \text{if } p < n \\ \dfrac{\mathbf{S}_n^{+} \mathbf{1}_p}{\mathbf{1}_p^\top \mathbf{S}_n^{+} \mathbf{1}_p} & \text{if } p > n \end{cases} \tag{10}$$

and

$$\hat{\mathbf{Q}}_n = \begin{cases} \mathbf{S}_n^{-1} - \dfrac{\mathbf{S}_n^{-1} \mathbf{1}_p \mathbf{1}_p^\top \mathbf{S}_n^{-1}}{\mathbf{1}_p^\top \mathbf{S}_n^{-1} \mathbf{1}_p} & \text{if } p < n \\ \mathbf{S}_n^{+} - \dfrac{\mathbf{S}_n^{+} \mathbf{1}_p \mathbf{1}_p^\top \mathbf{S}_n^{+}}{\mathbf{1}_p^\top \mathbf{S}_n^{+} \mathbf{1}_p} & \text{if } p > n \end{cases}, \tag{11}$$

where $\mathbf{S}_n^{+}$ is the Moore-Penrose inverse of $\mathbf{S}_n$ (see Ben-Israel and Greville 2003; Bodnar and Parolya 2024; Harville 1997; Meyer 1973; Penrose 1955; G. Wang et al. 2018).

Similarly, the sample estimator of the efficient frontier is obtained and it is given by

$$(R - \hat{R}_{GMV})^2 = \hat{s}(V - \hat{V}_{GMV}), \tag{12}$$

where

$$\hat{R}_{GMV} = \begin{cases} \dfrac{\bar{\mathbf{x}}_n^\top \mathbf{S}_n^{-1}\mathbf{1}_p}{\mathbf{1}_p^\top \mathbf{S}_n^{-1}\mathbf{1}_p} & \text{if } p < n \\ \dfrac{\bar{\mathbf{x}}_n^\top \mathbf{S}_n^{+}\mathbf{1}_p}{\mathbf{1}_p^\top \mathbf{S}_n^{+}\mathbf{1}_p} & \text{if } p > n \end{cases}, \quad \hat{V}_{GMV} = \begin{cases} \dfrac{1}{\mathbf{1}_p^\top \mathbf{S}_n^{-1}\mathbf{1}_p} & \text{if } p < n \\ \dfrac{1}{\mathbf{1}_p^\top \mathbf{S}_n^{+}\mathbf{1}_p} & \text{if } p > n \end{cases}, \quad \hat{s} = \bar{\mathbf{x}}_n^\top \hat{\mathbf{Q}}_n \bar{\mathbf{x}}_n. \tag{13}$$

The exact and asymptotic distributional properties of the estimators (9) to (13) under the classical and the high-dimensional asymptotic regimes are derived by Okhrin and Schmid (2006), Bodnar and Schmid (2009), Bodnar et al. (2021) and Bodnar, Dette, et al. (2022).

3. Improved shrinkage estimators

Traditional sample estimators perform usually poorly in practice, especially when the portfolio size is relatively large with respect to the sample size (see, e.g. Bodnar, Okhrin, and Parolya 2023; Bodnar, Parolya, and Schmid 2018). To improve the performance of the sample optimal portfolios, several approaches have been proposed in the literature. One of them suggests replacing the unknown mean vector $\boldsymbol{\mu}$ and covariance matrix Σ in (3) and (5) by improved estimators of the mean vector and covariance matrix, e.g. by their shrinkage estimators (cf., Kourtis, Dotsis, and Markellos 2012; Ledoit and Wolf 2004).

The shrinkage-type estimators were originally suggested by Stein (1956) for the estimation of the mean vector of a multivariate normal distribution and since that time the approach has become widely used in different fields of statistics, for example for estimating the covariance and precision matrices (see, e.g. Bodnar, Gupta, and Parolya 2014, 2016; Ledoit and Wolf 2020).

3.1. Shrinkage estimators for the mean vector

In package `HDShOP` (High-Dimensional Shrinkage Optimal Portfolio, see Bodnar, Dmytriv, et al. 2024) the following shrinkage estimators for the mean vector are available:

- James-Stein shrinkage estimator (Jorion 1986):

$$\hat{\boldsymbol{\mu}}_{n,JS} = (1 - \hat{\alpha}_{m,JS})\bar{\mathbf{x}}_n + \hat{\alpha}_{m,JS} x_0 \mathbf{1}_p, \tag{14}$$

  where x_0 is the shrinkage target and

$$\hat{\alpha}_{m,JS} = \min\left\{1, \frac{(p-2)/n}{(\bar{\mathbf{x}}_n - x_0\mathbf{1}_p)^\top \mathbf{S}_n^{-1}(\bar{\mathbf{x}}_n - x_0\mathbf{1}_p)}\right\}. \tag{15}$$

- Bayes-Stein shrinkage estimator (Jorion 1986):

$$\hat{\boldsymbol{\mu}}_{n,BS} = (1 - \hat{\alpha}_{m,BS})\bar{\mathbf{x}}_n + \hat{\alpha}_{m,BS} \hat{x}_0 \mathbf{1}_p, \tag{16}$$

  where

$$\hat{\alpha}_{m,BS} = \frac{p+2}{p+2+n(\bar{\mathbf{x}}_n - \hat{x}_0\mathbf{1}_p)^\top \mathbf{S}_n^{-1}(\bar{\mathbf{x}}_n - \hat{x}_0\mathbf{1}_p)} \quad \text{and} \quad \hat{x}_0 = \frac{\bar{\mathbf{x}}_n^\top \mathbf{S}_n^{-1}\mathbf{1}_p}{\mathbf{1}_p^\top \mathbf{S}_n^{-1}\mathbf{1}_p} = \hat{R}_{GMV}. \tag{17}$$

  Moreover, Jorion (1986) suggests the application of $\frac{n-1}{n-p-2}\mathbf{S}_n$ instead of $\mathbf{S}_n$ in the above formulas.

- Optimal linear shrinkage estimator (Bodnar, Okhrin, and Parolya 2019):

$$\hat{\boldsymbol{\mu}}_{n,BOP} = \hat{\alpha}_{m,BOP}\bar{\mathbf{x}}_n + \hat{\beta}_{m,BOP}\boldsymbol{\mu}_0\,, \tag{18}$$

where

$$\hat{\alpha}_{m,BOP} = \frac{\left(\bar{\mathbf{x}}_n^\top \mathbf{S}_n^{-1}\bar{\mathbf{x}}_n - \frac{p/n}{1-p/n}\right)\boldsymbol{\mu}_0^\top \mathbf{S}_n^{-1}\boldsymbol{\mu}_0 - (\bar{\mathbf{x}}_n^\top \mathbf{S}_n^{-1}\boldsymbol{\mu}_0)^2}{\bar{\mathbf{x}}_n^\top \mathbf{S}_n^{-1}\bar{\mathbf{x}}_n\boldsymbol{\mu}_0^\top \mathbf{S}_n^{-1}\boldsymbol{\mu}_0 - (\bar{\mathbf{x}}_n^\top \mathbf{S}_n^{-1}\boldsymbol{\mu}_0)^2} \tag{19}$$

and

$$\hat{\beta}_{m,BOP} = (1-\hat{\alpha}_{m,BOP})\frac{\bar{\mathbf{x}}_n^\top \mathbf{S}_n^{-1}\boldsymbol{\mu}_0}{\boldsymbol{\mu}_0^\top \mathbf{S}_n^{-1}\boldsymbol{\mu}_0}. \tag{20}$$

3.2. Shrinkage estimators for the covariance and precision matrices

A linear shrinkage estimator for the covariance matrix is given by (see, Bodnar, Gupta, and Parolya 2014)

$$\hat{\Sigma}_{n;BGP} = \hat{\alpha}_{cov}\mathbf{S}_n + \hat{\beta}_{cov}\Sigma_0, \tag{21}$$

with

$$\hat{\alpha}_{cov} = 1 - \frac{\frac{1}{n}||\mathbf{S}_n||_{tr}^2||\Sigma_0||_F^2}{||\mathbf{S}_n||_F^2||\Sigma_0||_F^2 - (\mathrm{tr}(\mathbf{S}_n\Sigma_0))^2} \tag{22}$$

and

$$\hat{\beta}_{cov} = \frac{\mathrm{tr}(\mathbf{S}_n\Sigma_0)}{||\Sigma_0||_F^2}\left(1-\hat{\alpha}_{cov}\right), \tag{23}$$

where $||\mathbf{A}||_{tr} = \mathrm{tr}[(\mathbf{A}\mathbf{A}^\top)^{1/2}]$ denotes for the trace norm and Σ_0 is assumed to possess the bounded trace norm.

Recently, a non-linear shrinkage estimator of the covariance matrix has been suggested in the literature by Ledoit and Wolf (2017, 2020). Following Ledoit and Wolf (2020) it is given by

$$\hat{\Sigma}_{n,LWnonlin} = \mathbf{H}\mathrm{diag}(d_1^{or},\dots,d_p^{or})\mathbf{H}^\top,$$
$$d_i^{or} = \begin{cases} \dfrac{d_i}{|1-p/n-p/nd_i\breve{m}_F(d_i)|^2}, & \text{if } d_i > 0,\\ \dfrac{1}{(p/n-1)\breve{m}_{\underline{F}}(0)}, & \text{if } d_i = 0, \end{cases} \quad \text{for } i \in \{1,\dots,p\}, \tag{24}$$

where $\mathbf{H}$ is the matrix with the sample eigenvectors of $\mathbf{S}_n$, $d_1,\dots d_p$ are the sample eigenvalues of $\mathbf{S}_n$, $\breve{m}_{\underline{F}}(x) = \lim_{z\to x} m_{\underline{F}}(z) = \lim_{z\to x}(\frac{c-1}{z} + cm_F(z))$ and $\breve{m}_F(x) = \lim_{z\to x} m_F(z)$ with $m_F(z)$ the limiting Stieltjes transform, which for a distribution function $G:\mathbb{R}\to\mathbb{R}$ it is defined by

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

Finally, a linear shrinkage estimator for the precision matrix (inverse covariance matrix) is given by

$$\hat{\mathbf{\Pi}}_{n,BGP} = \hat{\alpha}_{prec}\mathbf{S}_n^{-1} + \hat{\beta}_{prec}\mathbf{\Pi}_0, \tag{25}$$

with

$$\hat{\alpha}_{prec} = 1 - p/n - \frac{\frac{1}{n}||\mathbf{S}_n^{-1}||_{tr}^2||\mathbf{\Pi}_0||_F^2}{||\mathbf{S}_n^{-1}||_F^2||\mathbf{\Pi}_0||_F^2 - \left(\mathrm{tr}(\mathbf{S}_n^{-1}\mathbf{\Pi}_0)\right)^2} \tag{26}$$

and

$$\hat{\beta}_{prec} = \frac{\mathrm{tr}(\mathbf{S}_n^{-1}\mathbf{\Pi}_0)}{||\mathbf{\Pi}_0||_F^2}\left(1 - p/n - \hat{\alpha}_{prec}\right), \tag{27}$$

where $1/p\mathbf{\Pi}_0$ is assumed to have the bounded trace norm.

Other types of shrinkage estimators for the mean vector, covariance matrix, and precision matrix can be found in the review paper of Bodnar, Bodnar, and Parolya (2022).

3.3. Optimal shrinkage estimators

Another possibility to construct the shrinkage estimators of optimal portfolios is to define them directly for portfolio weights instead of using the shrinkage estimators of the mean vector and covariance matrix. The motivation behind this approach is twofold: (i) the application of the improved estimators for the mean vector and the covariance (precision) matrix does not necessarily lead to optimal estimators of optimal portfolios; (ii) constructing improved estimators for the mean vector and the covariance matrix requires accounting for the estimation error present in $p + p(p+1)/2$ parameters, while the aim is to obtain an estimator for a $p-1$ dimensional vector of portfolio weights.

Recently, optimal shrinkage estimators for the weights of the MV portfolio and GMV portfolio were constructed in Bodnar, Parolya, and Schmid (2018) and Bodnar, Okhrin, and Parolya (2023), obtained by minimizing the out-of-sample variance and the out-of-sample utility functions. The resulting shrinkage estimator for the MV portfolio is given by

$$\hat{\mathbf{w}}_{MV;oSh} = \hat{\alpha}_{n;MV}\hat{\mathbf{w}}_{MV;S} + (1 - \hat{\alpha}_{n;MV})\mathbf{b} \tag{28}$$

with

$$\hat{\alpha}_{n;MV} = \begin{cases} \dfrac{\gamma^{-1}(\hat{R}_{GMV} - \hat{R}_b)\left(1 + \frac{1}{1-p/n}\right) + (\hat{V}_b - \hat{V}_{GMV;c}) + \frac{\gamma^{-2}}{1-p/n}\hat{s}_c}{\frac{\hat{V}_{GMV;c}}{1-p/n} - 2\left(\hat{V}_{GMV;c} + \frac{\gamma^{-1}}{1-p/n}(\hat{R}_b - \hat{R}_{GMV})\right) + \gamma^{-2}\left(\frac{\hat{s}_c + p/n}{(1-p/n)^3}\right) + \hat{V}_b} & \text{if } p < n \\ \dfrac{\gamma^{-1}(\hat{R}_{GMV} - \hat{R}_b)\left(1 + \frac{1}{p/n(p/n-1)}\right) + (\hat{V}_b - \hat{V}_{GMV;c}) + \frac{\gamma^{-2}}{p/n(p/n-1)}\hat{s}_c}{\frac{(p/n)^2\hat{V}_{GMV;c}}{p/n-1} - 2\left(\hat{V}_{GMV;c} + \frac{\gamma^{-1}}{p/n(p/n-1)}(\hat{R}_b - \hat{R}_{GMV})\right) + \gamma^{-2}\left(\frac{\hat{s}_c + (p/n)^2}{(p/n-1)^3}\right) + \hat{V}_b} & \text{if } p > n \end{cases}, \tag{29}$$

where $\hat{R}_b$ and $\hat{V}_b$ are consistent estimators for the expected return $R_b = \boldsymbol{\mu}^\top\mathbf{b}$ and the variance $V_b = \mathbf{b}^\top\mathbf{\Sigma}\mathbf{b}$ of the target portfolio $\mathbf{b}$ given by

$$\hat{R}_b = \hat{\boldsymbol{\mu}}^\top\mathbf{b}, \quad \text{and} \quad \hat{V}_b = \mathbf{b}^\top\mathbf{S}_n\mathbf{b}, \tag{30}$$

while $\hat{V}_{GMV;c}$ and $\hat{s}_c$ are consistent estimator for V_{GMV} and s expressed as

$$\hat{V}_{GMV;c} = \begin{cases} \dfrac{1}{1-p/n}\hat{V}_{GMV} & \text{if } p < n \\ \dfrac{1}{p/n(p/n-1)}\hat{V}_{GMV} & \text{if } p > n \end{cases} \quad \text{and} \quad \hat{s}_c = \begin{cases} (1-p/n)\,\hat{s} - p/n & \text{if } p < n \\ p/n\,(p/n-1)\,\hat{s} - p/n & \text{if } p > n \end{cases}. \tag{31}$$

For the GMV portfolio we get

$$\hat{\mathbf{w}}_{GMV;oSh} = \hat{\alpha}_{n;GMV}\hat{\mathbf{w}}_{GMV;S} + (1 - \hat{\alpha}_{n;GMV})\mathbf{b} \tag{32}$$

with

$$\hat{\alpha}_{n;GMV} = \begin{cases} \dfrac{\hat{V}_b - \hat{V}_{GMV;c}}{\frac{\hat{V}_{GMV;c}}{1-p/n} - 2\hat{V}_{GMV;c} + \hat{V}_b} & \text{if } p < n \\ \dfrac{\hat{V}_b - \hat{V}_{GMV;c}}{\frac{(p/n)^2\hat{V}_{GMV;c}}{p/n-1} - 2\hat{V}_{GMV;c} + \hat{V}_b} & \text{if } p > n \end{cases}. \tag{33}$$

Remark 3.1: One of the important questions is the choice of the target vector/weights $\mathbf{b}$. It should be mentioned that every shrinkage method is, of course, sensitive to this choice to some extent. Still, there are some 'safe' targets which would never make your estimator worse than the traditional one, i.e. $\hat{\mathbf{w}}_{MV;S}$ given in (9). Let us consider the GMV portfolio (5) for simplicity. It is evident that the variance of this portfolio is of order $O(1/p)$ in case the population covariance matrix Σ has the maximum eigenvalue bounded away from infinity for increasing p. Indeed, using Ruhe's trace inequality (see, Ruhe 1970) we get

$$V_{GMV} = \frac{1}{\mathbf{1}_p^\top \Sigma^{-1} \mathbf{1}_p} \le \frac{\lambda_{\max}(\Sigma)}{\mathbf{1}_p^\top \mathbf{1}_p} = O(1/p),$$

where $\lambda_{\max}(\Sigma)$ is the maximum eigenvalue of Σ. Now, taking the naive portfolio as a target, i.e. $\mathbf{b} = \mathbf{1}_p/p$, is a safe option since using again the same trace inequality one gets

$$V_b = \frac{\mathbf{1}_p^\top \Sigma \mathbf{1}_p}{p^2} \le \frac{\mathbf{1}_p^\top \mathbf{1}_p}{p^2} \lambda_{\max}(\Sigma) = \frac{\lambda_{\max}(\Sigma)}{p} = O(1/p),$$

where the latter expression is bounded for a bounded $\lambda_{\max}(\Sigma)$, but even if $\lambda_{\max}(\Sigma)$ is unbounded and of order $O(p)$ (factor model), the variance of the target portfolio will still be bounded for increasing dimensions. Similarly, one can show that the target portfolios constructed from an equicorrelation target matrix of the form $\mathbf{\Omega} = (1-\rho)\mathbf{I} + \rho \mathbf{1}_p \mathbf{1}_p^\top$ for some suitable ρ also obey the variance of order $O(1/p)$.

The aforementioned discussion also implies that one should choose the target portfolio with a bounded variance; otherwise, the shrinkage estimation technique could fail. Indeed, due to (33) the shrinking coefficient $\hat{\alpha}_{n;GMV}$ will tend to one if V_b (and consequently $\hat{V}_b$) is tending to infinity; and the shrinkage will have no effect. But this is also very natural since nobody would shrink the sample portfolio to a target with (asymptotically) infinite variance. Otherwise, for finite V_b, the shrinkage intensity $\hat{\alpha}_{n;GMV}$ will be positive and utilize the best combination of two portfolios: sample and target ones. Thus, one should be careful with the data-driven target portfolios, especially if the behavior of their asymptotic variance is unknown.

3.4. High-dimensional test theory based on the shrinkage approach

Bodnar, Dmytriv, et al. (2019) and Bodnar et al. (2021) derived the high-dimensional asymptotic distributions of $\hat{\alpha}_{n;MV}$ and $\hat{\alpha}_{n;GMV}$, and utilized these distributions to develop high-dimensional tests assessing the optimality of a given portfolio with weights $\mathbf{w}_0$.

In the case of the MV portfolio, we aim to test the hypotheses

$$H_0 : \mathbf{w}_{MV} = \mathbf{w}_0 \quad \text{vs.} \quad H_1 : \mathbf{w}_{MV} \ne \mathbf{w}_0. \tag{34}$$

Bodnar et al. (2021) suggested the application of the following test statistics

$$T_{MV}(\mathbf{w}_0) = \sqrt{n} \frac{\hat{\alpha}_{n;MV}(\mathbf{w}_0)\hat{B}_n(\mathbf{w}_0)}{\sqrt{\mathbf{d}_0' \hat{\Omega}(\mathbf{w}_0)\mathbf{d}_0}}, \tag{35}$$

where $\hat{\alpha}_{n;MV}(\mathbf{w}_0)$ is the optimal shrinkage intensity as defined in (29) with $\mathbf{b} = \mathbf{w}_0$,

$$\hat{B}_n(\mathbf{w}_0) = \frac{\hat{V}_{GMV;c}}{1-c} - 2\left(\hat{V}_{GMV;c} + \frac{\gamma^{-1}}{1-c}(\hat{R}_{\mathbf{w}_0} - \hat{R}_{GMV})\right) + \gamma^{-2}\left(\frac{\hat{s}_c + c}{(1-c)^3}\right) + \hat{V}_{\mathbf{w}_0},$$

$$\mathbf{d}_0 = \begin{pmatrix} \gamma^{-1} + \dfrac{\gamma^{-1}}{1-c} \\ -1 \\ \dfrac{\gamma^{-2}}{1-c} \\ -\gamma^{-1} - \dfrac{\gamma^{-1}}{1-c} \\ 1 \end{pmatrix},$$

and

$$\hat{\Omega}(\mathbf{w}_0) = \begin{pmatrix} \dfrac{\hat{V}_{GMV;c}(\hat{s}_c+1)}{1-c} & 0 & 0 & \hat{V}_{GMV;c} & -2\hat{V}_{GMV;c}(\hat{R}_{\mathbf{w}_0} - \hat{R}_{GMV}) \\ 0 & 2\dfrac{\hat{V}^2_{GMV;c}}{1-c} & 0 & 0 & 2\hat{V}^2_{GMV;c} \\ 0 & 0 & 2\dfrac{((\hat{s}_c+1)^2 + c - 1)}{1-c} & 2(\hat{R}_{\mathbf{w}_0} - \hat{R}_{GMV}) & -2(\hat{R}_{\mathbf{w}_0} - \hat{R}_{GMV})^2 \\ \hat{V}_{GMV;c} & 0 & 2(\hat{R}_{\mathbf{w}_0} - \hat{R}_{GMV}) & \hat{V}_{\mathbf{w}_0} & 0 \\ -2\hat{V}_{GMV;c}(\hat{R}_{\mathbf{w}_0} - \hat{R}_{GMV}) & 2\hat{V}^2_{GMV;c} & -2(\hat{R}_{\mathbf{w}_0} - \hat{R}_{GMV})^2 & 0 & 2\hat{V}^2_{\mathbf{w}_0} \end{pmatrix}. \tag{36}$$

Under the null hypothesis in (34) it holds that (see, Bodnar et al. 2021)

$$T_{MV}(\mathbf{w}_0) \xrightarrow{d} \mathcal{N}(0,1) \quad \text{for } p/n \to c \in [0,1) \text{ as } n \to \infty.$$

Hence, the hypothesis that $\mathbf{w}_0$ are the weights of the MV optimal portfolio is rejected if $|T_{MV}(\mathbf{w}_0)| > z_{1-\delta/2}$, where $z_{1-\delta/2}$ is the $(1-\delta/2)$ quantile of the standard normal distribution.

Similarly, for the GMV portfolio, we aim to check whether the portfolio with weights $\mathbf{w}_0$ can be equal to the optimal portfolio with the smallest variance, that is

$$H_0 : \mathbf{w}_{GMV} = \mathbf{w}_0 \quad \text{vs.} \quad H_1 : \mathbf{w}_{GMV} \neq \mathbf{w}_0. \tag{37}$$

To test (37), Bodnar, Dmytriv, et al. (2019) derived the following test statistic

$$T_{GMV}(\mathbf{w}_0) = \sqrt{n}\frac{(1-c)\,\hat{L}_{\mathbf{w}_0}}{c + (1-c)\,\hat{L}_{\mathbf{w}_0}}, \tag{38}$$

where

$$\hat{L}_{\mathbf{w}_0} = (1-c)\,\mathbf{w}_0^\top \mathbf{S}_n \mathbf{w}_0 \mathbf{1}_p^\top \mathbf{S}_n^{-1} \mathbf{1}_p - 1.$$

Under the null hypothesis in (37) we get

$$T_{GMV}(\mathbf{w}_0) \xrightarrow{d} \mathcal{N}\left(0, 2\frac{1-c}{c}\right)$$

for $p/n \to c \in [0,1)$ as $n \to \infty$ and the null hypothesis in (37) is rejected as soon as

$$|T_{GMV}(\mathbf{w}_0)| > \sqrt{2\frac{1-c}{c}} z_{1-\delta/2}.$$

To this end, we note that the above two approaches are used to test the efficiency of a given portfolio, i.e. they are the tests on the whole vector of portfolio weights. Additionally to these two tests, Bodnar, Dette, et al. (2022)

developed a high-dimensional test for each element of the target vector $\mathbf{w}_0$, i.e. a test on the weight of each asset separately.

For the MV portfolio, the hypotheses on the i-th weight are formulated as

$$H_0 : w_{MV,i} = w_{0,i} \quad \text{vs.} \quad H_1 : w_{MV,i} \neq w_{0,i}. \tag{39}$$

Let $\mathbf{e}_i = (0, \ldots, 0, \underbrace{1}_{i}, 0, \ldots, 0)^\top$. Then the test statistic is given by (see Bodnar, Dette, et al. 2022)

$$T_{MV;i} = (n-p)\frac{(\hat{w}_{MV,i;c} - w_{0,i})^2}{v_i}, \tag{40}$$

where

$$\hat{w}_{MV,i;c} = \hat{w}_{GMV;S,i} + \gamma^{-1}(1-c_n)\,\mathbf{e}_i^\top \hat{\mathbf{Q}}_n \bar{\mathbf{x}}_n, \quad \text{with } \hat{w}_{GMV;S,i} = \frac{\mathbf{e}_i^\top \mathbf{S}_n^{-1} \mathbf{1}_p}{\mathbf{1}_p^\top \mathbf{S}_n^{-1} \mathbf{1}_p} \tag{41}$$

and

$$v_i = \left(\gamma^{-2}(\hat{s}_c + 1) + \hat{V}_{GMV;c}\right)(1-c_n)\mathbf{e}_i^\top \hat{\mathbf{Q}}_n \mathbf{e}_i + \gamma^{-2}(1-c_n)^2(\mathbf{e}_i^\top \hat{\mathbf{Q}}_n \bar{\mathbf{x}}_n)^2, \tag{42}$$

where $c_n = p/n$, and $\hat{w}_{GMV;S,i}$ denotes the i-th element of $\hat{\mathbf{w}}_{GMV;S}$ as defined in (10).

Under the null hypothesis in (39), it holds that

$$T_{MV;i} \xrightarrow{d} \chi_1^2$$

for $p/n \to c \in [0,1)$ as $n \to \infty$ and, consequently, the null hypothesis in (39) is rejected if $T_{MV;i} > q_{1-\delta}$, where $q_{1-\delta}$ is the $(1-\delta)$ quantile of the χ^2-distribution with one degree of freedom.

Using the duality between the test theory and the interval estimation (see, e.g. Aitchison 1964), the $(1-\delta)$ confidence interval for the i-th weight of the MV portfolio is obtained and it is given by

$$\left[\hat{w}_{MV,i;c} - z_{1-\delta/2}\frac{\sqrt{v_i}}{\sqrt{n-p}}, \hat{w}_{MV,i;c} + z_{1-\delta/2}\frac{\sqrt{v_i}}{\sqrt{n-p}}\right]. \tag{43}$$

Finally, it is noted that testing the i-th weight of the GMV portfolio and deriving a confidence interval can be accomplished, as demonstrated in (40) to (43), by setting $\gamma = \infty$.

4. Package structure

In this section, we discuss the internal structure of the package, explain how it works, and provide some examples. We begin with a description of S3 classes, which form the foundation of the package. We then proceed to the creation of portfolios with shrunk weights. The section concludes with the creation of custom portfolios, particularly those based on shrinkage estimators of mean vectors and covariance matrices. The following setup will be used throughout all the examples in this section:

```
R> library(HDShOP)
R> n <- 3e2 # number of realizations
R> p <- 0.5*n # number of assets
R> gamma <- 1
R> # the target for covariance shrinkage
R> TM <- matrix(0, p, p)
R> diag(TM) <- 1
R> # the target for shrinkage of the mean vector
R> mu_0 <- rep(0, p)
R> mu_0[1:10] <- 10:1
R> b <- rep(1/p,p)
```

All data matrices in the package are handled in a way that the variables are in rows and observations – in columns.

```
R> set.seed(5)
R> x <- t(SP_daily_asset_returns[1:n,2:(p+1)])
R> x <- as.matrix(x)
```

4.1. S3 classes for Mean-Variance portfolios

Traditional, weight-shrunk, and custom mean-variance portfolios are returned as objects of the class `MeanVar_portfolio`. A portfolio with shrunk weights has a subclass inheriting from it: the MV shrinkage portfolio (28) has subclass `MV_portfolio_weights_BDOPS21` and the GMV portfolio (32) – `GMV_portfolio_weights_BDPS19`. All objects of class `MeanVar_portfolio` carry at least eight slots.

- Slot `call` contains the function call that creates the portfolio object.
- `cov_mtrx` and `means` store the final estimates of the covariance matrix and the mean vector of the asset returns.
- `inv_cov_mtrx` is the inverse of the covariance matrix.
- `weights` is the final estimate of the portfolio weights.
- `Port_Var` and `Port_mean_return` are the portfolio variance and the expected portfolio return, respectively.
- `Sharpe` is the Sharpe ratio computed by using `Port_Var` and `Port_mean_return`.

The structure of the class is depicted in Figure 1. For example, in the case of the traditional portfolio, slot `weights` contains the weights computed according to formula (9), while in the case of the MV shrinkage portfolio, this slot contains the shrunk weights. In both cases, `cov_mtrx` and `means` are correspondingly the sample covariance matrix and the sample mean vector of the asset returns. When considering a mean-variance portfolio based on shrunk mean vector and covariance matrix, slots `cov_mtrx` and `means` contain the corresponding shrinkage estimates.

In the case $p < n$, it is possible to perform the test on the marginal significance of each weight for the MV and the GMV shrinkage portfolios (39)–(42), and to construct univariate confidence intervals for each weight at the specified confidence level as given in (43). To enable this, objects of classes `MV_portfolio_weights_BDOPS21` and `GMV_portfolio_weights_BDPS19` are returned with an additional slot – `weight_intervals`. It is a dataframe, in which column `weight` contains estimates of individual weights (41), columns `low_bound` and `upp_bound` are the boundaries of confidence intervals from (43) corresponding to those weights, T statistics is defined in formula (40), and `p_value` is the column of p-values of those individual tests.

For `MeanVar_portfolio` methods `summary` and `plot` are available.

```
R> portfolio_BDPS19 <- MVShrinkPortfolio(x=x, gamma=10, type='traditional')
R> summary(portfolio_BDPS19)
```

```
$call
MVShrinkPortfolio(x = x, gamma = 10, type = "traditional")
$Port_Var
[1] 0.09841004
$Port_mean_return
[1] 0.06524613

$Sharpe
[1] 0.2079865
```

```
R> plot(portfolio_BDPS19)
```

The results of the `plot` function are depicted in Figures 2 and 3.

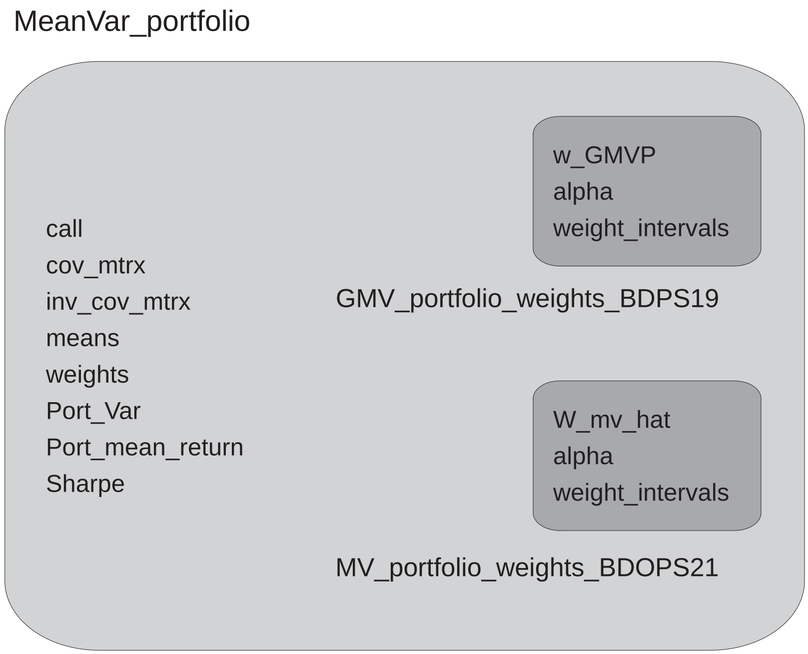

Figure 1. Structure of S3 classes for portfolio objects.

4.2. Mean-variance portfolios with shrunk weights

The main function performing the creation of shrinkage portfolios is `MVShrinkPortfolio`. Essentially, it serves as a function dispatcher, selecting a method based on the arguments `type`, `gamma`, and the ratio p/n of the original data. The argument `type` is a switch indicating whether a shrinkage or traditional portfolio is to be returned.

The function returns either a portfolio in the form of S3 class `MeanVar_portfolio` or a human-readable error message. Generally, `MVShrinkPortfolio` should be used in interactive programming as it is more user-friendly and slower than underlying methods, while the methods themselves should be exploited for writing more complicated but at the same time time-efficient code (Table 1).

Weight-shrunk MV portfolios (28) from Bodnar, Okhrin, and Parolya (2023) are constructed by setting `type='shrinkage'` and `gamma` being equal to a non-infinite real number. For the values of `p` and `n`, set in the beginning of this section, the result of execution of

```
R> portfolio_BDOPS21 <-
+ MVShrinkPortfolio(x=x, gamma=gamma, type='shrinkage', b=b, beta = 0.05)
```

is the same as that of `new_MV_portfolio_weights_BDOPS21` which was eventually dispatched by `MVShrinkPortfolio`:

```
R> portfolio_BDOPS21 <-
+ new_MV_portfolio_weights_BDOPS21(x=x, gamma=gamma, b=b, beta = 0.05)
```

Table 1. Functions dispatched by `MVShrinkPortfolio` based on the arguments `type`, `gamma`, and the p/n ratio of the input data `x`.

Function	Paper	`type`	`gamma`	p/n
`new_MV_portfolio_weights_BDOPS21`	Bodnar, Okhrin, and Parolya (2023)	shrinkage	$<$ Inf	<1
`new_MV_portfolio_weights_BDOPS21_pgn`	Bodnar, Okhrin, and Parolya (2023)	shrinkage	$<$ Inf	>1
`new_GMV_portfolio_weights_BDPS19`	Bodnar, Parolya, and Schmid (2018)	shrinkage	Inf	<1
`new_GMV_portfolio_weights_BDPS19_pgn`	Bodnar, Parolya, and Schmid (2018)	shrinkage	Inf	>1
`new_MV_portfolio_traditional`		traditional	$<$ Inf	<1
`new_MV_portfolio_traditional_pgn`		traditional	$<$ Inf	>1

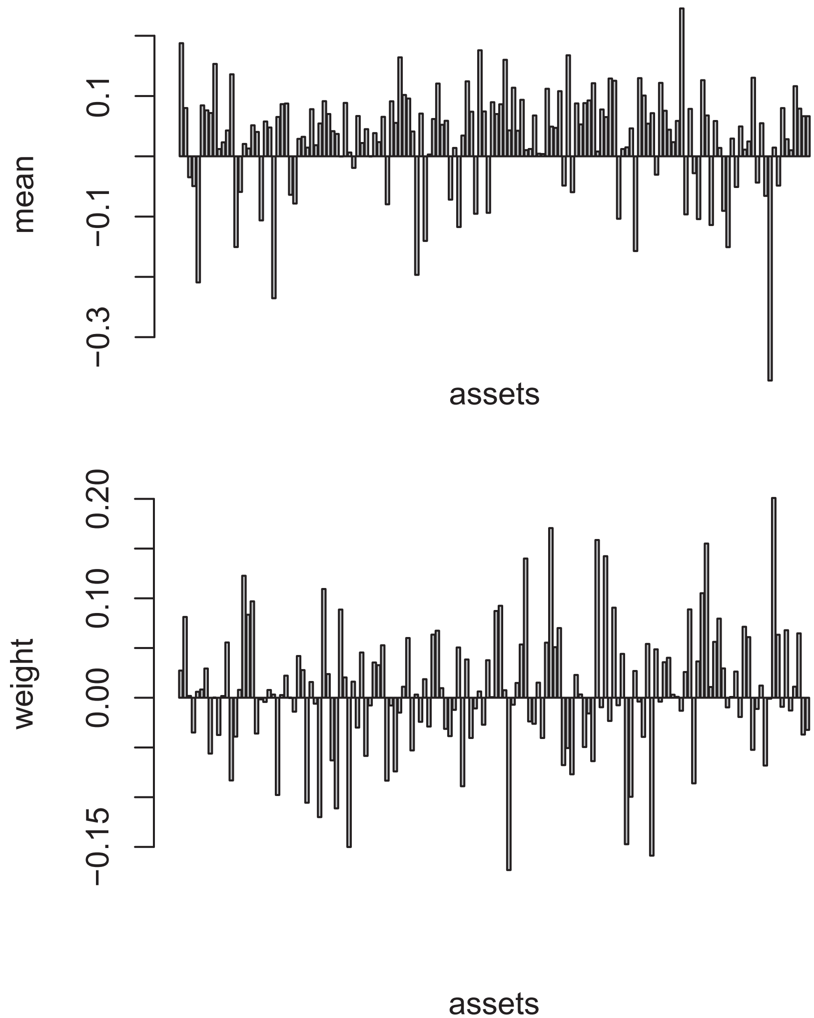

Figure 2. Asset mean returns and portfolio weights

If $p > n$, MVShrinkPortfolio dispatches the function new_MV_portfolio_weights_BDOPS21_pgn. Similarly, for the global minimum variance portfolio (32), setting gamma = Inf will dispatch either new_GMV_portfolio_weights_BDPS19 or new_GMV_portfolio_weights_BDPS19_pgn, depending on the p/n ratio:

```
R> portfolio_BDPS19 <-
+ MVShrinkPortfolio(x=x, gamma=Inf, type='shrinkage', b=b, beta = 0.05)
```

4.3. Custom mean-variance portfolios

Custom mean-variance portfolios can be created with prespecified mean vector and covariance matrix of asset returns. HDShOP includes three functions related to this purpose, namely new_MeanVar_portfolio which is a constructor of objects of class MeanVar_portfolio, validation function for such objects validate_MeanVar_portfolio, and helper MeanVar_portfolio. Conventionally, the constructor acts as a basic creator of class objects, the validator performs checks to confirm the legitimacy of the class object, and the helper function combines these functionalities. Following this concept,

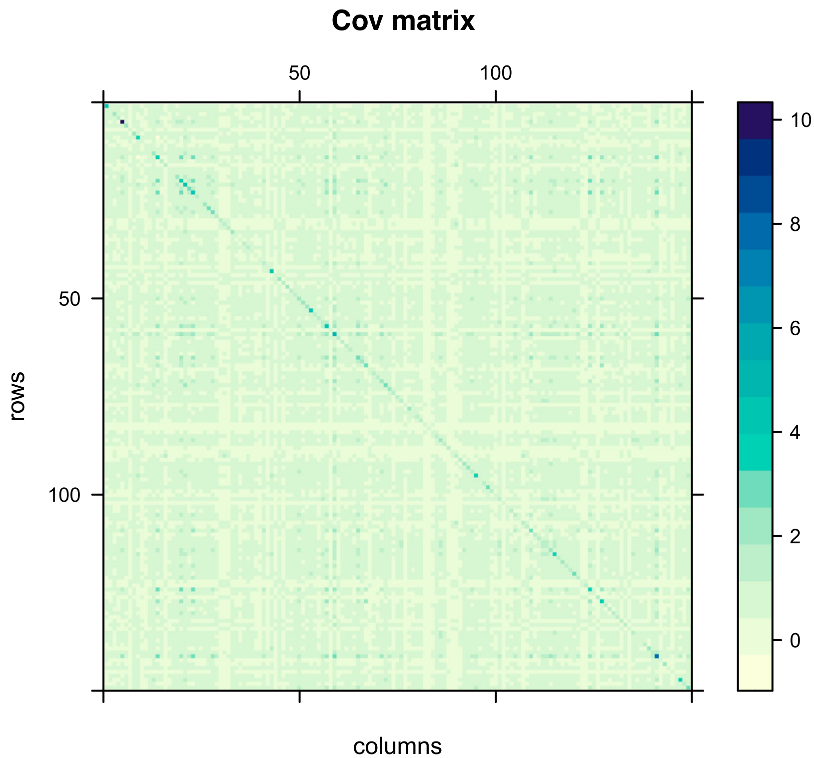

Figure 3. Asset covariance matrix.

`new_MeanVar_portfolio` takes arguments `mean_vec`, `cov_mtrx`, `gamma` and computes the weights accordingly to (3) where the population mean vector and covariance matrix are replaced by their estimators. Currently, `validate_MeanVar_portfolio` checks if all the necessary slots are present in the object, its weight and mean vectors have the same length as the number of rows in the covariance matrix, and that the covariance matrix is square.

Below is an example demonstrating how to use the constructor and validator to create a portfolio using shrinkage estimators for the mean vector and covariance matrix.

```
R> # Defining a target matrix
R> TM <- matrix(0, nrow=p, ncol=p)
R> diag(TM) <- 1/p
R> # Computing the covariance matrix of x
R> Sigma_shr <- CovarEstim(x, type="BGP14", TM=TM)
R> # Computing the means of x
R> means <- MeanEstim(x, type="BOP19", mu_0=mu_0)
R> # Portfolio constructor
R> cust_port_BS_LW <- new_MeanVar_portfolio(mean_vec=means,
+                       cov_mtrx=Sigma_shr, gamma=2)
R> # Validator
R> cust_port_BS_LW <- validate_MeanVar_portfolio(cust_port_BS_LW)
```

The last two lines are equivalent to the following.

```
R> cust_port_BS_LW_2 <- MeanVar_portfolio(mean_vec=means,
+                       cov_mtrx=Sigma_shr, gamma=2)
```

The following example of an improperly designed portfolio, created by modifying the previous one, will trigger the validator to report an error:

```
R> SCM <- matrix(0, nrow=p, ncol=2*p)
R> diag(SCM) <- 1/p
```

```
R> cust_port_BS_LW_3 <- cust_port_BS_LW_2
R> cust_port_BS_LW_3$cov_mtrx <- SCM
R> tools::assertError(
+ validate_MeanVar_portfolio(cust_port_BS_LW_3),
+ classes = "error", verbose = TRUE
+ )
```

```
## Asserted error: The covariance matrix is not square
```

To simplify the creation of portfolios with custom mean vectors and covariance matrices, two dispatchers are provided: `MeanEstim` and `CovarEstim`.

```
R> Mtrx_naive <- CovarEstim(x, type="trad")
R> Mtrx_bgp <- CovarEstim(x, type="BGP14", TM=TM, SCM=Mtrx_naive)
R> Mean_BOP <- MeanEstim(x, type="BOP19", mu_0=mu_0)
R> Mean_BOP <- MeanEstim(x, type="bs")
```

5. Empirical illustration

In this section, we illustrate the methods of the package on real data. For this purpose, we analyze daily log-returns of S&P 500 constituents from January 2005 to January 2025 (5230 observations). The data were obtained from Thomson Reuters Eikon. This time period is characterized by significant financial turbulence, covering multiple periods of crises. Since not all 500 assets were tradable over the whole period, we chose 300 stocks with the longest history.

To emphasize the importance of various estimation techniques, we start with a full sample analysis. Figure 4 shows the individual assets (as dots) in the mean-standard deviation space with the parameters estimated in the traditional way. The optimal portfolio weights depend on the mean vector and the covariance matrix of asset returns. These can be estimated either traditionally or by relying on robust alternatives described above. The corresponding efficient frontiers for different pairs of such estimators are visualized in the figure. The legend indicates first the estimator for the mean and then the estimator for the covariance matrix. For the shrinkage-based estimators, we choose the vector proportional to the vector with ones as the target for expected returns and the identity matrix as the target for the covariance matrix. Naturally, other choices, such as zero-correlation or equal-correlation covariance matrices, can also be considered.

From a financial perspective, the most desirable efficient frontier is the highest and leftmost frontier, as it contains portfolios with the highest returns for a given level of risk. However, for empirical data, this frontier may be affected by biases and noise, leading to extremely poor out-of-sample performance. In contrast, with simulated data, we can compare the estimated efficient frontiers to the true frontier based on known parameters

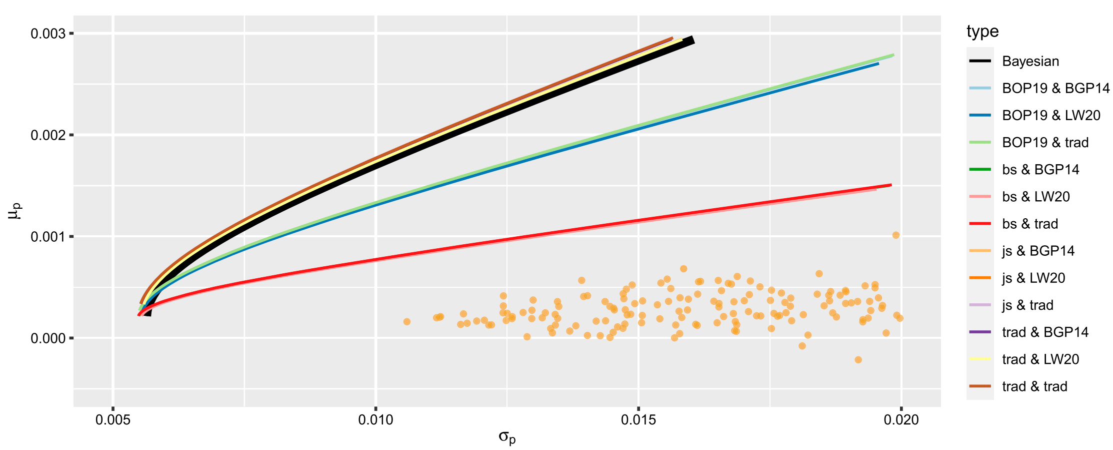

Figure 4. Efficient frontiers based on various estimators, utilizing daily returns from 300 stocks between January 2005 and January 2025.

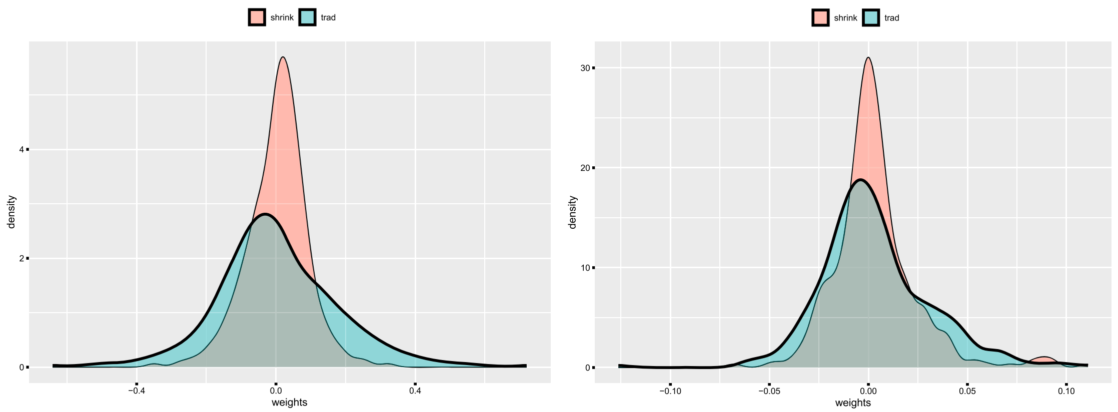

Figure 5. KDE of traditional and shrinkage estimators of the portfolio weights for MV with $\gamma = 5$ (left) and $\gamma = 50$ (right). The equally weighted portfolio serves as a target. Daily returns for 300 stocks for the full sample are used.

– an approach that is not possible with real data. To address this issue, we compare the various frontiers to a consistent estimator of the frontier. The black solid line shows the consistent Bayesian estimator suggested in Bauder et al. (2021). Given that the curves produced by the same mean estimator appear highly similar, it is evident that the shape of the efficient frontier is predominantly influenced by the mean estimator rather than the covariance estimator. Furthermore, some estimators lead to a substantial bias in the frontier, for example, the James-Stein estimator. The traditional estimator, however, shows a good performance leading to a frontier that is very close to the Bayesian one.

Next, we consider the shrinkage type estimators for the portfolio weights with the equally weighted portfolio (with weights $1/p = 0.0033$) as the target. It is well known, that the shrinkage approach stabilizes the portfolio weights. To visualize this effect, we plot the KDEs (kernel density estimators) of the traditional and shrinkage estimated portfolio weights of the 300 assets in Figure 5. The KDE is defined at point y by

$$\hat{f}_h(y) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{y - y_n}{h}\right)$$

for a given sample $y_1, \ldots, y_n$ where $K(\cdot)$ is a kernel, i.e. $K(\cdot)$ is a density symmetric around zero. The quantity h is the bandwidth parameter and is usually chosen via cross-validation in practice. Further details about the kernel density estimation can be found in Wasserman (2006). The left hand-side plot in Figure 5 depicts the KDEs for the minimum-variance portfolio for $\gamma = 5$, whereas the right figure is for the portfolio with $\gamma = 50$. The KDEs for the shrinkage estimator are noticeably narrower, demonstrating the stabilizing impact of the shrinkage technique.

The dynamics of the estimated shrinkage intensity are of great importance for the assessment of the estimation risk. For this purpose, we estimate the shrinkage parameter as in (29) using a 750-day moving estimation window and setting the risk aversion coefficient to 50. Figure 6 shows the dynamics of the shrinkage intensity and allows for additional financial insights. We see that the shrinkage intensity is high during the European debt crisis and the COVID-19 pandemic. This seems to be surprising since the volatility of returns is high in these periods and the equally weighted portfolio is believed to reduce the risk. However, the increase in the volatility is so high that minimum-variance portfolios appear to be more stable and less risky compared to the equally weighted portfolio.

The package offers a possibility to test the significance of the shrinkage intensity using the test in (34)–(35). Utilizing the asymptotic distribution of the estimator, we construct and illustrate pointwise confidence intervals in Figure 6. Despite of a short estimation window and large portfolio, the confidence intervals are relatively narrow.

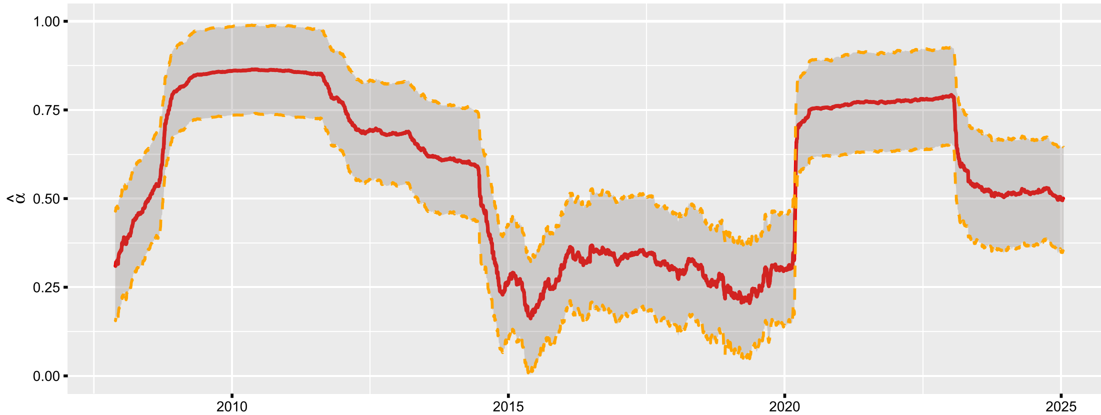

Figure 6. Moving window estimator of optimal shrinkage intensity for the shrinkage portfolio with corresponding 95% confidence intervals. The estimation window is set to 750 days and the risk aversion coefficient is 100.

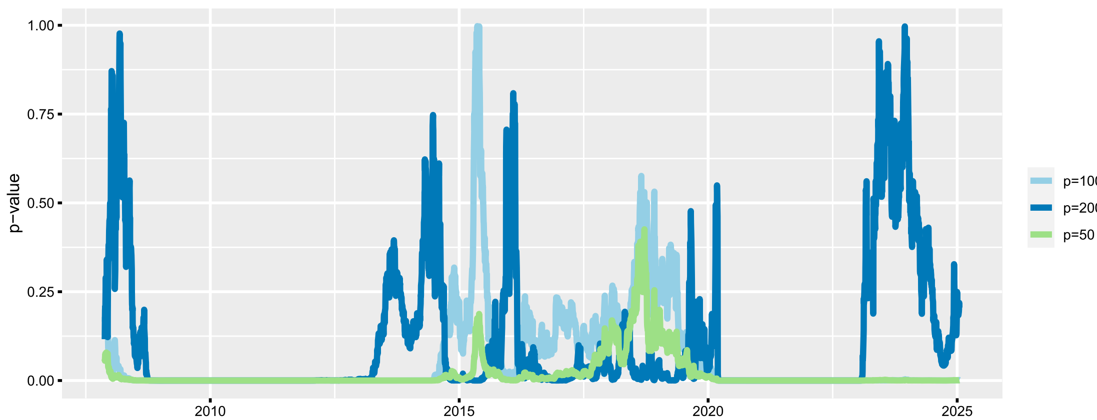

Figure 7. *p*-values of the test in (34) for different portfolio sizes and risk aversion coefficients. The estimation window is set to 750 days and the risk aversion coefficient to 50.

The significance of the difference between the mean-variance portfolio and the equally weighted portfolio over time is shown in Figure 7. Here, we plot the *p*-value of the significance test for different portfolio sizes and a risk aversion coefficient of $\gamma = 50$. As above, we use an estimation window of 750 days. Note that changing the portfolio size results in a completely different portfolio, and the shrinkage intensities are not directly comparable. Nevertheless, the figure provides interesting insights into the sensitivity of shrinkage estimation to market volatility and portfolio size.

First, the intensities are highly significant for all portfolio sizes during crisis periods. This implies that minimizing variance is crucial for all portfolio sizes in turbulent times. Second, in calm phases, the equally weighted portfolio leads to lower variance and can be preferred or is equivalent to large minimum-variance portfolios, while small portfolios remain significantly different. Third, the results heavily depend on the choice of stocks. As seen in the period 2016-2017, portfolios with 50 or 200 assets are significantly different from the equally weighted portfolio, whereas the portfolio with 100 stocks is not. This suggests that it is not possible to make a uniform recommendation regarding the optimal portfolio size.

To assess the out-of-sample economic performance of different estimators, we consider a dynamic portfolio strategy. We consider two holding periods of 5 or 25 days. At the end of every holding period the parameters are reestimated using the last 750 observations. The returns at the end of every holding period are use to calculate various performance and risk measures summarized in Tables 2 and 3 for holding periods of 5 and 25

Table 2. Out-of-sample performance measures of different estimation techniques for portfolio selection with 300 assets and estimation window of 750 trading days.

	trad/trad	trad/BGP14	trad/LW20	bs/trad	bs/BGP14	bs/LW20	js/trad	js/BGP14	js/LW20	BOP19/trad	BOP19/BGP14	BOP19/LW20
$\gamma = 20$												
mean	1e-04	2e-04	4e-04	3e-04	3e-04	3e-04	1e-04	2e-04	4e-04	0.0011	0.001	6e-04
sd	0.0456	0.0432	0.0356	0.0175	0.0167	0.0143	0.0456	0.0432	0.0356	0.0475	0.0448	0.0364
VaR 0.95	−0.0629	−0.0611	−0.0535	−0.0263	−0.0253	−0.022	−0.0629	−0.0611	−0.0535	−0.0719	−0.0681	−0.0561
VaR 0.99	−0.2121	−0.193	−0.141	−0.0756	−0.0706	−0.0585	−0.2121	−0.193	−0.141	−0.2319	−0.2007	−0.1283
ES 0.95	−0.0629	−0.0703	−0.0829	−0.0429	−0.0422	−0.0384	−0.0629	−0.0703	−0.0829	−0.1302	−0.12	−0.089
ES 0.99	−0.2121	−0.193	−0.141	−0.0756	−0.0706	−0.0585	−0.2121	−0.193	−0.141	−0.2319	−0.2007	−0.1283
Sharpe	0.0029	0.0042	0.0104	0.0161	0.0177	0.0242	0.0029	0.0042	0.0104	0.0232	0.0226	0.0172
Sortino	0.0041	0.0059	0.0146	0.0225	0.0248	0.0338	0.0041	0.0059	0.0146	0.0335	0.0328	0.025
$\gamma = 50$												
mean	2e-04	2e-04	3e-04	3e-04	3e-04	3e-04	2e-04	2e-04	3e-04	6e-04	6e-04	4e-04
sd	0.0197	0.0188	0.0159	0.0104	0.0101	0.0092	0.0197	0.0188	0.0159	0.0206	0.0195	0.0163
VaR 0.95	−0.0293	−0.0283	−0.0247	−0.0164	−0.0159	−0.0145	−0.0293	−0.0283	−0.0247	−0.0319	−0.0304	−0.0258
VaR 0.99	−0.0828	−0.0762	−0.0613	−0.0438	−0.0428	−0.0417	−0.0828	−0.0762	−0.0613	−0.103	−0.0925	−0.0679
ES 0.95	−0.0443	−0.0451	−0.0423	−0.0313	−0.0306	−0.0294	−0.0443	−0.0451	−0.0423	−0.0651	−0.0608	−0.0488
ES 0.99	−0.0828	−0.0762	−0.0613	−0.0438	−0.0428	−0.0417	−0.0828	−0.0762	−0.0613	−0.103	−0.0925	−0.0679
Sharpe	0.0109	0.0125	0.0198	0.0265	0.0278	0.0332	0.0109	0.0125	0.0198	0.0293	0.029	0.0256
Sortino	0.0152	0.0174	0.0276	0.0369	0.0388	0.0461	0.0152	0.0174	0.0276	0.0418	0.0415	0.0365
$\gamma = 100$												
mean	2e-04	3e-04	3e-04	3e-04	3e-04	3e-04	2e-04	3e-04	3e-04	4e-04	4e-04	3e-04
sd	0.0197	0.0118	0.0104	0.0089	0.0087	0.0082	0.0122	0.0118	0.0104	0.0126	0.0122	0.0106
VaR 0.95	−0.0293	−0.0184	−0.0165	−0.0143	−0.014	−0.013	−0.0191	−0.0184	−0.0165	−0.0201	−0.0193	−0.0169
VaR 0.99	−0.0828	−0.047	−0.0432	−0.0416	−0.0419	−0.0437	−0.0492	−0.047	−0.0432	−0.0669	−0.0644	−0.0577
ES 0.95	−0.0443	−0.0331	−0.0312	−0.0302	−0.0301	−0.0297	−0.0339	−0.0331	−0.0312	−0.0461	−0.0444	−0.0396
ES 0.99	−0.0828	−0.047	−0.0432	−0.0416	−0.0419	−0.0437	−0.0492	−0.047	−0.0432	−0.0669	−0.0644	−0.0577
Sharpe	0.0109	0.0214	0.0283	0.0304	0.0314	0.0354	0.0198	0.0214	0.0283	0.0344	0.0344	0.0325
Sortino	0.0152	0.0299	0.0394	0.0422	0.0435	0.0487	0.0276	0.0299	0.0394	0.0484	0.0483	0.0454

Note: The holding period is 5 days.

Table 3. Out-of-sample performance measures of different estimation techniques for portfolio selection with 300 assets and estimation window of 750 trading days.

	trad/trad	trad/BGP14	trad/LW20	bs/trad	bs/BGP14	bs/LW20	js/trad	js/BGP14	js/LW20	BOP19/trad	BOP19/BGP14	BOP19/LW20
$\gamma = 20$												
mean	6e-04	6e-04	6e-04	4e-04	4e-04	4e-04	6e-04	6e-04	6e-04	5e-04	4e-04	3e-04
sd	0.0465	0.0440	0.0362	0.0183	0.0175	0.0149	0.0465	0.0440	0.0362	0.0474	0.0448	0.0364
VaR 0.95	−0.0651	−0.0633	−0.0559	−0.0270	−0.0259	−0.0227	−0.0651	−0.0633	−0.0559	−0.0726	−0.0688	−0.0564
VaR 0.99	−0.2150	−0.1961	−0.1463	−0.0899	−0.0849	−0.0714	−0.2150	−0.1961	−0.1463	−0.2346	−0.2033	−0.1300
ES 0.95	−0.0721	−0.0841	−0.0961	−0.0433	−0.0430	−0.0428	−0.0721	−0.0841	−0.0961	−0.1330	−0.1224	−0.0898
ES 0.99	−0.215	−0.1961	−0.1463	−0.0899	−0.0849	−0.0714	−0.2150	−0.1961	−0.1463	−0.2346	−0.2033	−0.1300
Sharpe	0.0124	0.0135	0.0178	0.0230	0.0243	0.0289	0.0124	0.0135	0.0178	0.0102	0.0100	0.0071
Sortino	0.0174	0.0189	0.0248	0.0322	0.0340	0.0403	0.0174	0.0189	0.0248	0.0146	0.0143	0.0103
$\gamma = 50$												
mean	4e-04	4e-04	4e-04	3e-04	3e-04	3e-04	4e-04	4e-04	4e-04	4e-04	3e-04	3e-04
sd	0.0204	0.0195	0.0164	0.0111	0.0108	0.0097	0.0204	0.0195	0.0164	0.0206	0.0196	0.0164
VaR 0.95	−0.0301	−0.0290	−0.0255	−0.0160	−0.0155	−0.0141	−0.0301	−0.0290	−0.0255	−0.0325	−0.0310	−0.0262
VaR 0.99	−0.0944	−0.0880	−0.0719	−0.0674	−0.0668	−0.0641	−0.0944	−0.088	−0.0719	−0.1046	−0.0948	−0.0719
ES 0.95	−0.0463	−0.0475	−0.0477	−0.0247	−0.0231	−0.0228	−0.0463	−0.0475	−0.0477	−0.0684	−0.0639	−0.0513
ES 0.99	−0.0944	−0.0880	−0.0719	−0.0674	−0.0668	−0.0641	−0.0944	−0.0880	−0.0719	−0.1046	−0.0948	−0.0719
Sharpe	0.0195	0.0208	0.0259	0.0304	0.0313	0.0349	0.0195	0.0208	0.0259	0.0175	0.0176	0.0165
Sortino	0.0273	0.0291	0.0361	0.0424	0.0437	0.0484	0.0273	0.0291	0.0361	0.0247	0.0249	0.0233
$\gamma = 100$												
mean	4e-04	3e-04	4e-04	3e-04	3e-04	3e-04	3e-04	3e-04	4e-04	3e-04	3e-04	3e-04
sd	0.0204	0.0124	0.0110	0.0096	0.0094	0.0088	0.0129	0.0124	0.0110	0.0129	0.0124	0.0109
VaR 0.95	−0.0301	−0.0184	−0.0165	−0.0136	−0.0132	−0.0122	−0.0191	−0.0184	−0.0165	−0.0206	−0.0197	−0.0171
VaR 0.99	−0.0944	−0.0666	−0.0614	−0.0674	−0.0686	−0.069	−0.0684	−0.0666	−0.0614	−0.0715	−0.0707	−0.0678
ES 0.95	−0.0463	−0.0313	−0.0311	−0.0172	−0.0148	−0.0127	−0.0321	−0.0313	−0.0311	−0.0485	−0.0466	−0.0398
ES 0.99	−0.0944	−0.0666	−0.0614	−0.0674	−0.0686	−0.0690	−0.0684	−0.0666	−0.0614	−0.0715	−0.0707	−0.0678
Sharpe	0.0195	0.0275	0.0321	0.0321	0.0328	0.0354	0.0263	0.0275	0.0321	0.0249	0.0251	0.0252
Sortino	0.0273	0.0384	0.0446	0.0446	0.0454	0.0487	0.0368	0.0384	0.0446	0.0346	0.0349	0.0348

Note: The holding period is 25 days.

days respectively. We conclude that the Bayesian estimator for the mean return and the Ledoit and Wolf (2020) estimator for the covariance matrix lead to the best results. The second best model is the combination of the mean estimator of Bodnar, Okhrin, and Parolya (2019) and the covariance estimator of Bodnar, Gupta, and Parolya (2014). This evidence is consistent over different risk aversion coefficients and holding periods. However, as the risk aversion coefficient increases the the discrepancy between the models diminishes. Note that the results and conclusions depend on the targets for the shrinkage estimators, but shrinkage estimators generally improve the performance of portfolios compared to the traditional estimators.

6. Summary

Mean-variance analysis is a fundamental tool in portfolio theory, enriched by numerous applications and extensions detailed in the literature. It has recently been acknowledged that estimation risk can equal or surpass model risk when models are practically applied, and parameters necessary for constructing optimal portfolios are estimated using historical asset return data. This issue intensifies in high-dimensional settings, i.e. when the dimension of the holding portfolio is comparable to the sample size.

In this paper, we present theoretical results developed to determine the structure of high-dimensional optimal portfolios. The approach outlined utilizes shrinkage estimation and shrinkage-based test theory. We explore various shrinkage estimators, including those derived for the mean vector, covariance matrix, and precision matrix, as well as those directly formulated for the weights of optimal portfolios. These findings are also instrumental in constructing shrinkage-based tests on the mean-variance efficiency of a given portfolio.

The newly developed high-dimensional optimal portfolio theory is implemented in the R package `HDShOP`, which is detailed in this paper. The `HDShOP` package leverages recent theoretical advancements to determine high-dimensional optimal portfolios and provides quantitative tools needed to construct improved estimators of optimal portfolio weights. Additionally, it includes statistical tests on the mean-variance efficiency of the holding portfolio. The application of the `HDShOP` package to real-life problems is demonstrated through several practical examples based on data from returns on stocks included in the S&P 500 index.

The results of the paper can be extended in several directions. One of possible lines of future research deals with introducing constraints on the structure of the weights of the high-dimensional mean-variance optimal portfolios. For example, imposing linear constraints may significantly reduce the estimation error present in the sample portfolio weights as documented in Jagannathan and Ma (2003). Another line of research may be related to the construction of the high-dimensional dynamic shrinkage estimators generalizing the recent findings of Bodnar, Parolya, and Thorsén (2023) which are derived for the weights of the high-dimensional global minimum variance portfolio.

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.

Disclosure statement

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

Data availability statement

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

ORCID

Taras Bodnar http://orcid.org/0000-0001-7855-8221
Solomiia Dmytriv http://orcid.org/0000-0003-1855-3044
Yarema Okhrin http://orcid.org/0000-0003-4704-5233
Dmitry Otryakhin http://orcid.org/0000-0002-4700-7221
Nestor Parolya http://orcid.org/0000-0003-2147-2288

References

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. J. 2015. "Statistical Properties and Tests of Efficient Frontier Portfolios." In *Quantitative Financial Risk Management*, edited by C. Zopounidis and E. Galariotis, 242–269. New Jersey: John Wiley & Sons, Ltd.

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.

Aitchison, J. 1964. "Confidence-Region Tests." *Journal of the Royal Statistical Society: Series B (Methodological)* 26 (3): 462–476. https://doi.org/10.1111/j.2517-6161.1964.tb00576.x.

Alexander, G. J., and A. M. Baptista. 2002. "Economic Implications of Using a Mean-Var Model for Portfolio Selection: A Comparison with Mean-Variance Analysis." *Journal of Economic Dynamics and Control* 26 (7): 1159–1193. https://doi.org/10.1016/S0165-1889(01)00041-0.

Alexander, G. J., and A. M. Baptista. 2004. "A Comparison of Var and Cvar Constraints on Portfolio Selection with the Mean-Variance Model." *Management Science* 50 (9): 1261–1273. https://doi.org/10.1287/mnsc.1040.0201.

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., and J. W. Silverstein. 2010. *Spectral Analysis of Large Dimensional Random Matrices*. New York: Springer.

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.

Ben-Israel, A., and T. N. Greville. 2003. *Generalized Inverses: Theory and Applications*. Vol. 15. New York: Springer Science & Business Media.

Best, M. J., and R. R. Grauer. 1991. "On the Sensitivity of Mean-Variance-efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results." *The Review of Financial Studies* 4 (2): 315–342. https://doi.org/10.1093/rfs/4.2.315.

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.

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 (01): 2250008. https://doi.org/10.1142/S2010326322500083.

Bodnar, T., S. Dmytriv, Y. Okhrin, D. Otryakhin, and N. Parolya. 2024. *HDShOP: High-Dimensional Shrinkage Optimal Portfolios*. R package version 0.1.5.

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., A. K. 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. K. Gupta, and N. Parolya. 2016. "Direct Shrinkage Estimation of Large Dimensional Precision Matrix." *Journal of Multivariate Analysis* 146:223–236. https://doi.org/10.1016/j.jmva.2015.09.010.

Bodnar, T., N. Hautsch, Y. Okhrin, and N. Parolya. 2025. "Consistent Estimation of the High-Dimensional Efficient Frontier." *The European Journal of Finance*, to appear. https://doi.org/10.48550/arXiv.2409.15103.

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., O. Okhrin, and N. Parolya. 2019. "Optimal Shrinkage Estimator for High-Dimensional Mean Vector." *Journal of Multivariate Analysis* 170:63–79. https://doi.org/10.1016/j.jmva.2018.07.004.

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., and N. Parolya. 2024. "Reviving Pseudo-Inverses: Asymptotic Properties of Large Dimensional Moore-Penrose and Ridge-Type Inverses with Applications." *arXiv preprint arXiv:2403.15792*.

Bodnar, T., N. Parolya, and W. Schmid. 2013. "On the Equivalence of Quadratic Optimization Problems Commonly Used in Portfolio Theory." *European Journal of Operational Research* 229 (3): 637–644. https://doi.org/10.1016/j.ejor.2013.03.002.

Bodnar, T., N. Parolya, and W. Schmid. 2015. "On the Exact Solution of the Multi-Period Portfolio Choice Problem for An Exponential Utility under Return Predictability." *European Journal of Operational Research* 246 (2): 528–542. https://doi.org/10.1016/j.ejor.2015.04.039.

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 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.

Britten-Jones, M. 1999. "The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights." *The Journal of Finance* 54 (2): 655–671. https://doi.org/10.1111/jofi.1999.54.issue-2.

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.

Chen, Y., A. Wiesel, Y. C. Eldar, and A. O. Hero. 2010. "Shrinkage Algorithms for Mmse Covariance Estimation." *IEEE Transactions on Signal Processing* 58 (10): 5016–5029. https://doi.org/10.1109/TSP.2010.2053029.

Chopra, V. K., and W. T. Ziemba. 1993. "The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice." *The Journal of Portfolio Management* 19 (2): 6–11. https://doi.org/10.3905/jpm.1993.409440.

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.

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.

Galanos, A., and B. Pfaff. 2022. `parma`: *Portfolio Allocation and Risk Management Applications*. R package version 1.7.

Garcia, F., J. González-Bueno, J. Oliver, and R. Tamošiūnienė. 2019. "A Credibilistic Mean-Semivariance-per Portfolio Selection Model for Latin America." *Journal of Business Economics and Management* 20 (2): 225–243. https://doi.org/10.3846/jbem.2019.8317.

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.

Harville, D. A. 1997. *Matrix Algebra from a Statistician's Perspective*. Vol. 1. New York: Springer.

Huang, X., and D. Ma. 2023. "Uncertain Mean-Chance Model for Portfolio Selection with Multiplicative Background Risk." *International Journal of Systems Science: Operations & Logistics* 10 (1): 2158443.

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.

Jorion, P. 1986. "Bayes-Stein Estimation for Portfolio Analysis." *Journal of Financial and Quantitative Analysis* 21 (3): 279–292. https://doi.org/10.2307/2331042.

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.

Kourtis, A., G. Dotsis, and R. N. Markellos. 2012. "Parameter Uncertainty in Portfolio Selection: Shrinking the Inverse Covariance Matrix." *Journal of Banking & Finance* 36 (9): 2522–2531. https://doi.org/10.1016/j.jbankfin.2012.05.005.

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.

Ledoit, O., and M. Wolf. 2003. "Improved Estimation of the Covariance Matrix of Stock Returns with An Application to Portfolio Selection." *Journal of Empirical Finance* 10 (5): 603–621. https://doi.org/10.1016/S0927-5398(03)00007-0.

Ledoit, O., and M. Wolf. 2004. "A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices." *Journal of Multivariate Analysis* 88 (2): 365–411. https://doi.org/10.1016/S0047-259X(03)00096-4.

Ledoit, O., and M. Wolf. 2015. "Spectrum Estimation: A Unified Framework for Covariance Matrix Estimation and PCA in Large Dimensions." *Journal of Multivariate Analysis* 139:360–384. https://doi.org/10.1016/j.jmva.2015.04.006.

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

Ledoit, 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.

Lee, K., and K. You. 2021. `CovTools`: *Statistical Tools for Covariance Analysis*. R package version 0.5.4.

Markowitz, H. 1952. "Portfolio Selection." *The Journal of Finance* 7 (1): 77–91.

Merton, R. C. 1972. "An Analytic Derivation of the Efficient Portfolio Frontier." *The Journal of Financial and Quantitive Analysis* 7 (4): 1851–1872. https://doi.org/10.2307/2329621.

Merton, R. C. 1980. "On Estimating the Expected Return on the Market: An Exploratory Investigation." *Journal of Financial Economics* 8 (4): 323–361. https://doi.org/10.1016/0304-405X(80)90007-0.

Meyer Jr, C. D. 1973. "Generalized Inversion of Modified Matrices." *SIAM Journal on Applied Mathematics* 24 (3): 315–323. https://doi.org/10.1137/0124033.

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.

Palczewski, A. 2019. `PortfolioOptim`: *Small/Large Sample Portfolio Optimization*. R package version 1.1.1.

Pav, S. E. 2023. `MarkowitzR`: *Statistical Significance of the Markowitz Portfolio*. R package version 1.0.3.

Penrose, R. 1955. "A Generalized Inverse for Matrices." In *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. 51, 406–413. Cambridge University Press. https://doi.org/10.1017/S0305004100030401.

Ramprasad, P. 2016. `nlshrink`: *Non-Linear Shrinkage Estimation of Population Eigenvalues and Covariance Matrices*. R package version 1.0.1.

Reckziegel, B. 2023. `epo`: *Enhanced Portfolio Optimization*. R package version 0.1.0.

Ruhe, A. 1970. "Perturbation Bounds for Means of Eigenvalues and Invariant Subspaces." *BIT Numerical Mathematics* 10 (3): 343–354. https://doi.org/10.1007/BF01934203.

Stein, C. 1956. "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution." In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, 197–206. University of California Press: Berkeley, California.

Touloumis, A. 2015. "Nonparametric Stein-Type Shrinkage Covariance Matrix Estimators in High-Dimensional Settings." *Computational Statistics & Data Analysis* 83:251–261. https://doi.org/10.1016/j.csda.2014.10.018.

Wang, X., B. Wang, T. Li, H. Li, and J. Watada. 2023. "Multi-Criteria Fuzzy Portfolio Selection Based on Three-Way Decisions and Cumulative Prospect Theory." *Applied Soft Computing* 134:110033. https://doi.org/10.1016/j.asoc.2023.110033.

Wang, G., Y. Wei, S. Qiao, P. Lin, and Y. Chen. 2018. *Generalized Inverses: Theory and Computations*. Vol. 53. Singapore: Springer.

Wasserman, L. 2006. *All of Nonparametric Statistics*. New York: Springer Science & Business Media.

Wuertz, D., T. Setz, Y. Chalabi, and S. Theussl. 2023. `fPortfolio`: *Rmetrics – Portfolio Selection and Optimization*. R package version 4023.84.