WORKING PAPER 15/2022 (ECONOMICS AND STATISTICS)

Estimation of optimal portfolio compositions for small sample and singular covariance matrix

Taras Bodnar, Stepan Mazur and Hoang Nguyen

ISSN 1403-0586

Örebro University School of Business
SE-701 82 Örebro, Sweden

Estimation of optimal portfolio compositions for small sample and singular covariance matrix

Taras Bodnar[a], Stepan Mazur[b,c,1] and Hoang Nguyen[b]

[a] *Department of Mathematics, Stockholm University, Stockholm, Sweden*

[b] *Unit of Statistics, Unit of Statistics, School of Business, Örebro University, Örebro, Sweden*

[c] *Department of Economics and Statistics, School of Business and Economics, Linnaeus University, Sweden*

Abstract

In the paper we consider the optimal portfolio choice problem under parameter uncertainty when the covariance matrix of asset returns is singular. Very useful stochastic representations are deduced for the characteristics of the expected utility optimal portfolio. Using these stochastic representations, we derive the moments of higher order of the estimated expected return and the estimated variance of the expected utility optimal portfolio. Another line of applications leads to their asymptotic distributions obtained in the high-dimensional setting. Via a simulation study, it is shown that the derived high-dimensional asymptotic distributions provide good approximations of the exact ones even for moderate sample sizes.

JEL-codes: G11

Keywords: singular Wishart distribution, mean-variance portfolio, Moore-Penrose inverse

[1]Corresponding author. E-mail address: stepan.mazur@oru.se

1 Introduction

Mean-variance analysis of Harry Markowitz [32] plays important role in portfolio analysis. It is widely used by practitioners and it is also a hot topic by researchers of financial sector. The original idea of Markowitz [32] was to select a portfolio which minimizes the variance for a given level of the expected return. Recently, it was shown that different optimization problems lead to the same set of optimal portfolios (see, [18]). In particular, the set of mean-variance optimal portfolios, the so-called efficient frontier, can also be obtained by an investor who maximizes the exponential utility function under the assumption that the vector of asset returns follows a multivariate normal distribution. In this case, the optimization problem is given by

$$\max_{\mathbf{w}:\mathbf{w}^\top \mathbf{1}_k=1} \left[\mathbf{w}^\top \boldsymbol{\mu} - \frac{\alpha}{2}\mathbf{w}^\top \boldsymbol{\Sigma}\mathbf{w}\right], \tag{1}$$

where $\boldsymbol{\mu}$ denotes the vector of the k-dimensional expected asset returns, $\boldsymbol{\Sigma}$ is the $k \times k$ covariance matrix, and α is the risk-aversion coefficient which describes the investor attitude towards risk. Changing the risk aversion coefficient from zero to infinity, we get the same set of optimal portfolios as the one obtained by Markowitz's mean-variance optimization problem.

The solution of (1) is given by

$$\mathbf{w}_{EU} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}_k}{\mathbf{1}_k^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}_k} + \alpha^{-1}\mathbf{R}\boldsymbol{\mu}, \tag{2}$$

where

$$\mathbf{R} = \boldsymbol{\Sigma}^{-1} - \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}_k\mathbf{1}_k^\top \boldsymbol{\Sigma}^{-1}}{\mathbf{1}_k^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}_k}.$$

The portfolio with weights (2) is known in financial literature as the expected utility (EU) optimal portfolio (see, e.g., [29], [34], [17]).

In practice, however, the weights of the EU optimal portfolio (2) cannot be directly implemented, since the formula of the weights depends on the unknown quantities $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. As such, the investor should use the historical data of asset returns to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ before the optimal portfolio is constructed. The most commonly used estimators for the mean vector of the asset returns and for the covariance matrix are their sample counterparts expressed as

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i \quad \text{and} \quad \mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top, \tag{3}$$

where $\mathbf{x}_1, ..., \mathbf{x}_n$ is the sample of the asset returns. Then, the weights of the EU optimal portfolio are estimated by

$$\widehat{\mathbf{w}}_{EU} = \frac{\mathbf{S}^{-1}\mathbf{1}_k}{\mathbf{1}_k^\top \mathbf{S}^{-1}\mathbf{1}_k} + \alpha^{-1}\widehat{\mathbf{R}}\bar{\mathbf{x}} \quad \text{with} \quad \widehat{\mathbf{R}} = \mathbf{S}^{-1} - \frac{\mathbf{S}^{-1}\mathbf{1}_k\mathbf{1}_k^\top\mathbf{S}^{-1}}{\mathbf{1}_k^\top \mathbf{S}^{-1}\mathbf{1}_k}. \tag{4}$$

Similarly,

$$\widehat{R}_{EU} = \frac{\bar{\mathbf{x}}^\top\mathbf{S}^{-1}\mathbf{1}_k}{\mathbf{1}_k^\top \mathbf{S}^{-1}\mathbf{1}_k} + \alpha^{-1}\bar{\mathbf{x}}^\top\widehat{\mathbf{R}}\bar{\mathbf{x}} \quad \text{and} \quad \widehat{V}_{EU} = \frac{1}{\mathbf{1}_k^\top \mathbf{S}^{-1}\mathbf{1}_k} + \alpha^{-2}\bar{\mathbf{x}}^\top\widehat{\mathbf{R}}\bar{\mathbf{x}} \tag{5}$$

are the sample estimators of the expected return and the variance of the EU optimal portfolio expressed as

$$R_{EU} = \frac{\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}_k}{\mathbf{1}_k^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}_k} + \alpha^{-1}\boldsymbol{\mu}^\top\mathbf{R}\boldsymbol{\mu} \quad \text{and} \quad V_{EU} = \frac{1}{\mathbf{1}_k^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}_k} + \alpha^{-2}\boldsymbol{\mu}^\top\mathbf{R}\boldsymbol{\mu}. \tag{6}$$

The first two moments of the sample estimator of the EU portfolio weights together with their asymptotic distribution were obtained in [34], while the exact distribution of (4) was derived in [21]. Recently, the asymptotic distribution of the estimated weights as well as their consistent estimator in the high-dimensional setting were deduced in [11]. The exact distributions of the sample estimator for the expected return and the variance of the EU portfolio were provided in [20]. Finally, a shrinkage-based estimator and test theory for the EU portfolio weights were developed in [17] and [10], who extended the shrinkage-based approach applied to other optimal portfolio weights in [25], [24] and [19]. For the practical implementation of the derived estimators the investor can use the R package *HDShOP* ([9]), while [8] presented the detailed review of the estimation procedures.

Another line of research related to the estimation of optimal portfolio weights and characteristics leads to the Bayesian statistics. The research in this direction started with the papers of [40], [2] and [31]. Both informative and noninformative priors have been used in the literature. While the hyperparameter prior approach closely related to the Bayes–Stein shrinkage prior was considered in [30], the economical motivation for the usage of an informative prior was provided in [7]. Other informative priors were considered in [36], [37], [38], among other. Noninformative priors were employed in the papers of [13] and [4]. Finally, the Bayesian estimator of the efficient frontier was considered in the papers of [3] and [5].

We contribute to the existent literature by deriving the properties of the estimated EU optimal portfolio when the covariance matrix is singular. While the previous studies assume that the covariance matrix is positive definite, recently several papers deal with the singular case (see, e.g., [35], [26], [27]). The estimated global minimum variance portfolio with singular covariance matrix was considered in [15], while the results for the tangency portfolio were derived in [16]. Finally, several important properties of singular Wishart distribution were derived in [12] which were implemented in portfolio theory by

[14].

The rest of the paper is organised as follows. In the next section, the main findings are presented. In Section 2.1 the results for finite-sample case are provided, while the asymptotic distributions of the estimated EU optimal portfolio characteristics, obtained in the high-dimensional setting, are given in Section 2.2. The quality of the asymptotic approximation is investigated via simulations in Section 3. The final remarks are present in Section 4.

2 Sample EU optimal portfolio for singular covariance matrix

In the following it is assumed that the population covariance matrix $\mathbf{\Sigma}$ is singular with $rank(\mathbf{\Sigma}) = r < n$, where n is the sample size. Since $\mathbf{\Sigma}$ is not invertible, the formulas for the EU weights and its characteristics (2) and (6) cannot be used unless the ordinary inverse is replace by a generalized inverse, for example by the Moore-Penrose inverse $\mathbf{\Sigma}^+$.

The Moore-Penrose inverse of a matrix $\mathbf{A}$ is defined as the matrix $\mathbf{A}^+$ which fulfills the following four conditions

(i) $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$,

(ii) $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$,

(iii) $(\mathbf{A}^+\mathbf{A})^\top = \mathbf{A}^+\mathbf{A}$,

(iv) $(\mathbf{A}\mathbf{A}^+)^\top = \mathbf{A}\mathbf{A}^+$.

It is important to note that the Moore-Penrose inverse is uniquely defined for every matrix $\mathbf{A}$. Replacing the ordinary inverse of $\mathbf{\Sigma}$ by its Moore-Penrose inverse we get the weights of the EU portfolio expressed as

$$\mathbf{w}_{EU}^+ = \frac{\mathbf{\Sigma}^+\mathbf{1}_k}{\mathbf{1}_k^\top\mathbf{\Sigma}^+\mathbf{1}_k} + \alpha^{-1}\mathbf{R}^+\boldsymbol{\mu}, \quad \text{with} \quad \mathbf{R}^+ = \mathbf{\Sigma}^+ - \frac{\mathbf{\Sigma}^+\mathbf{1}_k\mathbf{1}_k^\top\mathbf{\Sigma}^+}{\mathbf{1}_k^\top\mathbf{\Sigma}^+\mathbf{1}_k} \tag{7}$$

and its expected return and variance given by

$$R_{EU}^+ = \frac{\boldsymbol{\mu}^\top\mathbf{\Sigma}^+\mathbf{1}_k}{\mathbf{1}_k^\top\mathbf{\Sigma}^+\mathbf{1}_k} + \alpha^{-1}\boldsymbol{\mu}^\top\mathbf{R}^+\boldsymbol{\mu} \quad \text{and} \quad V_{EU}^+ = \frac{1}{\mathbf{1}_k^\top\mathbf{\Sigma}^+\mathbf{1}_k} + \alpha^{-2}\boldsymbol{\mu}^\top\mathbf{R}^+\boldsymbol{\mu}, \tag{8}$$

respectively.

Similarly, the application of the Moore-Penrose inverse of the sample covariance matrix $\mathbf{S}^+$ leads to the estimated portfolio weights for a singular covariance matrix expressed as

$$\widehat{\mathbf{w}}_{EU}^+ = \frac{\mathbf{S}^+\mathbf{1}_k}{\mathbf{1}_k^\top\mathbf{S}^+\mathbf{1}_k} + \alpha^{-1}\widehat{\mathbf{R}}^+\bar{\mathbf{x}} \quad \text{with} \quad \widehat{\mathbf{R}}^+ = \mathbf{S}^+ - \frac{\mathbf{S}^+\mathbf{1}_k\mathbf{1}_k^\top\mathbf{S}^+}{\mathbf{1}_k^\top\mathbf{S}^+\mathbf{1}_k}. \tag{9}$$

The sample estimators for the expected return and the variance of the EU optimal portfolio are then given by

$$\widehat{R}_{EU}^{+}=\frac{\bar{\mathbf{x}}^\top\mathbf{S}^{+}\mathbf{1}_k}{\mathbf{1}_k^\top\mathbf{S}^{+}\mathbf{1}_k}+\alpha^{-1}\bar{\mathbf{x}}^\top\widehat{\mathbf{R}}^{+}\bar{\mathbf{x}} \quad \text{and} \quad \widehat{V}_{EU}^{+} = \frac{1}{\mathbf{1}_k^\top\mathbf{S}^{+}\mathbf{1}_k}+\alpha^{-2}\bar{\mathbf{x}}^\top\widehat{\mathbf{R}}^{+}\bar{\mathbf{x}}. \tag{10}$$

2.1 Results for finite sample

Let

$$R_{GMV}^{+} = \frac{\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{+}\mathbf{1}_k}{\mathbf{1}_k^\top\boldsymbol{\Sigma}^{+}\mathbf{1}_k} \quad \text{and} \quad V_{GMV}^{+}=\frac{1}{\mathbf{1}_k^\top\boldsymbol{\Sigma}^{+}\mathbf{1}_k}$$

be the expected return and the variance of the global minimum variance portfolio when the covariance matrix $\boldsymbol{\Sigma}$ is singular and let

$$s^{+}=\boldsymbol{\mu}^\top\mathbf{R}^{+}\boldsymbol{\mu}$$

be the slope parameter of the efficient frontier.

Let the symbol $\mathcal{N}_m(\mathbf{a},\mathbf{B})$ denote the m-dimensional normal distribution with mean vector $\mathbf{a}$ and covariance matrix $\mathbf{B}$, the symbol χ^2_m stand for the χ^2-distribution with m degrees of freedom and the symbol $\mathcal{F}_{m_1,m_2,c^2}$ denote the noncentral F-distribution with m_1 and m_2 degrees of freedom and noncentrality parameter c^2. In Theorem 1 we derive very useful stochastic representations of $\widehat{R}_{EU}^{+}$ and $\widehat{V}_{EU}^{+}$.

Theorem 1. *Let $\mathbf{x}_1,\ldots,\mathbf{x}_n$ be i.i.d. random vectors with $\mathbf{x}_1\sim\mathcal{N}_k(\boldsymbol{\mu},\boldsymbol{\Sigma})$, $k\geq n$ and $rank(\boldsymbol{\Sigma})=r<n$. Then, stochastic representations of $\widehat{R}_{EU}^{+}$ and $\widehat{V}_{EU}^{+}$ are given by*

$$\widehat{R}_{EU}^{+}\overset{d}{=}R_{GMV}^{+}+\alpha^{-1}\frac{(n-1)(r-1)}{n(n-r+1)}\xi+\sqrt{\frac{1}{n}\left(1+\frac{r-1}{n-r+1}\xi\right)}\sqrt{V_{GMV}^{+}}z_0$$

and

$$\widehat{V}_{EU}^{+}\overset{d}{=}\frac{V_{GMV}^{+}}{n-1}\eta+\alpha^{-2}\frac{(n-1)(r-1)}{n(n-r+1)}\xi,$$

where $\xi\sim\mathcal{F}_{r-1,n-r+1,ns^{+}}$, $\eta\sim\chi^2_{n-r}$, and $z_0\sim\mathcal{N}(0,1)$. Moreover, ξ, η, and z_0 are mutually independently distributed.

Proof of Theorem 1: We start by deriving a stochastic representation of $\widehat{R}_{EU}^{+}$. From Theorem 4 (c) of [14] we know that $\bar{\mathbf{x}}$ and $\mathbf{S}$ are independently distributed. Then, it holds that the distribution of $\widehat{R}_{EU}^{+}|\bar{\mathbf{x}}=\bar{\mathbf{x}}^{*}$ is equal to the distribution of $\widehat{R}_{EU}^{*}$ defined by

$$\widehat{R}_{EU}^{*}=\frac{\mathbf{1}_k^\top\mathbf{S}^{+}\bar{\mathbf{x}}^{*}}{\mathbf{1}_k^\top\mathbf{S}^{+}\mathbf{1}_k}+\alpha^{-1}\bar{\mathbf{x}}^{*\top}\widehat{\mathbf{R}}^{+}\bar{\mathbf{x}}^{*}.$$

Let $\mathbf{L}^\top = (\bar{\mathbf{x}}^*, \mathbf{1}_k)$ such that $rank(\mathbf{L}) = 2$. Then, $\widetilde{\mathbf{\Sigma}} = \mathbf{L}\mathbf{\Sigma}^+\mathbf{L}^\top = \{\tilde{\sigma}_{ij}\}_{i,j=1,2}$ with $\tilde{\sigma}_{11} = \bar{\mathbf{x}}^{*\top}\mathbf{\Sigma}^+\bar{\mathbf{x}}^*$, $\tilde{\sigma}_{12} = \tilde{\sigma}_{21} = \bar{\mathbf{x}}^{*\top}\mathbf{\Sigma}^+\mathbf{1}_k$, and $\tilde{\sigma}_{22} = \mathbf{1}_k^\top\mathbf{\Sigma}^+\mathbf{1}_k$. Similarly, let $\widetilde{\mathbf{S}} = \mathbf{L}\mathbf{S}^+\mathbf{L}^\top = \{\tilde{s}_{ij}\}_{i,j=1,2}$ with $\tilde{s}_{11} = \bar{\mathbf{x}}^{*\top}\mathbf{S}^+\bar{\mathbf{x}}^*$, $\tilde{s}_{12} = \tilde{s}_{21} = \bar{\mathbf{x}}^{*\top}\mathbf{S}^+\mathbf{1}_k$, and $\tilde{s}_{22} = \mathbf{1}_k^\top\mathbf{S}^+\mathbf{1}_k$. Also, let $\mathbf{V} = \widetilde{\mathbf{\Sigma}}^{-1} = \{v_{ij}\}_{i,j=1,2}$ and $\widehat{\mathbf{V}} = \widetilde{\mathbf{S}}^{-1} = \{\hat{v}_{ij}\}_{i,j=1,2}$. Then, the application of Theorem 8.5.11 of [28] leads to

$$\begin{aligned}
\hat{v}_{11} &= \left(\bar{\mathbf{x}}^{*\top}\mathbf{S}^+\bar{\mathbf{x}}^* - \frac{(\bar{\mathbf{x}}^{*\top}\mathbf{S}^+\mathbf{1}_k)^2}{\mathbf{1}_k^\top\mathbf{S}^+\mathbf{1}_k}\right)^{-1} = (\bar{\mathbf{x}}^{*\top}\widehat{\mathbf{R}}^+\bar{\mathbf{x}}^*)^{-1}, \\
\hat{v}_{12} &= -\frac{\bar{\mathbf{x}}^{*\top}\mathbf{S}^+\mathbf{1}_k}{\mathbf{1}_k^\top\mathbf{S}^+\mathbf{1}_k\bar{\mathbf{x}}^{*\top}\widehat{\mathbf{R}}^+\bar{\mathbf{x}}^*}.
\end{aligned}$$

From Theorem 4 (a) of [14] we have that $(n-1)\mathbf{S} \sim \mathcal{W}_k(n-1, \mathbf{\Sigma})$. Then applying Theorem 1 of [14] it holds that

$$(n-1)\widehat{\mathbf{V}} \sim \mathcal{W}_2(n-r+1, \mathbf{V}).$$

From Theorem 3.2.10 of [33] we obtain that $\hat{v}_{22} - \hat{v}_{12}^2\hat{v}_{11}^{-1} = \left(\mathbf{1}_k^\top\mathbf{S}^+\mathbf{1}_k\right)^{-1}$ is independent of $\hat{v}_{11} = (\bar{\mathbf{x}}^{*\top}\widehat{\mathbf{R}}^+\bar{\mathbf{x}}^*)^{-1}$. Consequently, $\mathbf{1}_k^\top\mathbf{S}^+\mathbf{1}_k$ and $\bar{\mathbf{x}}^{*\top}\widehat{\mathbf{R}}^+\bar{\mathbf{x}}^*$ are independently distributed. Moreover, from Corollary 1 of [14] and Theorem 3.2.10 of [33] we get that

$$(n-1)\frac{\mathbf{1}_k^\top\mathbf{\Sigma}^+\mathbf{1}_k}{\mathbf{1}_k^\top\mathbf{S}^+\mathbf{1}_k} \sim \chi^2_{n-r} \quad \text{and} \quad (n-1)\frac{\bar{\mathbf{x}}^{*\top}\mathbf{R}^+\bar{\mathbf{x}}^*}{\bar{\mathbf{x}}^{*\top}\widehat{\mathbf{R}}^+\bar{\mathbf{x}}^*} \sim \chi^2_{n-r+1}. \tag{11}$$

The application of Theorem 3.2.10 by [33] leads to

$$\frac{\bar{\mathbf{x}}^{*\top}\mathbf{S}^+\mathbf{1}_k}{\mathbf{1}_k^\top\mathbf{S}^+\mathbf{1}_k}\frac{n-1}{\bar{\mathbf{x}}^{*\top}\widehat{\mathbf{R}}^+\bar{\mathbf{x}}^*}\Bigg| \frac{n-1}{\bar{\mathbf{x}}^{*\top}\widehat{\mathbf{R}}^+\bar{\mathbf{x}}^*} = u \sim \mathcal{N}\left(\frac{\bar{\mathbf{x}}^{*\top}\mathbf{\Sigma}^+\mathbf{1}_k}{\mathbf{1}_k^\top\mathbf{\Sigma}^+\mathbf{1}_k}u, \frac{1}{\mathbf{1}_k^\top\mathbf{\Sigma}^+\mathbf{1}_k}u\right).$$

Then,

$$\begin{aligned}
&\frac{\bar{\mathbf{x}}^{*\top}\widehat{\mathbf{R}}^+\bar{\mathbf{x}}^*}{n-1}\left(\frac{\bar{\mathbf{x}}^{*\top}\mathbf{S}^+\mathbf{1}_k}{\mathbf{1}_k^\top\mathbf{S}^+\mathbf{1}_k}\frac{n-1}{\bar{\mathbf{x}}^{*\top}\widehat{\mathbf{R}}^+\bar{\mathbf{x}}^*} + (n-1)\alpha^{-1}\right)\Bigg| \frac{n-1}{\bar{\mathbf{x}}^{*\top}\widehat{\mathbf{R}}^+\bar{\mathbf{x}}^*} = u \\
\sim\ &\mathcal{N}\left(\frac{\bar{\mathbf{x}}^{*\top}\mathbf{\Sigma}^+\mathbf{1}_k}{\mathbf{1}_k^\top\mathbf{\Sigma}^+\mathbf{1}_k} + \frac{n-1}{\alpha u}, \frac{1}{\mathbf{1}_k^\top\mathbf{\Sigma}^+\mathbf{1}_k}\frac{1}{u}\right).
\end{aligned}$$

Using Theorem 4 (b) of [14] and following the proof of Theorem 3 of [16], we get

$$\bar{\mathbf{x}}^\top\mathbf{\Sigma}^+\mathbf{1}_k \sim \mathcal{N}\left(\boldsymbol{\mu}^\top\mathbf{\Sigma}^+\mathbf{1}_k, \frac{1}{n}\mathbf{1}_k^\top\mathbf{\Sigma}^+\mathbf{1}_k\right), \tag{12}$$

$$\frac{n(n-r+1)}{(n-1)(r-1)}\bar{\mathbf{x}}^\top\widehat{\mathbf{R}}^+\bar{\mathbf{x}} \sim \mathcal{F}_{r-1,n-r+1,ns} \tag{13}$$

and, moreover, $\bar{\mathbf{x}}^\top\mathbf{\Sigma}^+\mathbf{1}_k$ is independent of $\bar{\mathbf{x}}^\top\widehat{\mathbf{R}}^+\bar{\mathbf{x}}$. Thus, summarising the previous findings we get a stochastic representation of $\widehat{R}^+_{EU}$ as given in the statement of the theorem.

Next, we proceed with the derivation of a stochastic representation for $\widehat{V}^+_{EU}$. Let us recall that $\mathbf{1}_k^\top \mathbf{S}^+ \mathbf{1}_k$ is independent of $\bar{\mathbf{x}}^{*\top}\widehat{\mathbf{R}}^+\bar{\mathbf{x}}^*$. Therefore, since $\mathbf{1}_k^\top \mathbf{S}^+ \mathbf{1}_k$ does not depend on $\bar{\mathbf{x}}$, we also get that $\mathbf{1}_k^\top \mathbf{S}^+ \mathbf{1}_k$ and $\bar{\mathbf{x}}^\top\widehat{\mathbf{R}}^+\bar{\mathbf{x}}$ are independent. Using this fact together with (11), we arrive at a stochastic representation of $\widehat{V}^+_{EU}$ given in the statement of the theorem. ■

The results proved in Theorem 1 are fundamental and fully describe the stochastic behaviour of the estimated expected return and the estimated variance of the EU optimal portfolio when the covariance matrix is singular and the asset returns are independent and identically normally distributed. Moreover, these findings possess a number of important applications, which will be present in this section. In particular, they can be used to derive the moments of higher order of $\widehat{R}^+_{EU}$ and $\widehat{V}^+_{EU}$. We present the corresponding findings in Theorem 2.

In the formulation of the statement of Theorem 2, the confluent hypergeometric function is used which is defined by (see, Chapter 4 in [1].)

$$ {}_1F_1(a;b;x) = \frac{\Gamma(b)}{\Gamma(a)} \sum_{i=0}^{\infty} \frac{\Gamma(a+i)}{\Gamma(b+i)} \frac{x^i}{i!}. \tag{14} $$

Theorem 2. *Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be i.i.d. random vectors with $\mathbf{x}_1 \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $k \geq n$ and $rank(\boldsymbol{\Sigma}) = r < n$. Then the p^{th} order moments of $\widehat{R}^+_{EU}$ and $\widehat{V}^+_{EU}$ are given by*

$$ \begin{aligned} \mu^{(p)}_{R^+_{EU}} &:= \mathbb{E}\left[(\widehat{R}^+_{EU})^p\right] = e^{-\frac{ns^+}{2}} \sum_{j=0}^{\lfloor p/2 \rfloor} \left(\frac{V^+_{GMV}}{n}\right)^j \frac{(2j)!}{2^j j!} \binom{p}{2j} \\ &\times \sum_{j_1=0}^{p-2j} \sum_{j_2=0}^{j} (R^+_{GMV})^{p-2j-j_1} \left(\frac{n-1}{\alpha n}\right)^{j_1} \binom{p-2j}{j_1} \binom{j}{j_2} \\ &\times \frac{\Gamma\left(\frac{n-r+1}{2} - j_1 - j_2\right)}{\Gamma\left(\frac{n-r+1}{2}\right)} \frac{\Gamma\left(\frac{r-1}{2} + j_1 + j_2\right)}{\Gamma\left(\frac{r-1}{2}\right)} \, {}_1F_1\left(\frac{r-1}{2} + j_1 + j_2; \frac{r-1}{2}; \frac{ns^+}{2}\right) \end{aligned} $$

and

$$ \begin{aligned} \mu^{(p)}_{V^+_{EU}} &:= \mathbb{E}\left[(\widehat{V}^+_{EU})^p\right] = \sum_{i=0}^{p} (V^+_{GMV})^{p-i} \binom{p}{i} (n-1)^{2i-p} \left(\alpha^2 n\right)^{-i} \\ &\times (n-r)(n-r+2)(n-r+4)\dots(n-r+2(p-i)-2) \\ &\times e^{-\frac{ns^+}{2}} \frac{\Gamma\left(\frac{n-r+1}{2} - i\right)}{\Gamma\left(\frac{n-r+1}{2}\right)} \frac{\Gamma\left(\frac{r-1}{2} + i\right)}{\Gamma\left(\frac{r-1}{2}\right)} \, {}_1F_1\left(\frac{r-1}{2} + i; \frac{r-1}{2}; \frac{ns^+}{2}\right). \end{aligned} $$

Proof of Theorem 2: First, we derive the p^{th} order moment of $\widehat{R}^+_{EU}$. For that, we make use of the stochastic representation of $\widehat{R}^+_{EU}$ which is obtained in Theorem 1 and it is given by

$$\widehat{R}^+_{EU} \overset{d}{=} R^+_{GMV} + \alpha^{-1}\frac{(n-1)(r-1)}{n(n-r+1)}\xi + \sqrt{\frac{1}{n}\left(1+\frac{r-1}{n-r+1}\xi\right)}\sqrt{V^+_{GMV}}z_0,$$

where $\xi \sim \mathcal{F}_{r-1,n-r+1,ns^+}$ and $z_0 \sim \mathcal{N}(0,1)$. Moreover, ξ and z_0 are independent. The application of the binomial formula (see, [6, p. 129]) leads to

$$\begin{aligned}
&(\mu^+_{R_{EU}})^{(p)} := \mathbb{E}\left[(\widehat{R}^+_{EU})^p\right] \\
=\ & \mathbb{E}\left[\left(R^+_{GMV} + \alpha^{-1}\frac{(n-1)(r-1)}{n(n-r+1)}\xi + \sqrt{\frac{1}{n}\left(1+\frac{r-1}{n-r+1}\xi\right)}\sqrt{V^+_{GMV}}z_0\right)^p\right] \\
=\ & \mathbb{E}\left[\sum_{i=0}^{p}\binom{p}{i}\left(R^+_{GMV} + \alpha^{-1}\frac{(n-1)(r-1)}{n(n-r+1)}\xi\right)^{p-i}\left(\sqrt{\frac{1}{n}\left(1+\frac{r-1}{n-r+1}\xi\right)}\sqrt{V^+_{GMV}}z_0\right)^i\right] \\
=\ & \sum_{i=0}^{p}\left(\frac{V^+_{GMV}}{n}\right)^{\frac{i}{2}}\binom{p}{i}\mathbb{E}\left[z_0^i\right] \\
\times\ & \mathbb{E}\left[\left(R^+_{GMV} + \alpha^{-1}\frac{(n-1)(r-1)}{n(n-r+1)}\xi\right)^{p-i}\left(1+\frac{r-1}{n-r+1}\xi\right)^{i/2}\right].
\end{aligned}$$

Using the fact that the odd moments of z_0 are equal to zero and the even moments are given by (see, Chapter 34.2 in [39])

$$\mathbb{E}\left[z_0^{2l}\right] = \frac{(2l)!}{2^l l!} \quad \text{for} \quad l \geq 1,$$

we get that

$$\begin{aligned}
\mu^{(p)}_{R^+_{EU}} &= \sum_{j=0}^{\lfloor p/2 \rfloor}\left(\frac{V^+_{GMV}}{n}\right)^j\frac{(2j)!}{2^j j!}\binom{p}{2j} \\
&\times\ \mathbb{E}\left[\left(R^+_{GMV} + \alpha^{-1}\frac{(n-1)(r-1)}{n(n-r+1)}\xi\right)^{p-2j}\left(1+\frac{r-1}{n-r+1}\xi\right)^j\right].
\end{aligned} \tag{15}$$

Applying binomial formula again and the formula for higher order moments of noncentral $\mathcal{F}$-distribution (see, [39, Chapter 32.2]), we obtain that

$$\begin{aligned}
& \mathbb{E}\left[\left(R_{GMV}^{+}+\alpha^{-1}\frac{(n-1)(r-1)}{n(n-r+1)}\xi\right)^{p-2j}\left(1+\frac{r-1}{n-r+1}\xi\right)^{j}\right] \\
= & \ \mathbb{E}\left[\left[\sum_{j_1=0}^{p-2j}\left(\alpha^{-1}\frac{(n-1)(r-1)}{n(n-r+1)}\right)^{j_1}(R_{GMV}^{+})^{p-2j-j_1}\binom{p-2j}{j_1}\xi^{j_1}\right]\right. \\
\times & \left.\left[\sum_{j_2=0}^{j}\left(\frac{r-1}{n-r+1}\right)^{j_2}\binom{j}{j_2}\xi^{j_2}\right]\right] = \sum_{j_1=0}^{p-2j}\sum_{j_2=0}^{j}R_{GMV}^{p-2j-j_1} \\
\times & \left(\frac{n-1}{\alpha n}\right)^{j_1}\left(\frac{r-1}{n-r+1}\right)^{j_1+j_2}\binom{p-2j}{j_1}\binom{j}{j_2}\mathbb{E}\left[\xi^{j_1+j_2}\right] \\
= & \ e^{-\frac{ns^{+}}{2}}\sum_{j_1=0}^{p-2j}\sum_{j_2=0}^{j}(R_{GMV}^{+})^{p-2j-j_1}\left(\frac{n-1}{\alpha n}\right)^{j_1}\binom{p-2j}{j_1}\binom{j}{j_2}\frac{\Gamma\left(\frac{n-r+1}{2}-j_1-j_2\right)}{\Gamma\left(\frac{n-r+1}{2}\right)} \\
\times & \sum_{m=0}^{\infty}\frac{1}{m!}\left(\frac{ns^{+}}{2}\right)^{m}\frac{\Gamma\left(\frac{r-1}{2}+m+j_1+j_2\right)}{\Gamma\left(\frac{r-1}{2}+m\right)}.
\end{aligned}$$

Substituting the last equality in (15) and applying the formula for the confluent hypergeometric function (14), we get the expression of the higher order moments of $\widehat{R}_{EU}^{+}$ given in the statement of the theorem.

Next, we derive the p^{th} order moment of $\widehat{V}_{EU}^{+}$. From Theorem 1, a stochastic representation of $\widehat{V}_{EU}^{+}$ is expressed as

$$\widehat{V}_{EU}^{+} \overset{d}{=} \frac{V_{GMV}^{+}}{n-1}\eta+\alpha^{-2}\frac{(n-1)(r-1)}{n(n-r+1)}\xi,$$

where $\xi \sim \mathcal{F}_{r-1,n-r+1,ns^{+}}$ and $\eta \sim \chi^2_{n-r}$. Moreover, ξ and η are independently distributed. Applying binomial formula (see, [6, p. 129]) and the fact that ξ and η are independent, we obtain that

$$\begin{aligned}
\mu_{V_{EU}^{+}}^{(p)} & := \mathbb{E}\left[(\widehat{V}_{EU}^{+})^{p}\right]=\mathbb{E}\left[\left(\frac{V_{GMV}^{+}}{n-1}\eta+\alpha^{-2}\frac{(n-1)(r-1)}{n(n-r+1)}\xi\right)^{p}\right] \\
& = \mathbb{E}\left[\sum_{i=0}^{p}\binom{p}{i}\left(\frac{V_{GMV}^{+}}{n-1}\eta\right)^{p-i}\left(\alpha^{-2}\frac{(n-1)(r-1)}{n(n-r+1)}\xi\right)^{i}\right] \\
& = \sum_{i=0}^{p}\binom{p}{i}\left(\frac{V_{GMV}^{+}}{n-1}\right)^{p-i}\left(\alpha^{-2}\frac{(n-1)(r-1)}{n(n-r+1)}\right)^{i}\mathbb{E}\left[\eta^{p-i}\right]\mathbb{E}\left[\xi^{i}\right] \\
& = \sum_{i=0}^{p}\binom{p}{i}(V_{GMV}^{+})^{p-i}(n-1)^{2i-p}\left(\alpha^{-2}\frac{r-1}{n(n-r+1)}\right)^{i}\mathbb{E}\left[\eta^{p-i}\right]\mathbb{E}\left[\xi^{i}\right].
\end{aligned}$$

From Chapters 8.2 and 32.2 in [39], we get

$$
\begin{aligned}
\mathbb{E}\left[\eta^{p-i}\right] &= (n-r)(n-r+2)(n-r+4)\dots(n-r+2(p-i)-2),\\
\mathbb{E}\left[\xi^{i}\right] &= e^{-\frac{ns^+}{2}}\left(\frac{n-r+1}{r-1}\right)^{i}\frac{\Gamma\left(\frac{n-r+1}{2}-i\right)}{\Gamma\left(\frac{n-r+1}{2}\right)}\sum_{j=0}^{\infty}\frac{1}{j!}\left(\frac{ns^+}{2}\right)^{j}\frac{\Gamma\left(\frac{r-1}{2}+j+i\right)}{\Gamma\left(\frac{r-1}{2}+j\right)}.
\end{aligned}
$$

Summarising the above results we arrive at the expression for the p^{th} order moment of $\widehat{V}^+_{EU}$. ■

As special cases of Theorem 2, we obtain the expected values and the variances of $\widehat{R}^+_{EU}$ and $\widehat{V}^+_{EU}$ in Corollary 1.

Corollary 1. *Let* $\mathbf{x}_1,\dots,\mathbf{x}_n$ *be i.i.d. random vectors with* $\mathbf{x}_1\sim\mathcal{N}_k(\boldsymbol{\mu},\boldsymbol{\Sigma})$, $k\geq n$ *and* $rank(\boldsymbol{\Sigma})=r<n$. *Then the mean and variance of* $\widehat{R}_{EU}$ *and* $\widehat{V}_{EU}$ *are given by*

$$
\begin{aligned}
\mathbb{E}[\widehat{R}^+_{EU}] &= R^+_{GMV}+\alpha^{-1}\frac{(n-1)(r-1+ns^+)}{n(n-r-1)},\\
\mathbb{E}[\widehat{V}^+_{EU}] &= \frac{n-r}{n-1}V^+_{GMV}+\alpha^{-2}\frac{(n-1)(r-1+ns^+)}{n(n-r-1)},\\
\mathbb{V}ar[\widehat{R}^+_{EU}] &= \frac{n(s+1)-2}{n(n-r-1)}V^+_{GMV}\\
&+ 2\alpha^{-2}\frac{(n-1)^2}{n^2}\frac{(r-1+ns^+)^2+(r-1+2ns^+)(n-r-1)}{(n-r-1)^2(n-r-3)},\\
\mathbb{V}ar[\widehat{V}^+_{EU}] &= \frac{2(n-r)}{(n-1)^2}(V^+_{GMV})^2\\
&+ 2\alpha^{-4}\frac{(n-1)^2}{n^2}\frac{(r-1+ns^+)^2+(r-1+2ns^+)(n-r-1)}{(n-r-1)^2(n-r-3)}.
\end{aligned}
$$

Let $f_{\chi^2_m}(\cdot)$ denote for the density of the χ^2-distribution with m degrees of freedom and let $f_{\mathcal{F}_{m_1,m_2,c^2}}(\cdot)$ be the density of the noncentral F-distribution with m_1 and m_2 degrees of freedom and noncentrality parameter c^2. Another application of the results derived in Theorem 1 leads to the formulas of the exact densities of $\widehat{R}^+_{EU}$ and $\widehat{V}^+_{EU}$ presented in Theorem 3.

Theorem 3. *Let* $\mathbf{x}_1,\dots,\mathbf{x}_n$ *be i.i.d. random vectors with* $\mathbf{x}_1\sim\mathcal{N}_k(\boldsymbol{\mu},\boldsymbol{\Sigma})$, $k\geq n$ *and* $rank(\boldsymbol{\Sigma})=r<n$. *Then,*

i) the density function of $\widehat{R}^+_{EU}$ *is given by*

$$
\begin{aligned}
f_{\widehat{R}^+_{EU}}(x) &= e^{-\frac{ns}{2}}\sqrt{\frac{n}{2\pi V^+_{GMV}}}\frac{1}{B\left(\frac{n-r+1}{2},\frac{r-1}{2}\right)}\\
&\times \int_0^1 (1-u)^{\frac{n-r}{2}}u^{\frac{r-3}{2}}\ {}_1F_1\left(\frac{n}{2};\frac{r-1}{2};\frac{ns^+}{2}u\right)\\
&\times \exp\left\{-\frac{n(1-u)}{2V^+_{GMV}}\left(x-R^+_{GMV}-\alpha^{-1}\frac{(n-1)u}{n(1-u)}\right)^2\right\}du;
\end{aligned}
$$

ii) the density function of $\widehat{V}_{EU}^{+}$ *is given by*

$$
\begin{aligned}
f_{\widehat{V}_{EU}^{+}}(x) &= \frac{n(n-r+1)}{(r-1)V_{GMV}^{+}}\int_0^\infty f_{\chi^2_{n-r}}\left(\frac{n-1}{V_{GMV}^{+}}(x-\alpha^{-2}u)\right) \\
&\times\ f_{\mathcal{F}_{r-1,n-r+1,ns^{+}}}\left(\frac{n(n-r+1)}{(n-1)(r-1)}u\right)du.
\end{aligned}
$$

2.2 Results under the high-dimensional setting

The exact results derived in Theorem 1 are also very useful to study the behaviour of $\widehat{R}_{EU}^{+}$ and $\widehat{V}_{EU}^{+}$ in the large-dimensional setting. Since the population covariance matrix $\mathbf{\Sigma}$ is singular, we consider the stochastic behaviour of $\widehat{R}_{EU}^{+}$ and $\widehat{V}_{EU}^{+}$ for $r/n \to c \in (0,1)$ as $n \to \infty$. This appears to be a better asymptotic regime than the one based on ratio of k over n, since the singular population covariance matrix impacts the actual dimension of the data-generating process. Namely, the density of $\mathbf{x}_t$ is singular and it is nonsingular on a subspace of $\mathbb{R}^k$ of dimension r where $r = rank(\mathbf{\Sigma})$. As such, the actual dimension of the data-generating model for $\mathbf{x}_t$ is r, which motivates the asymptotic regime r/n. To this end, we note that the dimension of the data-generating k could be really very large and it can even be larger than the sample size n in our study.

In order to specify the asymptotic behaviour of $\widehat{R}_{EU}^{+}$ and $\widehat{V}_{EU}^{+}$. We also need to impose a condition on the model parameters $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$. Namely, in the following it is assumed that

(A1) There exist constants m and M such that $0 < m \le \mathbf{1}_k^\top \mathbf{\Sigma}^{+}\mathbf{1}_k \le M < \infty$ and $0 < m \le \boldsymbol{\mu}^\top\mathbf{\Sigma}^{+}\boldsymbol{\mu} \le M < \infty$ uniformly on r.

In Theorem 4 we provide the asymptotic behaviour of the estimated expected return and the estimated variance of the EU optimal portfolio.

Theorem 4. *Let* $\mathbf{x}_1, \ldots, \mathbf{x}_n$ *be i.i.d. random vectors with* $\mathbf{x}_1 \sim \mathcal{N}_k(\boldsymbol{\mu}, \mathbf{\Sigma})$*,* $k \ge n$ *and* $rank(\mathbf{\Sigma}) = r < n$*. Assume (A1). Then, it holds that*

$$
\sqrt{n}\left(\widehat{R}_{EU}^{+} - \left(R_{GMV}^{+} + \alpha^{-1}\frac{c+s^{+}}{1-c}\right)\right) \to \mathcal{N}(0, \sigma^2_{\widehat{R}_{EU}^{+}})
$$

and

$$
\sqrt{n}\left(\widehat{V}_{EU}^{+} - \left((1-c)V_{GMV}^{+} + \alpha^{-2}\frac{c+s^{+}}{1-c}\right)\right) \to \mathcal{N}(0, \sigma^2_{\widehat{V}_{EU}^{+}})
$$

for $c_n := r/n \to c \in (0,1)$ *as* $n \to \infty$*, where*

$$
\sigma^2_{\widehat{R}_{EU}^{+}} = \frac{1+s^{+}}{1-c}V_{GMV}^{+} + \frac{2\alpha^{-2}}{(1-c)^3}\left[c + s^{+}(2+s^{+})\right]
$$

and

$$\sigma^2_{\hat{V}^+_{EU}} = 2(1-c)(V^+_{GMV})^2 + \frac{2\alpha^{-2}}{(1-c)^3}\left[c + s^+(2+s^+)\right].$$

Proof of Theorem 4: From Lemma 3 in [22], we obtain that

$$\sqrt{n}\left(\begin{pmatrix} \xi \\ \eta/(n-r) \\ z_0/\sqrt{n} \end{pmatrix} - \begin{pmatrix} 1+s^+/c \\ 1 \\ 0 \end{pmatrix}\right) \to \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \sigma^2_\xi & 0 & 0 \\ 0 & 2/(1-c) & 0 \\ 0 & 0 & 1 \end{pmatrix}\right)$$

for $r/n \to c \in (0,1)$ as $n \to \infty$ with

$$\sigma^2_\xi = \frac{2}{c}\left(1 + 2\frac{s^+}{c}\right) + \frac{2}{1-c}\left(1 + \frac{s^+}{c}\right)^2.$$

Therefore, it holds that

$$\sqrt{n}\left(\begin{pmatrix} (r-1)\xi/(n-r+1) \\ \eta/(n-1) \\ z_0/\sqrt{n} \end{pmatrix} - \begin{pmatrix} (c+s^+)/(1-c) \\ 1-c \\ 0 \end{pmatrix}\right)$$
$$\to \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} c^2\sigma^2_\xi/(1-c)^2 & 0 & 0 \\ 0 & 2(1-c) & 0 \\ 0 & 0 & 1 \end{pmatrix}\right)$$

for $r/n \to c \in (0,1)$ as $n \to \infty$. The application of the delta method (see, [23, Theorem 3.7]) leads to the statement of the theorem. ■

3 Finite-sample performance

In this section, we study the properties of the high-dimensional asymptotic approximations of the standardised sample estimators of the expected return and the variance of the EU optimal portfolio via simulations.

We consider the sample size $n = \{50, 120, 250, 500\}$ corresponding to one year, two years, five years and ten years of weekly financial returns and the dimension of assets grows with the sample size, $k = 1.5n$. The concentration ratio is $c = \{0.5, 0.8\}$ so that the rank of the singular covariance matrix $\mathbf{\Sigma}$ is $r = cn$, and the risk aversion is $\alpha = 100$. We sample the vector of expected returns $\boldsymbol{\mu}$ and the population singular covariance matrix $\mathbf{\Sigma}$ following [16]. Each element in the vector $\boldsymbol{\mu}$ is drawn from the uniform distribution on $[-1, 1]$. The population singular covariance matrix $\mathbf{\Sigma}$ is obtained by generating eigenvalues and eigenvectors of $\mathbf{\Sigma}$. Firstly, the r non-zero eigenvalues of $\mathbf{\Sigma}$ are drawn from the uniform distribution in the unit domain and the remain $k-r$ eigenvalues are set to zero. Secondly, the eigenvectors of $\mathbf{\Sigma}$ is generated from the Haar distribution by calculating eigenvectors of a random matrix following a Wishart distribution with identity scale matrix and k degrees of freedom.

The sample mean and sample variance of the EU portfolio are calculated using $N = 100000$ repetitions using the same generated values of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ as follows:

- **Step 1**: Calculate the expected return R^+_{EU} and variance V^+_{EU} of the EU portfolio using (8).
- **Step 2**: Generate samples of $\widehat{R}^+_{EU}$ and variance $\widehat{V}^+_{EU}$ from their exact distributions by using the stochastic representations in Theorem 1,

$$\widehat{R}^+_{EU} \overset{d}{=} R^+_{GMV} + \alpha^{-1}\frac{(n-1)(r-1)}{n(n-r+1)}\xi + \sqrt{\frac{1}{n}\left(1+\frac{r-1}{n-r+1}\xi\right)}\sqrt{V^+_{GMV}}z_0$$

  and

$$\widehat{V}^+_{EU} \overset{d}{=} \frac{V^+_{GMV}}{n-1}\eta + \alpha^{-2}\frac{(n-1)(r-1)}{n(n-r+1)}\xi,$$

  where $\xi \sim \mathcal{F}_{r-1,n-r+1,ns^+}$, $\eta \sim \chi^2_{n-r}$, and $z_0 \sim \mathcal{N}(0,1)$.

- **Step 3**: Evaluate the standardized variables in Theorem 4 by

$$\begin{aligned}\bar{R}^+_{EU} &= \sqrt{\frac{n}{\sigma^2_{\widehat{R}^+_{EU}}}}(\widehat{R}^+_{EU} - R^+_{EU}),\\ \bar{V}^+_{EU} &= \sqrt{\frac{n}{\sigma^2_{\widehat{V}^+_{EU}}}}(\widehat{V}^+_{EU} - V^+_{EU}).\end{aligned}$$

Figures 1 and 2 show the asymptotic distribution (solid line) and the kernel density estimator (dashed line) of the finite-sample distribution of standardised $\bar{R}^+_{EU}$ and $\bar{V}^+_{EU}$ for $c = \{0.5, 0.8\}$ respectively. There is a large difference between the finite sample and the asymptotic distributions when the number of sample is small, however the finite sample density coincides with the asymptotic density when the sample size becomes larger, $n \geq 250$. These findings are inline with Table 1 where the sample mean and sample variance of standardised $\bar{R}^+_{EU}$ and $\bar{V}^+_{EU}$ are shown with different values of sample sizes and concentration ratios. Both sample quantities reach the asymptotic standardised normal distribution as sample size increases.

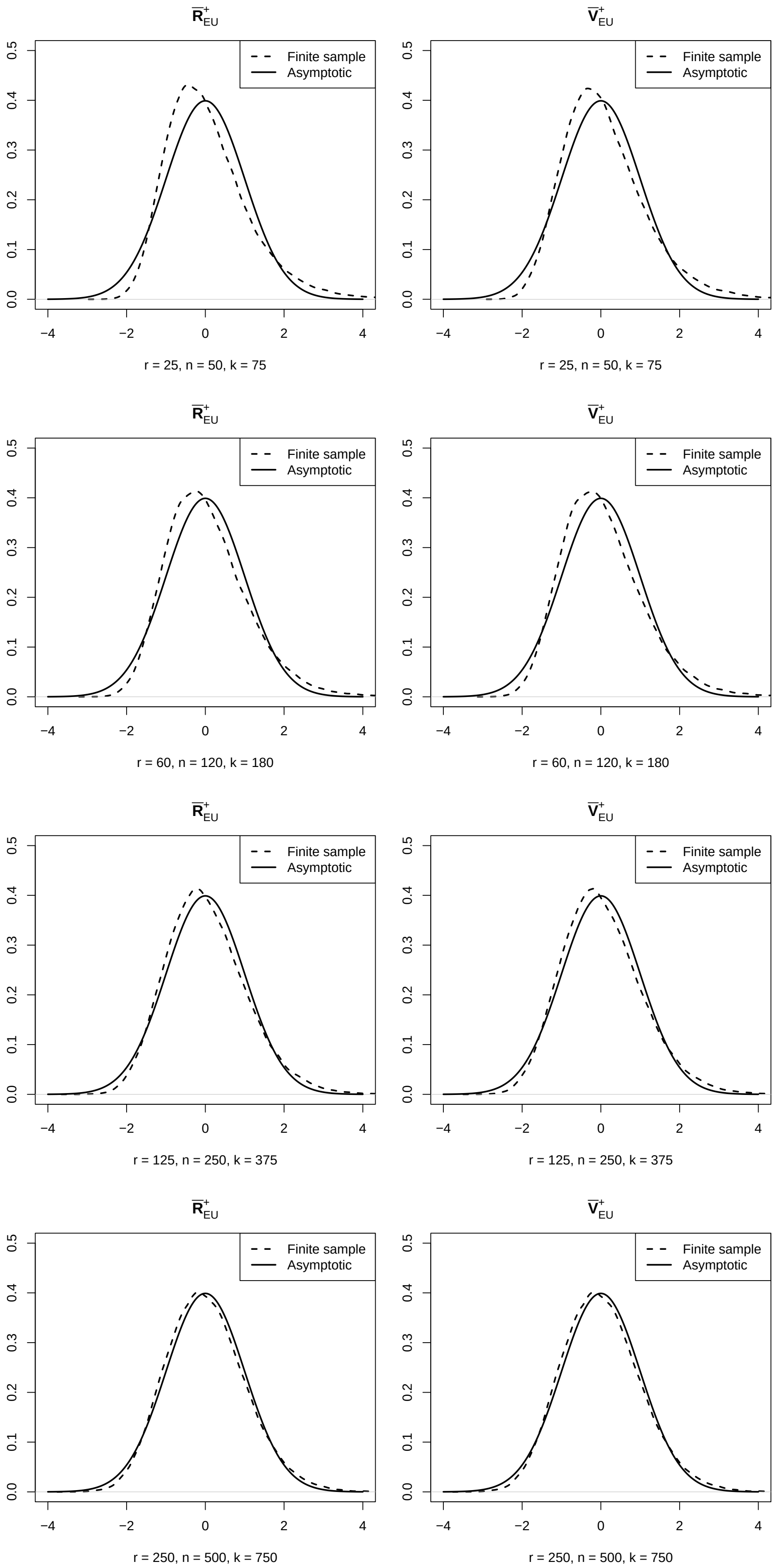

Figure 1: Asymptotic distribution and the kernel density estimator of the finite-sample distribution of standardised $\bar{R}^{+}_{EU}$ and $\bar{V}^{+}_{EU}$ for $c = 0.5$.

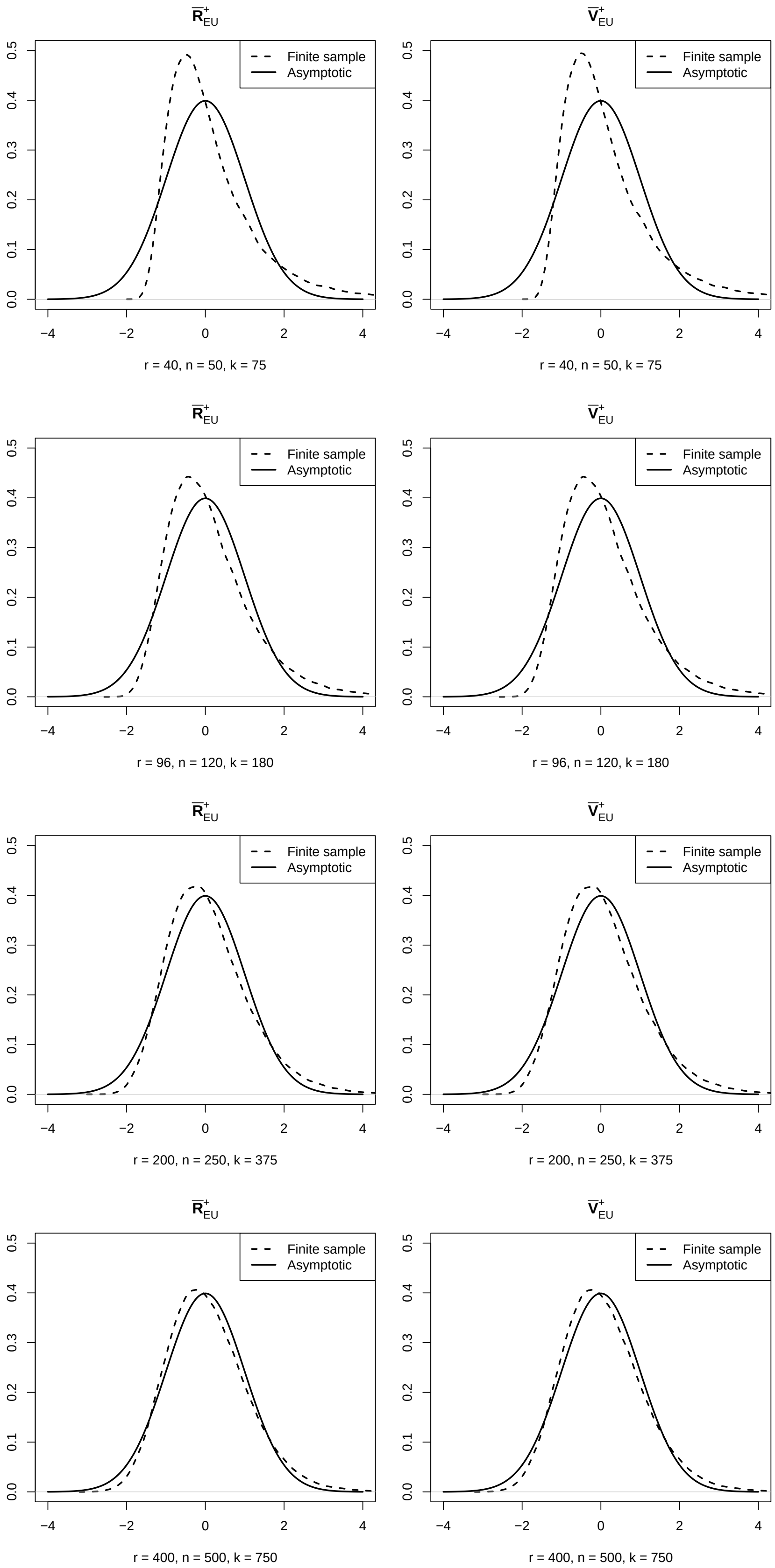

Figure 2: Asymptotic distribution and the kernel density estimator of the finite-sample distribution of standardised $\bar{R}^{+}_{EU}$ and $\bar{V}^{+}_{EU}$ for $c = 0.8$.

Table 1: Sample mean and sample variance of standardised $\bar{R}^{+}_{EU}$ and $\bar{V}^{+}_{EU}$

		$n=50$	$n=120$	$n=250$	$n=500$
Panel (a) - $\bar{R}^{+}_{EU}$					
$c=0.50$	Sample mean	0.053429	0.045069	0.036535	0.022397
	Sample variance	1.122271	1.064992	1.035233	1.020149
$c=0.80$	Sample mean	0.190551	0.125916	0.080326	0.054470
	Sample variance	1.701651	1.226259	1.093531	1.049863
Panel (b) - $\bar{V}^{+}_{EU}$					
$c=0.50$	Sample mean	0.098852	0.046847	0.037295	0.022291
	Sample variance	1.095661	1.067031	1.034508	1.020210
$c=0.80$	Sample mean	0.195541	0.126426	0.080406	0.054534
	Sample variance	1.714786	1.226477	1.093548	1.049995

4 Summary

Optimal portfolio selection plays an important role in both theory and practice of financial market research. While the most of the results in the theory are derived under the assumption that the population covariance matrix is nonsingular, we extend the existent findings to the case of singular covariance matrix. We deal with the problem of estimating the expected return and the variance of the EU optimal portfolio and with the characterisation of their sampling distribution. Very useful stochastic representations of the estimated expected return and for the estimated variance of the EU portfolio are deduced, which are later used in the derivation of their higher-order moments, of their density function, and of their asymptotic distributions in the high-dimensional setting. Via simulations, it is shown that the high-dimensional asymptotic distributions provide good approximations already for samples of moderate sizes.

References

[1] Andrews, G.E., Askey, R., Roy, R.: Special Functions. Cambridge University Press, Cambridge (2000)

[2] Barry, C.B.: Portfolio analysis under uncertain means, variances, and covariances. Journal of Finance **29**, 515–522 (1974)

[3] Bauder, D., Bodnar, R., Bodnar, T., Schmid, W.: Bayesian estimation of the efficient frontier. Scandinavian Journal of Statistics **46**, 802–830 (2019)

[4] Bauder, D., Bodnar, T., Mazur, S., Okhrin, Y.: Bayesian inference for the tangent portfolio. International Journal of Theoretical and Applied Finance **21**, 1850054 (2018)

[5] Bauder, D., Bodnar, T., Parolya, N., Schmid, W.: Bayesian mean–variance analysis: optimal portfolio selection under parameter uncertainty. Quantitative Finance **21**, 221–242 (2021)

[6] Biggs, N.L.: The roots of combinatorics. Historia Mathematica **6**, 109–136 (1979)

[7] Black, F., Litterman, R.: Global portfolio optimization. Financial Analysts Journal **48**, 28–43 (1992)

[8] Bodnar, O., Bodnar, T., Parolya, N.: Recent advances in shrinkage-based high-dimensional inference. Journal of Multivariate Analysis **188**, 104826 (2022)

[9] Bodnar, T., Dmytriv, S., Okhrin, Y., Otryakhin, D., Parolya, N.: HDShOP: High-dimensional shrinkage optimal portfolios. R package version 0.1.2 (2021)

[10] Bodnar, T., Dmytriv, S., Okhrin, Y., Parolya, N., Schmid, W.: Statistical inference for the expected utility portfolio in high dimensions. IEEE Transactions on Signal Processing **69**, 1–14 (2021)

[11] Bodnar, T., Dette, H., Parolya, N., Thorsén, E.: Sampling distributions of optimal portfolio weights and characteristics in small and large dimensions. Random Matrices: Theory and Applications **11**, 2250008 (2022)

[12] Bodnar, T., Mazur, S., Okhrin, Y.: Distribution of the product of a singular Wishart matrix and a normal vector. Theory of Probability and Mathematical Statistics **91**, 1–15 (2015)

[13] Bodnar, T., Mazur, S., Okhrin, Y.: Bayesian estimation of the global minimum variance portfolio. European Journal of Operational Research **256**, 292–307 (2017)

[14] Bodnar, T., Mazur, S., Podgórski, K.: Singular inverse Wishart distribution and its application to portfolio theory. Journal of Multivariate Analysis **143**, 314–326 (2016)

[15] Bodnar, T., Mazur, S., Podgórski, K.: A test for the global minimum variance portfolio for small sample and singular covariance. AStA Advances in Statistical Analysis **101**, 253–265 (2017)

[16] Bodnar, T., Mazur, S., Podgórski, K., Tyrcha, J.: Tangency portfolio weights for singular covariance matrix in small and large dimensions: Estimation and test theory. Journal of Statistical Planning and Inference **201**, 40–57 (2019)

[17] Bodnar, T., Okhrin, Y., Parolya, N.: Optimal shrinkage-based portfolio selection in high dimensions. Journal of Business & Economic Statistics, to appear (2022)

[18] Bodnar, T., Parolya, N., Schmid, W.: On the equivalence of quadratic optimization problems commonly used in portfolio theory. European Journal of Operational Research **229**, 637–644 (2013)

[19] Bodnar, T., Parolya, N., Schmid, W.: Estimation of the global minimum variance portfolio in high dimensions. European Journal of Operational Research **266**, 371–390 (2018)

[20] Bodnar, T., Schmid, W.: Estimation of optimal portfolio compositions for Gaussian returns. Statistics & Decisions **26**, 179–201 (2008)

[21] Bodnar, T., Schmid, W.: On the exact distribution of the estimated expected utility portfolio weights: Theory and applications. Statistics & Risk Modeling **28**, 319–342 (2011)

[22] Bodnar, T., Reiss, M.: Exact and asymptotic tests on a factor model in low and large dimensions with applications. Journal of Multivariate Analysis **150**, 125–151 (2016)

[23] DasGupta, A.: Asymptotic Theory of Statistics and Probability. Springer, New York (2008)

[24] Frahm, G., Memmel, C.: Dominating estimators for minimum-variance portfolios. Journal of Econometrics **159**, 289–302 (2010)

[25] Golosnoy, V., Okhrin, Y.: Multivariate shrinkage for optimal portfolio weights. The European Journal of Finance **13**, 441–458 (2007)

[26] Gulliksson, M., Mazur, S.: An iterative approach to ill-conditioned optimal portfolio selection. Computational Economics **56**, 773–794 (2020)

[27] Gulliksson, M., Oleynik, A., Mazur, S.: Portfolio selection with a rank-deficient covariance matrix. Working paper **12**, Örebro University (2021)

[28] Harville, D.A.: Matrix Algebra from Statistician's Perspective. Springer, New York (1997)

[29] Ingersoll J.E.: Theory of Financial Decision Making. Rowman & Littlefield Publishers, Lanham (1987)

[30] Jorion, P.: Bayes-Stein estimation for portfolio analysis. Journal of Financial and Quantitative analysis **21**, 279–292 (1986)

[31] Klein, R., Bawa, V.: The effect of estimation risk on optimal portfolio choice. Journal of Financial Economics **3**, 215–231 (1976)

[32] Markowitz, H: Portfolio selection. The Journal of Finance **7**, 77–91 (1952)

[33] Muirhead, R.J.: Aspects of Multivariate Statistical Theory. Wiley, New York (1982)

[34] Okhrin, Y., Schmid, W.: Distributional properties of portfolio weights. Journal of Econometrics **134**, 235–256 (2006)

[35] Pappas, D., Kiriakopoulos, K., Kaimakamis, G.: Optimal portfolio selection with singular covariance matrix. International Mathematical Forum **5**, 2305–2318, (2010)

[36] Pástor, L.: Portfolio selection and asset pricing models. Journal of Finance **55**, 179–223 (2000)

[37] Pástor, L., Stambaugh, R.F.: Comparing asset pricing models: An investment perspective. Journal of Financial Economics **56**, 335–381 (2000)

[38] Tu, J., Zhou, G.: Incorporating economic objectives into Bayesian priors: Portfolio choice under parameter uncertainty. Journal of Financial and Quantitative Analysis **45**, 959–986 (2010)

[39] Walck, C.: Hand-book on statistical distributions for experimentalists. Internal Report SUF-PFY/96-01 of Stockholm University, Stockholm (1996)

[40] Winkler, R. L.: Bayesian models for forecasting future security prices. Journal of Financial and Quantitative Analysis **8**, 387–405 (1973)