Optimal Shrinkage Estimator for High-Dimensional Mean Vector

Taras Bodnar

Ostap Okhrin

Nestor Parolya

Postal address:
Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:
http://www.math.su.se

Optimal Shrinkage Estimator for High-Dimensional Mean Vector

November 2016

Taras Bodnar[a,1], Ostap Okhrin[b] and Nestor Parolya[c]

[a] *Department of Mathematics, Stockholm University, SE-10691 Stockholm, Sweden*
[b] *Chair of Econometrics and Statistics esp. Transportation, Dresden University of Technology, D-01062 Dresden, Germany*
[c] *Institute of Statistics, Leibniz University of Hannover, D-30167 Hannover, Germany*

Abstract

In this paper we derive the optimal linear shrinkage estimator for the large-dimensional mean vector using random matrix theory. The results are obtained under the assumption that both the dimension p and the sample size n tend to infinity such that $n^{-1}p^{1-\gamma} \to c \in (0, +\infty)$ and $\gamma \in [0, 1)$. Under weak conditions imposed on the the underlying data generating process, we find the asymptotic equivalents to the optimal shrinkage intensities, prove their asymptotic normality, and estimate them consistently. The obtained non-parametric estimator for the high-dimensional mean vector has a simple structure and is proven to minimize asymptotically with probability 1 the quadratic loss in the case of $c \in (0, 1)$. For $c \in (1, +\infty)$ we modify the suggested estimator by using a feasible estimator for the precision covariance matrix. At the end, an exhaustive simulation study and an application to real data are provided where the proposed estimator is compared with known benchmarks from the literature.

1 Introduction

High-dimensional problems found a front place in the modern statistics with the development of high performance and high storage computers. The latter forces the collection of huge amounts

---

[1]* Corresponding author. E-mail address: taras.bodnar@math.su.se

of information which should be understood and used in the prediction of the specific needed features. This leads to the development of new mathematical models, since classical ones from the multivariate statistics do not necessarily lead to proper results because of specific features of big data, like noise accumulation, spurious correlation, heterogeneity and others (see, e.g., Bai and Silverstein (2010)). For a review of challenges in modeling big data we refer readers to Fan et al. (2014).

Huge amount of information has been used in various fields of human activity, like genomics with hundreds of thousands of microarrays (Shalon et al. (1996)), neurosciences with high-precision (large dimension) fMRI data of dozens of persons (Issa et al. (2013)), financial time series (Tsay (2005)), etc. In most of the models developed for the problems listed above an important place is settled to the estimation of the high-dimensional mean vector when the sample size is much smaller than the dimension. The usual estimator of the location under the quadratic loss function, the sample mean vector, is known for decades to be non-admissible even for $p \geq 3$ and $p \leq n$. First solutions to the problem of the improved estimation of the mean vector were proposed by James and Stein (1961) for the multivariate normal random variables with identity covariance matrix in the case of $p > 2$. Later Baranchik (1970) extended this estimator to the case of covariance matrix being diagonal with all variances being equal. Further research by Lin and Tsai (1973), Berger and Bock (1976), Berger et al. (1977), Gleser (1986), Fourdrinier et al. (2003) lead to the estimators developed under a general unknown covariance matrix for $n \geq p \geq 3$.

A high-dimensional version of the James-Stein type estimator was proposed by Chételat and Wells (2012) for $p > n \geq 3$ using an unbiased estimator of the risk difference. Chételat and Wells (2012) suggested the so-called positive-parttype James-Stein estimator which was shown, via a simulation study, to dominate the high-dimensional James-Stein estimator under invariant loss. Wang et al. (2014) considered an optimal shrinkage estimator towards the unity vector by minimizing the expected quadratic loss function. However, the resulting estimator appears to be very computationally demanding for large dimensions because nontrivial sums are present in the resulting expression. In practice, the authors suggested the application of the limiting expression of this estimator.

Our paper contributes to the literature on the estimation of the location by deriving the optimal shrinkage estimator (towards any fixed target) by means of random matrix theory. The new estimator relies on weaker assumptions than the ones which are used in the literature. We prove the asymptotic normality of the suggested estimator and derive its limit behavior. The results of the simulation study show the dominance of the given estimator over the benchmark methods in terms of both the values of the quadratic loss function and the values of the computational time.

This paper is structured as follows. Section 2 introduces the optimal shrinkage estimator for the mean vector and provides its asymptotic equivalence to a nonrandom quantity. We further prove its asymptotic normality. In Section 3 we provide a bona fide estimator and investigate its asymptotic behavior. Section 4 discusses benchmark procedures used in the simulation study provided in Section 5. Section 6 provides an financial application, whereas Section 7 concludes. The technical derivations are moved to the appendix.

2 Optimal shrinkage estimator for the mean vector

In this section we construct an optimal shrinkage estimator for the mean vector under high-dimensional asymptotics.

Let the $p \times n$ matrix $\mathbf{Y}_n$ be the observation matrix of p-dimensional vectors of random variables taken at time points $1, \ldots, n$. The mean vector of each column of $\mathbf{Y}_n$ is denoted by $\boldsymbol{\mu}_n$, while $\boldsymbol{\Sigma}_n$ stands for its covariance matrix. Under the large-dimensional asymptotics both the dimension p and the sample size n tend to infinity. For this reason, it is natural to assume, without loss of generality, that the dimension $p \equiv p(n)$ is the function of the sample size n. Later on, we assume that the observation matrix in distribution is equal to

$$\mathbf{Y}_n \stackrel{d}{=} \boldsymbol{\Sigma}_n^{\frac{1}{2}} \mathbf{X}_n + \boldsymbol{\mu}_n \mathbf{1}_n^\top, \tag{2.1}$$

where the symbol '$\stackrel{d}{=}$' denotes the equality in distribution and the $p \times n$ matrix $\mathbf{X}_n$ contains independent and identically distributed (i.i.d.) real random variables with zero mean and unit variance, while $\mathbf{1}_n$ is the n-dimensional vector of ones. Only the matrix $\mathbf{Y}_n$ is observable. Neither $\mathbf{X}_n$ nor $\boldsymbol{\Sigma}_n$ with $\boldsymbol{\mu}_n$ are known.

It must be noted that the observation matrix $\mathbf{Y}_n$ has dependent rows but independent columns. The assumption of the independence imposed on its columns can further be weakened to dependent elements of $\mathbf{X}_n$ by controlling the growth of the number of dependent entries while their joint distribution can be arbitrary (see, Friesen et al. (2013)). For that reason the assumption of independence is present only for technical reasons in order to make the proofs of the main theorems better readable.

Next, we present the main assumptions which are used in the paper

(A1) There exists $\lambda_0 > 0$ such that $\lambda_0 \leq \lambda_{min}(\boldsymbol{\Sigma}_n)$ uniformly on p, where $\lambda_{min}(\mathbf{A})$ denotes the smallest eigenvalue of the square matrix $\mathbf{A}$. Similarly, the largest eigenvalue of $\mathbf{A}$ is denoted by $\lambda_{max}(\mathbf{A})$.

(A2) There exists $\gamma \in [0, 1)$, $M_l > 0$, $M_u > 0$ such that $\lim_{p \to \infty} p^{-\gamma} ||\boldsymbol{\mu}_n||^2 = M_1$ and $\lim_{p \to \infty} p^{-\gamma} ||\boldsymbol{\mu}_0||^2 = M_0$ with $0 < M_l \leq M_0, M_1 \leq M_u < \infty$.

(A3) It holds that $n^{-1} p^{1-\gamma} \to c \in (0, +\infty)$ with $\gamma \in [0, 1)$ as $n \longrightarrow \infty$.

(A4) The elements of the matrix $\mathbf{X}_n$ have uniformly bounded $2 + \varepsilon$ moments with $\varepsilon > 0$.

All of these regularity assumptions are very general and fit many practical situations. Assumption (A1) controls the behavior of the smallest eigenvalue of the population covariance matrix. It is remarkable that no condition is imposed on the largest eigenvalue of $\boldsymbol{\Sigma}_n$ which could also increase to infinity as p becomes larger. This, in particular, allows the application of the obtained results to high-dimensional factor models, which are very popular in economics and finance (see, e.g., Chamberlain and Rothschild (1983), Bai and Ng (2002, 2008), Fan et al. (2008), Fan et al. (2012), Fan et al. (2013), Bodnar and Reiss (2016)). The assumption on the lower bound of the smallest eigenvalue of $\boldsymbol{\Sigma}_n$ can be avoided via structural assumptions on the true covariance matrix $\boldsymbol{\Sigma}_n$ like tapering, c.f. Cai et al. (2010). In this paper we assume all eigenvalues being strictly positive. The increase in the norms of the unknown mean vector and of the target vector are monitored by Assumption (A2) which only requires that they are of the same order. Finally, Assumptions (A2) is a technical one and can be relaxed in some cases

(see, Rubio et al. (2012)). If $\gamma > 0$ in Assumption (A3), then we get the case $p/n \to \infty$, while $\gamma = 0$ leads to $p/n = O(1)$. Hence, γ controls the growth rate of dimension p.

The general linear shrinkage estimator of the mean vector $\boldsymbol{\mu}_n$ is given by

$$\hat{\boldsymbol{\mu}}_{GSE} = \alpha_n \bar{\mathbf{y}}_n + \beta_n \boldsymbol{\mu}_0, \tag{2.2}$$

where the target vector $\boldsymbol{\mu}_0$ satisfies (A2). It can also be random but independent of the actual information set $\mathbf{Y}_n$. The symbol $\bar{\mathbf{y}}_n = n^{-1} \mathbf{Y}_n \mathbf{1}_n$ stands for the sample mean vector.

The aim is to find the optimal shrinkage intensities which minimize the quadratic loss for a given target vector $\boldsymbol{\mu}_0$ expressed as

$$L = ||\boldsymbol{\Sigma}_n^{-1/2}(\hat{\boldsymbol{\mu}}_{GSE} - \boldsymbol{\mu}_n)||^2 = (\hat{\boldsymbol{\mu}}_{GSE} - \boldsymbol{\mu}_n)^\top \boldsymbol{\Sigma}_n^{-1}(\hat{\boldsymbol{\mu}}_{GSE} - \boldsymbol{\mu}_n), \tag{2.3}$$

The application of (2.2) leads to the following optimization problem

$$\alpha_n^2 \bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1} \bar{\mathbf{y}}_n + \beta_n^2 \boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0 + 2\alpha_n \beta_n \bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0 - 2\alpha_n \bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_n - 2\beta_n \boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0 \longrightarrow \min_{\alpha_n, \beta_n}, \tag{2.4}$$

Taking the derivatives of L with respect to α_n and β_n and setting them equal to zero we get

$$\frac{\partial L}{\partial \alpha_n} = \alpha_n \bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1} \bar{\mathbf{y}}_n + \beta_n \bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0 - \bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_n = 0, \tag{2.5}$$

$$\frac{\partial L}{\partial \beta_n} = \beta_n \boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0 + \alpha_n \bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0 - \boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0 = 0. \tag{2.6}$$

The Hessian of L is given by

$$\mathbf{H} = \begin{pmatrix} \bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1} \bar{\mathbf{y}}_n & \bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0 \\ \bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0 & \boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0 \end{pmatrix}, \tag{2.7}$$

which is a positive definite matrix with probability 1, since $\bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1} \bar{\mathbf{y}}_n > 0$ with probability 1 and

$$\det(\mathbf{H}) = \bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1} \bar{\mathbf{y}}_n \boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0 - (\bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0)^2 > 0, \tag{2.8}$$

with probability 1 following the Cauchy-Schwarz inequality applied to the vectors $\boldsymbol{\Sigma}_n^{-1/2} \bar{\mathbf{y}}_n$ and $\boldsymbol{\Sigma}_n^{-1/2} \boldsymbol{\mu}_0$.

Thus, the optimal shrinkage intensities are given by

$$\alpha_n^* = \alpha_n^*(\bar{\mathbf{y}}_n, \boldsymbol{\Sigma}_n, \boldsymbol{\mu}_n, \boldsymbol{\mu}_0) = \frac{\bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_n \boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0 - \boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0 \bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0}{\bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1} \bar{\mathbf{y}}_n \boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0 - (\bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0)^2}, \tag{2.9}$$

$$\beta_n^* = \beta_n^*(\bar{\mathbf{y}}_n, \boldsymbol{\Sigma}_n, \boldsymbol{\mu}_n, \boldsymbol{\mu}_0) = \frac{\bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1} \bar{\mathbf{y}}_n \boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0 - \bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0 \bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_n}{\bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1} \bar{\mathbf{y}}_n \boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0 - (\bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0)^2}. \tag{2.10}$$

In Theorem 2.1, we show that the optimal shrinkage intensities α_n^* and β_n^* are almost surely asymptotically equivalent to nonrandom quantities α^* and β^* under the large-dimensional asymptotics $n^{-1} p^{1-\gamma} \longrightarrow c \in (0, +\infty)$ as $n \longrightarrow \infty$.

Theorem 2.1. *Assume that (A1) - (A4) hold. Then*

$$|\alpha_n^* - \alpha^*| \xrightarrow{a.s.} 0, \tag{2.11}$$

$$|\beta_n^* - \beta^*| \xrightarrow{a.s.} 0, \tag{2.12}$$

4

for $n^{-1}p^{1-\gamma} \longrightarrow c > 0$ as $n \longrightarrow \infty$, where

$$\alpha^* = \alpha^*(\boldsymbol{\Sigma}_n, \boldsymbol{\mu}_n, \boldsymbol{\mu}_0) = 1 - \frac{p^\gamma c \boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0}{(p^\gamma c + \boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_n) \boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0 - (\boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0)^2}, \tag{2.13}$$

$$\beta^* = \beta^*(\boldsymbol{\Sigma}_n, \boldsymbol{\mu}_n, \boldsymbol{\mu}_0) = (1 - \alpha^*) \frac{\boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0}{\boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0}. \tag{2.14}$$

Note that $\alpha^* \in (0, 1)$ due to inequality (2.8). Furthermore, using the results of Theorem 2.1, we are able to estimate α^* and β^* consistently at least for $c \in (0, 1)$ which is shown in Theorem 3.1 below.

It is remarkable that the proposed procedure is very different to the one suggested by Wang et al. (2014), where the authors minimized the expected quadratic loss and estimated optimal shrinkage intensities. Wang et al. (2014) found the estimators for the optimal shrinkage intensities which converge in probability while our aim is to construct consistent estimators which converge almost surely. It is worth pointing out the following remarks.

Remark 1. *If $\gamma > 0$ and the large dimensional asymptotic chosen such that $n^{-1}p \longrightarrow c$ as $n \longrightarrow \infty$ instead of $n^{-1}p^{1-\gamma} \longrightarrow c$ as $n \longrightarrow \infty$, then from the expression of α^* we get that $\alpha^* \longrightarrow 1$ a.s. Hence, the shrinkage estimator tends to the sample mean in this case.*

Remark 2. *The technical assumption (A2) with the same γ's is really important. If this condition does not hold, i.e. there exist γ_1 and γ_2 such that $\lim_{p \to \infty} p^{-\gamma_1} ||\boldsymbol{\mu}_n||^2 = M_1$ and $\lim_{p \to \infty} p^{-\gamma_2} ||\boldsymbol{\mu}_0||^2 = M_0$ with $0 < M_l \leq M_0, M_1 \leq M_u < \infty$, then from the expression of β_n^* we have that the rate for the numerator is $\gamma_1 + (\gamma_1 + \gamma_2)/2$ whereas the rate for the denominator is $\gamma_1 + \gamma_2$. Consequently, if $\gamma_1 \neq \gamma_2$ then*

$$\beta_n^* \xrightarrow{a.s.} \begin{cases} 0 & \text{for } n^{-1}p \to c > 0 \text{ as } n \to \infty & \text{if } \gamma_1 < \gamma_2, \\ \infty & \text{for } n^{-1}p \to c > 0 \text{ as } n \to \infty & \text{if } \gamma_1 > \gamma_2. \end{cases}$$

Let $q_{ij} = p^{-\gamma} \boldsymbol{\mu}_i^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_j$, for $i, j \in \{0, n\}$ and $d = q_{00}q_{nn} - q_{0n}^2$. In Theorem 2.2, we prove that α_n^* and β_n^* are asymptotically normally distributed under the high-dimensional asymptotic regime.

Theorem 2.2. *Assume that (A1) - (A4) hold. Then*

$$\sqrt{p^\gamma n} \sigma_\alpha^{-1} (\alpha_n^* - \alpha^*) \xrightarrow{d} N(0, 1), \tag{2.15}$$

$$\sqrt{p^\gamma n} \sigma_\beta^{-1} (\beta_n^* - \beta^*) \xrightarrow{d} N(0, 1), \tag{2.16}$$

for $n^{-1}p^{1-\gamma} \to c > 0$ as $n \to \infty$, where

$$\sigma_\alpha^2 = \frac{(cq_{00} - d)^2 q_{00} d + cd^2 q_{00}^2}{(cq_{00} + d)^4} \tag{2.17}$$

$$\sigma_\beta^2 = \frac{1}{(cq_{00} + d)^4} \{(d - cq_{00})^2 q_{0n}q_{nn} + (cq_{0n}^2 - cd - dq_{nn})^2 q_{00} + cd^2 q_{0n}^2$$
$$+ 2(cq_{0n}^2 - cd - dq_{nn})(d - cq_{00})q_{0n}^2\}. \tag{2.18}$$

3 Bona fide estimator

This section presents consistent estimators for α^* and β^*, i.e. for the deterministic equivalent quantities to the optimal shrinkage estimators α_n^* and β_n^*, which we denote by $\hat{\alpha}^*$ and $\hat{\beta}^*$. This procedure allows us to construct the bona fide estimators for the unknown shrinkage intensities. Using recent results from random matrix theory, we further prove that $\hat{\alpha}^*$ and $\hat{\beta}^*$ are consistent and asymptotically normally distributed. In this section we work under asymptotic regime $n^{-1}p^{1-\gamma} \longrightarrow c \in (0,1)$ as $n \longrightarrow \infty$.

Let

$$\mathbf{S}_n = n^{-1} \left(\mathbf{Y}_n - \bar{\mathbf{y}}_n \mathbf{1}_n^\top \right) \left(\mathbf{Y}_n - \bar{\mathbf{y}}_n \mathbf{1}_n^\top \right)^\top = n^{-1} \mathbf{Y}_n \mathbf{Y}_n^\top - \bar{\mathbf{y}}_n \bar{\mathbf{y}}_n^\top \tag{3.1}$$

be the sample covariance matrix. In Theorem 3.1 below we present the consistent estimators for α^* and β^* under the large-dimensional asymptotics.

Theorem 3.1. *Assume (A1)-(A2) and let the elements of $\mathbf{X}_n$ possess uniformly bounded $4+\varepsilon$ moments with $\varepsilon > 0$. Let $n^{-1}p^{1-\gamma} \to c \in (0,1)$ for $n \to \infty$. Then the consistent estimators for α^* and β^* are given by*

$$\hat{\alpha}^* = \hat{\alpha}^*(\bar{\mathbf{y}}_n, \mathbf{S}_n, \boldsymbol{\mu}_0) = 1 - \frac{(1-c)^{-1} c p^\gamma \boldsymbol{\mu}_0^\top \mathbf{S}_n^{-1} \boldsymbol{\mu}_0}{\{\frac{c(p^\gamma-1)}{1-c} + \bar{\mathbf{y}}_n^\top \mathbf{S}_n^{-1} \bar{\mathbf{y}}_n\} \boldsymbol{\mu}_0^\top \mathbf{S}_n^{-1} \boldsymbol{\mu}_0 - (\bar{\mathbf{y}}_n^\top \mathbf{S}_n^{-1} \boldsymbol{\mu}_0)^2}, \tag{3.2}$$

$$\hat{\beta}^* = \hat{\beta}^*(\bar{\mathbf{y}}_n, \mathbf{S}_n, \boldsymbol{\mu}_0) = (1 - \hat{\alpha}^*) \frac{\bar{\mathbf{y}}_n^\top \mathbf{S}_n^{-1} \boldsymbol{\mu}_0}{\boldsymbol{\mu}_0^\top \mathbf{S}_n^{-1} \boldsymbol{\mu}_0}. \tag{3.3}$$

Next, we prove that the consistent estimators for the shrinkage intensities are asymptotically normally distributed. This result is investigated under an additional condition imposed on the distribution of the entries of $\mathbf{X}_n$ which are assumed to be standard normally distributed. It has to be noted that only in Theorem 3.2 the assumption of normality is used, whereas the existence of the second (fourth) moments is only required for all other results.

Theorem 3.2. *Assume (A1)-(A2) and let the elements of $\mathbf{X}_n$ are standard normally distributed. Let $n^{-1}p \to c \in (0,1)$ (i.e., $\gamma = 0$) for $n \to \infty$. Then*

$$\sqrt{n}\boldsymbol{\Omega}^{-1/2} \left(\begin{array}{c} \hat{\alpha}^* - \alpha^* \\ \hat{\beta}^* - \beta^* \end{array} \right) \xrightarrow{d} N \left(\left(\begin{array}{c} 0 \\ 0 \end{array} \right), \mathbf{I} \right)$$

where

$$\boldsymbol{\Omega} = \left(\begin{array}{cc} \frac{c^2 \sigma_s^2}{(c+s)^4} & \frac{c^2 \sigma_s^2}{(c+s)^4} R \\ \frac{c^2 \sigma_s^2}{(c+s)^4} R & \frac{c^2 \sigma_s^2}{(c+s)^4} R^2 + \frac{c^2}{(c+s)^2} \frac{1+\frac{s+c}{1-c}}{\boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0} \end{array} \right)$$

where

$$s = \boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_n - \frac{(\boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_n)^2}{\boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0}, \quad R = \frac{\boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_n}{\boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0} \quad and \quad \sigma_s^2 = 2(c+2s) + \frac{2}{1-c}(c+s)^2.$$

The bona fide optimal shrinkage estimator for the mean vector in the case $c < 1$ is constructed by

$$\hat{\boldsymbol{\mu}}_{OLSE} = \hat{\alpha}^* \bar{\mathbf{y}}_n + \hat{\beta}^* \boldsymbol{\mu}_0, \tag{3.4}$$

where the optimal shrinkage intensities are given by (3.2) and (3.3), respectively. This estimator obeys almost surely smallest quadratic loss under the large-dimensional asymptotics. We refer to it as the Optimal Linear Shrinkage Estimator (OLSE) for the high-dimensional mean vector. It is obvious that the OLSE estimator (3.4) dominates the sample estimator uniformly if both p and n tend to infinity and $n^{-1}p^{1-\gamma} \to c < 1$.

In case $c > 1$ the sample covariance matrix is not invertible anymore and we need other techniques to estimate the unknown quantities given in (2.13) and (2.14). Here, we apply the generalized inverse of the sample covariance matrix $\mathbf{S}_n$. Particularly, we use the following generalized inverse of the sample covariance matrix $\mathbf{S}_n$

$$\mathbf{S}_n^- = \boldsymbol{\Sigma}_n^{-1/2} \left(\frac{1}{n} \mathbf{X}_n \mathbf{X}_n^\top - \bar{\mathbf{x}}_n \bar{\mathbf{x}}_n^\top \right)^+ \boldsymbol{\Sigma}_n^{-1/2}, \tag{3.5}$$

where $'+'$ denotes the Moore-Penrose inverse. It can be shown that $\mathbf{S}_n^-$ is a generalized inverse of $\mathbf{S}_n$ satisfying $\mathbf{S}_n^- \mathbf{S}_n \mathbf{S}_n^- = \mathbf{S}_n^-$ and $\mathbf{S}_n \mathbf{S}_n^- \mathbf{S}_n = \mathbf{S}_n$. However, $\mathbf{S}_n^-$ is not exactly equal to the Moore-Penrose inverse because it does not satisfy the conditions $(\mathbf{S}_n^- \mathbf{S}_n)^\top = \mathbf{S}_n^- \mathbf{S}_n$ and $(\mathbf{S}_n \mathbf{S}_n^-)^\top = \mathbf{S}_n \mathbf{S}_n^-$. In case $c < 1$ the generalized inverse $\mathbf{S}_n^-$ coincides with the usual inverse $\mathbf{S}_n^{-1}$. Moreover, if $\boldsymbol{\Sigma}_n$ is a multiple of identity matrix then $\mathbf{S}_n^-$ is equal to the Moore-Penrose inverse $\mathbf{S}_n^+$. So, it could be expected that if $\boldsymbol{\Sigma}_n$ is a sparse matrix, then both the inverses are very close. This conjecture is not shown here and it is left for future research.

In Theorem 3.3 below we present the consistent estimators for α^* and β^* under large-dimensional asymptotics in the case $c > 1$ utilizing the generalized inverse $\mathbf{S}_n^-$.

Theorem 3.3. *Assume (A1)-(A2) and let the elements of $\mathbf{X}_n$ possess uniformly bounded $4 + \varepsilon$ moments with $\varepsilon > 0$. Let $n^{-1}p^{1-\gamma} \to c \in (1, +\infty)$ for $n \to \infty$. Then the consistent estimators for α^* and β^* are given by*

$$\hat{\alpha}^* = \hat{\alpha}^*(\bar{\mathbf{y}}_n, \mathbf{S}_n, \boldsymbol{\mu}_0) = 1 - \frac{(c-1)^{-1} p^\gamma \boldsymbol{\mu}_0^\top \mathbf{S}_n^- \boldsymbol{\mu}_0}{\{ \frac{(p^\gamma - 1)}{c-1} + \bar{\mathbf{y}}_n^\top \mathbf{S}_n^- \bar{\mathbf{y}}_n \} \boldsymbol{\mu}_0^\top \mathbf{S}_n^- \boldsymbol{\mu}_0 - (\bar{\mathbf{y}}_n^\top \mathbf{S}_n^- \boldsymbol{\mu}_0)^2}, \tag{3.6}$$

$$\hat{\beta}^* = \hat{\beta}^*(\bar{\mathbf{y}}_n, \mathbf{S}_n, \boldsymbol{\mu}_0) = (1 - \hat{\alpha}^*) \frac{\bar{\mathbf{y}}_n^\top \mathbf{S}_n^- \boldsymbol{\mu}_0}{\boldsymbol{\mu}_0^\top \mathbf{S}_n^- \boldsymbol{\mu}_0}. \tag{3.7}$$

Because $\mathbf{S}_n^-$ depends on the unknown quantities we will approximate it by the Moore-Penrose inverse $\mathbf{S}_n^+$. The asymptotic properties of the Moore-Penrose inverse under high-dimensional settings are investigated in Bodnar et al. (2016a). Worth mentioning is that this procedure does not obviously lead to the optimal shrinkage estimator as it was for $c < 1$. This method is only suboptimal, but nevertheless it dominates in most of cases the existent estimators for the high-dimensional mean vector given in the literature as it is justified in the next sections via an extensive simulation study and in an empirical illustration.

4 Benchmark methods

This section introduces approaches used as benchmarks in the simulation study of Section 5. Probably, the most commonly used estimator for the mean vector in the literature is the sample mean vector expressed as

$$\bar{\mathbf{y}}_n = n^{-1} \mathbf{Y}_n \mathbf{1}_n. \tag{4.1}$$

Although this estimator is known to be inadmissible under quadratic loss, we nevertheless use the sample mean in our comparison study.

The original James-Stein estimator was derived in the form $\hat{\boldsymbol{\mu}}_{n,JSN} = (1 - \frac{p-2}{n\bar{\mathbf{y}}_n^\top \bar{\mathbf{y}}_n})\bar{\mathbf{y}}_n$ for $\boldsymbol{\Sigma}_n = \mathbf{I}_p$ and $n > p > 2$. In the comparison study we make use of a modified version of this estimator given by

$$\hat{\boldsymbol{\mu}}_{n,JS} = \left\{ 1 - \frac{(p-2)/(n-p-3)}{\bar{\mathbf{y}}_n^\top \widetilde{\mathbf{S}}_n^{-1} \bar{\mathbf{y}}_n} \right\} \bar{\mathbf{y}}_n$$

for $c < 1$ and an estimator $\widetilde{\mathbf{S}}_n \sim Wishart_p(n, \boldsymbol{\Sigma})$ of the covariance matrix $\boldsymbol{\Sigma}_n$ with $n \geq p \geq 3$. In the case of $p > n \geq 3$, we compare our estimator with those proposed by Chételat and Wells (2012) defined in the Baranchik type estimator as follows

$$\hat{\boldsymbol{\mu}}_{n,JS(p>n)} = \left\{ I_p - \frac{a\widetilde{\mathbf{S}}_n \widetilde{\mathbf{S}}_n^+}{\bar{\mathbf{y}}_n^\top \widetilde{\mathbf{S}}_n^+ \bar{\mathbf{y}}_n} \right\} \bar{\mathbf{y}}_n$$

with $0 \leq a \leq \frac{2(n-2)}{p-n+3}$ and $\widetilde{\mathbf{S}}_n^+$ the Moore-Penrose inverse of $\widetilde{\mathbf{S}}_n$. In the whole study we set $a = \frac{2(n-2)}{p-n+3}$. As shown by the simulation study of Chételat and Wells (2012), the so-called positive-parttype James-Stein estimator of the form

$$\hat{\boldsymbol{\mu}}_{n,JS+} = (I_p + \widetilde{\mathbf{S}}_n \widetilde{\mathbf{S}}_n^+)\bar{\mathbf{y}}_n + \left\{ I_p - \frac{a}{\bar{\mathbf{y}}_n^\top \widetilde{\mathbf{S}}_n^+ \bar{\mathbf{y}}_n} \right\}_+ \widetilde{\mathbf{S}}_n \widetilde{\mathbf{S}}_n^+ \bar{\mathbf{y}}_n,$$

with $b_+ = \max(b, 0)$ and $a = \frac{n-2}{p-n+3}$ dominates $\hat{\boldsymbol{\mu}}_{n,JS(p>n)}$ under the invariant loss.

Another benchmark estimator is taken from Wang et al. (2014) which is a shrinkage estimator with the unity target vector and shrinkage coefficients found by the minimization of the expected quadratic loss. Thus, this shrinkage estimator is given by (c.f. Wang et al. (2014))

$$\hat{\boldsymbol{\mu}}_{n,W} = \frac{Z_{1,n} - Z_{4,n}}{Z_{1,n} + Z_{2,n}Z_{4,n}}\bar{\mathbf{y}}_n + \frac{Z_{2,n}}{Z_{1,n} + Z_{2,n}Z_{4,n}}Z_{3,n}\mathbf{1}_n, \tag{4.2}$$

with

$$Z_{1,n} = \frac{1}{p(n-1)} \sum_{i \neq j} \mathbf{Y}_{n,i}^\top \widetilde{\mathbf{S}}_n^+ \mathbf{Y}_{n,j}, \quad Z_{2,n} = \frac{1}{np} \left(\sum_{k=1}^n \mathbf{Y}_{n,k}^\top \widetilde{\mathbf{S}}_n^+ \mathbf{Y}_{n,k} - \frac{1}{n-1} \sum_{i \neq j} \mathbf{Y}_{n,i}^\top \widetilde{\mathbf{S}}_n^+ \mathbf{Y}_{n,j} \right),$$

$$Z_{3,n} = \frac{1}{n\mathbf{1}_n^\top \widetilde{\mathbf{S}}_n^+ \mathbf{1}_n} \sum_{k=1}^n \mathbf{1}_n^\top \widetilde{\mathbf{S}}_n^+ \mathbf{Y}_{n,k}, \quad Z_{4,n} = \frac{1}{p(n-1)\mathbf{1}_n^\top \widetilde{\mathbf{S}}_n^+ \mathbf{1}_n} \sum_{i \neq j} \mathbf{1}_n^\top \widetilde{\mathbf{S}}_n^+ \mathbf{Y}_{n,i} \mathbf{Y}_{n,j}^\top \widetilde{\mathbf{S}}_n^+ \mathbf{1}_n,$$

for $n^{-1}p > 1$. The estimator 4.2 has a computationally complicated form because of the double sum over p and n, being therefore very time consuming for large dimensions and large sample sizes. Therefore, in practice, its asymptotic counterpart is considered (see, e.g., Wang et al. (2014)).

5   Finite-sample performance

This section provides an extensive simulation study, to test the validity of Theorems 2.2 and 3.2 as well as to compare the quality of the proposed OLSE estimator with the considered benchmark methods. In all simulations the true mean vector $\boldsymbol{\mu}_n$ has been simulated from the uniform distribution on $[-p^{-1/2}, p^{-1/2}]$. This choice of the mean vector is motivated by the

empirical illustration of Section 6, where the obtained theoretical results are applied to financial data consisting of asset returns which usually possess small expected values. Eigenvalues of true covariance matrix $\boldsymbol{\Sigma}_n$ are distributed as follows: 20% are equal to 1, 40% equals 3 and rest 40% equals to 10. The target mean vector is selected being uniformly distributed with the same norm as the $\boldsymbol{\mu}_n$. Moreover, we set $\gamma = 0$ in all of the considered cases.

First, we considered the finite sample behavior of the optimal shrinkage coefficients α_n^* and β_n^* and their bone fide estimators $\hat{\alpha}^*$ and $\hat{\beta}^*$ and compared them to the corresponding asymptotic distributions which are presented in Theorems 2.2 and 3.2. For $p \in \{20, 100, 250, 500\}$ and $c \in \{0.5, 0.9, 2.0\}$ we simulated $\boldsymbol{\mu}_n, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_n$ as described above and drew the columns of $\mathbf{Y}_n$ from $\mathrm{N}(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$ in each simulation run. Then, α_n^* and β_n^* were calculated and standardized by the corresponding asymptotic variances from Theorem 2.2 for $c \in \{0.5, 0.9, 2.0\}$. In the case of $c \in \{0.5, 0.9\}$, we additionally computed $\hat{\alpha}^*$ and $\hat{\beta}^*$ and standardized them by their asymptotic variances as presented in Theorem 3.2. The procedure was repeated $N = 1000$ times resulting in N values of $\alpha_n^*, \beta_n^*, \hat{\alpha}^*$, and $\hat{\beta}^*$ for chosen p and c. Kernel density estimators (KDE) for α_n^* and β_n^* (with Epanechnikov kernel and a bandwidth selected through the Silverman's rule of thumb) are depicted in Figure 1, while Figure 2 shows the corresponding results in the case of $\hat{\alpha}^*$ and $\hat{\beta}^*$. It is remarkable that Figure 1 supports the findings of Theorem 2.2 by resembling almost perfectly the shape of the standard normal distribution (shown in black) even for $p = 100$ with the exception for $c = 2$ in the case of β_n^*. Some deviations at the peak of the distribution are visible for small dimension $p = 20$ and sample sizes $n = 40$, $n = 22$ and $n = 10$ for $c = 0.5$, $c = 0.9$ and $c = 2.0$, respectively. The finite-sample performance of the asymptotic results presented in Theorem 3.2 are investigated in Figure 2 for $c < 1$. Here, we observe a perfect fit to the normal distribution even for small sample sizes and small dimensions.

Next, the quality of the estimators were measured by the quadratic loss expressed as

$$L(\widehat{\boldsymbol{\mu}}, \boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n) = (\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}_n)^\top \boldsymbol{\Sigma}_n^{-1} (\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}_n),$$

where $\widehat{\boldsymbol{\mu}}$ is an estimator for $\boldsymbol{\mu}_n$. The results are depicted in Figures 3 and 4 in the first and in the second columns, while the third columns present the time in log-seconds needed to compute each of the estimators (except the asymptotic one). The left columns show the results for the distribution of eigenvalues of $\boldsymbol{\Sigma}_n$ in proportions 20%, 40%, 40% for values 1, 3 and 10, while in the middle column $\lambda_{max} = p$ was chosen which corresponds to the extreme lambda case. For each p and c the procedure has been repeated $N = 100$ times. As shown in Figure 3, the optimal shrinkage estimator and its asymptotic counterparts with α_n^* and β_n^* replaced by α^* and β^* showed the best performance. This behavior is not surprising since they both contain unobservable information (true values of $\boldsymbol{\Sigma}_n$ or $\boldsymbol{\mu}_n$). The proposed bona fide estimator appears to be equivalent to the James-Stein estimator. For the case of $c > 1$, the Moore-Penrose inverse of the sample covariance matrix was employed. We observed that $\widehat{\boldsymbol{\mu}}_{n,W}$ performs pretty unstable for moderate dimensions and it was extremely time consuming for large dimensions. The new estimator dominates the rest of competitors. Worth mentioning that the loss function for $\bar{\mathbf{y}}_n$ is close to c in all of the considered cases.
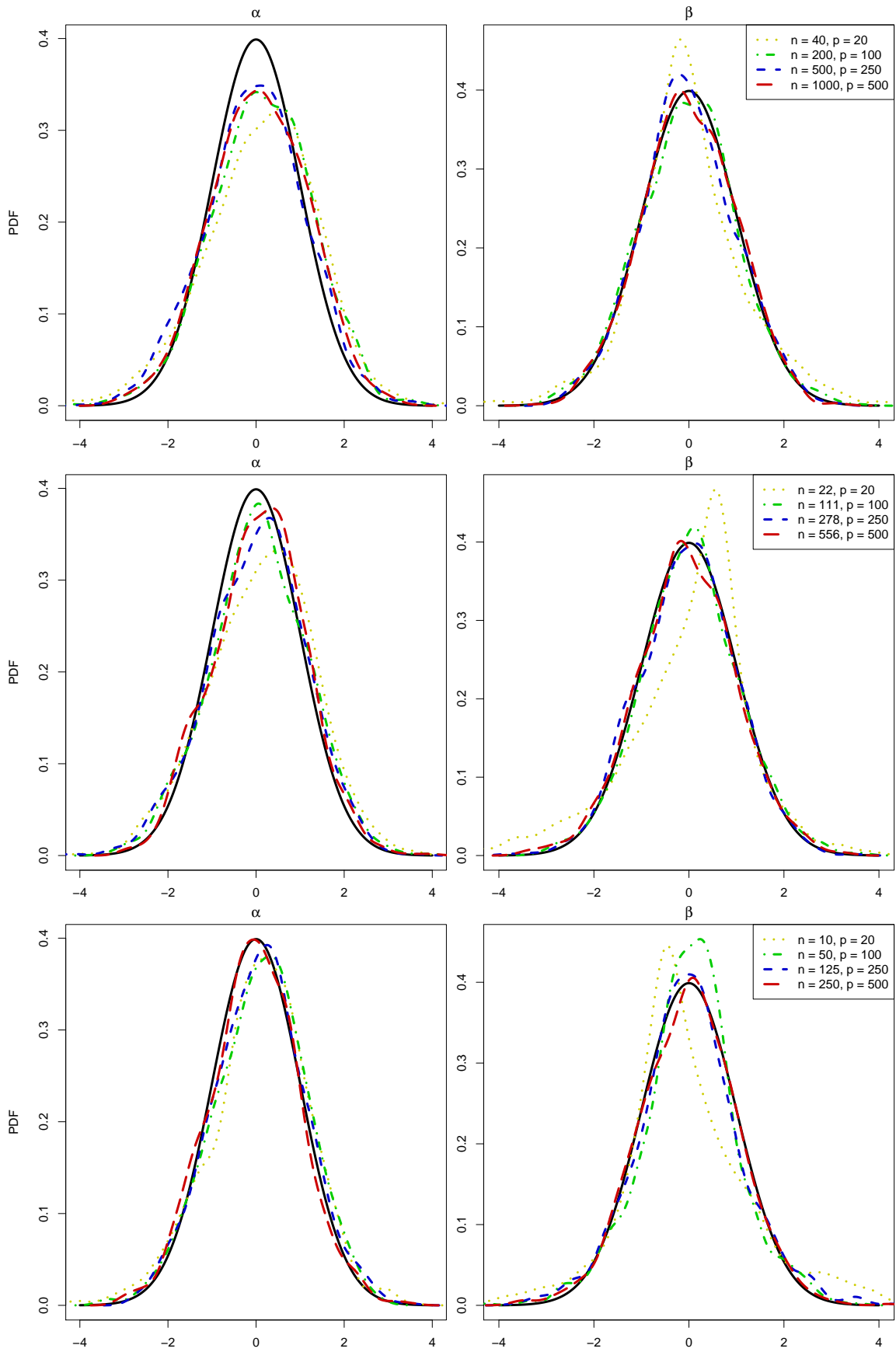
Figure 1: KDE for α_n^* and β_n^* in the case of $\gamma = 0$. We set c = 0.5 (top panel), c = 0.9 (middle panel) and c = 2.0 (bottom panel). The solid black line corresponds to the density of the standard normal distribution.
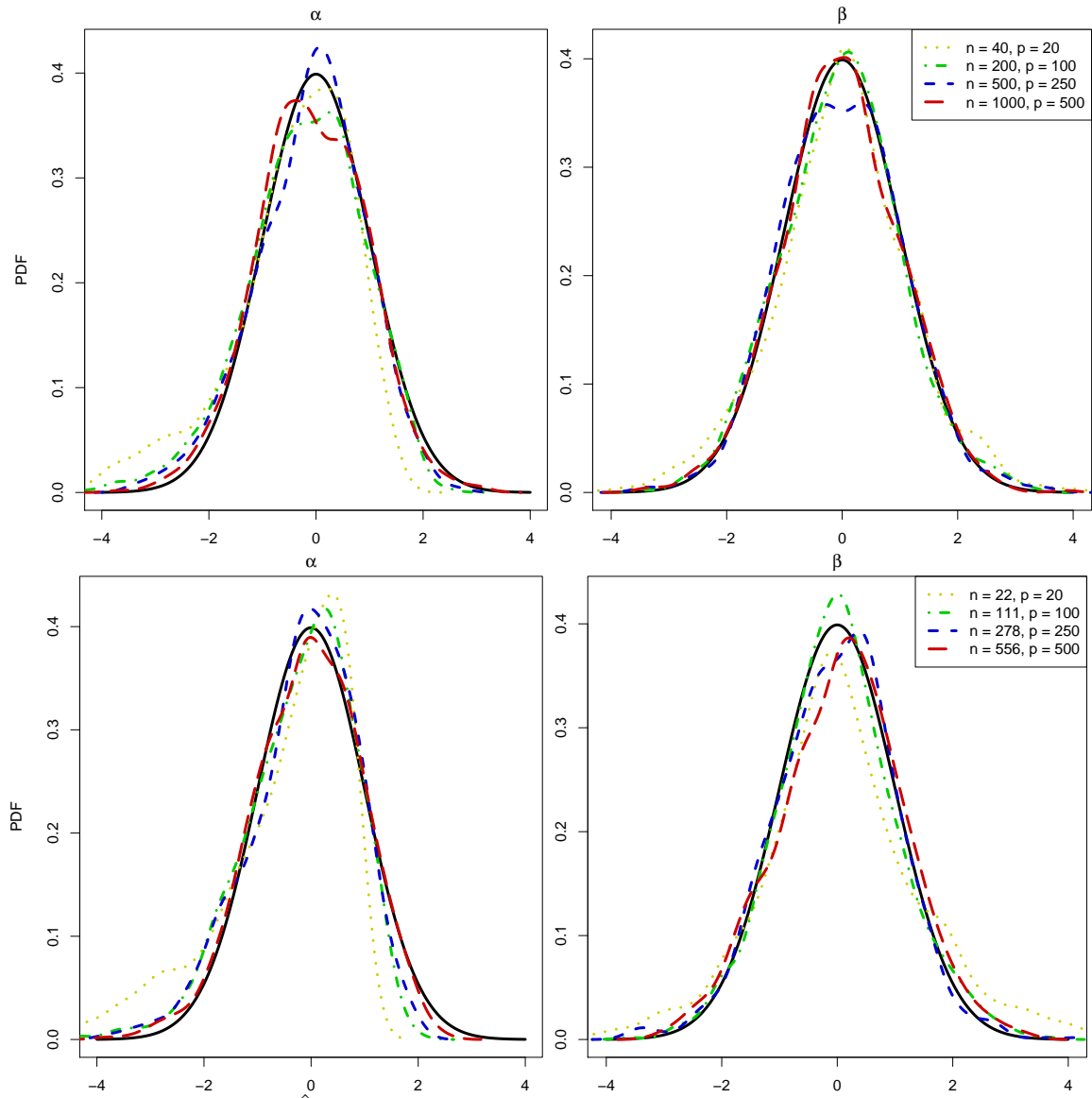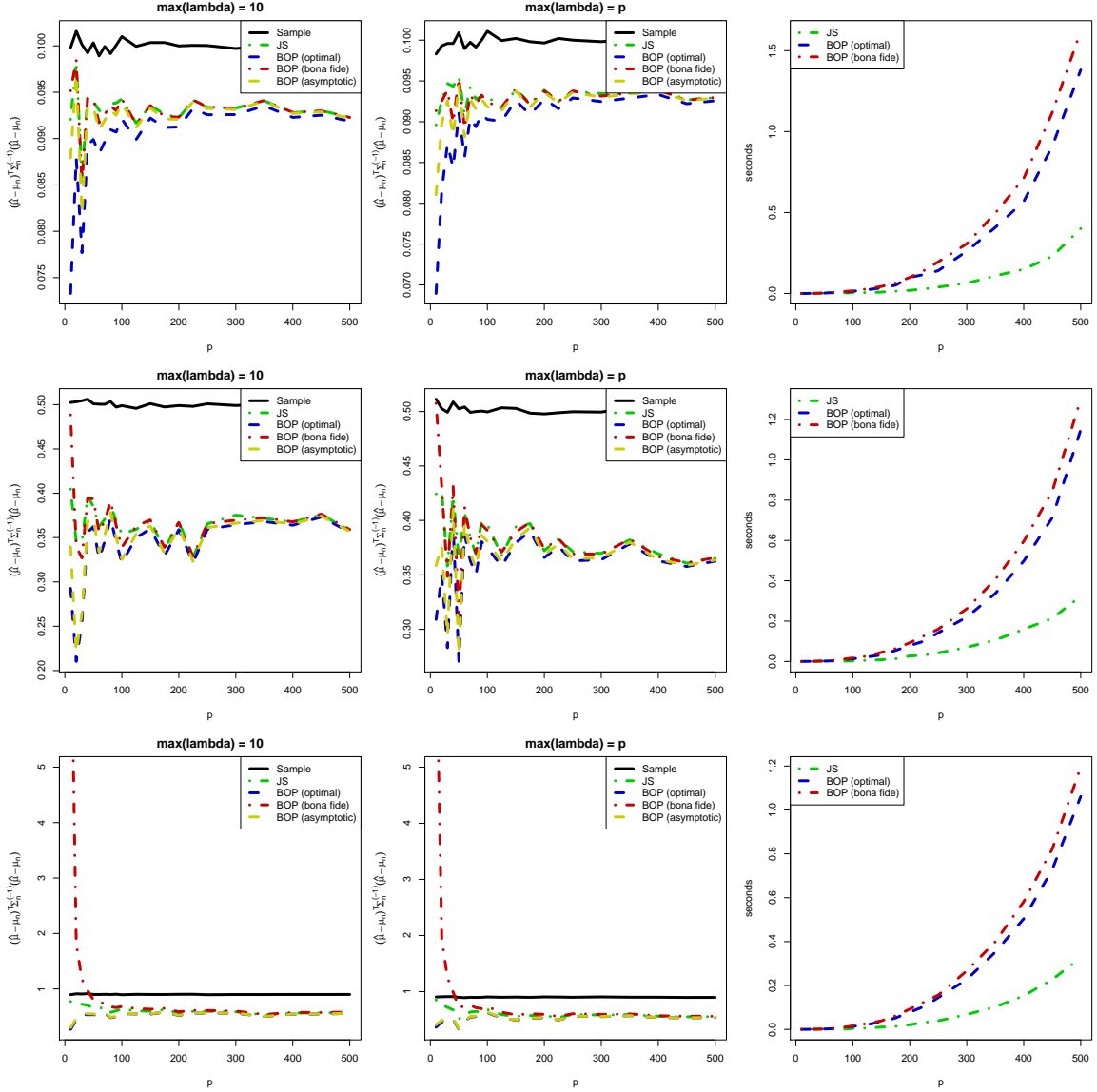
Figure 2: KDE for $\hat{\alpha}^*$ and $\hat{\beta}^*$ in the case of $\gamma = 0$. We set c = 0.5 (top panel), c = 0.9 (middle panel) and c = 2.0 (bottom panel). The solid black line corresponds to the density of the standard normal distribution.
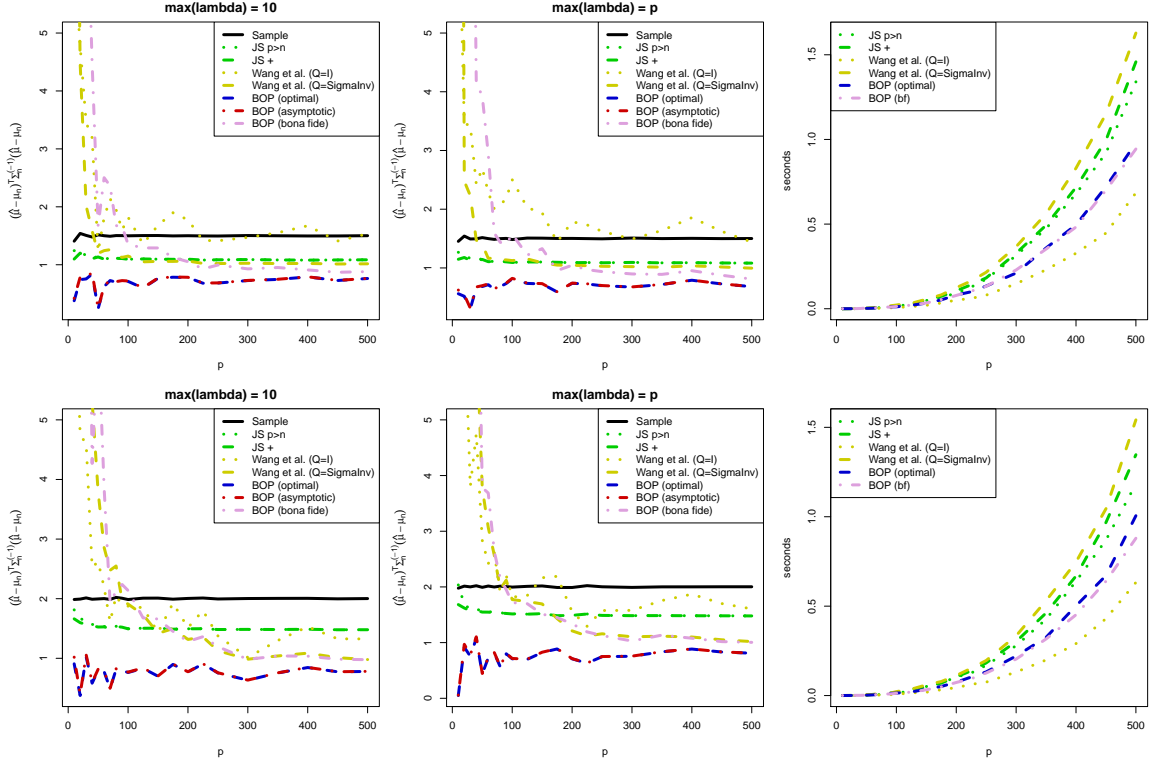
Figure 3: Quadratic loss function for estimators $\bar{\mathbf{y}}_n$ (black), $\hat{\boldsymbol{\mu}}_n$ (bona fide, red), $\boldsymbol{\mu}_n^*$ (optimal, blue), $\boldsymbol{\mu}^*$ (asymptotic, yellow), $\widehat{\boldsymbol{\mu}}_{n,JS}$ (green) of $\boldsymbol{\mu}_n$, performed with $N = 100$ iterations for $c = 0.1$ (top), $c = 0.5$ (middle) and $c = 0.9$ (bottom).

Figure 4: Quadratic loss function for estimators $\bar{\mathbf{y}}_n$ (black), $\boldsymbol{\mu}^*$ (asymptotic, red), $\boldsymbol{\mu}_n^*$ (optimal, blue), $\hat{\boldsymbol{\mu}}_n$ (bona fide, purple), $\hat{\boldsymbol{\mu}}_{n,W}$ (yellow, dotted for $\mathbf{Q} = \mathbf{I}$ and dashed for $\mathbf{Q} = \mathbf{S}^+$), $\hat{\boldsymbol{\mu}}_{n,JS(p>n)}$ (green, dotted), $\hat{\boldsymbol{\mu}}_{n,JS+}$ (green, dashed) of $\boldsymbol{\mu}_n$, performed with $N = 100$ iterations, for $c = 1.5$ (top) and $c = 2.0$ (bottom).

6   Financial application

In order to implement the suggested procedure to real data, we considered weekly log-returns $\mathbf{y}_t = \{y_{1t}, \ldots, y_{pt}\}^\top$ of the $p = 412$ constituents of the S&P500 index traded over 783 weeks in the period from 12.06.2001 till 09.06.2016. We compared the ability of the considered estimators for the mean vector to predict the future realisation of the expected return of the equally-weighted portfolio which is a popular trading strategy in the high-dimensional case (see, e.g. DeMiguel et al. (2009), Tu and Zhou (2011)). For that reason we applied the rolling window estimator with window sizes of $n \in \{25, 50, 75, 100\}$ leading to $c \in \{16.48, 8.24, 5.49, 4.12\}$. Namely, for each fixed date t we took historical data of asset returns with n most recent observations $\mathbf{Y}_{t,n} = \{y_{i\tau}\}_{\tau=\{t-n,\ldots,t\}, i=\{1,\ldots,p\}}$ and estimated the expected return of the equally-weighted portfolio which was then compared to the realized return of this portfolio on day $t + 1$.

n	c	$\bar{\mathbf{y}}_n$	$\widehat{\boldsymbol{\mu}}_{JS}$		$\widehat{\boldsymbol{\mu}}_{n,W}$		$\widehat{\boldsymbol{\mu}}_n$ (bf)
			$\widehat{\boldsymbol{\mu}}_{JS(p>n)}$	$\widehat{\boldsymbol{\mu}}_{JS+}$	$\mathbf{Q} = \mathbf{I}$	$\mathbf{Q} = \mathbf{S}^+$	
25	16.48	7.882	7.832	7.832	7.882	7.931	*7.831*
50	8.24	7.813	7.721	7.721	7.813	7.748	*7.716*
75	5.49	7.761	7.664	7.665	7.761	*7.660*	*7.660*
100	4.12	7.723	7.646	7.646	7.723	*7.630*	7.636

Table 1: Performance of different mean estimators on the basis of equally weighted portfolio for weekly returns for different sample sizes. With italic font we marked best models for each sample size.

For a given estimator $\hat{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}_n$, the estimated expected return of the equally-weighted portfolio is calculated by $\hat{R}_{t,n}(\hat{\boldsymbol{\mu}}) = p^{-1}\mathbf{1}^\top\hat{\boldsymbol{\mu}}$. Then, the performance of each estimator was measured by the average quadratic deviation (times 10^4 for visualization) from the realized return of the equally-weighted portfolio expressed as

$$L = 10^4 \cdot T^{-1} \sum_{t=1}^{T} \{\hat{R}_{t,n}(\hat{\boldsymbol{\mu}}) - p^{-1}\mathbf{1}^\top\mathbf{y}_{t+1}\}^2,$$

where $T = 683$ corresponds to the number of rolling windows. The results are depicted in Table 1. A very good performance of the suggested shrinkage estimator is observed which outperforms the other estimators in three out of four cases. Only for larger sample size $n = 100$ (with $c = 4.12$), the estimator $\widehat{\boldsymbol{\mu}}_{n,W}$ (with $\mathbf{Q} = \mathbf{S}^+$) performs slightly better than the new estimator. On the other hand, $\widehat{\boldsymbol{\mu}}_{n,W}$ (with $\mathbf{Q} = \mathbf{S}^+$) possesses the largest value of the average quadratic deviation if $n = 25$. Also, we point out a very bad performance of the sample mean vector which is always one of the worst estimation strategy. Interestingly, the estimator $\widehat{\boldsymbol{\mu}}_{n,W}$ (with $\mathbf{Q} = \mathbf{I}$) has always almost the same values of the average quadratic deviation as ones obtained in the case of the sample mean vector.

7 Summary

Nowadays, modern scientific data sets involve the large number of sample points which is often comparable or even less than the dimension (number of features) of the data generating process. In this situation many statistical, financial and genetic problems require simple and feasible estimators of the mean vector. Although the most of the classical multivariate procedures are based on the limit theorems assuming that the dimension p is fixed and the sample size n increases, it has been pointed out by numerous authors that this assumption does not yield precise estimators for commonly used quantities in large dimensions, and that better estimators can be obtained considering scenarios where the dimension tends to infinity as well (see, e.g., Bai and Silverstein (2004) and references therein).

In our paper we estimate the mean vector using the optimal shrinkage technique utilizing the theory of random matrices. We prove the asymptotic normality of the optimal shrinkage intensities as well as their almost sure limit convergence under the high-dimensional asymptotic regime. Both the simulation study and the empirical application justify the dominance of the suggested estimation technique over the benchmark methods in terms of both the quadratic loss and the computational time.

Acknowledgment

Appendix

Here are given the proofs of theorems.

Proof of Theorem 2.1 Let

$$\eta_1 = p^{-\gamma}\bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\xi} = n^{-1}p^{-\gamma}\mathbf{1}_n^\top \mathbf{Y}_n^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\xi} = n^{-1}p^{-\gamma}\mathbf{1}_n^\top \mathbf{X}_n^\top \boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{\xi} + p^{-\gamma}\boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\xi}, \qquad (7.1)$$

$$\begin{aligned}\eta_2 &= p^{-\gamma}\bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n = n^{-2}p^{-\gamma}\mathbf{1}_n^\top \mathbf{Y}_n^\top \boldsymbol{\Sigma}_n^{-1}\mathbf{Y}_n\mathbf{1}_n \\ &= n^{-2}p^{-\gamma}\mathbf{1}_n^\top (\mathbf{X}_n^\top \mathbf{X}_n)\mathbf{1}_n + 2n^{-1}p^{-\gamma}\mathbf{1}_n^\top \mathbf{X}_n^\top \boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{\mu}_n + p^{-\gamma}\boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n\,,\end{aligned} \qquad (7.2)$$

where the vector $\boldsymbol{\xi}$ is either $\boldsymbol{\mu}_0$ or $\boldsymbol{\mu}_n$. It holds that

$$n^{-1}p^{-\gamma}\mathbf{1}_n^\top \mathbf{X}_n^\top \boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{\xi} = n^{-1}p^{-\gamma}(\mathbf{1}_n^\top \otimes \boldsymbol{\xi}^\top \boldsymbol{\Sigma}_n^{-1/2})\operatorname{vec}(\mathbf{X}_n)\,, \qquad (7.3)$$

$$n^{-2}p^{-\gamma}\mathbf{1}_n^\top (\mathbf{X}_n^\top \mathbf{X}_n)\mathbf{1}_n = \frac{p^{1-\gamma}}{n}\frac{1}{p}\left\|\frac{1}{\sqrt{n}}\mathbf{X}_n\mathbf{1}_n\right\|^2\,. \qquad (7.4)$$

In (7.3) the sum of pn independent random variables is present with zero means and variances

15

equal to $(\boldsymbol{\xi}^\top\boldsymbol{\Sigma}_n^{-1/2}\mathbf{e}_i)^2$, $i = 1, ..., p$, where $\mathbf{e}_i$ is the i-th basis vector in $I\!\!R^p$. Because

$$\mathrm{Var}\left\{n^{-1}p^{-\gamma}(\mathbf{1}_n^\top\otimes\boldsymbol{\xi}^\top\boldsymbol{\Sigma}_n^{-1/2})\,\mathrm{vec}(\mathbf{X}_n)\right\} = \frac{p^{-2\gamma}n}{n^2}\sum_{i=1}^{p}(\boldsymbol{\xi}^\top\boldsymbol{\Sigma}_n^{-1/2}\mathbf{e}_i)^2 = \frac{p^{-2\gamma}}{n}\boldsymbol{\xi}^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\xi}$$

$$\leq \lambda_{max}(\boldsymbol{\Sigma}_n^{-1})\frac{p^{-\gamma}}{n}p^{-\gamma}\|\boldsymbol{\xi}\|^2 \leq \frac{\lambda_{max}(\boldsymbol{\Sigma}_n^{-1})M_u}{np^\gamma} < \infty\,,$$

the application of Kolmogorov's strong law of large numbers leads to

$$n^{-1}\mathbf{1}_n^\top\mathbf{X}_n^\top\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{\xi} \xrightarrow{a.s.} 0 \quad \text{for } p, n \to \infty\,. \tag{7.5}$$

Next, we consider the elements of the vector $\mathbf{X}_n\mathbf{1}_n$ which are determined by the sums of independently and identically distributed random variables with zero means and unit variances. Hence, from the central limit theorem the components of the vector $\frac{1}{\sqrt{n}}\mathbf{X}_n\mathbf{1}_n$ are independent and standard normally distributed. Consequently, their squares are independent and χ^2-distributed with one degree of freedom. Because the expectation of a χ_1^2-distributed random variable is one and the variance is finite we get that

$$\frac{1}{p}\left\|\frac{1}{\sqrt{n}}\mathbf{X}_n\mathbf{1}_n\right\|^2 \xrightarrow{a.s.} 1 \quad \text{for } p, n \to \infty\,,$$

and, consequently,

$$n^{-2}p^{-\gamma}\mathbf{1}_n^\top(\mathbf{X}_n^\top\mathbf{X}_n)\mathbf{1}_n \xrightarrow{a.s.} c \quad \text{for } n^{-1}p^{1-\gamma} \longrightarrow c > 0 \text{ as } n \longrightarrow \infty\,. \tag{7.6}$$

Hence,

$$\begin{aligned}
\alpha_n^* &= \frac{\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 - \boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0}{\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 - (\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0)^2} \\
&= \frac{p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n p^{-\gamma}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 - p^{-\gamma}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0}{p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n p^{-\gamma}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 - (p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0)^2} \\
&\to \frac{p^{-\gamma}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n p^{-\gamma}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 - p^{-\gamma}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 p^{-\gamma}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0}{(c + p^{-\gamma}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n)p^{-\gamma}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 - (p^{-\gamma}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0)^2} \\
&= \frac{\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 - (\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0)^2}{(p^\gamma c + \boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n)\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 - (\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0)^2} = \alpha^*\,,
\end{aligned}$$

and, similarly,

$$\begin{aligned}
\beta_n^* &= \frac{\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 - \bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n}{\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 - (\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0)^2} \\
&= \frac{p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n p^{-\gamma}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 - p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n}{p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n p^{-\gamma}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 - (p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0)^2} \\
&\to \frac{(c + p^{-\gamma}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n)p^{-\gamma}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 - p^{-\gamma}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 p^{-\gamma}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n}{(c + p^{-\gamma}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n)p^{-\gamma}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 - (p^{-\gamma}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0)^2} = (1 - \alpha^*)\frac{\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0}{\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0}\,.
\end{aligned}$$

The theorem is proved.

In the proof of Theorem 2.2 we make use of the following lemma.

Lemma 7.1. *Let* $\mathbf{Z} \sim \mathrm{N}_m(\mathbf{0}, \mathbf{I}_m)$ *and define*

$$Q_1(\mathbf{Z}) = \sqrt{m}\left(\frac{1}{m}\mathbf{Z}^\top\mathbf{Z} - 1\right) \quad and \quad Q_2(\mathbf{Z}) = \frac{1}{\sqrt{m}}\mathbf{a}^\top\mathbf{Z} \ for\ nonrandom\ \mathbf{a}\ with\ \lim_{m\to\infty}\frac{\mathbf{a}^\top\mathbf{a}}{m} = \tilde{\sigma}^2 \,.$$

Then

$$\begin{pmatrix} Q_1(\mathbf{Z}) \\ Q_2(\mathbf{Z}) \end{pmatrix} \xrightarrow{d} \mathrm{N}_2\left(\mathbf{0}, \begin{pmatrix} 2 & 0 \\ 0 & \tilde{\sigma}^2 \end{pmatrix}\right) \quad as\ m \to \infty. \tag{7.7}$$

Proof of Lemma 7.1 From Lemma 2.1 of Schöne and Schmid (2000) we get the moment generating function of $(Q_1(\mathbf{Z}), Q_2(\mathbf{Z}))^\top$ expressed as

$$\begin{aligned}
M(t_1, t_2) &= \mathrm{E}\left[\exp\{t_1 Q_1(\mathbf{Z}) + t_2 Q_2(\mathbf{Z})\}\right] \\
&= \exp(-\sqrt{m}\,t_1)\left(1 - \frac{2t_1}{\sqrt{m}}\right)^{-m/2}\exp\left(\frac{t_2^2}{2}\frac{\mathbf{a}^\top\mathbf{a}/m}{1 - 2t_1/\sqrt{m}}\right).
\end{aligned}$$

Hence,

$$\lim_{m\to\infty} M(t_1, t_2) = \exp\left(\frac{\tilde{\sigma}^2 t_2^2}{2}\right)\lim_{m\to\infty}\exp(-\sqrt{n}\,t_1)\left(1 - \frac{2t_1}{\sqrt{n}}\right)^{-m/2} = \exp\left(\frac{\tilde{\sigma}^2 t_1^2}{2} + t_1^2\right),$$

where the last equality follows from the facts that the function under the limit is the moment generating function of $\sqrt{m}(\chi/m - 1)$ with $\chi \sim \chi_m^2$ and $\sqrt{m}(\chi/m - 1) \xrightarrow{d} \mathrm{N}(0, 2)$ as $m \to \infty$. The lemma is proved.

Proof of Theorem 2.2 Let $q_{ij} = p^{-\gamma}\boldsymbol{\mu}_i^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_j$, for $i, j \in \{0, n\}$ and $d = q_{00}q_{nn} - q_{0n}^2$. First, we prove the result in case of α_n^*. It holds that

$$\sqrt{p^\gamma n}(\alpha_n^* - \alpha^*) = \frac{1}{(cq_{00} + d)^2}\frac{cq_{00} + d}{q_{00}p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n - (p^{-\gamma}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n)^2}Z_\alpha,$$

where

$$\begin{aligned}
Z_\alpha &= \sqrt{p^\gamma n}\Big[(cq_{00} + d)(q_{00}p^{-\gamma}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n - q_{0n}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n) \\
&\quad - d\{q_{00}p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n - (p^{-\gamma}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n)^2\}\Big].
\end{aligned}$$

The application of (7.1) and (7.2) leads to

$$\begin{aligned}
Z_\alpha &= \sqrt{p^\gamma n}\Big((cq_{00} + d)\{q_{00}(q_{nn} + n^{-1}p^{-\gamma}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1/2}\mathbf{X}_n\mathbf{1}_n) \\
&\quad - q_{0n}(q_{0n} + n^{-1}p^{-\gamma}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1/2}\mathbf{X}_n\mathbf{1}_n)\} \\
&\quad - d[q_{00}\{q_{nn} + 2n^{-1}p^{-\gamma}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1/2}\mathbf{X}_n\mathbf{1}_n + n^{-2}p^{-\gamma}\mathbf{1}_n^\top(\mathbf{X}_n^\top\mathbf{X}_n)\mathbf{1}_n\} \\
&\quad - \{q_{0n}^2 + 2q_{0n}n^{-1}p^{-\gamma}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1/2}\mathbf{X}_n\mathbf{1}_n + n^{-2}p^{-2\gamma}(\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1/2}\mathbf{X}_n\mathbf{1}_n)^2\}]\Big) \\
&= \Big\{(cq_{00} - d)(q_{00}p^{-\gamma/2}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1/2} - q_{0n}p^{-\gamma/2}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1/2})(n^{-1/2}\mathbf{X}_n\mathbf{1}_n) \\
&\quad + dp^{-\gamma/2}n^{-1/2}\{p^{-\gamma/2}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1/2}(n^{-1/2}\mathbf{X}_n\mathbf{1}_n)\}^2 + \sqrt{c}dq_{00}\sqrt{p}\left(\frac{1}{p}\|n^{-1/2}\mathbf{X}_n\mathbf{1}_n\|^2 - 1\right)\Big\}.
\end{aligned}$$

From the proof of Theorem 2.1, we get that the elements of $n^{-1/2}\mathbf{X}_n\mathbf{1}_n$ are independent and asymptotically standard normally distributed. As a result, we get that

$$p^{-\gamma/2}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1/2}\left(n^{-1/2}\mathbf{X}_n\mathbf{1}_n\right) \xrightarrow{d} \mathrm{N}(0, q_{00}) \quad as\ p, n \to \infty$$

17

and, consequently,

$$\frac{1}{\sqrt{p^\gamma n}}\left\{p^{-\gamma/2}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1/2}(n^{-1/2}\mathbf{X}_n\mathbf{1}_n)\right\}^2 \xrightarrow{a.s.} 0 \quad\text{as}\quad p,n\to\infty\,.$$

Furthermore, the application of Lemma 7.1 leads to

$$Z_\alpha \xrightarrow{d} \mathrm{N}\left(0,\sigma_{Z_\alpha}^2\right) \quad\text{for}\quad n^{-1}p^{1-\gamma}\longrightarrow c>0 \text{ as } n\longrightarrow\infty$$

with

$$\sigma_{Z_\alpha}^2 \;=\; (cq_{00}-d)^2(q_{00}^2q_{nn}+q_{0n}^2q_{00}-2q_{00}q_{0n}^2)+cd^2q_{00}^2 = (cq_{00}-d)^2q_{00}d+cd^2q_{00}^2\,.$$

Finally, from the proof of Theorem 2.1 we get that

$$\frac{cq_{00}+d}{q_{00}p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n-(p^{-\gamma}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n)^2} \xrightarrow{a.s.} 1 \quad\text{for}\quad n^{-1}p^{1-\gamma}\longrightarrow c>0 \text{ as } n\longrightarrow\infty$$

which together with Slutsky's theorem (cf. Lehmann (1999)) implies

$$\sqrt{p^\gamma n}(\alpha_n^*-\alpha^*) \xrightarrow{d} \mathrm{N}\left(0,\frac{(cq_{00}-d)^2q_{00}d+cd^2q_{00}^2}{(cq_{00}+d)^4}\right) \quad\text{for}\quad n^{-1}p^{1-\gamma}\longrightarrow c>0 \text{ as } n\longrightarrow\infty\,.$$

Similarly, for β_n^* we get

$$\sqrt{p^\gamma n}(\beta_n^*-\beta^*) \;=\; \frac{1}{(cq_{00}+d)^2}\frac{cq_{00}+d}{q_{00}p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n-(p^{-\gamma}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n)^2}Z_\beta$$

with

$$\begin{aligned}
Z_\beta \;=\;& \sqrt{p^\gamma n}\left\{(cq_{00}+d)(p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_nq_{0n}-p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n)\right. \\
& \left. -cq_{0n}(p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_nq_{00}-(p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0)^2)\right\} \\
\;=\;& \sqrt{p^\gamma n}\{dq_{0n}p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n+cq_{0n}(p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0)^2-(cq_{00}+d)p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n\}\,.
\end{aligned}$$

Using (7.1) and (7.2) we get

$$\begin{aligned}
Z_\beta \;=\;& \sqrt{p^\gamma n}[dq_{0n}\{q_{nn}+2n^{-1}p^{-\gamma}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1/2}\mathbf{X}_n\mathbf{1}_n+n^{-2}p^{-\gamma}\mathbf{1}_n^\top(\mathbf{X}_n^\top\mathbf{X}_n)\mathbf{1}_n\} \\
& +cq_{0n}\{q_{0n}^2+2q_{0n}n^{-1}p^{-\gamma}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1/2}\mathbf{X}_n\mathbf{1}_n+n^{-2}p^{-2\gamma}(\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1/2}\mathbf{X}_n\mathbf{1}_n)^2\} \\
& -(cq_{00}+d)\{q_{0n}q_{nn}+q_{0n}n^{-1}p^{-\gamma}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1/2}\mathbf{X}_n\mathbf{1}_n \\
& +q_{nn}n^{-1}p^{-\gamma}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1/2}\mathbf{X}_n\mathbf{1}_n+n^{-2}p^{-2\gamma}(\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1/2}\mathbf{X}_n\mathbf{1}_n)(\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1/2}\mathbf{X}_n\mathbf{1}_n)\}] \\
\;=\;& \{(d-cq_{00})q_{0n}p^{-\gamma/2}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1/2}+(cq_{0n}^2-cd-dq_{nn})p^{-\gamma/2}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1/2}\}\left(n^{-1/2}\mathbf{X}_n\mathbf{1}_n\right) \\
& +\sqrt{c}dq_{0n}\sqrt{p}\left(\frac{1}{p}\|n^{-1/2}\mathbf{X}_n\mathbf{1}_n\|^2-1\right)+cq_{0n}n^{-2}p^{-2\gamma}(\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1/2}\mathbf{X}_n\mathbf{1}_n)^2 \\
& -(cq_{00}+d)n^{-2}p^{-2\gamma}(\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1/2}\mathbf{X}_n\mathbf{1}_n)(\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1/2}\mathbf{X}_n\mathbf{1}_n)\,.
\end{aligned}$$

Following the proof in case of α_n^*, we get

$$cq_{0n}\frac{p^{-2\gamma}}{n^2}(\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1/2}\mathbf{X}_n\mathbf{1}_n)^2-(cq_{00}+d)\frac{p^{-2\gamma}}{n^2}(\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1/2}\mathbf{X}_n\mathbf{1}_n)(\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1/2}\mathbf{X}_n\mathbf{1}_n) \xrightarrow{a.s.} 0$$

for $n^{-1}p^{1-\gamma} \longrightarrow c > 0$ as $n \longrightarrow \infty$. Then, the application of Lemma 7.1 leads to

$$Z_\beta \xrightarrow{d} \mathrm{N}\left(0, \sigma_{Z_\beta}^2\right) \quad \text{for } n^{-1}p^{1-\gamma} \longrightarrow c > 0 \text{ as } n \longrightarrow \infty$$

with

$$\sigma_{Z_\beta}^2 = (d - cq_{00})^2 q_{0n}^2 q_{nn} + (cq_{0n}^2 - cd - dq_{nn})^2 q_{00} + 2(cq_{0n}^2 - cd - dq_{nn})(d - cq_{00})q_{0n}^2 + cd^2 q_{0n}^2.$$

Finally, using Slutsky's theorem (cf. Lehmann (1999)) we get

$$\sqrt{p^\gamma n}(\beta_n^* - \beta^*) \xrightarrow{d} \mathrm{N}\left(0, \frac{\sigma_{Z_\beta}^2}{(cq_{00} + d)^4}\right) \quad \text{for } n^{-1}p^{1-\gamma} \longrightarrow c > 0 \text{ as } n \longrightarrow \infty.$$

The theorem is proved.

In the proof of Theorem 3.1 we make use of the following lemma.

Lemma 7.2. *Assume (A1)-(A2) and let the elements of $\mathbf{X}_n$ possess uniformly bounded $4 + \varepsilon$, $\varepsilon > 0$, moments. Let $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ be the universal nonrandom vectors from the set $\{\boldsymbol{\mu}_0, \boldsymbol{\mu}_n\}$. Then it holds that*

$$p^{-\gamma}\left|\boldsymbol{\xi}^\top \mathbf{S}_n^{-1}\boldsymbol{\theta} - (1-c)^{-1}\boldsymbol{\xi}^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\theta}\right| \xrightarrow{a.s.} 0, \tag{7.8}$$

$$\bar{\mathbf{x}}_n^\top \boldsymbol{\Sigma}_n^{1/2}\mathbf{S}_n^{-1}\boldsymbol{\Sigma}_n^{1/2}\bar{\mathbf{x}}_n \xrightarrow{a.s.} \frac{c}{1-c}, \tag{7.9}$$

$$p^{-\gamma/2}\bar{\mathbf{x}}_n^\top \boldsymbol{\Sigma}_n^{1/2}\mathbf{S}_n^{-1}\boldsymbol{\theta} \xrightarrow{a.s.} 0 \tag{7.10}$$

for $n^{-1}p \longrightarrow c \in (0, +\infty)$ as $n \to \infty$, where $\bar{\mathbf{x}}_n = n^{-1}\mathbf{X}_n\mathbf{1}_n$ stands for the sample mean vector calculated from $\mathbf{X}_n$.

Proof of Lemma 7.2: For $\boldsymbol{\theta} \in \{\boldsymbol{\mu}_0, \boldsymbol{\mu}_n\}$ it holds that

$$\left\|\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{\theta}\right\|^2 \leq \lambda_{max}(\boldsymbol{\Sigma}_n^{-1})\|\boldsymbol{\theta}\|^2 < \infty$$

following assumptions (A1) and (A2). The rest of the proof follows from Lemma 6.3 in Bodnar et al. (2014).

Proof of Theorem 3.1: For $\boldsymbol{\theta}, \boldsymbol{\xi} \in \{\boldsymbol{\mu}_0, \boldsymbol{\mu}_n\}$, the application of Lemma 7.2 leads to

$$p^{-\gamma}\left|\boldsymbol{\mu}_0^\top \mathbf{S}_n^{-1}\boldsymbol{\mu}_0 - (1-c)^{-1}\boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0\right| \xrightarrow{a.s.} 0,$$

$$p^{-\gamma}\left|\boldsymbol{\mu}_0^\top \mathbf{S}_n^{-1}\bar{\mathbf{y}}_n - (1-c)^{-1}\boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n\right|$$
$$\leq p^{-\gamma}\left|\boldsymbol{\mu}_0^\top \mathbf{S}_n^{-1}\bar{\mathbf{x}}_n\right| + p^{-\gamma}\left|\boldsymbol{\mu}_0^\top \mathbf{S}_n^{-1}\boldsymbol{\mu}_n - (1-c)^{-1}\boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n\right| \xrightarrow{a.s.} 0,$$

$$p^{-\gamma}\left|\bar{\mathbf{y}}_n^\top \mathbf{S}_n^{-1}\bar{\mathbf{y}}_n - (1-c)^{-1}\boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n - \frac{c}{1-c}\right|$$
$$\leq p^{-\gamma}\left|\boldsymbol{\mu}_n^\top \mathbf{S}_n^{-1}\boldsymbol{\mu}_n - (1-c)^{-1}\boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n\right| + p^{-\gamma}\left|\bar{\mathbf{x}}_n^\top \mathbf{S}_n^{-1}\bar{\mathbf{x}}_n - \frac{c}{1-c}\right| + 2p^{-\gamma}\left|\boldsymbol{\mu}_n^\top \mathbf{S}_n^{-1}\bar{\mathbf{x}}_n\right| \xrightarrow{a.s.} 0.$$

Hence,

$$\hat{\alpha}^* = 1 - \frac{(1-c)^{-1}cp^{-\gamma}\boldsymbol{\mu}_0^\top \mathbf{S}_n^{-1}\boldsymbol{\mu}_0}{(1-c)^{-1}c(1-p^{-\gamma})p^{-\gamma}\boldsymbol{\mu}_0^\top \mathbf{S}_n^{-1}\boldsymbol{\mu}_0 + p^{-\gamma}\bar{\mathbf{y}}_n^\top \mathbf{S}_n^{-1}\bar{\mathbf{y}}_n p^{-\gamma}\boldsymbol{\mu}_0^\top \mathbf{S}_n^{-1}\boldsymbol{\mu}_0 - (p^{-\gamma}\bar{\mathbf{y}}_n^\top \mathbf{S}_n^{-1}\boldsymbol{\mu}_0)^2}$$

$$\xrightarrow{a.s.} 1 - \frac{cp^\gamma \boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0}{c(p^\gamma - 1)\boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 + \left(\boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n + c\right)\boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 - (\boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0)^2} = \alpha^*$$

and, similarly,

$$\hat{\beta}^* = (1 - \hat{\alpha}^*) \frac{\bar{\mathbf{y}}_n^\top \mathbf{S}_n^{-1} \boldsymbol{\mu}_0}{\boldsymbol{\mu}_0^\top \mathbf{S}_n^{-1} \boldsymbol{\mu}_0} \xrightarrow{a.s.} (1 - \alpha^*) \frac{\boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0}{\boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0} = \beta^* .$$

for $n^{-1} p \longrightarrow c \in (0, +\infty)$ as $n \to \infty$. The theorem is proved.

The proof of Theorem 3.2 is based on Lemma 7.3. Let

$$\hat{s} = \bar{\mathbf{x}}_n^\top \mathbf{S}_n^{-1} \bar{\mathbf{x}}_n - \frac{(\boldsymbol{\mu}_0^\top \mathbf{S}_n^{-1} \bar{\mathbf{x}}_n)^2}{\boldsymbol{\mu}_0^\top \mathbf{S}_n^{-1} \boldsymbol{\mu}_0} , \qquad\qquad \hat{R} = \frac{\boldsymbol{\mu}_0^\top \mathbf{S}_n^{-1} \bar{\mathbf{x}}_n}{\boldsymbol{\mu}_0^\top \mathbf{S}_n^{-1} \bar{\mathbf{x}}_n} .$$

Lemma 7.3. *Let* $\mathbf{y}_1, \ldots, \mathbf{y}_n$ *be a random sample of independent vectors such that* $\mathbf{y}_i \sim N_p(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$ *for* $i = 1, \ldots, n$. *Then for any* p *and* n *with* $n > p$ *it holds that*

(a) $\frac{n(n-p+1)}{(n-1)(p-1)} \hat{s} \sim F_{p-1, n-p+1, n\,s}$ *(non-central F-distribution with* $p - 1$ *and* $n - p + 1$ *degrees of freedom and non-centrality parameter* s*).*

(b) $\hat{R} | \hat{s} = y \sim N \left(R, \frac{1 + \frac{n}{n-1} y}{n} \frac{1}{\boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0} \right).$

Proof of Lemma 7.3: The results are obtained following the proof of Theorem 3.1 in Bodnar and Schmid (2008), where similar statements are presented in case $\boldsymbol{\mu}_0 = \mathbf{1}_p$.

Proof of Theorem 3.2: Let

$$\hat{s}_c = (1 - c)\hat{s} - c .$$

Then it holds that

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

The application of Lemma 7.3.a and Lemma 5.3.b in Bodnar et al. (2014) leads to

$$\sqrt{n} \left(\hat{s}_c - s \right) \xrightarrow{d} N \left(0, \sigma_s^2 \right) \quad \text{with} \quad \sigma_s^2 = 2 \left(c + 2s \right) + \frac{2}{1 - c} \left(c + s \right)^2 . \tag{7.11}$$

From Theorem 3.1, we get

$$\hat{\alpha}^* = 1 - \frac{c}{1 - c} \frac{1}{\hat{s}} = 1 - \frac{c}{c + \hat{s}_c} ,$$

and

$$\hat{\beta}^* = \frac{c}{c + \hat{s}_c} \hat{R} \xrightarrow{d} \frac{c}{c + \hat{s}_c} \left(R + \frac{\sqrt{1 + \frac{n}{n-1} \frac{\hat{s}_c + c}{1 - c}}}{\sqrt{\boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0}} z_0 \right) ,$$

where

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

with σ_s^2 as in (7.11). Now, the application of the δ-method (see, e. g., Theorem 3.7 in DasGupta (2008)) leads to

$$\sqrt{n} \left(\begin{array}{c} \hat{\alpha}^* - \alpha^* \\ \hat{\beta}^* - \beta^* \end{array} \right) \xrightarrow{d} N \left(\left(\begin{array}{c} 0 \\ 0 \end{array} \right), \boldsymbol{\Omega} \right)$$

with

$$\boldsymbol{\Omega} = \begin{pmatrix} \frac{c}{(c+s)^2} & 0 \\ \frac{c}{(c+s)^2}R & \frac{c}{c+s}\sqrt{\frac{1+\frac{s+c}{1-c}}{\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0}} \end{pmatrix} \begin{pmatrix} \sigma_s^2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{c}{(c+s)^2} & \frac{c}{(c+s)^2}R \\ 0 & \frac{c}{c+s}\sqrt{\frac{1+\frac{s+c}{1-c}}{\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0}} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{c^2\sigma_s^2}{(c+s)^4} & \frac{c^2\sigma_s^2}{(c+s)^4}R \\ \frac{c^2\sigma_s^2}{(c+s)^4}R & \frac{c^2\sigma_s^2}{(c+s)^4}R^2 + \frac{c^2}{(c+s)^2}\frac{1+\frac{s+c}{1-c}}{\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0} \end{pmatrix}.$$

The theorem is proved.

In the proof of Theorem 3.2 we make use of the following lemma.

Lemma 7.4. *Assume (A1)-(A2) and let the elements of $\mathbf{X}_n$ possess uniformly bounded $4+\varepsilon$, $\varepsilon > 0$, moments. Let $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ be the universal nonrandom vectors from the set $\{\boldsymbol{\mu}_0, \boldsymbol{\mu}_n\}$. Then it holds that*

$$p^{-\gamma}\left|\boldsymbol{\xi}^\top\mathbf{S}_n^-\boldsymbol{\theta} - c^{-1}(c-1)^{-1}\boldsymbol{\xi}^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\theta}\right| \xrightarrow{a.s.} 0, \tag{7.12}$$

$$\bar{\mathbf{x}}_n^\top\boldsymbol{\Sigma}_n^{1/2}\mathbf{S}_n^-\boldsymbol{\Sigma}_n^{1/2}\bar{\mathbf{x}}_n \xrightarrow{a.s.} \frac{1}{c-1}, \tag{7.13}$$

$$p^{-\gamma/2}\bar{\mathbf{x}}_n^\top\boldsymbol{\Sigma}_n^{1/2}\mathbf{S}_n^-\boldsymbol{\theta} \xrightarrow{a.s.} 0 \tag{7.14}$$

for $n^{-1}p^{-\gamma} \longrightarrow c \in (1, +\infty)$ as $n \to \infty$, where $\bar{\mathbf{x}}_n = n^{-1}\mathbf{X}_n\mathbf{1}_n$ stands for the sample mean vector calculated from $\mathbf{X}_n$.

Proof of Lemma 7.4: For $\boldsymbol{\theta} \in \{\boldsymbol{\mu}_0, \boldsymbol{\mu}_n\}$ it holds that

$$\left\|\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{\theta}\right\|^2 \leq \lambda_{max}(\boldsymbol{\Sigma}_n^{-1})\|\boldsymbol{\theta}\|^2 < \infty$$

following assumptions (A1) and (A2). The rest of the proof follows from Lemma 5.6 in Bodnar et al. (2016b).

Proof of Theorem 3.3: For $\boldsymbol{\theta}, \boldsymbol{\xi} \in \{\boldsymbol{\mu}_0, \boldsymbol{\mu}_n\}$, the application of Lemma 7.4 leads to

$$p^{-\gamma}\left|\boldsymbol{\mu}_0^\top\mathbf{S}_n^-\boldsymbol{\mu}_0 - c^{-1}(c-1)^{-1}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0\right| \xrightarrow{a.s.} 0,$$

$$p^{-\gamma}\left|\boldsymbol{\mu}_0^\top\mathbf{S}_n^-\bar{\mathbf{y}}_n - c^{-1}(c-1)^{-1}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n\right|$$

$$\leq p^{-\gamma}\left|\boldsymbol{\mu}_0^\top\mathbf{S}_n^-\bar{\mathbf{x}}_n\right| + p^{-\gamma}\left|\boldsymbol{\mu}_0^\top\mathbf{S}_n^-\boldsymbol{\mu}_n - c^{-1}(c-1)^{-1}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n\right| \xrightarrow{a.s.} 0,$$

$$p^{-\gamma}\left|\bar{\mathbf{y}}_n^\top\mathbf{S}_n^-\bar{\mathbf{y}}_n - c^{-1}(c-1)^{-1}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n - \frac{1}{c-1}\right|$$

$$\leq p^{-\gamma}\left|\boldsymbol{\mu}_n^\top\mathbf{S}_n^-\boldsymbol{\mu}_n - c^{-1}(c-1)^{-1}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n\right| + p^{-\gamma}\left|\bar{\mathbf{x}}_n^\top\mathbf{S}_n^-\bar{\mathbf{x}}_n - \frac{1}{c-1}\right| + 2p^{-\gamma}\left|\boldsymbol{\mu}_n^\top\mathbf{S}_n^-\bar{\mathbf{x}}_n\right| \xrightarrow{a.s.} 0.$$

Hence,

$$\hat{\alpha}^* = 1 - \frac{(c-1)^{-1}p^{-\gamma}\boldsymbol{\mu}_0^\top\mathbf{S}_n^-\boldsymbol{\mu}_0}{(c-1)^{-1}(1-p^{-\gamma})p^{-\gamma}\boldsymbol{\mu}_0^\top\mathbf{S}_n^-\boldsymbol{\mu}_0 + p^{-\gamma}\bar{\mathbf{y}}_n^\top\mathbf{S}_n^-\bar{\mathbf{y}}_n p^{-\gamma}\boldsymbol{\mu}_0^\top\mathbf{S}_n^-\boldsymbol{\mu}_0 - (p^{-\gamma}\bar{\mathbf{y}}_n^\top\mathbf{S}_n^-\boldsymbol{\mu}_0)^2}$$

$$\xrightarrow{a.s.} 1 - \frac{cp^\gamma\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0}{c(p^\gamma - 1)\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 + \left(\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n + c\right)\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 - (\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0)^2} = \alpha^*$$

and, similarly,

$$\hat{\beta}^* = (1 - \hat{\alpha}^*)\frac{\bar{\mathbf{y}}_n^\top \mathbf{S}_n^- \boldsymbol{\mu}_0}{\boldsymbol{\mu}_0^\top \mathbf{S}_n^- \boldsymbol{\mu}_0} \xrightarrow{a.s.} (1 - \alpha^*)\frac{\boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0}{\boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0} = \beta^* \ .$$

for $n^{-1}p^{1-\gamma} \longrightarrow c \in (1, +\infty)$ as $n \to \infty$. The theorem is proved.

References

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. Econometrica, 70:191–221.

Bai, J. and Ng, S. (2008). Large dimensional factor analysis. Foundations and Trends in Econometrics, 3:89–163.

Bai, Z. D. and Silverstein, J. W. (2004). CLT for linear spectral statistics of large dimensional sample covariance matrices. Annals of Probability, 32:553–605.

Bai, Z. D. and Silverstein, J. W. (2010). Spectral Analysis of Large Dimensional Random Matrices. Springer, New York.

Baranchik, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. The Annals of Mathematical Statistics, 41:642–645.

Berger, J., Bock, M. E., Brown, L. D., Casella, G., and Gleser, L. (1977). Minimax estimation of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. The Annals of Statistics, 5:763–771.

Berger, J. O. and Bock, M. E. (1976). Combining independent normal mean estimation problems with unknown variances. The Annals of Statistics, 4:642–648.

Bodnar, T., Dette, H., and Parolya, N. (2016a). Spectral analysis of the moore–penrose inverse of a large dimensional sample covariance matrix. Journal of Multivariate Analysis, 148:160–172.

Bodnar, T., Okhrin, Y., and Parolya, N. (2016b). Optimal shrinkage-based portfolio selection in high dimensions. Working paper.

Bodnar, T., Parolya, N., and Schmid, W. (2014). Estimation of the global minimum variance portfolio in high dimensions. Working paper, arxiv:1406.0437.

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

Bodnar, T. and Schmid, W. (2008). Estimation of optimal portfolio compositions for gaussian returns. Statistics & Risk Modeling, 26:179 – 201.

Cai, T. T., Zhang, C.-H., and Zhou, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. The Annals of Statistics, 38:2118–2144.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. Econometrica, 51:1281–1304.

Chételat, D. and Wells, M. T. (2012). Improved multivariate normal mean estimation with unknown covariance when p is greater than n. The Annals of Statistics, 40:3137–3160.

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

DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? Review of Financial Studies, 22(5):1915–1953.

Fan, J., Fan, Y., and Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. Journal of Econometrics, 147:186–197.

Fan, J., Han, F., and Lui, H. (2014). Challenges of big data analysis. National Science Review, 1:293–314.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. Journal of the Royal Statistical Society: Series B, 75:603–680.

Fan, J., Zhang, J., and Yu, K. (2012). Vast portfolio selection with gross-exposure constraints. Journal of American Statistical Association, 107:592–606.

Fourdrinier, D., Strawderman, W. E., and Wells, M. T. (2003). Robust shrinkage estimation for elliptically symmetric distributions with unknown covariance matrix. Journal of Multivariate Analysis, 85:24 – 39.

Friesen, O., Löwe, M., and Stolz, M. (2013). Gaussian fluctuations for sample covariance matrices with dependent data. Journal of Multivariate Analysis, 114:270–287.

Gleser, L. J. (1986). Minimax estimators of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. The Annals of Statistics, 14:1625–1633.

Issa, E. B., Papanastassiou, A. M., and DiCarlo, J. J. (2013). Large-scale, high-resolution neurophysiological maps underlying fmri of macaque temporal lobe. The Journal of Neuroscience, 33:15207–15219.

James, W. and Stein, C. (1961). Estimation with quadratic loss. Proc. Fourth Berkeley Symp. on Math. Statist. and Prob., 1:361–379.

Lehmann, E. L. (1999). Elements of Large-Sample Theory. Springer, New York.

Lin, P.-E. and Tsai, H.-L. (1973). Generalized bayes minimax estimators of the multivariate normal mean with unknown covariance matrix. The Annals of Statistics, 1:142–145.

Rubio, F., Mestre, X., and Palomar, D. P. (2012). Performance analysis and optimal selection of large minimum variance portfolios under estimation risk. J. Sel. Topics Signal Processing, pages 337–350.

Schöne, A. and Schmid, W. (2000). On the joint distribution of a quadratic and a linear form in normal variables. Journal of Multivariate Analysis, 72:163 – 182.

Shalon, D., Smith, S. J., and Brown, P. O. (1996). A dna microarray system for analyzing complex dna samples using two-color fluorescent probe hybridization. Genome Research, 6:639–645.

Tsay, R. S. (2005). Analysis of Financial Time Series. Wiley series in probability and statistics. Wiley-Interscience, Hoboken (N.J.).

Tu, J. and Zhou, G. (2011). Markowitz meets talmud: A combination of sophisticated and naive diversification strategies. Journal of Financial Economics, 99(1):204–215.

Wang, C., Tong, T., Cao, L., and Miao, B. (2014). Non-parametric shrinkage mean estimation for quadratic loss functions with unknown covariance matrices. Journal of Multivariate Analysis, 125:222 – 232.