Accepted Manuscript

Optimal shrinkage estimator for high-dimensional mean vector

Taras Bodnar, Ostap Okhrin, Nestor Parolya

PII: S0047-259X(17)30603-6
DOI: https://doi.org/10.1016/j.jmva.2018.07.004
Reference: YJMVA 4383

To appear in: *Journal of Multivariate Analysis*

Received date : 9 October 2017

Please cite this article as: T. Bodnar, O. Okhrin, N. Parolya, Optimal shrinkage estimator for high-dimensional mean vector, *Journal of Multivariate Analysis* (2018), https://doi.org/10.1016/j.jmva.2018.07.004

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Optimal shrinkage estimator for high-dimensional mean vector

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

[a]*Department of Mathematics, Stockholm University, SE-10691 Stockholm, Sweden*
[b]*Chair of Econometrics and Statistics esp. Transportation, Technische Universität Dresden, 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 high-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 in such a way that $p/n \to c \in (0, \infty)$. Under weak conditions imposed on the underlying data generating mechanism, we find the asymptotic equivalents to the optimal shrinkage intensities and estimate them consistently. The proposed nonparametric estimator for the high-dimensional mean vector has a simple structure and is proven to minimize asymptotically, with probability 1, the quadratic loss when $c \in (0, 1)$. When $c \in (1, \infty)$ we modify the estimator by using a feasible estimator for the precision covariance matrix. To this 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. It turns out that the existing estimators of the mean vector, including the new proposal, converge to the sample mean vector when the true mean vector has an unbounded Euclidean norm.

Keywords: Large-dimensional asymptotics, Mean vector estimation, Random matrix theory, Shrinkage estimator
2010 MSC: 60B20, 62H12, 62G20, 62G30

1. Introduction

Functional data analysis features prominently in modern statistical theory; see [15, 19, 27, 32, 37, 38]. As emphasized in [26, 31], functional data can often be regarded as high-dimensional vectors. This point of view has been explored, e.g., in [1, 2, 3, 44]. High-dimensional problems generate a great deal of interest on their own, thanks to the availability of high performance and high storage computers. These advances invite the collection of large amounts of information and stimulate the development of new methods, since classical ones may not be adequate because of specific features of Big Data such as noise accumulation, spurious correlation, heterogeneity and others; see, e.g., [7]. For a review of challenges in modeling Big Data, see [22].

Vast amounts of information are used in various fields, e.g., genomics with hundreds of thousands of microarrays [41], neurosciences with high-precision fMRI data of only dozens of persons [33], financial time series [42], etc. In these contexts, one must often deal with the problem of estimating a high-dimensional mean vector when the sample size is much smaller than the dimension. The usual estimator of location under quadratic loss, namely the sample mean vector, has been known for decades to be unsatisfactory because convergence under the quadratic loss function fails when p increases with n, even when $p \leq n$. An improved estimator was proposed early on by James and Stein [34] for p-dimensional normal random vectors with identity covariance matrix when $p > 2$. Baranchik [8] then extended this estimator to the case of a diagonal covariance matrix with equal variances. Further research reported in [9, 10, 28, 30, 36] led to estimators that can be used for arbitrary, unknown covariance matrix for $n \geq p \geq 3$.

More recently, high-dimensional version of the James–Stein type estimator was proposed by Chételat and Wells [18] for $p > n \geq 3$ using an unbiased estimator of the risk difference. They suggested the so-called positive-part type James–Stein estimator and showed through simulations that it dominates the high-dimensional James–Stein

*Corresponding author
Email address: taras.bodnar@math.su.se (Taras Bodnar)

ACCEPTED MANUSCRIPT

estimator under invariant loss. Wang et al. [45] considered an optimal shrinkage estimator towards the unit vector by minimizing expected quadratic loss. However, their estimator is computationally demanding in large dimensions because it involves nontrivial sums. They then suggested that a limiting form of their estimator be used in practice.

This paper contributes to the literature on location parameter estimation by deriving the optimal shrinkage estimator (towards any fixed target) using random matrix theory. The new estimator relies on weaker assumptions than commonly used. We prove its asymptotic normality, derive its limit behavior and show through simulations that it dominates the benchmark methods in terms of both quadratic loss function and computational time.

This paper is structured as follows. The optimal shrinkage estimator for the mean vector is introduced in Section 2, where it is shown to be asymptotically equivalent to a nonrandom quantity and Gaussian in the limit. In Section 3 we provide a bona fide estimator and investigate its asymptotic behavior. Section 4 discusses benchmark procedures used in the simulation study whose results are reported in Section 5. A financial application is presented in Section 6 and conclusions can be found in Section 7. The technical derivations are relegated 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 large-dimensional asymptotics. Let $\mathbf{Y}_n$ be a $p \times n$ matrix of p-dimensional random vectors observed 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 asymptotic paradigm, both the dimension p and the sample size n tend to infinity in such a way that $p/n \to c \in (0, \infty)$ as $n \to \infty$. For this reason, it is natural to assume, without loss of generality, that $p \equiv p(n)$ is a function of n. Later on, we assume that the observation matrix is equal in distribution to

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

where the $p \times n$ matrix $\mathbf{X}_n$ contains iid random variables with zero mean and unit variance, while $\mathbf{1}_n$ is the n-dimensional vector of 1s. 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 independence imposed on its columns can be weakened further 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 [29]. Thus the independence assumption is only used below for technical convenience.

Next, we present the main assumptions which are used in this 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 exist constants $\gamma > 0$, $M_\ell > 0$, $M_u > 0$ such that $\lim_{p\to\infty} p^{-\gamma} \|\boldsymbol{\mu}_n\|^2 = M_\ell$ and $\lim_{p\to\infty} p^{-\gamma} \|\boldsymbol{\mu}_0\|^2 = M_0$ with $0 < M_\ell \leq M_0, M_\ell \leq M_u < \infty$. Here, $\boldsymbol{\mu}_0$ denotes the target mean vector.

(A3) The elements of the matrix $\mathbf{X}_n$ have uniformly bounded $2 + \varepsilon$ moments for some $\varepsilon > 0$.

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 our results to high-dimensional factor models, which are very popular in economics and finance; see, e.g., [4, 5, 12, 17, 23, 24, 25]. 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; see [16]. In this paper we assume that all eigenvalues are strictly positive. The increase in the norms of the unknown mean vector $\boldsymbol{\mu}_n$ and of the target vector $\boldsymbol{\mu}_0$ are monitored by Assumption (A2), which only requires that they are of the same order. Finally, Assumption (A3) is a technical one and can be relaxed in some cases; see [39].

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

$$\hat{\boldsymbol{\mu}}_{GSE} = \alpha_n \bar{\mathbf{y}}_n + \beta_n \boldsymbol{\mu}_0, \tag{2}$$

ACCEPTED MANUSCRIPT

where $\bar{\mathbf{y}}_n = n^{-1}\mathbf{Y}_n\mathbf{1}_n$ stands for the sample mean vector. The target vector $\boldsymbol{\mu}_0$ can also be random but independent of the actual information set $\mathbf{Y}_n$. The aim is to find the optimal shrinkage intensities which minimize the quadratic loss for a given target vector $\boldsymbol{\mu}_0$ expressed as

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

The application of (2) leads to the following optimization problem:

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

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

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

The Hessian of L is given by

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

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

$$\det(\mathbf{H}) = \bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n \boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 - (\bar{\mathbf{y}}_n^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0)^2 > 0 \tag{3}$$

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

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

In Theorem 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 $p/n \to c \in (0, \infty)$ as $n \to \infty$.

Theorem 1. *Assume that (A1)–(A3) hold. Then* $|\alpha_n^* - \alpha^*| \xrightarrow{\text{a.s.}} 0$ *and* $|\beta_n^* - \beta^*| \xrightarrow{\text{a.s.}} 0$ *for* $p/n \to c > 0$ *as* $n \to \infty$*, where*

$$\alpha^* = \alpha^*(\boldsymbol{\Sigma}_n, \boldsymbol{\mu}_n, \boldsymbol{\mu}_0) = \frac{\boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n \boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 - (\boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0)^2}{(c + \boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n)\boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 - (\boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0)^2}, \tag{4}$$

$$\beta^* = \beta^*(\boldsymbol{\Sigma}_n, \boldsymbol{\mu}_n, \boldsymbol{\mu}_0) = (1 - \alpha^*)\frac{\boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0}{\boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0}. \tag{5}$$

Note that $\alpha^* \in (0, 1)$ due to inequality (3). Furthermore, using the results of Theorem 1, we are able to estimate α^* and β^* consistently at least for $c \in (0, 1)$, which is shown in Theorem 3 below.

It is remarkable that the proposed procedure is very different to the one suggested by Wang et al. [45], who minimized the expected quadratic loss and estimated optimal shrinkage intensities. While they found the estimators for the optimal shrinkage intensities which converge in probability, our aim is to construct consistent estimators which converge almost surely. It is worth pointing out the following remark.

Remark 1. The technical assumption (A2) with the same γs is key. If this condition fails, i.e., there exist γ_1 and γ_2 such that $\lim_{p\to\infty} p^{-\gamma_1}\|\boldsymbol{\mu}_n\|^2 = M_\ell$ and $\lim_{p\to\infty} p^{-\gamma_2}\|\boldsymbol{\mu}_0\|^2 = M_0$ with $0 < M_\ell \leq M_0, M_\ell \leq M_u < \infty$, 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

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

ACCEPTED MANUSCRIPT

Let $\tilde{c} = p^{-\gamma}c$, $q_{ij} = p^{-\gamma}\boldsymbol{\mu}_i^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_j$, for $i, j \in \{0, n\}$ and $d = q_{00}q_{nn} - q_{0n}^2$. In Theorem 2, we prove that α_n^* and β_n^* are asymptotically Gaussian under the large-dimensional asymptotic regime. It must be noted that the normality assumption is used only in Theorems 2 and 4, whereas the existence of the second (fourth) moments is only required for the rest of our results.

Theorem 2. *Assume (A1)–(A2) and let the elements of* $\mathbf{X}_n$ *be standard normally distributed. Then*

$$\sqrt{p^\gamma n}\,\sigma_\alpha^{-1}\,(\alpha_n^* - \alpha^*) \rightsquigarrow \mathcal{N}(0,1), \quad \sqrt{p^\gamma n}\,\sigma_\beta^{-1}\,(\beta_n^* - \beta^*) \rightsquigarrow \mathcal{N}(0,1),$$

for $p/n \to c > 0$ *as* $n \to \infty$, *where* $\sigma_\alpha^2 = \{(\tilde{c}q_{00} - d)^2 q_{00} d + \tilde{c}d^2 q_{00}^2\}(\tilde{c}q_{00} + d)^4$ *and*

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

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 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.

Let

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

be the sample covariance matrix. In Theorem 3 below we present consistent estimators for α^* and β^* under large-dimensional asymptotics.

Theorem 3. *Assume (A1)–(A2) and let the elements of* $\mathbf{X}_n$ *possess uniformly bounded* $4+\varepsilon$ *moments with* $\varepsilon > 0$. *Then the consistent estimators for* α^* *and* β^* *are given by*

$$\hat{\alpha}^* = \hat{\alpha}^*(\bar{\mathbf{y}}_n, \mathbf{S}_n, \boldsymbol{\mu}_0) = \frac{\{\bar{\mathbf{y}}_n^\top \mathbf{S}_n^{-1}\bar{\mathbf{y}}_n - p/(n-p)\}\boldsymbol{\mu}_0^\top \mathbf{S}_n^{-1}\boldsymbol{\mu}_0 - (\bar{\mathbf{y}}_n^\top \mathbf{S}_n^{-1}\boldsymbol{\mu}_0)^2}{\bar{\mathbf{y}}_n^\top \mathbf{S}_n^{-1}\bar{\mathbf{y}}_n \boldsymbol{\mu}_0^\top \mathbf{S}_n^{-1}\boldsymbol{\mu}_0 - (\bar{\mathbf{y}}_n^\top \mathbf{S}_n^{-1}\boldsymbol{\mu}_0)^2}, \tag{6}$$

$$\hat{\beta}^* = \hat{\beta}^*(\bar{\mathbf{y}}_n, \mathbf{S}_n, \boldsymbol{\mu}_0) = (1 - \hat{\alpha}^*)\,\frac{\bar{\mathbf{y}}_n^\top \mathbf{S}_n^{-1}\boldsymbol{\mu}_0}{\boldsymbol{\mu}_0^\top \mathbf{S}_n^{-1}\boldsymbol{\mu}_0}. \tag{7}$$

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.

Theorem 4. *Assume (A1)–(A2) and let the elements of* $\mathbf{X}_n$ *be standard normally distributed. Then*

$$\sqrt{n}\,\boldsymbol{\Omega}^{-1/2}\begin{pmatrix} \hat{\alpha}^* - \alpha^* \\ \hat{\beta}^* - \beta^* \end{pmatrix} \rightsquigarrow \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{I}\right]$$

where

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

and

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

The results of Theorem 4 can be used to construct an asymptotic test on both α^* and β^*. For example, testing the null hypothesis $\mathcal{H}_0 : \alpha^* = 1$ will imply that in terms of the quadratic loss, the best estimator for the mean vector is the sample mean vector because $\beta^* = 0$ as soon as $\alpha^* = 1$.

ACCEPTED MANUSCRIPT

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

$$\hat{\boldsymbol{\mu}}_{OLSE} = \hat{\alpha}^* \bar{\mathbf{y}}_n + \hat{\beta}^* \boldsymbol{\mu}_0, \tag{8}$$

where the optimal shrinkage intensities are given by (6) and (7), respectively. This estimator has 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 (8) dominates the sample estimator in terms of minimum quadratic loss uniformly if both p and n tend to infinity and $p/n \to c < 1$.

For $c > 1$ the sample covariance matrix is no longer invertible and we need other techniques to estimate the unknown quantities given in (4) and (5). 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 $\mathbf{S}_n$:

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

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 $(\mathbf{S}_n \mathbf{S}_n^-)^\top = \mathbf{S}_n \mathbf{S}_n^-$. When $c < 1$, 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; it is left for future research.

In Theorem 5 below we present the consistent estimators for α^* and β^* under large-dimensional asymptotics in the case of $c > 1$ utilizing the generalized inverse $\mathbf{S}_n^-$.

Theorem 5. *Assume (A1)–(A2) and let the elements of* $\mathbf{X}_n$ *possess uniformly bounded* $4 + \varepsilon$ *moments with* $\varepsilon > 0$. *Let* $p/n \to c \in (1, \infty)$ *for* $n \to \infty$. *Then consistent estimators for* α^* *and* β^* *are given by*

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

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

Because $\mathbf{S}_n^-$ depends on 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 [11]. It is worth mentioning that changing $\mathbf{S}^-$ to $\mathbf{S}^+$ does not in general lead to the optimal shrinkage estimator. This is in contrast to the case $c < 1$, where it obviously holds that $\mathbf{S}^{-1} = \mathbf{S}^- = \mathbf{S}^+$. Consequently, the approach suggested for $p > n$ is only suboptimal, but nevertheless it dominates in most cases the existent estimators for the high-dimensional mean vector given in the literature. This fact is justified in the next sections via a simulation study and an empirical illustration.

3.1. Choice of $\boldsymbol{\mu}_0$

An important issue is the choice of the nonrandom target vector $\boldsymbol{\mu}_0$ which should satisfy Assumption (A2). This depends on the underlying data because the choice of the target vector is equivalent to the choice of the hyperparameter for the prior distribution of $\boldsymbol{\mu}_n$. This problem is well-known in Bayesian statistics. Different priors lead to different results. So it is crucial to choose the one which works satisfactory for most cases. The naive choice is $\boldsymbol{\mu}_0 = p^{(\gamma-1)/2}\mathbf{1}$, where $\mathbf{1}$ is the p-dimensional vector of 1s. Obviously, the perfect $\boldsymbol{\mu}_0$ is the true mean vector $\boldsymbol{\mu}_n$; we then get from Theorem 1 that $\alpha^* = 0$ and $\beta^* = 1$. Hence, the proposed optimal shrinkage estimator will tend to the true mean almost surely as both p and n tend to infinity; setting $\boldsymbol{\mu}_0$ close to $\boldsymbol{\mu}_n$ will improve the resulting shrinkage estimator.

Another quantity present in the naive target $\boldsymbol{\mu}_0 = p^{(\gamma-1)/2}\mathbf{1}$ is γ, which measures how quickly the Euclidean norm of the true mean vector goes to infinity. As can be seen from the simulation study, if $\gamma \neq 0$, then all of the considered estimators including our proposal converge to the sample mean vector.

ACCEPTED MANUSCRIPT

4. Benchmark methods

This section introduces approaches used as benchmarks in the simulation study of Section 5. The most commonly used estimator for the mean vector in the literature is the sample mean vector expressed as $\bar{\mathbf{y}}_n = \mathbf{Y}_n\mathbf{1}_n/n$. Although this estimator is known to be not converging under quadratic loss for high-dimensional data, we nevertheless use the sample mean in our comparison study.

The original James–Stein estimator was expressed in the form

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

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

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

for $c < 1$ and an estimator $\widetilde{\mathbf{S}}_n \sim \mathcal{W}_p(n, \boldsymbol{\Sigma}_n)$ of the covariance matrix $\boldsymbol{\Sigma}_n$ with $n \geq p \geq 3$. When $p > n \geq 3$, we compare our estimator with those proposed by Chételat and Wells [18], who defined a Baranchik type estimator as

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

with $0 \leq a \leq 2(n-2)/(p-n+3)$ and $\widetilde{\mathbf{S}}_n^+$ the Moore–Penrose inverse of $\widetilde{\mathbf{S}}_n$. In our study we set $a = 2(n-2)/(p-n+3)$. As shown by the simulation study in [18], the so-called positive-part type James–Stein estimator of the form

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

with $b_+ = \max(b, 0)$ dominates $\widehat{\boldsymbol{\mu}}_{n,JS(p>n)}$ under the invariant loss.

Another benchmark estimator is taken from [45]. It is a shrinkage estimator with unit target vector and shrinkage coefficients found by the minimization of the expected quadratic loss. This shrinkage estimator is given by

$$\widehat{\boldsymbol{\mu}}_{n,W} = \frac{Z_{1,n} - Z_{4,n}}{Z_{1,n} + Z_{2,n}Z_{4,n}}\bar{\mathbf{y}}_n + \frac{Z_{2,n}}{Z_{1,n} + Z_{2,n}Z_{4,n}}Z_{3,n}\mathbf{1}_n, \tag{9}$$

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 $p/n > 1$. The estimator (9) 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. That is why, in practice, its asymptotic counterpart is considered; see, e.g., [45].

5. Finite-sample performance

This section provides an extensive simulation study, to test the validity of Theorems 1, 2, 4, and 5 as well as to compare the quality of the proposed OLSE estimator with the considered benchmark methods. As discussed in Section 3.1, γ controls the speed of divergence of the Euclidean norms of $\boldsymbol{\mu}_n$ and $\boldsymbol{\mu}_0$ to infinity. Therefore, we consider two extreme scenarios, namely $\gamma = 0$ and $\gamma = 1$.

ACCEPTED MANUSCRIPT

When $\gamma = 0$, the true mean vector $\boldsymbol{\mu}_n$ and the target mean vector $\boldsymbol{\mu}_0$ have been independently simulated from the uniform distribution on $[-p^{-1/2}, p^{-1/2}]$. This scenario keeps the norms of both $\boldsymbol{\mu}_n$ and $\boldsymbol{\mu}_0$ bounded. In contrast, only very little prior information about $\boldsymbol{\mu}_n$ is used in the selection of $\boldsymbol{\mu}_0$. This choice of $\boldsymbol{\mu}_n$ is also 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.

When $\gamma = 1$, the true mean vector $\boldsymbol{\mu}_n$ contains randomly interchanging values 1 and -1, while the $\boldsymbol{\mu}_0$ is set to the unit vector. This scenario forces the norms of both mean vectors to explode with the dimension p and it uses no information about the true mean vector $\boldsymbol{\mu}_n$ when $\boldsymbol{\mu}_0$ is chosen.

For all of the considered setups we also construct a bona fide shrinkage estimator with $\boldsymbol{\mu}_0 = \boldsymbol{\mu}_n$ which corresponds to the case when the optimal shrinkage estimator converges to $\boldsymbol{\mu}_n$ almost surely; see Section 3.1. The corresponding results are depicted in all figures as the solid red line. The eigenvalues of the true covariance matrix $\boldsymbol{\Sigma}_n$ are chosen in the following way: 20% of all eigenvalues are equal to 1, 40% of them equal 3 and the rest 40% are 10. The eigenvectors are extracted via the QR decomposition of a standard normally distributed random matrix. The obtained eigenvalues and eigenvectors are finally wrapped via the singular value decomposition into a covariance matrix $\boldsymbol{\Sigma}_n$. We further note that the quantities $\boldsymbol{\mu}_n$, $\boldsymbol{\mu}_0$, and $\boldsymbol{\Sigma}_n$ are fixed in each simulation study, i.e., the same values of $\boldsymbol{\mu}_n$, $\boldsymbol{\mu}_0$, and $\boldsymbol{\Sigma}_n$ are used to generate samples which follow model (1).

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 $\hat{\beta}^*$. We compared them to the corresponding asymptotic distributions presented in Theorems 2 and 4. 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, considered only the case $\gamma = 0$, and drew the columns of $\mathbf{Y}_n$ from $\mathcal{N}(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$ in each simulation run. Then, α_n^* and β_n^* were calculated and standardized by the corresponding asymptotic variances from Theorem 2 for $c \in \{0.5, 0.9, 2.0\}$. For $c \in \{0.5, 0.9\}$, we also computed $\hat{\alpha}^*$ and $\hat{\beta}^*$ and standardized them by their asymptotic variances as presented in Theorem 4. The procedure was repeated $N = 1000$ times resulting in N values of α_n^*, β_n^*, $\hat{\alpha}^*$, and $\hat{\beta}^*$ for chosen p and c.

The Q-Q plots for the normal distribution 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: the constructed Q-Q plots look like straight lines already for $p = 100$ with the exception in the case of β_n^* for $c = 2$. Moderate deviations are present only for small dimension $p = 20$ and sample sizes $n = 22$ and $n = 10$ for $c = 0.9$ and $c = 2.0$, respectively. The finite-sample performance of the asymptotic results presented in Theorem 4 were investigated in Figure 2 for $c < 1$. Here, we observed a perfect fit to the normal distribution even for moderate sample sizes and moderate dimensions.

The scatter plots of the optimal shrinkage intensities and the corresponding bona fide estimators are provided in Figure 3. We highlighted with triangles the corresponding asymptotic values α^* and β^*. It is important to note that all the values converge to the asymptotic ones and all the clouds are centered around triangles with the dispersion decreasing with n. It is remarkable that α^* should be positive. In contrast, both α_n and $\hat{\alpha}^*$ can be negative. Additional insight is provided by Table 1, where relative frequencies of α_n^* and $\hat{\alpha}^*$ being below zero for several values of p and c are presented. It is remarkable that they decrease as p increases.

Next, the quality of the estimators was 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 4 and 5 in the first, second and third rows, while the fourth rows present the time in seconds needed to compute each of the estimators (except the asymptotic one). The first and the second rows show the results for the distribution of eigenvalues of $\boldsymbol{\Sigma}_n$ in proportions 20%, 40%, 40% for values 1, 3 and 10 as mentioned in the beginning of this section, while in the third row, $\lambda_{\max} = p$ was chosen which corresponds to the extreme lambda case. The first and second rows differ in the choice of γ; $\gamma = 1$ in the first row and $\gamma = 0$ in the second row. For each p and c the procedure was repeated $N = 1000$ times and averaged over simulation runs quadratic losses are plotted.

If $\gamma > 0$, then $\alpha^* \to 1$ a.s. Hence, the shrinkage estimator tends to the sample mean vector in this case. Interestingly, a similar behavior is also present for all of the considered benchmark estimators. This result is clearly visible in the first rows on both figures. As shown in Figure 4, the optimal shrinkage estimator and its asymptotic counterparts with α_n^* and β_n^* replaced by α^* and β^* exhibited 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 $c < 1$. When $c > 1$, the Moore–Penrose inverse of the sample covariance matrix was

ACCEPTED MANUSCRIPT

Table 1: Relative frequencies of optimal shrinkage intensity $\alpha*_n$ and its bona fide estimator $\hat{\alpha}^*$ to be negative.

	α_n^*			$\hat{\alpha}^*$		
$p \setminus c$	0.5	0.9	2.0	0.5	0.9	2.0
20	0.082	0.071	0.180	0.505	0.555	0.621
100	0.000	0.000	0.014	0.192	0.397	0.200
250	0.000	0.000	0.000	0.036	0.335	0.043
500	0.000	0.000	0.000	0.006	0.268	0.010

employed. The new estimator strongly dominates the rest of the competitors for $c = 1.5$ and shows almost equivalent performance to the $\hat{\boldsymbol{\mu}}_{n,W}$ for $c = 2.0$. It is worth mentioning that the loss function for $\bar{\mathbf{y}}_n$ is close to c in all of the considered cases. The proposed estimator is also 30% faster to compute than $\hat{\boldsymbol{\mu}}_{n,W}$. In all cases, we observe very fast convergence of quadratic loss to zero when the target mean vector $\boldsymbol{\mu}_0$ coincides with the true mean vector $\boldsymbol{\mu}_n$, as shown by the red solid lines in Figures 4 and 5.

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 extending from June 12, 2001 to June 9, 2016. We compared the ability of the estimators of the mean vector to predict the future realization of the expected return of the equally-weighted portfolio, which is a popular trading strategy in the high-dimensional case; see, e.g., [21], [43]. For this 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 the n most recent observations, viz. $\mathbf{Y}_{t,n} = \{y_{i\tau} : \tau \in \{t-n, \ldots, t\}, i \in \{1, \ldots, p\}\}$, and estimated the expected return of the equally-weighted portfolio which was then compared to the realized return of this portfolio on day $t + 1$.

For the proposed bona fide estimator $\hat{\boldsymbol{\mu}}_n(\mathrm{bf})$, we considered three different choices of the target mean value $\boldsymbol{\mu}_0$. In the first case, the elements of $\boldsymbol{\mu}_0$ is drawn randomly from the uniform distribution on the interval covering the smallest and the largest average of weekly returns for the corresponding stock (over respective rolling window). In the second case, $\boldsymbol{\mu}_0$ consists of randomly interchanging values of 1 and -1, whereas in the third case we set $\boldsymbol{\mu}_0 = \mathbf{1}_p$.

For a given estimator $\hat{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}_n$, the estimated expected return of the equally-weighted portfolio is calculated by $\hat{R}_{t,n}(\hat{\boldsymbol{\mu}}) = \mathbf{1}^\top \hat{\boldsymbol{\mu}}/p$. 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

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

where $T = 683$ corresponds to the number of rolling windows. The results are depicted in Table 2. The proposed shrinkage estimator is seen to outperform the other estimators in three out of four cases, particularly in the case where the target mean vector $\boldsymbol{\mu}_0$ contains randomly interchanging values of 1 and -1. Only for larger sample size $n = 100$ (with $c = 4.12$), the estimator $\widehat{\boldsymbol{\mu}}_{n,W}$ performs better than the new estimator with any target. In contrast, $\widehat{\boldsymbol{\mu}}_{n,W}$ has the largest value of the average quadratic deviation if $n = 25$. Note also the poor performance of the sample mean vector, which is always one of the worst estimation strategies.

7. Summary

Nowadays, modern scientific data sets involve sample sizes that are often comparable or even lower than the dimension (number of features) of the data generating mechanism. In this situation many statistical, financial and genetic problems require simple and feasible estimators of the mean vector. Although most of the classical multivariate procedures are based on limit theorems assuming that the dimension p is fixed and the sample size n increases, it has

ACCEPTED MANUSCRIPT

Table 2: Performance of different mean estimators on the basis of equally weighted portfolio for weekly returns for different sample sizes. The best models are underlined for each sample size.

n	c	$\bar{\mathbf{y}}_n$	$\widehat{\boldsymbol{\mu}}_{JS}$		$\widehat{\boldsymbol{\mu}}_{n,W}$	$\boldsymbol{\mu}_0$	$\widehat{\boldsymbol{\mu}}_n$ (bf)	$\widehat{\boldsymbol{\mu}}_n$ (bf)	$\widehat{\boldsymbol{\mu}}_n$ (bf)
			$\widehat{\boldsymbol{\mu}}_{JS(p>n)}$	$\widehat{\boldsymbol{\mu}}_{JS+}$			uniform	sample$(1,-1)$	$(1,\ldots,1)$
25	16.48	7.882	7.832	7.832	7.931		7.831	<u>7.824</u>	7.847
50	8.24	7.813	7.721	7.721	7.748		7.716	<u>7.715</u>	7.736
75	5.49	7.769	7.672	7.673	7.669		<u>7.660</u>	7.668	7.687
100	4.12	7.723	7.646	7.646	<u>7.630</u>		7.636	7.644	7.654

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., [6] and references therein.

In this paper, we showed how to estimate the mean vector using the optimal shrinkage technique based on random matrix theory. We proved the asymptotic normality of the optimal shrinkage intensities as well as their almost sure limit convergence under the large-dimensional asymptotic regime. Both the simulation study and the empirical application illustrate the dominance of the proposed estimation technique over the benchmark methods in terms of quadratic loss and computational time. Following the recent developments in functional data analysis where a data curve can be treated as a high-dimensional vector of its discretized values on a finite grid [2, 3], the results derived here have potential applications in dealing with estimation problems arising in functional data settings.

Appendix

This appendix contains proofs of this paper's results.

Proof of Theorem 1. Let

$$\eta_1 = p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\xi} = n^{-1}p^{-\gamma}\mathbf{1}_n^\top\mathbf{Y}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\xi} = n^{-1}p^{-\gamma}\mathbf{1}_n^\top\mathbf{X}_n^\top\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{\xi} + p^{-\gamma}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\xi}, \tag{10}$$

$$\begin{aligned}\eta_2 &= p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n = n^{-2}p^{-\gamma}\mathbf{1}_n^\top\mathbf{Y}_n^\top\boldsymbol{\Sigma}_n^{-1}\mathbf{Y}_n\mathbf{1}_n \\ &= n^{-2}p^{-\gamma}\mathbf{1}_n^\top(\mathbf{X}_n^\top\mathbf{X}_n)\mathbf{1}_n + 2n^{-1}p^{-\gamma}\mathbf{1}_n^\top\mathbf{X}_n^\top\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{\mu}_n + p^{-\gamma}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n,\end{aligned} \tag{11}$$

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

$$n^{-1}p^{-\gamma}\mathbf{1}_n^\top\mathbf{X}_n^\top\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{\xi} = n^{-1}p^{-\gamma}(\mathbf{1}_n^\top\otimes\boldsymbol{\xi}^\top\boldsymbol{\Sigma}_n^{-1/2})\mathrm{vec}(\mathbf{X}_n) \tag{12}$$

and $n^{-2}\mathbf{1}_n^\top(\mathbf{X}_n^\top\mathbf{X}_n)\mathbf{1}_n = \|\mathbf{X}_n\mathbf{1}_n/\sqrt{n}\|^2/n$.

In Eq. (12), the sum of pn independent random variables is present with zero means and variances equal to $(\boldsymbol{\xi}^\top\boldsymbol{\Sigma}_n^{-1/2}\mathbf{e}_i)^2$ for $i\in\{1,\ldots,p\}$, where $\mathbf{e}_i$ is the ith basis vector in $\mathbb{R}^p$. Because

$$\begin{aligned}\mathrm{var}\{n^{-1}p^{-\gamma}(\mathbf{1}_n^\top\otimes\boldsymbol{\xi}^\top\boldsymbol{\Sigma}_n^{-1/2})\mathrm{vec}(\mathbf{X}_n)\} &= \frac{p^{-2\gamma}n}{n^2}\sum_{i=1}^{p}(\boldsymbol{\xi}^\top\boldsymbol{\Sigma}_n^{-1/2}\mathbf{e}_i)^2 = \frac{p^{-2\gamma}}{n}\boldsymbol{\xi}^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\xi} \\ &\le \lambda_{\max}(\boldsymbol{\Sigma}_n^{-1})\frac{p^{-\gamma}}{n}p^{-\gamma}\|\boldsymbol{\xi}\|^2 \le \frac{\lambda_{\max}(\boldsymbol{\Sigma}_n^{-1})M_u}{np^\gamma} < \infty,\end{aligned}$$

the application of Kolmogorov's Strong Law of Large Numbers implies that, as $p,n\to\infty$, $p^{-\gamma}n^{-1}\mathbf{1}_n^\top\mathbf{X}_n^\top\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{\xi}\xrightarrow{\text{a.s.}}0$.

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 $n^{-1/2}\mathbf{X}_n\mathbf{1}_n$ are asymptotically independent and standard Gaussian. Consequently, their squares are asymptotically independent and χ^2- distributed with one degree of freedom. Because the expectation of a

ACCEPTED MANUSCRIPT

χ^2_1-distributed random variable is 1 and the variance is finite we get that, as $p, n \to \infty$, $\|\mathbf{X}_n\mathbf{1}_n/\sqrt{n}\|^2/p \xrightarrow{\text{a.s.}} 1$ and hence, as $p, n \to \infty$, $n^{-2}\mathbf{1}_n^\top(\mathbf{X}_n^\top\mathbf{X}_n)\mathbf{1}_n \xrightarrow{\text{a.s.}} c$. Hence,

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

and, similarly,

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

for $p/n \to c > 0$ as $n \to \infty$. The theorem is thus proved. □

In the proof of Theorem 2 we make use of the following lemma.

Lemma 1. *Let* $\mathbf{Z} \sim \mathcal{N}_m(\mathbf{0}, \mathbf{I}_m)$ *and define* $Q_1(\mathbf{Z}) = \sqrt{m}\,(\mathbf{Z}^\top/m\mathbf{Z} - 1)$ *and* $Q_2(\mathbf{Z}) = \mathbf{a}^\top\mathbf{Z}/\sqrt{m}$ *for nonrandom* $\mathbf{a}$ *with* $\lim_{m\to\infty}\mathbf{a}^\top\mathbf{a}/m = \tilde{\sigma}^2$. *Then, as* $m \to \infty$,

$$\begin{pmatrix} Q_1(\mathbf{Z}) \\ Q_2(\mathbf{Z}) \end{pmatrix} \rightsquigarrow \mathcal{N}_2\left[\mathbf{0}, \begin{pmatrix} 2 & 0 \\ 0 & \tilde{\sigma}^2 \end{pmatrix}\right].$$

Proof of Lemma 1. From Lemma 2.1 of [40] we get the moment generating function of $(Q_1(\mathbf{Z}), Q_2(\mathbf{Z}))^\top$ expressed as

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

Hence,

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

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 $\chi \sim \chi^2_m$ and $\sqrt{m}\,(\chi/m - 1) \rightsquigarrow \mathcal{N}(0, 2)$ as $m \to \infty$. The lemma is thus proved. □

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

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

where

$$Z_\alpha = \sqrt{p^\gamma n}\,[(\tilde{c}q_{00} + d)(q_{00}p^{-\gamma}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n - q_{0n}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n) - d\{q_{00}p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n - (p^{-\gamma}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n)^2\}].$$

The application of (10) and (11) leads to

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

ACCEPTED MANUSCRIPT

Because the elements of $n^{-1/2}\mathbf{X}_n\mathbf{1}_n$ are independent and standard Gaussian, we get that, as $p, n \to \infty$,

$$p^{-\gamma/2}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1/2}(n^{-1/2}\mathbf{X}_n\mathbf{1}_n) \rightsquigarrow \mathcal{N}(0, q_{00}) \to \infty$$

and, consequently, as $p, n \to \infty$,

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

Furthermore, the application of Lemma 1 leads to $Z_\alpha \rightsquigarrow \mathcal{N}(0, \sigma_{Z_\alpha}^2)$ for $p/n \to c > 0$ as $n \to \infty$ with

$$\sigma_{Z_\alpha}^2 = (\tilde{c}q_{00} - d)^2(q_{00}^2 q_{nn} + q_{0n}^2 q_{00} - 2q_{00}q_{0n}^2) + \tilde{c}d^2 q_{00}^2 = (\tilde{c}q_{00} - d)^2 q_{00} d + \tilde{c}d^2 q_{00}^2.$$

Finally, from the proof of Theorem 1 we get that, for $p/n \to c > 0$ as $n \to \infty$,

$$\frac{\tilde{c}q_{00} + d}{q_{00}p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n - (p^{-\gamma}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n)^2} \xrightarrow{\text{a.s.}} 1,$$

which together with Slutsky's theorem (see [35]) implies that, for $p/n \to c > 0$ as $n \to \infty$,

$$\sqrt{p^\gamma n}\,(\alpha_n^* - \alpha^*) \rightsquigarrow \mathcal{N}\left[0, \frac{(\tilde{c}q_{00} - d)^2 q_{00} d + \tilde{c}d^2 q_{00}^2}{(\tilde{c}q_{00} + d)^4}\right].$$

Similarly, for β_n^* we get

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

with

$$\begin{aligned} Z_\beta =\ & \sqrt{p^\gamma n}\left[(\tilde{c}q_{00} + d)(p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n q_{0n} - p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n) - \tilde{c}q_{0n}\{p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n q_{00} - (p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0)^2\}\right] \\ =\ & \sqrt{p^\gamma n}\{dq_{0n}p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\bar{\mathbf{y}}_n + \tilde{c}q_{0n}(p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0)^2 - (\tilde{c}q_{00} + d)p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 p^{-\gamma}\bar{\mathbf{y}}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n\}. \end{aligned}$$

Using (10) and (11) we get

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

Proceeding as in the proof pertaining to α_n^*, we get

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

for $p/n \to c > 0$ as $n \to \infty$. Then, the application of Lemma 1 leads to $Z_\beta \rightsquigarrow \mathcal{N}(0, \sigma_{Z_\beta}^2)$ for $p/n \to c > 0$ as $n \to \infty$ with

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

Finally, using Slutsky's theorem (see [35]) we get $\sqrt{p^\gamma n}\,(\beta_n^* - \beta^*) \rightsquigarrow \mathcal{N}[0, \sigma_{Z_\beta}^2/(\tilde{c}q_{00} + d)^4]$ for $p/n \to c > 0$ as $n \to \infty$. The theorem is proved. □

In the proof of Theorem 3 we make use of the following lemma.

ACCEPTED MANUSCRIPT

Lemma 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 Then*

$$p^{-\gamma}\left|\boldsymbol{\xi}^\top \mathbf{S}_n^{-1}\boldsymbol{\theta} - (1-c)^{-1}\boldsymbol{\xi}^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\theta}\right| \xrightarrow{\text{a.s.}} 0, \quad \bar{\mathbf{x}}_n^\top\boldsymbol{\Sigma}_n^{1/2}\mathbf{S}_n^{-1}\boldsymbol{\Sigma}_n^{1/2}\bar{\mathbf{x}}_n \xrightarrow{\text{a.s.}} \frac{c}{1-c}, \quad p^{-\gamma/2}\bar{\mathbf{x}}_n^\top\boldsymbol{\Sigma}_n^{1/2}\mathbf{S}_n^{-1}\boldsymbol{\theta} \xrightarrow{\text{a.s.}} 0$$

for $p/n \to c \in (0,\infty)$ *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 2. For $\boldsymbol{\theta}\in\{\boldsymbol{\mu}_0,\boldsymbol{\mu}_n\}$, one has $\|\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{\theta}\|^2 \le \lambda_{\max}(\boldsymbol{\Sigma}_n^{-1})\|\boldsymbol{\theta}\|^2 < \infty$ in view of assumptions (A1) and (A2). The rest of the proof follows from Lemma 5.3 in [14]. □

Proof of Theorem 3. For $\boldsymbol{\theta},\boldsymbol{\xi}\in\{\boldsymbol{\mu}_0,\boldsymbol{\mu}_n\}$, the application of Lemma 2 leads to

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

$$p^{-\gamma}|\boldsymbol{\mu}_0^\top\mathbf{S}_n^{-1}\bar{\mathbf{y}}_n - (1-c)^{-1}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n| \le p^{-\gamma}|\boldsymbol{\mu}_0^\top\mathbf{S}_n^{-1}\boldsymbol{\Sigma}_n^{1/2}\bar{\mathbf{x}}_n| + p^{-\gamma}|\boldsymbol{\mu}_0^\top\mathbf{S}_n^{-1}\boldsymbol{\mu}_n - (1-c)^{-1}\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n| \xrightarrow{\text{a.s.}} 0,$$

and

$$\begin{aligned}p^{-\gamma}|\bar{\mathbf{y}}_n^\top\mathbf{S}_n^{-1}\bar{\mathbf{y}}_n - (1-c)^{-1}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n - c(1-c)^{-1}| &\le p^{-\gamma}|\boldsymbol{\mu}_n^\top\mathbf{S}_n^{-1}\boldsymbol{\mu}_n - (1-c)^{-1}\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n|\\ &\quad + p^{-\gamma}|\bar{\mathbf{x}}_n^\top\boldsymbol{\Sigma}_n^{1/2}\mathbf{S}_n^{-1}\boldsymbol{\Sigma}_n^{1/2}\bar{\mathbf{x}}_n - c(1-c)^{-1}| + 2p^{-\gamma}|\boldsymbol{\mu}_n^\top\mathbf{S}_n^{-1}\boldsymbol{\Sigma}_n^{1/2}\bar{\mathbf{x}}_n| \xrightarrow{\text{a.s.}} 0.\end{aligned}$$

Hence,

$$\hat{\alpha}^* = \frac{\{\bar{\mathbf{y}}_n^\top\mathbf{S}_n^{-1}\bar{\mathbf{y}}_n - c/(1-c)\}\boldsymbol{\mu}_0^\top\mathbf{S}_n^{-1}\boldsymbol{\mu}_0 - (\bar{\mathbf{y}}_n^\top\mathbf{S}_n^{-1}\boldsymbol{\mu}_0)^2}{\bar{\mathbf{y}}_n^\top\mathbf{S}_n^{-1}\bar{\mathbf{y}}_n\boldsymbol{\mu}_0^\top\mathbf{S}_n^{-1}\boldsymbol{\mu}_0 - (\bar{\mathbf{y}}_n^\top\mathbf{S}_n^{-1}\boldsymbol{\mu}_0)^2} \xrightarrow{\text{a.s.}} \frac{\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 - (\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0)^2}{(c+\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n)\boldsymbol{\mu}_0^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 - (\boldsymbol{\mu}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0)^2} = \alpha^*$$

and, similarly,

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

for $p/n\to c\in(0,\infty)$ as $n\to\infty$. Thus the theorem is proved. □

The proof of Theorem 4 is based on Lemma 3. Let $\hat{s} = \bar{\mathbf{x}}_n^\top\mathbf{S}_n^{-1}\bar{\mathbf{x}}_n - (\boldsymbol{\mu}_0^\top\mathbf{S}_n^{-1}\bar{\mathbf{x}}_n)^2/\boldsymbol{\mu}_0^\top\mathbf{S}_n^{-1}\boldsymbol{\mu}_0$ and $\hat{R} = \boldsymbol{\mu}_0^\top\mathbf{S}_n^{-1}\bar{\mathbf{x}}_n/\boldsymbol{\mu}_0^\top\mathbf{S}_n^{-1}\boldsymbol{\mu}_0$.

Lemma 3. *Let* $\mathbf{y}_1,\ldots,\mathbf{y}_n$ *be a random sample of independent vectors such that* $\mathbf{y}_i\sim\mathcal{N}_p(\boldsymbol{\mu}_n,\boldsymbol{\Sigma}_n)$ *for all* $i\in\{1,\ldots,n\}$. *Then for any* p *and* n *with* $n>p$, *the following statements hold.*

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

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

Proof of Lemma 3. The results are obtained following the proof of Theorem 3.1 in [13], where similar statements are presented in the case $\boldsymbol{\mu}_0=\mathbf{1}_p$. □

Proof of Theorem 4. Let $\hat{s}_c=(1-c)\hat{s}-c$. Then $\sqrt{n}\,(\hat{s}_c-s) = \sqrt{n}\,\{(1-c)\hat{s}-c-s\} = \sqrt{n}\,\{(1-c)\hat{s}/c-1-s/c\})c$. An application of Lemma 3.a and Lemma 3 in [12] leads to

$$\sqrt{n}(\hat{s}_c-s)\rightsquigarrow\mathcal{N}(0,\sigma_s^2) \quad\text{with}\quad \sigma_s^2 = 2\,(c+2s) + \frac{2}{1-c}\,(c+s)^2\,. \tag{13}$$

From Theorem 3, we get $\hat{\alpha}^* = 1-c(1-c)^{-1}/\hat{s} = 1-c/(c+\hat{s}_c)$ and

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

ACCEPTED MANUSCRIPT

where

$$\sqrt{n}\begin{pmatrix} \hat{s}_c - s \\ z_0 \end{pmatrix} \rightsquigarrow \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_s^2 & 0 \\ 0 & 1 \end{pmatrix}\right].$$

with σ_s^2 as in (13). Now, an application of the Delta method (see, e.g., Theorem 3.7 in [20]) leads to

$$\sqrt{n}\begin{pmatrix} \hat{\alpha}^* - \alpha^* \\ \hat{\beta}^* - \beta^* \end{pmatrix} \rightsquigarrow \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Omega}\right]$$

with

$$\boldsymbol{\Omega} = \begin{pmatrix} \frac{c}{(c+s)^2} & 0 \\ \frac{c}{(c+s)^2}R & \frac{c}{c+s}\frac{\sqrt{1+\frac{s+c}{1-c}}}{\sqrt{\boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0}} \end{pmatrix} \begin{pmatrix} \sigma_s^2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{c}{(c+s)^2} & \frac{c}{(c+s)^2}R \\ 0 & \frac{c}{c+s}\frac{\sqrt{1+\frac{s+c}{1-c}}}{\sqrt{\boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0}} \end{pmatrix}$$
$$= \begin{pmatrix} \frac{c^2\sigma_s^2}{(c+s)^4} & \frac{c^2\sigma_s^2}{(c+s)^4}R \\ \frac{c^2\sigma_s^2}{(c+s)^4}R & \frac{c^2\sigma_s^2}{(c+s)^4}R^2 + \frac{c^2}{(c+s)^2}\frac{1+\frac{s+c}{1-c}}{\boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0} \end{pmatrix}.$$

Hence the theorem is proved. □

In the proof of Theorem 4 we make use of the following lemma.

Lemma 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*

$$p^{-\gamma}\left|\boldsymbol{\xi}^\top \mathbf{S}_n^- \boldsymbol{\theta} - c^{-1}(c-1)^{-1}\boldsymbol{\xi}^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\theta}\right| \xrightarrow{\text{a.s.}} 0, \quad \bar{\mathbf{x}}_n^\top \boldsymbol{\Sigma}_n^{1/2}\mathbf{S}_n^- \boldsymbol{\Sigma}_n^{1/2}\bar{\mathbf{x}}_n \xrightarrow{\text{a.s.}} \frac{1}{c-1}, \quad p^{-\gamma/2}\bar{\mathbf{x}}_n^\top \boldsymbol{\Sigma}_n^{1/2}\mathbf{S}_n^- \boldsymbol{\theta} \xrightarrow{\text{a.s.}} 0$$

for $p/n \to c \in (1,\infty)$ *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 4. For $\boldsymbol{\theta} \in \{\boldsymbol{\mu}_0, \boldsymbol{\mu}_n\}$, one has $\left\|\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{\theta}\right\|^2 \le \lambda_{\max}(\boldsymbol{\Sigma}_n^{-1})\,\|\boldsymbol{\theta}\|^2 < \infty$ following assumptions (A1) and (A2). The rest of the proof follows from Lemma 5.6 in [14]. □

Proof of Theorem 5. For $\boldsymbol{\theta}, \boldsymbol{\xi} \in \{\boldsymbol{\mu}_0, \boldsymbol{\mu}_n\}$, the application of Lemma 4 leads to

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

$$p^{-\gamma}|\boldsymbol{\mu}_0^\top \mathbf{S}_n^- \bar{\mathbf{y}}_n - c^{-1}(c-1)^{-1}\boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n| \le p^{-\gamma}|\boldsymbol{\mu}_0^\top \mathbf{S}_n^- \boldsymbol{\Sigma}_n^{1/2}\bar{\mathbf{x}}_n| + p^{-\gamma}|\boldsymbol{\mu}_0^\top \mathbf{S}_n^- \boldsymbol{\mu}_n - c^{-1}(c-1)^{-1}\boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n| \xrightarrow{\text{a.s.}} 0,$$

and

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

Hence,

$$\hat{\alpha}^* = \frac{\{\bar{\mathbf{y}}_n^\top \mathbf{S}_n^- \bar{\mathbf{y}}_n - (c-1)^{-1}\}\boldsymbol{\mu}_0^\top \mathbf{S}_n^- \boldsymbol{\mu}_0 - (\bar{\mathbf{y}}_n^\top \mathbf{S}_n^- \boldsymbol{\mu}_0)^2}{\bar{\mathbf{y}}_n^\top \mathbf{S}_n^- \bar{\mathbf{y}}_n \boldsymbol{\mu}_0^\top \mathbf{S}_n^- \boldsymbol{\mu}_0 - (\bar{\mathbf{y}}_n^\top \mathbf{S}_n^- \boldsymbol{\mu}_0)^2} \xrightarrow{\text{a.s.}} \frac{\boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n \boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 - (\boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0)^2}{(c + \boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n)\boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0 - (\boldsymbol{\mu}_n^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_0)^2} = \alpha^*$$

and, similarly,

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

for $p/n \to c \in (1,\infty)$ as $n \to \infty$. This completes the proof of the theorem. □

ACCEPTED MANUSCRIPT

Acknowledgments

We are grateful to Professors Christian Genest and Philippe Vieu as well as two anonymous reviewers for their helpful suggestions. We also gratefully acknowledge the comments from the participants at the Fourth International Workshop on Functional and Operatorial Statistics (A Coruña), the European Meeting of Statisticians (Amsterdam), the WISE-CASE Workshop on Econometrics and Statistics (Xiamen, China), and the Joint Meeting of the German Mathematical Society and the Polish Mathematical Society (Poznan). Finally, we thank Cheng Wang for providing us the MATLAB code for his estimation method.

References

[1] G. Aneiros, P. Vieu, Variable selection in infinite-dimensional problems, Statist. Probab. Lett. 94 (2014) 12–20.
[2] G. Aneiros, P. Vieu, Partial linear modelling with multi-functional covariates, Comput. Statist. 30 (2015) 647–671.
[3] G. Aneiros, P. Vieu, Sparse nonparametric model for regression with functional covariate, J. Nonparam. Statist. 28 (2016) 839–859.
[4] J. Bai, S. Ng, Determining the number of factors in approximate factor models, Econometrica 70 (2002) 191–221.
[5] J. Bai, S. Ng, Large dimensional factor analysis, Foundations and Trends in Econometrics 3 (2008) 89–163.
[6] Z.D. Bai, J.W. Silverstein, CLT for linear spectral statistics of large dimensional sample covariance matrices, Ann. Probab. 32 (2004) 553–605.
[7] Z.D. Bai, J.W. Silverstein, Spectral Analysis of Large Dimensional Random Matrices, Springer, New York, 2010.
[8] A.J. Baranchik, A family of minimax estimators of the mean of a multivariate normal distribution, Ann. Math. Statist. 41 (1970) 642–645.
[9] J.O. Berger, M.E. Bock, Combining independent normal mean estimation problems with unknown variances, Ann. Statist. 4 (1976) 642–648.
[10] J.O. Berger, M.E. Bock, L.D. Brown, G. Casella, L. Gleser, Minimax estimation of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix, Ann. Statist. 5 (1977) 763–771.
[11] T. Bodnar, H. Dette, N. Parolya, Spectral analysis of the Moore–Penrose inverse of a large dimensional sample covariance matrix, J. Multivariate Anal. 148 (2016) 160–172.
[12] T. Bodnar, M. Reiss, Exact and asymptotic tests on a factor model in low and large dimensions, J. Multivariate Anal. 150 (2016) 125–151.
[13] T. Bodnar, W. Schmid, Estimation of optimal portfolio compositions for Gaussian returns, Statistics & Risk Modeling 26 (2008) 179–201.
[14] T. Bodnar, Y. Okhrin, N. Parolya, Optimal shrinkage-based portfolio selection in high dimensions, Working Paper, 2017.
[15] E.G. Bongiorno, E. Salinelli, A. Goia, P. Vieu, Contributions in Infinite-Dimensional Statistics and Related Topics, Società Editrice Esculapio, Bologna, Italy, 2014.
[16] T.T. Cai, C.-H. Zhang, H.H. Zhou, Optimal rates of convergence for covariance matrix estimation, Ann. Statist. 38 (2010) 2118–2144.
[17] G. Chamberlain, M. Rothschild, Arbitrage, factor structure, and mean-variance analysis on large asset markets, Econometrica 51 (1983) 1281–1304.
[18] D. Chételat, M.T. Wells, Improved multivariate normal mean estimation with unknown covariance when p is greater than n, Ann. Statist. 40 (2012) 3137–3160.
[19] A. Cuevas, A partial overview of the theory of statistics with functional data, J. Statist. Plann. Inf. 147 (2014) 1–23.
[20] A. DasGupta, Asymptotic Theory of Statistics and Probability, Springer, New York, 2008.
[21] V. DeMiguel, L. Garlappi, R. Uppal, Optimal versus naive diversification: How inefficient is the $1/N$ portfolio strategy? Review of Financial Studies 22 (2009) 1915–1953.
[22] J. Fan, F. Han, H. Lui, Challenges of Big Data analysis, Nat. Sci. Rev. 1 (2014) 293–314.
[23] J. Fan, Y. Fan, J. Lv, High dimensional covariance matrix estimation using a factor model, J. Econometrics 147 (2008) 186–197.
[24] J. Fan, Y. Liao, M. Mincheva, Large covariance estimation by thresholding principal orthogonal complements, J. R. Stat. Soc. Ser. B 75 (2013) 603–680.
[25] J. Fan, J. Zhang, K. Yu, Vast portfolio selection with gross-exposure constraints, J. Amer. Statist. Assoc. 107 (2012) 592–606.
[26] F. Ferraty, P. Hall, P. Vieu, Most-predictive design points for functional data predictors, Biometrika 97 (2010) 807–824.
[27] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis: Theory and Practice, Springer, Berlin, 2006.
[28] D. Fourdrinier, W.E. Strawderman, M.T. Wells, Robust shrinkage estimation for elliptically symmetric distributions with unknown covariance matrix, J. Multivariate Anal. 85 (2003) 24–39.
[29] O. Friesen, M. Löwe, M. Stolz, Gaussian fluctuations for sample covariance matrices with dependent data, J. Multivariate Anal. 114 (2013) 270–287.
[30] L.J. Gleser, Minimax estimators of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix, Ann. Statist. 14 (1986) 1625–1633.
[31] A. Goia, P. Vieu, An introduction to recent advances in high/infinite dimensional statistics, J. Multivariate Anal. 146 (2016) 1–6.
[32] L. Horváth, P. Kokoszka, Inference for Functional Data with Applications, Springer, New York, 2012.
[33] E.B. Issa, A.M. Papanastassiou, J.J. DiCarlo, Large-scale, high-resolution neurophysiological maps underlying fMRI of Macaque temporal lobe, J. Neuroscience 33 (2013) 15207–15219.
[34] W. James, C. Stein, Estimation with quadratic loss, Proc. Fourth Berkeley Symp. on Math. Statist. and Prob. 1 (1961) 361–379.
[35] E.L. Lehmann, Elements of Large-Sample Theory, Springer, New York, 1999.
[36] P.-E. Lin, H.-L. Tsai, Generalized Bayes minimax estimators of the multivariate normal mean with unknown covariance matrix, Ann. Statist. 1 (1973) 142–145.
[37] J.O. Ramsay, Functional Data Analysis, Wiley Online Library, 2006.
[38] J.O. Ramsay, B.W. Silverman, Applied Functional Data Analysis: Methods and Case Studies, Springer, 2007.
[39] F. Rubio, X. Mestre, D.P. Palomar, Performance analysis and optimal selection of large minimum variance portfolios under estimation risk, J. Sel. Topics Signal Proc. (2012) 337–350.

ACCEPTED MANUSCRIPT

[40] A. Schöne, W. Schmid, On the joint distribution of a quadratic and a linear form in normal variables, J. Multivariate Anal. 72 (2000) 163 – 182.
[41] D. Shalon, S.J. Smith, P.O. Brown, A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization, Genome Research 6 (1996) 639–645.
[42] R.S. Tsay, Analysis of Financial Time Series, Wiley-Interscience, Hoboken, NJ, 2005.
[43] J. Tu, G. Zhou, Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies, J. Financial Econom. 99 (2011) 204–215.
[44] P. Vieu, On dimension reduction models for functional data, Statist. Probab. Lett. 136 (2018) 134–138.
[45] C. Wang, T. Tong, L. Cao, B. Miao, Non-parametric shrinkage mean estimation for quadratic loss functions with unknown covariance matrices, J. Multivariate Anal. 125 (2014) 222–232.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

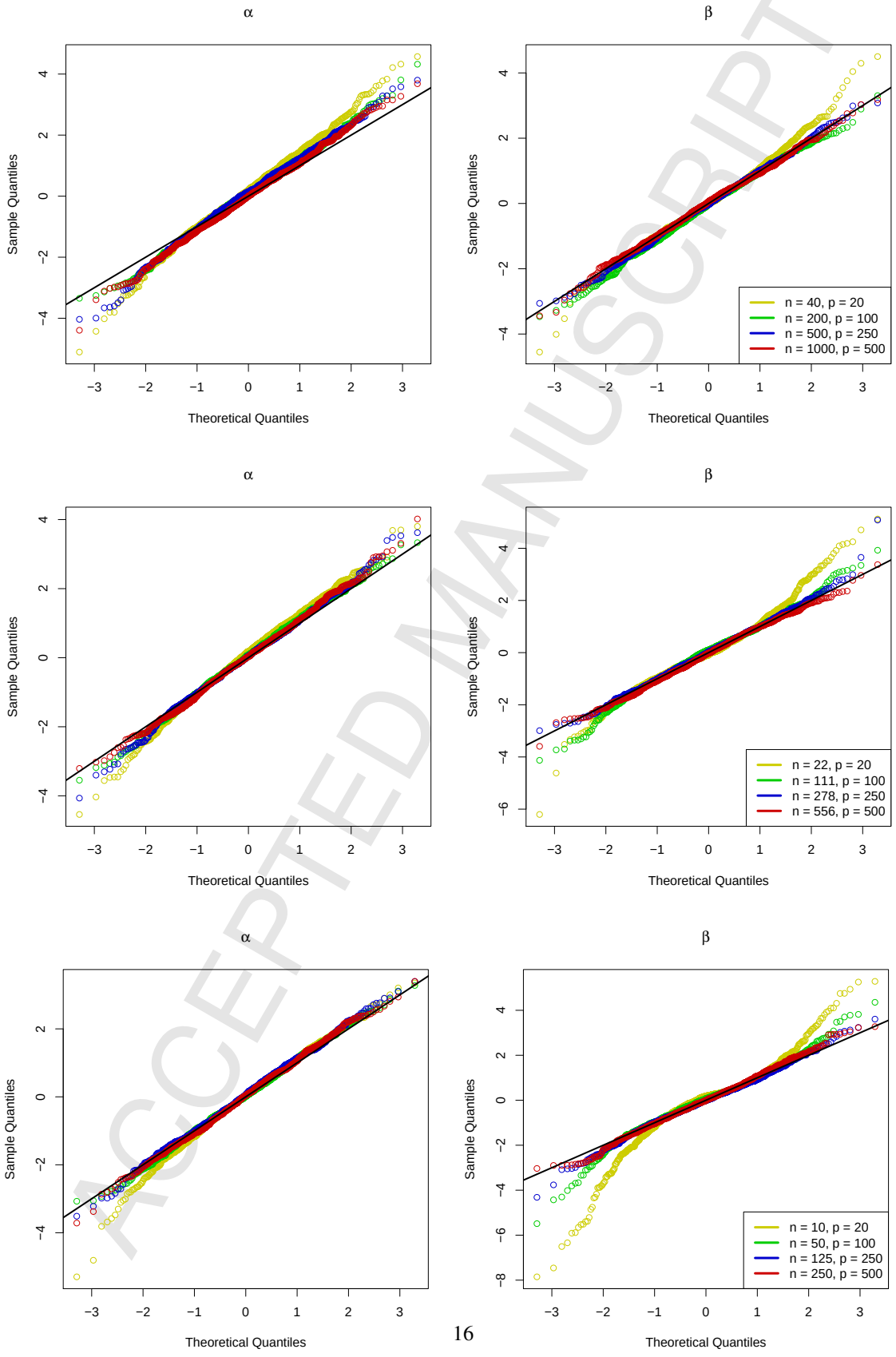

Figure 1: Q-Q plots for α_n^* and β_n^* in the case of $\gamma = 0$. We set $c = 0.5$ (top panel), $c = 0.9$ (middle panel) and $c = 2.0$ (bottom panel).

ACCEPTED MANUSCRIPT

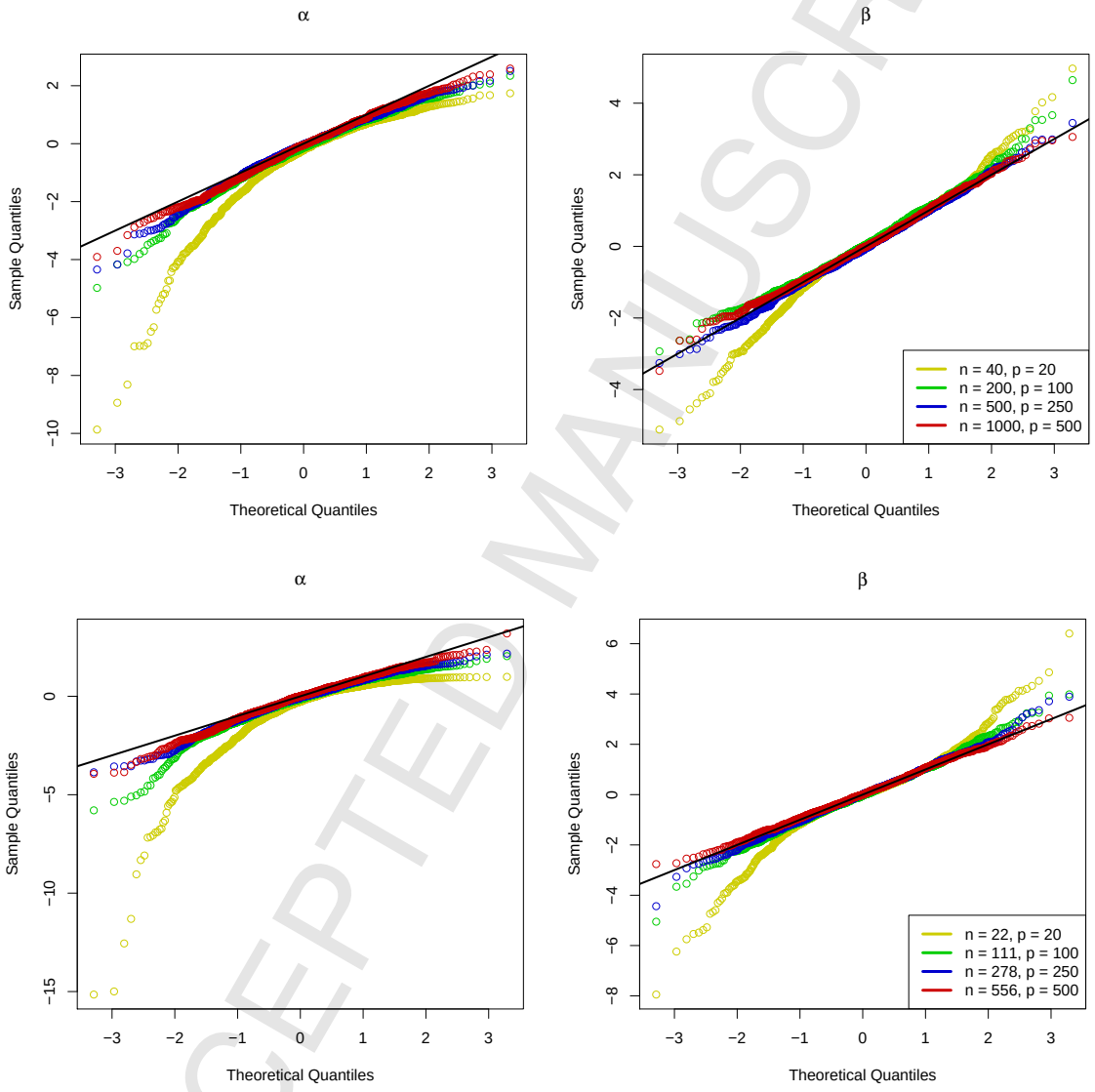

Figure 2: Q-Q plots for $\hat{\alpha}^*$ and $\hat{\beta}^*$ in the case of $\gamma = 0$. We set $c = 0.5$ (top panel) and $c = 0.9$ (bottom panel).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

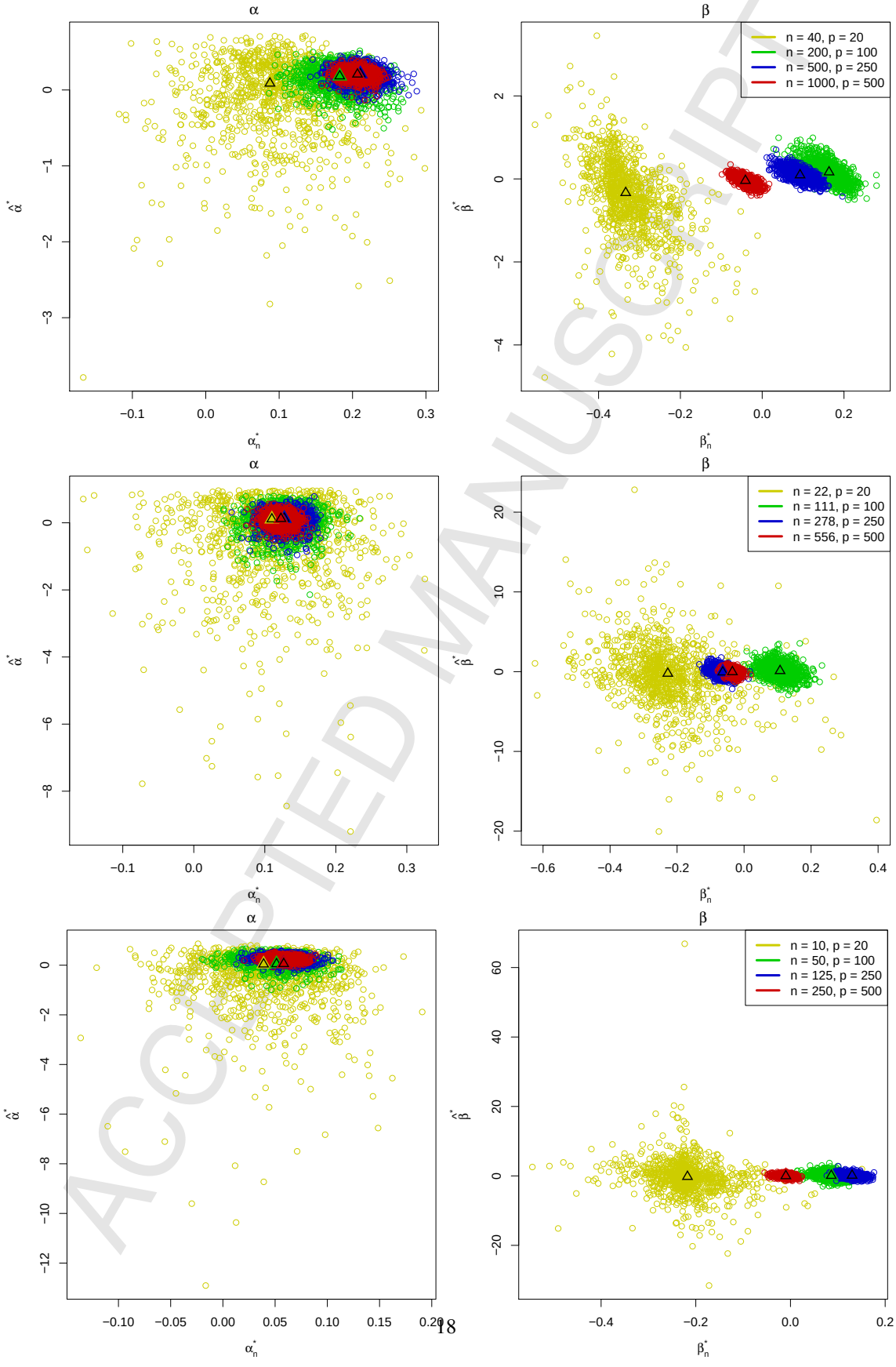

Figure 3: Scatter plots for $(\alpha_n^*, \hat{\alpha}^*)$ and $(\beta_n^*, \hat{\beta}^*)$ in the case of $\gamma = 0$. The triangular shows the asymptotic values (α^*, α^*) and (β^*, β^*). We set $c = 0.5$ (top panel), $c = 0.9$ (middle panel) and $c = 2.0$ (bottom panel).

ACCEPTED MANUSCRIPT

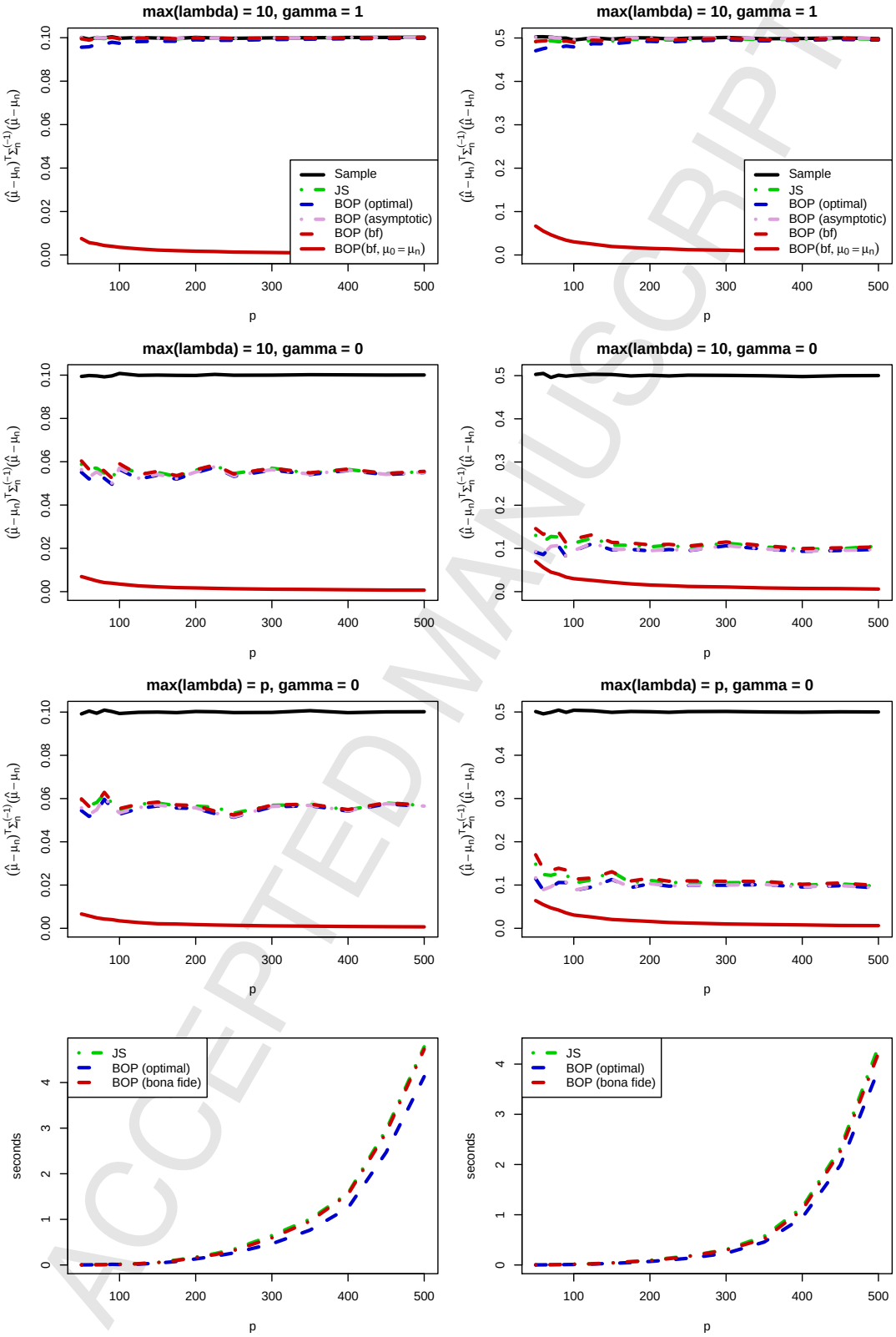

Figure 4: Averaged quadratic loss for estimators $\bar{\mathbf{y}}_n$ (black), $\hat{\boldsymbol{\mu}}_n$ (bona fide, red, dashed for $\boldsymbol{\mu}_0 \neq \boldsymbol{\mu}_n$ and solid for $\boldsymbol{\mu}_0 = \boldsymbol{\mu}_n$), $\boldsymbol{\mu}_n^*$ (optimal, blue), $\boldsymbol{\mu}^*$ (asymptotic, plum), $\widehat{\boldsymbol{\mu}}_{n,JS}$ (green) of $\boldsymbol{\mu}_n$, performed with $N = 1000$ iterations for $c = 0.1$ (left) and $c = 0.5$ (right).

ACCEPTED MANUSCRIPT

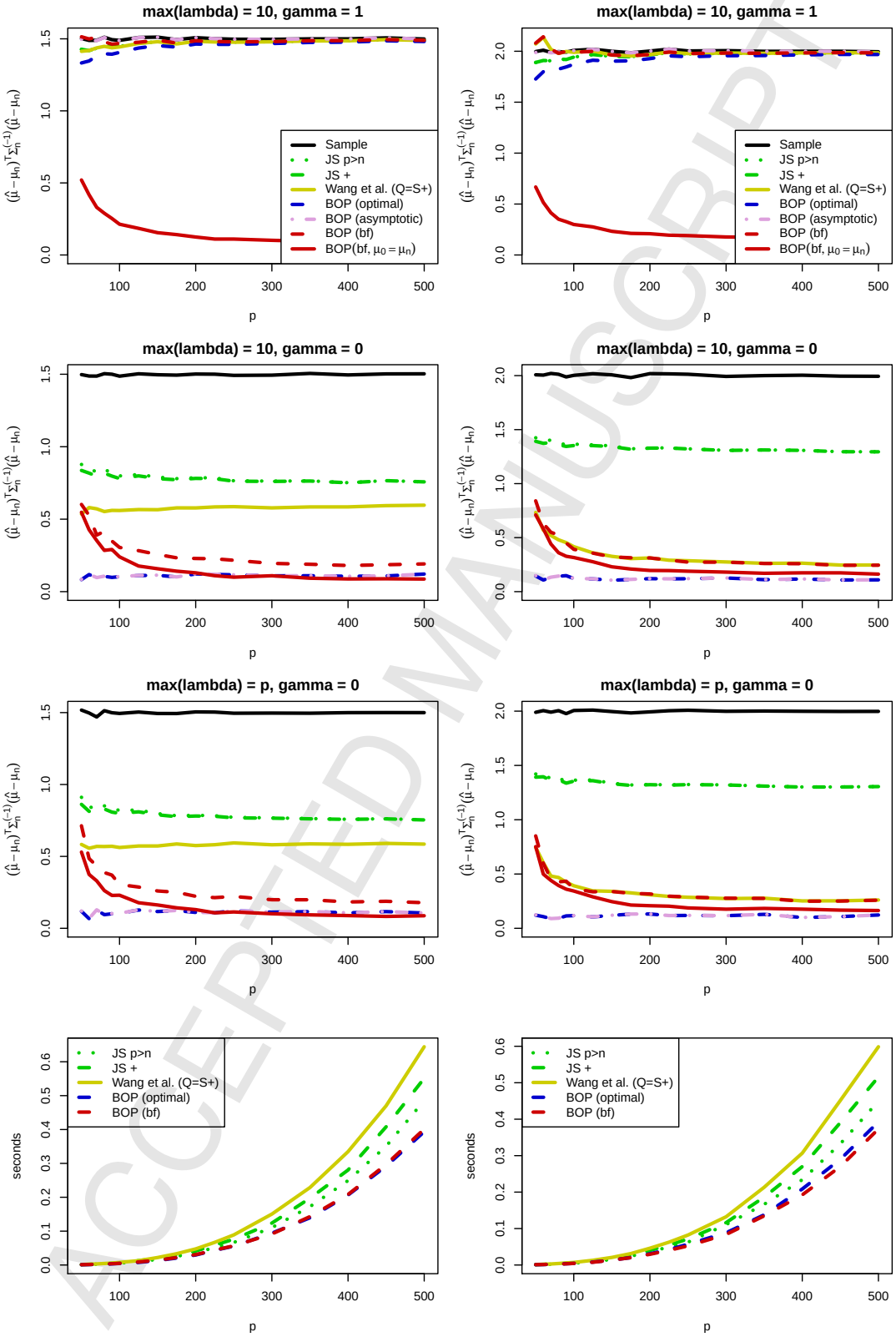

Figure 5: Quadratic loss function for estimators $\bar{\mathbf{y}}_n$ (black), $\boldsymbol{\mu}^*$ (asymptotic, plum), $\boldsymbol{\mu}_n^*$ (optimal, blue), $\hat{\boldsymbol{\mu}}_n$ (bona fide, red, dashed for $\boldsymbol{\mu}_0 \neq \boldsymbol{\mu}_n$ and solid for $\boldsymbol{\mu}_0 = \boldsymbol{\mu}_n$), $\hat{\boldsymbol{\mu}}_{n,W}$ (yellow), $\hat{\boldsymbol{\mu}}_{n,JS(p>n)}$ (green, dotted), $\hat{\boldsymbol{\mu}}_{n,JS+}$ (green, dashed) of $\boldsymbol{\mu}_n$, performed with $N = 1000$ iterations, for $c = 1.5$ (left) and $c = 2.0$ (right).