Mathematical Statistics
Stockholm University

Direct Shrinkage Estimation of Large Dimensional Precision Matrix

Taras Bodnar
Arjun K. Gupta
Nestor Parolya

Research Report 2015:1

ISSN 1650-0377

Postal address:

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

Internet:

http://www.math.su.se

Mathematical Statistics
Stockholm University
Research Report **2015:1**,
http://www.math.su.se

Direct Shrinkage Estimation of Large Dimensional Precision Matrix

January 2015

Taras Bodnar[a,1], Arjun K. Gupta[b] and Nestor Parolya[c]

[a] *Department of Mathematics, Stockholm University, SE-10691 Stockholm, Sweden*
[b] *Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403, USA*
[c] *Institute of Empirical Economics, Leibniz University of Hannover, D-30167 Hannover, Germany*

Abstract

In this work we construct an optimal shrinkage estimator for the precision matrix in high dimensions. We consider the general asymptotics when the number of variables $p \to \infty$ and the sample size $n \to \infty$ so that $p/n \to c \in (0, +\infty)$. The precision matrix is estimated directly, without inverting the corresponding estimator for the covariance matrix. The recent results from the random matrix theory allow us to find the asymptotic deterministic equivalents of the optimal shrinkage intensities and estimate them consistently. The resulting distribution-free estimator has almost surely the minimum Frobenius loss. Additionally, we prove that the Frobenius norms of the inverse and of the pseudo-inverse sample covariance matrices tend almost surely to deterministic quantities and estimate them consistently. At the end, a simulation is provided where the suggested estimator is compared with the estimators for the precision matrix proposed in the literature. The optimal shrinkage estimator shows significant improvement and robustness even for non-normally distributed data.

AMS 2010 subject classifications: 60B20, 62H12, 62G20, 62G30
Keywords: large-dimensional asymptotics, random matrix theory, precision matrix estimation.

[1]Corresponding author: Taras Bodnar. E-mail address: taras.bodnar@math.su.se. The first author is partly supported by the German Science Foundation (DFG) via the Research Unit 1735 "Structural Inference in Statistics: Adaptation and Efficiency".

1 Introduction

The estimation of the covariance matrix, as well as its inverse (the precision matrix), plays an important role in many disciplines from finance and genetics to wireless communications and engineering. In fact, having a suitable estimator for the precision matrix we are able to construct a good estimator for different types of optimal portfolios (see, Markowitz (1952), Elton et al. (2009)). Similarly, in the array processing, the beamformer or the so-called minimum variance distortionless response spatial filter is defined in terms of the precision matrix (see, e.g., Van Trees (2002)). In practice, however, the true precision matrix is unknown and a feasible estimator, constructed from data, must be used.

If the number of variables p is much smaller than the sample size n we can use the sample estimator which is biased but a consistent estimator for the precision matrix (see, e.g., Bai and Shi (2011)). This case is known in the multivariate statistics as the "standard asymptotics" (see, Le Cam and Yang (2000)). There are many findings on the estimation of the precision matrix when a particular distribution assumption is imposed. For example, the estimation of the precision matrix under the multivariate normal distribution was considered by Krishnamoorthy and Gupta (1989), Gupta and Ofori-Nyarko (1994, 1995a, 1995b), Kubokawa (2005) and Tsukuma and Konno (2006). The results in the case of multivariate Pearson type II distribution as well as the multivariate elliptically contoured stable distribution are obtained by Sarr and Gupta (2009) as well as by Bodnar and Gupta (2011) and Gupta et al. (2013), respectively.

Unfortunately, in practice p is often comparable in size to n or even is greater than n, i.e., we are in the situation when both the sample size n and the dimension p tend to infinity but their ratio keeps (tends to) a positive constant. This case often arises in finance when the number of assets is comparable or even greater than the number of observations for each asset. Similarly, in genetics, the data set can be huge comparable to the number of patients. Both examples illustrate the importance of the results obtained for $p, n \to \infty$.

We deal with this type of asymptotics, called the "large dimensional asymptotics" and also known as the "Kolmogorov asymptotics", in the present paper. More precisely, it is assumed that the dimension $p \equiv p(n)$ is a function of the sample size n and $p/n \to c \in (0, +\infty)$ as $n \to \infty$. This general type of asymptotics was intensively studied by several authors (see, Girko (1990, 1995), Bühlmann and van Geer (2011) etc.). In this asymptotics the usual estimators for the precision matrix perform poorly and are not consistent anymore. There are some techniques which can be used to handle the problem. Assuming that the covariance (precision) matrix has a sparse structure, significant improvements have already been achieved (see, Cai et al. (2011), Cai and Shen (2011), Cai and Zhou (2012)). For the low-rank covariance matrices see the work of Rohde and Tsybakov (2011). An interesting nonparanormal graphic model was recently proposed by Xue and Zou (2012). Also, in order to estimate the large dimensional covariance matrix the method of block thresholding can be applied (see, Cai and Yuan (2012)). If the covariance matrix has a factor structure then the progress has been made by Fan et al. (2008).

However, if neither the assumption about the structure of covariance (precision) matrix nor about a particular distribution is imposed, not many results are known in the literature which are based on the shrinkage estimators in high-dimensional setting (cf., Ledoit and Wolf (2004), Kubokawa and Srivastava (2008), Bodnar et al. (2014) and Wang et al. (2014)). The shrinkage estimator was first developed by Stein (1956) and forms a linear combination of the sample estimator and some target. The corresponding shrinkage coefficients are often called shrinkage intensities. Ledoit and Wolf (2004) proposed to shrink the sample covariance matrix to the identity matrix and showed that the resulting estimator is well-behaved in large dimensions.

This estimator is called the linear shrinkage estimator because it shrinks the eigenvalues of the sample covariance matrix linearly. Recently, Bodnar et al. (2014) proposed a generalization of the linear shrinkage estimator, where the shrinkage target was chosen to be an arbitrary nonrandom matrix and they showed the almost sure convergence of the derived estimator to its oracle.

The aim of our paper is to construct a feasible estimator for the precision matrix using the linear shrinkage technique and the random matrix theory. In contrast to well-known procedures, we shrink the inverse of the sample covariance matrix itself instead of shrinking the sample covariance matrix and then inverting it. The direct shrinkage estimation of the precision matrix can be used in several important practical situations where the application of the inverse of the shrinkage estimator of the covariance matrix does not perform well. For instance, this could happen when the data generating process follows a factor model which is very popular in economics and finance (cf., Bai (2003), Fan et al. (2008), Fan et al. (2012), Fan et al. (2013)). In this case the largest eigenvalue of the covariance matrix is of order p and, consequently, the inverse of the linear shrinkage estimator for the covariance matrix does not work well. In the case when $c > 1$ the pseudo inverse of the sample covariance matrix is taken. The recent results from the random matrix theory allow us to find the asymptotics of the optimal shrinkage intensities and estimate them consistently.

The random matrix theory is a very fast growing branch of the probability theory with many applications in statistics. It studies the asymptotic behavior of the eigenvalues of the different random matrices under general asymptotics (see, e.g., Anderson et al. (2010), Bai and Silverstein (2010)). The asymptotic behavior of the functionals of the sample covariance matrices was studied by Mačenko and Pastur (1967), Yin (1986), Girko and Gupta (1994, 1996a, 1996b), Silverstein (1995), Bai et al. (2007), Bai and Silverstein (2010), Rubio and Mestre (2011) etc.

We extend these results in the present paper by establishing the almost sure convergence of the optimal shrinkage intensities and the Frobenius norm of the inverse sample covariance matrix. Moreover, we construct a general linear shrinkage estimator for the precision matrix which has *almost surely* the smallest Frobenius loss when both the dimension p and the sample size n increase together and $p/n \to c \in (0, +\infty)$ as $n \to \infty$.

The suggested approach can potentially be applied in functional data analysis (cf., Ramsay and Silverman (2005), Ferraty and Vieu (2006)). For instance, Ferraty et al. (2010) pointed out that functional data can be seen as a special case of a high-dimensional vector. The estimation of the covariance (precision) matrix of this high-dimensional vector can be used in determining the prediction for the dependent variable as well as the corresponding predictive design points.

The rest of the paper is organized as follows. In Section 2 we present some preliminary results from the random matrix theory and formulate the assumptions used throughout the paper. In Section 3 we construct the *oracle* linear shrinkage estimator for the precision matrix and verify the main asymptotic results about the shrinkage intensities and the Frobenius norm of the inverse and pseudo-inverse sample covariance matrices. Section 4 is dedicated to the *bona fide* linear shrinkage estimator for the precision matrix while Section 5 contains the results of the simulation study. Here, the performance of the derived estimator is compared with other known estimators for the large dimensional precision matrices. Section 6 includes the summary, while the proofs of the theorems are presented in the appendix (Section 7).

2 Assumptions and notations

The "large dimensional asymptotics" or "Kolmogorov asymptotics" include $\frac{p}{n} \to c \in (0, +\infty)$ as both the number of variables $p \equiv p(n)$ and the sample size n tend to infinity. In this case the traditional sample estimator performs poorly or very poorly. The inverse of the sample covariance matrix $\mathbf{S}_n^{-1}$ is biased, inconsistent for $\frac{p}{n} \to c > 0$ as $n \to \infty$ and it does not exist for $c > 1$. For example, under the normality assumption $\mathbf{S}_n^{-1}$ has an inverse Wishart distribution if $c < 1$ and (cf. Gupta and Nagar (2000))

$$E(\mathbf{S}_n^{-1}) = \frac{n}{n-p-2}\mathbf{\Sigma}_n^{-1}$$

In particular, for $p = n/2 + 2$ we have that $c = 1/2$ and $E(\mathbf{S}_n^{-1}) = 2\mathbf{\Sigma}_n^{-1}$. In general, as c increases the sample estimator of the precision matrix becomes worse.

We use the following notations in the paper:

- $\mathbf{\Sigma}_n$ stands for the covariance matrix, $\mathbf{S}_n$ denotes the corresponding sample covariance matrix.[2] The population covariance matrix $\mathbf{\Sigma}_n$ is a nonrandom p-dimensional positive definite matrix.

- $||\boldsymbol{A}||_F^2 = \text{tr}(\boldsymbol{A}\boldsymbol{A}')$ denotes the Frobenius norm of a square matrix $\boldsymbol{A}$, $||\boldsymbol{A}||_{tr} = \text{tr}\left[(\boldsymbol{A}\boldsymbol{A}')^{1/2}\right]$ stands for its trace norm, while $||\boldsymbol{A}||_2$ is the spectral norm.

- The pairs $(\tau_i, \boldsymbol{\nu}_i)$ for $i = 1, \ldots, p$ denote the collection of eigenvalues and the corresponding orthonormal eigenvectors of the covariance matrix $\mathbf{\Sigma}_n$.

- $H_n(t)$ is the empirical distribution function (e.d.f.) of the eigenvalues of $\mathbf{\Sigma}_n$, i.e.,

$$H_n(t) = \frac{1}{p}\sum_{i=1}^{p} \mathbb{1}_{\{\tau_i \le t\}} \tag{2.1}$$

  where $\mathbb{1}_{\{\cdot\}}$ is the indicator function.

- Let $\mathbf{X}_n$ be a $p \times n$ matrix which consists of independent and identically distributed (i.i.d.) real random variables with zero mean and unit variance. The observation matrix is defined as

$$\mathbf{Y}_n = \mathbf{\Sigma}_n^{\frac{1}{2}}\mathbf{X}_n. \tag{2.2}$$

  Only the matrix $\mathbf{Y}_n$ is observable. We know neither $\mathbf{X}_n$ nor $\mathbf{\Sigma}_n$ itself.

- The pairs $(\lambda_i, \mathbf{u}_i)$ for $i = 1, \ldots, p$ are the eigenvalues and the corresponding orthonormal eigenvectors of the sample covariance matrix[3]

$$\mathbf{S}_n = \frac{1}{n}\mathbf{Y}_n\mathbf{Y}_n' = \frac{1}{n}\mathbf{\Sigma}_n^{\frac{1}{2}}\mathbf{X}_n\mathbf{X}_n'\mathbf{\Sigma}_n^{\frac{1}{2}}. \tag{2.3}$$

- Similarly, the (e.d.f.) of the eigenvalues of the sample covariance matrix $\mathbf{S}_n$ is defined as

$$F_n(\lambda) = \frac{1}{p}\sum_{i=1}^{p} \mathbb{1}_{\{\lambda_i \le \lambda\}} \quad \forall\ \lambda \in \mathbb{R}. \tag{2.4}$$

[2]Since the dimension $p \equiv p(n)$ is a function of the sample size n, the covariance matrix $\mathbf{\Sigma}_n$ also depends on n via $p(n)$. That is why we make use of the subscript n for all of the considered objects in order to emphasize this fact and to simplify the notation in the paper.

[3]The sample mean vector $\bar{\mathbf{x}}$ was omitted because the 1-rank matrix $\bar{\mathbf{x}}\bar{\mathbf{x}}'$ does not influence the asymptotic behavior of the spectrum of sample covariance matrix (see, Bai and Silverstein (2010), Theorem A.44).

- In order to handle the case when $c > 1$ we introduce the dual sample covariance matrix defined as

$$\bar{\mathbf{S}}_n = \frac{1}{n}\mathbf{Y}_n'\mathbf{Y}_n = \frac{1}{n}\mathbf{X}_n'\mathbf{\Sigma}_n\mathbf{X}_n \tag{2.5}$$

with the corresponding (e.d.f.) defined by

$$\bar{F}_n(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{\lambda_i \leq \lambda\}} \quad \forall\ \lambda \in \mathbb{R}\,. \tag{2.6}$$

Note that the matrix $\bar{\mathbf{S}}_n$ has the same nonzero eigenvalues as $\mathbf{S}_n$, they differ only in $|p-n|$ zero eigenvalues.

The main assumptions, which we mention throughout the paper, are as follows

(A1) We assume that $H_n(t)$ converges to a limit $H(t)$ at all points of continuity of H.

(A2) The elements of the matrix $\mathbf{X}_n$ have uniformly bounded $4+\varepsilon$, $\varepsilon > 0$ moments.

(A3) For all n large enough there exists the compact interval $[h_0, h_1]$ in $(0, +\infty)$ which contains the support of H_n.

All of these assumptions are quite general and are satisfied in many practical situations. The assumption (A1) is essential to prove the Marčhenko-Pastur equation (see, e.g., Silverstein (1995)) which is used for studying the asymptotic behavior of the spectrum of general random matrices (see, e.g., Bai and Silverstein (2010)). The fourth moment is needed for the proof of Theorem 3.2 and Theorem 3.3. The assumption (A3) ensures that both the matrix $\mathbf{\Sigma}_n$ and its inverse $\mathbf{\Sigma}_n^{-1}$ have uniformly bounded spectral norms at infinity. It means that $\mathbf{\Sigma}_n$ has the uniformly bounded maximum eigenvalue and its minimum eigenvalue is greater than zero. Rubio et al. (2012) pointed out that (A2) and (A3) are only some technical conditions which can be further violated.

In order to investigate the (e.d.f) $F_n(\lambda)$ the Stieltjes transform is used. For nondecreasing function with bounded variation G the Stieltjes transform is defined as

$$\forall z \in \mathbb{C}^+ \quad m_G(z) = \int_{-\infty}^{+\infty}\frac{1}{\lambda - z}dG(\lambda)\,. \tag{2.7}$$

In our notation $\mathbb{C}^+ = \{z \in \mathbb{C} : \mathbf{Im}(z) > 0\}$ is the half-plane of complex numbers with strictly positive imaginary part and any complex number is defined as $z = \mathbf{Re}(z) + i\mathbf{Im}(z)$. More about the Stieltjes transform and its properties can be found in Silverstein (2009).

The Stieltjes transform of the sample (e.d.f.) $F_n(\lambda)$ for all $z \in \mathbb{C}^+$ is given by

$$m_{F_n}(z) \;=\; \frac{1}{p}\sum_{i=1}^{p}\int_{-\infty}^{+\infty}\frac{1}{\lambda - z}\delta(\lambda - \lambda_i)d\lambda = \frac{1}{p}\mathrm{tr}\{(\mathbf{S}_n - z\mathbf{I})^{-1}\} \tag{2.8}$$

where $\mathbf{I}$ is a suitable identity matrix and $\delta(\cdot)$ is the Dirac delta function.

3 Optimal linear shrinkage estimator for the precision matrix

3.1 Case $c < 1$

In this section we construct an optimal linear shrinkage estimator for the precision matrix under high-dimensional asymptotics. The estimator is an *oracle* one, i.e., it depends on unknown quantities. The corresponding *bona fide* estimator is given in Section 4. We use a procedure similar to Bodnar et al. (2014) where the optimal linear shrinkage estimator for the covariance matrix was constructed. The general linear shrinkage estimator of the precision matrix $\mathbf{\Sigma}_n^{-1}$ for $c < 1$ is given by

$$\widehat{\mathbf{\Pi}}_{GSE} = \alpha_n \mathbf{S}_n^{-1} + \beta_n \mathbf{\Pi}_0 \quad \text{with} \ \sup_p \frac{1}{p}||\mathbf{\Pi}_0||_F^2 \le M\,. \tag{3.1}$$

Note that we need the condition $c < 1$ to keep the sample covariance matrix $\mathbf{S}_n$ invertible. The assumption[4] that the target matrix $\mathbf{\Pi}_0$ has a uniformly bounded normalized trace norm, i.e., there exists $M > 0$ such that $\sup_p \frac{1}{p}||\mathbf{\Pi}_0||_F^2 \le M$, is rather general and it is actually needed to keep the coefficient β_n bounded for large dimensions p. This condition can be replaced with an equivalent assumption on β_n. Note that the target matrix can also be random but independent of $\mathbf{Y}_n$. In practice, $\mathbf{\Pi}_0 = \mathbf{I}$ is used when no information about the precision matrix is available. On the other hand, the information about the structure of the precision matrix can be included into $\mathbf{\Pi}_0$, like, e.g., sparsity. In order to present the results in the most general case, throughout this section we assume only that $\mathbf{\Pi}_0$ satisfies $\sup_p \frac{1}{p}||\mathbf{\Pi}_0||_F^2 \le M$. The choice of the target matrix $\mathbf{\Pi}_0$ is not treated in the paper in detail and it is left for future research.[5]

Our aim is now to find the optimal shrinkage intensities which minimize the Frobenius-norm loss for a given nonrandom target matrix $\mathbf{\Pi}_0$ expressed as

$$L_F^2 = ||\widehat{\mathbf{\Pi}}_{GSE} - \mathbf{\Sigma}_n^{-1}||_F^2 = ||\mathbf{\Sigma}_n^{-1}||_F^2 + ||\widehat{\mathbf{\Pi}}_{GSE}||_F^2 - 2\text{tr}\left(\widehat{\mathbf{\Pi}}_{GSE}\mathbf{\Sigma}_n^{-1}\right)\,. \tag{3.2}$$

As a result, using (3.1) the following optimization problem has to be solved

$$\alpha_n^2||\mathbf{S}_n^{-1}||_F^2 + 2\alpha_n\beta_n\text{tr}(\mathbf{S}_n^{-1}\mathbf{\Pi}_0) + \beta_n^2||\mathbf{\Pi}_0||_F^2 - 2\alpha_n\text{tr}(\mathbf{S}_n^{-1}\mathbf{\Sigma}_n^{-1}) - 2\beta_n\text{tr}(\mathbf{\Sigma}_n^{-1}\mathbf{\Pi}_0) \longrightarrow \min$$
$$\text{with respect to } \alpha_n \text{ and } \beta_n\,.$$

Next, taking the derivatives of L_F^2 with respect to α_n and β_n and setting them equal to zero we get

$$\frac{\partial L_F^2}{\partial \alpha_n} = \alpha_n||\mathbf{S}_n^{-1}||_F^2 + \beta_n\text{tr}(\mathbf{S}_n^{-1}\mathbf{\Pi}_0) - \text{tr}(\mathbf{S}_n^{-1}\mathbf{\Sigma}_n^{-1}) = 0\,, \tag{3.3}$$

$$\frac{\partial L_F^2}{\partial \beta_n} = \alpha_n\text{tr}(\mathbf{S}_n^{-1}\mathbf{\Pi}_0) + \beta_n||\mathbf{\Pi}_0||_F^2 - \text{tr}(\mathbf{\Sigma}_n^{-1}\mathbf{\Pi}_0) = 0\,. \tag{3.4}$$

The Hessian of the L_F^2 has the form

$$\mathbf{H} = \begin{pmatrix} ||\mathbf{S}_n^{-1}||_F^2 & \text{tr}(\mathbf{S}_n^{-1}\mathbf{\Pi}_0) \\ \text{tr}(\mathbf{S}_n^{-1}\mathbf{\Pi}_0) & ||\mathbf{\Pi}_0||_F^2 \end{pmatrix} \tag{3.5}$$

which is always positive definite, since

$$\begin{aligned} \det(\mathbf{H}) &= ||\mathbf{S}_n^{-1}||_F^2||\mathbf{\Pi}_0||_F^2 - (\text{tr}(\mathbf{S}_n^{-1}\mathbf{\Pi}_0))^2 \\ &\ge ||\mathbf{S}_n^{-1}||_F^2||\mathbf{\Pi}_0||_F^2 - ||\mathbf{S}_n^{-1}||_2^2(\text{tr}(\mathbf{\Pi}_0))^2 \overset{Jensen}{\ge} (||\mathbf{S}_n^{-1}||_F^2 - ||\mathbf{S}_n^{-1}||_2^2)||\mathbf{\Pi}_0||_F^2 > 0\,, \end{aligned} \tag{3.6}$$

where the last inequality in (3.6) is well-known (see, e.g., Horn and Johnson (1985)).

[4]The similar assumption is presented by Bodnar et al. (2014) but with $\sup_p ||\mathbf{\Pi}_0||_F^2 \le M$. Note that the assumption $\sup_p 1/p||\mathbf{\Pi}_0||_F^2 \le M$ is more general.

[5]In practice, however, one has to be careful with the choice of the target matrix $\mathbf{\Pi}_0$. If it is close in some sense to $\mathbf{S}_n^{-1}$, negative shrinkage intensities might occur.

Thus, the optimal α_n^* and β_n^* are given by

$$\alpha_n^* = \frac{\text{tr}(\mathbf{S}_n^{-1}\mathbf{\Sigma}_n^{-1})||\mathbf{\Pi}_0||_F^2 - \text{tr}(\mathbf{\Sigma}_n^{-1}\mathbf{\Pi}_0)\text{tr}(\mathbf{S}_n^{-1}\mathbf{\Pi}_0)}{||\mathbf{S}_n^{-1}||_F^2||\mathbf{\Pi}_0||_F^2 - \left(\text{tr}(\mathbf{S}_n^{-1}\mathbf{\Pi}_0)\right)^2}\,, \tag{3.7}$$

$$\beta_n^* = \frac{\text{tr}(\mathbf{\Sigma}_n^{-1}\mathbf{\Pi}_0)||\mathbf{S}_n^{-1}||_F^2 - \text{tr}(\mathbf{S}_n^{-1}\mathbf{\Sigma}_n^{-1})\text{tr}(\mathbf{S}_n^{-1}\mathbf{\Pi}_0)}{||\mathbf{S}_n^{-1}||_F^2||\mathbf{\Pi}_0||_F^2 - \left(\text{tr}(\mathbf{S}_n^{-1}\mathbf{\Pi}_0)\right)^2}\,. \tag{3.8}$$

Now, we formulate our first main result in Theorem 3.1 which states that the normalized Frobenius norm of the inverse sample covariance matrix $1/p||\mathbf{S}_n^{-1}||_F^2$ tends almost surely to a nonrandom quantity.

Theorem 3.1. *Assume that (A1) and (A3) hold and* $\frac{p}{n} \to c \in (0,1)$ *for* $n \to \infty$*. Then the normalized Frobenius norm of the inverse sample covariance matrix* $\psi_n = \frac{1}{p}||\mathbf{S}_n^{-1}||_F^2$ *almost surely tends to a nonrandom* ψ *which is given by*

$$\psi = \frac{1}{(1-c)^2}\int_{-\infty}^{+\infty}\frac{dH(\tau)}{\tau^2} + \frac{c}{(1-c)^3}\left(\int_{-\infty}^{+\infty}\frac{dH(\tau)}{\tau}\right)^2\,. \tag{3.9}$$

The proof is given in the Appendix. Theorem 3.1 presents an important result which indicates that the Frobenius norm of the inverse sample covariance matrix is asymptotically nonrandom as well as it depends on H and concentration ratio c only. Moreover, Theorem 3.1 gives us an intuitive hint how to find the asymptotic equivalent representation of $||\mathbf{S}_n^{-1}||_F^2$. The corresponding result is presented in Theorem 3.2.

Theorem 3.2. *Let the assumptions (A1)-(A3) hold and* $\frac{p}{n} \to c \in (0,1)$*. Then as* $n \to \infty$*,*

$$\frac{1}{p}\left|||\mathbf{S}_n^{-1}||_F^2 - \left(\frac{1}{(1-c)^2}||\mathbf{\Sigma}_n^{-1}||_F^2 + \frac{c}{p(1-c)^3}||\mathbf{\Sigma}_n^{-1}||_{tr}^2\right)\right| \underset{a.s.}{\longrightarrow} 0\,. \tag{3.10}$$

Additionally, for the quantity $tr(\mathbf{S}_n^{-1}\mathbf{\Theta})$ *with a symmetric positive definite matrix* $\mathbf{\Theta}$ *which has uniformly bounded trace norm as* $n \to \infty$*,*

$$\left|tr(\mathbf{S}_n^{-1}\mathbf{\Theta}) - \frac{1}{1-c}tr(\mathbf{\Sigma}_n^{-1}\mathbf{\Theta})\right| \underset{a.s.}{\longrightarrow} 0 \quad for \quad \frac{p}{n} \to c \in (0,1)\,. \tag{3.11}$$

The theorem is proved in the Appendix. Theorem 3.2 provides us the asymptotic behavior of the Frobenius norm of the inverse sample covariance matrix and of the functional $\text{tr}(\mathbf{S}_n^{-1}\mathbf{\Theta})$. It shows that the consistent estimator for the Frobenius norm of the precision matrix under the general asymptotics is not equal to its sample counterpart. Using Theorem 3.2 we can easily determine the asymptotic bias of the sample estimator which consists of the two types of biases. The multiplicative bias is violated by multiplying $||\mathbf{S}_n^{-1}||_F^2$ by $(1-c)^2$. After that, the additive bias is dealt by subtracting $\frac{c}{p(1-c)}||\mathbf{\Sigma}_n^{-1}||_{tr}^2$. The sample estimator of the functional $\text{tr}(\mathbf{S}_n^{-1}\mathbf{\Theta})$ is also not a consistent estimator for $\text{tr}(\mathbf{\Sigma}_n^{-1}\mathbf{\Theta})$. The consistent estimator is obtained by multiplying $\text{tr}(\mathbf{S}_n^{-1}\mathbf{\Theta})$ by the constant $(1-c)$.

Results similar to those given in Theorem 3.1 and Theorem 3.2 are also available for the estimation of the population covariance matrix (cf. Bodnar et al. (2014)). However, in the case of the covariance matrix, the sample estimator for the Frobenius norm possesses only the additive bias $\frac{c}{p}\text{tr}(||\mathbf{\Sigma}_n^{-1}||_{tr})$, while $\text{tr}(\mathbf{S}_n\mathbf{\Theta})$ is a consistent estimator for $\text{tr}(\mathbf{\Sigma}_n\mathbf{\Theta})$.

Next, we show that the optimal shrinkage intensities α_n^* and β_n^* are almost surely asymptotic equivalent to nonrandom quantities α^* and β^* under the large-dimensional asymptotics $\frac{p}{n} \to c \in (0,1)$.

Corollary 3.1. *Assume that (A1)-(A3) hold and $\frac{p}{n} \to c \in (0,1)$ for $n \to \infty$. Then for the optimal shrinkage intensities α_n^* and β_n^**

$$|\alpha_n^* - \alpha^*| \longrightarrow 0 \quad a.\ s.\,, \tag{3.12}$$

where

$$\alpha^* = (1-c)\frac{||\mathbf{\Sigma}_n^{-1}||_F^2||\mathbf{\Pi}_0||_F^2 - \left(tr(\mathbf{\Sigma}_n^{-1}\mathbf{\Pi}_0)\right)^2}{\left(||\mathbf{\Sigma}_n^{-1}||_F^2 + \frac{c}{p(1-c)}||\mathbf{\Sigma}_n^{-1}||_{tr}^2\right)||\mathbf{\Pi}_0||_F^2 - \left(tr(\mathbf{\Sigma}_n^{-1}\mathbf{\Pi}_0)\right)^2} \tag{3.13}$$

and

$$|\beta_n^* - \beta^*| \longrightarrow 0 \quad a.\ s.\,, \tag{3.14}$$

with

$$\beta^* = \frac{tr(\mathbf{\Sigma}_n^{-1}\mathbf{\Pi}_0)}{||\mathbf{\Pi}_0||_F^2}\left(1 - \frac{\alpha^*}{1-c}\right) \tag{3.15}$$

Note that both the asymptotic optimal intensities α^* and β^* are always positive as well as $\alpha^* \in (0, 1-c)$ due to inequality (3.6) and $c \in (0,1)$. Using these results we are immediately able to estimate α^*, β^* consistently which is shown in Section 4.

3.2 Case $c > 1$

In this subsection we deal with the problem of the estimation of the precision matrix when the dimension p is greater than the sample size n, i.e., $c > 1$. This case is very difficult to handle because of the loss of information as c becomes greater than one. Moreover, the sample covariance matrix $\mathbf{S}_n$ is not invertible and thus the estimator $\mathbf{S}_n^{-1}$ must be replaced by a suitable one. This is usually done by using the generalized inverse matrix $\mathbf{S}_n^+$ instead of $\mathbf{S}_n^{-1}$. In this case the general shrinkage estimator has the form

$$\widehat{\mathbf{\Pi}}_{GSE} = \tilde{\alpha}_n\mathbf{S}_n^+ + \tilde{\beta}_n\mathbf{\Pi}_0 \quad \text{with } \sup_p \frac{1}{p}||\mathbf{\Pi}_0||_F^2 \le M\,. \tag{3.16}$$

The optimal shrinkage intensities $\tilde{\alpha}_n^*$ and $\tilde{\beta}_n^*$ are determined following the procedure of Section 3.1. They are given by

$$\tilde{\alpha}_n^* = \frac{\text{tr}(\mathbf{S}_n^+\mathbf{\Sigma}_n^{-1})||\mathbf{\Pi}_0||_F^2 - \text{tr}(\mathbf{\Sigma}_n^{-1}\mathbf{\Pi}_0)\text{tr}(\mathbf{S}_n^+\mathbf{\Pi}_0)}{||\mathbf{S}_n^+||_F^2||\mathbf{\Pi}_0||_F^2 - \left(\text{tr}(\mathbf{S}_n^+\mathbf{\Pi}_0)\right)^2}\,, \tag{3.17}$$

$$\tilde{\beta}_n^* = \frac{\text{tr}(\mathbf{\Sigma}_n^{-1}\mathbf{\Pi}_0)||\mathbf{S}_n^+||_F^2 - \text{tr}(\mathbf{S}_n^+\mathbf{\Sigma}_n^{-1})\text{tr}(\mathbf{S}_n^+\mathbf{\Pi}_0)}{||\mathbf{S}_n^+||_F^2||\mathbf{\Pi}_0||_F^2 - \left(\text{tr}(\mathbf{S}_n^+\mathbf{\Pi}_0)\right)^2}\,. \tag{3.18}$$

In Theorem 3.3 we derive the asymptotic properties of two quantities used in (3.17) and (3.18), namely $\text{tr}(\mathbf{\Theta}\mathbf{S}_n^+)$ and $||\mathbf{S}_n^+||_F^2$.

Theorem 3.3. *Let the assumptions (A1)-(A3) hold and* $\frac{p}{n} \to c \in (1, +\infty)$*. Then as* $n \to \infty$*,*

$$\left| \frac{1}{p} ||\mathbf{S}_n^{+}||_F^2 - c^{-1} x'(0) \right| \xrightarrow[a.s.]{} 0, \quad \textit{where } x'(0) = \frac{1}{\frac{1}{x^2(0)} - \frac{c}{p} tr\left[\left(\mathbf{\Sigma}_n^{-1} + x(0)\mathbf{I}\right)^{-2}\right]} \tag{3.19}$$

and $x(0)$ *is the unique solution of the equation*

$$\frac{1}{x(0)} = \frac{c}{p} tr\left[\left(\mathbf{\Sigma}_n^{-1} + x(0)\mathbf{I}\right)^{-1}\right] \tag{3.20}$$

Additionally, for the quantity $tr(\mathbf{\Theta}\mathbf{S}_n^{+})$ *with a symmetric positive definite matrix* $\mathbf{\Theta}$ *which has uniformly bounded spectral norm, as* $n \to \infty$*,*

$$\left| \frac{1}{p} tr(\mathbf{\Theta}\mathbf{S}_n^{+}) - c^{-1} y(\mathbf{\Theta}) \right| \xrightarrow[a.s.]{} 0 \quad \textit{for} \quad \frac{p}{n} \to c \in (1, +\infty)\,, \tag{3.21}$$

where $y(\mathbf{\Theta})$ *is the solution of*

$$\frac{1}{y(\mathbf{\Theta})} = \frac{c}{p} tr\left[\left(\mathbf{\Sigma}_n^{-1/2}\mathbf{\Theta}\mathbf{\Sigma}_n^{-1/2} + y(\mathbf{\Theta})\mathbf{I}\right)^{-1}\right] \tag{3.22}$$

The proof of Theorem 3.3 is given in the Appendix. The results of Theorem 3.3 show that using the generalized inverse technique it is not clear how to estimate the functionals of $\mathbf{\Sigma}_n^{-1}$ consistently. The asymptotic values obtained in Theorem 3.3 are far away from the desired ones. In order to correct these biases, we need to solve the non-linear equations (3.20) and (3.22), respectively, which appears to be a difficult task. Finally, we notice, that the quantities $x(0)$ and $x'(0)$, however, can be estimated consistently using Theorem 3.3.

In an important special case when the matrix $\mathbf{\Theta} = \boldsymbol{\xi}\boldsymbol{\eta}'$ for some $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ with bounded Euclidean norms we get the following result summarized in Proposition 3.1 which is proved in the Appendix.

Proposition 3.1. *Under the assumptions of Theorem 3.3 and* $\mathbf{\Theta} = \boldsymbol{\xi}\boldsymbol{\eta}'$ *for some* $\boldsymbol{\xi}$ *and* $\boldsymbol{\eta}$ *with bounded Euclidean norms it holds*

$$\frac{1}{p}\left| \boldsymbol{\eta}'\mathbf{S}_n^{+}\boldsymbol{\xi} - \frac{c^{-1}}{c-1}\boldsymbol{\eta}'\mathbf{\Sigma}_n^{-1}\boldsymbol{\xi} \right| \xrightarrow[a.s.]{} 0 \quad \textit{for} \quad \frac{p}{n} \to c \in (1, +\infty)\,, \tag{3.23}$$

It is remarkable to note that the results of Proposition 3.1 are very similar to those presented in Theorem 3.2 if $\mathbf{\Theta} = \boldsymbol{\xi}\boldsymbol{\xi}'$. Here, the constant $1 - c$ need to be replaced by $c(c-1)$.

Next we use the asymptotic results of Theorem 3.3 for finding the asymptotic equivalents to the optimal shrinkage intensities $\tilde{\alpha}_n^*$ and $\tilde{\beta}_n^*$ given in (3.17) and (3.18), respectively.

Corollary 3.2. *Assume that (A1)-(A3) hold and* $\frac{p}{n} \to c \in (1, +\infty)$ *for* $n \to \infty$*. Then for the optimal shrinkage intensities* α_n^* *and* β_n^* *from (3.17) and (3.18) holds*

$$|\alpha_n^* - \alpha^*| \longrightarrow 0 \quad a.\ s.\,, \tag{3.24}$$

where

$$\alpha^* = \frac{y(\mathbf{\Sigma}_n^{-1})||\mathbf{\Pi}_0||_F^2 - y(\mathbf{\Pi}_0) tr(\mathbf{\Sigma}_n^{-1}\mathbf{\Pi}_0)}{x'(0)||\mathbf{\Pi}_0||_F^2 - c^{-1}y^2(\mathbf{\Pi}_0)} \tag{3.25}$$

and

$$|\beta_n^* - \beta^*| \longrightarrow 0 \quad a.\ s.\,, \tag{3.26}$$

with

$$\beta^* = \frac{tr(\mathbf{\Sigma}_n^{-1}\mathbf{\Pi}_0)x'(0) - y(\mathbf{\Sigma}_n^{-1})y(\mathbf{\Pi}_0)}{cx'(0)||\mathbf{\Pi}_0||_F^2 - y^2(\mathbf{\Pi}_0)} \tag{3.27}$$

Even if the target matrix $\mathbf{\Pi}_0$ is chosen as a one-rank matrix, i.e., $\mathbf{\Pi}_0 = \boldsymbol{\xi}\boldsymbol{\eta}'$, we are not able to provide consistent estimates for α^* and β^* without an additional assumption imposed on $\mathbf{\Sigma}_n$. One of possible assumptions for which α^* and β^* are consistently estimable is $\mathbf{\Sigma}_n = \sigma\mathbf{I}$ as illustrated in Corollary 3.4 below. If $\mathbf{\Sigma}_n = \sigma\mathbf{I}$, then for $\frac{1}{p}||\mathbf{S}_n^+||_F^2$ and $\frac{1}{p}\text{tr}(\mathbf{S}_n^+)$ we get

Corollary 3.3. *Under the assumptions of Theorem 3.3 assume additionally that $\mathbf{\Sigma}_n = \sigma\mathbf{I}$. Then it holds as $n \to \infty$*

$$\left|\frac{1}{p}||\mathbf{S}_n^+||_F^2 - \frac{\sigma^{-2}}{(c-1)^3}\right| \underset{a.s.}{\longrightarrow} 0\,, \tag{3.28}$$

Additionally, for the quantity $tr(\mathbf{S}_n^+)$ as $n \to \infty$ the norm

$$\left|\frac{1}{p}tr(\mathbf{S}_n^+) - \frac{c^{-1}}{(c-1)}\sigma^{-1}\right| \underset{a.s.}{\longrightarrow} 0 \quad for \quad \frac{p}{n} \to c \in (1, +\infty)\,. \tag{3.29}$$

The proof of Corollary 3.3 is based on the fact that the equation (3.20) has the explicit solution $x(0) = \dfrac{\sigma^{-1}}{c-1}$ if $\mathbf{\Sigma}_n = \sigma\mathbf{I}$. The rest calculations are only technical ones. It is interesting to note that the result of Corollary 3.3 coincides with the corresponding one of Theorem 3.2 for $c < 1$ if $\mathbf{\Sigma}_n = \sigma\mathbf{I}$.

Next we apply Corollary 3.3 with $\mathbf{\Sigma}_n = \sigma\mathbf{I}$ and $\mathbf{\Pi}_0 = \mathbf{I}$ to construct the asymptotic equivalents to the optimal shrinkage intensities $\tilde{\alpha}_n^*$ and $\tilde{\beta}_n^*$ given in (3.17) and (3.18), respectively.

Corollary 3.4. *Assume that (A1)-(A3) hold, $\mathbf{\Sigma}_n = \sigma\mathbf{I}$, $\mathbf{\Pi}_0 = \mathbf{I}$ and $\frac{p}{n} \to c \in (1, +\infty)$ for $n \to \infty$. Then for the optimal shrinkage intensities α_n^* and β_n^* from (3.17) and (3.18) holds*

$$\tilde{\alpha}_n^* \longrightarrow 0 \quad a.\ s. \quad and \quad \tilde{\beta}_n^* \longrightarrow \sigma^{-1} \quad a.\ s.\,. \tag{3.30}$$

Corollary 3.4 implies that the oracle optimal shrinkage estimator for the precision matrix in the case $c > 1$ and $\mathbf{\Sigma}_n = \sigma\mathbf{I}$ is equal to

$$\widehat{\mathbf{\Pi}}_{GSE} = \mathbf{\Sigma}_n^{-1} = \sigma^{-1}\mathbf{I}\,. \tag{3.31}$$

The quantity $\sigma^{-1} = \frac{1}{p}\text{tr}(\mathbf{\Sigma}_n^{-1})$ can be easily estimated using the result of Corollary 3.3. Namely, the consistent estimator of σ^{-1} is given by

$$\hat{\sigma}^{-1} = p/n\frac{p/n - 1}{p}\text{tr}(\mathbf{S}_n^+)\,. \tag{3.32}$$

However, in the general case when $\mathbf{\Sigma}_n$ and $\mathbf{\Pi}_0$ are arbitrary, the results of Corollary 3.3 and Corollary 3.4 do not hold anymore. For this reason, we consider the oracle estimator given by (3.16) in the simulation study of Section 5.

4 Estimation of unknown parameters

In this section we present consistent estimators for the asymptotic optimal shrinkage intensities derived in Section 3. The results of Theorem 3.2 allow us to estimate consistently the functionals of type $\text{tr}(\mathbf{\Sigma}_n^{-1}\mathbf{\Theta})$ and the Frobenius norm of the precision matrix. The consistent estimator for the functional $\theta_n(\mathbf{\Theta}) = \text{tr}(\mathbf{\Sigma}_n^{-1}\mathbf{\Theta})$ is given by

$$\hat{\theta}_n(\mathbf{\Theta}) = (1 - p/n)\text{tr}(\mathbf{S}_n^{-1}\mathbf{\Theta}) \tag{4.1}$$

which is a generalization of the so-called $G3$-estimator obtained by Girko (1995). In particular, in the case when $\mathbf{\Theta} = \boldsymbol{\xi}\boldsymbol{\eta}'$ for some vectors $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ with bounded Euclidean norm, Girko (1995) showed that the corresponding estimator $\hat{\theta}_n(\boldsymbol{\xi}\boldsymbol{\eta}')$ tends to $\theta_n(\boldsymbol{\xi}\boldsymbol{\eta}')$ in probability. In contrast, Theorem 3.2 ensures the consistency of $\hat{\theta}_n(\mathbf{\Theta})$ for more general forms of $\mathbf{\Theta}$ which should not be of rank 1.

Again, using (4.1) and Theorem 3.2 we construct a consistent estimator for $\rho_n = 1/p||\mathbf{\Sigma}_n^{-1}||_F^2$ which is given by

$$\hat{\rho}_n = \frac{(1-p/n)^2}{p}||\mathbf{S}_n^{-1}||_F^2 - \frac{1-p/n}{pn}||\mathbf{S}_n^{-1}||_{tr}^2 . \tag{4.2}$$

Note that the result (4.2) is entirely new and it was not mentioned in the literature up to now. Moreover, it is noted that for the derivation of (4.2) we do not need the existence of the 4th moment (see, the assumption (A2) in Section 2).

Using both the estimators (4.1) and (4.2), we are able now to construct the optimal linear shrinkage estimator (OLSE) for the precision matrix which is given by

$$\widehat{\mathbf{\Pi}}_{OLSE} = \widehat{\alpha}_n^* \mathbf{S}_n^{-1} + \widehat{\beta}_n^* \mathbf{\Pi}_0 \quad \text{with } \sup_p 1/p||\mathbf{\Pi}_0||_{tr} \le M \,, \tag{4.3}$$

where

$$\begin{aligned}
\widehat{\alpha}_n^* &= (1-p/n)\frac{p\hat{\rho}_n||\mathbf{\Pi}_0||_F^2 - \hat{\theta}_n^2(\mathbf{\Pi}_0)}{\left(p\hat{\rho}_n + \dfrac{p/n}{p(1-p/n)}\hat{\theta}_n^2(\mathbf{I})\right)||\mathbf{\Pi}_0||_F^2 - \hat{\theta}_n^2(\mathbf{\Pi}_0)} \\
&= (1-p/n)\left(1 - \frac{\dfrac{p/n}{p(1-p/n)}\hat{\theta}_n^2(\mathbf{I})||\mathbf{\Pi}_0||_F^2}{\left(p\hat{\rho}_n + \dfrac{p/n}{p(1-p/n)}\hat{\theta}_n^2(\mathbf{I})\right)||\mathbf{\Pi}_0||_F^2 - \hat{\theta}_n^2(\mathbf{\Pi}_0)}\right) \\
&= 1 - p/n - \frac{\dfrac{1}{n}||\mathbf{S}_n^{-1}||_{tr}^2||\mathbf{\Pi}_0||_F^2}{||\mathbf{S}_n^{-1}||_F^2||\mathbf{\Pi}_0||_F^2 - (\text{tr}(\mathbf{S}_n^{-1}\mathbf{\Pi}_0))^2} \,.
\end{aligned} \tag{4.4}$$

and

$$\widehat{\beta}_n^* = \frac{\hat{\theta}_n(\mathbf{\Pi}_0)}{||\mathbf{\Pi}_0||_F^2}\left(1 - \frac{\widehat{\alpha}_n^*}{1-p/n}\right) = \frac{\text{tr}(\mathbf{S}_n^{-1}\mathbf{\Pi}_0)}{||\mathbf{\Pi}_0||_F^2}\left(1 - p/n - \widehat{\alpha}_n^*\right) . \tag{4.5}$$

The bona fide OLSE estimator (4.3) is optimal in the sense that it minimizes the Frobenius loss. It means that the estimators $\widehat{\alpha}_n^*$ and $\widehat{\beta}_n^*$ tend almost surely to their oracle asymptotic values (3.13) and (3.15) as $n \to \infty$, respectively. According to Corollary 3.1 the oracle optimal intensities α_n^* and β_n^* given in (3.7) and (3.8) behave similarly. It is a remarkable property of the OLSE estimator which indicates that the bona fide OLSE estimator tends almost surely to its oracle one. Moreover, using the inequality (3.6) it can be easily verified that the estimated optimal shrinkage intensities $\widehat{\alpha}_n^*$ and $\widehat{\beta}_n^*$ are almost surely positive and $\widehat{\alpha}_n^*$ has the support

$(0, 1-p/n)$. Only in the case when $p/n \to c = 0$ as $n \to \infty$ the shrinkage intensities satisfy $\widehat{\alpha}_n^* \to 1$ and $\widehat{\beta}_n^* \to 0$. In this case the OLSE estimator coincides with the sample estimator which is consistent for the standard asymptotics.

If we compare the estimates of the optimal shrinkage intensities given in (4.4) and (4.5) with the corresponding ones calculated for the population covariance matrix given in Bodnar et al. (2014) then we conclude that the corresponding estimators are different although the structure remains somewhat similar. In the simulation study of Section 5 both estimators for the precision matrix are compared with each other and it is shown that it is better to shrink the inverse sample covariance matrix itself than to shrink the sample covariance matrix and then to invert it.

4.1 Choice of $\mathbf{\Pi}_0$.

The next question is the choice of the nonrandom target matrix $\mathbf{\Pi}_0$ which should be positive definite with uniformly bounded normalized trace norm. Unfortunately, the answer on this question depends on the underlying data because the choice of the target matrix is equivalent to the choice of the hyperparameter for the prior distribution of $\mathbf{\Sigma}_n^{-1}$. This problem is well-known in Bayesian statistics. The application of different priors leads to different results. So it is very important to choose the one which works fine in most cases. The naive one is $\mathbf{\Pi}_0 = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix. Obviously, the oracle OLSE estimator has the prior matrix as the true precision matrix $\mathbf{\Sigma}_n^{-1}$ and is a consistent estimator for the precision matrix as shown in Proposition 4.1. Obviously, including some new information about the true covariance matrix $\mathbf{\Sigma}_n$ into the prior can lead to a significant increase of performance (see, Bodnar et al. (2014)). In our simulation study, however, we take $\mathbf{\Pi}_0 = \mathbf{I}$ as a reasonable prior in the case when no additional information is available.

Consider the OLSE estimator as a matrix function $\widehat{\mathbf{\Pi}}_{OLSE}(\mathbf{\Pi}_0) : M_p \to \tilde{M}_p$, where M_p is the space of p-dimensional positive definite symmetric matrices and $\tilde{M}_p$ is the corresponding space of the p-dimensional positive definite symmetric random matrices. In the following proposition we present some properties of the OLSE estimator as a function of the prior matrix $\mathbf{\Pi}_0$.

Proposition 4.1. *For the OLSE estimator $\widehat{\mathbf{\Pi}}_{OLSE}(\mathbf{\Pi}_0)$ it holds that*

i). the function $\widehat{\mathbf{\Pi}}_{OLSE}(\mathbf{\Pi}_0)$ is scale invariant, i.e., for arbitrary $\sigma > 0$ $\widehat{\mathbf{\Pi}}_{OLSE}(\sigma\mathbf{\Pi}_0) = \widehat{\mathbf{\Pi}}_{OLSE}(\mathbf{\Pi}_0)$.

ii). $\widehat{\mathbf{\Pi}}_{OLSE}(1/p\mathbf{I})$ is a consistent estimator for the precision matrix $\mathbf{\Sigma}_n^{-1} = \sigma\mathbf{I}$ for arbitrary $\sigma > 0$ and $c \in (0, +\infty)$.

iii). $\widehat{\mathbf{\Pi}}_{OLSE}(\mathbf{\Sigma}_n^{-1})$ is a consistent estimator for the precision matrix $\mathbf{\Sigma}_n^{-1}$.

5 Simulation Study

In this section we provide a Monte Carlo simulation study to investigate the performance of the suggested OLSE estimator for the precision matrix.

Before we proceed, four benchmark estimators for the precision matrix used in the simulations are introduced. The first one has been suggested by Bodnar et al. (2014) and it is called the OLSE estimator for the population covariance matrix.

Recall, that the optimal linear shrinkage estimator (OLSE) for the covariance matrix $\mathbf{\Sigma}_n$ for $c \in (0, +\infty)$ is given by

$$\widehat{\mathbf{\Sigma}}_{OLSE} = \tilde{\alpha}^* \mathbf{S}_n + \tilde{\beta}^* \mathbf{\Sigma}_0 \quad \text{with } ||\mathbf{\Sigma}_0||_{tr} \le M\,, \tag{5.1}$$

where $\mathbf{\Sigma}_0$ is the positive definite symmetric target matrix,

$$\tilde{\alpha}^* = 1 - \frac{\frac{1}{n}||\mathbf{S}_n||^2_{tr}||\mathbf{\Sigma}_0||^2_F}{||\mathbf{S}_n||^2_F||\mathbf{\Sigma}_0||^2_F - \left(\mathrm{tr}(\mathbf{S}_n\mathbf{\Sigma}_0)\right)^2} \quad \text{and} \quad \tilde{\beta}^* = \frac{\mathrm{tr}(\mathbf{S}_n\mathbf{\Sigma}_0)}{||\mathbf{\Sigma}_0||^2_F}\left(1-\tilde{\alpha}^*\right) . \tag{5.2}$$

Bodnar et al. (2014) proved that their $\widehat{\mathbf{\Sigma}}_{OLSE}$ estimator possesses asymptotically almost surely the smallest Frobenius loss over all linear shrinkage estimators. Moreover, they showed by simulations that if the target matrix is $\mathbf{\Sigma}_0 = 1/p\mathbf{I}$ then the estimator $\widehat{\mathbf{\Sigma}}_{OLSE}$ is asymptotically equivalent to the linear shrinkage estimator proposed by Ledoit and Wolf (2004) for normally distributed data. Of course, in order to compare this estimator with the suggested OLSE estimator for the precision matrix $\widehat{\mathbf{\Pi}}_{OLSE}$ from (4.3) we have to invert $\widehat{\mathbf{\Sigma}}_{OLSE}$.

The second considered estimator for the precision matrix is the scaled standard estimator (SSE), which is presented by several authors Stein (1975), Mestre and Lagunas (2006), Srivastava (2007) and Kubokawa and Srivastava (2008). It is given by

$$\widehat{\mathbf{\Pi}}_{SSE} = \frac{n-p-2}{n-1}\mathbf{S}_n^{-1}\delta_{(p<n)} + \frac{p}{n-1}\mathbf{S}_n^{+}\delta_{(p\geq n)} \tag{5.3}$$

where $\mathbf{S}_n^+$ is the Moore-Penrose inverse of $\mathbf{S}_n$ and $\delta_{(\cdot)}$ is a Dirac delta function. Next we consider the estimator proposed by Efron and Morris (1976)

$$\widehat{\mathbf{\Pi}}_{EM} = \frac{n-p-2}{n-1}\mathbf{S}_n^{-1} + \frac{p^2+p-2}{(n-1)\mathrm{tr}(\mathbf{S}_n)}\mathbf{I}\,. \tag{5.4}$$

The last one is the empirical Bayes estimator proposed by Kubokawa and Srivastava (2008)

$$\widehat{\mathbf{\Pi}}_{KS} = p\left((n-1)\mathbf{S}_n + \mathrm{tr}(\mathbf{S}_n)\mathbf{I}\right)^{-1} . \tag{5.5}$$

Next, we compare the performance of the considered estimators. As a performance measure we take the PRIAL (Percentage Relative Improvement in Average Loss). For an arbitrary estimator of the precision matrix, $\widehat{\mathbf{M}}$, the PRIAL is defined by

$$\mathrm{PRIAL}(\widehat{\mathbf{M}}) = \left(1 - \frac{E||\widehat{\mathbf{M}} - \mathbf{\Sigma}_n^{-1}||^2_F}{E||\mathbf{S}_n^{-1} - \mathbf{\Sigma}_n^{-1}||^2_F}\right)\cdot 100\%\,. \tag{5.6}$$

Thus, by definition (5.6), PRIAL($\mathbf{S}_n^{-1}$) is equal to zero and PRIAL($\mathbf{\Sigma}_n^{-1}$) is equal to 100%.

In our simulations we take $\mathbf{\Sigma}_n$ as a non-diagonal matrix. We separate its spectrum in three parts with 20% of the eigenvalues equal to 1, 40% equal to 3 and 40% equal to 10. In terms of the corresponding cumulative distribution function of the eigenvalues of $\mathbf{\Sigma}_n$ (cf. Section 2) it holds that

$$H_n^{\mathbf{\Sigma}_n}(t) = 1/5\delta_{[1,\,\infty)}(t) + 2/5\delta_{[3,\,\infty)}(t) + 2/5\delta_{[10,\,\infty)}(t)\,, \tag{5.7}$$

where δ is the Dirac delta function. The matrix of eigenvectors of the population covariance matrix $\mathbf{\Sigma}_n$ we simulate from the Haar distribution[6]. Doing so we leave the spectral structure of population covariance matrix unchanged for all dimensions p whereas the eigenvectors may vary.

[6]If $\mathbf{V}$ has a Haar measure over the orthogonal matrices, then for any unit vector $\mathbf{x} \in \mathbb{R}^p$, $\mathbf{Vx}$ has a uniform distribution over the unit sphere $S_p = \{\mathbf{x} \in \mathbb{R}^p; ||\mathbf{x}|| = 1\}$.

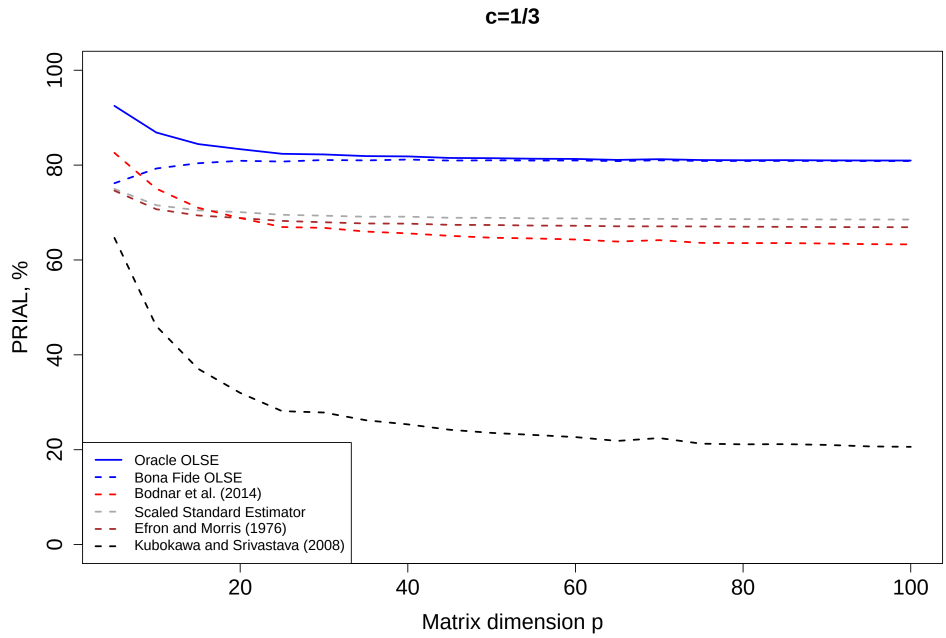

Figure 1: PRIALs for the oracle and bona fide estimator $\widehat{\mathbf{\Pi}}_{OLSE}$ with the prior (target) matrix $\mathbf{\Pi}_0 = \mathbf{I}$, the inverse of the bona fide estimator $\widehat{\mathbf{\Sigma}}_{OLSE}$ with the prior $1/p\mathbf{I}$, the scaled standard estimator (SSE), EM (Efron and Morris (1976)) and KS (Kubokawa and Srivastava (2008)) for $p = 5k, k \in \{1, \ldots, 20\}$, $c = 1/3$. The results are based on 1000 independent realizations.

In Figure 1 we present the results of the first simulation study in case of the normally distributed data and $c = 1/3$. The suggested oracle estimator is shown as a solid blue line, while the corresponding bona fide estimator is a dashed blue line. For the bona fide estimator we observe a fast convergence rate to its, i.e., it converges in PRIAL already for the dimension $p \geq 50$. It is remarkable that the bona fide OLSE estimator for the precision matrix $\widehat{\mathbf{\Pi}}_{OLSE}$ with the naive prior $\mathbf{\Pi}_0 = \mathbf{I}$ dominates the corresponding inverted OLSE estimator $\widehat{\mathbf{\Sigma}}^{-1}_{OLSE}$. The standard scaled estimator (SSE) and the one proposed by Efron and Morris (1976) lie close and they are a little better than the inverted OLSE estimator $\widehat{\mathbf{\Sigma}}^{-1}_{OLSE}$. The KS estimator (Kubokawa and Srivastava (2008)) seems to be the worst one in this case.

In Figure 2 we present the results of simulations under the normal distribution for $c = 1/2$, whereas Figure 3 shows the results in case $c = 0.8$. For $c = 1/2$ we observe a better overall performance for all of the considered estimators for the precision matrix. The OLSE estimator (4.3) is again the best one for $c = 1/2$. The inverted OLSE estimator is becoming better than the other competitors. The worst one in this case is the KS estimator.

Figure 3 for $c = 0.8$ shows that all estimators provide the superior performance. The inverse of the OLSE estimator from Bodnar et al. (2014) is surprisingly the best one for $c = 0.8$ and converges to the oracle OLSE estimator. It must be noted that the bona fide OLSE estimator converges to its oracle slower than the inverted OLSE estimator. The KS estimator is slightly behind the inverted OLSE estimator, whereas the SSE and the EM estimators have roughly the same behavior and they are clearly worse than the others. Nevertheless, from the results of the last simulation show that all of the considered estimators are close to the true precision matrix.

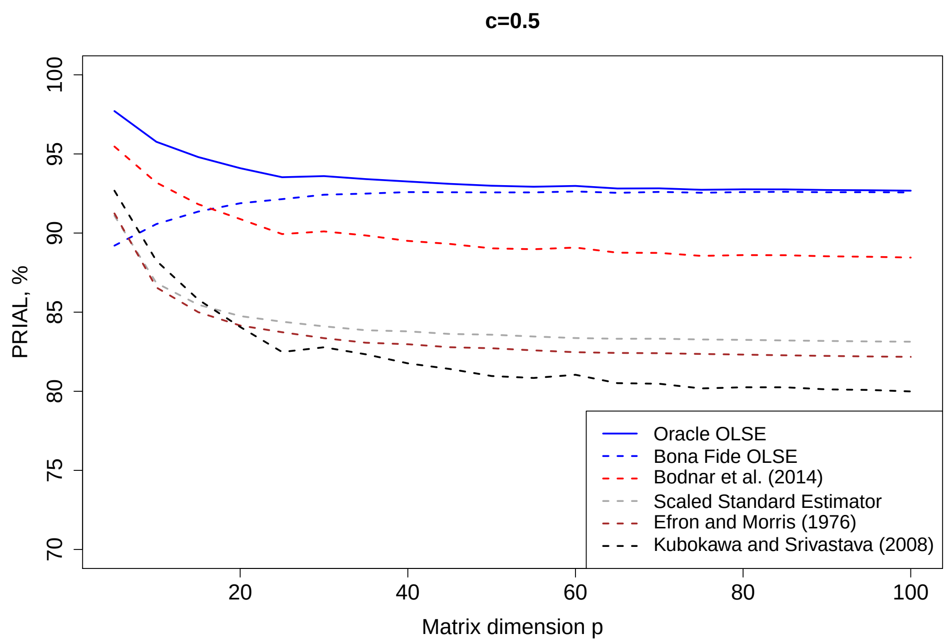

Figure 2: PRIALs for the oracle and bona fide estimator $\widehat{\mathbf{\Pi}}_{OLSE}$ with the prior (target) matrix $\mathbf{\Pi}_0 = \mathbf{I}$, the inverse of the bona fide estimator $\widehat{\mathbf{\Sigma}}_{OLSE}$ with the prior $1/p\mathbf{I}$, the scaled standard estimator (SSE), EM (Efron and Morris (1976)) and KS (Kubokawa and Srivastava (2008)) for $p = 5k, k \in \{1, \ldots, 20\}$, $c = 1/2$. The results are based on 1000 independent realizations.

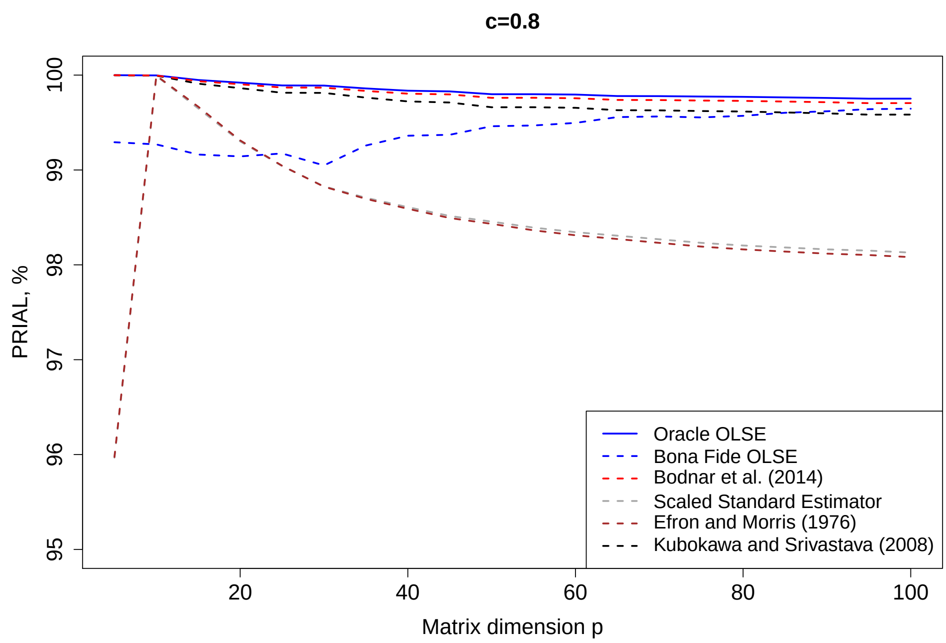

Figure 3: PRIALs for the oracle and bona fide estimator $\widehat{\mathbf{\Pi}}_{OLSE}$ with the prior (target) matrix $\mathbf{\Pi}_0 = \mathbf{I}$, the inverse of the bona fide estimator $\widehat{\mathbf{\Sigma}}_{OLSE}$ with the prior $1/p\mathbf{I}$, the scaled standard estimator (SSE), EM (Efron and Morris (1976)) and KS (Kubokawa and Srivastava (2008)) for $p = 5k, k \in \{1, \ldots, 20\}$, $c = 0.8$. The results are based on 1000 independent realizations.

Figure 4 is dedicated to the case when the underlying distribution departs from the normal one. Here, we consider the t-distribution with 10 degrees of freedom. The estimators and their priors are the same as used in Figure 1. For $c = 1/3$ we observe that the overall performance of all considered estimators is even better as in the case of the normal distribution in Figure 1. As usual, the OLSE estimator with the prior $\mathbf{I}$ is ranked first. It is remarkable that the suggested bona fide OLSE estimator converges much slower to its oracle. In contrast to Figure 1 where the convergence was very fast ($p \geq 50$), Figure 4 ensures the convergence for at least $p \geq 100$. It seems that the convergence rate is influenced by heavy tails, i.e., the heavier the tails are the slower is the convergence of the bona fide OLSE estimator to its oracle. The second place belongs to the inverted bona fide $\widehat{\mathbf{\Sigma}}_{OLSE}$ with the prior $1/p\mathbf{I}$. On the third and on the forth places we rank the SSE and the EM estimators. The worst one, as in Figure 1, is the KS estimator. As a result, even in the non-normal case the proposed OLSE estimator for the precision matrix shows remarkable stability and robustness. The simulation results for $c = 1/2$ and $c = 0.8$ (see, Figures 5 and 6) are very similar to those obtained for the normal distribution in Figures 2 and 3. Only the convergence of the bona fide estimator to its oracle is slower.

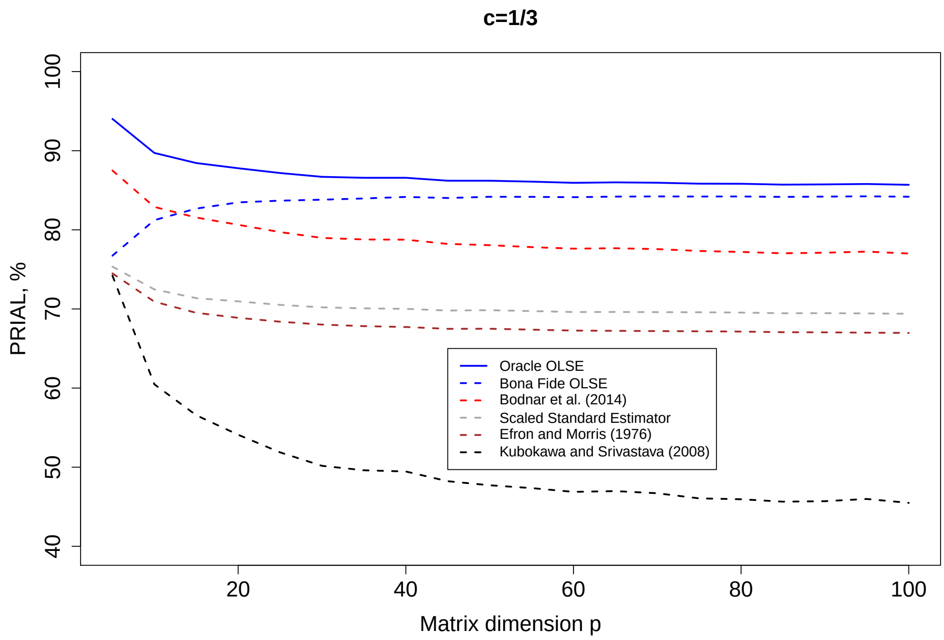

Figure 4: PRIALs for the oracle and bona fide estimator $\widehat{\mathbf{\Pi}}_{OLSE}$ with the prior(target) matrix $\mathbf{\Pi}_0 = \mathbf{I}$, the inverse of the bona fide estimator $\widehat{\mathbf{\Sigma}}_{OLSE}$ with the prior $1/p\mathbf{I}$, the scaled standard estimator (SSE), EM (Efron and Morris (1976)) and KS (Kubokawa and Srivastava (2008)) for $p = 5k, k \in \{1, \ldots, 20\}$, $c = 1/3$. The data are generated from the t-distribution with 10 degrees of freedom. The results are based on 1000 independent realizations.

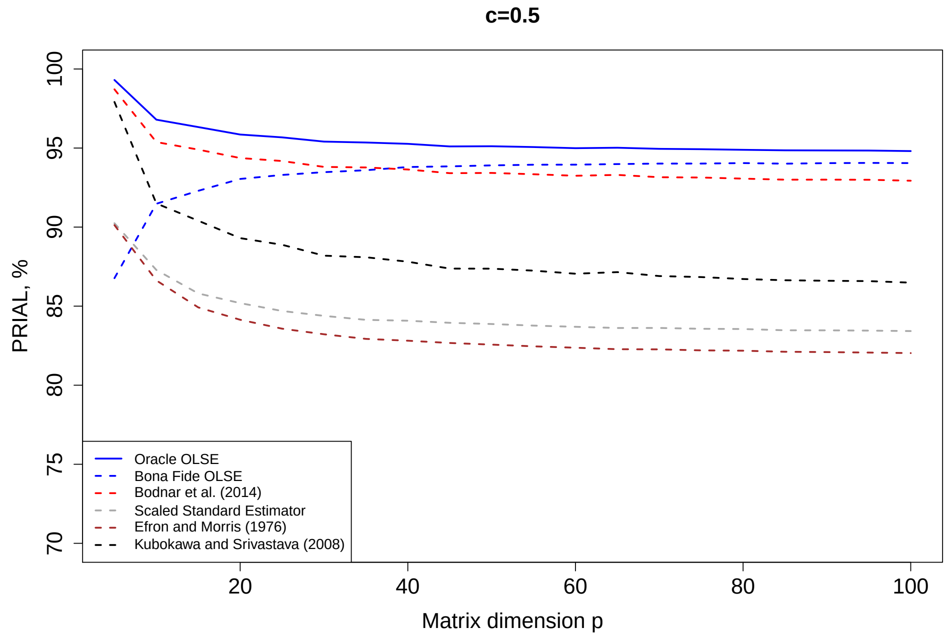

Figure 5: PRIALs for the oracle and bona fide estimator $\widehat{\mathbf{\Pi}}_{OLSE}$ with the prior(target) matrix $\mathbf{\Pi}_0 = \mathbf{I}$, the inverse of the bona fide estimator $\widehat{\mathbf{\Sigma}}_{OLSE}$ with the prior $1/p\mathbf{I}$, the scaled standard estimator (SSE), EM (Efron and Morris (1976)) and KS (Kubokawa and Srivastava (2008)) for $p = 5k, k \in \{1, \dots, 20\}$, $c = 1/2$. The data are generated from the t-distribution with 10 degrees of freedom. The results are based on 1000 independent realizations.

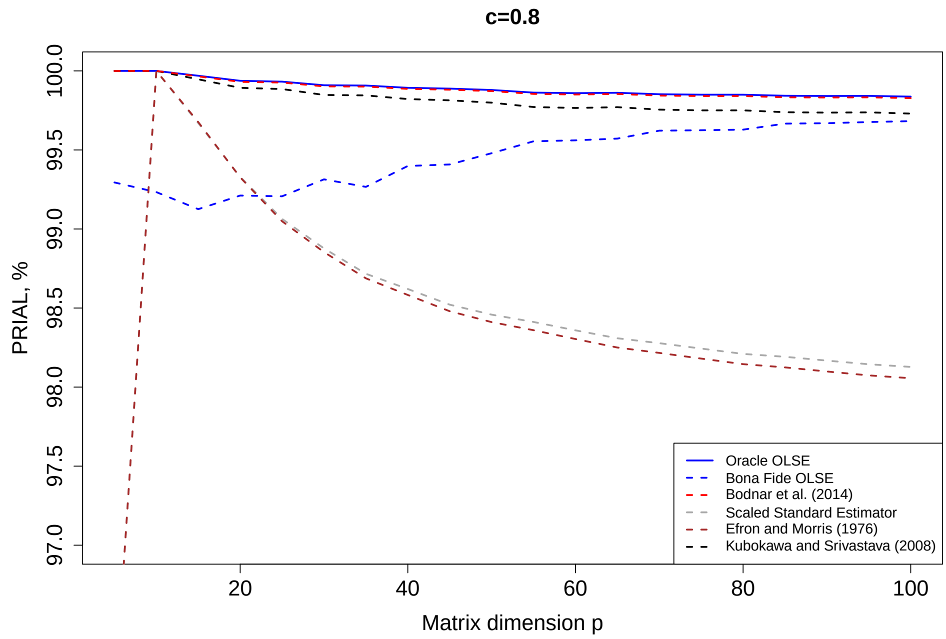

Figure 6: PRIALs for the oracle and bona fide estimator $\widehat{\mathbf{\Pi}}_{OLSE}$ with the prior(target) matrix $\mathbf{\Pi}_0 = \mathbf{I}$, the inverse of the bona fide estimator $\widehat{\mathbf{\Sigma}}_{OLSE}$ with the prior $1/p\mathbf{I}$, the scaled standard estimator (SSE), EM (Efron and Morris (1976)) and KS (Kubokawa and Srivastava (2008)) for $p = 5k, k \in \{1, \dots, 20\}$, $c = 0.8$. The data are generated from the t-distribution with 10 degrees of freedom. The results are based on 1000 independent realizations.

At last, Figure 7 contains the example when $c = 2.5$ for the normally distributed data. Here we use the oracle OLSE estimator proposed in Section 3.2 and compare it with the bona fide inverse OLSE, the SSE and the KS estimators. The PRIAL is defined in the similar way as (5.6), only the matrix $\mathbf{S}^{-1}$ is changed to the generalized inverse $\mathbf{S}_n^+$. Similarly, in case of $\mathbf{S}_n^+$ the PRIAL is equal to zero. The best estimator is the oracle OLSE with the simple prior $\mathbf{I}$. The inverted OLSE is not far behind. The KS estimator is placed on the third place, while the SSE estimator seems to be worse than the Moore-Penrose inverse of $\mathbf{S}_n^+$ and, therefore, it is not visible on the picture. The result of the suggested oracle OLSE estimator is remarkable because it shows that if we could consistently estimate the shrinkage intensities given by (3.17) and (3.18), then the resulting bona fide OLSE estimator would probably converge to the oracle one and dominates the other estimators. This task is not an easy one and it is left for future research. Nevertheless, from Figure 7 we conclude that the inverted OLSE estimator proposed is still a very good feasible alternative for $c > 1$.

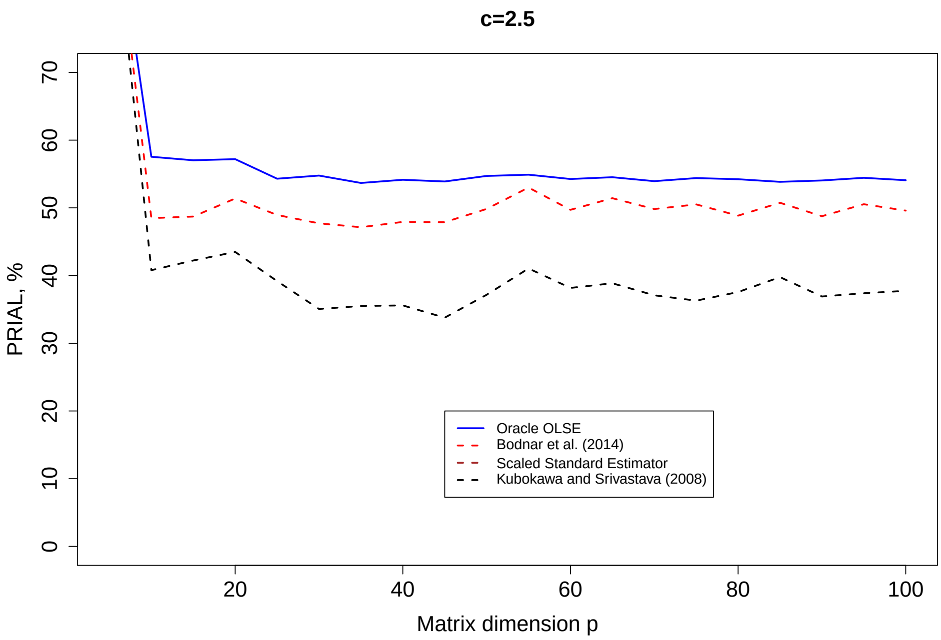

Figure 7: PRIALs for the oracle estimator $\widehat{\mathbf{\Pi}}_{OLSE}$ with the prior $\mathbf{\Pi}_0 = \mathbf{I}$, the inverse of the bona fide $\widehat{\mathbf{\Sigma}}_{OLSE}$ with the prior $1/p\mathbf{I}$, the scaled standard estimator (SSE), EM (Efron and Morris (1976)) and KS (Kubokawa and Srivastava (2008)) for $p = 5k, k \in \{1, \dots, 20\}$, $c = 2.5$. The data are generated from the normal distribution. The results are based on 1000 independent realizations.

As a result, the simulation results as well as the theoretical findings show that the OLSE estimator $\widehat{\mathbf{\Pi}}_{OLSE}$ is a great alternative not only to the sample estimator, to the inverted linear shrinkage estimator proposed by Ledoit and Wolf (2004) and its generalization of Bodnar et. al (2014), but also to the other estimators suggested in literature. The case of $c > 1$ is even more important for the practical purposes but it seems to be more difficult to handle analytically. This can be done in an efficient way if the population covariance matrix is a multiple of identity. In general case a good alternative would be the inverse OLSE estimator given in (5.1), but it is not optimal for the precision matrix. This point will be treated in future research.

6 Summary

In this paper, we deal with the problem of the estimation of the precision matrix for large dimensional data is considered. Our particular interest is the case when both the dimension of the precision matrix $p \to \infty$ and the sample size $n \to \infty$ such that $p/n \to c \in (0, +\infty)$. Using the results from the random matrix theory and the linear shrinkage technique we develop an estimator for the precision matrix which is distribution-free, robust and possesses almost surely the smallest Frobenius loss asymptotically. In particular, we prove that the Frobenius norms of the inverse and of the generalized inverse sample covariance matrices as well as of the optimal shrinkage intensities tend to the nonrandom quantities under high dimensional asymptotics. In order to get the optimal linear shrinkage estimator for the precision matrix we estimate the unknown quantities consistently. The performance of the suggested OLSE estimator is compared with other known estimators for the precision matrix via the simulation study.

Acknowledgments

The authors would like to thank the participants of the "3rd International Workshop on Functional and Operatorial Statistics (IWFOS 2014)", the conferences "11th German Probability and Statistics Days 2014", "Random Matrix Theory: Foundations and Applications 2014" and "Statistische Woche 2014" for fruitful comments and discussions. We also thank Dr. C. Guillaume for his valuable comments on the choice of target matrix and the practical insights of the shrinkage estimation.

7 Appendix

Here the proofs of the theorems are given.

Proof of Theorem 3.1. The proof of the theorem is based on the Marchenko-Pastur theorem proved by Silverstein (1995).

Theorem 7.1. ***[Silverstein (1995)]*** *Assume that on the common probability space assumption (A1) is satisfied for $\frac{p}{n} \to c \in (0, +\infty)$ as $n \to \infty$. Then almost surely $F_n(t) \stackrel{a.s.}{\Rightarrow} F(t)$ as $n \to \infty$. Moreover, the Stieltjes transform of F satisfies the following equation*

$$m_F(z) = \int\limits_{-\infty}^{+\infty} \frac{1}{\tau(1 - c - czm_F(z)) - z} dH(\tau)\,, \tag{7.1}$$

in the sense that $m_F(z)$ is the unique solution of (7.1) for all $z \in \mathbb{C}^+$.

Consider the asymptotics of the quantity

$$\frac{1}{p}\mathrm{tr}\left(\mathbf{S}_n^{-2}\right) = \frac{\partial}{\partial z}\frac{1}{p}\mathrm{tr}\left[(\mathbf{S}_n - z\mathbf{I})^{-1}\right]\Big|_{z=0} = \frac{\partial}{\partial z} m_{F_n}(z)\Big|_{z=0}\,. \tag{7.2}$$

Using Theorem 7.1 we note that $m_{F_n}(z)$ tends almost surely to a nonrandom limit function $m_F(z)$ which is the unique solution of the MP equation (7.1) First, we show that the limit $z \to 0^+$ can be taken under the integral sign in (7.1). Let $\underline{m}(z) = -\dfrac{1-c}{z} + cm_F(z)$ then using the assumption (A3) we rewrite equality (7.1) in the following way

$$m_F(z) = \int\limits_{-\infty}^{+\infty} \frac{-1}{z(\tau\underline{m}(z) + 1)} dH(\tau) = \int\limits_{h_0}^{h_1} \frac{-1}{z(\tau\underline{m}(z) + 1)} dH(\tau) = \int\limits_{h_0}^{h_1} \frac{s(z, \tau)}{\tau} dH(\tau)\,, \tag{7.3}$$

where the function $s(z,\tau) = \dfrac{-\tau}{z(\tau \underline{m}(z)+1)}$ is the Stieltjes transform of a positive measure on $\mathbb{R}^+$ with total mass τ (see, Paul and Silverstein (2009)). Using the properties of Stieltjes transform $\overline{s(z,\tau)} = s(\overline{z},\tau)$ and $\lim\limits_{z\to 0^+} s(z,\tau) = \lim\limits_{z\to 0^+} s(\overline{z},\tau)$ we get

$$\gamma^2(c) \equiv \lim_{z\to 0^+} |s(z,\tau)|^2 = \lim_{z\to 0^+} s(z,\tau)s(\overline{z},\tau) = \left(\lim_{z\to 0^+} s(z,\tau)\right)^2 = \left(\frac{1}{1-c}\right)^2 . \tag{7.4}$$

The last equality in (7.4) follows from boundedness of $m_F(0) = \lim\limits_{z\to 0^+} m_F(z)$ for $c<1$ (see, Silverstein and Choi (1995)).

Let $\mathcal{B}$ be a compact ball around 0. Then, the application of the inequality (cf. Silverstein (2009))

$$0 \le |s(z,\tau)| \le \frac{1}{\mathbf{Im}(z)} . \tag{7.5}$$

leads to

$$0 \le \left|\frac{s(z,\tau)}{\tau}\right| \le \frac{\gamma(c)}{h_0} \quad \forall z \in \mathcal{B} . \tag{7.6}$$

Hence, using the inequality (7.6) together with (7.3) and the dominated convergence theorem we conclude that the limit $z \to 0^+$ can be moved under the integral sign in (7.1). This fact together with $m_F(0) < \infty$ implies that

$$m_F(0) = \lim_{z\to 0^+} m_F(z) = \frac{1}{1-c}\int_{-\infty}^{+\infty} \frac{dH(\tau)}{\tau} , \tag{7.7}$$

where the integral exists due to assumption (A3).

The function $m_F(z)$ is analytic in $\mathbb{C}^+$ thus we take the derivative with respect to z from both sides of equation (7.1) and get

$$m_F'(z) = \int_{-\infty}^{+\infty} \frac{\tau c(m_F(z) + z m_F'(z)) + 1}{(\tau(1-c-czm_F(z)) - z)^2} dH(\tau) . \tag{7.8}$$

Rearranging terms in (7.8) we get

$$m_F'(z)\left(1 - \int_{-\infty}^{+\infty} \frac{cz\tau}{(\tau(1-c-czm_F(z)) - z)^2} dH(\tau)\right) \tag{7.9}$$

$$= \int_{-\infty}^{+\infty} \frac{\tau c m_F(z) + 1}{(\tau(1-c-czm_F(z)) - z)^2} dH(\tau) .$$

The right side of (7.9) exists as $z \to 0^+$ due to (7.6), $m_F(0) < \infty$, and the dominated convergence theorem, thus the left hand side must also exist as $z \to 0^+$ and it implies that

$$m_F'(0) = \frac{1}{(1-c)^2}\left(c m_F(0)\int_{-\infty}^{+\infty} \frac{dH(\tau)}{\tau} + \int_{-\infty}^{+\infty} \frac{dH(\tau)}{\tau^2}\right) . \tag{7.10}$$

The application of (7.7) completes the proof of Theorem 3.1.

Proof of Theorem 3.2. In order to prove Theorem 3.2 we need the following lemma.

Lemma 7.1. ***[Lemma B.26, Bai and Silverstein (2010)]*** *Let $\boldsymbol{A}$ be a $p \times p$ nonrandom matrix and let $\mathbf{X} = (x_1, \ldots, x_p)'$ be a random vector with independent entries. Assume that $E(x_i) = 0$, $E|x_i|^2 = 1$, and $E|x_i|^l \leq \nu_l$. Then, for any $k \geq 1$,*

$$E|\mathbf{X}'\boldsymbol{A}\mathbf{X} - tr(\boldsymbol{A})|^k \leq C_k \left((\nu_4 tr(\boldsymbol{A}\boldsymbol{A}'))^{\frac{k}{2}} + \nu_{2k} tr(\boldsymbol{A}\boldsymbol{A}')^{\frac{k}{2}} \right) , \tag{7.11}$$

where C_k is some constant which depends only on k.

Rubio and Mestre (2011) studied the asymptotics of the functionals $\mathrm{tr}(\boldsymbol{\Theta}(\mathbf{S}_n - z\mathbf{I})^{-1})$ for a deterministic matrix $\boldsymbol{\Theta}$ with bounded trace norm at infinity. It is noted that the results of Theorem 1 by Rubio and Mestre (2011) also hold under the assumption of the existence of 4th moments which is weaker than the one given in the original paper. This statement is obtained by using Lemma B.26 of Bai and Silverstein (2010) on quadratic forms which we recall for presentation purposes as Lemma 7.1 above.

In order to obtain the statement of Theorem 1 by Rubio and Mestre (2011) under the weaker assumption imposed on the moments, we replace Lemma 2 of Rubio and Mestre (2011) by Lemma 7.1 in the case of $k \geq 1$. This implies that Lemma 3 of Rubio and Mestre (2011) holds also for $k \geq 1$. Lemma 4 of Rubio and Mestre (2011) has already been proved under the assumption that there exist $4+\varepsilon$ moments. The last step is the application of Lemma 1, 2 and 3 of Rubio and Mestre (2011) with $k \geq 1$. Finally, it can be easily checked that further steps of the proof of Theorem 1 by Rubio and Mestre (2011) hold under the existence of $4+\varepsilon$ moments. A partial case of the result proved by Rubio and Mestre (2011) is summarized in Theorem 7.2.

Theorem 7.2. ***[Rubio and Mestre (2011)]*** *Assume that (A2) and (A3) hold and additionally some nonrandom matrix $\boldsymbol{\Theta}$ has uniformly bounded trace norm at infinity then for $p/n \to c > 0$ as $n \to \infty$*

$$\left| tr\left(\boldsymbol{\Theta}(\mathbf{S}_n - z\mathbf{I})^{-1}\right) - tr\left(\boldsymbol{\Theta}(x(z)\boldsymbol{\Sigma}_n - z\mathbf{I})^{-1}\right) \right| \longrightarrow 0 \quad a.\ s. \tag{7.12}$$

where $x(z)$ is a unique solution in $\mathbb{C}^+$ of the following equation

$$\frac{1-x(z)}{x(z)} = \frac{c}{p} tr\left(x(z)\mathbf{I} - z\boldsymbol{\Sigma}_n^{-1}\right)^{-1} . \tag{7.13}$$

Next we prove Theorem 3.2 directly. We investigate the asymptotic behavior of the following quantities

$$\gamma_1 = \mathrm{tr}(\mathbf{S}_n^{-1}\boldsymbol{\Theta}), \tag{7.14}$$

$$\gamma_2 = \frac{1}{p}||\mathbf{S}_n^{-1}||_F^2 = \frac{1}{p}\mathrm{tr}(\mathbf{S}_n^{-2}) . \tag{7.15}$$

First, we consider the quantity γ_1 given in (7.14) and rewrite it as $\gamma_1(z) = \mathrm{tr}((\mathbf{S}_n - z\mathbf{I})^{-1}\boldsymbol{\Theta})$ for all $z \in \mathbb{C}^+$. Using Theorem 7.2 we get that for $\frac{p}{n} \to c \in (0,1)$ as $n \to \infty$ holds

$$\left|\gamma_1(z) - \mathrm{tr}(\boldsymbol{\Theta}(x(z)\boldsymbol{\Sigma}_n - z\mathbf{I})^{-1})\right| \longrightarrow 0 \quad \mathrm{a.\ s.} \tag{7.16}$$

and $x(z)$ is the unique solution in $\mathbb{C}^+$ of the equation

$$\frac{1-x(z)}{x(z)} = \frac{c}{p}\mathrm{tr}(x(z)\mathbf{I} - z\boldsymbol{\Sigma}_n^{-1})^{-1} . \tag{7.17}$$

First, we show that

$$x(0) = \lim_{z \to 0^+} x(z) = 1 - c . \tag{7.18}$$

If we assume that $x(0) = \infty$ then we get immediately the contradiction due to equation (7.17). Similarly, from (7.17) we conclude that $x(0) \neq 0$. This implies that $0 < x(0) < \infty$ and thus taking the limit $z \to 0^+$ from both sides of (7.17) we get (7.18).

Note that $\lim\limits_{z\to 0^+} \gamma_1(z) = \gamma_1$ which together with (7.18) and (7.19) implies

$$\left|\gamma_1 - \frac{1}{1-c}\text{tr}(\mathbf{\Sigma}_n^{-1}\mathbf{\Theta})\right| \longrightarrow 0 \quad \text{a. s.} \tag{7.19}$$

for $\dfrac{p}{n} \to c \in (0,1)$ as $n \to \infty$.

Next, we prove the following statement

$$\left|\gamma_2 - \frac{1}{p}\frac{1}{(1-c)^2}\left(||\mathbf{\Sigma}_n^{-1}||_F^2 + \frac{c}{p(1-c)}(||\mathbf{\Sigma}_n^{-1}||_{tr}^2\right)\right| \longrightarrow 0 \quad \text{a. s.} \tag{7.20}$$

for $\frac{p}{n} \to c \in (0,1)$ as $n \to \infty$.

Using the triangle inequality we rewrite the difference in (7.20) in the following way

$$\begin{aligned}\left|\gamma_2 - \frac{1}{p}\frac{1}{(1-c)^2}\left(||\mathbf{\Sigma}_n^{-1}||_F^2 + \frac{c}{p(1-c)}||\mathbf{\Sigma}_n^{-1}||_{tr}^2\right)\right| &\leq \left|\gamma_2 - \psi\right| \\ + \left|\psi - \frac{1}{p}\frac{1}{(1-c)^2}\left(||\mathbf{\Sigma}_n^{-1}||_F^2 + \frac{c}{p(1-c)}||\mathbf{\Sigma}_n^{-1}||_{tr}^2\right)\right|&,\end{aligned} \tag{7.21}$$

where $\psi = \dfrac{1}{(1-c)^2}\int\limits_{-\infty}^{+\infty}\dfrac{1}{\tau^2}dH(\tau) + \dfrac{c}{(1-c)^3}\left(\int\limits_{-\infty}^{+\infty}\dfrac{1}{\tau}dH(\tau)\right)^2$ is given in Theorem 3.1. Next we show that the right side of (7.21) vanishes almost surely as $n \to \infty$. Using Theorem 3.1 we get

$$\left|\gamma_2 - \psi\right| \longrightarrow 0 \text{ a. s. for } n \to \infty\,. \tag{7.22}$$

Next, we show that the second nonrandom term in (7.21) approaches to zero as $n \to \infty$. Using assumption (A1) it holds that $H_n(t)$ tends to $H(t)$ at all continuity points of $H(t)$. Thus,

$$\frac{1}{p}||\mathbf{\Sigma}_n^{-1}||_F^2 = \frac{1}{p}\text{tr}(\mathbf{\Sigma}_n^{-2}) = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{\tau_i^2} = \int\limits_{-\infty}^{+\infty}\frac{1}{\tau^2}dH_n(\tau) \overset{n\to\infty}{\longrightarrow} \int\limits_{-\infty}^{+\infty}\frac{1}{\tau^2}dH(\tau)\,, \tag{7.23}$$

where the last integral in (7.23) exists due to assumption (A3). Similarly, it holds that

$$\frac{1}{p^2}||\mathbf{\Sigma}_n^{-1}||_{tr}^2 \overset{n\to\infty}{\longrightarrow} \left(\int\limits_{-\infty}^{+\infty}\frac{1}{\tau}dH(\tau)\right)^2\,. \tag{7.24}$$

Using (7.23) and (7.24) we get

$$\left|\psi - \frac{1}{p}\frac{1}{(1-c)^2}\left(||\mathbf{\Sigma}_n^{-1}||_F^2 + \frac{c}{p(1-c)}||\mathbf{\Sigma}_n^{-1}||_{tr}^2\right)\right| \longrightarrow 0 \quad \text{for } n \to \infty \tag{7.25}$$

As a result, (7.22) and (7.25) complete the proof of Theorem 3.2.

Proof of Corollary 3.1. It holds that

$$\alpha_n^* = \frac{\frac{1}{p}\mathrm{tr}(\mathbf{S}_n^{-1}\mathbf{\Sigma}_n^{-1})\frac{1}{p}||\mathbf{\Pi}_0||_F^2 - \frac{1}{p}\mathrm{tr}(\mathbf{\Sigma}_n^{-1}\mathbf{\Pi}_0)\frac{1}{p}\mathrm{tr}(\mathbf{S}_n^{-1}\mathbf{\Pi}_0)}{\frac{1}{p}||\mathbf{S}_n^{-1}||_F^2\frac{1}{p}||\mathbf{\Pi}_0||_F^2 - \left(\frac{1}{p}\mathrm{tr}(\mathbf{S}_n^{-1}\mathbf{\Pi}_0)\right)^2} . \tag{7.26}$$

Using $\|\frac{1}{p}\mathbf{\Pi}_0\|_{tr} \leq \sqrt{p}\|\frac{1}{p}\mathbf{\Pi}_0\|_F = \frac{1}{\sqrt{p}}\|\mathbf{\Pi}_0\|_F \leq \sqrt{M}$ and $\|\frac{1}{p}\mathbf{\Sigma}_n^{-1}\|_{tr} \leq \infty$ from Assumption (A3), we get from Theorem 3.2

$$\begin{aligned}
&\frac{1}{p}||\mathbf{S}_n^{-1}||_F^2 \underset{\text{a.s.}}{\longrightarrow} \frac{1}{p}\left(\frac{1}{(1-c)^2}||\mathbf{\Sigma}_n^{-1}||_F^2 + \frac{c}{p(1-c)^3}||\mathbf{\Sigma}_n^{-1}||_{tr}^2\right),\\
&\frac{1}{p}\mathrm{tr}(\mathbf{S}_n^{-1}\mathbf{\Sigma}_n^{-1}) \underset{\text{a.s.}}{\longrightarrow} \frac{1}{p}\frac{1}{1-c}\mathrm{tr}(\mathbf{\Sigma}_n^{-2}) = \frac{1}{p}\frac{1}{1-c}\|\mathbf{\Sigma}_n^{-1}\|_F^2,\\
&\frac{1}{p}\mathrm{tr}(\mathbf{S}_n^{-1}\mathbf{\Pi}_0) \underset{\text{a.s.}}{\longrightarrow} \frac{1}{p}\frac{1}{1-c}\mathrm{tr}(\mathbf{\Sigma}_n^{-1}\mathbf{\Pi}_0)
\end{aligned}$$

as $p/n \longrightarrow c \in (0,1)$ for $n \longrightarrow \infty$. Substituting these results into (7.26), we obtain the first statement of the corollary. The second statement is proved in the same way.

Proof of Theorem 3.3. In the proof, we use a special case of the results derived by Rubio and Mestre (2011) which are summarized in Theorem 7.3.

Theorem 7.3. ***[Rubio and Mestre (2011)]*** *Assume that (A2) and (A3) hold and additionally some nonrandom $n \times n$ matrix $\mathbf{\Theta}$ has uniformly bounded trace norm at infinity. Let $\bar{\mathbf{S}}_n = \mathbf{X}_n'\mathbf{\Sigma}_n\mathbf{X}_n$. Then for $p/n \to c > 0$ as $n \to \infty$*

$$\left|tr\left(\mathbf{\Theta}(\bar{\mathbf{S}}_n - z\mathbf{I})^{-1}\right) - tr\left(\mathbf{\Theta}\right)x(z)\right| \longrightarrow 0 \quad a.\ s. \tag{7.27}$$

where $x(z)$ is a unique solution in $\mathbb{C}^+$ of the following equation

$$\frac{1+zx(z)}{x(z)} = \frac{c}{p}tr\left(x(z)\mathbf{I} + \mathbf{\Sigma}_n^{-1}\right)^{-1} . \tag{7.28}$$

First we consider for $c > 1$ the asymptotics of the quantity

$$\kappa_1 = \frac{1}{p}\mathrm{tr}(\mathbf{S}_n^{+}) = \frac{c^{-1}}{n}\mathrm{tr}(\bar{\mathbf{S}}_n^{-1}) = \lim_{z\to 0^+}\frac{c^{-1}}{n}\mathrm{tr}\left((\bar{\mathbf{S}}_n - z\mathbf{I})^{-1}\right) , \tag{7.29}$$

where the $n\times n$ matrix $\bar{\mathbf{S}}_n$ is defines in Theorem 7.3. The second equality in (7.29) follows from the fact that the matrices $\mathbf{S}_n$ and $\bar{\mathbf{S}}_n$ possess the same nonzero eigenvalues (see, e.g., Silverstein (2009)). Note that the limit in (7.29) exists because $c > 1$. Using Theorem 7.3 for $\mathbf{\Theta} = 1/n\mathbf{I}$ and setting $z \to 0^+$ we get almost surely for $p/n \to c \in (1,+\infty)$

$$\left|\kappa_1 - c^{-1}x(0)\right| \to 0 \quad \text{as } n \to \infty \tag{7.30}$$

and $x(0) = \lim\limits_{z\to 0^+} x(z)$ satisfies the equation

$$\frac{1}{x(0)} = \frac{c}{p}\mathrm{tr}\left(x(0)\mathbf{I} + \mathbf{\Sigma}_n^{-1}\right)^{-1} . \tag{7.31}$$

Note that $x(0)$ always exists in case of $c > 1$. In order to find the asymptotics for the quantity $\kappa_1(\mathbf{\Theta}) = 1/n\mathrm{tr}(\mathbf{S}_n^+\mathbf{\Theta})$ we recall the properties of the generalized inverse. Namely, if $\mathbf{\Theta}$ is positive definite then $\mathbf{\Theta}^{1/2}\mathbf{S}_n^+\mathbf{\Theta}^{1/2}$ is a generalized inverse of $\mathbf{\Theta}^{-1/2}\mathbf{S}_n\mathbf{\Theta}^{-1/2}$ which has the same non-zero eigenvalues as $\bar{\mathbf{S}}^* = \mathbf{X}'\mathbf{\Sigma}_n^{1/2}\mathbf{\Theta}^{-1}\mathbf{\Sigma}_n^{1/2}\mathbf{X}$. Consequently, since $tr(\mathbf{S}^+\mathbf{\Theta}) = tr(\mathbf{\Theta}^{1/2}\mathbf{S}_n^+\mathbf{\Theta}^{1/2})$ we get the same result as in case of $\kappa_1 = 1/n\mathrm{tr}(\mathbf{S}_n^+)$ with $\mathbf{\Sigma}_n^{-1}$ be replaced by

$$(\mathbf{\Sigma}_n^{1/2}\mathbf{\Theta}^{-1}\mathbf{\Sigma}_n^{1/2})^{-1} = \mathbf{\Sigma}_n^{-1/2}\mathbf{\Theta}\mathbf{\Sigma}_n^{-1/2}\,. \tag{7.32}$$

The second quantity of interest is

$$\kappa_2 = \frac{1}{p}||\mathbf{S}_n^+||_F^2 = \frac{c^{-1}}{n}||\bar{\mathbf{S}}_n^{-1}||_F^2 = c^{-1}\lim_{z\to 0^+}\frac{\partial}{\partial z}\frac{1}{n}\mathrm{tr}\left((\bar{\mathbf{S}}_n - z\mathbf{I})^{-1}\right)\,. \tag{7.33}$$

Again, we use Theorem 7.3 for $\frac{1}{n}\mathrm{tr}\left((\bar{\mathbf{S}}_n - z\mathbf{I})^{-1}\right)$, calculate the derivative with respect to z and set it to zero. As a result, we obtain the following identity for $p/n \to c \in (1, +\infty)$

$$\left|\kappa_2 - c^{-1}x'(0)\right| \to 0 \quad \text{as } n \to \infty\,, \tag{7.34}$$

where

$$\frac{1}{x'(0)} = \frac{1}{x^2(0)} - \frac{c}{p}\mathrm{tr}\left(x(0)\mathbf{I} + \mathbf{\Sigma}_n^{-1}\right)^{-2} \tag{7.35}$$

and $x(0)$ satisfies the equation (7.31). From (7.30) and (7.34) follows the statement of Theorem 3.3.

Proof of Proposition 3.1. Using Theorem 3.3 we get that $\frac{1}{p}\boldsymbol{\eta}'\mathbf{S}_n^+\boldsymbol{\xi} = \frac{1}{p}\mathrm{tr}(\boldsymbol{\xi}\boldsymbol{\eta}'\mathbf{S}_n^+) \underset{\text{a.s.}}{\to} c^{-1}y(\boldsymbol{\xi}\boldsymbol{\eta}')$, where $y(\boldsymbol{\xi}\boldsymbol{\eta}')$ satisfies the equation

$$\frac{1}{y(\boldsymbol{\xi}\boldsymbol{\eta}')} = \frac{c}{p}\mathrm{tr}\left[\left(\mathbf{\Sigma}_n^{-1/2}\boldsymbol{\xi}\boldsymbol{\eta}'\mathbf{\Sigma}_n^{-1/2} + y(\boldsymbol{\xi}\boldsymbol{\eta}')\mathbf{I}\right)^{-1}\right]\,. \tag{7.36}$$

Using the Sherman-Morrison formula (see, e.g., Horn and Johnson (1986)) we can rewrite the the right-hand side of (7.36) in the following way

$$\frac{c}{p}\mathrm{tr}\left[\frac{1}{y(\boldsymbol{\xi}\boldsymbol{\eta}')}\mathbf{I} - \frac{1}{y^2(\boldsymbol{\xi}\boldsymbol{\eta}')}\frac{\mathbf{\Sigma}_n^{-1/2}\boldsymbol{\xi}\boldsymbol{\eta}'\mathbf{\Sigma}_n^{-1/2}}{1 + \frac{1}{y(\boldsymbol{\xi}\boldsymbol{\eta}')}\boldsymbol{\eta}'\mathbf{\Sigma}_n^{-1}\boldsymbol{\xi}}\right] = \frac{c}{y(\boldsymbol{\xi}\boldsymbol{\eta}')}\left(1 - \frac{\boldsymbol{\eta}'\mathbf{\Sigma}_n^{-1}\boldsymbol{\xi}}{y(\boldsymbol{\xi}\boldsymbol{\eta}') + \boldsymbol{\eta}'\mathbf{\Sigma}_n^{-1}\boldsymbol{\xi}}\right)\,. \tag{7.37}$$

Combining (7.36) with (7.37) and multiplying both sides by $y(\boldsymbol{\xi}\boldsymbol{\eta}')$, we get

$$1 - c^{-1} = \frac{\boldsymbol{\eta}'\mathbf{\Sigma}_n^{-1}\boldsymbol{\xi}}{y(\boldsymbol{\xi}\boldsymbol{\eta}') + \boldsymbol{\eta}'\mathbf{\Sigma}_n^{-1}\boldsymbol{\xi}}\,, \tag{7.38}$$

which is a linear equation in $y(\boldsymbol{\xi}\boldsymbol{\eta}')$ with the solution given by

$$y(\boldsymbol{\xi}\boldsymbol{\eta}') = \frac{c^{-1}}{1 - c^{-1}}\boldsymbol{\eta}'\mathbf{\Sigma}_n^{-1}\boldsymbol{\xi} = \frac{1}{c-1}\boldsymbol{\eta}'\mathbf{\Sigma}_n^{-1}\boldsymbol{\xi}\,. \tag{7.39}$$

The last equality finishes the proof of Proposition 3.1.

Proof of Proposition 4.1. We write the proof in the case $c < 1$ because the case $c > 1$ is already handled in Section 3.2.

i). Consider the functionals $\widehat{\alpha}_n^*(\sigma\mathbf{\Pi}_0)$ and $\widehat{\beta}_n^*(\sigma\mathbf{\Pi}_0)$. Using (4.4) and (4.5) it holds that

$$\widehat{\alpha}_n^*(\sigma\mathbf{\Pi}_0) = \widehat{\alpha}_n^*(\mathbf{\Pi}_0) \tag{7.40}$$

$$\widehat{\beta}_n^*(\sigma\mathbf{\Pi}_0) = \frac{1}{\sigma}\widehat{\beta}_n^*(\mathbf{\Pi}_0)\,. \tag{7.41}$$

Putting (7.40) and (7.41) together with $\sigma\mathbf{\Pi}_0$ in (4.3) completes the proof of the first part of Proposition 4.1.

ii). From Corollary 3.1 it holds that

$$|\widehat{\alpha}_n^*(1/p\mathbf{I}) - \alpha^*(1/p\mathbf{I})| \longrightarrow 0 \quad \text{a. s.} \quad \text{for } n \to \infty \tag{7.42}$$

with

$$\alpha^*(1/p\mathbf{I}) = (1-c)\frac{||\mathbf{\Sigma}_n^{-1}||_F^2||1/p\mathbf{I}||_F^2 - \left(\operatorname{tr}(\mathbf{\Sigma}_n^{-1}1/p\mathbf{I})\right)^2}{\left(||\mathbf{\Sigma}_n^{-1}||_F^2 + \dfrac{c}{p(1-c)}||\mathbf{\Sigma}_n^{-1}||_{tr}^2\right)||1/p\mathbf{I}||_F^2 - \left(\operatorname{tr}(\mathbf{\Sigma}_n^{-1}1/p\mathbf{I})\right)^2} \tag{7.43}$$

Using (7.43) and noting that $\mathbf{\Sigma}_n^{-1} = \sigma\mathbf{I}$ we get that

$$\alpha^*(1/p\mathbf{I}) = 0. \tag{7.44}$$

Similarly,

$$\left|\widehat{\beta}_n^*(1/p\mathbf{I}) - \beta^*(1/p\mathbf{I})\right| \longrightarrow 0 \quad \text{a. s.} \quad \text{for } n \to \infty \tag{7.45}$$

and

$$\beta^*(1/p\mathbf{I}) = \frac{\operatorname{tr}(\mathbf{\Sigma}_n^{-1}1/p\mathbf{I})}{||1/p\mathbf{I}||_F^2}\left(1 - \frac{\alpha^*(1/p\mathbf{I})}{1-c}\right), \tag{7.46}$$

which together with (7.44) and $\mathbf{\Sigma}_n^{-1} = \sigma\mathbf{I}$ implies that

$$\beta^*(1/p\mathbf{I}) = p\sigma. \tag{7.47}$$

The equalities (7.44) and (7.47) with (4.3) complete the proof of the second part of Proposition 4.1.

iii). From (3.13) and (3.15) in Corollary 3.1 it follows that

$$\alpha_n^*(\mathbf{\Sigma}_n^{-1}) = 0 \quad \text{and} \quad \beta_n^*(\mathbf{\Sigma}_n^{-1}) = 1 \tag{7.48}$$

and, hence,

$$\widehat{\alpha}_n^*(\mathbf{\Sigma}_n^{-1}) \longrightarrow 0 \quad \text{and} \quad \widehat{\beta}_n^*(\mathbf{\Sigma}_n^{-1}) \longrightarrow 1 \text{ a.s. for } n \to \infty\,. \tag{7.49}$$

Substituting (7.49) in (4.3) completes the proof of the third part of Proposition 4.1.

References

[1] Anderson G.W., Guionnet A., and O. Zeitouni, (2010). An Introduction to Random Matrices. Cambridge University Press: New York, USA.

[2] Bai, J. (2003), Inferential theory for factor models of large dimensions, *Econometrica* 71, 135-171.

[3] Bai J., and S. Shi, (2011), Estimating high dimensional covariance matrices and its applications, *Annals of Economics and Finance* 12-2, 199-215.

[4] Bai Z.D., and J. W. Silverstein, (2010), Spectral Analysis of Large Dimensional Random Matrices, Springer: New York; Dordrecht; Heidelberg; London.

[5] Bai Z.D., Miao B.Q., and G.M. Pan, (2007), On asymptotics of eigenvectors of large sample covariance matrix, *Annals of Probability* 35(4), 1532-1572.

[6] Bodnar T., Gupta A.K., and N. Parolya (2014), On the strong convergence of the optimal linear shrinkage estimator for large dimensional covariance matrix, *Journal of Multivariate Analysis* 132, 215-228.

[7] Bühlmann, P. and S. van de Geer, (2011). Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer: Heidelberg; New York.

[8] Cai T., Lui W., and X. Luo, (2011), A constrained l_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association* 106, 594-607.

[9] Cai T., and M. Yuan, (2012), Adaptive covariance matrix estimation through block thresholding. *Annals of Statistics* 40, 2014-2042.

[10] Cai T., and H. Zhou, (2012), Minimax estimation of large covariance matrices under l_1 norm. *Statistica Sinica* 22, 1319-1378.

[11] Cai T., and X. Shen (2011), High-Dimensional Data Analysis. *Frontiers of Statistics*: Volume 2, World Scientific: Singapore.

[12] Efron, B. and Morris, C. (1976). Multivariate empirical Bayes and estimation of covariance matrices. *Annals of Statistics* 4, 22-32.

[13] Elton E. J., Gruber M.J., Brown S. J., and W. N. Goetzmann (2009), Modern Portfolio Theory and Investment Analysis. John Wiley & Sons Inc.: NY, USA.

[14] Fan, J., Fan, Y. and Lv, J. (2008), High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics* 147, 186-197.

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

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

[17] Ferraty, F., P. Hall, P. Vieu (2006). Most-predictive design points for functional data predictors. *Biometrika* 97, 807-824.

[18] Ferraty, F., P. Vieu (2006), Nonparametric Functional Data Analysis. Springer: New York.

[19] Girko V.L., (1990), Theory of Random Determinants. Kluwer Academic Publishers: Dordrecht.

[20] Girko V.L., and A. K. Gupta, (1994), Asymptotic behaviour of spectral function of empirical covariance matrices,*Random Operators and Stochastic Equations* 2, 43-60.

[21] Girko V.L., (1995), Statistical Analysis of Observations of Increasing Dimension, Kluwer Academic Publishers: Dordrecht; Boston; London.

[22] Girko V.L., and A. K. Gupta, (1996a), Canonical equation for the resolvent of empirical covariance matrices pencil. *Multidimensional Statistical Analysis and Theory of Random Matrices* (eds. A. K. Gupta and V. L. Girko), VSP Publishers: Netherlands.

[23] Girko V.L., and A. K. Gupta, (1996b), Multivariate elliptically contoured linear models and some aspects of the theory of random matrices. *Multidimensional Statistical Analysis and Theory of Random Matrices* (eds. A. K. Gupta and V. L. Girko), VSP Publishers: Netherlands.

[24] Gupta A.K., and S. Ofori-Nyarko, (1994), Estimation of generalized variance, precision and covariance matrices using the Pitman nearness criterion, *South Afr. Statist. J.* 28, 1-16.

[25] Gupta A.K., and S. Ofori-Nyarko, (1995a), Improved minimax estimators of normal covariance and precision matrices, *Statistics* 26 (1), 19-25.

[26] Gupta A.K., and S. Ofori-Nyarko, (1995b), Improved estimation of generalized variance and precision, *Statistics* 26(2), 99-109.

[27] Gupta A.K., and D.K. Nagar, (2000), Matrix Variate Distributions, Chapman and Hall/CRC: Boca Raton.

[28] Gupta A.K., T. Varga, and T. Bodnar, (2013), Elliptically Contoured Models in Statistics and Portfolio Theory: Theory and Applications. Springer: New York.

[29] Horn R., and C. Johnson, (1985), Matrix Analysis. Cambridge University Press:Cambridge, London, NY, New Rochelle, Melbourne, Sydney.

[30] Jing B.-Y., Pan G. and Shao Q.-M., and W. Zhou, (2010), Nonparametric estimate of spectral density functions of sample covariance matrices: A first step, *Annals of Statistics* 38(6), 3724-3750.

[31] Krishnamoorthy K., and A.K. Gupta, (1989), Improved minimax estimation of a normal precision matrix, *Can. J. Statist.* 17(1), 91-102.

[32] Kubokawa T., (2005), A revisit to estimation of the precision matrix of the Wishart distribution, *J. Statist. Res.* 39, 91-114.

[33] Kubokawa, T. and Srivastava, M. (2008). Estimation of the precision matrix of a singular Wishart distribution and its application in high-dimensional data. *Journal of Multivariate Analysis* 99, 1906-1928.

[34] Le Cam, L., and G. Lo Yang (2000), Asymptotics in Statistics: Some Basic Concepts. Springer: New York.

[35] Ledoit, O. and Wolf, M. (2004), A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* 88, 365-411.

[36] Marčenko, V. A. and Pastur, L. A. (1967), Distribution of eigenvalues for some sets of random matrices. *Sbornik: Mathematics* 1 457-483.

[37] Markowitz, H., (1952), Portfolio selection, *The Journal of Finance* 7, 77-91.

[38] Mestre, X. and Lagunas, M. (2006). Finite sample size effect on minimum variance beamformers: Optimum diagonal loading factor for large arrays. *IEEE Transactions on Signal Processing* 54, 69-82.

[39] Ramsay, J. O. and B. Silverman (2005), Functional Data Analysis. Springer: New York.

[40] Paul D., and J.W. Silverstein, (2009), No eigenvalues outside the support of limiting empirical spectral distribution of a separable covariance matrix. *Journal of Multivariate Analysis* 100, 37-57.

[41] Rohde A., and A. B. Tsybakov (2011), Estimation of high dimensional low-rank matrices. *The Annals of Statistics* 39, 887-930.

[42] Rubio F., and X. Mestre, (2011), Spectral convergence for a general class of random matrices, *Statistics and Probability Letters* 81, 592-602.

[43] Rubio F., Mestre, X., and D. Palomar, (2012), Performance Analysis and Optimal Selection of Large Minimum Variance Portfolios Under Estimation Risk, *IEEE Journal of Selected Topics in Signal Processing* 6, 4, 337-350.

[44] Sarr A., and A.K. Gupta, (2009), Estimation of the precision matrix of multivariate Kotz type model, *Journal of Multivariate Analysis* 100, 742-752.

[45] Silverstein, J. W., (1995), Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices *Journal of Multivariate Analysis* 55, 331-339.

[46] Silverstein, J. W. and Choi, S.-I. (1995), Analysis of the limiting spectral distribution of large-dimensional random matrices, *Journal of Multivariate Analysis* 54, 295-309.

[47] Silverstein, J. W., and Z.D. Bai, (1995), On the empirical distribution of eigenvalues of a class of large dimensional random matrices, *Journal of Multivariate Analysis* 54(2), pp. 175-192.

[48] Silverstein, J. W., (2009), The Stieltjes transform and its role in eigenvalue behavior of large dimensional random matrices. Random Matrix Theory and its Applications, Lecture Notes Series, World Scientific: Singapore.

[49] Srivastava, M. (2007). Multivariate theory for analyzing high dimensional data. *Journal of the Japan Statistical Society* 37, 53-86.

[50] Stein, C., (1956), Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954-1955, Vol. I 197-206. Univ. California Press, Berkeley.

[51] Stein, C. (1975). Estimation of a covariance matrix. 39th Annual Meeting IMS Rietz Lecture.

[52] Tsukuma H., and Y. Konno, (2006), On improved estimation of normal precision matrix and discriminant coefficients, *Journal of Multivariate Analysis* 97, 1477-1500.

[53] Van Trees H.L. (2002), Optimum Array Processing. John Wiley & Sons: NY, USA.

[54] Wang, C., Pan, G. M., Tong T. J. and Zhu, L. X (2014). Shrinkage estimator of large dimensional precision matrix using random matrix theory. *Statistica Sinica*. To appear.

[55] Yin, Y.Q., (1986), Limiting spectral distribution for a class of random matrices. *Journal of Multivariate Analysis* 20, 50-68.

[56] Xue L., and H. Zou, (2012), Regularized rank-based estimation of high-dimensional nonparanormal graphical models, *The Annals of Statistics* 40/5, 2541-2571.