

**Document Version**

Proof

**Citation (APA)**

Bodnar, T., & Parolya, N. (2026). Reviving pseudo-inverses: Asymptotic properties of large dimensional Moore–Penrose and ridge-type inverses with applications. *Annals of Statistics*, 54(2), 1053-1079. <https://doi.org/10.1214/25-AOS2602>

**Important note**

To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**

In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership. Unless copyright is transferred by contract or statute, it remains with the copyright holder.

**Sharing and reuse**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

# REVIVING PSEUDO-INVERSES: ASYMPTOTIC PROPERTIES OF LARGE DIMENSIONAL MOORE-PENROSE AND RIDGE-TYPE INVERSES WITH APPLICATIONS

BY TARAS BODNAR<sup>1,a</sup> AND NESTOR PAROLYA<sup>2,b</sup>

<sup>1</sup>Department of Management and Engineering, Linköping University, SE-581 83 Linköping, Sweden, <sup>a</sup>[taras.bodnar@liu.se](mailto:taras.bodnar@liu.se)

<sup>2</sup>Department of Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands, <sup>b</sup>[n.parolya@tudelft.nl](mailto:n.parolya@tudelft.nl)

In this paper, we derive high-dimensional asymptotic properties of the Moore-Penrose inverse and, as a byproduct, of various ridge-type inverses of the sample covariance matrix. In particular, the analytical expressions of the asymptotic behavior of the weighted sample trace moments of generalized inverse matrices are deduced in terms of the partial exponential Bell polynomials which can be easily computed in practice. The existent results for pseudo-inverses are extended in several directions: (i) First, the population covariance matrix is not assumed to be a multiple of the identity matrix; (ii) Second, the assumption of normality is not used in the derivation; (iii) Third, the asymptotic results are derived under the high-dimensional asymptotic regime. Our findings provide universal methodology for construction of fully data-driven improved shrinkage estimators of the precision matrix, optimal portfolio weights and beyond. It is found that the Moore-Penrose inverse acts asymptotically as a certain regularizer of the true covariance matrix and it seems that its proper transformation (shrinkage) performs similarly to or even outperforms the existing benchmarks in many applications, while keeping the computational time as minimal as possible.

**1. Introduction.** The covariance matrix and its inverse, the precision matrix, are present in many different applications. The covariance matrix is typically regarded as a multivariate measure of uncertainty and as a measure of linear dependence structure (see [Rencher and Christensen \(2012\)](#)), while the precision matrix is present in the formulae of the weights of optimal portfolios in finance (see [Ao et al. \(2019\)](#), [Cai et al. \(2020\)](#), [Kan et al. \(2022\)](#), [Bodnar et al. \(2023a\)](#), [Lassance et al. \(2024\)](#)), in the expression of the minimum variance filter in signal processing (see [Feng and Palomar \(2016\)](#)), in high-dimensional time-series analysis (cf., [Heiny \(2019\)](#), [Heiny and Mikosch \(2021\)](#)), in prediction and test theory in multivariate and high-dimensional statistics (see [Chen et al. \(2010\)](#), [Cai and Jiang \(2011\)](#), [Yao et al. \(2015\)](#), [Bodnar et al. \(2019\)](#), [Shi et al. \(2022\)](#)).

In practice, the unknown true covariance matrix is commonly estimated by the sample covariance matrix, while one of the most used estimators of the precision matrix is the inverse of the sample covariance matrix. The sample estimator of the covariance matrix possesses several important properties. First, it is an unbiased estimator of the true covariance matrix. Second, when the dimension of the data-generating model is fixed and the sample size tends to infinity, then both the sample covariance matrix and its inverse are consistent estimators for the population covariance matrix and the population precision matrix, respectively (see [Muirhead \(1982\)](#)).

The situation becomes challenging in the high-dimensional setting, i.e., when the model dimension is proportional to the sample size, especially if it is larger than the sample size.

---

*MSC2020 subject classifications:* Primary 60B20, 15A09, 62R07; secondary 62H12, 62F12.

*Keywords and phrases:* Moore-Penrose inverse, Bell polynomials, Sample covariance matrix, Random matrix theory, High-dimensional asymptotics.

It is known as the high-dimensional asymptotic regime or Kolmogorov asymptotics (see [Bai and Silverstein \(2010\)](#)). Neither the sample covariance matrix nor its inverse are consistent estimators of the corresponding population quantities without imposing some restrictions on the structure of the true covariance/precision matrix. The issue is even more difficult when the observation vectors are taken from a heavy-tailed distribution (see, e.g., [Heiny and Yao \(2022\)](#)) and the dimension of the data-generating model is larger than the sample size. In the latter case, the sample covariance matrix is singular and its inverse cannot be constructed (cf., [Muirhead \(1982\)](#), [Srivastava \(2003\)](#)).

Matrix algebra proposes several ways how a generalized or pseudo-inverse of a singular matrix can be defined (see, e.g., [Penrose \(1955\)](#), [Rao and Mitra \(1972\)](#), [Ben-Israel and Greville \(2003\)](#), [Wang et al. \(2018\)](#)). Although the generalized inverse is not uniquely determined in general, there exists a specific type of pseudo-inverse matrices, which is uniquely defined. This is the Moore-Penrose inverse, which is also a least squares generalized inverse (see, e.g., [Harville \(1997\)](#)).

Even though the properties of the Moore-Penrose inverse of deterministic matrices have been studied in the literature (cf., [Meyer \(1973\)](#), [Harville \(1997\)](#)), only a few results are available for the Moore-Penrose inverse of the sample covariance matrix which were derived under very strict assumptions imposed on the data-generating model. Regarding the finite-sample properties, the sample covariance matrix has a singular Wishart distribution under the assumption of normality ([Srivastava \(2003\)](#)). The density function of the Moore-Penrose inverse of a singular Wishart distributed random matrix was derived in [Bodnar and Okhrin \(2008\)](#). The resulting expression appears to be very complicated. As such, the moments (i.e., expectations of powers of the matrix) of the Moore-Penrose inverse of the sample covariance matrix cannot even be obtained under the assumption of normality imposed on the data-generating model when the true population covariance matrix is not restricted to be proportional to the identity matrix. Only the expressions of the upper and lower limits for the mean matrix and covariance matrix of the Moore-Penrose inverse of the sample covariance matrix were derived in [Imori and von Rosen \(2020\)](#) in the general case, while [Cook and Forzani \(2011\)](#) provided the exact mean matrix and the covariance matrix in the very restrictive special case when the true covariance matrix is proportional to the identity matrix. Both these results are non-asymptotic and were obtained when the data-generating model follows a multivariate normal distribution. No other results have been derived either for the Moore-Penrose inverse or for the ridge-type inverse in the literature in the non-asymptotic setting to the best of our knowledge, even though both matrices are widely used in practice (see, e.g., [Ben-Israel and Greville \(2003\)](#), [Wang et al. \(2018\)](#)). The ridge-type inverse was used for constructing shrinkage estimators of the precision matrix in [Kubokawa and Srivastava \(2008\)](#) and [Wang et al. \(2015\)](#). While [Kubokawa and Srivastava \(2008\)](#) derived a shrinkage estimator by assuming that the observation matrix is drawn from the multivariate normal distribution, [Wang et al. \(2015\)](#) derived the results in the general case by using the methods of random matrix theory. Lastly, while the limiting spectral distribution and central limit theorem (CLT) for linear spectral statistics of the Moore-Penrose inverse are discussed in [Bodnar et al. \(2016a\)](#), explicit results involving eigenvectors are not available. To date, this remains the only known asymptotic analysis of the large-sample behavior of the Moore-Penrose inverse of the sample covariance matrix.

We contribute to the existing literature in several directions. First, the asymptotic limits of higher-order weighted empirical trace moments (see, e.g., [Golub and Strakoš \(1994\)](#) and [Bai et al. \(1996\)](#)) of Moore-Penrose and ridge-type inverses are derived by imposing no specific distributional assumption on the data-generating model. This addresses a well-known challenge in multivariate statistics, particularly when the data dimension exceeds the sample size. Our general weighting scheme enables insights into the asymptotic behavior of the matrix

elements, the interaction between eigenvalues and eigenvectors, and lays the groundwork for developing new shrinkage estimators. Second, we assume only the existence of fourth moments and impose no structural assumptions on the population covariance matrix beyond positive definiteness and boundedness of its spectrum. Third, our results are established under a high-dimensional asymptotic regime, where both the dimension and sample size tend to infinity. Fourth, we extend existing results by deriving asymptotic properties of higher-order weighted trace moments for ridge-type inverses. Fifth, as a by-product of our methodology, we provide closed-form expressions for the asymptotic equivalents of the higher-order weighted trace moments of the sample covariance matrix itself. Finally, the developed theory is applied in the derivation of improved shrinkage estimators of the precision matrix and optimal portfolio weights.

The rest of the paper is structured as follows. In Section 2, the main theoretical results are presented. The weighted sample trace moments of the sample Moore-Penrose inverse are given in Section 2.1, while Section 2.2 and Section 2.3 present similar findings derived for the ridge-type inverse and the Moore-Penrose-ridge inverse. Section S.3 in the supplement (Bodnar and Parolya (2025)) provides additional results derived for the sample covariance matrix. The obtained theoretical results are used in the derivation of shrinkage estimators of the precision matrix in Section 3.1 and optimal portfolio weights in Section 3.2. The finite-sample performance of the theoretical findings is investigated via an extensive simulation study in Section 4. Finally, the proofs and some additional simulations are postponed to the supplementary material (Bodnar and Parolya (2025)).

**2. Asymptotic properties of pseudo-inverse matrices.** Let  $\mathbf{Y}_n = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)$  be the  $p \times n$  observation matrix with  $p > n$  and let  $\mathbb{E}(\mathbf{y}_i) = \boldsymbol{\mu}$  and  $\text{Cov}(\mathbf{y}_i) = \boldsymbol{\Sigma}$  for  $i \in 1, \dots, n$ . Throughout the paper, it is assumed that there exists a  $p \times n$  random matrix  $\mathbf{X}_n = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$  which consists of independent and identically distributed (i.i.d.) random variables with zero mean and unit variance such that

$$(1) \quad \mathbf{Y}_n \stackrel{d}{=} \boldsymbol{\mu} \mathbf{1}_n^\top + \boldsymbol{\Sigma}^{1/2} \mathbf{X}_n,$$

where  $\mathbf{1}_n$  denotes the  $n$ -dimensional vector of ones and the symbol  $\stackrel{d}{=}$  denotes the equality in distribution. Using the observation matrix  $\mathbf{Y}_n$ , the sample estimator of the covariance matrix  $\boldsymbol{\Sigma}$  is given by

$$(2) \quad \mathbf{S}_n = \frac{1}{n} \mathbf{Y}_n \mathbf{Y}_n^\top - \bar{\mathbf{y}}_n \bar{\mathbf{y}}_n^\top \quad \text{with} \quad \bar{\mathbf{y}}_n = \frac{1}{n} \mathbf{Y}_n \mathbf{1}_n.$$

In many statistical and machine learning applications, one seeks best possible estimators for certain functions of the high-dimensional precision matrix  $\boldsymbol{\Sigma}^{-1}$  using the sample covariance matrix  $\mathbf{S}_n$ . This problem becomes particularly challenging for the case  $p > n$ , due to the singularity of  $\mathbf{S}_n$ , as one must somehow “invert” a non-invertible matrix. There are plenty of ways to “invert”  $\mathbf{S}_n$  when  $p > n$ . For a while, we assume that one of the methods has been chosen and denote the corresponding generalized inverse by  $\mathbf{S}_n^\#(t)$  where  $t$  is a tuning parameter. Since none of these inverses presents a consistent estimator of  $\boldsymbol{\Sigma}^{-1}$  in the high-dimensional setting, much of the research focuses on finding an appropriate matrix- or vector-valued function  $f(\mathbf{S}_n^\#(t))$  that asymptotically minimizes the distance to the target transformation of the precision matrix  $g(\boldsymbol{\Sigma}^{-1})$  for some prespecified function  $g$  (see, e.g., Lam (2020) and references therein). This can be formalized in the following way

$$(3) \quad \underset{f \in \mathcal{C}_f}{\operatorname{argmin}} \|f(\mathbf{S}_n^\#(t)) - g(\boldsymbol{\Sigma}^{-1})\|_{l_q}^q$$

with  $\mathcal{C}_f$  being a class of functions of interest and  $\|\cdot\|_{l_q}$  denoting the  $l_q$ -norm. For example, by taking the quadratic norm  $l_2$  and  $\mathcal{C}_f$  as a class of linear functions, which act on the eigenvalues

of  $\mathbf{S}_n$ , one can be interested in the estimation of  $\Sigma^{-1}$  itself. In that case, the well-established linear shrinkage techniques presented, e.g., in [Ledoit and Wolf \(2004\)](#), [Bodnar et al. \(2014a\)](#) and [Bodnar et al. \(2016b\)](#) among many others, can be recovered. Similarly, nonlinear shrinkage estimators were derived in [Ledoit and Wolf \(2012\)](#), [Ledoit and Wolf \(2020\)](#) and [Ledoit and Wolf \(2022\)](#) for a nonlinear function  $f$ . Also, the functions  $f$  and  $g$  can be vector valued, which happens often in financial applications, e.g., shrinkage estimation of the optimal portfolio weights (see [Frahm and Memmel \(2010\)](#), [Golosnoy and Okhrin \(2007\)](#), [Bodnar et al. \(2018\)](#) among others). Taking a closer look at the aforementioned papers and the optimization problem (3) itself, we deal very often with the moments of the specific estimator of  $\Sigma^{-1}$ . More precisely, one has to establish the asymptotic behavior of the weighted trace moments alike  $\text{tr} \left[ (\mathbf{S}_n^\#(t))^m \Theta \right]$  with  $m = 1, 2, 3, \dots$  for some initial estimator  $\mathbf{S}_n^\#(t)$  of the precision matrix and a weighting matrix  $\Theta$  possibly dependent on  $\Sigma$ . More details can be found in Section 3, which is concerned with the application of the theoretical results of this paper to the shrinkage estimation techniques.

In this section, we investigate the properties of the weighted sample trace moments of three pseudo-inverses of the sample covariance matrix  $\mathbf{S}_n$  under the high-dimensional asymptotic regime. Namely, we deal with  $\text{tr} \left[ (\mathbf{S}_n^\#(t))^m \Theta \right]$  for  $m = 1, 2, \dots$  where  $\mathbf{S}_n^\#(t)$  is one of the pseudo-inverses considered: (i) Moore-Penrose inverse  $\mathbf{S}_n^+$  in Section 2.1, (ii) ridge-type inverse  $\mathbf{S}_n^-(t)$  in Section 2.2, and (iii) Moore-Penrose-ridge inverse  $\mathbf{S}_n^\pm(t)$  in Section 2.3. Both the ridge-type inverse and the Moore-Penrose-ridge inverse are functions on a tuning parameter  $t$  whose optimal value depends on the application at hand (see Section 3.1 and Section 3.2 for details).

Theoretical results are derived by imposing the following three assumptions.

**(A1)**  $\Sigma$  is a nonrandom positive definite matrix, which satisfies

$$0 < \inf_p \lambda_{\min}(\Sigma) \leq \sup_p \lambda_{\max}(\Sigma) < \infty.$$

**(A2)** The elements of  $\mathbf{X}_n$  have bounded  $4 + \varepsilon$  moments for some  $\varepsilon > 0$ .

**(A3)**  $\Theta$  is symmetric and

- (i) it has finite rank  $k$ , i.e.,  $\Theta = \sum_{i=1}^k \boldsymbol{\theta}_i \boldsymbol{\theta}_i^\top$ , and  $\sup_p \boldsymbol{\theta}_i^\top \boldsymbol{\theta}_i < \infty$  for all  $i = 1, \dots, k$ ;
- (ii) or  $\lambda_{\max}(\Theta) = o(1)$  as  $p \rightarrow \infty$ , otherwise.

The first assumption ensures that the smallest and the largest eigenvalues of the population covariance matrix are uniformly bounded in  $p$  away from zero and infinity. As such, the only source of the singularity of the sample covariance matrix  $\mathbf{S}_n$  is the lack of data, that is,  $n < p$ . This is a classical technical assumption imposed in random matrix theory (see, e.g., [Pan \(2014\)](#), [Ledoit and Wolf \(2020\)](#)). Assumption **(A2)** imposes no distributional assumptions on the data-generating model and presents the classical conditions used in random matrix theory (cf., [Bai and Silverstein \(2010\)](#), [Wang et al. \(2015\)](#), [Ledoit and Wolf \(2021\)](#)).

**REMARK 2.1** (Discussion on Assumption **(A3)**). While the matrix  $\Theta$  can be chosen with some flexibility, its definition inherently includes a normalization factor. For instance, in the case of sample trace moments of a generalized inverse, the choice  $\Theta = \mathbf{I}_p/p$  leads to all eigenvalues being equal to  $1/p$ . When  $\Theta$  has finite rank, a bounded Euclidean norm is assumed for each vector  $\boldsymbol{\theta}_i$ ,  $i = 1, \dots, k$ . Consequently, if these vectors are not sparse, their entries are expected to be relatively small. It is important to highlight that the assumption on  $\Theta$  is often of a technical nature. In many applications, expressions involving  $\mathbf{S}_n^\#(t)$  appear as ratios (see Section 3), and adjustments in the definition of  $\Theta$  can be made accordingly. Moreover, Assumption **(A3)** may be replaced by a slightly stronger but more convenient condition: that  $\Theta$  has bounded trace norm. This alternative assumption requires that

$$(A3') \sup_p \text{tr} [(\Theta^\top \Theta)^{1/2}] < \infty,$$

and is often easier to verify in practice. If  $\Theta$  is not symmetric, it should be replaced by its symmetrized version  $(\Theta + \Theta^\top)/2$ . Alternatively, bilinear forms can be addressed by using the identity

$$\theta\eta^\top + \eta\theta^\top = ((\theta + \eta)(\theta + \eta)^\top - (\theta - \eta)(\theta - \eta)^\top)/2$$

for any vectors  $\theta$  and  $\eta$  in  $\mathbb{R}^p$ . Therefore, Assumption (A3) is quite general and can either be readily verified or satisfied by normalizing  $\Theta$  via  $\text{tr} [(\Theta^\top \Theta)^{1/2}]$ . Indeed, if (A3) is not met, then the quantities of interest,  $\text{tr} [(\mathbf{S}_n^\#(t))^m \Theta]$ , either diverge or exhibit persistent fluctuations (e.g., asymptotic normality). The latter case will be explored in future research.

The typical examples of the weighting matrix  $\Theta$  include several key instances that are crucial in random matrix theory and statistical applications:

- Setting  $\Theta$  equal to  $\frac{1}{p}\mathbf{I}_p$  and  $m = 1$  leads to the Stieltjes transform restricted to positive real line, which is a concept often associated with classical random matrix theory. This transform is fundamental in understanding the spectral distribution of different types of random matrices (see [Bai and Silverstein \(2010\)](#)).
- When  $\Theta$  is defined as  $\theta\eta^\top$ , it pertains to bilinear or quadratic forms that involve eigenvectors. This approach focuses on the interaction between eigenvalues and eigenvectors.
- Choosing  $\Theta$  as  $\frac{1}{p}g(\Sigma)$  for some  $g(\cdot)$  corresponds to the Ledoit-Péché functionals, which involve nonlinear shrinkage formulas. These objects are vital for their ability to adjust covariance matrix estimators in the high-dimensional setting (see [Ledoit and Péché \(2011\)](#)).

The significance of exploring these different forms of  $\Theta$  includes the construction of new shrinkage estimators for cases where the ratio of the dimensionality to the sample size exceeds one, which will be elaborated upon later in Section 3.1 and Section 3.2. The specification of  $\Theta$  as  $\theta\theta^\top$  unveils information about the eigenvectors and arises frequently in portfolio analysis (see [Bodnar et al. \(2023a\)](#)). Finally, utilizing  $\Theta$  as  $\mathbf{e}_i\mathbf{e}_j^\top$ , where  $\mathbf{e}_i$  is a unit vector with one in the  $i$ -th position, helps in deriving the asymptotics of the matrix entries.

In the formulation of the theoretical results, we use the partial exponential Bell polynomials, which are defined by (see [Bell \(1928, 1934\)](#))

$$(4) \quad B_{m,k}(x_1, x_2, \dots, x_{m-k+1}) = \sum \frac{m!}{j_1!j_2!\dots j_{m-k+1}!} \left(\frac{x_1}{1!}\right)^{j_1} \left(\frac{x_2}{2!}\right)^{j_2} \dots \left(\frac{x_{m-k+1}}{(m-k+1)!}\right)^{j_{m-k+1}},$$

where the sum is taken over all sequences  $j_1, \dots, j_{m-k+1}$  of non-negative integers such that  $\sum_{l=1}^{m-k+1} j_l = k$  and  $\sum_{l=1}^{m-k+1} lj_l = m$ . In practice, the Bell polynomials can be easily computed in the R-package *kStatistics*, see also [Di Nardo et al. \(2008\)](#).

Further, throughout the paper, we will need a specific function  $v(t) : \mathbb{R}^+ \rightarrow \mathbb{R}$ , which satisfies the following asymptotic equation

$$(5) \quad \frac{1}{p} \text{tr} \left[ (v(t)\Sigma + \mathbf{I}_p)^{-1} \right] = \frac{c_n - 1 + tv(t)}{c_n},$$

with  $c_n = p/n$ . It is easy to verify that there exists the limit of  $v(t)$  as  $t \rightarrow 0$  and, consequently, we denote it by  $v(0) = \lim_{t \rightarrow 0} v(t)$ . Moreover, from (5), it is straightforward to find that  $v(0)$  is a unique solution of the equation

$$(6) \quad \frac{1}{p} \text{tr} \left[ (v(0)\Sigma + \mathbf{I}_p)^{-1} \right] = \frac{c_n - 1}{c_n} \quad \text{with} \quad c_n = \frac{p}{n}.$$

Although the function  $v(t)$  is defined as a solution of a nonlinear equation (5) and it is not available in closed form, we show that it is a decreasing function in Proposition 2.1 whose proof is given in the supplement.

**PROPOSITION 2.1.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A3),  $v(t)$  is strictly decreasing for  $t \geq 0$  and  $c_n > 1$ .*

All results in this paper are expressed in terms of the function  $v(t)$ , its value at zero  $v(0)$ , and its derivatives  $v^{(1)}(t), v^{(2)}(t), \dots$ , along with their respective values at zero,  $v^{(1)}(0), v^{(2)}(0), \dots$ , and so on. Furthermore, we demonstrate how these quantities can be consistently estimated in the high-dimensional asymptotic regime where  $p/n \rightarrow c > 1$  as  $n \rightarrow \infty$ . Accordingly, we now present the consistent estimators of the function  $v(t)$ , its derivatives, and their values at zero for  $m = 0, 1, 2, \dots$  expressed as

$$(7) \quad \hat{v}^{(m)}(t) = (-1)^m m! c_n \left( \frac{1}{p} \text{tr} [(\mathbf{S}_n^-(t))^{m+1}] - t^{-(m+1)} \frac{c_n - 1}{c_n} \right),$$

$$(8) \quad \hat{v}^{(m)}(0) = (-1)^m m! c_n \frac{1}{p} \text{tr} [(\mathbf{S}_n^+)^{m+1}],$$

with  $\hat{v}^{(0)}(t) = \hat{v}(t)$  and  $\hat{v}^{(0)}(0) = \hat{v}(0)$ . Here,  $\mathbf{S}_n^+$  and  $\mathbf{S}_n^-(t)$  are the Moore–Penrose and the ridge-type inverse, defined at the beginning of Sections 2.1 and 2.2, respectively. The expressions (7) and (8) are deduced in Corollaries 2.2 and 2.5.

Throughout the paper, we will assume that the dimension  $p$  is a function of the sample size  $n$ , i.e.,  $p = p(n)$ . All limits are for  $n \rightarrow \infty$ , unless explicitly stated otherwise. At the beginning of the supplement, an abstract *sketch of the proof* is presented, which applies to all theorems proved in Section 2.

**2.1. Weighted moments of the sample Moore–Penrose inverse.** The Moore–Penrose inverse of  $\mathbf{S}_n$  is uniquely defined as the matrix  $\mathbf{S}_n^+$  which fulfills the following four conditions:

- (i)  $\mathbf{S}_n^+ \mathbf{S}_n \mathbf{S}_n^+ = \mathbf{S}_n^+$ ,      (ii)  $\mathbf{S}_n \mathbf{S}_n^+ \mathbf{S}_n = \mathbf{S}_n$ ,
- (iii)  $(\mathbf{S}_n^+ \mathbf{S}_n)^\top = \mathbf{S}_n^+ \mathbf{S}_n$ ,      (iv)  $(\mathbf{S}_n \mathbf{S}_n^+)^\top = \mathbf{S}_n \mathbf{S}_n^+$ .

Theorem 2.1 presents the behavior of the weighted sample trace moments of the Moore–Penrose inverse,  $\text{tr}((\mathbf{S}_n^+)^m \Theta)$ , under the high-dimensional asymptotic regime.

**THEOREM 2.1.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A3), it holds for  $m = 1, 2, \dots$  that*

$$(9) \quad \left| \text{tr} [(\mathbf{S}_n^+)^m \Theta] - s_m(\Theta) \right| \xrightarrow{a.s.} 0 \quad \text{for } p/n \rightarrow c \in (1, \infty) \quad \text{as } n \rightarrow \infty,$$

where

$$(10) \quad s_m(\Theta) = \sum_{k=1}^m \frac{(-1)^{m+k+1} k!}{m!} d_k(\Theta) B_{m,k} \left( v^{(1)}(0), \dots, v^{(m-k+1)}(0) \right),$$

with

$$(11) \quad d_k(\Theta) = \text{tr} \left\{ (v(0)\Sigma + \mathbf{I}_p)^{-1} \left[ \Sigma (v(0)\Sigma + \mathbf{I}_p)^{-1} \right]^k \Theta \right\}, \quad k = 1, 2, \dots,$$

$v(0)$  the unique solution of the equation (6) and its derivatives satisfy

$$(12) \quad v^{(1)}(0) = - \frac{1}{\frac{1}{v(0)^2} - c_n \frac{1}{p} \text{tr} \left\{ \left[ \Sigma (v(0)\Sigma + \mathbf{I}_p)^{-1} \right]^2 \right\}},$$

$v^{(2)}(0), \dots, v^{(m)}(0)$  are computed recursively by

$$(13) \quad v^{(m)}(0) = -v^{(1)}(0) \sum_{k=2}^m (-1)^k k! h_{k+1} B_{m,k} \left( v^{(1)}(0), \dots, v^{(m-k+1)}(0) \right) \quad \text{with}$$

$$(14) \quad h_k = \frac{1}{[v(0)]^k} - c_n \frac{1}{p} \text{tr} \left\{ \left[ \Sigma (v(0)\Sigma + \mathbf{I}_p)^{-1} \right]^k \right\}, \quad k = 1, 2, \dots$$

In Corollary 2.1, we compute closed-form expressions in case  $m = 1, 2, 3, 4$ .

**COROLLARY 2.1.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A3) we get (9) with*

$$\begin{aligned} s_1(\Theta) &= -v^{(1)}(0)d_1(\Theta), \quad s_2(\Theta) = \frac{1}{2}v^{(2)}(0)d_1(\Theta) - [v^{(1)}(0)]^2d_2(\Theta), \\ s_3(\Theta) &= -\frac{1}{6}v^{(3)}(0)d_1(\Theta) + v^{(1)}(0)v^{(2)}(0)d_2(\Theta) - [v^{(1)}(0)]^3d_3(\Theta), \\ s_4(\Theta) &= \frac{1}{24}v^{(4)}(0)d_1(\Theta) - \frac{1}{12} \left( 4v^{(1)}(0)v^{(3)}(0) + 3[v^{(2)}(0)]^2 \right) d_2(\Theta) \\ &\quad + \frac{3}{2}[v^{(1)}(0)]^2v^{(2)}(0)d_3(\Theta) - [v^{(1)}(0)]^4d_4(\Theta), \end{aligned}$$

for  $p/n \rightarrow c \in (1, \infty)$  as  $n \rightarrow \infty$  where  $v(0)$  is the unique solution of (6),

$$\begin{aligned} v^{(1)}(0) &= -h_2^{-1}, \quad v^{(2)}(0) = -2[v^{(1)}(0)]^3h_3, \quad v^{(3)}(0) = -6[v^{(1)}(0)]^2v^{(2)}(0)h_3 + 6[v^{(1)}(0)]^4h_4, \\ v^{(4)}(0) &= - \left( 8[v^{(1)}(0)]^2v^{(3)}(0) + 6v^{(1)}(0)[v^{(2)}(0)]^2 \right) h_3 + 36[v^{(1)}(0)]^3v^{(2)}(0)h_4 - 24[v^{(1)}(0)]^5h_5, \end{aligned}$$

and  $d_k(\Theta)$  and  $h_k$  are defined in (11) and (14), respectively.

The scenario with  $m = 1$  illustrates that the Moore-Penrose inverse inherently functions as a particular form of regularization. Specifically, it exhibits asymptotic behavior similar/proportional to

$$(v(0)\Sigma + \mathbf{I}_p)^{-1} \Sigma (v(0)\Sigma + \mathbf{I}_p)^{-1} = \frac{1}{v(0)^2} (\Sigma + v^{-1}(0)\mathbf{I}_p)^{-1} \Sigma (\Sigma + v^{-1}(0)\mathbf{I}_p)^{-1}.$$

This observation prompts the exploration of an alternative type of inverse in Section 2.3.

Corollary 2.2 provides the results for a special case  $\Theta = (1/p)\mathbf{I}_p$  with all eigenvalues equal to  $1/p$ . Consequently, it fulfills the condition about the weighting matrix  $\Theta$  stated in Assumption (A3).

**COROLLARY 2.2.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A2) it holds for  $m = 1, 2, \dots$  that*

$$(15) \quad \left| \frac{1}{p} \text{tr} [(\mathbf{S}_n^+)^m] - \frac{(-1)^{m-1} v^{(m-1)}(0)}{(m-1)! c_n} \right| \xrightarrow{a.s.} 0 \quad \text{for } p/n \rightarrow c \in (1, \infty) \quad \text{as } n \rightarrow \infty,$$

where  $v^{(m)}(0)$  is defined in (6), (12), and (13) for  $m = 0, 1, \dots$  with  $v^{(0)}(0) = v(0)$ .

The results of Corollary 2.2 have a number of important applications. In particular, they allow us to approximate  $v(0)$  and  $v^{(m)}(0)$ ,  $m = 1, 2, \dots$  even though  $v(0)$  is defined as a solution of the nonlinear equation (6) which can be solved only numerically and the definition

of  $v^{(m)}(0)$ ,  $m = 1, 2, \dots$ , depends on  $v(0)$ . Using the findings of Corollary 2.2, we get the closed-form expressions of consistent estimators for  $v^{(m)}(0)$ ,  $m = 0, 1, \dots$  expressed with the formula (8), namely

$$\hat{v}^{(m)}(0) = (-1)^m m! c_n \frac{1}{p} \text{tr} [(\mathbf{S}_n^+)^{m+1}].$$

In Figure 1, the finite-sample performance of the estimators in (8) is depicted when  $m = 0$  and  $m = 1$  under the assumption of the normal distribution and  $t$ -distribution for  $p/n \in (1, 5]$  and  $n \in \{100, 250, 500\}$ . We observe that the estimators of  $v(0)$  and  $v^{(1)}(0)$  converge to their true values already for small sample size  $n = 100$  when the concentration ratio  $c_n$  is larger than 2. When  $c_n \in (1, 2]$ , then a larger sample size is needed. These findings hold independently of the distribution used to generate the elements of  $\mathbf{X}_n$ , while the convergence is slower for the  $t$ -distribution. Finally, we note that the convergence is faster for  $v(0)$ .

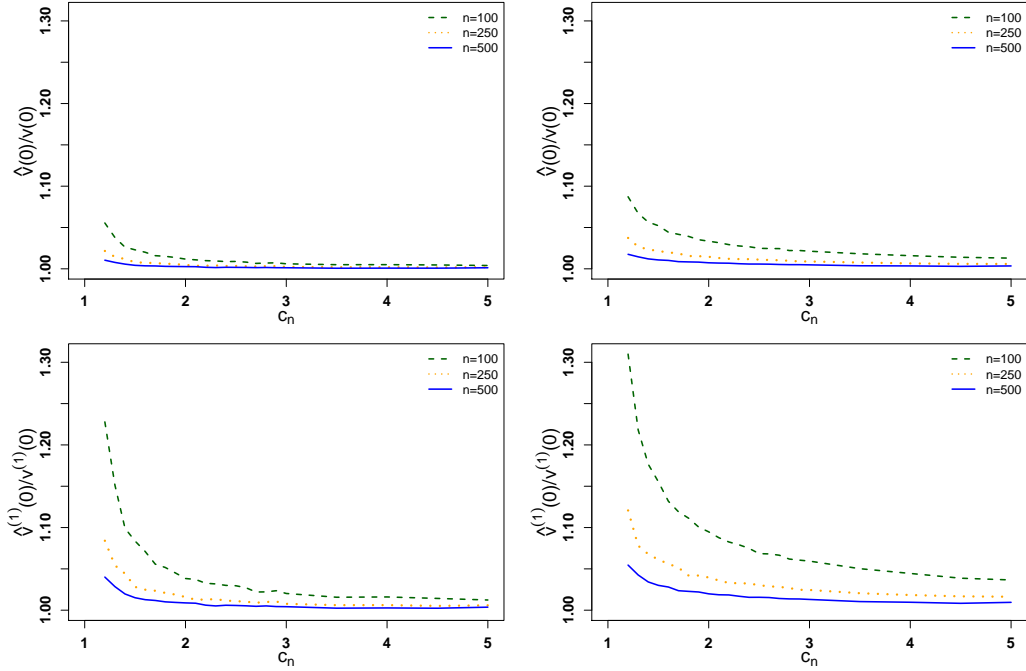


FIG 1. Finite-sample performance of the estimators for  $v(0)$  and  $v^{(1)}(0)$  when  $c_n \in (1, 5]$ ,  $n \in \{100, 250, 500\}$ , and the elements of  $\mathbf{X}_n$  drawn from the normal (first column) and scale  $t$ -distribution (second column).

Next, we present the general results of Theorem 2.1 for  $\Sigma = \mathbf{I}_p$ .

**COROLLARY 2.3.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1) with  $\Sigma = \mathbf{I}_p$ . Then, under Assumptions (A2)-(A3) it holds that*

$$(16) \quad \left| \text{tr}((\mathbf{S}_n^+)^m \Theta) - \frac{(-1)^{m-1} v^{(m-1)}(0)}{(m-1)! c_n} \text{tr}(\Theta) \right| \xrightarrow{a.s.} 0$$

for  $p/n \rightarrow c \in (1, \infty)$  as  $n \rightarrow \infty$  with

$$(17) \quad v(0) = \frac{1}{c_n - 1}, \quad v^{(1)}(0) = \frac{-c_n}{(c_n - 1)^3} \quad \text{and}$$

$$(18) \quad v^{(m)}(0) = \frac{c_n}{(c_n - 1)^3} \sum_{k=2}^m (-1)^k k! (c_n - 1)^{(k+1)} (1 - c_n^{-k}) B_{m,k} \left( v^{(1)}(0), \dots, v^{(m-k+1)}(0) \right).$$

As a special case of Corollary 2.3, we get

$$\begin{aligned} \left| \operatorname{tr}(\mathbf{S}_n^+ \Theta) - \frac{1}{(c_n - 1)c_n} \operatorname{tr}(\Theta) \right| &\xrightarrow{a.s.} 0, & \left| \operatorname{tr}((\mathbf{S}_n^+)^2 \Theta) - \frac{1}{(c_n - 1)^3} \operatorname{tr}(\Theta) \right| &\xrightarrow{a.s.} 0, \\ \left| \operatorname{tr}((\mathbf{S}_n^+)^3 \Theta) - \frac{c_n + 1}{(c_n - 1)^5} \operatorname{tr}(\Theta) \right| &\xrightarrow{a.s.} 0, & \left| \operatorname{tr}((\mathbf{S}_n^+)^4 \Theta) - \frac{c_n^2 + 3c_n + 1}{(c_n - 1)^7} \operatorname{tr}(\Theta) \right| &\xrightarrow{a.s.} 0, \end{aligned}$$

for  $p/n \rightarrow c \in (1, \infty)$  as  $n \rightarrow \infty$ .

The results of Corollary 2.3 extend the previous findings of Cook and Forzani (2011) and Imori and von Rosen (2020). Under the additional assumption that the elements of the matrix  $\mathbf{Y}_n$  are normally distributed, we get that  $(n - 1)\mathbf{S}_n$  has a  $p$ -dimensional singular Wishart distribution with  $n - 1$  degrees of freedom and identity covariance matrix (see, e.g., Srivastava (2003), Bodnar and Okhrin (2008)). Cook and Forzani (2011) derived the exact expression of the mean matrix of the singular Wishart distribution under the assumption that the population covariance matrix is proportional to the identity matrix. They proved that

$$\mathbb{E}(\mathbf{S}_n^+) = (n - 1)\mathbb{E}(((n - 1)\mathbf{S}_n)^+) = \frac{(n - 1)^2}{p(p - n)} \mathbf{I}_p.$$

Furthermore, it holds that  $\frac{(n - 1)^2}{p(p - n)} \rightarrow \frac{1}{c(c - 1)}$  for  $p/n \rightarrow c \in (1, \infty)$  as  $n \rightarrow \infty$ . On the other side, the application of Corollary 2.3 leads to the same result when one defines  $\Theta = \mathbf{e}_i \mathbf{e}_j^\top$ ,  $i, j = 1, \dots, p$ , where  $\mathbf{e}_i$  is the vector of zeros except the  $i$ -th element which is one.

Imori and von Rosen (2020) extended the results of Cook and Forzani (2011) by considering a singular Wishart distribution with an arbitrary covariance matrix and deriving the lower and upper limits of the mean matrix and covariance matrix. The findings of Corollary 2.3 provide further generalization by establishing the limiting behaviour of the elements of the Moore-Penrose inverse of the sample covariance matrix. Moreover, the results of Theorem 2.1, Corollary 2.1, Corollary 2.2, and Corollary 2.3 are derived in the general case without the assumption of normality imposed on the elements of  $\mathbf{Y}_n$ .

*2.2. Weighted moments of the sample ridge-type inverse.* The proof of Theorem 2.1 can be adjusted to derive the weighted trace moments of the ridge-type inverse of the sample covariance matrix, expressed as

$$(19) \quad \mathbf{S}_n^-(t) = (\mathbf{S}_n + t\mathbf{I}_p)^{-1} \quad \text{for } t > 0.$$

The findings are summarized in Theorem 2.2. In contrast to the statement of Theorem 2.1, the results of Theorem 2.2 are derived for  $c \in (0, +\infty)$ .

**THEOREM 2.2.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A3) for any  $t > 0$  and  $m = 0, 1, 2, \dots$ , it holds that*

$$(20) \quad \left| \operatorname{tr}((\mathbf{S}_n^-(t))^{m+1} \Theta) - \tilde{s}_{m+1}(t, \Theta) \right| \xrightarrow{a.s.} 0 \quad \text{for } p/n \rightarrow c \in (0, +\infty) \quad \text{as } n \rightarrow \infty,$$

where

$$(21) \quad \begin{aligned} \tilde{s}_{m+1}(t, \Theta) &= \sum_{l=1}^m t^{-(m-l)-1} \sum_{k=1}^l \frac{(-1)^{l+k} k!}{l!} d_k(t, \Theta) B_{l,k} \left( v^{(1)}(t), v^{(2)}(t), \dots, v^{(l-k+1)}(t) \right) \\ &+ t^{-m-1} d_0(t, \Theta) \quad \text{with} \end{aligned}$$

$$(22) \quad d_k(t, \Theta) = \operatorname{tr} \left\{ (v(t)\Sigma + \mathbf{I}_p)^{-1} \left[ \Sigma (v(t)\Sigma + \mathbf{I}_p)^{-1} \right]^k \Theta \right\}, \quad k = 0, 1, 2, \dots,$$

$v(t)$  is the unique solution of the equation (5) and its derivatives satisfy

$$(23) \quad v^{(1)}(t) = -\frac{1}{v(t)^{-2} - c_n \frac{1}{p} \operatorname{tr} \left\{ \left[ \Sigma(v(t)\Sigma + \mathbf{I}_p)^{-1} \right]^2 \right\}},$$

$v^{(2)}(t), \dots, v^{(m)}(t)$  are computed recursively by

$$(24) \quad v^{(m)}(t) = -v^{(1)}(t) \sum_{k=2}^m (-1)^k k! h_{k+1}(t) B_{m,k} \left( v^{(1)}(t), \dots, v^{(m-k+1)}(t) \right), \quad \text{with}$$

$$(25) \quad h_k(t) = [v(t)]^{-k} - c_n \frac{1}{p} \operatorname{tr} \left\{ \left[ \Sigma(v(t)\Sigma + \mathbf{I}_p)^{-1} \right]^k \right\}, \quad k = 1, 2, \dots$$

The almost sure convergence stated in Theorem 2.2 remains valid and, moreover, it is uniform on the larger domain  $\mathbb{C} \setminus \mathbb{R}^-$ . This extension follows from the Weierstrass convergence theorem Ahlfors (1953). Alternatively, the same conclusion can be justified using arguments found in Bai and Silverstein (2004), Rubio and Mestre (2011), Rubio et al. (2012), or Bodnar et al. (2023a), which rely on Vitali's theorem concerning the uniform convergence of sequences of uniformly bounded holomorphic functions to a holomorphic limit (see Rudin (1987), Hille (2002)).

Corollary 2.4 provides the formulae for  $\tilde{s}_m(t, \Theta)$  for  $m = 0, 1, 2, 3$ .

**COROLLARY 2.4.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A3) for any  $t > 0$ , we get (9) with*

$$\begin{aligned} \tilde{s}_1(t, \Theta) &= t^{-1} d_0(t, \Theta), \quad \tilde{s}_2(t, \Theta) = t^{-2} d_0(t, \Theta) + t^{-1} v^{(1)}(t) d_1(t, \Theta), \\ \tilde{s}_3(t, \Theta) &= t^{-3} d_0(t, \Theta) + t^{-2} v^{(1)}(t) d_1(t, \Theta) + t^{-1} \left\{ [v^{(1)}(t)]^2 d_2(t, \Theta) - \frac{1}{2} v^{(2)}(t) d_1(t, \Theta) \right\}, \\ \tilde{s}_4(t, \Theta) &= t^{-4} d_0(t, \Theta) + t^{-3} v^{(1)}(t) d_1(t, \Theta) + t^{-2} \left\{ [v^{(1)}(t)]^2 d_2(t, \Theta) - \frac{1}{2} v^{(2)}(t) d_1(t, \Theta) \right\} \\ &\quad + t^{-1} \left\{ \frac{1}{6} v^{(3)}(t) d_1(t, \Theta) - v^{(1)}(t) v^{(2)}(t) d_2(t, \Theta) + [v^{(1)}(t)]^3 d_3(t, \Theta) \right\}, \end{aligned}$$

for  $p/n \rightarrow c \in (0, \infty)$  as  $n \rightarrow \infty$  where  $d_k(t, \Theta)$  and  $h_k(t)$  are defined in (22) and (25),  $v(t)$  is the unique solution of (5) and

$$v^{(1)}(t) = -h_2(t)^{-1}, \quad v^{(2)}(t) = -2[v^{(1)}(t)]^3 h_3(t), \quad v^{(3)}(t) = -6[v^{(1)}(t)]^2 v^{(2)}(t) h_3(t) + 6[v^{(1)}(t)]^4 h_4(t).$$

It is interesting to note that  $\tilde{s}_{m+1}(t, \Theta)$  in (21) can be recursively computed by

$$(26) \quad \tilde{s}_{m+1}(t, \Theta) = \frac{1}{t} \tilde{s}_m(t, \Theta) + \frac{1}{t} \sum_{k=1}^m \frac{(-1)^{m+k} k!}{m!} d_k(t, \Theta) B_{m,k} \left( v^{(1)}(t), v^{(2)}(t), \dots, v^{(m-k+1)}(t) \right).$$

Using (26), we present the results in a special case for  $\Theta = (1/p)\mathbf{I}_p$  in Corollary 2.5.

**COROLLARY 2.5.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A2) for any  $t > 0$ , it holds for  $m = 0, 1, \dots$  that*

$$\left| \frac{1}{p} \operatorname{tr} \left[ (\mathbf{S}_n^-(t))^{m+1} \right] - t^{-(m+1)} \frac{c_n - 1}{c_n} - \frac{(-1)^m v^{(m)}(t)}{m! c_n} \right| \xrightarrow{\text{a.s.}} 0 \text{ for } p/n \rightarrow c \in (0, \infty) \text{ as } n \rightarrow \infty,$$

where  $v^{(m)}(t)$  is defined in (5), (23), and (24) for  $m = 0, 1, \dots$  with  $v^{(0)}(t) = v(t)$ .

Several interesting results follow from the statement of Corollary 2.5. The sample trace moments  $\text{tr}((\mathbf{S}_n^-(t))^{m+1})$  are not defined for  $t = 0$ , while their limiting values can be well approximated by the expressions of the form  $at^{-(m+1)} + b(t)$ , where the constant  $a$  depends on  $c_n$  only. Moreover, equations (5), (23), and (24) in Theorem 2.2 which are used in the computations of  $v^{(m)}(t)$ ,  $m = 0, 1, \dots$ , are well defined at  $t = 0$  and they coincide with the corresponding expressions (6), (12), and (13) derived in Theorem 2.1 for the Moore-Penrose inverse.

Another important application of Corollary 2.5 provides consistent estimators for  $v^{(m)}(t)$ ,  $m = 0, 1, \dots$ , for  $t > 0$  despite these functions are not defined in the closed form in Theorem 2.2. Their consistent estimators are given by formula (7), namely

$$\hat{v}^{(m)}(t) = (-1)^m m! c_n \left( \frac{1}{p} \text{tr} [(\mathbf{S}_n^-(t))^{m+1}] - t^{-(m+1)} \frac{c_n - 1}{c_n} \right).$$

Figure 2 depicts the estimators of  $v(t)$  and  $v^{(1)}(t)$ , normalized by the corresponding true values for  $t \in [0.1, 5]$ . Two data-generating models are considered, which are the normal distribution and the scaled  $t$ -distribution with five degrees of freedom. The covariance matrix  $\Sigma$  is set following the procedure described in Section 4. In the figure, we observe that  $\hat{v}(t)$  and  $\hat{v}^{(1)}(t)$  converge to  $v(t)$  and  $v^{(1)}(t)$ , respectively, already for  $n = 100$ . The results are similar for all considered values of the concentration ratio  $c_n$  and the distributional models used to generate  $\mathbf{X}_n$ . It is interesting that for small values of  $t$ , the convergence is slower. Also, the convergence is faster when data are generated from the normal distribution.

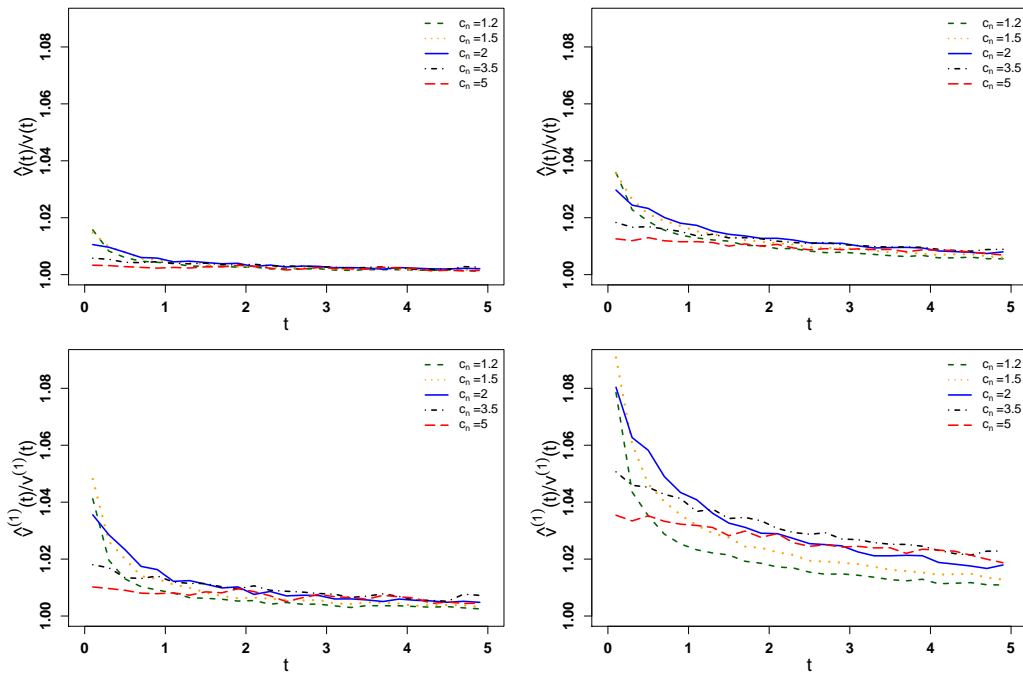


FIG 2. Finite-sample performance of the estimators for  $v(t)$  and  $v^{(1)}(t)$  when  $t \in [0.1, 5]$ ,  $c_n \in \{1.2, 1.5, 2, 3.5, 5\}$ ,  $n = 100$ , and the elements of  $\mathbf{X}_n$  drawn from the normal (first column) and scale  $t$ -distribution (second column).

Corollary 2.6 presents the findings under the additional assumption that the true population covariance matrix is the identity matrix.

**COROLLARY 2.6.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1) with  $\Sigma = \mathbf{I}_p$ . Then, under Assumptions (A2)-(A3) for any  $t > 0$ , it holds that*

$$(27) \quad \left| \text{tr}((\mathbf{S}_n^-(t))^{m+1} \Theta) - \left( t^{-(m+1)} \frac{c_n - 1}{c_n} + \frac{(-1)^m v^{(m)}(t)}{m! c_n} \right) \text{tr}(\Theta) \right| \xrightarrow{a.s.} 0$$

for  $p/n \rightarrow c \in (0, +\infty)$  as  $n \rightarrow \infty$  with

$$(28) \quad v(t) = \frac{2}{(c_n - 1 + t) + \sqrt{(c_n - 1 + t)^2 + 4t}}, \quad v^{(1)}(t) = \frac{-1}{v(t)^{-2} - c_n(v(t) + 1)^{-2}},$$

$$(29) \quad v^{(m)}(t) = -v^{(1)}(t) \sum_{k=2}^m (-1)^k k! \left( v(t)^{-(k+1)} - c_n(v(t) + 1)^{-(k+1)} \right) B_{m,k} \left( v^{(1)}(t), \dots, v^{(m-k+1)}(t) \right).$$

**2.3. Weighted moments of the sample Moore-Penrose-ridge inverse.** Taking into account the properties of the Moore-Penrose and ridge-type inverses, one can take the advantage of both estimators and build a new inverse for any  $t > 0$  by

$$(30) \quad \mathbf{S}_n^\pm(t) = (\mathbf{S}_n + t\mathbf{I}_p)^{-1} \mathbf{S}_n (\mathbf{S}_n + t\mathbf{I}_p)^{-1}.$$

This matrix inherits the properties of the Moore-Penrose inverse for  $t \rightarrow 0$  because by the spectral decomposition, we have

$$\lim_{t \rightarrow 0} \mathbf{S}_n^\pm(t) = \lim_{t \rightarrow 0} (\mathbf{S}_n + t\mathbf{I}_p)^{-1} \mathbf{S}_n (\mathbf{S}_n + t\mathbf{I}_p)^{-1} = \lim_{t \rightarrow 0} \mathbf{U} \text{diag} \left\{ \frac{d_1}{(d_1 + t)^2}, \dots, \frac{d_p}{(d_p + t)^2} \right\} \mathbf{U}^\top = \mathbf{S}_n^+,$$

where  $\mathbf{S}_n = \mathbf{U} \text{diag} \{d_1, \dots, d_p\} \mathbf{U}^\top$  is the spectral decomposition of  $\mathbf{S}_n$ , with  $d_1, d_2, \dots, d_p$  denoting the eigenvalues of  $\mathbf{S}_n$ , and  $\mathbf{U}$  the corresponding matrix of eigenvectors.

On the other side, it is related to the ridge inverse since

$$(31) \quad \mathbf{S}_n^\pm(t) = (\mathbf{S}_n + t\mathbf{I}_p)^{-1} - t(\mathbf{S}_n + t\mathbf{I}_p)^{-2}.$$

In fact, we have already all of the instruments to find the asymptotic behavior for the weighted sample trace moments of  $\mathbf{S}^\pm(t)$ . Indeed, using binomial theorem we get

$$(32) \quad \begin{aligned} (\mathbf{S}_n^\pm(t))^m &= ((\mathbf{S}_n + t\mathbf{I}_p)^{-1} - t(\mathbf{S}_n + t\mathbf{I}_p)^{-2})^m = \sum_{k=0}^m \binom{m}{k} (\mathbf{S}_n + t\mathbf{I}_p)^{-(m-k)} (-1)^k t^k (\mathbf{S}_n + t\mathbf{I}_p)^{-2k} \\ &= \sum_{k=0}^m (-1)^k t^k \binom{m}{k} (\mathbf{S}_n + t\mathbf{I}_p)^{-(m+k)}. \end{aligned}$$

Thus, in order to find the limit of  $\text{tr}((\mathbf{S}_n^\pm(t))^m \Theta)$  we only need to know the limits of the functionals  $\text{tr}((\mathbf{S}_n + t\mathbf{I}_p)^{-(m+k)} \Theta)$ , which have already been derived in Theorem 2.2. As a result, we get the following theorem.

**THEOREM 2.3.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A3) for any  $t > 0$  and  $m = 1, 2, \dots$ , it holds that*

$$(33) \quad \begin{aligned} & \left| \text{tr}((\mathbf{S}_n^\pm(t))^m \Theta) - \dot{\mathfrak{s}}_m(t, \Theta) \right| \xrightarrow{a.s.} 0 \quad \text{for } p/n \rightarrow c \in (0, +\infty) \quad \text{as } n \rightarrow \infty \quad \text{with} \\ & \dot{\mathfrak{s}}_m(t, \Theta) = \sum_{k=0}^m (-1)^k t^k \binom{m}{k} \tilde{\mathfrak{s}}_{m+k}(t, \Theta), \end{aligned}$$

where  $\tilde{\mathfrak{s}}_{m+k}(t, \Theta)$  can be computed using (21) for  $k = 0, 1, \dots, m$  and  $m = 1, 2, \dots$

In Corollary 2.7 we present another expression of  $\dot{\mathfrak{s}}_m(t, \Theta)$  which follows from the recursive computation of  $\tilde{\mathfrak{s}}_{m+k}(t, \Theta)$  as given in (26). The proof of Corollary 2.7 is given in the supplement.

COROLLARY 2.7. *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A3) for any  $t > 0$ , it holds that*

$$(34) \quad \dot{\tilde{s}}_m(t, \Theta) = -D_m(t, \Theta) + \sum_{k=2}^m (-1)^k \binom{m}{k} \sum_{j=1}^{k-1} t^j D_{m+j}(t, \Theta), \text{ where}$$

$$(35) \quad D_m(t, \Theta) = \sum_{k=1}^m \frac{(-1)^{m+k} k!}{m!} d_k(t, \Theta) B_{m,k} \left( v^{(1)}(t), v^{(2)}(t), \dots, v^{(m-k+1)}(t) \right).$$

Interestingly, the summands which are unbounded for  $\tilde{s}_m(t, \Theta)$  as  $t \rightarrow 0$  are all canceled in the limiting behavior of for  $\dot{\tilde{s}}_m(t, \Theta)$ . Since all derivatives of  $v(t)$  and the functions  $d_k(t, \Theta)$  for  $k \geq 1$  are all well-defined and bounded in  $t = 0$ , it holds that the limits of  $\dot{\tilde{s}}_m(t, \Theta)$  are also well defined at zero and, moreover,

$$(36) \quad \dot{\tilde{s}}_m(0, \Theta) = -D_m(0, \Theta) = s_m(\Theta).$$

Hence, the limiting behavior of  $\dot{\tilde{s}}_m(0, \Theta)$  coincides with the one, obtained in the case of the Moore-Penrose inverse.

In Corollary 2.8, the results for  $m = 1$  and  $m = 2$  are summarized.

COROLLARY 2.8. *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A3) for any  $t > 0$  we get (9) with*

$$(37) \quad \dot{\tilde{s}}_1(t, \Theta) = -v^{(1)}(t) d_1(t, \Theta),$$

$$(38) \quad \begin{aligned} \dot{\tilde{s}}_2(t, \Theta) = & - \left\{ [v^{(1)}(t)]^2 d_2(t, \Theta) - \frac{1}{2} v^{(2)}(t) d_1(t, \Theta) \right\} \\ & + t \left\{ \frac{1}{6} v^{(3)}(t) d_1(t, \Theta) - v^{(1)}(t) v^{(2)}(t) d_2(t, \Theta) + [v^{(1)}(t)]^3 d_3(t, \Theta) \right\}, \end{aligned}$$

for  $p/n \rightarrow c \in (0, +\infty)$  as  $n \rightarrow \infty$  where  $d_k(t, \Theta)$  and  $h_k(t)$  are defined in (22) and (25), and  $v^{(m)}(t)$ ,  $m = 0, 1, \dots$ , are defined in (5), (23), and (24) with  $v^{(0)}(t) = v(t)$ .

Next, we present the results for two other important special cases  $\Theta = (1/p)\mathbf{I}_p$  and  $\Sigma = \mathbf{I}_p$  in Corollary 2.9 and Corollary 2.10, respectively. The proof of Corollary 2.9 is given in the supplement.

COROLLARY 2.9. *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A2) for any  $t > 0$ , it holds for  $m = 1, 2, \dots$  that*

$$\left| \frac{1}{p} \text{tr} [(\mathbf{S}_n^\pm(t))^m] - \frac{(-1)^{m-1}}{c_n} \sum_{k=0}^m \binom{m}{k} \frac{v^{(m+k-1)}(t)}{(m+k-1)!} t^k \right| \xrightarrow{a.s.} 0 \text{ for } p/n \rightarrow c \in (0, +\infty) \text{ as } n \rightarrow \infty,$$

where  $v^{(m)}(t)$  is defined in (5), (23), and (24) for  $m = 0, 1, \dots$  with  $v^{(0)}(t) = v(t)$ .

COROLLARY 2.10. *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1) with  $\Sigma = \mathbf{I}_p$ . Then, under Assumptions (A2)-(A3) for any  $t > 0$ , it holds that*

$$(39) \quad \left| \text{tr}((\mathbf{S}_n^\pm(t))^{m+1} \Theta) - \frac{(-1)^{m-1}}{c_n} \sum_{k=0}^m \binom{m}{k} \frac{v^{(m+k-1)}(t)}{(m+k-1)!} t^k \text{tr}(\Theta) \right| \xrightarrow{a.s.} 0$$

for  $p/n \rightarrow c \in (0, +\infty)$  as  $n \rightarrow \infty$  with  $v(t)$  and  $v^{(m)}(t)$  defined in (28) and (29).

The results of Section 2 show that the three generalized inverses have different limiting behavior. This is easily visualized with the findings presented in Corollary 2.3, Corollary 2.6, and Corollary 2.10 when  $\Sigma = \mathbf{I}_p$ . Namely, it holds that

$$\begin{aligned} & \left| \text{tr}(\mathbf{S}_n^+ \Theta) - \frac{1}{(c_n - 1)c_n} \text{tr}(\Theta) \right| \xrightarrow{a.s.} 0, \\ & \left| \text{tr}(\mathbf{S}_n^-(t) \Theta) - \left( \frac{v(t)}{c_n} + \frac{1}{t} \frac{c_n - 1}{c_n} \right) \text{tr}(\Theta) \right| \xrightarrow{a.s.} 0, \\ & \left| \text{tr}(\mathbf{S}_n^\pm(t) \Theta) - \left( \frac{v(t)}{c_n} + t \frac{v^{(1)}(t)}{c_n} \right) \text{tr}(\Theta) \right| \xrightarrow{a.s.} 0 \end{aligned}$$

for  $p/n \rightarrow c \in (1, +\infty)$  as  $n \rightarrow \infty$  where

$$v(t) = \frac{2}{(c_n - 1 + t) + \sqrt{(c_n - 1 + t)^2 + 4t}}, \quad v^{(1)}(t) = \frac{-1}{v(t)^{-2} - c_n(v(t) + 1)^{-2}}.$$

Interestingly, all three generalized inverses consistently estimate non-diagonal elements of the precision matrix elementwise, which follows directly by setting

$$(40) \quad \Theta = \frac{1}{2}(\mathbf{e}_i \mathbf{e}_j^\top + \mathbf{e}_j \mathbf{e}_i^\top) = \frac{1}{4} \left( (\mathbf{e}_i + \mathbf{e}_j)(\mathbf{e}_i + \mathbf{e}_j)^\top - (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^\top \right),$$

where  $1 \leq i < j \leq p$  and  $\mathbf{e}_i$  is a unit vector with one in the  $i$ -th position.

Another notable observation concerns the case  $t \rightarrow 0$ . When  $p > n$ , then the application of (17) shows that the Moore-Penrose inverse and the Moore-Penrose-ridge inverse exhibit the same limiting behavior, while the ridge-type inverse does not. This is not surprising, as it follows directly from their definitions:  $\lim_{t \rightarrow 0} \mathbf{S}_n^\pm(t) = \mathbf{S}_n^+$ , while  $\lim_{t \rightarrow 0} \mathbf{S}_n^-(t) = \infty$ . The last two equalities also hold for  $\Sigma \neq \mathbf{I}_p$  following (36). Thus, the Moore-Penrose-ridge estimator  $\mathbf{S}_n^\pm(t)$  effectively approximates the Moore-Penrose inverse for small values of  $t$ . This feature is particularly useful when only mild regularization is needed. For instance, when the concentration ratio  $c_n = p/n$  is close to one, a regime in which the Moore-Penrose inverse  $\mathbf{S}_n^+$  is known to be numerically unstable. This insight supports the practical relevance of the Moore-Penrose-ridge inverse  $\mathbf{S}_n^\pm$  in scenarios where  $p$  and  $n$  are nearly equal.

For  $p < n$ , the sample covariance matrix  $\mathbf{S}_n$  is nonsingular. In this case, the application of Theorem S.2.1 and Corollary S.2.4 from the supplement (Bodnar and Parolya (2025)), and (31) yields for  $\Sigma = \mathbf{I}_p$  that

$$\begin{aligned} & \left| \text{tr}(\mathbf{S}_n^{-1} \Theta) - \frac{1}{1 - c_n} \text{tr}(\Theta) \right| \xrightarrow{a.s.} 0, \quad \left| \text{tr}(\mathbf{S}_n^-(t) \Theta) - \frac{1}{t + w(t)} \text{tr}(\Theta) \right| \xrightarrow{a.s.} 0, \\ & \left| \text{tr}(\mathbf{S}_n^\pm(t) \Theta) - \left( \frac{1}{t + w(t)} - t \frac{w^{(1)}(t) + 1}{(t + w(t))^2} \right) \text{tr}(\Theta) \right| \xrightarrow{a.s.} 0 \end{aligned}$$

for  $p/n \rightarrow c \in (0, 1)$  as  $n \rightarrow \infty$  where

$$w(t) = \frac{(1 - c_n - t) + \sqrt{(1 - c_n - t)^2 + 4t}}{2}, \quad w^{(1)}(t) = \frac{c_n w(t)}{(t + w(t))^2 + t c_n},$$

with  $w(0) = 1 - c_n$  and  $w^{(1)}(0) = c_n/(1 - c_n)$ . Consequently, in contrast to the case  $p > n$ , all three inverses exhibit the same limiting behavior. Furthermore, the application of (40) yields that all three inverses consistently estimate non-diagonal elements of the precision matrix elementwise, similarly to the case  $p > n$ . Finally, the comparison of formulae after Corollary 2.3 and at the end of Section S.2 in the supplement (Bodnar and Parolya (2025)) leads to the conclusion that  $\text{tr}((\mathbf{S}_n^+)^m \Theta)$  and  $\text{tr}((\mathbf{S}_n^{-1})^m \Theta)$  have the same asymptotic behavior up to the sign of  $c_n - 1$  when  $m = 2, 3, 4$ .

**3. Applications.** The derived theoretical results have a lot of applications in high-dimensional statistics. Maybe, the most obvious one would be the approximation of the weighted functionals  $\text{tr}\left(f(\mathbf{S}_n^\#(t))\Theta\right)$  by Taylor expansion up to a specific order  $m$ , assuming that the test function  $f$  is smooth enough and  $\mathbf{S}_n^\#(t)$  denotes an estimator of the precision matrix. The most prominent examples, however, are the construction of fully data-driven shrinkage estimators for the precision matrix and for the weights of the global minimum variance portfolio in the case of a singular sample covariance matrix. Later on, we choose  $\mathbf{S}_n^\#(t)$  to be one of the three generalized inverses considered in Section 2, i.e.,  $\mathbf{S}_n^\#(t) \in \{\mathbf{S}_n^+, \mathbf{S}_n^-(t), \mathbf{S}_n^\pm(t)\}$ .

3.1. *Shrinkage estimator of the high-dimensional precision matrix.* Following [Bodnar et al. \(2016b\)](#) the general linear shrinkage estimator for the precision matrix is given by

$$(41) \quad \widehat{\Pi}_{GSE} = \alpha_n \mathbf{S}_n^\#(t) + \beta_n \Pi_0,$$

where  $\Pi_0$  is a target matrix, which typically is chosen to be equal to the identity matrix when no *a priori* information about the precision matrix  $\Sigma^{-1}$  is available. Potentially, the estimator (41) is nonlinear since all the three pseudo-inverses of  $\mathbf{S}_n$  are nonlinear functions of  $\Sigma$  and also they are nonlinear functions of the eigenvalues of  $\mathbf{S}_n$ . For example, the Moore-Penrose inverse inverts the non-zero eigenvalues and keeps zero eigenvalues untouched. A similar structure as (41) obeys the non-linear quadratic shrinkage estimator of [Ledoit and Wolf \(2022\)](#) with specific  $\alpha_n, \beta_n$  and the data-driven equivariant (with same eigenvectors as  $\mathbf{S}_n$ ) target  $\Pi_0$ , which come from the nonlinear Ledoit-Péché shrinkage formula (see [Ledoit and Péché, 2011](#)). A proper shrinkage of  $\mathbf{S}_n^\#(t)$  could result in a better estimator of  $\Sigma^{-1}$ . Moreover, it is interesting how it compares to the well-established nonlinear shrinkage technique. The choice of  $\Pi_0$  is free to some extent but we will concentrate us on  $\Pi_0 = \mathbf{I}_p$  in simulations to have a fair comparison with common rotation-equivariant estimators (see, e.g., [Ledoit and Wolf, 2004, 2012; Bodnar et al., 2014b](#)).

The optimal shrinkage intensities  $\alpha_n^*$  and  $\beta_n^*$  are determined by minimizing the Frobenius (quadratic) loss<sup>1</sup> for a given nonrandom target matrix  $\Pi_0$  given by (see [Haff, 1979; Krishnamoorthy and Gupta, 1989; Yang and Berger, 1994; Wang et al., 2015; Bodnar et al., 2016b](#))

$$(42) \quad L_{F;n}^2 = \|\widehat{\Pi}_{GSE} \Sigma - \mathbf{I}_p\|_F^2.$$

The solution is summarized in Theorem 3.1.

**THEOREM 3.1.** *For any fixed value of  $t > 0$ , the loss function  $L_{F;n}^2$  is minimized at*

$$(43) \quad \alpha_n^*(\mathbf{S}_n^\#(t)) = \frac{\text{tr}(\mathbf{S}_n^\#(t)\Sigma) - \text{tr}\left(\frac{\Sigma\Pi_0}{\|\Sigma\Pi_0\|_F}\right) \text{tr}\left(\mathbf{S}_n^\#(t) \frac{\Sigma^2\Pi_0}{\|\Sigma\Pi_0\|_F}\right)}{\|\mathbf{S}_n^\#(t)\Sigma\|_F^2 - \left[\text{tr}\left(\mathbf{S}_n^\#(t) \frac{\Sigma^2\Pi_0}{\|\Sigma\Pi_0\|_F}\right)\right]^2}$$

and

$$(44) \quad \beta_n^*(\mathbf{S}_n^\#(t)) = \frac{\text{tr}\left(\frac{\Sigma\Pi_0}{\|\Sigma\Pi_0\|_F}\right) \|\mathbf{S}_n^\#(t)\Sigma\|_F^2 - \text{tr}(\mathbf{S}_n^\#(t)\Sigma) \text{tr}\left(\mathbf{S}_n^\#(t) \frac{\Sigma^2\Pi_0}{\|\Sigma\Pi_0\|_F}\right)}{\|\mathbf{S}_n^\#(t)\Sigma\|_F^2 - \left[\text{tr}\left(\mathbf{S}_n^\#(t) \frac{\Sigma^2\Pi_0}{\|\Sigma\Pi_0\|_F}\right)\right]^2} \|\Sigma\Pi_0\|_F^{-1}.$$

<sup>1</sup>This loss function is a bit different than given in [Bodnar et al. \(2016b\)](#). The reason for that is the discussion in Section 5 of [Ledoit and Wolf \(2022\)](#) about the singular case, i.e.,  $\Sigma^{-1}$  should be avoided in the loss function because of potential numerical instabilities.

Additionally, when  $\mathbf{S}_n^\#(t) \in \{\mathbf{S}_n^-(t), \mathbf{S}_n^\pm(t)\}$  the optimal value of  $t$  is found by maximizing

$$(45) \quad L_{F;n,2}^2(\mathbf{S}_n^\#(t)) = \frac{\left[ \text{tr}(\mathbf{S}_n^\#(t)\mathbf{\Sigma}) - \text{tr}\left(\mathbf{S}_n^\#(t) \frac{\mathbf{\Sigma}^2\mathbf{\Pi}_0}{\|\mathbf{\Sigma}\mathbf{\Pi}_0\|_F}\right) \text{tr}\left(\frac{\mathbf{\Sigma}\mathbf{\Pi}_0}{\|\mathbf{\Sigma}\mathbf{\Pi}_0\|_F}\right) \right]^2}{\|\mathbf{S}_n^\#(t)\mathbf{\Sigma}\|_F^2 - \left[ \text{tr}\left(\mathbf{S}_n^\#(t) \frac{\mathbf{\Sigma}^2\mathbf{\Pi}_0}{\|\mathbf{\Sigma}\mathbf{\Pi}_0\|_F}\right) \right]^2}.$$

The proof of Theorem 3.1 is given in the supplement. It is important to note that after dividing by  $p$  the numerator and denominator in the formulas for  $\alpha_n^*(\mathbf{S}_n^\#(t))$ ,  $\beta_n^*(\mathbf{S}_n^\#(t))$  and  $L_{F;n,2}^2(\mathbf{S}_n^\#(t))$ , no additional assumption on  $\mathbf{\Pi}_0$  is needed since all ratios in (43), (44) and (45) are self-normalized. For example, by taking

$$\Theta = \frac{\mathbf{\Sigma}\mathbf{\Pi}_0}{\sqrt{p}\|\mathbf{\Sigma}\mathbf{\Pi}_0\|_F}$$

in Corollary 2.1 and using Jensen's inequality, we get that  $\text{tr}(\Theta) \leq 1$ . Similarly, for  $\Theta = \frac{1}{p}\mathbf{\Sigma}^2$  it holds that  $\frac{1}{p}\text{tr}(\mathbf{\Sigma}^2) \leq \frac{p\lambda_{\max}^2(\mathbf{\Sigma})}{p} = \lambda_{\max}^2(\mathbf{\Sigma}) < \infty$ . In the same way, it can be shown that  $\Theta = \frac{1}{p}\mathbf{\Sigma}$  and  $\Theta = \frac{\mathbf{\Sigma}^2\mathbf{\Pi}_0}{\sqrt{p}\|\mathbf{\Sigma}\mathbf{\Pi}_0\|_F}$  possess bounded trace norms. The latter properties verify Assumption (A3).

Since  $\alpha_n^*(\mathbf{S}_n^\#(t))$ ,  $\beta_n^*(\mathbf{S}_n^\#(t))$  and  $L_{F;n,2}^2(\mathbf{S}_n^\#(t))$  depend on the unknown population covariance matrix  $\mathbf{\Sigma}$ , they cannot be computed in practice and thus we refer to  $\alpha_n^*(\mathbf{S}_n^\#(t))$  and  $\beta_n^*(\mathbf{S}_n^\#(t))$  as the *oracle* estimators. In the high-dimensional singular setting, i.e.,  $p > n$ , Bodnar et al. (2016b) found the almost sure asymptotic equivalents for  $\alpha_n^*(\mathbf{S}_n^+)$  and  $\beta_n^*(\mathbf{S}_n^+)$  and estimate them consistently only in the case of  $\mathbf{\Sigma} = \sigma^2\mathbf{I}_p$ . The results of Section 2 allow us to establish the limiting behaviour of  $\alpha_n^*(\mathbf{S}_n^\#(t))$  and  $\beta_n^*(\mathbf{S}_n^\#(t))$ , and to estimate consistently their limiting values  $\alpha^*$  and  $\beta^*$  in a very general case. It is enough to find the exact limits of  $\frac{1}{p}\|\mathbf{S}_n^\#\mathbf{\Sigma}\|_F^2 = \frac{1}{p}\text{tr}\left([\mathbf{S}_n^\#(t)]^2\mathbf{\Sigma}^2\right)$ ,  $\frac{1}{p}\text{tr}(\mathbf{S}_n^\#(t)\mathbf{\Sigma})$  and  $\frac{1}{p}\text{tr}\left(\mathbf{S}_n^\#(t) \frac{\mathbf{\Sigma}^2\mathbf{\Pi}_0}{\|\mathbf{\Sigma}\mathbf{\Pi}_0\|_F}\right)$  by using Corollary 2.1, Corollary 2.4 and Corollary 2.8, and to estimate them consistently, together with some other quantities which depend on the population covariance matrix, not on its inverse. Thus, it is straightforward to determine the almost sure equivalents of the optimal shrinkage intensities. We present the results in the following subsections for all three pseudo-inverses.

**3.1.1. Results for the Moore-Penrose inverse.** In the case of the Moore-Penrose inverse, the optimization problem and the resulting optimal shrinkage intensities do not depend on the tuning parameter  $t$ . This fact simplifies considerably the construction of the optimal shrinkage estimator. Let  $\alpha_{MP;n}^* = \alpha_n^*(\mathbf{S}_n^+)$  and  $\beta_{MP;n}^* = \beta_n^*(\mathbf{S}_n^+)$ . The asymptotic deterministic equivalents to  $\alpha_{MP;n}^*$  and  $\beta_{MP;n}^*$  are derived in Theorem S.5.1 from the supplementary material (see Bodnar and Parolya (2025)) by using the results presented in Corollary 2.1.

As a result, consistent estimators  $\hat{\alpha}_{MP}^*$  and  $\hat{\beta}_{MP}^*$  are obtained, which are summarized in Theorem 3.2. Interestingly, all formulas are given in the closed form and, as such, no numerical computation is needed to construct consistent estimators for  $\alpha_{MP}^*$  and  $\beta_{MP}^*$ .

**THEOREM 3.2.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A2), consistent estimators for  $\alpha_{MP}^*$  and  $\beta_{MP}^*$  are given by*

$$(46) \quad \hat{\alpha}_{MP}^* = \frac{\hat{d}_1\left(\frac{1}{p}\mathbf{\Sigma}\right)\hat{q}_2\left(\frac{1}{p}\mathbf{\Pi}_0^2\right) - \hat{d}_1\left(0, \frac{1}{p}\mathbf{\Sigma}^2\mathbf{\Pi}_0\right)\hat{q}_1\left(\frac{1}{p}\mathbf{\Pi}_0\right)}{-\frac{1}{\hat{h}_2}\left(\hat{d}_2\left(\frac{1}{p}\mathbf{\Sigma}^2\right) - \hat{d}_1\left(\frac{1}{p}\mathbf{\Sigma}^2\right)\frac{\hat{h}_3}{\hat{h}_2}\right)\hat{q}_2\left(\frac{1}{p}\mathbf{\Pi}_0^2\right) - \frac{1}{\hat{h}_2}\hat{d}_1^2\left(0, \frac{1}{p}\mathbf{\Sigma}^2\mathbf{\Pi}_0\right)},$$

$$(47) \quad \hat{\beta}_{MP}^* = \frac{\left( \hat{d}_2 \left( \frac{1}{p} \Sigma^2 \right) - \hat{d}_1 \left( \frac{1}{p} \Sigma^2 \right) \frac{\hat{h}_3}{\hat{h}_2} \right) \hat{q}_1 \left( \frac{1}{p} \Pi_0 \right) + \hat{d}_1 \left( \frac{1}{p} \Sigma \right) \hat{d}_1 \left( 0, \frac{1}{p} \Sigma^2 \Pi_0 \right)}{\left( \hat{d}_2 \left( \frac{1}{p} \Sigma^2 \right) - \hat{d}_1 \left( \frac{1}{p} \Sigma^2 \right) \frac{\hat{h}_3}{\hat{h}_2} \right) \hat{q}_2 \left( \frac{1}{p} \Pi_0^2 \right) + \hat{d}_1^2 \left( 0, \frac{1}{p} \Sigma^2 \Pi_0 \right)},$$

with

$$(48) \quad \hat{d}_1 \left( \frac{1}{p} \Sigma \right) = \frac{1}{\hat{v}(0)} \left( \frac{1}{c_n \hat{v}(0)} - \hat{d}_1 \left( \frac{1}{p} \mathbf{I}_p \right) \right),$$

$$(49) \quad \hat{d}_1 \left( \frac{1}{p} \Sigma^2 \right) = \frac{1}{[\hat{v}(0)]^2} \left( \frac{1}{p} \text{tr}[\mathbf{S}_n] + \hat{d}_1 \left( \frac{1}{p} \mathbf{I}_p \right) - \frac{2}{c_n \hat{v}(0)} \right),$$

$$(50) \quad \hat{d}_1 \left( 0, \frac{1}{p} \Sigma^2 \Pi_0 \right) = \frac{1}{[\hat{v}(0)]^2} \left( \frac{1}{p} \text{tr}[\mathbf{S}_n \Pi_0] + \hat{d}_1 \left( \frac{1}{p} \Pi_0 \right) \right) - \frac{2}{[\hat{v}(0)]^3} \left( \frac{1}{p} \text{tr}[\Pi_0] - \hat{d}_0 \left( 0, \frac{1}{p} \Pi_0 \right) \right),$$

$$(51) \quad \hat{d}_2 \left( \frac{1}{p} \Sigma^2 \right) = \frac{1}{\hat{v}(0)} \hat{d}_1 \left( \frac{1}{p} \Sigma^2 \right) - \frac{1}{[\hat{v}(0)]^2} \left( \hat{d}_1 \left( \frac{1}{p} \Sigma \right) - \hat{d}_2 \left( \frac{1}{p} \mathbf{I}_p \right) \right),$$

where  $\hat{v}(0)$ ,  $\hat{h}_2$ ,  $\hat{h}_3$ ,  $\hat{d}_1 \left( \frac{1}{p} \mathbf{I}_p \right)$ ,  $\hat{d}_1 \left( \frac{1}{p} \Pi_0 \right)$ ,  $\hat{d}_2 \left( \frac{1}{p} \mathbf{I}_p \right)$ ,  $\hat{d}_0 \left( 0, \frac{1}{p} \Pi_0 \right)$ ,  $\hat{q}_1 \left( \frac{1}{p} \Pi_0 \right)$  and  $\hat{q}_2 \left( \frac{1}{p} \Pi_0^2 \right)$  are given in (8) and in (S.38), (S.39), (S.40), (S.41), (S.42), (S.45), (S.47), and (S.48) in the supplement (Bodnar and Parolya (2025)), respectively.

The proof of Theorem 3.2 is given in the supplementary material. If  $\Pi_0 = \mathbf{I}_p$ , then  $\hat{d}_0 \left( 0, \frac{1}{p} \mathbf{I}_p \right) = \frac{c_n - 1}{c_n}$  and the results of Theorem 3.2 can be significantly simplified. Using Theorem 3.2, a bona fide shrinkage estimator for the precision matrix is deduced as

$$(52) \quad \hat{\Pi}_{MP;BF} = \hat{\alpha}_{MP}^* \mathbf{S}_n^+ + \hat{\beta}_{MP}^* \Pi_0.$$

3.1.2. *Results for the ridge-type inverse.* Similarly, a shrinkage estimator for the precision matrix can be constructed by using the ridge-type inverse of the sample covariance matrix. First, we consider the case when the tuning parameter  $t$  is fixed to some preselected value  $t_0$  and then extend the results by minimizing the loss function (42) additionally with respect to the tuning parameter  $t$ . In the former case the first two summands in (42) are considered and the solution is given by

$$\alpha_{R;n}^*(t_0) = \alpha_n^*(\mathbf{S}_n^-(t_0)) \quad \text{and} \quad \beta_{R;n}^*(t_0) = \beta_n^*(\mathbf{S}_n^-(t_0)),$$

while in the latter case  $L_{R;n,2}^2(t) = L_{F;n,2}^2(\mathbf{S}_n^-(t))$  is additionally maximized with respect to the tuning parameter  $t$ , which is then used in the determination of the optimal shrinkage intensities.

In Theorem S.6.1 from Bodnar and Parolya (2025), the deterministic equivalents of  $\alpha_{R;n}^*(t_0)$  and  $\beta_{R;n}^*(t_0)$  in the high-dimensional setting are presented by using Corollary 2.4, while their consistent estimators are given in Theorem 3.3. Moreover, Theorem S.6.2 from the supplementary material (Bodnar and Parolya (2025)) presents the deterministic asymptotic equivalent  $L_{R;2}^2$  to  $L_{R;n,2}^2(t)$  by applying the results of Corollary 2.4, which is consistently estimated in the second statement of Theorem 3.3 (see, the formula (59)).

**THEOREM 3.3.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A2) for any  $t_0 > 0$ , consistent estimators for  $\alpha_{R;n}^*(t_0)$  and  $\beta_{R;n}^*(t_0)$  are given by*

$$(53) \quad \hat{\alpha}_{R;n}^*(t_0) = \frac{\hat{d}_0 \left( t_0, \frac{1}{p} \Sigma \right) \hat{q}_2 \left( \frac{1}{p} \Pi_0^2 \right) - \hat{d}_0 \left( t_0, \frac{1}{p} \Sigma^2 \Pi_0 \right) \hat{q}_1 \left( \frac{1}{p} \Pi_0 \right)}{\left( t_0^{-1} \hat{d}_0 \left( t_0, \frac{1}{p} \Sigma^2 \right) + \hat{v}^{(1)}(t_0) \hat{d}_1 \left( t_0, \frac{1}{p} \Sigma^2 \right) \right) \hat{q}_2 \left( \frac{1}{p} \Pi_0^2 \right) - t_0^{-1} \hat{d}_0^2 \left( t_0, \frac{1}{p} \Sigma^2 \Pi_0 \right)},$$

$$(54) \quad \hat{\beta}_{R;n}^*(t_0) = \frac{\left( t_0^{-1} \hat{d}_0 \left( t_0, \frac{1}{p} \Sigma^2 \right) + \hat{v}^{(1)}(t_0) \hat{d}_1 \left( t_0, \frac{1}{p} \Sigma^2 \right) \right) \hat{q}_1 \left( \frac{1}{p} \Pi_0 \right) - t_0^{-1} \hat{d}_0 \left( t_0, \frac{1}{p} \Sigma \right) \hat{d}_0 \left( t_0, \frac{1}{p} \Sigma^2 \Pi_0 \right)}{\left( t_0^{-1} \hat{d}_0 \left( t_0, \frac{1}{p} \Sigma^2 \right) + \hat{v}^{(1)}(t_0) \hat{d}_1 \left( t_0, \frac{1}{p} \Sigma^2 \right) \right) \hat{q}_2 \left( \frac{1}{p} \Pi_0^2 \right) - t_0^{-1} \hat{d}_0^2 \left( t_0, \frac{1}{p} \Sigma^2 \Pi_0 \right)},$$

with

$$(55) \quad \hat{d}_0 \left( t_0, \frac{1}{p} \boldsymbol{\Sigma} \right) = \frac{1}{c_n \hat{v}(t_0)} - \frac{t_0}{c_n},$$

$$(56) \quad \hat{d}_0 \left( t_0, \frac{1}{p} \boldsymbol{\Sigma}^2 \right) = \frac{1}{\hat{v}(t_0)} \left( \frac{1}{p} \text{tr}[\mathbf{S}_n] - \frac{1}{c_n \hat{v}(t_0)} + \frac{t_0}{c_n} \right),$$

$$(57) \quad \hat{d}_0 \left( t_0, \frac{1}{p} \boldsymbol{\Sigma}^2 \boldsymbol{\Pi}_0 \right) = \frac{1}{\hat{v}(t_0)} \frac{1}{p} \text{tr}[\mathbf{S}_n \boldsymbol{\Pi}_0] - \frac{1}{[\hat{v}(t_0)]^2} \left( \frac{1}{p} \text{tr}[\boldsymbol{\Pi}_0] - \hat{d}_0 \left( t_0, \frac{1}{p} \boldsymbol{\Pi}_0 \right) \right),$$

$$(58) \quad \hat{d}_1 \left( t_0, \frac{1}{p} \boldsymbol{\Sigma}^2 \right) = \frac{1}{[\hat{v}(t_0)]^2} \left( \frac{1}{p} \text{tr}[\mathbf{S}_n] + \hat{d}_1 \left( t_0, \frac{1}{p} \mathbf{I}_p \right) - \frac{2}{c_n \hat{v}(t_0)} + \frac{2t_0}{c_n} \right),$$

where  $\hat{v}(t_0)$ ,  $\hat{v}^{(1)}(t_0)$ ,  $\hat{d}_0 \left( t_0, \frac{1}{p} \boldsymbol{\Pi}_0 \right)$ ,  $\hat{d}_1 \left( t_0, \frac{1}{p} \mathbf{I}_p \right)$ ,  $\hat{q}_1 \left( \frac{1}{p} \boldsymbol{\Pi}_0 \right)$  and  $\hat{q}_2 \left( \frac{1}{p} \boldsymbol{\Pi}_0^2 \right)$  are given in (7) and in (S.45), (S.46), (S.47), and (S.48) from the supplement (Bodnar and Parolya (2025)), respectively.

Moreover, for any  $t > 0$ , it holds that

$$(59) \quad \frac{1}{p} \left| \hat{L}_{R;2}^2(t) - L_{R;2}^2(t) \right| \xrightarrow{a.s.} 0, \text{ for } p/n \rightarrow c \in (1, \infty) \text{ as } n \rightarrow \infty \text{ with}$$

$$\hat{L}_{R;2}^2(t) = \frac{1}{\hat{q}_2 \left( \frac{1}{p} \boldsymbol{\Pi}_0^2 \right)} \frac{\left[ \hat{d}_0 \left( t, \frac{1}{p} \boldsymbol{\Sigma} \right) \hat{q}_2 \left( \frac{1}{p} \boldsymbol{\Pi}_0^2 \right) - \hat{d}_0 \left( t, \frac{1}{p} \boldsymbol{\Sigma}^2 \boldsymbol{\Pi}_0 \right) \hat{q}_1 \left( \frac{1}{p} \boldsymbol{\Pi}_0 \right) \right]^2}{\left( \hat{d}_0 \left( t, \frac{1}{p} \boldsymbol{\Sigma}^2 \right) + t \hat{v}^{(1)}(t) \hat{d}_1 \left( t, \frac{1}{p} \boldsymbol{\Sigma}^2 \right) \right) \hat{q}_2 \left( \frac{1}{p} \boldsymbol{\Pi}_0^2 \right) - \hat{d}_0^2 \left( t, \frac{1}{p} \boldsymbol{\Sigma}^2 \boldsymbol{\Pi}_0 \right)},$$

where  $\hat{v}^{(1)}(t)$ ,  $\hat{q}_1 \left( \frac{1}{p} \boldsymbol{\Pi}_0 \right)$ ,  $\hat{q}_2 \left( \frac{1}{p} \boldsymbol{\Pi}_0^2 \right)$ ,  $\hat{d}_0 \left( t, \frac{1}{p} \boldsymbol{\Sigma} \right)$ ,  $\hat{d}_0 \left( t, \frac{1}{p} \boldsymbol{\Sigma}^2 \right)$ ,  $\hat{d}_0 \left( t, \frac{1}{p} \boldsymbol{\Sigma}^2 \boldsymbol{\Pi}_0 \right)$  and  $\hat{d}_1 \left( t, \frac{1}{p} \boldsymbol{\Sigma}^2 \right)$  are given in (7), (55), (56), (57), and (58), and in (S.47) and (S.48) from Bodnar and Parolya (2025).

The proofs of both statements of Theorem 3.3 follow from the theoretical results derived in Section 2.2 and the consistent estimators are presented in Section S.4 of the supplement (Bodnar and Parolya (2025)). The bona fide shrinkage estimator for the precision matrix under the application of the ridge-type inverse is given by

$$(60) \quad \hat{\boldsymbol{\Pi}}_{R;BF}(t_0) = \hat{\alpha}_R^*(t_0) \mathbf{S}_n^-(t_0) + \hat{\beta}_R^*(t_0) \boldsymbol{\Pi}_0 \quad \text{for some } t_0 > 0.$$

It is noted that, here we have additionally to optimize the function (42) for the tuning parameter  $t$ , or, equivalently due to Theorem 3.1, maximize the function  $L_{F;n,2}^2(\mathbf{S}_n^-(t))$  from (45). The results of the second statement of Theorem 3.3 are very important from the practical viewpoint. In particular, they allow us to find the optimal value of the tuning parameter  $t$  by maximizing  $\hat{L}_{R;2}^2(t)$ , which can be easily computed, instead of  $L_{R;n,2}^2(t)$  that depends on the unknown population covariance matrix. Both results of Theorem 3.3 ensure that the tuning parameter found by maximizing  $\hat{L}_{R;2}^2(t)$  will be closed when the original loss function is optimized. Furthermore, since  $\hat{L}_{R;2}^2(t)$  is available in the closed form, no cross-validation is needed which by its definition reduces the sample size.

Let  $t^* = \arg\max_{t>0} \hat{L}_{R;2}^2(t)$ . Then, the bona-fide optimal shrinkage estimator of the precision matrix with the ridge-type estimator is expressed as

$$(61) \quad \hat{\boldsymbol{\Pi}}_{R;BF}(t^*) = \hat{\alpha}_R^*(t^*) \mathbf{S}_n^-(t^*) + \hat{\beta}_R^*(t^*) \boldsymbol{\Pi}_0,$$

where  $\hat{\alpha}_R^*(t^*)$  and  $\hat{\beta}_R^*(t^*)$  are given in (53) and (54), respectively. Finally, we note that for practical purposes one can consider to replace  $t$  by  $u$  with  $t = \tan(u)$  and to maximize  $\hat{L}_{R;2}^2(\tan(u))$  over the finite interval  $(0, \pi/2)$ .

3.1.3. *Results for the Moore-Penrose-ridge inverse.* When the shrinkage estimator of the precision matrix is based on the Moore-Penrose-ridge inverse, we proceed in the same way as in the case of the ridge inverse. First, we investigate the limiting behavior of the optimal shrinkage intensities

$$\alpha_{MPR;n}^*(t_0) = \alpha_n^*(\mathbf{S}_n^\pm(t_0)) \quad \text{and} \quad \beta_{MPR;n}^*(t_0) = \beta_n^*(\mathbf{S}_n^\pm(t_0)),$$

for a fixed value of the tuning parameter  $t_0$  and then, present the procedure for how the optimal value of the tuning parameter should be chosen. Since this procedure mimics the one derived for the ridge inverse the results are moved to Section S.7 in [Bodnar and Parolya \(2025\)](#) (see Theorem S.7.1, Theorem S.7.2, Theorem S.7.3, and Theorem S.7.4 for details).

3.2. *Shrinkage estimator for minimum variance portfolio.* Following [Bodnar et al. \(2018\)](#), the general linear shrinkage estimator for the global minimum variance (GMV) portfolio is constructed by

$$(62) \quad \widehat{\mathbf{w}}_{GSE} = \alpha_n \mathbf{w}_{\mathbf{S}_n^\#(t)} + (1 - \alpha_n) \mathbf{b},$$

where  $\mathbf{b}$  is a target portfolio, i.e.,  $\mathbf{b}^\top \mathbf{1} = 1$ , which typically is chosen to be equal to the naive portfolio  $\frac{1}{p}$  when no *a priori* information about the true weights of the GMV portfolio

$$(63) \quad \mathbf{w}_{GMV} = \frac{\boldsymbol{\Sigma}^{-1} \mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}}$$

is available. Let  $\mathbf{S}_n^\#(t)$  denote an estimator of the precision matrix. Then, the traditional (sample) estimator of  $\mathbf{w}_{GMV}$  is given by

$$(64) \quad \mathbf{w}_{\mathbf{S}_n^\#(t)} = \frac{\mathbf{S}_n^\#(t) \mathbf{1}}{\mathbf{1}^\top \mathbf{S}_n^\#(t) \mathbf{1}}$$

is obtained by replacing  $\boldsymbol{\Sigma}^{-1}$  with  $\mathbf{S}_n^\#(t)$  in (63).

Due to [Bodnar et al. \(2018\)](#) and [Frahm and Memmel \(2010\)](#), the shrinkage estimator of the form (62) dominates the sample estimator  $\mathbf{w}_{\mathbf{S}_n^\#(t)}$  for  $c < 1$  and  $\mathbf{S}_n^\#(t) = \mathbf{S}_n^{-1}$ . Moreover, the estimator proposed by [Bodnar et al. \(2018\)](#) is proved to be asymptotically dominant over [Frahm and Memmel \(2010\)](#) in case  $c < 1$  and becomes superior if  $c$  gets closer to one (see [Bodnar et al. \(2023b\)](#)). Nevertheless, only the case  $c < 1$  was efficiently covered in the aforementioned works. In case  $c > 1$ , [Bodnar et al. \(2018\)](#) provided the approximate formulae for the shrinkage intensities in the case of Moore-Penrose inverse. However, those formulae appear to be accurate only if  $c < 2$  and close to 1, otherwise they could be far away from the optimal ones. The theoretical results of Section 2 provide us tools to deal with this problem efficiently.

The aim is to estimate the shrinkage intensity  $\alpha_n$  and the ridge parameter  $t$  from the following normalized optimization problem (see, e.g., [Bodnar et al. \(2024b\)](#))

$$(65) \quad \min_{\alpha_n, t} \frac{\widehat{\mathbf{w}}_{GSE}^\top \boldsymbol{\Sigma} \widehat{\mathbf{w}}_{GSE}}{\mathbf{b}^\top \boldsymbol{\Sigma} \mathbf{b}}.$$

In the numerator, we have the usual performance measure, the so-called out-of-sample variance defined by  $\widehat{\mathbf{w}}_{GSE}^\top \boldsymbol{\Sigma} \widehat{\mathbf{w}}_{GSE}$  for the portfolio with weights  $\widehat{\mathbf{w}}_{GSE}$ , and the denominator  $\mathbf{b}^\top \boldsymbol{\Sigma} \mathbf{b}$  is merely for technical reasons, i.e., it is a normalization, which makes the loss bounded for increasing portfolio dimension  $p$ . The following theorem gives the optimal values of  $\alpha_n$  and  $t$ .

**THEOREM 3.4.** *For given  $t$ , the out-of-sample variance (65) is minimized with respect to  $\alpha$  at*

$$(66) \quad \alpha_n^*(t) = \frac{\mathbf{b}^\top \boldsymbol{\Sigma} \left( \mathbf{b} - \mathbf{w}_{\mathbf{S}_n^\#(t)} \right)}{\left( \mathbf{b} - \mathbf{w}_{\mathbf{S}_n^\#(t)} \right)^\top \boldsymbol{\Sigma} \left( \mathbf{b} - \mathbf{w}_{\mathbf{S}_n^\#(t)} \right)},$$

while the optimal  $t$  is found by maximizing

$$(67) \quad L_{n;2}(t) = \frac{1}{\mathbf{b}^\top \boldsymbol{\Sigma} \mathbf{b}} \frac{\left( \mathbf{b}^\top \boldsymbol{\Sigma} \left( \mathbf{b} - \mathbf{w}_{\mathbf{S}_n^\#(t)} \right) \right)^2}{\left( \mathbf{b} - \mathbf{w}_{\mathbf{S}_n^\#(t)} \right)^\top \boldsymbol{\Sigma} \left( \mathbf{b} - \mathbf{w}_{\mathbf{S}_n^\#(t)} \right)}.$$

The proof of Theorem 3.4 is given in the supplementary material. Similarly as before, to find the value  $t^*$  together with  $\alpha_n^*(t)$ , which minimizes the loss function, we proceed in three steps. First, we find the deterministic equivalent to  $L_{n;2}(t)$ , and estimate it consistently in the second step. Finally, we minimize the obtained consistent estimator in the last step. This procedure has to be done for the generalized inverses  $\mathbf{S}_n^\#(t) \in \{\mathbf{S}_n^-(t), \mathbf{S}_n^\pm(t)\}$  except  $\mathbf{S}_n^\#(t) = \mathbf{S}_n^+$  because the latter one is independent of the tuning parameter  $t$ . Thus, it is enough to find the asymptotic equivalent<sup>2</sup>, i.e.,  $\alpha^*$ , for the optimal shrinkage intensity  $\alpha_n^*(t) \equiv \alpha_n^*$  and to estimate it consistently. The results regarding the asymptotic equivalent  $\alpha^*$  in case of the Moore-Penrose inverse  $\mathbf{S}_n^+$  are summarized in Theorem S.8.1 in Bodnar and Parolya (2025), while Theorem 3.5 provides the bona-fide estimator, i.e., a consistent estimator of  $\alpha^*$ .

**THEOREM 3.5.** *Let  $\mathbf{Y}_n$  possess the stochastic representation as in (1). Then, under the assumption (A1) and (A2), a consistent estimator of  $\alpha^*$  is given by*

$$(68) \quad \hat{\alpha}^* = \frac{p \mathbf{b}^\top \mathbf{S}_n \mathbf{b} - \frac{\hat{d}_1(\mathbf{1b}^\top \boldsymbol{\Sigma})}{\hat{d}_1\left(\frac{\mathbf{11}^\top}{p}\right)}}{p \mathbf{b}^\top \mathbf{S}_n \mathbf{b} - 2 \frac{\hat{d}_1(\mathbf{1b}^\top \boldsymbol{\Sigma})}{\hat{d}_1\left(\frac{\mathbf{11}^\top}{p}\right)} + \frac{\hat{d}_3\left(\frac{\mathbf{11}^\top}{p}\right)}{\hat{d}_1^2\left(\frac{\mathbf{11}^\top}{p}\right)}},$$

where

$$(69) \quad \hat{d}_1(\mathbf{1b}^\top \boldsymbol{\Sigma}) = \frac{1}{\hat{v}(0)} \left[ \frac{1}{\hat{v}(0)} \left( 1 - \hat{d}_0(0, \mathbf{1b}^\top) \right) - \hat{d}_1(\mathbf{1b}^\top) \right]$$

with  $\hat{v}(0)$ ,  $\hat{d}_1\left(\frac{\mathbf{11}^\top}{p}\right)$ ,  $\hat{d}_1(\mathbf{1b}^\top)$ ,  $\hat{d}_3\left(\frac{\mathbf{11}^\top}{p}\right)$ , and  $\hat{d}_0(0, \mathbf{1b}^\top)$  given in (8) and in (S.41), (S.44), and (S.45) from the supplement (Bodnar and Parolya (2025)), respectively.

In the same way as in Section 3.1, the shrinkage estimators for the GMV portfolio weights can be constructed by using the ridge-type inverse and the Moore-Penrose-ridge inverse. As mentioned above, one should consider first the case when the tuning parameter  $t$  is fixed to some preselected value  $t_0$  and then extend the results by minimizing the loss function (67) additionally with respect to the tuning parameter  $t$ . The case of the ridge inverse has already been covered recently in Bodnar et al. (2024b). The case of Moore-Penrose-ridge inverse can be done similarly, however with very tedious computations.

<sup>2</sup>Note that the asymptotic shrinkage intensity  $\alpha^*$  clearly depends on  $n$ , due to the dependence of both  $\mathbf{b}$  and  $\boldsymbol{\Sigma}$  on  $p = p(n)$ . However, this dependence was intentionally omitted in the notation to avoid overwhelming the reader and to keep the presentation of the results as clear and simple as possible.

**4. Finite sample performance.** In this section, we study the finite-sample performance of the considered estimators via simulations. Two stochastic models will be considered. In the first scenario, we assume that the elements of the matrix  $\mathbf{X}_n$  are standard normally distributed, while they are assumed to be scaled  $t$ -distributed with 5 degrees of freedom in the second scenario. The scale factor in the case of the  $t$ -distribution is set to  $\sqrt{3/5}$  to ensure that the variances of the elements of  $\mathbf{X}_n$  are all one. Finally, the mean vector  $\boldsymbol{\mu}$  is taken to be zero. The eigenvectors of the population covariance matrix are drawn from the Haar distribution (see, e.g., [Muirhead \(1982\)](#)), while its eigenvalues are chosen in the following way: 20% of eigenvalues are equal to one, 40% of eigenvalues equal to three, and the rest equal to ten. Finally, we set  $n \in \{100, 250\}$  and  $c \in (1, 5]$ . In the case of the ridge estimator and the Moore-Penrose-ridge estimator, we consider  $t \in [0.1, 10]$ . Within the simulation study, we compare the performance of the introduced shrinkage estimators with the existing benchmark approaches. The figures with the computation times are provided in the supplement.

4.1. *Results for the estimation of precision matrix.* In this section, we compare the derived shrinkage approaches for the estimation of the precision matrix (see Section 3.1) with the Moore-Penrose inverse and several benchmark methods, which are used in the statistical literature. These approaches are the **empirical Bayes** estimator of [Kubokawa and Srivastava \(2008\)](#), the **optimal ridge** estimator of [Wang et al. \(2015\)](#) and the **inverse nonlinear (NL) shrinkage** estimator of [Ledoit and Wolf \(2020\)](#). More details about the considered benchmark approaches can be found in Section S.9.1 of the supplement ([Bodnar and Parolya \(2025\)](#)). Besides that, we consider the **oracle nonlinear shrinkage** estimator, which is also introduced in Section S.9.1 of the supplement. Although the latter estimator cannot be computed in practice since it depends on the unobservable population covariance matrix, it will be used in the simulation study as the global benchmark strategy. The target matrix is set to the identity matrix, i.e.,  $\boldsymbol{\Pi}_0 = \mathbf{I}_p$ . In fact, the linear shrinkage estimator (41) becomes purely a function of the eigenvalues of the sample covariance  $\mathbf{S}_n$  for the considered target matrix and, thus, the oracle nonlinear shrinkage estimator is the best one can do in that specific situation.

Next, we compare the three shrinkage approaches suggested in Section 3.1 with the considered benchmark methods. As a performance measure, we use the percentage relative improvement in average loss (PRIAL) defined by

$$(70) \quad \text{PRIAL}(\widehat{\boldsymbol{\Pi}}) = \left( 1 - \frac{\mathbb{E} \|\widehat{\boldsymbol{\Pi}}\boldsymbol{\Sigma} - \mathbf{I}_p\|_F^2}{\mathbb{E} \|\mathbf{S}_n^+ \boldsymbol{\Sigma} - \mathbf{I}_p\|_F^2} \right) \cdot 100\%,$$

where  $\widehat{\boldsymbol{\Pi}}$  is an estimator of the precision matrix  $\boldsymbol{\Sigma}^{-1}$ . By definition, the PRIAL provides the percentage improvement of each strategy in comparison to the one based on the Moore-Penrose inverse. Larger values of the PRIAL indicate better performance. Moreover, it holds that  $\text{PRIAL}(\boldsymbol{\Sigma}^{-1}) = 100\%$ .

Figure 3 depicts the values of the PRIAL, computed for the introduced three shrinkage approaches and for the two benchmarks: the optimal ridge estimator and the oracle nonlinear shrinkage estimator. The results of the empirical Bayes estimator and of the inverse nonlinear shrinkage estimator are considerably worse and are available in Section S.9.2 of the supplement (see Figure S.1 in [Bodnar and Parolya \(2025\)](#)).

In Figure 3, we observe that the ridge shrinkage and the optimal ridge estimators outperform the other competitors with the ridge shrinkage estimator having a slightly better performance. This observation holds for the two considered models of the data-generating process and for the selected values of  $n$  and  $c_n$ . The Moore-Penrose shrinkage and the Moore-Penrose ridge shrinkage estimators are ranked in the third and fourth places with the Moore-Penrose ridge shrinkage estimator performing better in almost all cases with the exception of large

values of  $c_n$  when  $n = 250$ . These two estimators are followed by the empirical Bayes ridge which performs better than the inverse nonlinear shrinkage estimator (see Figure S.1 in [Bodnar and Parolya \(2025\)](#)). However, it has to be noted that the considered inverse nonlinear shrinkage estimator was derived under the other loss function, which explains the considerable differences between the oracle and bona fide nonlinear shrinkage estimators. Finally, it is remarkable that both the ridge shrinkage and the optimal ridge attain the global benchmark, which is the oracle nonlinear shrinkage estimator and, as such, they can be considered optimal for the chosen data-generating models. Although the Moore-Penrose shrinkage estimator performs slightly worse than the shrinkage approaches based on the ridge estimators, it is present in closed-form and requires no numerical approximation in practice. This fact is illustrated in Figure S.2 of [Bodnar and Parolya \(2025\)](#), which compares the computational time of the estimators. Among them, the MP shrinkage estimator is the fastest, second only to the Moore-Penrose inverse itself.

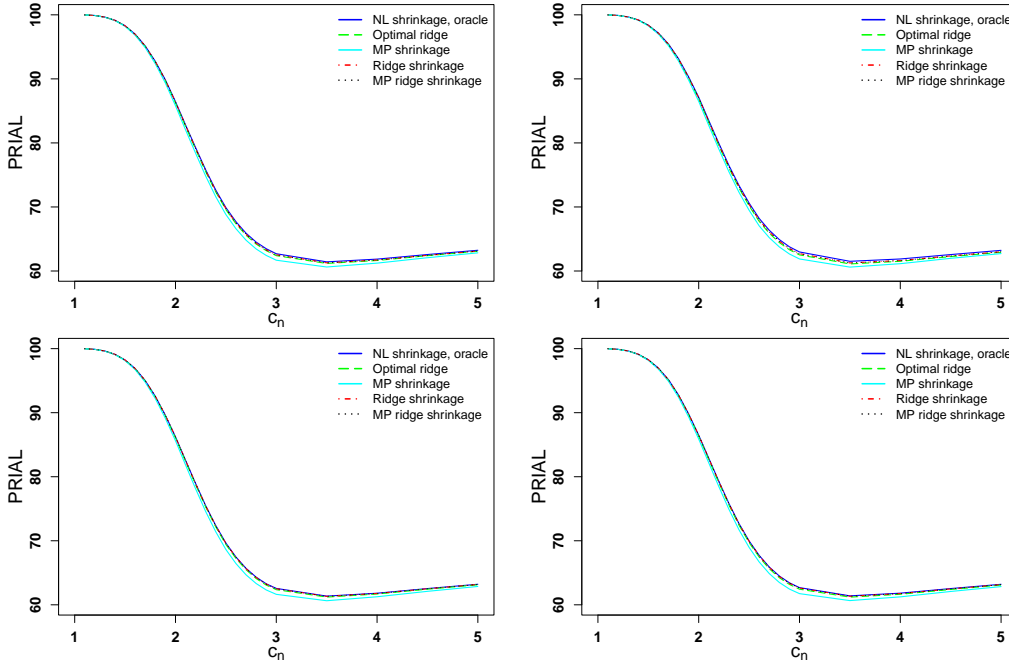


FIG 3. *PRIAL* for  $c_n \in (1, 5]$ ,  $n = 100$  (first row) and  $n = 250$  (second row) when the elements of  $\mathbf{X}_n$  are drawn from the normal distribution (first column) and scale  $t$ -distribution (second column).

4.2. *Results for the estimation of the global minimum variance portfolio.* In this section, we compare the derived shrinkage approaches for the GMV portfolio for estimating the GMV portfolio (see Section 3.2) with the **traditional** sample estimator based on the Moore-Penrose inverse, the **reflexive inverse estimator** introduced in [Bodnar et al. \(2018\)](#), and the **double shrinkage estimator** or **ridge shrinkage estimator** recently derived in [Bodnar et al. \(2024b\)](#). All benchmarks are described in Section S.9.3 of the supplement ([Bodnar and Parolya \(2025\)](#)). In the simulation study, the target portfolio  $\mathbf{b}$  is set to the equally weighted portfolio.

As a performance measure, we use the relative out-of-sample variance (rOSV) defined as

$$(71) \quad \text{rOSV}(\hat{\mathbf{w}}) = \frac{\hat{\mathbf{w}}^\top \hat{\Sigma} \hat{\mathbf{w}}}{V_{GMV}} - 1,$$

where  $\hat{\mathbf{w}}$  is any estimator of the GMV portfolio, and  $V_{GMV}$  is the variance of the true (population) GMV portfolio given by  $V_{GMV} = \frac{1}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}}$ . By definition, rOSV is always greater than zero and equals zero only if  $\hat{\mathbf{w}} = \mathbf{w}_{GMV}$  given by (63). Therefore, a smaller rOSV indicates better performance of the portfolio estimator.

Figure 4 presents the rOSV values computed for the proposed Moore-Penrose portfolio shrinkage approach and the two benchmarks: the double shrinkage estimator and the reflexive inverse estimator. Results for the traditional estimator, which are available in Section S.9.4 (see Figure S.4 in Bodnar and Parolya (2025)), are omitted here due to its poor performance.

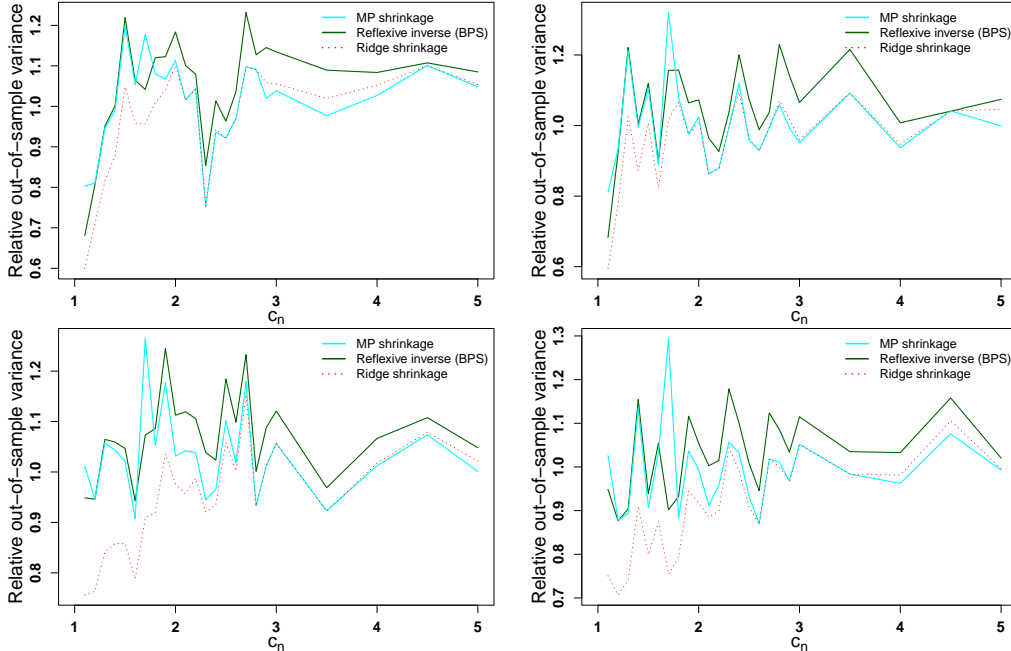


FIG 4. rOSV for  $c_n \in (1, 5]$ ,  $n = 100$  (first row) and  $n = 250$  (second row) when the elements of  $\mathbf{X}_n$  are drawn from the normal distribution (first column) and scale  $t$ -distribution (second column).

From Figure 4, we observe that all approaches exhibit similar behavior, with the proposed MP shrinkage and double (ridge) shrinkage performing slightly better as the ratio  $c = p/n$  increases. The reflexive inverse estimator, as shown in Bodnar et al. (2018), serves as a good approximation to the Moore-Penrose inverse only when  $c$  is less than two, a result that is supported here. Additionally, the recently proposed double shrinkage approach by Bodnar et al. (2024b) outperforms other strategies for smaller values of the concentration ratio  $c$ .

It is worth noting, however, that the proposed MP shrinkage approach avoids the need for numerical optimization of the penalty parameter, unlike the double shrinkage strategy, which requires a quite involved nonlinear optimization routine. The computational advantages of the MP shrinkage estimator are depicted in Figure S.5 of Bodnar and Parolya (2025).

**Acknowledgement.** The authors would like to thank Professor Enno Mammen, Professor Hans-Georg Müller, the Associate Editor, and the two anonymous Reviewers for their constructive comments that improved the quality of this paper. We gratefully acknowledge the comments from the participants at the conference 'International Conference on Matrix Analysis and its Applications' (Bendlewo) 2023, 'Computational and Methodological Statistics' (CMStatistics) 2024 (Berlin), 'Stochastic models, Statistics and their Applications' (SMSA)

2024 (Delft), the 11th Tartu Conference on Multivariate Statistics (Tartu), the 11th World Congress in Probability and Statistics (Bochum) and the 2nd Joint Conference on Statistics and Data Science in China. The authors are also thankful to Prof. Holger Dette, Prof. Raymond Kan, Prof. Olivier Ledoit, Prof. Mark Podolskij, Dr. habil. Maryna Prus, Prof. Dietrich von Rosen, and Prof. Jianfeng Yao for fruitful discussions, and Dr. Alexis Derumigny for creating and developing the R package `UniversalShrink`, used extensively in this work. Taras Bodnar was partly supported by the Swedish Research Council (VR) via the grant 2025-04704 ”*Reviving generalized inverses: High-dimensional statistical analysis with insufficient data*”.

**SUPPLEMENT TO “REVIVING PSEUDO-INVERSES:  
ASYMPTOTIC PROPERTIES OF LARGE DIMENSIONAL MOORE-PENROSE  
AND RIDGE-TYPE INVERSES WITH APPLICATIONS”**

BY TARAS BODNAR<sup>1,a,a</sup> AND NESTOR PAROLYA<sup>2,b,b</sup>

<sup>1</sup>*Department of Management and Engineering, Linköping University, SE-581 83 Linköping, Sweden, <sup>a</sup>[taras.bodnar@liu.se](mailto:taras.bodnar@liu.se)*

<sup>2</sup>*Department of Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands, <sup>b</sup>[n.parolya@tudelft.nl](mailto:n.parolya@tudelft.nl)*

CONTENTS

1	Introduction . . . . .	1
2	Asymptotic properties of pseudo-inverse matrices . . . . .	3
2.1	Weighted moments of the sample Moore-Penrose inverse . . . . .	6
2.2	Weighted moments of the sample ridge-type inverse . . . . .	9
2.3	Weighted moments of the sample Moore-Penrose-ridge inverse . . . . .	12
3	Applications . . . . .	15
3.1	Shrinkage estimator of the high-dimensional precision matrix . . . . .	15
3.1.1	Results for the Moore-Penrose inverse . . . . .	16
3.1.2	Results for the ridge-type inverse . . . . .	17
3.1.3	Results for the Moore-Penrose-ridge inverse . . . . .	19
3.2	Shrinkage estimator for minimum variance portfolio . . . . .	19
4	Finite sample performance . . . . .	21
4.1	Results for the estimation of precision matrix . . . . .	21
4.2	Results for the estimation of the global minimum variance portfolio . . . . .	22
	Acknowledgement . . . . .	23
S.5	Proofs of the main theorems and corollaries . . . . .	25
S.5.1	Proof of Proposition 2.1 . . . . .	26
S.5.2	Proof of Theorem 2.1 . . . . .	26
S.5.3	Proof of Corollary 2.1 . . . . .	30
S.5.4	Proof of Corollary 2.2 . . . . .	30
S.5.5	Proof of Corollary 2.3 . . . . .	32
S.5.6	Proof of Theorem 2.2 . . . . .	32
S.5.7	Proof of 2.4 . . . . .	33
S.5.8	Proof of Corollary 2.5 . . . . .	34
S.5.9	Proof of Corollary 2.6 . . . . .	35
S.5.10	Proof of Corollary 2.7 . . . . .	35
S.5.11	Proof of Corollary 2.9 . . . . .	36

S.5.12 Proof of Theorem 3.1 . . . . . 36

S.5.13 Proof of Theorem 3.4 . . . . . 37

S.6 Weighted moments of the ordinary inverse of the sample covariance matrix . . . . . 38

S.7 Weighted moments of the sample covariance matrix . . . . . 45

S.8 Consistent estimators for  $h_m$  and  $d_m(\cdot, \cdot)$  . . . . . 48

S.9 Shrinkage estimation of the precision matrix with the Moore-Penrose inverse . . . . . 49

    S.9.1 Oracle estimator . . . . . 49

    S.9.2 Bona-fide estimator: Proof of Theorem 3.2 . . . . . 49

S.10 Shrinkage estimation of the precision matrix with the ridge inverse . . . . . 50

    S.10.1 Oracle estimators . . . . . 50

    S.10.2 Bona-fide estimators: Proof of Theorem 3.3 . . . . . 50

S.11 Shrinkage estimation of the precision matrix with the Moore-Penrose-ridge inverse . . . . . 51

    S.11.1 Oracle estimators . . . . . 51

    S.11.2 Bona-fide estimators . . . . . 52

S.12 Oracle estimator of the GMV portfolio weights with the Moore-Penrose inverse . . . . . 54

S.13 Additional results of numerical studies . . . . . 56

    S.13.1 Benchmarks: Precision matrix . . . . . 56

    S.13.2 Additional figures: Precision matrix . . . . . 57

    S.13.3 Benchmarks: Global minimum variance portfolio . . . . . 57

    S.13.4 Additional figures: Global minimum variance portfolio . . . . . 60

References . . . . . 60

**S.5. Proofs of the main theorems and corollaries.** Throughout the supplement, we use the notation  $A_n \xrightarrow{d.a.s.} B_n$  to denote the convergence in difference almost surely, that is,  $A_n - B_n \xrightarrow{a.s.} 0$ . All limits are for  $n \rightarrow \infty$  unless explicitly stated otherwise. An abstract *sketch of the proof*, which applies to all theorems proved in Section 2 of the paper, can be summarized in the following way:

**Step 1: Centering.** In the first step, we show that the asymptotic behaviour of the weighted trace moments of the Moore-Penrose inverse (and other inverses) with centering by the sample mean is the same as without centering. This is achieved by rewriting the sample covariance matrix in terms of a 'new' observation matrix  $\tilde{\mathbf{Y}}_n = \mathbf{Y}_n \mathbf{H}_n$ , where  $\mathbf{H}_n$  is a specific orthogonal matrix of rank  $n - 1$ .

**Step 2: Finding a nonconventional moment generating function.** In the second step, we use the well-known Woodbury matrix identity (matrix inversion lemma, see, e.g., [Horn and Johnson \(1985\)](#)) to derive a nonconventional moment generating function for the weighted traces of the considered generalized inverses.

**Step 3: Rewriting the nonconventional moment generating function.** Since the dependence on the true covariance matrix  $\Sigma$  is complex, we rewrite the moment generating function in a form that facilitates its high-dimensional analysis. This is done by algebraic manipulation of the weighted traces and by 'pushing' the dependence on  $\Sigma$  into the weighting matrix  $\Theta$ , thereby reducing the problem to a classical setup with an isotropic case, i.e.,  $\Sigma = \mathbf{I}_p$ .

**Step 4: Generalized Marchenko-Pastur theorem.** With the nonconventional moment generating function in a form where the dependence on  $\Sigma$  is absorbed by the weighting matrix, we apply the generalized Marchenko-Pastur result (see [Rubio and Mestre \(2011\)](#) and Lemma 6 in [Bodnar et al. \(2024b\)](#)) to obtain the asymptotic behaviour of the nonconventional moment generating function for  $p/n \rightarrow c > 1$  as  $p, n \rightarrow \infty$ . A similar approach is used for the case  $c < 1$ , although the function must be represented differently to avoid singularity at zero.

**Step 5: Faá di Bruno formula.** The partial derivatives of the limiting nonconventional moment generating function and its uniform convergence yield all the moments. This step involves applying the Faá di Bruno formula (see Section 1.2 in [Krantz and Parks \(2002\)](#)), which introduces the partial exponential Bell polynomials from (4).

**Step 6: Playing with Bell polynomials.** As the general formula for the asymptotic moments is complicated, we exploit properties of the partial exponential Bell polynomials to derive the final closed-form expressions. These polynomials are essential to obtain explicit results.

### S.5.1. Proof of Proposition 2.1.

PROOF. Let  $t_1 < t_2$ . The application of (5) yields

$$(S.1) \quad \frac{1}{p} \operatorname{tr} \left[ (v(t_1)\boldsymbol{\Sigma} + \mathbf{I}_p)^{-1} \right] - \frac{1}{p} \operatorname{tr} \left[ (v(t_2)\boldsymbol{\Sigma} + \mathbf{I}_p)^{-1} \right] = \frac{t_1 v(t_1) - t_2 v(t_2)}{c_n}.$$

Assume that the statement of the theorem is wrong for the considered  $t_1$  and  $t_2$ , i.e.,  $v(t_1) \leq v(t_2)$ . Then, the left hand-side of (S.1) is nonnegative. On the other side, the right hand-side of (S.1) is nonnegative only if

$$\frac{t_1}{t_2} \geq \frac{v(t_2)}{v(t_1)},$$

which contradicts that  $v(t_1) \leq v(t_2)$ . The theorem is proved.  $\square$

### S.5.2. Proof of Theorem 2.1.

PROOF. Let  $\mathbf{J}_n = \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top$ . Then,  $\mathbf{J}_n$  is a projection matrix of rank  $n - 1$ . As such, all  $n - 1$  nonzero eigenvalues of  $\mathbf{J}_n$  are equal to one, and its singular value decomposition is expressed as

$$\mathbf{J}_n = \mathbf{H}_n \mathbf{H}_n^\top,$$

where  $\mathbf{H}_n$  is a  $n \times (n - 1)$  orthogonal matrix, i.e.,  $\mathbf{H}_n^\top \mathbf{H}_n = \mathbf{I}_{n-1}$ . Then,

$$(S.2) \quad \mathbf{S}_n = \frac{1}{n} \mathbf{Y}_n \mathbf{J}_n \mathbf{Y}_n^\top = \frac{1}{n} \tilde{\mathbf{Y}}_n \tilde{\mathbf{Y}}_n^\top \quad \text{with} \quad \tilde{\mathbf{Y}}_n = \mathbf{Y}_n \mathbf{H}_n.$$

Moreover, since  $\mathbf{H}_n$  is orthogonal and the rank of  $\mathbf{Y}_n$  is equal to  $n$ , the application of Sylvester's rank inequality leads to the conclusion that the rank of  $\tilde{\mathbf{Y}}_n$  is  $n - 1$  (see Section 4.3.3 in [Lütkepohl \(1996\)](#)). Hence, we get

$$(S.3) \quad \mathbf{S}_n^+ = \left( \frac{1}{n} \tilde{\mathbf{Y}}_n \tilde{\mathbf{Y}}_n^\top \right)^+ = \frac{1}{\sqrt{n}} \tilde{\mathbf{Y}}_n \left( \frac{1}{n} \tilde{\mathbf{Y}}_n^\top \tilde{\mathbf{Y}}_n \right)^{-2} \frac{1}{\sqrt{n}} \tilde{\mathbf{Y}}_n^\top$$

and, similarly,

$$(S.4) \quad (\mathbf{S}_n^+)^m = \frac{1}{\sqrt{n}} \tilde{\mathbf{Y}}_n \left( \frac{1}{n} \tilde{\mathbf{Y}}_n^\top \tilde{\mathbf{Y}}_n \right)^{-(m+1)} \frac{1}{\sqrt{n}} \tilde{\mathbf{Y}}_n^\top \quad \text{for} \quad m = 2, 3, \dots$$

It holds that

$$\begin{aligned} & \operatorname{tr} \left[ \frac{1}{\sqrt{n}} \tilde{\mathbf{Y}}_n \left( \frac{1}{n} \tilde{\mathbf{Y}}_n^\top \tilde{\mathbf{Y}}_n \right)^{-(m+1)} \frac{1}{\sqrt{n}} \tilde{\mathbf{Y}}_n^\top \boldsymbol{\Theta} \right] \\ &= \frac{(-1)^m}{m!} \frac{\partial^m}{\partial t^m} \operatorname{tr} \left[ \frac{1}{\sqrt{n}} \tilde{\mathbf{Y}}_n \left( \frac{1}{n} \tilde{\mathbf{Y}}_n^\top \tilde{\mathbf{Y}}_n + t \mathbf{I}_n \right)^{-1} \frac{1}{\sqrt{n}} \tilde{\mathbf{Y}}_n^\top \boldsymbol{\Theta} \right] \Bigg|_{t=0} \end{aligned}$$

for  $m = 1, 2, \dots$

The application of the Woodbury formula (matrix inversion lemma, see, e.g., [Horn and Johnson \(1985\)](#)), leads to

$$\begin{aligned} & \frac{1}{\sqrt{n}} \tilde{\mathbf{Y}}_n \left( \frac{1}{n} \tilde{\mathbf{Y}}_n^\top \tilde{\mathbf{Y}}_n + t \mathbf{I}_n \right)^{-1} \frac{1}{\sqrt{n}} \tilde{\mathbf{Y}}_n^\top = \mathbf{I}_p - t \left( \frac{1}{n} \tilde{\mathbf{Y}}_n \tilde{\mathbf{Y}}_n^\top + t \mathbf{I}_p \right)^{-1} \\ & = \mathbf{I}_p - t \boldsymbol{\Sigma}^{-1/2} \left( \frac{1}{n} \mathbf{X}_n \mathbf{X}_n^\top + t \boldsymbol{\Sigma}^{-1} - \bar{\mathbf{x}}_n \bar{\mathbf{x}}_n^\top \right)^{-1} \boldsymbol{\Sigma}^{-1/2}, \end{aligned}$$

with  $\bar{\mathbf{x}}_n = \frac{1}{n} \mathbf{X}_n \mathbf{1}_n$  where we use that

$$\frac{1}{n} \tilde{\mathbf{Y}}_n \tilde{\mathbf{Y}}_n^\top = \mathbf{S}_n = \frac{1}{n} \boldsymbol{\Sigma}^{1/2} \mathbf{X}_n \mathbf{X}_n^\top \boldsymbol{\Sigma}^{1/2} - \boldsymbol{\Sigma}^{1/2} \bar{\mathbf{x}}_n \bar{\mathbf{x}}_n^\top \boldsymbol{\Sigma}^{1/2}.$$

Hence,

$$\begin{aligned} \text{tr}((\mathbf{S}_n^+)^m \boldsymbol{\Theta}) &= \frac{(-1)^{m+1}}{m!} \frac{\partial^m}{\partial t^m} t \text{tr} \left[ \left( \frac{1}{n} \mathbf{X}_n \mathbf{X}_n^\top + t \boldsymbol{\Sigma}^{-1} \right)^{-1} \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\Theta} \boldsymbol{\Sigma}^{-1/2} \right] \Bigg|_{t=0} \\ &+ \frac{(-1)^{m+1}}{m!} \frac{\partial^m}{\partial t^m} t \frac{\bar{\mathbf{x}}_n^\top \left( \frac{1}{n} \mathbf{X}_n \mathbf{X}_n^\top + t \boldsymbol{\Sigma}^{-1} \right)^{-1} \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\Theta} \boldsymbol{\Sigma}^{-1/2} \left( \frac{1}{n} \mathbf{X}_n \mathbf{X}_n^\top + t \boldsymbol{\Sigma}^{-1} \right)^{-1} \bar{\mathbf{x}}_n}{1 - \bar{\mathbf{x}}_n^\top \left( \frac{1}{n} \mathbf{X}_n \mathbf{X}_n^\top + t \boldsymbol{\Sigma}^{-1} \right)^{-1} \bar{\mathbf{x}}_n} \Bigg|_{t=0}. \end{aligned}$$

Following [Pan \(2014, Eq. \(2.28\)\)](#), the quantity

$$\frac{1}{1 - \bar{\mathbf{x}}_n^\top \left( \frac{1}{n} \mathbf{X}_n \mathbf{X}_n^\top + t \boldsymbol{\Sigma}^{-1} \right)^{-1} \bar{\mathbf{x}}_n} = \frac{1}{1 - \bar{\mathbf{x}}_n^\top \boldsymbol{\Sigma}^{1/2} \left( \frac{1}{n} \boldsymbol{\Sigma}^{1/2} \mathbf{X}_n \mathbf{X}_n^\top \boldsymbol{\Sigma}^{1/2} + t \mathbf{I}_p \right)^{-1} \boldsymbol{\Sigma}^{1/2} \bar{\mathbf{x}}_n}$$

is uniformly bounded in  $t$ . Moreover, we have that

$$(S.5) \quad \left| \boldsymbol{\xi}^\top \left( \frac{1}{n} \boldsymbol{\Sigma}^{1/2} \mathbf{X}_n \mathbf{X}_n^\top \boldsymbol{\Sigma}^{1/2} + t \mathbf{I}_p \right)^{-1} \boldsymbol{\Sigma}^{1/2} \bar{\mathbf{x}}_n \right| \xrightarrow{a.s.} 0$$

following [Pan \(2014, p. 673\)](#) where  $\boldsymbol{\xi}^\top \boldsymbol{\xi}$  is assumed to be uniformly bounded in  $p$ . As a result, if  $\boldsymbol{\Theta} = \sum_{i=1}^k \boldsymbol{\theta}_i \boldsymbol{\theta}_i^\top$  with finite  $k$ , then

$$\begin{aligned} & \bar{\mathbf{x}}_n^\top \left( \frac{1}{n} \mathbf{X}_n \mathbf{X}_n^\top + t \boldsymbol{\Sigma}^{-1} \right)^{-1} \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\Theta} \boldsymbol{\Sigma}^{-1/2} \left( \frac{1}{n} \mathbf{X}_n \mathbf{X}_n^\top + t \boldsymbol{\Sigma}^{-1} \right)^{-1} \bar{\mathbf{x}}_n \\ &= \sum_{i=1}^k \left( \boldsymbol{\theta}_i^\top \boldsymbol{\Sigma}^{-1/2} \left( \frac{1}{n} \mathbf{X}_n \mathbf{X}_n^\top + t \boldsymbol{\Sigma}^{-1} \right)^{-1} \bar{\mathbf{x}}_n \right)^2 \\ &= \sum_{i=1}^k \left( \boldsymbol{\theta}_i^\top \left( \frac{1}{n} \boldsymbol{\Sigma}^{1/2} \mathbf{X}_n \mathbf{X}_n^\top \boldsymbol{\Sigma}^{1/2} + t \mathbf{I}_p \right)^{-1} \boldsymbol{\Sigma}^{1/2} \bar{\mathbf{x}}_n \right)^2 \xrightarrow{a.s.} 0. \end{aligned}$$

If the rank of  $\boldsymbol{\Theta}$  is not finite and we assume that  $\lambda_{\max}(\boldsymbol{\Theta}) = o(1)$ , then we get

$$\begin{aligned} & t \bar{\mathbf{x}}_n^\top \left( \frac{1}{n} \mathbf{X}_n \mathbf{X}_n^\top + t \boldsymbol{\Sigma}^{-1} \right)^{-1} \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\Theta} \boldsymbol{\Sigma}^{-1/2} \left( \frac{1}{n} \mathbf{X}_n \mathbf{X}_n^\top + t \boldsymbol{\Sigma}^{-1} \right)^{-1} \bar{\mathbf{x}}_n \\ & \leq \lambda_{\max} \left( t \left( \frac{1}{n} \mathbf{X}_n \mathbf{X}_n^\top + t \boldsymbol{\Sigma}^{-1} \right)^{-1/2} \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\Theta} \boldsymbol{\Sigma}^{-1/2} \left( \frac{1}{n} \mathbf{X}_n \mathbf{X}_n^\top + t \boldsymbol{\Sigma}^{-1} \right)^{-1/2} \right) \\ & \times \bar{\mathbf{x}}_n^\top \left( \frac{1}{n} \mathbf{X}_n \mathbf{X}_n^\top + t \boldsymbol{\Sigma}^{-1} \right)^{-1} \bar{\mathbf{x}}_n \end{aligned}$$

$$\begin{aligned}
&\leq \lambda_{\max}(\Theta) \lambda_{\max} \left( t \Sigma^{-1} \left( \frac{1}{n} \mathbf{X}_n \mathbf{X}_n^\top + t \Sigma^{-1} \right)^{-1} \right) \bar{\mathbf{x}}_n^\top \left( \frac{1}{n} \mathbf{X}_n \mathbf{X}_n^\top + t \Sigma^{-1} \right)^{-1} \bar{\mathbf{x}}_n \\
&\leq \lambda_{\max}(\Theta) \bar{\mathbf{x}}_n^\top \left( \frac{1}{n} \mathbf{X}_n \mathbf{X}_n^\top + t \Sigma^{-1} \right)^{-1} \bar{\mathbf{x}}_n \xrightarrow{a.s.} 0.
\end{aligned}$$

Thus,

$$\text{tr}((\mathbf{S}_n^+)^m \Theta) = \frac{(-1)^{m+1}}{m!} \frac{\partial^m}{\partial t^m} t \text{tr} \left[ \left( \frac{1}{n} \mathbf{X}_n \mathbf{X}_n^\top + t \Sigma^{-1} \right)^{-1} \Sigma^{-1/2} \Theta \Sigma^{-1/2} \right] \Big|_{t=0}.$$

Lemma 6 in [Bodnar et al. \(2024b\)](#) yields

$$\begin{aligned}
&\left| \text{tr} \left[ \left( \frac{1}{n} \mathbf{X}_n \mathbf{X}_n^\top + t \Sigma^{-1} \right)^{-1} \Sigma^{-1/2} \Theta \Sigma^{-1/2} \right] \right. \\
&\quad \left. - \text{tr} \left[ (\tilde{v}(t) \mathbf{I}_p + t \Sigma^{-1})^{-1} \Sigma^{-1/2} \Theta \Sigma^{-1/2} \right] \right| \xrightarrow{a.s.} 0,
\end{aligned} \tag{S.6}$$

for  $p/n \rightarrow c \in [1, \infty)$  as  $n \rightarrow \infty$  where  $\tilde{v}(t)$  satisfies the following equation

$$\frac{1}{\tilde{v}(t)} - 1 = c_n \frac{1}{p} \text{tr} \left[ \Sigma (\tilde{v}(t) \Sigma + t \mathbf{I}_p)^{-1} \right]$$

with  $c_n = p/n$ .

Let  $v(t) = \tilde{v}(t)/t$ . Then,  $v(t)$  is the solution of the following equation

$$\frac{1}{v(t)} - t = c_n \frac{1}{p} \text{tr} \left[ \Sigma (v(t) \Sigma + \mathbf{I}_p)^{-1} \right], \tag{S.7}$$

which can be rewritten as

$$1 - tv(t) = c_n \frac{1}{p} \text{tr} \left[ v(t) \Sigma (v(t) \Sigma + \mathbf{I}_p)^{-1} \right] = c_n - c_n \frac{1}{p} \text{tr} \left[ (v(t) \Sigma + \mathbf{I}_p)^{-1} \right],$$

As a result,  $v(0)$  solves the equation

$$\frac{1}{p} \text{tr} \left[ (v(0) \Sigma + \mathbf{I}_p)^{-1} \right] = \frac{c_n - 1}{c_n}.$$

The left hand-side of the equation is a monotonically decreasing function in  $v(0)$  taking the values from one to zero, while the right hand-side is equal to a number smaller than one. As a result, there is a unique solution of this equation, which is bounded for the fixed value of  $c_n$ .

Let

$$f_1(v) = \frac{1}{v} \quad \text{and} \quad f_2(v; \mathbf{A}) = \frac{1}{p} \text{tr} \left[ (v \mathbf{I}_p + \Sigma^{-1})^{-1} \mathbf{A} \right]. \tag{S.8}$$

for a matrix  $\mathbf{A}$ . Then, the  $k$ -th order derivatives of these two functions are expressed as

$$f_1^{(k)}(v) = (-1)^k k! \frac{1}{v^{k+1}}$$

and

$$\begin{aligned}
f_2^{(k)}(v; \mathbf{A}) &= \frac{(-1)^k k!}{p} \text{tr} \left[ (v \mathbf{I}_p + \Sigma^{-1})^{-(k+1)} \mathbf{A} \right] \\
&= \frac{(-1)^k k!}{p} \text{tr} \left\{ (v \Sigma + \mathbf{I}_p)^{-1} \left[ \Sigma (v \Sigma + \mathbf{I}_p)^{-1} \right]^k \Sigma^{1/2} \mathbf{A} \Sigma^{1/2} \right\}
\end{aligned}$$

for  $k = 1, 2, \dots$ . Then, the application of the Faà di Bruno formula (see Section 1.3 in [Krantz and Parks \(2002\)](#)) leads to

$$\begin{aligned}
 & \operatorname{tr}((\mathbf{S}_n^+)^m \Theta) \xrightarrow{d.a.s.} \frac{(-1)^{m+1}}{m!} \frac{\partial^m}{\partial t^m} t \operatorname{tr} \left[ (\tilde{v}(t) \mathbf{I}_p + t \Sigma^{-1})^{-1} \Sigma^{-1/2} \Theta \Sigma^{-1/2} \right] \Big|_{t=0} \\
 &= \frac{(-1)^{m+1}}{m!} \frac{\partial^m}{\partial t^m} \operatorname{tr} \left[ (v(t) \mathbf{I}_p + \Sigma^{-1})^{-1} \Sigma^{-1/2} \Theta \Sigma^{-1/2} \right] \Big|_{t=0} \\
 &= \frac{(-1)^{m+1}}{m!} \sum_{k=1}^m f_2^{(k)}(v(0); \Sigma^{-1/2} \Theta \Sigma^{-1/2}) B_{m,k} \left( v^{(1)}(0), v^{(2)}(0), \dots, v^{(m-k+1)}(0) \right) \\
 &= \sum_{k=1}^m \frac{(-1)^{m+k+1} k!}{m! p} \operatorname{tr} \left\{ (v(0) \Sigma + \mathbf{I}_p)^{-1} \left[ \Sigma (v(0) \Sigma + \mathbf{I}_p)^{-1} \right]^k \Theta \right\} B_{m,k} \left( v^{(1)}(0), v^{(2)}(0), \dots, v^{(m-k+1)}(0) \right)
 \end{aligned}$$

for  $p/n \rightarrow c \in (1, \infty)$  as  $n \rightarrow \infty$  where the symbol  $B_{m,k}(v_1, v_2, \dots, v_{m-k+1})$  denotes the partial exponential Bell polynomials defined in Section 2.

To compute  $v^{(1)}(0), v^{(2)}(0), \dots, v^{(m-k+1)}(0)$ , we rewrite (S.7) as

$$(S.9) \quad \frac{1}{v(t)} - t = c_n \frac{1}{p} \operatorname{tr} \left[ (v(t) \mathbf{I}_p + \Sigma^{-1})^{-1} \right].$$

Hence,

$$-\frac{v^{(1)}(t)}{v(t)^2} - 1 = -c_n v^{(1)}(t) \frac{1}{p} \operatorname{tr} \left[ (v(t) \mathbf{I}_p + \Sigma^{-1})^{-2} \right] = -c_n v^{(1)}(t) \frac{1}{p} \operatorname{tr} \left\{ \left[ \Sigma (v(t) \Sigma + \mathbf{I}_p)^{-1} \right]^2 \right\},$$

from which we get

$$v^{(1)}(0) = - \frac{1}{\frac{1}{v(0)^2} - c_n \frac{1}{p} \operatorname{tr} \left\{ \left[ \Sigma (v(0) \Sigma + \mathbf{I}_p)^{-1} \right]^2 \right\}}$$

For the computation of the derivative of  $v(\cdot)$  of the order larger than one, we apply the Faà di Bruno formula to both the sides of (S.9). This yields

$$\sum_{k=1}^m f_1^{(k)}(v(t)) B_{m,k} \left( v^{(1)}(t), \dots, v^{(m-k+1)}(t) \right) = c_n \sum_{k=1}^m f_2^{(k)}(v(t); \mathbf{I}_p) B_{m,k} \left( v^{(1)}(t), \dots, v^{(m-k+1)}(t) \right)$$

or

$$\begin{aligned}
 & \sum_{k=1}^m \frac{(-1)^k k!}{v(t)^{k+1}} B_{m,k} \left( v^{(1)}(t), \dots, v^{(m-k+1)}(t) \right) \\
 &= c_n \sum_{k=1}^m \frac{(-1)^k k!}{p} \operatorname{tr} \left\{ \left[ \Sigma (v(t) \Sigma + \mathbf{I}_p)^{-1} \right]^{k+1} \right\} B_{m,k} \left( v^{(1)}(t), \dots, v^{(m-k+1)}(t) \right)
 \end{aligned}$$

Using that  $B_{m,1}(v^{(1)}(t), v^{(2)}(t), \dots, v^{(m)}(t)) = v^{(m)}(t)$ , we finally get

$$v^{(m)}(0) = \frac{\sum_{k=2}^m (-1)^k k! \left( \frac{1}{v(0)^{k+1}} - c_n \frac{1}{p} \operatorname{tr} \left\{ \left[ \Sigma (v(0) \Sigma + \mathbf{I}_p)^{-1} \right]^{k+1} \right\} \right) B_{m,k} \left( v^{(1)}(0), \dots, v^{(m-k+1)}(0) \right)}{\frac{1}{v(0)^2} - c_n \frac{1}{p} \operatorname{tr} \left\{ \left[ \Sigma (v(0) \Sigma + \mathbf{I}_p)^{-1} \right]^2 \right\}}$$

□

S.5.3. *Proof of Corollary 2.1.*

PROOF. From the properties of the partial exponential Bell polynomials we get that

$$\begin{aligned}
B_{1,1} \left( v^{(1)}(0) \right) &= v^{(1)}(0), \\
B_{2,1} \left( v^{(1)}(0), v^{(2)}(0) \right) &= v^{(2)}(0), \\
B_{3,1} \left( v^{(1)}(0), v^{(2)}(0), v^{(3)}(0) \right) &= v^{(3)}(0), \\
B_{4,1} \left( v^{(1)}(0), v^{(2)}(0), v^{(3)}(0), v^{(4)}(0) \right) &= v^{(4)}(0), \\
B_{2,2} \left( v^{(1)}(0) \right) &= [v^{(1)}(0)]^2, \\
B_{3,2} \left( v^{(1)}(0), v^{(2)}(0) \right) &= 3v^{(1)}(0)v^{(2)}(0), \\
B_{3,3} \left( v^{(1)}(0) \right) &= [v^{(1)}(0)]^3, \\
B_{4,2} \left( v^{(1)}(0), v^{(2)}(0), v^{(3)}(0) \right) &= 4v^{(1)}(0)v^{(3)}(0) + 3[v^{(2)}(0)]^2, \\
B_{4,3} \left( v^{(1)}(0), v^{(2)}(0) \right) &= 6[v^{(1)}(0)]^2v^{(2)}(0), \\
B_{4,4} \left( v^{(1)}(0) \right) &= [v^{(1)}(0)]^4.
\end{aligned}$$

The substitution of the above formulas for the partial exponential Bell polynomials to (10) and (13) and the application of (12) leads to the statement of the corollary.  $\square$

S.5.4. *Proof of Corollary 2.2.*

PROOF. Let

$$g_k = \frac{1}{p} \operatorname{tr} \left\{ \left[ \Sigma (v(0)\Sigma + \mathbf{I}_p)^{-1} \right]^k \right\}.$$

Then,

$$d_k \left( \frac{1}{p} \mathbf{I}_p \right) = \frac{1}{p} \operatorname{tr} \left\{ (v(0)\Sigma + \mathbf{I}_p)^{-1} \left[ \Sigma (v(0)\Sigma + \mathbf{I}_p)^{-1} \right]^k \mathbf{I}_p \right\} = g_k - v(0)g_{k+1}$$

and, consequently,

$$\begin{aligned}
s_m \left( \frac{1}{p} \mathbf{I}_p \right) &= \sum_{k=1}^m \frac{(-1)^{m+k+1} k!}{m!} (g_k - v(0)g_{k+1}) B_{m,k} \left( v^{(1)}(0), \dots, v^{(m-k+1)}(0) \right) \\
&= \frac{(-1)^m}{m!} (g_1 - v(0)g_2) B_{m,1} \left( v^{(1)}(0), \dots, v^{(m)}(0) \right) \\
&\quad + \sum_{k=2}^m \frac{(-1)^{m+k+1} k!}{m!} (g_k - v(0)g_{k+1}) B_{m,k} \left( v^{(1)}(0), \dots, v^{(m-k+1)}(0) \right).
\end{aligned}$$

The application of  $B_{m,1} \left( v^{(1)}(0), \dots, v^{(m)}(0) \right) = v^{(m)}(0)$  and (13) yields

$$s_m \left( \frac{1}{p} \mathbf{I}_p \right) = \frac{(-1)^{m+1}}{m!} \sum_{k=2}^m (-1)^k k! (g_1 - v(0)g_2) v^{(1)}(0) h_{k+1} B_{m,k} \left( v^{(1)}(0), \dots, v^{(m-k+1)}(0) \right)$$

$$\begin{aligned}
 & + \frac{(-1)^{m+1}}{m!} \sum_{k=2}^m (-1)^k k! (g_k - v(0)g_{k+1}) B_{m,k} \left( v^{(1)}(0), \dots, v^{(m-k+1)}(0) \right) \\
 & = \frac{(-1)^{m+1}}{m!} \sum_{k=2}^m (-1)^k k! \left[ (g_1 - v(0)g_2) v^{(1)}(0) h_{k+1} + g_k - v(0)g_{k+1} \right] B_{m,k} \left( v^{(1)}(0), \dots, v^{(m-k+1)}(0) \right).
 \end{aligned}$$

Using

$$\begin{aligned}
 & (g_1 - v(0)g_2) v^{(1)}(0) h_{k+1} + g_k - v(0)g_{k+1} \\
 & = \left( \frac{1}{p} \operatorname{tr} \left[ \Sigma(v(0)\Sigma + \mathbf{I}_p)^{-1} \right] - v(0)g_2 \right) \frac{-v(0)^2}{1 - c_n v(0)^2 g_2} h_{k+1} + g_k - v(0)g_{k+1} \\
 & = \left( \frac{1}{v(0)} \left( 1 - \frac{1}{p} \operatorname{tr} \left[ (v(0)\Sigma + \mathbf{I}_p)^{-1} \right] \right) - v(0)g_2 \right) \frac{-v(0)^2}{1 - c_n v(0)^2 g_2} h_{k+1} + g_k - v(0)g_{k+1} \\
 & = \left( \frac{1}{v(0)} \left( 1 - \frac{c_n}{c_n - 1} \right) - v(0)g_2 \right) \frac{-v(0)^2}{1 - c_n v(0)^2 g_2} h_{k+1} + g_k - v(0)g_{k+1} \\
 & = \frac{-v(0)}{c_n} h_{k+1} + g_k - v(0)g_{k+1} = \frac{-h_k}{c_n}
 \end{aligned}$$

and the equality (see Equation (1.4) in Cvijović (2011))

$$k! B_{m,k} \left( v^{(1)}(0), \dots, v^{(m-k+1)}(0) \right) = (k-1)! \sum_{r=1}^{m-1} \frac{m!}{r!(m-r)!} v^{(r)}(0) B_{m-r,k-1} \left( v^{(1)}(0), \dots, v^{(m-r-k+2)}(0) \right),$$

where  $B_{i,k-1}(\cdot) = 0$  for  $i < k-1$ , we get

$$\begin{aligned}
 s_m \left( \frac{1}{p} \mathbf{I}_p \right) & = \frac{(-1)^m}{m! c_n} \sum_{k=2}^m (-1)^k h_k (k-1)! \sum_{r=1}^{m-1} \frac{m!}{r!(m-r)!} v^{(r)}(0) B_{m-r,k-1} \left( v^{(1)}(0), \dots, v^{(m-r-k)}(0) \right) \\
 & = \frac{(-1)^m}{m! c_n} \sum_{r=1}^{m-1} \frac{m!}{r!(m-r)!} v^{(r)}(0) \sum_{k=2}^{m-r+1} (-1)^k h_k (k-1)! B_{m-r,k-1} \left( v^{(1)}(0), \dots, v^{(m-r-k+2)}(0) \right) \\
 & = \frac{(-1)^m}{m! c_n} \frac{m!}{(m-1)!} v^{(m-1)}(0) h_2 v^{(1)}(0) + \frac{(-1)^m}{m! c_n} \sum_{r=1}^{m-2} \frac{m!}{r!(m-r)!} v^{(r)}(0) \left( h_2 v^{(m-r)}(0) \right. \\
 & \quad \left. - \sum_{k=3}^{m-r+1} (-1)^{k-1} h_k (k-1)! B_{m-r,k-1} \left( v^{(1)}(0), \dots, v^{(m-r-k+2)}(0) \right) \right) \\
 & = \frac{(-1)^{m-1}}{(m-1)! c_n} v^{(m-1)}(0),
 \end{aligned}$$

where we used that  $v^{(1)}(0) = -h_2^{-1}$  and

$$\sum_{k=3}^{m-r+1} (-1)^{k-1} h_k (k-1)! B_{m-r,k-1} \left( v^{(1)}(0), \dots, v^{(m-r-k+2)}(0) \right) = -\frac{v^{(m-r)}(0)}{v^{(1)}(0)}.$$

□

### S.5.5. Proof of Corollary 2.3.

PROOF. If  $\Sigma = \mathbf{I}_p$ , then

$$d_k(\Theta) = (v(0) + 1)^{-(k+1)} \text{tr}(\Theta), \quad k = 1, 2, \dots$$

and  $s_m(\Theta) = a_m \text{tr}(\Theta)$  for  $m = 1, 2, \dots$  where  $a_m$  does not depend on  $\Theta$ . As such,  $a_m$  can be computed by setting  $\Theta = \frac{1}{p} \mathbf{I}_p$  and using the findings of Corollary 2.2. Thus,

$$s_m(\Theta) = s_m\left(\frac{1}{p} \mathbf{I}_p\right) \text{tr}(\Theta) = \frac{(-1)^{m-1} v^{(m-1)}(0)}{(m-1)! c_n} \text{tr}(\Theta),$$

where it holds from (6) that

$$\frac{1}{v(0) + 1} = \frac{c_n - 1}{c_n}$$

and, hence,

$$(S.10) \quad v(0) = \frac{1}{c_n - 1} \quad \text{and} \quad v^{(1)}(0) = -\frac{c_n}{(c_n - 1)^3},$$

where the second equality follows from (12).

Finally, from (13) we get

$$\begin{aligned} v^{(m)}(0) &= -v^{(1)}(0) \sum_{k=2}^m (-1)^k k! \left( v(0)^{-(k+1)} - c_n (v(0) + 1)^{-(k+1)} \right) B_{m,k} \left( v^{(1)}(0), \dots, v^{(m-k+1)}(0) \right) \\ &= \frac{c_n}{(c_n - 1)^3} \sum_{k=2}^m (-1)^k k! (c_n - 1)^{(k+1)} \left( 1 - c_n^{-k} \right) B_{m,k} \left( v^{(1)}(0), \dots, v^{(m-k+1)}(0) \right). \end{aligned}$$

□

### S.5.6. Proof of Theorem 2.2.

PROOF. Following the proof of Theorem 2.1, we get

$$\text{tr}((\mathbf{S}_n + t\mathbf{I}_p)^{-(m+1)} \Theta) = \frac{(-1)^m}{m!} \frac{\partial^m}{\partial t^m} \text{tr} \left[ \left( \frac{1}{n} \mathbf{X}_n \mathbf{X}_n^\top + t\Sigma^{-1} \right)^{-1} \Sigma^{-1/2} \Theta \Sigma^{-1/2} \right],$$

while the application of Lemma 6 in Bodnar et al. (2024b) leads to

$$\left| \text{tr} \left[ \left( \frac{1}{n} \mathbf{X}_n \mathbf{X}_n^\top + t\Sigma^{-1} \right)^{-1} \Sigma^{-1/2} \Theta \Sigma^{-1/2} \right] - \frac{1}{t} \text{tr} \left[ (v(t)\mathbf{I}_p + \Sigma^{-1})^{-1} \Sigma^{-1/2} \Theta \Sigma^{-1/2} \right] \right| \xrightarrow{a.s.} 0,$$

for  $p/n \rightarrow c \in (0, \infty)$  as  $n \rightarrow \infty$  where  $v(t)$  is the solution of (S.7). Hence,

$$\begin{aligned} &\text{tr}((\mathbf{S}_n + t\mathbf{I}_p)^{-(m+1)} \Theta) \xrightarrow{d.a.s.} \frac{(-1)^m}{m!} \frac{\partial^m}{\partial t^m} \frac{1}{t} \text{tr} \left[ (v(t)\mathbf{I}_p + \Sigma^{-1})^{-1} \Sigma^{-1/2} \Theta \Sigma^{-1/2} \right] \\ &= \frac{(-1)^m}{m!} \frac{\partial^m}{\partial t^m} \frac{1}{t} \text{tr} \left[ (v(t)\mathbf{I}_p + \Sigma^{-1})^{-1} \Sigma^{-1/2} \Theta \Sigma^{-1/2} \right] \\ &\quad + \frac{(-1)^m}{m!} \sum_{l=1}^m \frac{m!}{l!(m-l)!} \frac{\partial^{m-l}}{\partial t^{m-l}} \frac{1}{t} \frac{\partial^l}{\partial t^l} \text{tr} \left[ (v(t)\mathbf{I}_p + \Sigma^{-1})^{-1} \Sigma^{-1/2} \Theta \Sigma^{-1/2} \right] \\ &= t^{-m-1} \text{tr} \left[ (v(t)\mathbf{I}_p + \Sigma^{-1})^{-1} \Sigma^{-1/2} \Theta \Sigma^{-1/2} \right] \end{aligned}$$

$$\begin{aligned}
 & + \frac{(-1)^m}{m!} \sum_{l=1}^m \frac{m!}{l!(m-l)!} (-1)^{m-l} (m-l)! t^{-(m-l)-1} \sum_{k=1}^l f_2^{(k)}(v(t); \Sigma^{-1/2} \Theta \Sigma^{-1/2}) \\
 & \times B_{l,k} \left( v^{(1)}(t), v^{(2)}(t), \dots, v^{(l-k+1)}(t) \right) \\
 & = t^{-m-1} \text{tr} \left[ (v(t) \Sigma + \mathbf{I}_p)^{-1} \Theta \right] \\
 & + \sum_{l=1}^m t^{-(m-l)-1} \sum_{k=1}^l \frac{(-1)^{l+k} k!}{l!} \text{tr} \left\{ (v(t) \Sigma + \mathbf{I}_p)^{-1} \left[ \Sigma (v(t) \Sigma + \mathbf{I}_p)^{-1} \right]^k \Theta \right\} \\
 & \times B_{l,k} \left( v^{(1)}(t), v^{(2)}(t), \dots, v^{(l-k+1)}(t) \right)
 \end{aligned}$$

for  $p/n \rightarrow c \in (0, \infty)$  as  $n \rightarrow \infty$ . Furthermore,  $v(t)$  is the solution of (S.7) which can be rewritten as (5),

$$v^{(1)}(t) = - \frac{1}{\frac{1}{v(t)^2} - c_n \frac{1}{p} \text{tr} \left\{ \left[ \Sigma (v(t) \Sigma + \mathbf{I}_p)^{-1} \right]^2 \right\}},$$

and

$$v^{(m)}(t) = \frac{\sum_{k=2}^m (-1)^k k! \left( \frac{1}{v(t)^{k+1}} - c_n \frac{1}{p} \text{tr} \left\{ \left[ \Sigma (v(t) \Sigma + \mathbf{I}_p)^{-1} \right]^{k+1} \right\} \right) B_{m,k} \left( v^{(1)}(t), \dots, v^{(m-k+1)}(t) \right)}{\frac{1}{v(t)^2} - c_n \frac{1}{p} \text{tr} \left\{ \left[ \Sigma (v(t) \Sigma + \mathbf{I}_p)^{-1} \right]^2 \right\}}.$$

□

### S.5.7. Proof of 2.4.

PROOF. The application of Theorem 2.2 leads to

$$\begin{aligned}
 \tilde{s}_1(t, \Theta) &= t^{-1} d_0(t, \Theta), \\
 \tilde{s}_2(t, \Theta) &= t^{-2} d_0(t, \Theta) + t^{-1} d_1(t, \Theta) B_{1,1} \left( v^{(1)}(t) \right), \\
 \tilde{s}_3(t, \Theta) &= t^{-3} d_0(t, \Theta) + t^{-2} d_1(t, \Theta) B_{1,1} \left( v^{(1)}(t) \right) \\
 & + t^{-1} \left\{ -\frac{1}{2} v^{(2)}(t) d_1(t, \Theta) B_{2,1} \left( v^{(1)}(t), v^{(2)}(t) \right) + d_2(t, \Theta) B_{2,2} \left( v^{(1)}(t) \right) \right\}, \\
 \tilde{s}_4(t, \Theta) &= t^{-4} d_0(t, \Theta) + t^{-3} d_1(t, \Theta) B_{3,1} \left( v^{(1)}(t) \right) \\
 & + t^{-2} \left\{ -\frac{1}{2} d_1(t, \Theta) B_{2,1} \left( v^{(1)}(t), v^{(2)}(t) \right) + d_2(t, \Theta) B_{2,2} \left( v^{(1)}(t) \right) \right\} \\
 & + t^{-1} \left\{ \frac{1}{6} d_1(t, \Theta) B_{3,1} \left( v^{(1)}(t), v^{(2)}(t), v^{(3)}(t) \right) - \frac{1}{3} d_2(t, \Theta) B_{3,2} \left( v^{(1)}(t), v^{(2)}(t) \right) + d_3(t, \Theta) B_{3,3} \left( v^{(1)}(t) \right) \right\}.
 \end{aligned}$$

The rest of the proof follows by using the formulas for the partial exponential Bell polynomials as given in the proof of Corollary 2.1. □

S.5.8. *Proof of Corollary 2.5.*

PROOF. The proof of Corollary 2.5 is done by using the methods of mathematical induction. For  $m = 0$ , we get

$$\tilde{s}_1 \left( t, \frac{1}{p} \mathbf{I}_p \right) = t^{-1} d_k \left( t, \frac{1}{p} \mathbf{I}_p \right) = t^{-1} \frac{c_n - 1}{c_n} + \frac{v(t)}{c_n},$$

where the last equality follows from (5) and (22).

Next, we assume that the statement of the corollary is true for  $\tilde{s}_1 \left( t, \frac{1}{p} \mathbf{I}_p \right), \dots, \tilde{s}_m \left( t, \frac{1}{p} \mathbf{I}_p \right)$  and prove it for  $\tilde{s}_{m+1} \left( t, \frac{1}{p} \mathbf{I}_p \right)$ . Let

$$g_k(t) = \frac{1}{p} \operatorname{tr} \left\{ \left[ \Sigma(v(t)\Sigma + \mathbf{I}_p)^{-1} \right]^k \right\}.$$

Then,

$$d_k \left( t, \frac{1}{p} \mathbf{I}_p \right) = \frac{1}{p} \operatorname{tr} \left\{ (v(t)\Sigma + \mathbf{I}_p)^{-1} \left[ \Sigma(v(t)\Sigma + \mathbf{I}_p)^{-1} \right]^k \mathbf{I}_p \right\} = g_k(t) - v(t)g_{k+1}(t)$$

and  $h_k(t) = v(t)^{-k} - c_n g_k(t)$ . Moreover, from the definition of  $v^{(1)}(t)$  we have

$$v^{(1)}(t)h_2(t) = -1$$

and

$$v^{(1)}(t)(g_1(t) - v(t)g_2(t)) = \frac{-1}{v(t)^{-2} - c_n g_2(t)} \left( \frac{1 - tv(t)}{c_n v(t)} - v(t)g_2(t) \right) = -\frac{v(t)}{c_n} - \frac{tv^{(1)}(t)}{c_n}.$$

Using the recursion (26) and the assumption of the induction, we get

$$\begin{aligned} \tilde{s}_{m+1} \left( t, \frac{1}{p} \mathbf{I}_p \right) &= t^{-(m+1)} \frac{c_n - 1}{c_n} + t^{-1} \frac{(-1)^{m-1} v^{(m-1)}(t)}{(m-1)!c_n} \\ &\quad + t^{-1} \left( \frac{(-1)^{m-1}}{m!} (g_1(t) - v(t)g_2(t)) B_{m,1} \left( v^{(1)}(t), \dots, v^{(m)}(t) \right) \right. \\ &\quad \left. + \sum_{k=2}^m \frac{(-1)^{m+k} k!}{m!} (g_k(t) - v(t)g_{k+1}(t)) B_{m,k} \left( v^{(1)}(t), \dots, v^{(m-k+1)}(t) \right) \right) \\ &= t^{-(m+1)} \frac{c_n - 1}{c_n} + t^{-1} \frac{(-1)^{m-1} v^{(m-1)}(t)}{(m-1)!c_n} \\ &\quad + \frac{(-1)^m}{m!t} \sum_{k=2}^m (-1)^k k! \left[ (g_1(t) - v(t)g_2(t)) v^{(1)}(t) h_{k+1}(t) + g_k(t) - v(t)g_{k+1}(t) \right] \\ &\quad \times B_{m,k} \left( v^{(1)}(t), \dots, v^{(m-k+1)}(t) \right), \end{aligned}$$

following the proof of Corollary 2.2 where

$$\begin{aligned} &(g_1(t) - v(t)g_2(t)) v^{(1)}(t) h_{k+1}(t) + g_k(t) - v(t)g_{k+1}(t) \\ &= \left( -\frac{v(t)}{c_n} - \frac{tv^{(1)}(t)}{c_n} \right) h_{k+1}(t) + g_k(t) - v(t)g_{k+1}(t) = -\frac{h_k(t)}{c_n} - \frac{tv^{(1)}(t)}{c_n} h_{k+1}(t). \end{aligned}$$

Hence,

$$\begin{aligned}
 \tilde{s}_{m+1} \left( t, \frac{1}{p} \mathbf{I}_p \right) &= t^{-(m+1)} \frac{c_n - 1}{c_n} + t^{-1} \frac{(-1)^{m-1} v^{(m-1)}(t)}{(m-1)! c_n} \\
 &\quad + t^{-1} \frac{(-1)^{m-1}}{m! c_n} \sum_{k=2}^m (-1)^k k! h_k(t) B_{m,k} \left( v^{(1)}(t), \dots, v^{(m-k+1)}(t) \right) \\
 &\quad + \frac{(-1)^m}{m! c_n} (-1) v^{(1)}(t) \sum_{k=2}^m (-1)^k k! h_{k+1}(t) B_{m,k} \left( v^{(1)}(t), \dots, v^{(m-k+1)}(t) \right) \\
 &= t^{-(m+1)} \frac{c_n - 1}{c_n} + t^{-1} \frac{(-1)^m v^{(m)}(t)}{m! c_n},
 \end{aligned}$$

following the end of the proof of Corollary 2.2.  $\square$

### S.5.9. Proof of Corollary 2.6.

PROOF. If  $\Sigma = \mathbf{I}_p$ , then

$$d_k(t, \Theta) = (v(t) + 1)^{-(k+1)} \text{tr}(\Theta), \quad k = 0, 1, 2, \dots$$

As a result,  $\tilde{s}_{m+1}(t, \Theta) = a_{m+1}(t) \text{tr}(\Theta)$  for  $m = 0, 1, \dots$  where  $a_{m+1}(t)$ ,  $m = 0, 1, \dots$ , does not depend on  $\Theta$ . Consequently,  $a_{m+1}(t)$ ,  $m = 0, 1, \dots$ , can be computed by setting  $\Theta = \frac{1}{p} \mathbf{I}_p$  and using the results of Corollary 2.5. It leads to

$$\tilde{s}_{m+1}(t, \Theta) = \tilde{s}_{m+1} \left( t, \frac{1}{p} \mathbf{I}_p \right) \text{tr}(\Theta) = \left( t^{-(m+1)} \frac{c_n - 1}{c_n} + \frac{(-1)^m v^{(m)}(t)}{(m)! c_n} \right) \text{tr}(\Theta),$$

where  $v(t)$  is the solution of

$$\frac{1}{v(t) + 1} = \frac{c_n - 1 + tv(t)}{c_n}$$

and, hence,

$$v(t) = \frac{-(c_n - 1 + t) + \sqrt{(c_n - 1 + t)^2 + 4t}}{2t} = \frac{2}{(c_n - 1 + t) + \sqrt{(c_n - 1 + t)^2 + 4t}}.$$

Moreover, it holds from (23) and (24) that

$$v^{(1)}(t) = \frac{-1}{v(t)^{-2} - c_n(v(t) + 1)^{-2}}$$

and

$$v^{(m)}(t) = -v^{(1)}(t) \sum_{k=2}^m (-1)^k k! \left( v(t)^{-(k+1)} - c_n(v(t) + 1)^{-(k+1)} \right) B_{m,k} \left( v^{(1)}(t), \dots, v^{(m-k+1)}(t) \right).$$

$\square$

### S.5.10. Proof of Corollary 2.7.

PROOF. From the definition of  $D_m(t, \Theta)$  and the recursion (26), we get

$$\tilde{s}_{m+1}(t, \Theta) = \frac{1}{t} \tilde{s}_m(t, \Theta) + \frac{1}{t} D_m(t, \Theta)$$

and, similarly,

$$\tilde{s}_{m+k}(t, \Theta) = t^{-k} \tilde{s}_m(t, \Theta) + t^{-k} D_m(t, \Theta) + \sum_{j=1}^{k-1} t^{-(k-j)} D_{m+j}(t, \Theta) \text{ for } k \geq 2.$$

Hence,

$$\begin{aligned} \dot{s}_m(t, \Theta) &= \tilde{s}_m(t, \Theta) + \sum_{k=1}^m (-1)^k t^k \binom{m}{k} \left( t^{-k} \tilde{s}_m(t, \Theta) + t^{-k} D_m(t, \Theta) \right) \\ &+ \sum_{k=2}^m (-1)^k t^k \binom{m}{k} \sum_{j=1}^{k-1} t^{-(k-j)} D_{m+j}(t, \Theta) \\ &= (\tilde{s}_m(t, \Theta) + D_m(t, \Theta)) \sum_{k=0}^m (-1)^k \binom{m}{k} - D_m(t, \Theta) + \sum_{k=2}^m (-1)^k \binom{m}{k} \sum_{j=1}^{k-1} t^j D_{m+j}(t, \Theta). \end{aligned}$$

Since  $\sum_{k=0}^m (-1)^k \binom{m}{k} = (1-1)^m = 0$ , we get the statement of the corollary.  $\square$

### S.5.11. Proof of Corollary 2.9.

PROOF. The application of Corollary 2.5 for  $m = 1, 2, \dots$  leads to

$$\begin{aligned} \dot{s}_m\left(t, \frac{1}{p} \mathbf{I}_p\right) &= \sum_{k=0}^m (-1)^k t^k \binom{m}{k} t^{-(m+k)} \frac{c_n - 1}{c_n} + \sum_{k=0}^m (-1)^k t^k \binom{m}{k} \frac{(-1)^{m+k-1} v^{(m+k-1)}(t)}{(m+k-1)! c_n} \\ &= \frac{c_n - 1}{c_n} t^{-m} \underbrace{\sum_{k=0}^m (-1)^k \binom{m}{k}}_{(1-1)^m=0} + \frac{(-1)^{m-1}}{c_n} \sum_{k=0}^m \binom{m}{k} \frac{v^{(m+k-1)}(t)}{(m+k-1)!} t^k \end{aligned}$$

and the statement of the corollary follows.  $\square$

### S.5.12. Proof of Theorem 3.1.

PROOF. The loss function  $L_{F;n}^2$  from (42) can be rewritten in the following way

$$\begin{aligned} L_{F;n}^2 &= \|\widehat{\Pi}_{GSE} \Sigma - \mathbf{I}_p\|_F^2 = \text{tr} \left[ (\widehat{\Pi}_{GSE} \Sigma - \mathbf{I}_p) (\widehat{\Pi}_{GSE} \Sigma - \mathbf{I}_p)^\top \right] \\ &= \|\mathbf{I}_p\|_F^2 + \|\mathbf{S}_n^\#(t) \Sigma\|_F^2 \left( \alpha_n + \frac{\beta_n \text{tr}(\mathbf{S}_n^\#(t) \Sigma^2 \Pi_0) - \text{tr}(\mathbf{S}_n^\#(t) \Sigma)}{\|\mathbf{S}_n^\#(t) \Sigma\|_F^2} \right)^2 \\ &+ \frac{\|\mathbf{S}_n^\#(t) \Sigma\|_F^2 \|\Pi_0 \Sigma\|_F^2 - \left( \text{tr}(\mathbf{S}_n^\#(t) \Sigma^2 \Pi_0) \right)^2}{\|\mathbf{S}_n^\#(t) \Sigma\|_F^2} \\ &\times \left( \beta_n + \frac{\text{tr}(\mathbf{S}_n^\#(t) \Sigma^2 \Pi_0) \text{tr}(\mathbf{S}_n^\#(t) \Sigma) - \|\mathbf{S}_n^\#(t) \Sigma\|_F^2 \text{tr}(\Sigma \Pi_0)}{\|\mathbf{S}_n^\#(t) \Sigma\|_F^2 \|\Pi_0 \Sigma\|_F^2 - \left( \text{tr}(\mathbf{S}_n^\#(t) \Sigma^2 \Pi_0) \right)^2} \right)^2 \\ (S.11) \quad &- \left( \frac{[\text{tr}(\mathbf{S}_n^\#(t) \Sigma)]^2}{\|\mathbf{S}_n^\#(t) \Sigma\|_F^2} + \frac{\left[ \text{tr}(\mathbf{S}_n^\#(t) \Sigma^2 \Pi_0) \text{tr}(\mathbf{S}_n^\#(t) \Sigma) - \|\mathbf{S}_n^\#(t) \Sigma\|_F^2 \text{tr}(\Sigma \Pi_0) \right]^2}{\|\mathbf{S}_n^\#(t) \Sigma\|_F^2 \left[ \|\mathbf{S}_n^\#(t) \Sigma\|_F^2 \|\Pi_0 \Sigma\|_F^2 - \left( \text{tr}(\mathbf{S}_n^\#(t) \Sigma^2 \Pi_0) \right)^2 \right]} \right). \end{aligned}$$

Since  $\|\mathbf{S}_n^\#(t)\boldsymbol{\Sigma}\|_F^2 > 0$  and  $\|\mathbf{S}_n^\#(t)\boldsymbol{\Sigma}\|_F^2\|\boldsymbol{\Pi}_0\boldsymbol{\Sigma}\|_F^2 - \left(\text{tr}(\mathbf{S}_n^\#(t)\boldsymbol{\Sigma}^2\boldsymbol{\Pi}_0)\right)^2 \geq 0$  with equality to zero if and only if  $\mathbf{S}_n^\#(t) = \boldsymbol{\Pi}_0$ , then, for a fixed value of the tuning parameter  $t$ , the loss function  $L_{F;n}^2$  is minimized at

$$(S.12) \quad \begin{aligned} \beta_n^*(\mathbf{S}_n^\#(t)) &= \frac{\|\mathbf{S}_n^\#(t)\boldsymbol{\Sigma}\|_F^2 \text{tr}(\boldsymbol{\Sigma}\boldsymbol{\Pi}_0) - \text{tr}(\mathbf{S}_n^\#(t)\boldsymbol{\Sigma}^2\boldsymbol{\Pi}_0)\text{tr}(\mathbf{S}_n^\#(t)\boldsymbol{\Sigma})}{\|\mathbf{S}_n^\#(t)\boldsymbol{\Sigma}\|_F^2\|\boldsymbol{\Pi}_0\boldsymbol{\Sigma}\|_F^2 - \left(\text{tr}(\mathbf{S}_n^\#(t)\boldsymbol{\Sigma}^2\boldsymbol{\Pi}_0)\right)^2} \\ &= \frac{\text{tr}\left(\frac{\boldsymbol{\Sigma}\boldsymbol{\Pi}_0}{\|\boldsymbol{\Sigma}\boldsymbol{\Pi}_0\|_F}\right)\|\mathbf{S}_n^\#(t)\boldsymbol{\Sigma}\|_F^2 - \text{tr}(\mathbf{S}_n^\#(t)\boldsymbol{\Sigma})\text{tr}\left(\mathbf{S}_n^\#(t)\frac{\boldsymbol{\Sigma}^2\boldsymbol{\Pi}_0}{\|\boldsymbol{\Sigma}\boldsymbol{\Pi}_0\|_F}\right)}{\|\mathbf{S}_n^\#(t)\boldsymbol{\Sigma}\|_F^2 - \left[\text{tr}\left(\mathbf{S}_n^\#(t)\frac{\boldsymbol{\Sigma}^2\boldsymbol{\Pi}_0}{\|\boldsymbol{\Sigma}\boldsymbol{\Pi}_0\|_F}\right)\right]^2} \|\boldsymbol{\Sigma}\boldsymbol{\Pi}_0\|_F^{-1}. \end{aligned}$$

and

$$(S.13) \quad \alpha_n^*(\mathbf{S}_n^\#(t)) = \frac{\text{tr}(\mathbf{S}_n^\#(t)\boldsymbol{\Sigma}) - \text{tr}\left(\frac{\boldsymbol{\Sigma}\boldsymbol{\Pi}_0}{\|\boldsymbol{\Sigma}\boldsymbol{\Pi}_0\|_F}\right)\text{tr}\left(\mathbf{S}_n^\#(t)\frac{\boldsymbol{\Sigma}^2\boldsymbol{\Pi}_0}{\|\boldsymbol{\Sigma}\boldsymbol{\Pi}_0\|_F}\right)}{\|\mathbf{S}_n^\#(t)\boldsymbol{\Sigma}\|_F^2 - \left[\text{tr}\left(\mathbf{S}_n^\#(t)\frac{\boldsymbol{\Sigma}^2\boldsymbol{\Pi}_0}{\|\boldsymbol{\Sigma}\boldsymbol{\Pi}_0\|_F}\right)\right]^2},$$

When  $\mathbf{S}_n^\#(t) \in \{\mathbf{S}_n^-(t), \mathbf{S}_n^\pm(t)\}$ , the last summand in (S.11) should be minimized with respect to the tuning parameter  $t$ . The last row in (S.11) can be written by

$$\begin{aligned} &\frac{[\text{tr}(\mathbf{S}_n^\#(t)\boldsymbol{\Sigma})]^2}{\|\mathbf{S}_n^\#(t)\boldsymbol{\Sigma}\|_F^2} + \frac{[\text{tr}(\mathbf{S}_n^\#(t)\boldsymbol{\Sigma}^2\boldsymbol{\Pi}_0)\text{tr}(\mathbf{S}_n^\#(t)\boldsymbol{\Sigma}) - \|\mathbf{S}_n^\#(t)\boldsymbol{\Sigma}\|_F^2\text{tr}(\boldsymbol{\Sigma}\boldsymbol{\Pi}_0)]^2}{\|\mathbf{S}_n^\#(t)\boldsymbol{\Sigma}\|_F^2\left[\|\mathbf{S}_n^\#(t)\boldsymbol{\Sigma}\|_F^2\|\boldsymbol{\Pi}_0\boldsymbol{\Sigma}\|_F^2 - \left(\text{tr}(\mathbf{S}_n^\#(t)\boldsymbol{\Sigma}^2\boldsymbol{\Pi}_0)\right)^2\right]} \\ &= \frac{[\text{tr}(\mathbf{S}_n^\#(t)\boldsymbol{\Sigma})]^2\|\boldsymbol{\Pi}_0\boldsymbol{\Sigma}\|_F^2 - 2\text{tr}(\mathbf{S}_n^\#(t)\boldsymbol{\Sigma}^2\boldsymbol{\Pi}_0)\text{tr}(\mathbf{S}_n^\#(t)\boldsymbol{\Sigma})\text{tr}(\boldsymbol{\Sigma}\boldsymbol{\Pi}_0) + \|\mathbf{S}_n^\#(t)\boldsymbol{\Sigma}\|_F^2[\text{tr}(\boldsymbol{\Sigma}\boldsymbol{\Pi}_0)]^2}{\|\mathbf{S}_n^\#(t)\boldsymbol{\Sigma}\|_F^2\|\boldsymbol{\Pi}_0\boldsymbol{\Sigma}\|_F^2 - \left(\text{tr}(\mathbf{S}_n^\#(t)\boldsymbol{\Sigma}^2\boldsymbol{\Pi}_0)\right)^2} \\ &= L_{F;n,2}^2(\mathbf{S}_n^\#(t)) + \left[\text{tr}\left(\frac{\boldsymbol{\Sigma}\boldsymbol{\Pi}_0}{\|\boldsymbol{\Sigma}\boldsymbol{\Pi}_0\|_F}\right)\right]^2 \end{aligned}$$

with

$$(S.14) \quad L_{F;n,2}^2(\mathbf{S}_n^\#(t)) = \frac{\left[\text{tr}(\mathbf{S}_n^\#(t)\boldsymbol{\Sigma}) - \text{tr}\left(\mathbf{S}_n^\#(t)\frac{\boldsymbol{\Sigma}^2\boldsymbol{\Pi}_0}{\|\boldsymbol{\Sigma}\boldsymbol{\Pi}_0\|_F}\right)\text{tr}\left(\frac{\boldsymbol{\Sigma}\boldsymbol{\Pi}_0}{\|\boldsymbol{\Sigma}\boldsymbol{\Pi}_0\|_F}\right)\right]^2}{\|\mathbf{S}_n^\#(t)\boldsymbol{\Sigma}\|_F^2 - \left[\text{tr}\left(\mathbf{S}_n^\#(t)\frac{\boldsymbol{\Sigma}^2\boldsymbol{\Pi}_0}{\|\boldsymbol{\Sigma}\boldsymbol{\Pi}_0\|_F}\right)\right]^2}.$$

Then the optimal value of  $t$  is found by maximizing  $L_{F;n,2}^2(\mathbf{S}_n^\#(t))$ . This finishes the proof of the theorem.  $\square$

### S.5.13. Proof of Theorem 3.4.

PROOF. The out-of-sample variance, or the loss function  $L(\alpha, t) = \widehat{\mathbf{w}}_{GSE}^\top \boldsymbol{\Sigma} \widehat{\mathbf{w}}_{GSE}$ , can be simplified to

$$\begin{aligned} L(\alpha, t) &= \left(\alpha \mathbf{w}_{\mathbf{S}_n^\#(t)} + (1 - \alpha) \mathbf{b}\right)^\top \boldsymbol{\Sigma} \left(\alpha \mathbf{w}_{\mathbf{S}_n^\#(t)} + (1 - \alpha) \mathbf{b}\right) \\ &= \left(\mathbf{b} - \mathbf{w}_{\mathbf{S}_n^\#(t)}\right)^\top \boldsymbol{\Sigma} \left(\mathbf{b} - \mathbf{w}_{\mathbf{S}_n^\#(t)}\right) \left(\alpha - \frac{\mathbf{b}^\top \boldsymbol{\Sigma} \left(\mathbf{b} - \mathbf{w}_{\mathbf{S}_n^\#(t)}\right)}{\left(\mathbf{b} - \mathbf{w}_{\mathbf{S}_n^\#(t)}\right)^\top \boldsymbol{\Sigma} \left(\mathbf{b} - \mathbf{w}_{\mathbf{S}_n^\#(t)}\right)}\right)^2 \end{aligned}$$

$$(S.15) \quad - \frac{\left(\mathbf{b}^\top \boldsymbol{\Sigma} \left(\mathbf{b} - \mathbf{w}_{\mathbf{S}_n^\#(t)}\right)\right)^2}{\left(\mathbf{b} - \mathbf{w}_{\mathbf{S}_n^\#(t)}\right)^\top \boldsymbol{\Sigma} \left(\mathbf{b} - \mathbf{w}_{\mathbf{S}_n^\#(t)}\right)} + \mathbf{b}^\top \boldsymbol{\Sigma} \mathbf{b}.$$

Now, the result of Theorem 3.4 follows immediately.  $\square$

**S.6. Weighted moments of the ordinary inverse of the sample covariance matrix.** In this section, we provide the results for the weighted trace moments of the ordinary inverse of the sample covariance matrix when  $p < n$ . The derivation is based on the asymptotic properties of the ridge-type estimator of the inverse covariance matrix expressed as

$$\mathbf{S}_n^-(t) = (\mathbf{S}_n + t\mathbf{I}_p)^{-1} \quad \text{for } t \geq 0,$$

by considering the special case of  $t = 0$  when  $c_n = p/n < 1$ . In this case, the sample covariance matrix  $\mathbf{S}_n$  is non-singular and its ordinary inverse is a well-defined matrix. However, the results of Theorem 2.2 cannot be directly used when  $p < n$  and  $t = 0$ . An alternative representation of the asymptotic behavior of  $\text{tr}((\mathbf{S}_n^-(t))^{m+1}\boldsymbol{\Theta})$  is deduced in Theorem S.6.1.

**THEOREM S.6.1.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A3), it holds that*

$$(S.16) \quad \left| \text{tr}((\mathbf{S}_n^-(t))^{m+1}\boldsymbol{\Theta}) - \tilde{r}_{m+1}(t, \boldsymbol{\Theta}) \right| \xrightarrow{a.s.} 0 \quad \text{for } p/n \rightarrow c \in (0, 1) \quad \text{as } n \rightarrow \infty$$

where

$$(S.17) \quad \tilde{r}_1(t, \boldsymbol{\Theta}) = \frac{1}{w(t)} \text{tr} \left[ (\tilde{w}(t)\mathbf{I}_p + \boldsymbol{\Sigma})^{-1} \boldsymbol{\Theta} \right],$$

$$(S.18) \quad \begin{aligned} \tilde{r}_{m+1}(t, \boldsymbol{\Theta}) &= \frac{(-1)^m}{m!} \tilde{d}_0(t, \boldsymbol{\Theta}) \sum_{k=1}^m \frac{(-1)^k k!}{w(t)^{k+1}} B_{m,k} \left( w^{(1)}(t), \dots, w^{(m-k+1)}(t) \right) \\ &\quad + \frac{(-1)^m}{m!} \frac{1}{w(t)} \sum_{k=1}^m (-1)^k k! \tilde{d}_k(t, \boldsymbol{\Theta}) B_{m,k} \left( \tilde{w}^{(1)}(t), \dots, \tilde{w}^{(m-k+1)}(t) \right) \\ &\quad + \frac{(-1)^m}{m!} \sum_{l=1}^{m-1} \frac{m!}{l!(m-l)!} \sum_{k_1=1}^{m-l} \frac{(-1)^{k_1} k_1!}{w(t)^{k_1+1}} B_{m-l, k_1} \left( w^{(1)}(t), \dots, w^{(m-l-k_1+1)}(t) \right) \\ &\quad \times \sum_{k_2=1}^l (-1)^{k_2} k_2! \tilde{d}_{k_2}(t, \boldsymbol{\Theta}) B_{l, k_2} \left( \tilde{w}^{(1)}(t), \dots, \tilde{w}^{(l-k_2+1)}(t) \right), \end{aligned}$$

with

$$(S.19) \quad \tilde{d}_k(t, \boldsymbol{\Theta}) = \text{tr} \left[ (\tilde{w}(t)\mathbf{I}_p + \boldsymbol{\Sigma})^{-(k+1)} \boldsymbol{\Theta} \right], \quad k = 0, 1, 2, \dots,$$

$$(S.20) \quad \tilde{w}(t) = \frac{t}{w(t)}, \quad \tilde{w}^{(1)}(t) = \frac{1}{w(t)} - t \frac{w^{(1)}(t)}{[w(t)]^2},$$

$$(S.21) \quad \begin{aligned} \tilde{w}^{(m)}(t) &= t \sum_{k=1}^m \frac{(-1)^k k!}{w(t)^{k+1}} B_{m,k} \left( w^{(1)}(t), \dots, w^{(m-k+1)}(t) \right) \\ &\quad + m \sum_{k=1}^{m-1} \frac{(-1)^k k!}{w(t)^{k+1}} B_{m-1, k} \left( w^{(1)}(t), \dots, w^{(m-k)}(t) \right), \end{aligned}$$

for  $m = 2, 3, \dots$  where  $w(t)$  is the unique solution of

$$(S.22) \quad 1 - w(t) = c_n \frac{1}{p} \text{tr} \left[ (\tilde{w}(t) \mathbf{I}_p + \boldsymbol{\Sigma})^{-1} \boldsymbol{\Sigma} \right]$$

with its derivatives satisfy

$$(S.23) \quad w^{(1)}(t) = \frac{c_n \tilde{d}_1 \left( t, \frac{1}{p} \boldsymbol{\Sigma} \right) w(t)}{w(t)^2 + t c_n \tilde{d}_1 \left( t, \frac{1}{p} \boldsymbol{\Sigma} \right)},$$

and  $w^{(m)}(0)$  for  $m = 2, 3, \dots$  is computed recursively by

$$(S.24) \quad w^{(m)}(t) = \frac{w(t)^2 c_n}{w(t)^2 + t c_n \tilde{d}_1 \left( t, \frac{1}{p} \boldsymbol{\Sigma} \right)} \left( \tilde{d}_1 \left( t, \frac{1}{p} \boldsymbol{\Sigma} \right) t \sum_{k=2}^m \frac{(-1)^k k!}{w(t)^{k+1}} B_{m,k} \left( w^{(1)}(t), \dots, w^{(m-k+1)}(t) \right) \right. \\ \left. + \tilde{d}_1 \left( t, \frac{1}{p} \boldsymbol{\Sigma} \right) m \sum_{k=1}^{m-1} \frac{(-1)^k k!}{w(t)^{k+1}} B_{m-1,k} \left( w^{(1)}(t), \dots, w^{(m-k)}(t) \right) \right. \\ \left. - \sum_{k=2}^m (-1)^k k! \tilde{d}_k \left( t, \frac{1}{p} \boldsymbol{\Sigma} \right) B_{m,k} \left( \tilde{w}^{(1)}(t), \dots, \tilde{w}^{(m-k+1)}(t) \right) \right).$$

PROOF OF THEOREM S.6.1: Following the proof of Theorem 2.1, we get

$$\text{tr}((\mathbf{S}_n + t \mathbf{I}_p)^{-(m+1)} \boldsymbol{\Theta}) = \frac{(-1)^m}{m!} \frac{\partial^m}{\partial t^m} \text{tr} \left[ \left( \frac{1}{n} \mathbf{X}_n \mathbf{X}_n^\top + t \boldsymbol{\Sigma}^{-1} \right)^{-1} \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\Theta} \boldsymbol{\Sigma}^{-1/2} \right],$$

where the application of Lemma 6 in Bodnar et al. (2024b) yields

$$\left| \text{tr} \left[ \left( \frac{1}{n} \mathbf{X}_n \mathbf{X}_n^\top + t \boldsymbol{\Sigma}^{-1} \right)^{-1} \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\Theta} \boldsymbol{\Sigma}^{-1/2} \right] - \frac{1}{w(t)} \text{tr} \left[ \left( \frac{t}{w(t)} \mathbf{I}_p + \boldsymbol{\Sigma} \right)^{-1} \boldsymbol{\Theta} \right] \right| \xrightarrow{a.s.} 0,$$

for  $p/n \rightarrow c \in (0, 1)$  as  $n \rightarrow \infty$  with  $w(t)$  being the solution of (S.22).

Let  $\tilde{w}(t) = t/w(t)$ . Then, it holds that

$$\text{tr}((\mathbf{S}_n + t \mathbf{I}_p)^{-(m+1)} \boldsymbol{\Theta}) \xrightarrow{d.a.s.} \frac{(-1)^m}{m!} \frac{\partial^m}{\partial t^m} \frac{1}{w(t)} \text{tr} \left[ (\tilde{w}(t) \mathbf{I}_p + \boldsymbol{\Sigma})^{-1} \boldsymbol{\Theta} \right] \\ = \frac{(-1)^m}{m!} \frac{\partial^m}{\partial t^m} \left( \frac{1}{w(t)} \right) \text{tr} \left[ (\tilde{w}(t) \mathbf{I}_p + \boldsymbol{\Sigma})^{-1} \boldsymbol{\Theta} \right] + \frac{(-1)^m}{m!} \frac{1}{w(t)} \frac{\partial^m}{\partial t^m} \left( \text{tr} \left[ (\tilde{w}(t) \mathbf{I}_p + \boldsymbol{\Sigma})^{-1} \boldsymbol{\Theta} \right] \right) \\ + \frac{(-1)^m}{m!} \sum_{l=1}^{m-1} \frac{m!}{l!(m-l)!} \frac{\partial^{m-l}}{\partial t^{m-l}} \frac{1}{w(t)} \frac{\partial^l}{\partial t^l} \text{tr} \left[ (\tilde{w}(t) \mathbf{I}_p + \boldsymbol{\Sigma})^{-1} \boldsymbol{\Theta} \right] \\ = \frac{(-1)^m}{m!} \text{tr} \left[ (\tilde{w}(t) \mathbf{I}_p + \boldsymbol{\Sigma})^{-1} \boldsymbol{\Theta} \right] \sum_{k=1}^m \frac{(-1)^k k!}{w(t)^{k+1}} B_{m,k} \left( w^{(1)}(t), \dots, w^{(m-k+1)}(t) \right) \\ + \frac{(-1)^m}{m!} \frac{1}{w(t)} \sum_{k=1}^m (-1)^k k! \text{tr} \left[ (\tilde{w}(t) \mathbf{I}_p + \boldsymbol{\Sigma})^{-(k+1)} \boldsymbol{\Theta} \right] B_{m,k} \left( \tilde{w}^{(1)}(t), \dots, \tilde{w}^{(m-k+1)}(t) \right) \\ + \frac{(-1)^m}{m!} \sum_{l=1}^{m-1} \frac{m!}{l!(m-l)!} \sum_{k_1=1}^{m-l} \frac{(-1)^{k_1} k_1!}{w(t)^{k_1+1}} B_{m-l,k_1} \left( w^{(1)}(t), \dots, w^{(m-l-k_1+1)}(t) \right)$$

$$\times \sum_{k_2=1}^l (-1)^{k_2} k_2! \operatorname{tr} \left[ (\tilde{w}(t) \mathbf{I}_p + \boldsymbol{\Sigma})^{-(k_2+1)} \boldsymbol{\Theta} \right] B_{l, k_2} \left( \tilde{w}^{(1)}(t), \dots, \tilde{w}^{(l-k_2+1)}(t) \right),$$

where

$$\begin{aligned} \tilde{w}^{(1)}(t) &= \frac{1}{w(t)} - t \frac{w^{(1)}(t)}{[w(t)]^2}, \\ \tilde{w}^{(m)}(t) &= t \frac{\partial^m}{\partial t^m} \left( \frac{1}{w(t)} \right) + m \frac{\partial^{m-1}}{\partial t^{m-1}} \left( \frac{1}{w(t)} \right) \\ &= t \sum_{k=1}^m \frac{(-1)^k k!}{w(t)^{k+1}} B_{m, k} \left( w^{(1)}(t), \dots, w^{(m-k+1)}(t) \right) + m \sum_{k=1}^{m-1} \frac{(-1)^k k!}{w(t)^{k+1}} B_{m-1, k} \left( w^{(1)}(t), \dots, w^{(m-k)}(t) \right). \end{aligned}$$

Finally, (S.22) yields

$$\begin{aligned} -w^{(1)}(t) &= -c_n \frac{1}{p} \operatorname{tr} \left[ (\tilde{w}(t) \mathbf{I}_p + \boldsymbol{\Sigma})^{-2} \boldsymbol{\Sigma} \right] \left( \frac{1}{w(t)} - t \frac{w^{(1)}(t)}{[w(t)]^2} \right), \\ -w^{(m)}(t) &= c_n \sum_{k=1}^m \frac{(-1)^k k!}{p} \operatorname{tr} \left[ (\tilde{w}(t) \mathbf{I}_p + \boldsymbol{\Sigma})^{-(k+1)} \boldsymbol{\Sigma} \right] B_{m, k} \left( \tilde{w}^{(1)}(t), \dots, \tilde{w}^{(m-k+1)}(t) \right) \\ &= c_n \sum_{k=2}^m \frac{(-1)^k k!}{p} \operatorname{tr} \left[ (\tilde{w}(t) \mathbf{I}_p + \boldsymbol{\Sigma})^{-(k+1)} \boldsymbol{\Sigma} \right] B_{m, k} \left( \tilde{w}^{(1)}(t), \dots, \tilde{w}^{(m-k+1)}(t) \right) \\ &\quad - c_n \frac{1}{p} \operatorname{tr} \left[ (\tilde{w}(t) \mathbf{I}_p + \boldsymbol{\Sigma})^{-2} \boldsymbol{\Sigma} \right] \tilde{w}^{(m)}(t) \\ &= c_n \sum_{k=2}^m \frac{(-1)^k k!}{p} \operatorname{tr} \left[ (\tilde{w}(t) \mathbf{I}_p + \boldsymbol{\Sigma})^{-(k+1)} \boldsymbol{\Sigma} \right] B_{m, k} \left( \tilde{w}^{(1)}(t), \dots, \tilde{w}^{(m-k+1)}(t) \right) \\ &\quad - c_n \frac{1}{p} \operatorname{tr} \left[ (\tilde{w}(t) \mathbf{I}_p + \boldsymbol{\Sigma})^{-2} \boldsymbol{\Sigma} \right] t \sum_{k=2}^m \frac{(-1)^k k!}{w(t)^{k+1}} B_{m, k} \left( w^{(1)}(t), \dots, w^{(m-k+1)}(t) \right) \\ &\quad - c_n \frac{1}{p} \operatorname{tr} \left[ (\tilde{w}(t) \mathbf{I}_p + \boldsymbol{\Sigma})^{-2} \boldsymbol{\Sigma} \right] m \sum_{k=1}^{m-1} \frac{(-1)^k k!}{w(t)^{k+1}} B_{m-1, k} \left( w^{(1)}(t), \dots, w^{(m-k)}(t) \right) \\ &\quad + c_n \frac{1}{p} \operatorname{tr} \left[ (\tilde{w}(t) \mathbf{I}_p + \boldsymbol{\Sigma})^{-2} \boldsymbol{\Sigma} \right] \frac{t}{w(t)^2} w^{(m)}(t), \end{aligned}$$

from which we get (S.20) and (S.21). The theorem is proved.  $\square$

In Corollary S.6.1, the results are presented for the special case  $t = 0$ , i.e., for the ordinary inverse of the sample covariance matrix.

**COROLLARY S.6.1.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A3), it holds that*

$$(S.25) \quad \left| \operatorname{tr}(\mathbf{S}_n^{-(m+1)} \boldsymbol{\Theta}) - \tilde{r}_{m+1}(\boldsymbol{\Theta}) \right| \xrightarrow{a.s.} 0 \quad \text{for } p/n \rightarrow c \in (0, 1) \quad \text{as } n \rightarrow \infty$$

where

$$(S.26) \quad \tilde{r}_1(\boldsymbol{\Theta}) = \frac{1}{1 - c_n} \operatorname{tr} \left[ \boldsymbol{\Sigma}^{-1} \boldsymbol{\Theta} \right],$$

$$(S.27) \quad \tilde{r}_{m+1}(\Theta) = \frac{(-1)^{m+1}}{(m+1)!} \sum_{k=1}^{m+1} (-1)^k k! \tilde{d}_{k-1}(\Theta) B_{m+1,k} \left( \tilde{w}^{(1)}(0), \dots, \tilde{w}^{(m+1-k+1)}(0) \right),$$

with

$$(S.28) \quad \tilde{d}_k(\Theta) = \text{tr} \left[ \Sigma^{-(k+1)} \Theta \right], \quad k = 0, 1, 2, \dots,$$

$$(S.29) \quad \tilde{w}^{(1)}(0) = \frac{1}{1-c_n}, \quad \tilde{w}^{(m)}(0) = m \sum_{k=1}^{m-1} \frac{(-1)^k k!}{(1-c_n)^{k+1}} B_{m-1,k} \left( w^{(1)}(0), \dots, w^{(m-k)}(0) \right),$$

for  $m = 2, 3, \dots$  where  $w(0) = 1 - c_n$ ,

$$(S.30) \quad w^{(1)}(0) = \frac{c_n}{1-c_n} \tilde{d}_1 \left( \frac{1}{p} \Sigma \right),$$

and  $w^{(m)}(0)$  for  $m = 2, 3, \dots$  is computed recursively by

$$(S.31) \quad w^{(m)}(0) = -c_n \sum_{k=1}^m (-1)^k k! \tilde{d}_k \left( \frac{1}{p} \Sigma \right) B_{m,k} \left( \tilde{w}^{(1)}(0), \dots, \tilde{w}^{(m-k+1)}(0) \right).$$

**PROOF OF COROLLARY S.6.1:** The statement of the corollary follows from Theorem S.6.1 by noting that  $w(0) = 1 - c_n$ ,  $\tilde{w}(0) = 0$ ,  $w^{(1)}(0) = \frac{c_n}{1-c_n} \tilde{d}_1 \left( \frac{1}{p} \Sigma \right)$ ,  $\tilde{w}^{(1)}(0) = \frac{1}{1-c_n}$ ,

$$\tilde{w}^{(m)}(0) = m \sum_{k=1}^{m-1} \frac{(-1)^k k!}{(1-c_n)^{k+1}} B_{m-1,k} \left( w^{(1)}(0), \dots, w^{(m-k)}(0) \right),$$

and

$$\begin{aligned} w^{(m)}(0) &= c_n \left( \tilde{d}_1 \left( \frac{1}{p} \Sigma \right) m \sum_{k=1}^{m-1} \frac{(-1)^k k!}{(1-c_n)^{k+1}} B_{m-1,k} \left( w^{(1)}(0), \dots, w^{(m-k)}(0) \right) \right. \\ &\quad \left. - \sum_{k=2}^m (-1)^k k! \tilde{d}_k \left( \frac{1}{p} \Sigma \right) B_{m,k} \left( \tilde{w}^{(1)}(0), \dots, \tilde{w}^{(m-k+1)}(0) \right) \right) \\ &= -c_n \left( -\tilde{d}_k \left( \frac{1}{p} \Sigma \right) \tilde{w}^{(m)}(0) + \sum_{k=2}^m (-1)^k k! \tilde{d}_k \left( \frac{1}{p} \Sigma \right) B_{m,k} \left( \tilde{w}^{(1)}(0), \dots, \tilde{w}^{(m-k+1)}(0) \right) \right) \\ &= -c_n \sum_{k=1}^m (-1)^k k! \tilde{d}_k \left( \frac{1}{p} \Sigma \right) B_{m,k} \left( \tilde{w}^{(1)}(0), \dots, \tilde{w}^{(m-k+1)}(0) \right), \end{aligned}$$

where the equality  $\tilde{w}^{(m)}(0) = B_{m,1} \left( \tilde{w}^{(1)}(0), \dots, \tilde{w}^{(m)}(0) \right)$  is used.

Similarly, we get

$$\begin{aligned} \tilde{r}_{m+1}(\Theta) &= \frac{(-1)^m}{m!} \tilde{d}_0(\Theta) \sum_{k=1}^m \frac{(-1)^k k!}{(1-c_n)^{k+1}} B_{m,k} \left( w^{(1)}(0), \dots, w^{(m-k+1)}(0) \right) \\ &\quad + \frac{(-1)^m}{m!} \frac{1}{1-c_n} \sum_{k=1}^m (-1)^k k! \tilde{d}_k(\Theta) B_{m,k} \left( \tilde{w}^{(1)}(0), \dots, \tilde{w}^{(m-k+1)}(0) \right) \\ &\quad + \frac{(-1)^m}{m!} \sum_{l=1}^{m-1} \frac{m!}{l!(m-l)!} \sum_{k_1=1}^{m-l} \frac{(-1)^{k_1} k_1!}{(1-c_n)^{k_1+1}} B_{m-l,k_1} \left( w^{(1)}(0), \dots, w^{(m-l-k_1+1)}(0) \right) \end{aligned}$$

$$\begin{aligned}
& \times \sum_{k_2=1}^l (-1)^{k_2} k_2! \tilde{d}_{k_2}(\Theta) B_{l,k_2} \left( \tilde{w}^{(1)}(0), \dots, \tilde{w}^{(l-k_2+1)}(0) \right) \\
& = \frac{(-1)^m}{m!} \tilde{d}_0(\Theta) \frac{\tilde{w}^{(m+1)}(0)}{m+1} + \frac{(-1)^m}{m!} \frac{1}{1-c_n} \sum_{k=1}^m (-1)^k k! \tilde{d}_k(\Theta) B_{m,k} \left( \tilde{w}^{(1)}(0), \dots, \tilde{w}^{(m-k+1)}(0) \right) \\
& + \frac{(-1)^m}{m!} \sum_{l=1}^{m-1} \frac{m!}{l!(m-l)!} \frac{\tilde{w}^{(m-l+1)}(0)}{m-l+1} \sum_{k_2=1}^l (-1)^{k_2} k_2! \tilde{d}_{k_2}(\Theta) B_{l,k_2} \left( \tilde{w}^{(1)}(0), \dots, \tilde{w}^{(l-k_2+1)}(0) \right) \\
& = \frac{(-1)^m \tilde{w}^{(m+1)}(0)}{(m+1)!} \tilde{d}_0(\Theta) \\
& + \frac{(-1)^m}{(m+1)!} \sum_{l=1}^m \frac{(m+1)!}{l!(m+1-l)!} \tilde{w}^{(m-l+1)}(0) \sum_{k_2=1}^l (-1)^{k_2} k_2! \tilde{d}_{k_2}(\Theta) B_{l,k_2} \left( \tilde{w}^{(1)}(0), \dots, \tilde{w}^{(l-k_2+1)}(0) \right) \\
& = \frac{(-1)^m \tilde{w}^{(m+1)}(0)}{(m+1)!} \tilde{d}_0(\Theta) \\
& + \frac{(-1)^m}{(m+1)!} \sum_{k_2=1}^m (-1)^{k_2} \tilde{d}_{k_2}(\Theta) k_2! \sum_{l=k_2}^m \frac{(m+1)!}{l!(m+1-l)!} \tilde{w}^{(m-l+1)}(0) B_{l,k_2} \left( \tilde{w}^{(1)}(0), \dots, \tilde{w}^{(l-k_2+1)}(0) \right).
\end{aligned}$$

From Equation (1.4) in Cvijović (2011) and the fact that  $B_{i,k_2}(\cdot) = 0$  for  $i < k_2$ , we get

$$\begin{aligned}
& k_2! \sum_{l=k_2}^m \frac{(m+1)!}{l!(m+1-l)!} \tilde{w}^{(m-l+1)}(0) B_{l,k_2} \left( \tilde{w}^{(1)}(0), \dots, \tilde{w}^{(l-k_2+1)}(0) \right) \\
& = k_2! \sum_{r=1}^{m+1-k_2} \frac{(m+1)!}{r!(m+1-r)!} \tilde{w}^{(r)}(0) B_{m+1-r,k_2} \left( \tilde{w}^{(1)}(0), \dots, \tilde{w}^{(m+1-r-k_2+1)}(0) \right) \\
& = (k_2+1)! B_{m+1,k_2+1} \left( \tilde{w}^{(1)}(0), \dots, \tilde{w}^{(m-k_2+1)}(0) \right).
\end{aligned}$$

Hence,

$$\begin{aligned}
\tilde{r}_{m+1}(\Theta) & = \frac{(-1)^m \tilde{w}^{(m+1)}(0)}{(m+1)!} \tilde{d}_0(\Theta) \\
& - \frac{(-1)^m}{(m+1)!} \sum_{k_2=1}^m (-1)^{k_2+1} \tilde{d}_{k_2}(\Theta) (k_2+1)! B_{m+1,k_2+1} \left( \tilde{w}^{(1)}(0), \dots, \tilde{w}^{(m-k_2+1)}(0) \right) \\
& = \frac{(-1)^{m+1}}{(m+1)!} \sum_{k_2=0}^m (-1)^{k_2+1} \tilde{d}_{k_2}(\Theta) (k_2+1)! B_{m+1,k_2+1} \left( \tilde{w}^{(1)}(0), \dots, \tilde{w}^{(m-k_2+1)}(0) \right) \\
& = \frac{(-1)^{m+1}}{(m+1)!} \sum_{k=1}^{m+1} (-1)^k k! \tilde{d}_{k-1}(\Theta) B_{m+1,k} \left( \tilde{w}^{(1)}(0), \dots, \tilde{w}^{(m+1-k+1)}(0) \right),
\end{aligned}$$

which completes the proof of the corollary.  $\square$

In Corollary S.6.2, the results for  $m = 0, 1, 2, 3$  are summarized.

COROLLARY S.6.2. *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A3), we get (S.6.1) with*

$$\begin{aligned}\tilde{r}_1(\Theta) &= \frac{1}{1-c_n} \text{tr} [\Sigma^{-1} \Theta], \\ \tilde{r}_2(\Theta) &= \frac{c_n}{(1-c_n)^3} \text{tr} [\Sigma^{-1} \Theta] \frac{1}{p} \text{tr} [\Sigma^{-1}] + \frac{1}{(1-c_n)^2} \text{tr} [\Sigma^{-2} \Theta], \\ \tilde{r}_3(\Theta) &= \frac{\tilde{w}^{(3)}(0)}{6} \text{tr} [\Sigma^{-1} \Theta] - \frac{\tilde{w}^{(2)}(0)}{1-c_n} \text{tr} [\Sigma^{-2} \Theta] + \frac{1}{(1-c_n)^3} \text{tr} [\Sigma^{-3} \Theta], \\ \tilde{r}_4(\Theta) &= -\frac{\tilde{w}^{(4)}(0)}{24} \text{tr} [\Sigma^{-1} \Theta] + \left( \frac{\tilde{w}^{(3)}(0)}{3(1-c_n)} + \frac{[\tilde{w}^{(2)}(0)]^2}{4} \right) \text{tr} [\Sigma^{-2} \Theta] \\ &\quad - \frac{3\tilde{w}^{(2)}(0)}{2(1-c_n)^2} \text{tr} [\Sigma^{-3} \Theta] + \frac{1}{(1-c_n)^4} \text{tr} [\Sigma^{-4} \Theta],\end{aligned}$$

where

$$\begin{aligned}\tilde{w}^{(2)}(0) &= -\frac{2}{(1-c_n)^2} w^{(1)}(0), \\ \tilde{w}^{(3)}(0) &= -3 \left( \frac{1}{(1-c_n)^2} w^{(2)}(0) - \frac{2}{(1-c_n)^3} [w^{(1)}(0)]^2 \right), \\ \tilde{w}^{(4)}(0) &= -4 \left( \frac{1}{(1-c_n)^2} w^{(3)}(0) - \frac{6}{(1-c_n)^3} w^{(1)}(0) w^{(2)}(0) + \frac{6}{(1-c_n)^4} [w^{(1)}(0)]^3 \right),\end{aligned}$$

with

$$\begin{aligned}w^{(1)}(0) &= \frac{c_n}{1-c_n} \frac{1}{p} \text{tr} [\Sigma^{-1}], \\ w^{(2)}(0) &= c_n \left( \tilde{w}^{(2)}(0) \frac{1}{p} \text{tr} [\Sigma^{-1}] - 2 \frac{1}{(1-c_n)^2} \frac{1}{p} \text{tr} [\Sigma^{-2}] \right), \\ w^{(3)}(0) &= c_n \left( \tilde{w}^{(3)}(0) \frac{1}{p} \text{tr} [\Sigma^{-1}] - 6 \frac{\tilde{w}^{(2)}(0)}{1-c_n} \frac{1}{p} \text{tr} [\Sigma^{-2}] + 6 \frac{1}{(1-c_n)^3} \frac{1}{p} \text{tr} [\Sigma^{-3}] \right).\end{aligned}$$

PROOF OF COROLLARY S.6.2: The expression for  $m = 0$  is already present in (S.26) of Corollary S.6.1. The application of (S.27) for  $m = 1, 2, 3$  yields

$$\begin{aligned}\tilde{r}_2(\Theta) &= -\frac{1}{2} \tilde{d}_0(\Theta) \tilde{w}^{(2)}(0) + \tilde{d}_1(\Theta) [\tilde{w}^{(1)}(0)]^2 \\ \tilde{r}_3(\Theta) &= \frac{1}{6} \tilde{d}_0(\Theta) \tilde{w}^{(3)}(0) - \tilde{d}_1(\Theta) \tilde{w}^{(1)}(0) \tilde{w}^{(2)}(0) + \tilde{d}_2(\Theta) [\tilde{w}^{(1)}(0)]^3 \\ \tilde{r}_4(\Theta) &= -\frac{1}{24} \tilde{d}_0(\Theta) \tilde{w}^{(4)}(0) + \frac{1}{12} \tilde{d}_1(\Theta) (4\tilde{w}^{(1)}(0) \tilde{w}^{(3)}(0) + 3[\tilde{w}^{(2)}(0)]^2) \\ &\quad - \frac{3}{2} \tilde{d}_2(\Theta) [\tilde{w}^{(1)}(0)]^2 \tilde{w}^{(2)}(0) + \tilde{d}_3(\Theta) [\tilde{w}^{(1)}(0)]^4,\end{aligned}$$

where  $\tilde{w}^{(1)}(0) = \frac{1}{1-c_n}$  and  $\tilde{w}^{(2)}(0)$ ,  $\tilde{w}^{(3)}(0)$ , and  $\tilde{w}^{(4)}(0)$  are provided in the statement of the corollary.  $\square$

The formulas for  $\tilde{r}_1(\Theta)$  and  $\tilde{r}_2(\Theta)$  in Corollary S.6.2 coincide with the previous findings derived in Bodnar et al. (2016b). In Corollary S.6.3, we present the results derived for  $\Theta = \frac{1}{p}\mathbf{I}_p$ .

**COROLLARY S.6.3.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1) and let  $\Theta = \frac{1}{p}\mathbf{I}_p$ . Then, under Assumptions (A1)-(A2), it holds that*

$$\left| \frac{1}{p} \text{tr} \left[ \mathbf{S}_n^{-(m+1)} \right] - \frac{(-1)^m w^{(m+1)}(0)}{(m+1)! c_n} \right| \xrightarrow{\text{a.s.}} 0 \quad \text{for } p/n \rightarrow c \in (0, 1) \quad \text{as } n \rightarrow \infty,$$

where  $w^{(m+1)}(0)$ ,  $m = 0, 1, \dots$ , are defined in (S.31).

**PROOF OF COROLLARY S.6.3:** For  $m = 0$ , we get from Corollary S.6.1 that

$$\tilde{r}_1 \left( \frac{1}{p} \mathbf{I}_p \right) = \frac{1}{1 - c_n} \tilde{d}_0 \left( \frac{1}{p} \mathbf{I}_p \right) = \frac{1}{c_n} \frac{c_n}{1 - c_n} \tilde{d}_1 \left( \frac{1}{p} \Sigma \right) = \frac{w^{(1)}(0)}{c_n}.$$

The application of  $\tilde{d}_k \left( \frac{1}{p} \mathbf{I}_p \right) = \tilde{d}_{k+1} \left( \frac{1}{p} \Sigma \right)$  for  $m = 1, 2, \dots$  and Corollary S.6.1 yield

$$\begin{aligned} \tilde{r}_{m+1} \left( \frac{1}{p} \mathbf{I}_p \right) &= \frac{(-1)^{m+1}}{(m+1)!} \sum_{k=1}^{m+1} (-1)^k k! \tilde{d}_k \left( \frac{1}{p} \Sigma \right) B_{m+1,k} \left( \tilde{w}^{(1)}(0), \dots, \tilde{w}^{(m+1-k+1)}(0) \right) \\ &= \frac{(-1)^m}{(m+1)! c_n} w^{(m+1)}(0), \end{aligned}$$

where the second equality follows from (S.31).  $\square$

**COROLLARY S.6.4.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1) with  $\Sigma = \mathbf{I}_p$ . Then, under Assumptions (A2)-(A3), it holds that*

$$(S.32) \quad \left| \text{tr}(\mathbf{S}_n^{-(m+1)} \Theta) - \frac{(-1)^m w^{(m+1)}(0)}{(m+1)! c_n} \text{tr}(\Theta) \right| \xrightarrow{\text{a.s.}} 0$$

for  $p/n \rightarrow c \in (0, 1)$  as  $n \rightarrow \infty$ , where  $w^{(m+1)}(0)$ ,  $m = 0, 1, \dots$ , are given in (S.31) with  $\Sigma = \mathbf{I}_p$ .

**PROOF OF COROLLARY S.6.4:** The result follows from Corollary S.6.1 by noting that  $d_k(\Theta) = \text{tr}[\Theta]$  for  $k = 0, 1, \dots$  when  $\Sigma = \mathbf{I}_p$ .  $\square$

As a special case of Corollary S.6.4, we get

$$\begin{aligned} \left| \text{tr}(\mathbf{S}_n^{-1} \Theta) - \frac{1}{1 - c_n} \text{tr}(\Theta) \right| &\xrightarrow{\text{a.s.}} 0, & \left| \text{tr}(\mathbf{S}_n^{-2} \Theta) - \frac{1}{(1 - c_n)^3} \text{tr}(\Theta) \right| &\xrightarrow{\text{a.s.}} 0, \\ \left| \text{tr}(\mathbf{S}_n^{-3} \Theta) - \frac{1 + c_n}{(1 - c_n)^5} \text{tr}(\Theta) \right| &\xrightarrow{\text{a.s.}} 0, & \left| \text{tr}(\mathbf{S}_n^{-4} \Theta) - \frac{1 + 3c_n + c_n^2}{(1 - c_n)^7} \text{tr}(\Theta) \right| &\xrightarrow{\text{a.s.}} 0, \end{aligned}$$

for  $p/n \rightarrow c \in (0, 1)$  as  $n \rightarrow \infty$ .

**S.7. Weighted moments of the sample covariance matrix.** In this section, we present the auxiliary results for the weighted trace moments of the sample covariance matrix needed for the estimation of some quantities from Section 3. The results of this section are of the independent interest and are established in both cases  $p/n > 1$  and  $p/n < 1$ .

**THEOREM S.7.1.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A3), it holds that*

$$(S.33) \quad |\mathrm{tr}(\mathbf{S}_n^m \Theta) - \check{s}_m(\Theta)| \xrightarrow{a.s.} 0 \quad \text{for } p/n \rightarrow c \in (0, \infty) \quad \text{as } n \rightarrow \infty$$

where

$$(S.34) \quad \check{s}_m(\Theta) = \sum_{k=1}^m \frac{(-1)^{m+k} k!}{m!} \mathrm{tr} \left\{ \Sigma^k \Theta \right\} B_{m,k} \left( u^{(1)}(0), u^{(2)}(0), \dots, u^{(m-k+1)}(0) \right),$$

$u^{(1)}(0) = 1$  and  $u^{(2)}(0), \dots, u^{(m)}(0)$  are computed recursively by

$$(S.35) \quad u^{(m)}(0) = mc_n \sum_{k=1}^{m-1} \frac{(-1)^k k!}{p} \mathrm{tr} \left\{ \Sigma^k \right\} B_{m-1,k} \left( u^{(1)}(0), \dots, u^{(m-k)}(0) \right).$$

**PROOF OF THEOREM S.7.1:** Following the proof of Theorem 3.1 in Bodnar et al. (2014b), we get the following identity

$$\mathrm{tr}(\mathbf{S}_n^m \Theta) = \frac{(-1)^m}{m!} \frac{\partial^m}{\partial t^m} \mathrm{tr} \left[ (t\mathbf{S}_n + \mathbf{I}_p)^{-1} \Theta \right] \Big|_{t=0}$$

The proof of Theorem 2.1 and the application of Lemma 6 in Bodnar et al. (2024b) yields

$$\begin{aligned} \mathrm{tr}(\mathbf{S}_n^m \Theta) &= \frac{(-1)^m}{m!} \frac{\partial^m}{\partial t^m} \mathrm{tr} \left[ \left( \frac{t}{n} \mathbf{X}_n \mathbf{X}_n^\top + \Sigma^{-1} \right)^{-1} \Sigma^{-1/2} \Theta \Sigma^{-1/2} \right] \Big|_{t=0} \\ &\xrightarrow{a.s.} \frac{(-1)^m}{m!} \frac{\partial^m}{\partial t^m} \mathrm{tr} \left[ (u(t)\mathbf{I}_p + \Sigma^{-1})^{-1} \Sigma^{-1/2} \Theta \Sigma^{-1/2} \right] \Big|_{t=0}, \end{aligned}$$

for  $p/n \rightarrow c \in [0, \infty)$  as  $n \rightarrow \infty$  where  $u(t)$  solves the following equation

$$(S.36) \quad t - u(t) = c_n t - c_n t \frac{1}{p} \mathrm{tr} \left[ (u(t)\mathbf{I}_p + \Sigma^{-1})^{-1} \Sigma^{-1} \right].$$

From the proof of Theorem 2.1, we get

$$\begin{aligned} \mathrm{tr}(\mathbf{S}_n^m \Theta) &\xrightarrow{d.a.s.} \sum_{k=1}^m \frac{(-1)^{m+k} k!}{m!} \mathrm{tr} \left\{ (u(0)\Sigma + \mathbf{I}_p)^{-1} \left[ \Sigma (u(0)\Sigma + \mathbf{I}_p)^{-1} \right]^k \Theta \right\} \\ &\quad \times B_{m,k} \left( u^{(1)}(0), u^{(2)}(0), \dots, u^{(m-k+1)}(0) \right), \end{aligned}$$

for  $p/n \rightarrow c \in [0, \infty)$  as  $n \rightarrow \infty$ . Furthermore, the application of (S.36) yields  $u(0) = 0$  and  $u^{(1)}(0) = 1$ . Furthermore, it holds that

$$u^{(m)}(t) = c_n t \frac{\partial^m}{\partial t^m} \frac{1}{p} \mathrm{tr} \left[ (u(t)\mathbf{I}_p + \Sigma^{-1})^{-1} \Sigma^{-1} \right] + mc_n \frac{\partial^{m-1}}{\partial t^{m-1}} \frac{1}{p} \mathrm{tr} \left[ (u(t)\mathbf{I}_p + \Sigma^{-1})^{-1} \Sigma^{-1} \right]$$

and, hence,

$$u^{(m)}(0) = mc_n \sum_{k=1}^{m-1} \frac{(-1)^k k!}{p} \mathrm{tr} \left\{ \left[ \Sigma (u(0)\Sigma + \mathbf{I}_p)^{-1} \right]^{k+1} \Sigma^{-1} \right\} B_{m-1,k} \left( u^{(1)}(0), \dots, u^{(m-k)}(0) \right).$$

□

COROLLARY S.7.1. *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A3), we get (S.33) with*

$$\begin{aligned}\check{s}_1(\Theta) &= \text{tr}(\Sigma\Theta), \\ \check{s}_2(\Theta) &= c_n \frac{1}{p} \text{tr}(\Sigma) \text{tr}(\Sigma\Theta) + \text{tr}(\Sigma^2\Theta), \\ \check{s}_3(\Theta) &= c_n \left\{ c_n \left[ \frac{1}{p} \text{tr}(\Sigma) \right]^2 + \frac{1}{p} \text{tr}(\Sigma^2) \right\} \text{tr}(\Sigma\Theta) + 2c_n \frac{1}{p} \text{tr}(\Sigma) \text{tr}(\Sigma^2\Theta) + \text{tr}(\Sigma^3\Theta), \\ \check{s}_4(\Theta) &= c_n \left\{ c_n^2 \left[ \frac{1}{p} \text{tr}(\Sigma) \right]^3 + 3c_n \frac{1}{p} \text{tr}(\Sigma) \frac{1}{p} \text{tr}(\Sigma^2) + \frac{1}{p} \text{tr}(\Sigma^3) \right\} \text{tr}(\Sigma\Theta) \\ &\quad + c_n \left\{ 3c_n \left[ \frac{1}{p} \text{tr}(\Sigma) \right]^2 + 2 \frac{1}{p} \text{tr}(\Sigma^2) \right\} \text{tr}(\Sigma^2\Theta) + 3c_n \frac{1}{p} \text{tr}(\Sigma) \text{tr}(\Sigma^3\Theta) + \text{tr}(\Sigma^4\Theta),\end{aligned}$$

for  $p/n \rightarrow c \in [0, \infty)$  as  $n \rightarrow \infty$ .

PROOF OF COROLLARY S.7.1: The results of Theorem S.7.1 and the proof of Corollary 2.1 yield

$$\begin{aligned}\check{s}_1(\Theta) &= \text{tr}(\Sigma\Theta)B_{1,1} \left( u^{(1)}(0) \right) = \text{tr}(\Sigma\Theta)u^{(1)}(0), \\ \check{s}_2(\Theta) &= -\frac{1}{2} \text{tr}(\Sigma\Theta)B_{2,1} \left( u^{(1)}(0), u^{(2)}(0) \right) + \text{tr}(\Sigma^2\Theta)B_{2,2} \left( u^{(1)}(0) \right) \\ &= -\frac{1}{2} u^{(2)}(0) \text{tr}(\Sigma\Theta) + \text{tr}(\Sigma^2\Theta)[u^{(1)}(0)]^2, \\ \check{s}_3(\Theta) &= \frac{1}{6} \text{tr}(\Sigma\Theta)B_{3,1} \left( u^{(1)}(0), u^{(2)}(0), u^{(3)}(0) \right) - \frac{1}{3} \text{tr}(\Sigma^2\Theta)B_{3,2} \left( u^{(1)}(0), u^{(2)}(0) \right) + \text{tr}(\Sigma^3\Theta)B_{3,3} \left( u^{(1)}(0) \right) \\ &= \frac{1}{6} \text{tr}(\Sigma\Theta)u^{(3)}(0) - \text{tr}(\Sigma^2\Theta)u^{(1)}(0)u^{(2)}(0) + \text{tr}(\Sigma^3\Theta)[u^{(1)}(0)]^3, \\ \check{s}_4(\Theta) &= -\frac{1}{24} \text{tr}(\Sigma\Theta)B_{4,1} \left( u^{(1)}(0), u^{(2)}(0), u^{(3)}(0), u^{(4)}(0) \right) + \frac{1}{12} \text{tr}(\Sigma^2\Theta)B_{4,2} \left( u^{(1)}(0), u^{(2)}(0), u^{(3)}(0) \right) \\ &\quad - \frac{1}{4} \text{tr}(\Sigma^3\Theta)B_{4,3} \left( u^{(1)}(0), u^{(2)}(0) \right) + \text{tr}(\Sigma^4\Theta)B_{4,4} \left( u^{(1)}(0) \right) \\ &= -\frac{1}{24} \text{tr}(\Sigma\Theta)u^{(4)}(0) + \frac{1}{12} \text{tr}(\Sigma^2\Theta)[4u^{(1)}(0)u^{(3)}(0) + 3[u^{(2)}(0)]^2] \\ &\quad - \frac{3}{2} \text{tr}(\Sigma^3\Theta)[u^{(1)}(0)]^2 u^{(2)}(0) + \text{tr}(\Sigma^4\Theta)[u^{(1)}(0)]^4.\end{aligned}$$

Furthermore, we get from Theorem S.7.1 that

$$\begin{aligned}u^{(1)}(0) &= 1, \\ u^{(2)}(0) &= -2c_n \frac{1}{p} \text{tr}(\Sigma)B_{1,1} \left( u^{(1)}(0) \right) = -2c_n \frac{1}{p} \text{tr}(\Sigma), \\ u^{(3)}(0) &= 3c_n \left\{ -\frac{1}{p} \text{tr}(\Sigma)B_{2,1} \left( u^{(1)}(0), u^{(2)}(0) \right) + 2 \frac{1}{p} \text{tr}(\Sigma^2)B_{2,2} \left( u^{(1)}(0) \right) \right\} \\ &= 3c_n \left\{ -\frac{1}{p} \text{tr}(\Sigma)u^{(2)}(0) + 2 \frac{1}{p} \text{tr}(\Sigma^2)[u^{(1)}(0)]^2 \right\} = 6c_n \left\{ c_n \left[ \frac{1}{p} \text{tr}(\Sigma) \right]^2 + \frac{1}{p} \text{tr}(\Sigma^2) \right\},\end{aligned}$$

$$\begin{aligned}
 u^{(4)}(0) &= 4c_n \left\{ -\frac{1}{p} \text{tr}(\Sigma) B_{3,1} \left( u^{(1)}(0), u^{(2)}(0), u^{(3)}(0) \right) + 2\frac{1}{p} \text{tr}(\Sigma^2) B_{3,2} \left( u^{(1)}(0), u^{(2)}(0) \right) \right. \\
 &\quad \left. + 6\frac{1}{p} \text{tr}(\Sigma^3) B_{3,3} \left( u^{(1)}(0) \right) \right\} \\
 &= 4c_n \left\{ -\frac{1}{p} \text{tr}(\Sigma) u^{(3)}(0) + 6\frac{1}{p} \text{tr}(\Sigma^2) u^{(1)}(0) u^{(2)}(0) + 6\frac{1}{p} \text{tr}(\Sigma^3) [u^{(1)}(0)]^3 \right\} \\
 &= -24c_n \left\{ c_n^2 \left[ \frac{1}{p} \text{tr}(\Sigma) \right]^3 + 3c_n \frac{1}{p} \text{tr}(\Sigma) \frac{1}{p} \text{tr}(\Sigma^2) + \frac{1}{p} \text{tr}(\Sigma^3) \right\}.
 \end{aligned}$$

Substituting the expressions of  $u^{(1)}(0)$ ,  $u^{(2)}(0)$ ,  $u^{(3)}(0)$ , and  $u^{(4)}(0)$  into the formulas obtained for  $\check{s}_1, \dots, \check{s}_4$  leads to the statement of the corollary.  $\square$

In Corollary S.7.2 we present the results derived for  $\Theta = \frac{1}{p} \mathbf{I}_p$ .

**COROLLARY S.7.2.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A2), it holds for  $m = 1, 2, \dots$  that*

$$\left| \frac{1}{p} \text{tr}[\mathbf{S}_n^m] - \frac{(-1)^m u^{(m+1)}(0)}{(m+1)! c_n} \right| \xrightarrow{\text{a.s.}} 0 \quad \text{for } p/n \rightarrow c \in (0, \infty) \quad \text{as } n \rightarrow \infty,$$

where  $u^m(0)$ ,  $m = 1, \dots$ , are defined in (S.35).

**PROOF OF COROLLARY S.7.2:** From (S.35) we have that

$$\frac{u^{(m+1)}(0)}{(m+1)c_n} = \sum_{k=1}^m \frac{(-1)^k k!}{p} \text{tr}(\Sigma^k) B_{m,k} \left( u^{(1)}(0), \dots, u^{(m-k+1)}(0) \right),$$

which together with (S.34) computed for  $\Theta = \frac{1}{p} \mathbf{I}_p$  completes the proof of the theorem.  $\square$

**COROLLARY S.7.3.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1) with  $\Sigma = \mathbf{I}_p$ . Then, under Assumptions (A2)-(A3), it holds that*

$$(S.37) \quad \left| \text{tr}(\mathbf{S}_n^m \Theta) - \frac{(-1)^m u^{(m+1)}(0)}{(m+1)! c_n} \text{tr}\{\Theta\} \right| \xrightarrow{\text{a.s.}} 0$$

for  $p/n \rightarrow c \in (0, \infty)$  as  $n \rightarrow \infty$ , where  $u^{(m+1)}(0)$ ,  $m = 1, \dots$ , are given in (S.35) with  $\Sigma = \mathbf{I}_p$ .

**PROOF OF COROLLARY S.7.3:** If  $\Sigma = \mathbf{I}_p$ , then

$$\check{s}_m(\Theta) = \frac{(-1)^m \text{tr}\{\Theta\}}{m!} \sum_{k=1}^m (-1)^k k! B_{m,k} \left( u^{(1)}(0), u^{(2)}(0), \dots, u^{(m-k+1)}(0) \right),$$

which together with (S.35) completes the proof of the corollary.  $\square$

As a special case of Corollary S.7.3, we get

$$\begin{aligned}
 & \left| \text{tr}(\mathbf{S}_n \Theta) - \text{tr}(\Theta) \right| \xrightarrow{\text{a.s.}} 0, & \left| \text{tr}(\mathbf{S}_n^2 \Theta) - (c_n + 1) \text{tr}(\Theta) \right| \xrightarrow{\text{a.s.}} 0, \\
 & \left| \text{tr}(\mathbf{S}_n^3 \Theta) - (c_n^2 + 3c_n + 1) \text{tr}(\Theta) \right| \xrightarrow{\text{a.s.}} 0, & \left| \text{tr}(\mathbf{S}_n^4 \Theta) - (c_n^3 + 6c_n^2 + 6c_n + 1) \text{tr}(\Theta) \right| \xrightarrow{\text{a.s.}} 0,
 \end{aligned}$$

for  $p/n \rightarrow c \in (0, \infty)$  as  $n \rightarrow \infty$ .

**S.8. Consistent estimators for  $h_m$  and  $d_m(\cdot, \cdot)$ .** In this section we provide consistent estimators of all unknown quantities needed in Section 3. Indeed, the findings of Corollary 2.2 provide the closed-form formulae of consistent estimators for  $v^{(m)}(0)$ ,  $m = 0, 1, \dots$ , as summarized in (8). Using these results together with Corollary 2.1, we get consistent estimators for  $h_m$ ,  $d_m\left(\frac{1}{p}\mathbf{I}_p\right)$ , and, more generally,  $d_m(\Theta)$  for  $m = 1, 2, \dots$ . In particular, consistent estimators for  $h_2$ ,  $h_3$ ,  $d_1\left(\frac{1}{p}\mathbf{I}_p\right)$ ,  $d_2\left(\frac{1}{p}\mathbf{I}_p\right)$ ,  $d_1(\Theta)$ ,  $d_2(\Theta)$ , and  $d_3(\Theta)$  are given by

$$(S.38) \quad \hat{h}_2 = -\frac{1}{\hat{v}^{(1)}(0)} = \frac{1}{c_n \frac{1}{p} \text{tr}[(\mathbf{S}_n^+)^2]},$$

$$(S.39) \quad \hat{h}_3 = \frac{-\hat{v}^{(2)}(0)}{2[\hat{v}^{(1)}(0)]^3} = \frac{\frac{1}{p} \text{tr}[(\mathbf{S}_n^+)^3]}{c_n^2 \left\{ \frac{1}{p} \text{tr}[(\mathbf{S}_n^+)^2] \right\}^3},$$

$$(S.40) \quad \hat{d}_1\left(\frac{1}{p}\mathbf{I}_p\right) = -\frac{\hat{s}_1\left(\frac{1}{p}\mathbf{I}_p\right)}{\hat{v}^{(1)}(0)} = \frac{\frac{1}{p} \text{tr}[\mathbf{S}_n^+]}{c_n \frac{1}{p} \text{tr}[(\mathbf{S}_n^+)^2]},$$

$$(S.41) \quad \hat{d}_1(\Theta) = -\frac{\hat{s}_1(\Theta)}{\hat{v}^{(1)}(0)} = \frac{\text{tr}[\mathbf{S}_n^+ \Theta]}{c_n \frac{1}{p} \text{tr}[(\mathbf{S}_n^+)^2]},$$

$$(S.42) \quad \hat{d}_2\left(\frac{1}{p}\mathbf{I}_p\right) = \frac{\frac{1}{2}\hat{v}^{(2)}(0)\hat{d}_1\left(\frac{1}{p}\mathbf{I}_p\right) - \hat{s}_2\left(\frac{1}{p}\mathbf{I}_p\right)}{[\hat{v}^{(1)}(0)]^2} = \frac{\frac{1}{p} \text{tr}[\mathbf{S}_n^+] \frac{1}{p} \text{tr}[(\mathbf{S}_n^+)^3] - \left\{ \frac{1}{p} \text{tr}[(\mathbf{S}_n^+)^2] \right\}^2}{c_n^2 \left\{ \frac{1}{p} \text{tr}[(\mathbf{S}_n^+)^2] \right\}^3}$$

$$(S.43) \quad \hat{d}_2(\Theta) = \frac{\frac{1}{2}\hat{v}''(0)\hat{d}_1(\Theta) - \hat{s}_2(\Theta)}{[\hat{v}'(0)]^2} = \frac{\text{tr}[\mathbf{S}_n^+ \Theta] \frac{1}{p} \text{tr}[(\mathbf{S}_n^+)^3] - \frac{1}{p} \text{tr}[(\mathbf{S}_n^+)^2] \text{tr}[(\mathbf{S}_n^+)^2 \Theta]}{c_n^2 \left\{ \frac{1}{p} \text{tr}[(\mathbf{S}_n^+)^2] \right\}^3}$$

$$(S.44) \quad \hat{d}_3(\Theta) = \frac{\hat{v}'(0)\hat{v}''(0)\hat{d}_2(\Theta) - \frac{1}{6}\hat{v}'''(0)\hat{d}_1(\Theta) - \hat{s}_3(\Theta)}{[\hat{v}'(0)]^3}$$

$$= \frac{\text{tr}[(\mathbf{S}_n^+)^3 \Theta]}{c_n^3 \left\{ \frac{1}{p} \text{tr}[(\mathbf{S}_n^+)^2] \right\}^3} + \frac{2 \left( \frac{1}{p} \text{tr}[(\mathbf{S}_n^+)^3] \right)^2 \text{tr}[\mathbf{S}_n^+ \Theta]}{c_n^3 \left\{ \frac{1}{p} \text{tr}[(\mathbf{S}_n^+)^2] \right\}^5} - \frac{\text{tr}[(\mathbf{S}_n^+)^2 \Theta] + \frac{1}{p} \text{tr}[(\mathbf{S}_n^+)^4] \text{tr}[\mathbf{S}_n^+ \Theta]}{c_n^3 \left\{ \frac{1}{p} \text{tr}[(\mathbf{S}_n^+)^2] \right\}^4}.$$

Similarly, the findings of Theorem 2.2 and Theorem S.7.1 lead to the consistent estimators of  $d_0(t, \Theta)$ ,  $d_1(t, \Theta)$ ,  $q_1(\Theta) = \text{tr}[\Sigma \Theta]$  and  $q_2(\Theta) = \text{tr}[\Sigma^2 \Theta]$  expressed as

$$(S.45) \quad \hat{d}_0(t, \Theta) = \begin{cases} t \text{tr}[(\mathbf{S}_n + t\mathbf{I}_p)^{-1} \Theta] & \text{for } t > 0 \\ \text{tr}[(\mathbf{I}_p - \mathbf{S}_n \mathbf{S}_n^+) \Theta] & \text{for } t = 0 \end{cases},$$

$$(S.46) \quad \hat{d}_1(t, \Theta) = -\frac{t \text{tr}[(\mathbf{S}_n + t\mathbf{I}_p)^{-2} \Theta] - t^{-1} \hat{d}_0(t, \Theta)}{c_n \left( \frac{1}{p} \text{tr}[(\mathbf{S}_n + t\mathbf{I}_p)^{-2}] - t^{-2} \frac{c_n - 1}{c_n} \right)} \quad \text{for } t > 0,$$

$$(S.47) \quad \hat{q}_1(\Theta) = \text{tr}[\mathbf{S}_n \Theta],$$

$$(S.48) \quad \hat{q}_2(\Theta) = \text{tr}[\mathbf{S}_n^2 \Theta] - c_n \frac{1}{p} \text{tr}[\mathbf{S}_n] \text{tr}[\mathbf{S}_n \Theta].$$

Note that the expression of  $\hat{d}_0(0, \Theta)$  is obtained using continuity at  $t = 0$  of  $d_0$  and  $\hat{d}_0$  together with the Woodbury matrix identity in the following way

$$\hat{d}_0(0, \Theta) = \lim_{t \rightarrow 0} t \text{tr}[(\mathbf{S}_n + t\mathbf{I}_p)^{-1} \Theta] = \lim_{t \rightarrow 0} t \text{tr} \left[ \frac{1}{t} \Theta - t \frac{\mathbf{Y}_n}{\sqrt{n}} \left( \frac{1}{n} \mathbf{Y}_n^\top \mathbf{Y}_n + t\mathbf{I}_n \right)^{-1} \frac{\mathbf{Y}_n^\top}{\sqrt{n}} \Theta \right]$$

$$\begin{aligned}
 &= \text{tr}(\Theta) - \lim_{t \rightarrow 0} \text{tr} \left[ \frac{\mathbf{Y}_n}{\sqrt{n}} \left( \frac{1}{n} \mathbf{Y}_n^\top \mathbf{Y}_n + t \mathbf{I}_n \right)^{-1} \frac{\mathbf{Y}_n^\top}{\sqrt{n}} \Theta \right] \\
 &= \text{tr}(\Theta) - \text{tr} \left[ \mathbf{Y}_n (\mathbf{Y}_n^\top \mathbf{Y}_n)^{-1} \mathbf{Y}_n^\top \Theta \right] = \text{tr}(\Theta) - \text{tr} [\mathbf{S}_n \mathbf{S}_n^+ \Theta] \\
 &= \text{tr} [(\mathbf{I}_p - \mathbf{S}_n \mathbf{S}_n^+) \Theta],
 \end{aligned}$$

since the  $n \times n$  companion matrix  $\mathbf{Y}_n^\top \mathbf{Y}_n$  is non-singular for  $p > n$ .

### S.9. Shrinkage estimation of the precision matrix with the Moore-Penrose inverse.

S.9.1. *Oracle estimator.* The proof of the following theorem follows directly from Theorem 2.1 and Corollary 2.1.

**THEOREM S.9.1.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A2), it holds that*

$$|\alpha_{MP;n}^* - \alpha_{MP}^*| \xrightarrow{a.s.} 0, \text{ and } |\beta_{MP;n}^* - \beta_{MP}^*| \xrightarrow{a.s.} 0 \text{ for } p/n \rightarrow c \in (1, \infty) \text{ as } n \rightarrow \infty$$

with

$$(S.49) \quad \alpha_{MP}^* = \frac{d_1(\Sigma) - d_1 \left( \frac{\Sigma^2 \Pi_0}{\|\Sigma \Pi_0\|_F} \right) \text{tr} \left( \frac{\Sigma \Pi_0}{\|\Sigma \Pi_0\|_F} \right)}{-\frac{1}{h_2} \left( d_2(\Sigma^2) - d_1(\Sigma^2) \frac{h_3}{h_2} \right) - \frac{1}{h_2} d_1^2 \left( \frac{\Sigma^2 \Pi_0}{\|\Sigma \Pi_0\|_F} \right)},$$

$$(S.50) \quad \beta_{MP}^* = \frac{\left( d_2(\Sigma^2) - d_1(\Sigma^2) \frac{h_3}{h_2} \right) \text{tr} \left( \frac{\Sigma \Pi_0}{\|\Sigma \Pi_0\|_F} \right) + d_1(\Sigma) d_1 \left( \frac{\Sigma^2 \Pi_0}{\|\Sigma \Pi_0\|_F} \right)}{\left( d_2(\Sigma^2) - d_1(\Sigma^2) \frac{h_3}{h_2} \right) + d_1^2 \left( \frac{\Sigma^2 \Pi_0}{\|\Sigma \Pi_0\|_F} \right)} \|\Sigma \Pi_0\|_F^{-1},$$

where  $d_i$  and  $h_j$  are given in (11) and (14) for  $i = 1, 2$  and  $j = 2, 3$ , respectively.

### S.9.2. Bona-fide estimator: Proof of Theorem 3.2.

**PROOF.** With some tedious but straightforward computations, we get the following equalities

$$\begin{aligned}
 d_1 \left( \frac{1}{p} \Sigma \right) &= \frac{1}{v(0)} \left( \frac{1}{c_n v(0)} - d_1 \left( \frac{1}{p} \mathbf{I}_p \right) \right), \\
 d_1 \left( \frac{1}{p} \Sigma^2 \right) &= \frac{1}{[v(0)]^2} \left( \frac{1}{p} \text{tr}[\Sigma] + d_1 \left( \frac{1}{p} \mathbf{I}_p \right) - \frac{2}{c_n v(0)} \right), \\
 d_1 \left( \frac{1}{p} \Sigma^2 \Pi_0 \right) &= \frac{1}{[v(0)]^2} \left( \frac{1}{p} \text{tr}[\Sigma \Pi_0] + d_1 \left( \frac{1}{p} \Pi_0 \right) \right) - \frac{2}{[v(0)]^3} \left( \frac{1}{p} \text{tr}[\Pi_0] - d_0 \left( 0, \frac{1}{p} \Pi_0 \right) \right), \\
 d_2 \left( \frac{1}{p} \Sigma^2 \right) &= \frac{1}{v(0)} d_1 \left( \frac{1}{p} \Sigma^2 \right) - \frac{1}{[v(0)]^2} \left( d_1 \left( \frac{1}{p} \Sigma \right) - d_2 \left( \frac{1}{p} \mathbf{I}_p \right) \right),
 \end{aligned}$$

by using that  $\Sigma_n (v(0) \Sigma_n + \mathbf{I})^{-1} = \frac{1}{v(0)} (\mathbf{I}_p - (v(0) \Sigma_n + \mathbf{I}_p)^{-1})$ . Those simplifications show how these quantities can be consistently estimated using the results presented in Section S.8.  $\square$

### S.10. Shrinkage estimation of the precision matrix with the ridge inverse.

S.10.1. *Oracle estimators.* The proofs of the following theorems follow directly from Theorem 2.2 and Corollary 2.4.

**THEOREM S.10.1.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A2) for any  $t_0 > 0$ , it holds that*

$|\alpha_{R;n}^*(t_0) - \alpha_R^*(t_0)| \xrightarrow{a.s.} 0$ , and  $|\beta_{R;n}^*(t_0) - \beta_R^*(t_0)| \xrightarrow{a.s.} 0$  for  $p/n \rightarrow c \in (0, \infty)$  as  $n \rightarrow \infty$  with

$$\alpha_{R^*}^*(t_0) = \frac{t_0^{-1}d_0(t_0, \Sigma) - t_0^{-1}d_0\left(t_0, \frac{\Sigma^2\Pi_0}{\|\Sigma\Pi_0\|_F}\right) \operatorname{tr}\left(\frac{\Sigma\Pi_0}{\|\Sigma\Pi_0\|_F}\right)}{t_0^{-2}d_0(t_0, \Sigma^2) + t_0^{-1}v^{(1)}(t_0)d_1(t_0, \Sigma^2) - t_0^{-2}d_0^2\left(t_0, \frac{\Sigma^2\Pi_0}{\|\Sigma\Pi_0\|_F}\right)},$$

$$\beta_{R^*}^*(t_0) = \frac{\left(t_0^{-2}d_0(t_0, \Sigma^2) + t_0^{-1}v^{(1)}(t_0)d_1(t_0, \Sigma^2)\right) \operatorname{tr}\left(\frac{\Sigma\Pi_0}{\|\Sigma\Pi_0\|_F}\right) - t_0^{-2}d_0(t_0, \Sigma)d_0\left(t_0, \frac{\Sigma^2\Pi_0}{\|\Sigma\Pi_0\|_F}\right)}{t_0^{-2}d_0(t_0, \Sigma^2) + t_0^{-1}v^{(1)}(t_0)d_1(t_0, \Sigma^2) - t_0^{-2}d_0^2\left(t_0, \frac{\Sigma^2\Pi_0}{\|\Sigma\Pi_0\|_F}\right)} \|\Sigma\Pi_0\|_F^{-1},$$

where  $d_i(t_0, \cdot)$  and  $v^{(1)}(t_0)$  are given in (22) and (23) for  $i = 1, 2$ .

**THEOREM S.10.2.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A2) for any  $t > 0$ , it holds that*

$$\frac{1}{p} |L_{R;n,2}^2(t) - L_{R;2}^2(t)| \xrightarrow{a.s.} 0, \text{ for } p/n \rightarrow c \in (0, \infty) \text{ as } n \rightarrow \infty \quad \text{with}$$

$$L_{R;2}^2(t) = \frac{\left[d_0(t, \Sigma) - d_0\left(t, \frac{\Sigma^2\Pi_0}{\|\Sigma\Pi_0\|_F}\right) \operatorname{tr}\left(\frac{\Sigma\Pi_0}{\|\Sigma\Pi_0\|_F}\right)\right]^2}{d_0(t, \Sigma^2) + tv^{(1)}(t)d_1(t, \Sigma^2) - d_0^2\left(t, \frac{\Sigma^2\Pi_0}{\|\Sigma\Pi_0\|_F}\right)},$$

where  $d_i(t, \cdot)$  and  $v^{(1)}(t)$  are given in (22) and (23) for  $i = 1, 2$ .

#### S.10.2. Bona-fide estimators: Proof of Theorem 3.3.

**PROOF.** It holds that

$$d_0\left(t_0, \frac{1}{p}\Sigma\right) = \frac{1}{p} \operatorname{tr}\left[(v(t_0)\Sigma + \mathbf{I}_p)^{-1}\Sigma\right] = \frac{1}{c_n v(t_0)} - \frac{t_0}{c_n},$$

$$d_0\left(t_0, \frac{1}{p}\Sigma^2\right) = \frac{1}{p} \operatorname{tr}\left[(v(t_0)\Sigma + \mathbf{I}_p)^{-1}\Sigma^2\right] = \frac{1}{v(t_0)} \left(\frac{1}{p} \operatorname{tr}[\Sigma] - \frac{1}{c_n v(t_0)} + \frac{t_0}{c_n}\right),$$

$$d_0\left(t_0, \frac{1}{p}\Sigma^2\Pi_0\right) = \frac{1}{p} \operatorname{tr}\left[(v(t_0)\Sigma + \mathbf{I}_p)^{-1}\Sigma^2\Pi_0\right]$$

$$= \frac{1}{v(t_0)} \left(\frac{1}{p} \operatorname{tr}[\Sigma\Pi_0] - \frac{1}{p} \operatorname{tr}\left[(v(t_0)\Sigma + \mathbf{I}_p)^{-1}\Sigma\Pi_0\right]\right)$$

$$= \frac{1}{v(t_0)} \frac{1}{p} \operatorname{tr}[\Sigma\Pi_0] - \frac{1}{[v(t_0)]^2} \left(\frac{1}{p} \operatorname{tr}[\Pi_0] - d_0\left(t_0, \frac{1}{p}\Pi_0\right)\right),$$

$$\begin{aligned}
d_1\left(t_0, \frac{1}{p}\Sigma^2\right) &= \frac{1}{p}\text{tr}\left\{(v(t_0)\Sigma + \mathbf{I}_p)^{-1}\Sigma(v(t_0)\Sigma + \mathbf{I}_p)^{-1}\Sigma^2\right\} \\
&= \frac{1}{v(t_0)}\left(\frac{1}{p}\text{tr}\left[(v(t_0)\Sigma + \mathbf{I}_p)^{-1}\Sigma^2\right] - \frac{1}{p}\text{tr}\left[(v(t_0)\Sigma + \mathbf{I}_p)^{-1}\Sigma(v(t_0)\Sigma + \mathbf{I}_p)^{-1}\Sigma\right]\right) \\
&= \frac{1}{[v(t_0)]^2}\left(\frac{1}{p}\text{tr}[\Sigma] + d_1\left(t_0, \frac{1}{p}\mathbf{I}_p\right) - \frac{2}{c_n v(t_0)} + \frac{2t_0}{c_n}\right),
\end{aligned}$$

where  $v(t_0)$ ,  $v^{(1)}(t_0)$ ,  $d_0\left(t_0, \frac{1}{p}\Pi_0\right)$ ,  $d_1\left(t_0, \frac{1}{p}\mathbf{I}_p\right)$ ,  $q_1\left(\frac{1}{p}\Pi_0\right)$ ,  $q_2\left(\frac{1}{p}\Pi_0\right)$  and  $q_2\left(\frac{1}{p}\Pi_0^2\right)$  are consistently estimated as in (7), (S.45), (S.46), (S.47) and (S.48), respectively.  $\square$

### S.11. Shrinkage estimation of the precision matrix with the Moore-Penrose-ridge inverse.

S.11.1. *Oracle estimators.* The application of Corollary 2.8 yields

**THEOREM S.11.1.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A2) for any  $t_0 > 0$ , it holds that*

$$|\alpha_{MPR;n}^*(t_0) - \alpha_{MPR}^*(t_0)| \xrightarrow{a.s.} 0, \quad \text{and} \quad |\beta_{MPR;n}^*(t_0) - \beta_{MPR}^*(t_0)| \xrightarrow{a.s.} 0,$$

for  $p/n \rightarrow c \in (0, \infty)$  as  $n \rightarrow \infty$  with

$$\begin{aligned}
\alpha_{MPR}^*(t_0) &= \frac{-v^{(1)}(t_0)d_1(t_0, \Sigma) + v^{(1)}(t_0)d_1\left(t_0, \frac{\Sigma^2\Pi_0}{\|\Sigma\Pi_0\|_F}\right)\text{tr}\left(\frac{\Sigma\Pi_0}{\|\Sigma\Pi_0\|_F}\right)}{\dot{s}_2(t_0, \Sigma^2) - [v^{(1)}(t_0)]^2 d_1^2\left(t_0, \frac{\Sigma^2\Pi_0}{\|\Sigma\Pi_0\|_F}\right)}, \\
\beta_{MPR}^*(t_0) &= \frac{\dot{s}_2(t_0, \Sigma^2)\text{tr}\left(\frac{\Sigma\Pi_0}{\|\Sigma\Pi_0\|_F}\right) - [v^{(1)}(t_0)]^2 d_1(t_0, \Sigma)d_1\left(t_0, \frac{\Sigma^2\Pi_0}{\|\Sigma\Pi_0\|_F}\right)}{\dot{s}_2(t_0, \Sigma^2) - [v^{(1)}(t_0)]^2 d_1^2\left(t_0, \frac{\Sigma^2\Pi_0}{\|\Sigma\Pi_0\|_F}\right)} \|\Sigma\Pi_0\|_F^{-1},
\end{aligned}$$

where  $d_1(t_0, \cdot)$ ,  $v^{(1)}(t_0)$  and  $\dot{s}_2(t_0, \cdot)$  are given in (22), (23), and (38), respectively.

The optimal value of the tuning parameter  $t$  should be chosen by maximizing  $L_{MPR;n,2}^2(t) = L_{F;n,2}^2(\mathbf{S}_n^\pm(t))$ . Since the latter depends on the unknown population covariance matrix  $\Sigma$ , we first derive its asymptotic deterministic equivalent in Theorem S.11.2 by using Corollary 2.8.

**THEOREM S.11.2.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A2) for any  $t > 0$ , it holds that*

$$\frac{1}{p} \left| L_{MPR;n,2}^2(t) - L_{MPR;2}^2(t) \right| \xrightarrow{a.s.} 0, \quad \text{for } p/n \rightarrow c \in (0, \infty) \text{ as } n \rightarrow \infty \text{ with}$$

$$L_{MPR;2}^2(t) = \frac{\left[ -v^{(1)}(t)d_1(t, \Sigma) + v^{(1)}(t)d_1\left(t, \frac{\Sigma^2\Pi_0}{\|\Sigma\Pi_0\|_F}\right)\text{tr}\left(\frac{\Sigma\Pi_0}{\|\Sigma\Pi_0\|_F}\right) \right]^2}{\dot{s}_2(t, \Sigma^2) - [v^{(1)}(t)]^2 d_1^2\left(t, \frac{\Sigma^2\Pi_0}{\|\Sigma\Pi_0\|_F}\right)},$$

where  $d_1(t, \cdot)$ ,  $v^{(1)}(t)$  and  $\dot{s}_2(t, \cdot)$  are given in (22), (23), and (38), respectively.

S.11.2. *Bona-fide estimators.* Consistent estimators of  $\alpha_{MPR}^*(t_0)$  and  $\beta_{MPR}^*(t_0)$  are given in Theorem S.11.3, while Theorem S.11.4 provides a consistent estimator of  $L_{MPR;2}^2(t)$ .

THEOREM S.11.3. *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A2) for any  $t > 0$ , consistent estimators for  $\alpha_{MPR}^*(t_0)$  and  $\beta_{MPR}^*(t_0)$  are given by*

$$(S.51) \quad \hat{\alpha}_{MPR}^*(t_0) = \frac{-\hat{v}^{(1)}(t_0)\hat{d}_1\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}\right)\hat{q}_2\left(\frac{1}{p}\boldsymbol{\Pi}_0^2\right) + \hat{v}^{(1)}(t_0)\hat{d}_1\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\boldsymbol{\Pi}_0\right)\hat{q}_1\left(\frac{1}{p}\boldsymbol{\Pi}_0\right)}{\hat{s}_2\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\right)\hat{q}_2\left(\frac{1}{p}\boldsymbol{\Pi}_0^2\right) - [\hat{v}^{(1)}(t_0)]^2\hat{d}_1^2\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\boldsymbol{\Pi}_0\right)},$$

$$(S.52) \quad \hat{\beta}_{MPR}^*(t_0) = \frac{\hat{s}_2\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\right)\hat{q}_1\left(\frac{1}{p}\boldsymbol{\Pi}_0\right) - [\hat{v}^{(1)}(t_0)]^2\hat{d}_1\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}\right)\hat{d}_1\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\boldsymbol{\Pi}_0\right)}{\hat{s}_2\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\right)\hat{q}_2\left(\frac{1}{p}\boldsymbol{\Pi}_0^2\right) - [\hat{v}^{(1)}(t_0)]^2\hat{d}_1^2\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\boldsymbol{\Pi}_0\right)},$$

with

$$(S.53) \quad \hat{s}_2\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\right) = -\left\{[\hat{v}^{(1)}(t_0)]^2\hat{d}_2\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\right) - \frac{1}{2}\hat{v}^{(2)}(t_0)\hat{d}_1\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\right)\right\} \\ + t_0\left\{\frac{1}{6}\hat{v}^{(3)}(t_0)\hat{d}_1\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\right) - \hat{v}^{(1)}(t_0)\hat{v}^{(2)}(t_0)\hat{d}_2\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\right) + [\hat{v}^{(1)}(t_0)]^3\hat{d}_3\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\right)\right\},$$

$$(S.54) \quad \hat{d}_1\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}\right) = \frac{[\hat{v}(t_0)]^{-2} + [\hat{v}^{(1)}(t_0)]^{-1}}{c_n},$$

$$(S.55) \quad \hat{d}_1\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\right) = \frac{1}{\hat{v}(t_0)}\left\{d_0\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\right) - \hat{d}_1\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}\right)\right\},$$

$$(S.56) \quad \hat{d}_1\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\boldsymbol{\Pi}_0\right) = \frac{1}{\hat{v}(t_0)}\hat{d}_0\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\boldsymbol{\Pi}_0\right) + \frac{1}{[\hat{v}(t_0)]^2}\hat{d}_1\left(t_0, \frac{1}{p}\boldsymbol{\Pi}_0\right) \\ - \frac{1}{[\hat{v}(t_0)]^3}\left\{\frac{1}{p}\text{tr}\{\boldsymbol{\Pi}_0\} - \hat{d}_0\left(t_0, \frac{1}{p}\boldsymbol{\Pi}_0\right)\right\},$$

$$(S.57) \quad \hat{d}_2\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\right) = \frac{1}{\hat{v}(t_0)}\left\{\hat{d}_1\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\right) - \frac{1}{c_n}\left(\frac{1}{[\hat{v}(t_0)]^3} + \frac{\hat{v}^{(2)}(t_0)}{2[\hat{v}^{(1)}(t_0)]^3}\right)\right\},$$

$$(S.58) \quad \hat{d}_3\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\right) = \frac{1}{\hat{v}(t_0)}\left\{\hat{d}_2\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\right) - \frac{1}{c_n}\left(\frac{1}{[\hat{v}(t_0)]^4} + \frac{1}{2}\frac{[\hat{v}^{(2)}(t_0)]^2}{[\hat{v}^{(1)}(t_0)]^5} - \frac{1}{6}\frac{\hat{v}^{(3)}(t_0)}{[\hat{v}^{(1)}(t_0)]^4}\right)\right\},$$

where  $\hat{v}(t_0)$ ,  $\hat{v}^{(1)}(t_0)$ ,  $\hat{v}^{(2)}(t_0)$ ,  $\hat{v}^{(3)}(t_0)$ ,  $\hat{d}_0\left(t_0, \frac{1}{p}\boldsymbol{\Pi}_0\right)$ ,  $\hat{d}_1\left(t_0, \frac{1}{p}\boldsymbol{\Pi}_0\right)$ ,  $\hat{q}_1\left(\frac{1}{p}\boldsymbol{\Pi}_0\right)$ ,  $\hat{q}_2\left(\frac{1}{p}\boldsymbol{\Pi}_0^2\right)$ ,  $\hat{d}_0\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\right)$  and  $\hat{d}_0\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\boldsymbol{\Pi}_0\right)$  are given in (7), (S.45), (S.46), (S.47), (S.48), (56) and (57).

PROOF OF THEOREM S.11.3: It holds that

$$d_1\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}\right) = \frac{1}{p}\text{tr}\left\{(v(t_0)\boldsymbol{\Sigma} + \mathbf{I}_p)^{-1}\boldsymbol{\Sigma}(v(t_0)\boldsymbol{\Sigma} + \mathbf{I}_p)^{-1}\boldsymbol{\Sigma}\right\} \\ = \frac{[v(t_0)]^{-2} - h_2(t)}{c_n} = \frac{[v(t_0)]^{-2} + [v^{(1)}(t_0)]^{-1}}{c_n},$$

$$d_1\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\right) = \frac{1}{p}\text{tr}\left\{(v(t_0)\boldsymbol{\Sigma} + \mathbf{I}_p)^{-1}\boldsymbol{\Sigma}(v(t_0)\boldsymbol{\Sigma} + \mathbf{I}_p)^{-1}\boldsymbol{\Sigma}^2\right\} \\ = \frac{1}{v(t_0)}\left\{d_0\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}^2\right) - d_1\left(t_0, \frac{1}{p}\boldsymbol{\Sigma}\right)\right\},$$

$$\begin{aligned}
 d_1 \left( t_0, \frac{1}{p} \Sigma^2 \Pi_0 \right) &= \frac{1}{p} \text{tr} \left\{ (v(t_0) \Sigma + \mathbf{I}_p)^{-1} \Sigma (v(t_0) \Sigma + \mathbf{I}_p)^{-1} \Sigma^2 \Pi_0 \right\} \\
 &= \frac{1}{v(t_0)} d_0 \left( t_0, \frac{1}{p} \Sigma^2 \Pi_0 \right) + \frac{1}{[v(t_0)]^2} d_1 \left( t_0, \frac{1}{p} \Pi_0 \right) \\
 &\quad - \frac{1}{[v(t_0)]^3} \left\{ \frac{1}{p} \text{tr} \{ \Pi_0 \} - d_0 \left( t_0, \frac{1}{p} \Pi_0 \right) \right\},
 \end{aligned}$$

$$\begin{aligned}
 d_2 \left( t_0, \frac{1}{p} \Sigma^2 \right) &= \frac{1}{p} \text{tr} \left\{ (v(t_0) \Sigma + \mathbf{I}_p)^{-1} \Sigma (v(t_0) \Sigma + \mathbf{I}_p)^{-1} \Sigma (v(t_0) \Sigma + \mathbf{I}_p)^{-1} \Sigma^2 \right\} \\
 &= \frac{1}{v(t_0)} \left\{ d_1 \left( t_0, \frac{1}{p} \Sigma^2 \right) - \frac{[v(t_0)]^{-3} - h_3(t)}{c_n} \right\} \\
 &= \frac{1}{v(t_0)} \left\{ d_1 \left( t_0, \frac{1}{p} \Sigma^2 \right) - \frac{1}{c_n} \left( \frac{1}{[v(t_0)]^3} + \frac{v^{(2)}(t_0)}{2[v^{(1)}(t_0)]^3} \right) \right\},
 \end{aligned}$$

$$\begin{aligned}
 d_3 \left( t_0, \frac{1}{p} \Sigma^2 \right) &= \frac{1}{p} \text{tr} \left\{ (v(t_0) \Sigma + \mathbf{I}_p)^{-1} \Sigma (v(t_0) \Sigma + \mathbf{I}_p)^{-1} \Sigma (v(t_0) \Sigma + \mathbf{I}_p)^{-1} \Sigma (v(t_0) \Sigma + \mathbf{I}_p)^{-1} \Sigma^2 \right\} \\
 &= \frac{1}{v(t_0)} \left\{ d_2 \left( t_0, \frac{1}{p} \Sigma^2 \right) - \frac{[v(t_0)]^{-4} - h_4(t)}{c_n} \right\} \\
 &= \frac{1}{v(t_0)} \left\{ d_2 \left( t_0, \frac{1}{p} \Sigma^2 \right) - \frac{1}{c_n} \left( \frac{1}{[v(t_0)]^4} + \frac{1}{2} \frac{[v^{(2)}(t_0)]^2}{[v^{(1)}(t_0)]^5} - \frac{1}{6} \frac{v^{(3)}(t_0)}{[v^{(1)}(t_0)]^4} \right) \right\}.
 \end{aligned}$$

□

**THEOREM S.11.4.** *Let  $\mathbf{Y}_n$  fulfill the stochastic representation (1). Then, under Assumptions (A1)-(A2) for any  $t > 0$ , it holds that*

$$\frac{1}{p} \left| \hat{L}_{MPR;2}^2(t) - L_{MPR;2}^2(t) \right| \xrightarrow{a.s.} 0, \text{ for } p/n \rightarrow c \in (0, \infty) \text{ as } n \rightarrow \infty$$

with

$$\text{(S.59)} \quad \hat{L}_{MPR;2}^2(t) = \frac{1}{\hat{q}_2 \left( \frac{1}{p} \Pi_0^2 \right)} \frac{[\hat{v}^{(1)}(t)]^2 \left[ \hat{d}_1 \left( t, \frac{1}{p} \Sigma \right) \hat{q}_2 \left( \frac{1}{p} \Pi_0^2 \right) - \hat{d}_1 \left( t, \frac{1}{p} \Sigma^2 \Pi_0 \right) \hat{q}_1 \left( \frac{1}{p} \Pi_0 \right) \right]^2}{\hat{s}_2 \left( t, \frac{1}{p} \Sigma^2 \right) \hat{q}_2 \left( \frac{1}{p} \Pi_0^2 \right) - [\hat{v}^{(1)}(t)]^2 \hat{d}_1^2 \left( t, \frac{1}{p} \Sigma^2 \Pi_0 \right)},$$

where  $\hat{v}^{(1)}(t)$ ,  $\hat{q}_1 \left( \frac{1}{p} \Pi_0 \right)$ ,  $\hat{q}_2 \left( \frac{1}{p} \Pi_0^2 \right)$ ,  $\hat{s}_2(t, \Sigma^2)$ ,  $\hat{d}_1 \left( t, \frac{1}{p} \Sigma \right)$  and  $\hat{d}_1 \left( t, \frac{1}{p} \Sigma^2 \Pi_0 \right)$  are given in (7), (S.47), (S.48), (S.53), (S.54) and (S.56), respectively.

In the case of  $t = 0$ , a consistent estimator for  $L_{MPR;2}^2(0)$  is constructed by using the properties of the Moore-Penrose inverse. Namely, it holds that

$$\text{(S.60)} \quad \hat{L}_{MPR;2}^2(0) = \frac{1}{\hat{q}_2 \left( \frac{1}{p} \Pi_0^2 \right)} \frac{[\hat{v}^{(1)}(0)]^2 \left[ \hat{d}_1 \left( \frac{1}{p} \Sigma \right) \hat{q}_2 \left( \frac{1}{p} \Pi_0^2 \right) - \hat{d}_1 \left( 0, \frac{1}{p} \Sigma^2 \Pi_0 \right) \hat{q}_1 \left( \frac{1}{p} \Pi_0 \right) \right]^2}{\hat{s}_2 \left( \frac{1}{p} \Sigma^2 \right) \hat{q}_2 \left( \frac{1}{p} \Pi_0^2 \right) - [\hat{v}^{(1)}(0)]^2 \hat{d}_1^2 \left( 0, \frac{1}{p} \Sigma^2 \Pi_0 \right)}$$

with

$$\text{(S.61)} \quad \hat{s}_2 \left( \frac{1}{p} \Sigma^2 \right) = - \left\{ [\hat{v}^{(1)}(0)]^2 \hat{d}_2 \left( \frac{1}{p} \Sigma^2 \right) - \frac{1}{2} \hat{v}^{(2)}(0) \hat{d}_1 \left( \frac{1}{p} \Sigma^2 \right) \right\},$$

where  $\hat{v}^{(1)}(0)$ ,  $\hat{v}^{(2)}(0)$ ,  $\hat{q}_1\left(\frac{1}{p}\mathbf{\Pi}_0\right)$ ,  $\hat{q}_2\left(\frac{1}{p}\mathbf{\Pi}_0^2\right)$ ,  $\hat{d}_1\left(\frac{1}{p}\mathbf{\Sigma}\right)$ ,  $\hat{d}_1\left(0, \frac{1}{p}\mathbf{\Sigma}^2\mathbf{\Pi}_0\right)$ ,  $\hat{d}_1\left(\frac{1}{p}\mathbf{\Sigma}^2\right)$  and  $\hat{d}_2\left(\frac{1}{p}\mathbf{\Sigma}^2\right)$  are given in (8), (S.47), (S.48), (48), (49), (50) and (51), respectively. Furthermore, if  $\mathbf{\Pi}_0 = \mathbf{I}_p$ , then one can simplify the computation of  $\hat{d}_1\left(\frac{1}{p}\mathbf{\Sigma}^2\mathbf{\Pi}_0\right)$  by using  $d_0\left(0, \frac{1}{p}\mathbf{I}_p\right) = \frac{c_n-1}{c_n}$ .

To find the optimal value of the tuning parameter  $t$ , we first maximize  $\hat{L}_{MPR;2}^2(t)$  over the open interval  $(0, \infty)$  to get  $t^*$ . If  $\hat{L}_{MPR;2}^2(t^*) > \hat{L}_{MPR;2}^2(0)$ , then  $t^*$  should be used as the optimal value of the tuning parameter. Otherwise, one opts for the Moore-Penrose inverse and the optimal shrinkage intensities are determined as in Section 3.1.1.

**S.12. Oracle estimator of the GMV portfolio weights with the Moore-Penrose inverse.** Next, we derive the expression of the oracle estimator of the GMV portfolio weights.

**THEOREM S.12.1.** *Let  $\mathbf{Y}_n$  possess the stochastic representation as in (1). Then, under Assumptions (A1)-(A2), it holds that*

$$(S.62) \quad |\alpha_n^* - \alpha^*| \xrightarrow{a.s.} 0$$

for  $p/n \rightarrow c \in (1, +\infty)$  as  $n \rightarrow \infty$  with

$$(S.63) \quad \alpha^* = \begin{cases} \frac{p\mathbf{b}^\top \mathbf{\Sigma} \mathbf{b} - \frac{d_1(\mathbf{1b}^\top \mathbf{\Sigma})}{d_1\left(\frac{\mathbf{11}^\top}{p}\right)}}{p\mathbf{b}^\top \mathbf{\Sigma} \mathbf{b} - 2\frac{d_1(\mathbf{1b}^\top \mathbf{\Sigma})}{d_1\left(\frac{\mathbf{11}^\top}{p}\right)} + \frac{d_3\left(\frac{\mathbf{11}^\top}{p}\right)}{d_1^2\left(\frac{\mathbf{11}^\top}{p}\right)}} & \text{if } p\mathbf{b}^\top \mathbf{\Sigma} \mathbf{b} = O(1), \\ 1, & \text{otherwise.} \end{cases}$$

**PROOF OF THEOREM S.12.1:** First, it is noted that in case of Moore-Penrose inverse, i.e.,  $\mathbf{S}_n^\#(t) = \mathbf{S}_n^+$  the function  $L_{n;2}(t)$  is independent of  $t$  in the case of the Moore-Penrose inverse, i.e.,  $\mathbf{S}_n^\#(t) = \mathbf{S}_n^+$ , so we just need to find the asymptotic behavior of  $\alpha_n^*(t) = \alpha_n^*$  with

$$\alpha_n^* = \frac{\mathbf{b}^\top \mathbf{\Sigma} (\mathbf{b} - \mathbf{w}_{\mathbf{S}_n^+})}{(\mathbf{b} - \mathbf{w}_{\mathbf{S}_n^+})^\top \mathbf{\Sigma} (\mathbf{b} - \mathbf{w}_{\mathbf{S}_n^+})} = \frac{p\mathbf{b}^\top \mathbf{\Sigma} \mathbf{b} - \frac{\mathbf{b}^\top \mathbf{\Sigma} \mathbf{S}_n^+ \mathbf{1}}{\frac{1}{p}\mathbf{1}^\top \mathbf{S}_n^+ \mathbf{1}}}{p\mathbf{b}^\top \mathbf{\Sigma} \mathbf{b} - 2\frac{\mathbf{b}^\top \mathbf{\Sigma} \mathbf{S}_n^+ \mathbf{1}}{\frac{1}{p}\mathbf{1}^\top \mathbf{S}_n^+ \mathbf{1}} + \frac{\frac{1}{p}\mathbf{1}^\top \mathbf{S}_n^+ \mathbf{\Sigma} \mathbf{S}_n^+ \mathbf{1}}{(\frac{1}{p}\mathbf{1}^\top \mathbf{S}_n^+ \mathbf{1})^2}}$$

Next, we study the asymptotic behavior of  $\mathbf{b}^\top \mathbf{\Sigma} \mathbf{S}_n^+ \mathbf{1}$  and  $\frac{1}{p}\mathbf{1}^\top \mathbf{S}_n^+ \mathbf{1}$ . Since  $\text{tr}\left(\frac{\mathbf{11}^\top}{p}\right) = 1$ ,

$$\text{tr}\left(\mathbf{1b}^\top \mathbf{\Sigma}\right) = \text{tr}\left(1/2\left(\mathbf{1b}^\top + \mathbf{b1}^\top\right) \mathbf{\Sigma}\right) \leq \lambda_{\max}(\mathbf{\Sigma})\mathbf{b}^\top \mathbf{1} = \lambda_{\max}(\mathbf{\Sigma}) < \infty,$$

and using Theorem 2.1, we get for  $p, n \rightarrow \infty$  with  $p/n \rightarrow c > 1$  that

$$(S.64) \quad \frac{\mathbf{b}^\top \mathbf{\Sigma} \mathbf{S}_n^+ \mathbf{1}}{\frac{1}{p}\mathbf{1}^\top \mathbf{S}_n^+ \mathbf{1}} \xrightarrow{a.s.} \frac{d_1(\mathbf{1b}^\top \mathbf{\Sigma})}{d_1\left(\frac{\mathbf{11}^\top}{p}\right)},$$

where due to Theorem 2.1, it holds that

$$(S.65) \quad \text{tr}(\mathbf{S}_n^+ \mathbf{\Theta}) \xrightarrow{d.a.s.} (-v'(0))d_1(\mathbf{\Theta}) \quad \text{with } \mathbf{\Theta} \in \{\mathbf{1b}^\top \mathbf{\Sigma}, \frac{1}{p}\mathbf{11}^\top\}.$$

To examine the asymptotic behavior of  $\frac{1}{p} \mathbf{1}^\top \mathbf{S}_n^+ \Sigma \mathbf{S}_n^+ \mathbf{1}$ , we consider the eigenvalue decomposition of  $\Sigma = \sum_{i=1}^p \mathbf{v}_i \mathbf{v}_i^\top \tau_i$  with  $\tau_1, \dots, \tau_p$  eigenvalues of  $\Sigma$  and  $\mathbf{v}_1, \dots, \mathbf{v}_p$  the corresponding eigenvectors. It holds that

$$\frac{1}{p} \mathbf{1}^\top \mathbf{S}_n^+ \Sigma \mathbf{S}_n^+ \mathbf{1} = \frac{1}{p} \sum_{i=1}^p \mathbf{1}^\top \mathbf{S}_n^+ \mathbf{v}_i \mathbf{v}_i^\top \mathbf{S}_n^+ \mathbf{1} \tau_i = \sum_{i=1}^p \left( \frac{1}{\sqrt{p}} \mathbf{1}^\top \mathbf{S}_n^+ \mathbf{v}_i^\top \right)^2 \tau_i.$$

Since  $\text{tr} \left( \frac{1}{\sqrt{p}} \mathbf{v}_i \mathbf{1}^\top \right)$  is bounded by one due to the Cauchy-Schwarz inequality, the application of Theorem 2.1 yields

$$\left| \frac{1}{\sqrt{p}} \mathbf{1}^\top \mathbf{S}_n^+ \mathbf{v}_i - (-v'(0)) d_1 \left( \frac{\mathbf{v}_i \mathbf{1}^\top}{\sqrt{p}} \right) \right| a.s. \rightarrow 0 \quad \text{for } p/n \rightarrow c \in (1, \infty) \quad \text{as } n \rightarrow \infty.$$

Furthermore, due to  $\frac{\mathbf{v}_i^\top \mathbf{1}}{\sqrt{p}} = O(1)$ , one immediately gets for  $(p, n) \rightarrow \infty$  using the trace inequality that

$$(S.66) \quad \frac{1}{\sqrt{p}} \mathbf{1}^\top \mathbf{S}_n^+ \mathbf{v}_i - (-v'(0)) d_1 \left( \frac{\mathbf{v}_i \mathbf{1}^\top}{\sqrt{p}} \right) = o_{\mathbb{P}} \left( \frac{\mathbf{v}_i^\top \mathbf{1}}{\sqrt{p}} \right),$$

$$(S.67) \quad (-v'(0)) d_1 \left( \frac{\mathbf{v}_i \mathbf{1}^\top}{\sqrt{p}} \right) = O \left( \frac{\mathbf{v}_i^\top \mathbf{1}}{\sqrt{p}} \right).$$

The application of (S.66) and (S.67) results into

$$\begin{aligned} \sum_{i=1}^p \left( \frac{1}{\sqrt{p}} \mathbf{1}^\top \mathbf{S}_n^+ \mathbf{v}_i^\top \right)^2 \tau_i &= \sum_{i=1}^p \left( \frac{1}{\sqrt{p}} \mathbf{1}^\top \mathbf{S}_n^+ \mathbf{v}_i^\top - (-v'(0)) d_1 \left( \frac{\mathbf{v}_i \mathbf{1}^\top}{\sqrt{p}} \right) + (-v'(0)) d_1 \left( \frac{\mathbf{v}_i \mathbf{1}^\top}{\sqrt{p}} \right) \right)^2 \tau_i \\ &= \sum_{i=1}^p \underbrace{\left( \frac{1}{\sqrt{p}} \mathbf{1}^\top \mathbf{S}_n^+ \mathbf{v}_i^\top - (-v'(0)) d_1 \left( \frac{\mathbf{v}_i \mathbf{1}^\top}{\sqrt{p}} \right) \right)^2}_{o_{\mathbb{P}} \left( \frac{\mathbf{v}_i^\top \mathbf{1}}{\sqrt{p}} \right)^2} \tau_i \\ &\quad + 2 \sum_{i=1}^p \underbrace{\left( \frac{1}{\sqrt{p}} \mathbf{1}^\top \mathbf{S}_n^+ \mathbf{v}_i^\top - (-v'(0)) d_1 \left( \frac{\mathbf{v}_i \mathbf{1}^\top}{\sqrt{p}} \right) \right)}_{o_{\mathbb{P}} \left( \frac{\mathbf{v}_i^\top \mathbf{1}}{\sqrt{p}} \right)} \underbrace{(-v'(0)) d_1 \left( \frac{\mathbf{v}_i \mathbf{1}^\top}{\sqrt{p}} \right)}_{O \left( \frac{\mathbf{v}_i^\top \mathbf{1}}{\sqrt{p}} \right)} \tau_i \\ &\quad + \sum_{i=1}^p (-v'(0))^2 d_1^2 \left( \frac{\mathbf{v}_i \mathbf{1}^\top}{\sqrt{p}} \right) \tau_i \\ &= C \sum_{i=1}^p o_{\mathbb{P}} \left( \frac{\mathbf{1}^\top \mathbf{v}_i \mathbf{v}_i^\top \mathbf{1}}{p} \right) \tau_i + \sum_{i=1}^p (-v'(0))^2 d_1^2 \left( \frac{\mathbf{v}_i \mathbf{1}^\top}{\sqrt{p}} \right) \tau_i, \end{aligned}$$

where the constant  $C$  is independent of  $p$  and  $n$ , and

$$\sum_{i=1}^p o_{\mathbb{P}} \left( \frac{\mathbf{1}^\top \mathbf{v}_i \mathbf{v}_i^\top \mathbf{1}}{p} \right) \tau_i \leq o_{\mathbb{P}} \left( \sum_{i=1}^p \frac{\mathbf{1}^\top \mathbf{v}_i \mathbf{v}_i^\top \mathbf{1}}{p} \right) \tau_{max} = o_{\mathbb{P}} \left( \frac{\mathbf{1}^\top \sum_{i=1}^p \mathbf{v}_i \mathbf{v}_i^\top \mathbf{1}}{p} \right) \tau_{max}$$

$$= o_{\mathbb{P}} \left( \frac{1}{p} \mathbf{1}^{\top} \underbrace{[\mathbf{v}_1, \dots, \mathbf{v}_p][\mathbf{v}_1, \dots, \mathbf{v}_p]^{\top}}_{=\mathbf{I}_p} \mathbf{1} \right) \tau_{max} = o_{\mathbb{P}}(\tau_{max}).$$

Since the maximum eigenvalue of  $\Sigma$ , i.e.,  $\tau_{max} = \lambda_{max}(\Sigma)$ , is uniformly bounded in  $p$ , we get  $o_{\mathbb{P}}(\tau_{max}) = o_{\mathbb{P}}(1)$  and the definitions of  $d_1(\cdot)$  and  $d_3(\cdot)$  from (11) yield

$$\begin{aligned} & \sum_{i=1}^p \left( \frac{1}{\sqrt{p}} \mathbf{1}^{\top} \mathbf{S}_n^+ \mathbf{v}_i^{\top} \right)^2 \tau_i \xrightarrow{d.a.s.} \sum_{i=1}^p (-v'(0))^2 d_1^2 \left( \frac{\mathbf{v}_i \mathbf{1}^{\top}}{\sqrt{p}} \right) \tau_i \\ &= (v'(0))^2 \sum_{i=1}^p \text{tr}^2 \left\{ (v(0)\Sigma + \mathbf{I}_p)^{-1} \left[ \Sigma (v(0)\Sigma + \mathbf{I}_p)^{-1} \right] \frac{\mathbf{v}_i \mathbf{1}^{\top}}{\sqrt{p}} \right\} \tau_i \\ &= \frac{(v'(0))^2}{p} \sum_{i=1}^p \mathbf{1}^{\top} (v(0)\Sigma + \mathbf{I}_p)^{-1} \left[ \Sigma (v(0)\Sigma + \mathbf{I}_p)^{-1} \right] \mathbf{v}_i \mathbf{v}_i^{\top} (v(0)\Sigma + \mathbf{I}_p)^{-1} \left[ \Sigma (v(0)\Sigma + \mathbf{I}_p)^{-1} \right] \mathbf{1} \tau_i \\ &= \frac{(v'(0))^2}{p} \mathbf{1}^{\top} (v(0)\Sigma + \mathbf{I}_p)^{-1} \left[ \Sigma (v(0)\Sigma + \mathbf{I}_p)^{-1} \right] \underbrace{\left( \sum_{i=1}^p \mathbf{v}_i \mathbf{v}_i^{\top} \tau_i \right)}_{\Sigma} (v(0)\Sigma + \mathbf{I}_p)^{-1} \left[ \Sigma (v(0)\Sigma + \mathbf{I}_p)^{-1} \right] \mathbf{1} \\ &= (v'(0))^2 \text{tr} \left\{ (v(0)\Sigma + \mathbf{I}_p)^{-1} \left[ \Sigma (v(0)\Sigma + \mathbf{I}_p)^{-1} \right]^3 \frac{\mathbf{1} \mathbf{1}^{\top}}{p} \right\} = (v'(0))^2 d_3 \left( \frac{\mathbf{1} \mathbf{1}^{\top}}{p} \right). \end{aligned}$$

Finally, the application of the last equality, (S.64), (S.65) and the expression of  $\alpha_n^*$  finish the proof of the theorem.  $\square$

**REMARK S.12.1.** It should be noted that the condition  $p\mathbf{b}^{\top} \Sigma \mathbf{b} = O(1)$  actually implies that the variance of the target portfolio is of order  $O(1/p)$ . This situation is quite natural since the variance of the GMV portfolio is also typically at most  $1/p$ , similar to that of the equally weighted portfolio, represented as  $\mathbf{1}/p\mathbf{1}$ . On the other hand, if the rate of  $\mathbf{b}^{\top} \Sigma \mathbf{b}$  exceeds  $1/p$ , then  $\alpha_n^* \rightarrow 1$ , indicating that the target portfolio may be suboptimal.

### S.13. Additional results of numerical studies.

S.13.1. *Benchmarks: Precision matrix.* The following benchmark approaches are considered for the estimation of the precision matrix:

- **Empirical Bayes** ridge-type estimator of [Kubokawa and Srivastava \(2008\)](#) given by

$$(S.68) \quad \hat{\Pi}_{EBR} = p((n-1)\mathbf{S}_n + \text{tr}[\mathbf{S}_n]\mathbf{I}_p)^{-1}.$$

- **Optimal ridge** estimator of [Wang et al. \(2015\)](#) expressed as

$$(S.69) \quad \hat{\Pi}_{OR} = \hat{\alpha}_{OR;n} \left( \mathbf{S}_n + \hat{\beta}_{OR;n} \mathbf{I}_p \right)^{-1},$$

where  $\hat{\alpha}_{OR;n} = \hat{R}_{1,n}(\hat{\beta}_{OR;n}) / \hat{R}_{2,n}(\hat{\beta}_{OR;n})$ ,

$$\hat{R}_{1,n}(\lambda) = \frac{\hat{a}_{1;n}(\lambda)}{1 - c_n \hat{a}_{1;n}(\lambda)}, \quad \hat{R}_{2,n}(\lambda) = \frac{\hat{a}_{1;n}(\lambda)}{(1 - c_n \hat{a}_{1;n}(\lambda))^3} - \frac{\hat{a}_{2;n}(\lambda)}{(1 - c_n \hat{a}_{1;n}(\lambda))^4}$$

with  $\hat{a}_{1;n}(\lambda) = 1 - \frac{1}{p} \text{tr}[(\mathbf{S}_n/\lambda + \mathbf{I}_p)^{-1}]$ ,  $\hat{a}_{2;n}(\lambda) = \frac{1}{p} \text{tr}[(\mathbf{S}_n/\lambda + \mathbf{I}_p)^{-1}] - \frac{1}{p} \text{tr}[(\mathbf{S}_n/\lambda + \mathbf{I}_p)^{-2}]$ , and  $\hat{\beta}_{OR;n}$  minimizes  $L_{OR;n} = 1 - \hat{R}_{1,n}(\lambda)^2 / \hat{R}_{2,n}(\lambda)$ .

- **Inverse nonlinear shrinkage** estimator of the covariance matrix introduced in [Ledoit and Wolf \(2020\)](#). For  $i \in \{1, \dots, p\}$ , is given by

$$(S.70) \quad \mathbf{S}_{NLSh} = \mathbf{U} \text{diag}(d_1^{or}, \dots, d_p^{or}) \mathbf{U}^\top, \quad d_i^{or} = \begin{cases} \frac{d_i}{|1 - c - cd_i \check{m}_F(d_i)|^2}, & \text{if } d_i > 0, \\ \frac{1}{(c-1)\check{m}_F(0)}, & \text{if } d_i = 0, \end{cases}$$

where  $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_p)$  is the matrix with the sample eigenvectors of  $\mathbf{S}_n$ ,  $d_i$ ,  $i = 1, \dots$ , are the sample eigenvalues of  $\mathbf{S}_n$  and  $\check{m}_F(x) = \lim_{z \rightarrow x} m_F(z)$  with  $m_F(z)$  the limiting Stieltjes transform of the sample covariance matrix. A numerical approach to estimate  $\check{m}_F(x)$  is provided in [Ledoit and Wolf \(2020\)](#) and is available in the R-package *HDSHOP* (see [Bodnar et al. \(2024a\)](#)).

- **Oracle nonlinear shrinkage** estimator is derived for the loss function considered in [Ledoit and Wolf \(2021\)](#) and it is given by

$$(S.71) \quad \mathbf{S}_{oNLSh} = \mathbf{U} \text{diag}(\tilde{d}_1^{or}, \dots, \tilde{d}_p^{or}) \mathbf{U}^\top, \quad \tilde{d}_i^{or} = \frac{\mathbf{u}_i^\top \boldsymbol{\Sigma}^2 \mathbf{u}_i}{\mathbf{u}_i^\top \boldsymbol{\Sigma} \mathbf{u}_i}.$$

S.13.2. *Additional figures: Precision matrix.* In Figure S.5, we present the results of the simulation study obtained for the three suggested shrinkage estimators of the precision matrix and three benchmark approaches. The results of Figure S.5 complement the findings of Figure 3 by adding the values of the PRIAL computed for the empirical Bayes estimator and the inverse nonlinear shrinkage estimator. In Figure S.6, we report the average computational time of the most relevant estimators presented in Figure S.5 based on 100 repetitions. The MP estimator is used as a benchmark since it is essentially a plug-in estimator, requiring only the inversion of the nonzero eigenvalues. As expected, the MP shrinkage estimator performs best in terms of computational efficiency, being closest to the benchmark. The NL shrinkage estimator ranks second, while all ridge-type estimators are the most computationally demanding.

Figure S.7 depicts the kernel density estimators of the relative differences between the values of  $t$  which maximize  $\hat{L}_{R;2}^2(t)$  and  $L_{R;n,2}^2(t)$  in the case of the ridge estimator, and  $L_{MPR;n,2}^2(t)$  and  $\hat{L}_{MPR;2}^2(t)$  for the Moore-Penrose ridge estimator. The results are obtained when the elements of  $\mathbf{X}_n$  are drawn from the normal distribution and  $t$ -distribution with  $n \in \{100, 250\}$  and  $c_n \in \{1.1, 2, 4\}$ . We observe that the computed kernel densities are concentrated around zero in almost all of the considered cases, except for the extreme case when  $n = 100$  and  $c_n = 4$ . Furthermore, the kernel densities are skewed to the right with the exception when  $n = 250$  and  $c_n = 1.1$ . As such, it can be concluded that the optimization of the estimated loss function instead of the true one may lead to larger values of the tuning parameter  $t$ . However, as the sample size increases, the difference between the two values, as expected, becomes negligible. Finally, faster convergence is achieved when the matrix  $\mathbf{X}_n$  is drawn from the normal distribution.

S.13.3. *Benchmarks: Global minimum variance portfolio.* Three benchmark methods to estimate the weights of the global minimum variance portfolio are considered in Section 4.2:

- Traditional **sample estimator** is given by

$$(S.72) \quad \mathbf{w}_S = \frac{\mathbf{S}_n^+ \mathbf{1}}{\mathbf{1}^\top \mathbf{S}_n^+ \mathbf{1}},$$

where  $\mathbf{S}_n^+$  is the Moore-Penrose inverse of  $\mathbf{S}_n$ .

- **Reflexive inverse estimator** of [Bodnar et al. \(2018\)](#) is defined by

$$(S.73) \quad \mathbf{w}_{Ref} = \hat{\alpha}_{Ref}^* \mathbf{w}_S + (1 - \hat{\alpha}_{Ref}^*) \mathbf{b},$$

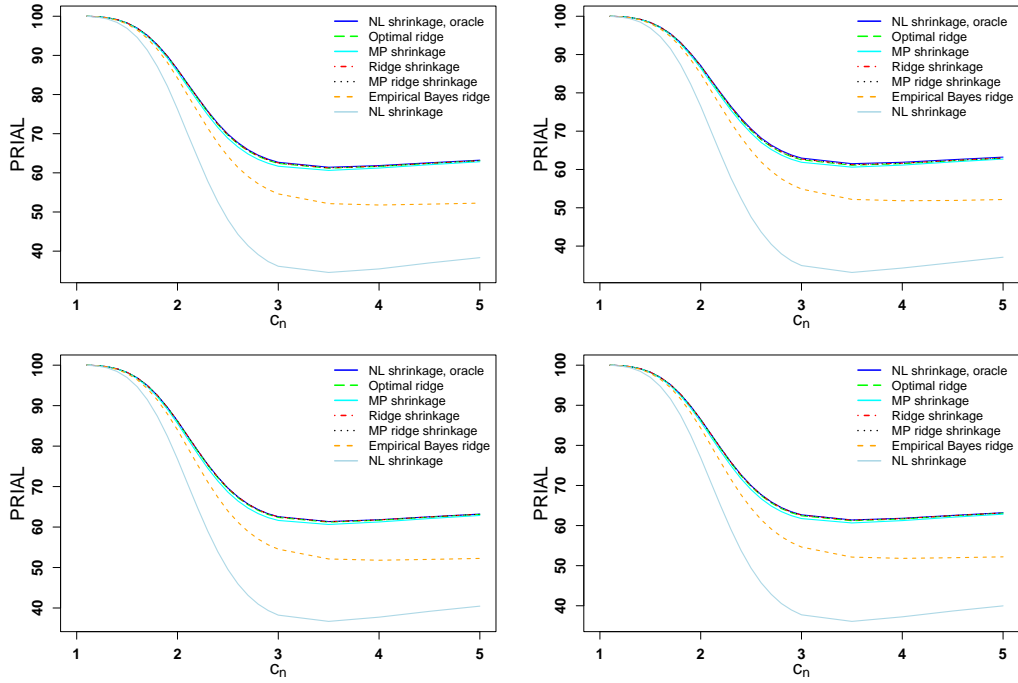


FIG S.5. *PRIAL* for  $c_n \in (1, 5]$ ,  $n = 100$  (first row) and  $n = 250$  (second row) when the elements of  $\mathbf{X}_n$  are drawn from the normal distribution (first column) and scale  $t$ -distribution (second column).

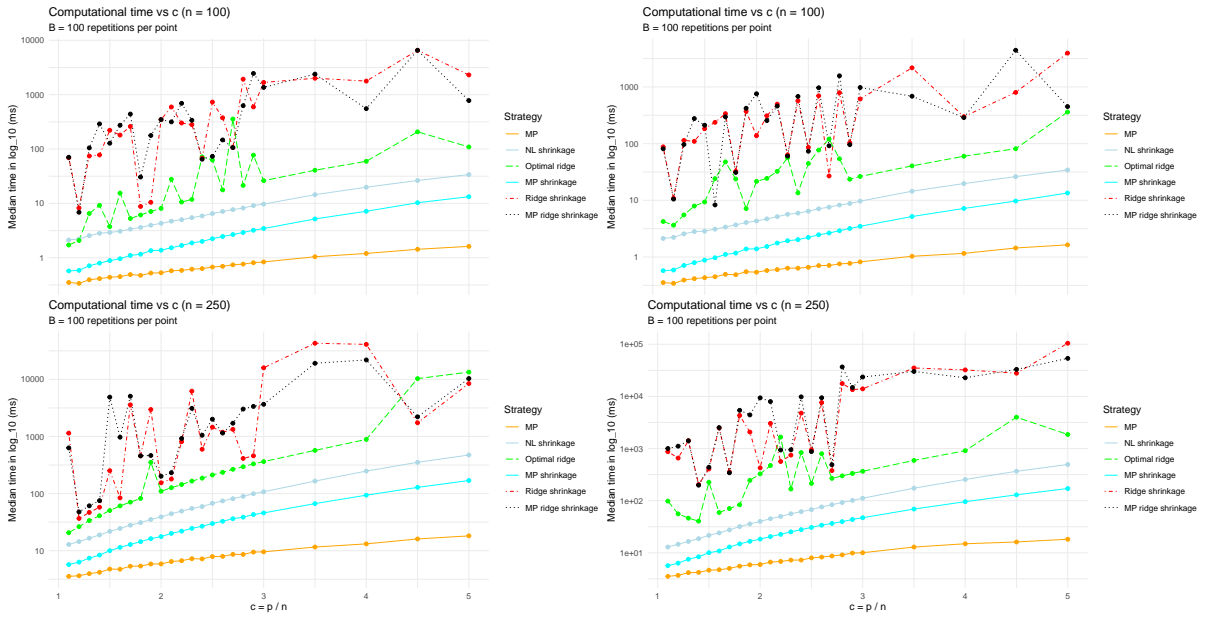


FIG S.6. Computational time (in logarithmic scale) for precision matrix estimators for  $c_n \in (1, 5]$ ,  $n = 100$  (first row) and  $n = 250$  (second row) when the elements of  $\mathbf{X}_n$  are drawn from the normal distribution (first column) and scale  $t$ -distribution (second column).

with

$$\hat{\alpha}_{Ref}^* = \frac{(p/n - 1)\hat{R}_{Ref}}{(p/n - 1)^2 + p/n + (p/n - 1)\hat{R}_{Ref}}$$

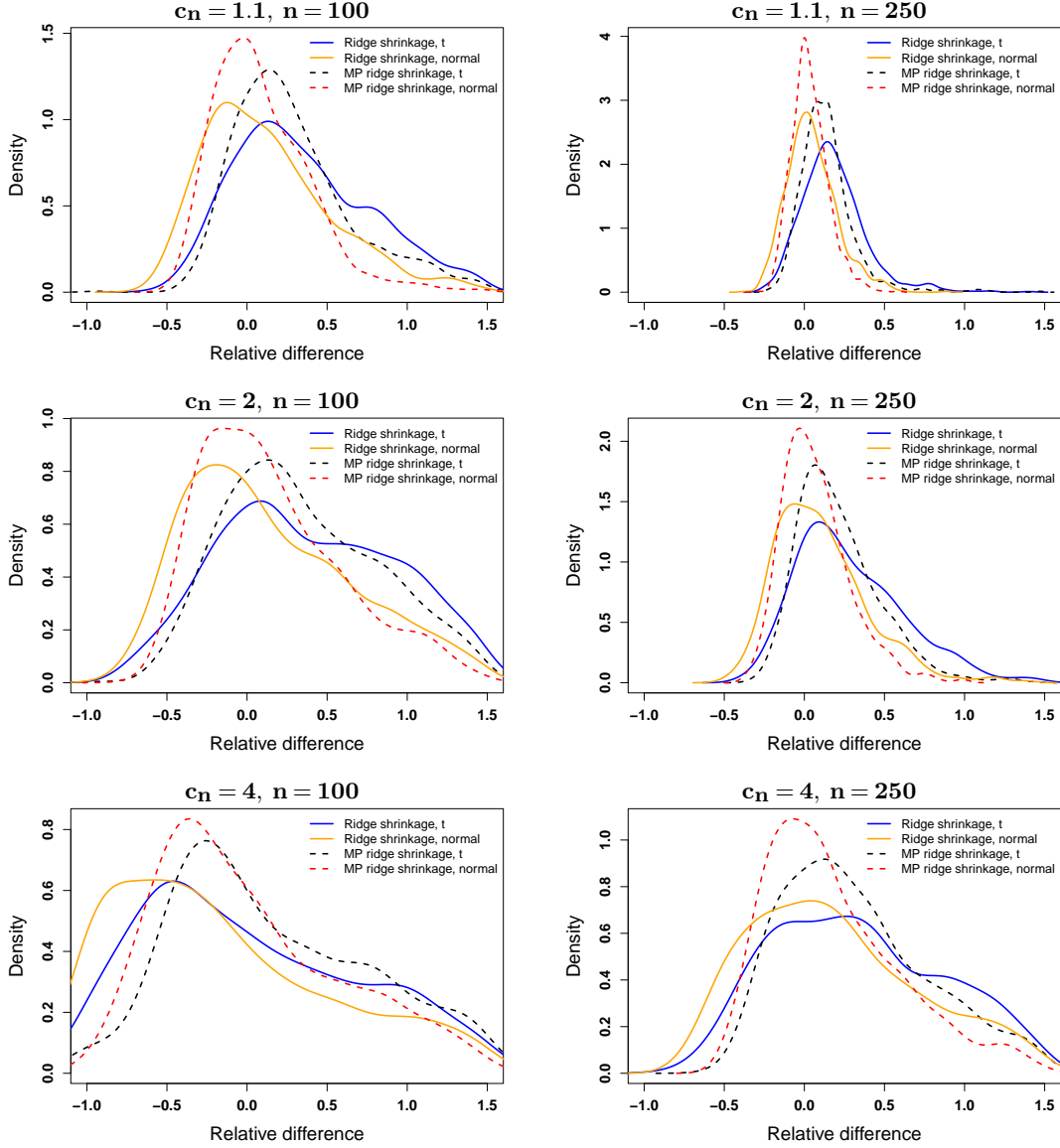


FIG S.7. Relative differences of  $t$ 's that maximize  $L_{R;n,2}^2(t)$  ( $L_{MPR;n,2}^2(t)$ ) and  $\hat{L}_{R;2}^2(t)$  ( $\hat{L}_{MPR;2}^2(t)$ ) for  $c_n \in \{1.1, 2, 4\}$ ,  $n = 100$  and  $n = 250$  when the elements of  $\mathbf{X}_n$  are drawn from the normal distribution and scale  $t$ -distribution.

where

$$\hat{R}_{Ref} = \frac{p}{n} \left( \frac{p}{n} - 1 \right) \mathbf{b}^\top \mathbf{S}_n \mathbf{b} \cdot \mathbf{1}^\top \mathbf{S}_n^+ \mathbf{1} - 1$$

In the simulation study, we use the equally weighted portfolio as the target portfolio  $\mathbf{b}$ , i.e.,  $\mathbf{b} = \frac{1}{p} \mathbf{1}$ .

- **Double shrinkage estimator** or **ridge shrinkage estimator** of Bodnar et al. (2024b) after some simplifications is expressed as

$$(S.74) \quad \mathbf{w}_{Rid} = \hat{\alpha}_{Rid}^*(\hat{\eta}) \frac{\mathbf{S}_n^-(\hat{\eta}) \mathbf{1}}{\mathbf{1}^\top \mathbf{S}_n^-(\hat{\eta}) \mathbf{1}} + (1 - \hat{\alpha}_{Rid}^*(\hat{\eta})) \mathbf{b},$$

with

$$\hat{\alpha}_{Rid}^*(\hat{\eta}) = \frac{1 - \frac{(1+\hat{\eta})^{-1}}{\mathbf{b}^\top \mathbf{S}_n \mathbf{b}} \frac{d_1(\hat{\eta})}{\mathbf{1}^\top \mathbf{S}_n^-(\hat{\eta}) \mathbf{1}}}{1 - \frac{2(1+\hat{\eta})^{-1}}{\mathbf{b}^\top \mathbf{S}_n \mathbf{b}} \frac{d_1(\hat{\eta})}{\mathbf{1}^\top \mathbf{S}_n^-(\hat{\eta}) \mathbf{1}} + \frac{(1+\hat{\eta})^{-2}}{\mathbf{b}^\top \mathbf{S}_n \mathbf{b}} \frac{(1-\hat{v}_2^{(1)}(\hat{\eta}, 0))d_2(\hat{\eta})}{(\mathbf{1}^\top \mathbf{S}_n^-(\hat{\eta}) \mathbf{1})^2}}$$

where

$$\begin{aligned} d_1(\eta) &= \frac{(1+\eta)}{\hat{v}(\eta, 0)} \left( 1 - \eta \mathbf{b}^\top \mathbf{S}_n^-(\eta) \mathbf{1} \right), \\ d_2(\eta) &= \frac{(1+\eta)^2}{\hat{v}(\eta, 0)} \left( \mathbf{1}^\top \mathbf{S}_n^-(\eta) \mathbf{1} - \eta \mathbf{1}^\top (\mathbf{S}_n^-(\eta))^2 \mathbf{1} \right), \\ \hat{v}(\eta, 0) &= 1 - \frac{p}{n} \left( 1 - \eta \frac{1}{p} \text{tr}(\mathbf{S}_n^-(\eta)) \right), \\ \hat{v}_1^{(1)}(\eta, 0) &= \hat{v}(\eta, 0) \frac{p}{n} \left( \frac{1}{p} \text{tr}(\mathbf{S}_n^-(\eta)) - \eta \frac{1}{p} \text{tr}((\mathbf{S}_n^-(\eta))^2) \right), \\ \hat{v}_2^{(1)}(\eta, 0) &= 1 - \frac{1}{\hat{v}(\eta, 0)} + \eta \frac{\hat{v}_1^{(1)}(\eta, 0)}{\hat{v}(\eta, 0)^2}, \end{aligned}$$

and  $\hat{\eta}$  maximizes of the following function

$$\hat{L}_{n;2}(\eta) = \frac{\left( 1 - \frac{(1+\eta)^{-1}}{\mathbf{b}^\top \mathbf{S}_n \mathbf{b}} \frac{d_1(\eta)}{\mathbf{1}^\top \mathbf{S}_n^-(\eta) \mathbf{1}} \right)^2}{1 - \frac{2(1+\eta)^{-1}}{\mathbf{b}^\top \mathbf{S}_n \mathbf{b}} \frac{d_1(\eta)}{\mathbf{1}^\top \mathbf{S}_n^-(\eta) \mathbf{1}} + \frac{(1+\eta)^{-2}}{\mathbf{b}^\top \mathbf{S}_n \mathbf{b}} \frac{(1-\hat{v}_2^{(1)}(\eta, 0))d_2(\eta)}{(\mathbf{1}^\top \mathbf{S}_n^-(\eta) \mathbf{1})^2}}.$$

In the simulation study, the target portfolio  $\mathbf{b}$  is set to be equal to the equally weighted portfolio.

S.13.4. *Additional figures: Global minimum variance portfolio.* In Figure S.8, we present the results of the simulation study for the proposed MP shrinkage estimator of the GMV portfolio weights alongside three benchmark approaches. The results in Figure S.8 complement those in Figure 4 by including the rOSV values computed for the traditional GMV portfolio estimator obtained by directly plugging in the Moore-Penrose inverse (see (S.72)). Figure S.9 complements the findings from Figure S.8 by presenting the corresponding computational time.

## REFERENCES

- Ahlfors, L. V. (1953). *Complex analysis : an introduction to the theory of analytic functions of one complex variable*. International series in pure and applied mathematics. McGraw-Hill, New York.
- Ao, M., Yingying, L., and Zheng, X. (2019). Approaching mean-variance efficiency for large portfolios. *The Review of Financial Studies*, 32(7):2890–2919.
- Bai, Z., Fahey, G., and Golub, G. (1996). Some large-scale matrix computation problems. *Journal of Computational and Applied Mathematics*, 74(1):71–89.
- Bai, Z. and Silverstein, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*. Springer.
- Bai, Z. D. and Silverstein, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *The Annals of Probability*, 32(1A):553 – 605.
- Bell, E. T. (1927-1928). Partition polynomials. *Annals of Mathematics*, 29(1/4):38–46.
- Bell, E. T. (1934). Exponential polynomials. *Annals of Mathematics*, 35(2):258–277.
- Ben-Israel, A. and Greville, T. N. (2003). *Generalized Inverses: Theory and Applications*. Springer.
- Bodnar, T., Dette, H., and Parolya, N. (2016a). Spectral analysis of the moore–penrose inverse of a large dimensional sample covariance matrix. *Journal of Multivariate Analysis*, 148:160 – 172.

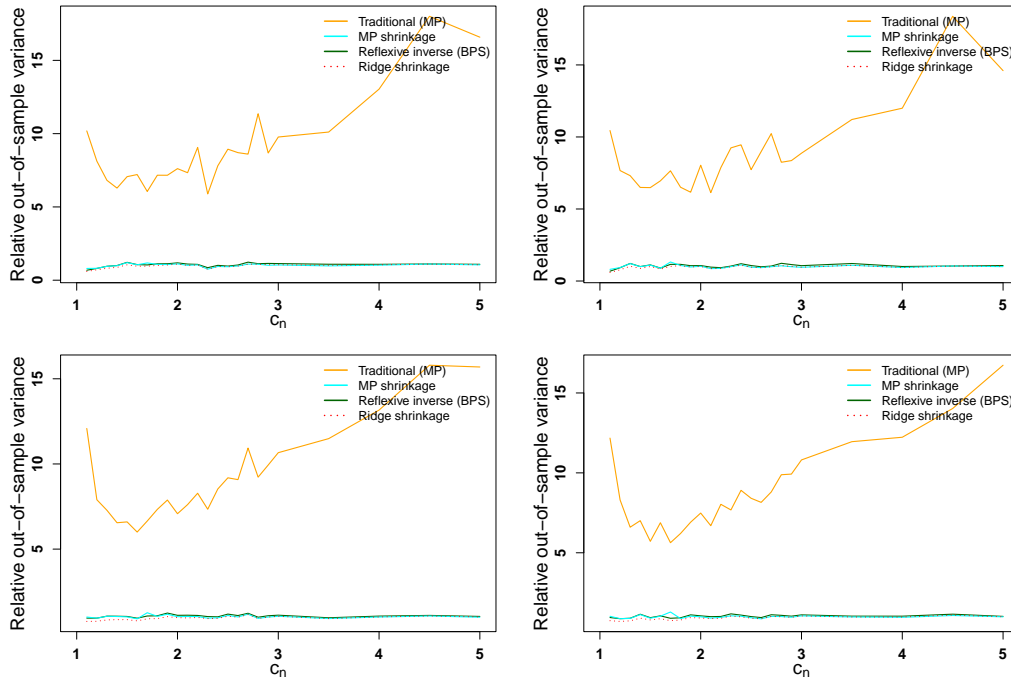


FIG S.8.  $rOSV$  for  $c_n \in (1, 5]$ ,  $n = 100$  (first row) and  $n = 250$  (second row) when the elements of  $\mathbf{X}_n$  are drawn from the normal distribution (first column) and scale  $t$ -distribution (second column).

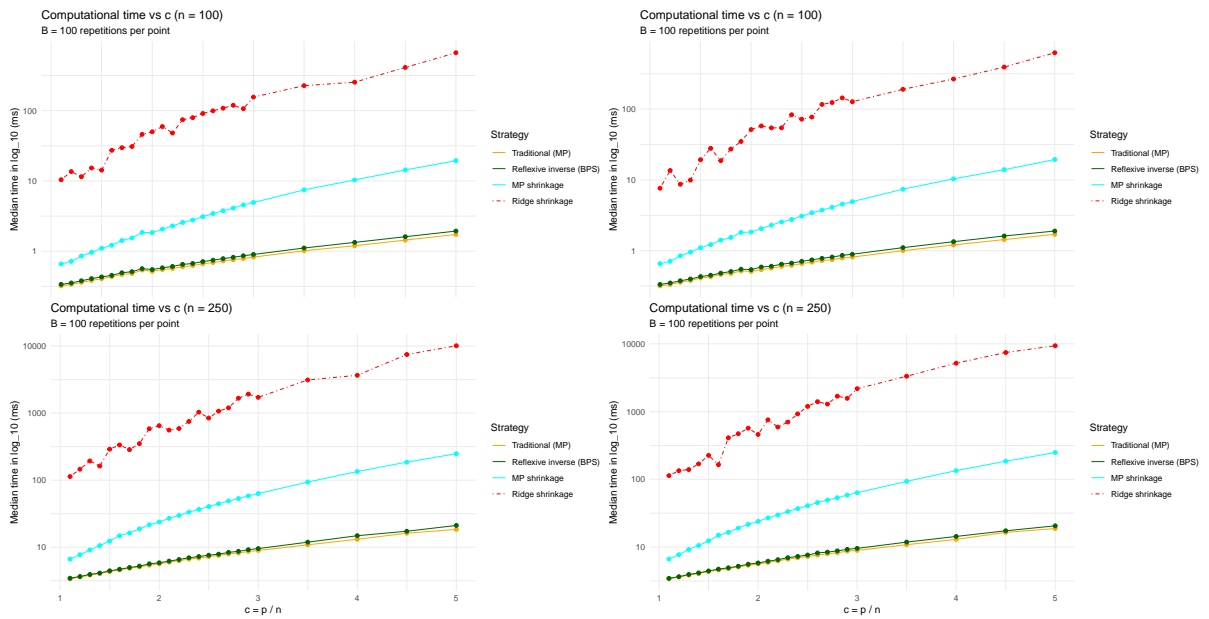


FIG S.9. Computational time (in logarithmic scale) for GMV estimators for  $c_n \in (1, 5]$ ,  $n = 100$  (first row) and  $n = 250$  (second row) when the elements of  $\mathbf{X}_n$  are drawn from the normal distribution (first column) and scale  $t$ -distribution (second column).

Bodnar, T., Dette, H., and Parolya, N. (2019). Testing for independence of large dimensional vectors. *The Annals of Statistics*, 47(5):2977–3008.

- Bodnar, T., Dmytriv, S., Okhrin, Y., Otryakhin, D., and Parolya, N. (2024a). *HDSHOP: High-Dimensional Shrinkage Optimal Portfolios*. R package version 0.1.5.
- Bodnar, T., Gupta, A. K., and Parolya, N. (2014a). On the strong convergence of the optimal linear shrinkage estimator for large dimensional covariance matrix. *Journal of Multivariate Analysis*, 132:215–228.
- Bodnar, T., Gupta, A. K., and Parolya, N. (2014b). On the strong convergence of the optimal linear shrinkage estimator for large dimensional covariance matrix. *Journal of Multivariate Analysis*, 132:215–228.
- Bodnar, T., Gupta, A. K., and Parolya, N. (2016b). Direct shrinkage estimation of large dimensional precision matrix. *Journal of Multivariate Analysis*, 146:223–236.
- Bodnar, T. and Okhrin, Y. (2008). Properties of the singular, inverse and generalized inverse partitioned Wishart distributions. *Journal of Multivariate Analysis*, 99(10):2389–2405.
- Bodnar, T., Okhrin, Y., and Parolya, N. (2023a). Optimal shrinkage-based portfolio selection in high dimensions. *Journal of Business & Economic Statistics*, 41(1):140–156.
- Bodnar, T. and Parolya, N. (2025). Supplement to “Reviving pseudo-inverses: Asymptotic properties of large dimensional Moore-Penrose and ridge-type inverses with applications”.
- Bodnar, T., Parolya, N., and Schmid, W. (2018). Estimation of the global minimum variance portfolio in high dimensions. *European Journal of Operational Research*, 266(1):371–390.
- Bodnar, T., Parolya, N., and Thorsén, E. (2023b). Is the empirical out-of-sample variance an informative risk measure for the high-dimensional portfolios? *Finance Research Letters*, 54:103807.
- Bodnar, T., Parolya, N., and Thorsén, E. (2024b). Two is better than one: regularized shrinkage of large minimum variance portfolios. *Journal of Machine Learning Research*, 25(173):1–32.
- Cai, T. T., Hu, J., Li, Y., and Zheng, X. (2020). High-dimensional minimum variance portfolio estimation based on high-frequency data. *Journal of Econometrics*, 214(2):482–494.
- Cai, T. T. and Jiang, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Annals of Statistics*, 39(3):1496–1525.
- Chen, S. X., Zhang, L.-X., and Zhong, P.-S. (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association*, 105(490):810–819.
- Cook, R. D. and Forzani, L. (2011). On the mean and variance of the generalized inverse of a singular Wishart matrix. *Electronic Journal of Statistics*, 5:146–158.
- Cvijović, D. (2011). New identities for the partial Bell polynomials. *Applied Mathematics Letters*, 24(9):1544–1547.
- Di Nardo, E., Guarino, G., and Senato, D. (2008). A unifying framework for k-statistics, polykays and their multivariate generalizations. *Bernoulli*, 14:440–468.
- Feng, Y. and Palomar, D. P. (2016). *A Signal Processing Perspective on Financial Engineering*. now Publishers Inc., Boston and Delft.
- Frahm, G. and Memmel, C. (2010). Dominating estimators for minimum-variance portfolios. *Journal of Econometrics*, 159:289–302.
- Golosnoy, V. and Okhrin, Y. (2007). Multivariate shrinkage for optimal portfolio weights. *The European Journal of Finance*, 13(5):441–458.
- Golub, G. H. and Strakoš, Z. (1994). Estimates in quadratic formulas. *Numerical Algorithms*, 8:241–268.
- Haff, L. (1979). An identity for the Wishart distribution with applications. *Journal of Multivariate Analysis*, 9(4):531–544.
- Harville, D. A. (1997). *Matrix Algebra from a Statistician’s Perspective*. Springer.
- Heiny, J. (2019). Random matrix theory for heavy-tailed time series. *Journal of Mathematical Sciences*, 237:652–666.
- Heiny, J. and Mikosch, T. (2021). Large sample autocovariance matrices of linear processes with heavy tails. *Stochastic Processes and their Applications*, 141:344–375.
- Heiny, J. and Yao, J. (2022). Limiting distributions for eigenvalues of sample correlation matrices from heavy-tailed populations. *The Annals of Statistics*, 50(6):3249–3280.
- Hille, E. (2002). *Analytic Function Theory*, volume 2. American Mathematical Soc.
- Horn, R. A. and Johnson, C. R. (1985). *Matrix Analysis*. Cambridge University Press, Cambridge.
- Imori, S. and von Rosen, D. (2020). On the mean and dispersion of the Moore-Penrose generalized inverse of a Wishart matrix. *The Electronic Journal of Linear Algebra*, 36:124–133.
- Kan, R., Wang, X., and Zhou, G. (2022). Optimal portfolio choice with estimation risk: No risk-free asset case. *Management Science*, 68(3):2047–2068.
- Krantz, S. G. and Parks, H. R. (2002). *A Primer of Real Analytic Functions*. Springer Science & Business Media.
- Krishnamoorthy, K. and Gupta, A. K. (1989). Improved minimax estimation of a normal precision matrix. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 17(1):91–102.
- Kubokawa, T. and Srivastava, M. S. (2008). Estimation of the precision matrix of a singular Wishart distribution and its application in high-dimensional data. *Journal of Multivariate Analysis*, 99(9):1906–1928.

- Lam, C. (2020). High-dimensional covariance matrix estimation. *WIREs Computational Statistics*, 12(2):e1485.
- Lassance, N., Vanderveken, R., and Vrins, F. (2024). On the combination of naive and mean-variance portfolio strategies. *Journal of Business & Economic Statistics*, 42(3):875–889.
- Ledoit, O. and Péché, S. (2011). Eigenvectors of some large sample covariance matrix ensembles. *Probability Theory and Related Fields*, 151(1-2):233–264.
- Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88:365–411.
- Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Annals of Statistics*, 40:1024–1060.
- Ledoit, O. and Wolf, M. (2020). Analytical nonlinear shrinkage of large-dimensional covariance matrices. *The Annals of Statistics*, 48(5):3043 – 3065.
- Ledoit, O. and Wolf, M. (2021). Shrinkage estimation of large covariance matrices: Keep it simple, statistician? *Journal of Multivariate Analysis*, 186:104796.
- Ledoit, O. and Wolf, M. (2022). Quadratic shrinkage for large covariance matrices. *Bernoulli*, 28(3):1519 – 1547.
- Lütkepohl, H. (1996). *Handbook of Matrices*. John Wiley & Sons.
- Meyer, Jr, C. D. (1973). Generalized inversion of modified matrices. *SIAM Journal on Applied Mathematics*, 24(3):315–323.
- Muirhead, R. (1982). *Aspects of Multivariate Statistical Theory*. New York: Wiley.
- Pan, G. (2014). Comparison between two types of large sample covariance matrices. *Annales de l’IHP Probabilités et Statistiques*, 50(2):655–677.
- Penrose, R. (1955). A generalized inverse for matrices. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 51, pages 406–413. Cambridge University Press.
- Rao, C. R. and Mitra, S. K. (1972). Generalized inverse of a matrix and its applications. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Theory of Statistics*, volume 6, pages 601–621. University of California Press.
- Rencher, A. and Christensen, W. (2012). *Methods of Multivariate Analysis*. Wiley.
- Rubio, F. and Mestre, X. (2011). Spectral convergence for a general class of random matrices. *Statistics & Probability Letters*, 81(5):592–602.
- Rubio, F., Mestre, X., and Palomar, D. P. (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.
- Rudin, W. (1987). *Real and complex analysis*. McGraw-Hill, Inc.
- Shi, H., Hallin, M., Drton, M., and Han, F. (2022). On universally consistent and fully distribution-free rank tests of vector independence. *The Annals of Statistics*, 50(4):1933–1959.
- Srivastava, M. S. (2003). Singular Wishart and multivariate beta distributions. *The Annals of Statistics*, 31(5):1537–1560.
- Wang, C., Pan, G., Tong, T., and Zhu, L. (2015). Shrinkage estimation of large dimensional precision matrix using random matrix theory. *Statistica Sinica*, 25:993–1008.
- Wang, G., Wei, Y., Qiao, S., Lin, P., and Chen, Y. (2018). *Generalized Inverses: Theory and Computations*. Springer.
- Yang, R. and Berger, J. O. (1994). Estimation of a covariance matrix using the reference prior. *The Annals of Statistics*, 22(3):1195 – 1211.
- Yao, J., Zheng, S., and Bai, Z. (2015). *Sample Covariance Matrices and High-Dimensional Data Analysis*. Cambridge University Press Cambridge.