Linear shrinkage-based hypothesis test for large-dimensional covariance matrix

Bodnar, Taras; Parolya, Nestor; Veldman, Frederik

DOI
10.1007/978-3-031-69111-9_12

Publication date
2024

Document Version
Final published version

Published in
Advanced Statistical Methods in Process Monitoring, Finance, and Environmental Science

Linear Shrinkage-Based Hypothesis Test for Large-Dimensional Covariance Matrix

Taras Bodnar, Nestor Parolya, and Frederik Veldman

Abstract The chapter is concerned with finding the asymptotic distribution of the estimated shrinkage intensity used in the definition of the linear shrinkage estimator of the covariance matrix, derived by Bodnar et al. (J Multivar Anal 132:215–228, 2014). As a result, a new test statistic is proposed which is deduced from the linear shrinkage estimator. This result is a ready-to-use multivariate hypothesis test in the large-dimensional asymptotic framework and constitutes the main result of the chapter. The theoretical findings are compared by means of a simulation study with existing tests, in particular with the commonly used corrected likelihood ratio test and the corrected John test, both derived by Wang and Yao (Electron J Stat 7:2164–2192, 2013).

1 Introduction

Estimating and testing the structure of the covariance matrix are import problems that have many applications in different fields of science, especially in economics and finance. For example, the covariance matrix plays a significant role in portfolio theory (see Markowitz 1952), where it is important to understand the relation and the variability of different assets included in a portfolio.

Challenging problems arise for estimating and testing the structure of the covariance matrix when the dimension p of the random sample $(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n)$

T. Bodnar
Department of Management and Engineering, Linköping University, Linköping, Sweden
e-mail: taras.bodnar@liu.se

N. Parolya (✉)
Delft Institute of Applied Mathematics, Delft University of Technology, Delft, The Netherlands
e-mail: n.parolya@tudelft.nl

F. Veldman
Delft University of Technology, Delft, The Netherlands
e-mail: f.j.g.veldman@student.tudelft.nl

S. Knoth et al. (eds.), *Advanced Statistical Methods in Process Monitoring, Finance, and Environmental Science*,
https://doi.org/10.1007/978-3-031-69111-9_12

is of similar order of even larger than the sample size n, where $\mathbf{y}_i \in \mathbb{R}^p$ for all $i \in \{0, \cdots, n\}$. This is because commonly used estimators and tests, such as the likelihood ratio test (see Anderson 1984) or John's test (see John 1971), are constructed under the assumption that the dimension p of a random sample stays fixed. However, it has been pointed out by numerous authors (see Wang and Yao (2013); Paul and Aue (2014); Yao et al. (2015); Bodnar et al. (2019), among others) that the assumption of a fixed dimension does not yield precise distributional approximations for commonly used statistics and that better approximations can be obtained considering the dimension p go to infinity as well. This leads to a new area in asymptotic statistics where the dimension p is no longer fixed, but tends to infinity together with the sample size n, i.e., $\frac{p}{n} \to c < \infty$ when $n \to \infty$ and $p \to \infty$. This framework is called the large-dimensional asymptotics (see Bai and Silverstein, 2010).

Many statistical tools that used to work for a fixed dimension did not work properly anymore in the large-dimensional asymptotic framework and needed to be altered. Moreover, new statistical tests have been developed and are still being developed. For instance, the likelihood ratio test and John's test are extended by Wang and Yao (2013) to work properly in the large-dimensional case. In addition, new estimators are proposed for the population covariance matrix because it is well known that the commonly used sample covariance matrix does not perform well and produces large errors in the large-dimensional case. To tackle this issue, new estimators are developed to better approximate the actual population covariance matrix in this setting. One of the commonly used approaches is the shrinkage estimator originally proposed by Stein (1956) for the mean vector of a normal distribution and extended in different directions. One of the most important among the existing estimators of the covariance matrix is the linear shrinkage estimator developed by Ledoit and Wolf (2004), further improved and generalized by Bodnar et al. (2014).

For numerous applications in finance one should perform a hypothesis test on large-dimensional covariance matrices. For example, one can test whether a large number of assets in a portfolio can be assumed to have the same risk profile or testing if the assets can be assumed to be independent of each other based on a large set of historical returns. The latter corresponds to testing the null hypothesis $H_0 : \boldsymbol{\Sigma}_n = \xi^2 \mathbf{I}$ versus the alternative hypothesis $H_1 : \boldsymbol{\Sigma}_n \neq \xi^2 \mathbf{I}$ for some $\xi^2 > 0$, where $\boldsymbol{\Sigma}_n$ is the population covariance matrix of asset returns and $\mathbf{I}$ the identity matrix.[1] This type of tests is called the sphericity test. Similarly, it can be of interest to explore if a large set of assets have a certain dependence structure, thus testing whether the population covariance matrix can be assumed to be equal to a prior

[1] The subindex n by the population covariance matrix $\boldsymbol{\Sigma}_n$ arises because of the large-dimensional framework, i.e., $p/n \to c > 0$. Here it is implicitly assumed that the dimension p is a function of the sample size n, namely, $p \equiv p(n)$.

believe $\boldsymbol{\Sigma}_0$ of the population covariance matrix. This corresponds to testing the hypotheses

$$H_0 : \boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}_0 \quad \text{against} \quad H_1 : \boldsymbol{\Sigma}_n \neq \boldsymbol{\Sigma}_0 \,. \tag{1}$$

Such a hypothesis test is an important tool in multivariate statistics and a lot of research has been conducted to extend the classical multivariate tests to work under the large-dimensional framework. This is also the goal of this chapter, namely, to construct a new hypothesis test for the population covariance matrix in the large-dimensional asymptotic framework based on the linear shrinkage estimator.

The new test that will be constructed in the chapter will be based on the linear shrinkage estimator. This is because it has been pointed out by Bodnar et al. (2014) that the linear shrinkage estimator performs very well in estimating the population covariance matrix in the large-dimensional case. This makes it interesting to explore whether linear shrinkage-based hypothesis tests also perform well, in terms of size and power, in the large-dimensional asymptotic framework.

Before the new test will be presented at the end of Sect. 2, we discuss the linear shrinkage estimator of the covariance matrix proposed in Bodnar et al. (2014) and derive the asymptotic distribution of the estimated shrinkage intensities. After this theoretical part, a simulation study to assess the finite-sample performance of the new test and to compare it with the already existing approaches will be conducted in Sect. 3. The already existing tests that will be used in the simulation study are the corrected likelihood ratio test (CLRT) and the corrected John test (CJ), both derived by Wang and Yao (2013). The test of Ledoit and Wolf (2002) will not be considered in the chapter because it is less powerful than the CJ test and therefore irrelevant to use in the comparison. The comparison will be made using the empirical size and the empirical power. Final remarks are provided in Sect. 4.

2 New Test Based on the Shrinkage Approach

In this section we present the linear shrinkage estimator of the covariance matrix and develop the framework used in the derivation of the test statistic and its asymptotic distribution.

2.1 Linear Shrinkage Estimator of the Covariance Matrix

Let $\mathbf{y} \in \mathbb{R}^p$ be a random vector of dimension p and let n denote the sample size of a random sample $(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n)$. Assume that $c_n = \frac{p}{n} \to c \in (0, +\infty)$ when $n \to \infty$ and $p \to \infty$. The data matrix $\mathbf{Y}_n = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n]$ is assumed to be a realization from the following stochastic model:

$$\mathbf{Y}_n = \boldsymbol{\Sigma}_n^{1/2} \mathbf{X}_n. \tag{2}$$

In (2), $\boldsymbol{\Sigma}_n$ denotes the population covariance matrix and the matrix $\mathbf{X}_n$ explains the randomness of the model. Since under the null hypothesis in (1) we have that $\boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}_0$, we rescale the data matrix $\mathbf{Y}_n$ by $\widetilde{\mathbf{Y}}_n = \boldsymbol{\Sigma}_0^{-1/2}\mathbf{Y}_n$. In addition, the following assumptions on the stochastic model (2) are imposed:

- **(A1)** The population covariance matrix $\boldsymbol{\Sigma}_n$ is a nonrandom p-dimensional positive definite matrix.
- **(A2)** $\mathbf{X}_n$ is an p $\times$ n matrix where the entries are independent and identically distributed (i.i.d.) random variables with mean zero, unit variance, and finite fourth moment equal to $E[|x_{i,j}|^4] = \beta + 1 + \kappa < \infty$, where $\kappa = 2$ in case of real variables and $\kappa = 1$ in case of complex variables, and also $E[x_{i,j}^2] = 0$ in case of complex variables.
- **(A3)** One only observes $\mathbf{Y}_n$ where $\mathbf{Y}_n = \boldsymbol{\Sigma}_n^{\frac{1}{2}}\mathbf{X}_n$.

These assumptions are maintained throughout this chapter. We further assume that $\mathbf{X}_n \in \mathbb{R}^{pxn}$; therefore, $\kappa = 2$. However, to keep the results as general as possible we present the results as a function of κ.

The sample covariance matrix is defined by

$$\mathbf{S}_n = \frac{1}{n}\mathbf{Y}_n\mathbf{Y}_n^T.$$

It is the most commonly used estimator of the covariance matrix which possesses nice distributional properties in the classical asymptotic regime. However, that is no longer the case in the large-dimensional setting and improved estimators are suggested in the literature.

The linear shrinkage estimator presents a widely spread approach for the estimation of the population covariance matrix in the large-dimensional framework. This estimator was first derived in Ledoit and Wolf (2004) and improved by Bodnar et al. (2014). The general linear shrinkage estimator $\hat{\boldsymbol{\Sigma}}_{GLSE}$ is expressed as

$$\hat{\boldsymbol{\Sigma}}_{GLSE} = \alpha_n\mathbf{S}_n + \beta_n\boldsymbol{\Omega}, \tag{3}$$

where $\boldsymbol{\Omega}$ is the shrinkage target which is assumed to be a matrix with bounded trace norm. The parameters α_n and β_n are called the shrinkage intensities because they basically shrink the matrices which they are multiplied with. Thus, $\hat{\boldsymbol{\Sigma}}_{GLSE}$ is essentially a linear combination between the sample covariance matrix $\mathbf{S}_n$ and the prior belief of the population covariance matrix $\boldsymbol{\Omega}$.

The optimal shrinkage intensities are found by minimizing the squared loss function given by

$$L_f^2 = ||\hat{\boldsymbol{\Sigma}}_{GLSE} - \boldsymbol{\Sigma}_n||_F^2,$$

where $||.||_F^2$ is the squared Frobenius norm. The loss function L_f^2 measures the distance between the estimator $\hat{\boldsymbol{\Sigma}}_{GLSE}$ and the population covariance matrix $\boldsymbol{\Sigma}_n$.

For the estimator to be working properly, this distance, thus the loss function, should be as small as possible. Bodnar et al. (2014) minimize this loss function L_f^2 and found that the optimal shrinkage estimators are equal to

$$\alpha_n^*(\boldsymbol{\Omega}) = \frac{\mathrm{tr}(\mathbf{S}_n\boldsymbol{\Sigma}_n)||\boldsymbol{\Omega}||_F^2 - \mathrm{tr}(\boldsymbol{\Sigma}_n\boldsymbol{\Omega})\mathrm{tr}(\mathbf{S}_n\boldsymbol{\Omega})}{||\mathbf{S}_n||_F^2||\boldsymbol{\Omega}||_F^2 - \left(\mathrm{tr}(\mathbf{S}_n\boldsymbol{\Omega})\right)^2}, \tag{4}$$

$$\beta_n^*(\boldsymbol{\Omega}) = \frac{\mathrm{tr}(\boldsymbol{\Sigma}_n\boldsymbol{\Omega})||\mathbf{S}_n||_F^2 - \mathrm{tr}(\mathbf{S}_n\boldsymbol{\Sigma}_n)\mathrm{tr}(\mathbf{S}_n\boldsymbol{\Omega})}{||\mathbf{S}_n||_F^2||\boldsymbol{\Omega}||_F^2 - \left(\mathrm{tr}(\mathbf{S}_n\boldsymbol{\Omega})\right)^2}. \tag{5}$$

Corollary 3.1 in Bodnar et al. (2014) presents the deterministic asymptotic equivalents to $\alpha_n^*(\boldsymbol{\Omega})$ and $\beta_n^*(\boldsymbol{\Omega})$ given by

$$\alpha^*(\boldsymbol{\Omega}) = 1 - \frac{\frac{c}{p}||\boldsymbol{\Sigma}_n||_{tr}^2||\boldsymbol{\Omega}||_F^2}{\left(||\boldsymbol{\Sigma}_n||_F^2 + \frac{c}{p}||\boldsymbol{\Sigma}_n||_{tr}^2\right)||\boldsymbol{\Omega}||_F^2 - \left(\mathrm{tr}(\boldsymbol{\Sigma}_n\boldsymbol{\Omega})\right)^2}, \tag{6}$$

$$\beta^*(\boldsymbol{\Omega}) = \frac{\mathrm{tr}(\boldsymbol{\Sigma}_n\boldsymbol{\Omega})}{||\boldsymbol{\Omega}||_F^2}\left(1 - \alpha^*(\boldsymbol{\Omega})\right), \tag{7}$$

where $||\mathbf{A}||_{tr}$ is the trace norm of matrix $\mathbf{A}$, while their consistent estimators in the large-dimensional framework are expressed as (see Section 4 in Bodnar et al. (2014))

$$\hat{\alpha}^*(\boldsymbol{\Omega}) = 1 - \frac{\frac{1}{n}||\mathbf{S}_n||_{tr}^2||\boldsymbol{\Omega}||_F^2}{||\mathbf{S}_n||_F^2||\boldsymbol{\Omega}||_F^2 - \left(\mathrm{tr}(\mathbf{S}_n\boldsymbol{\Omega})\right)^2}, \tag{8}$$

$$\hat{\beta}^*(\boldsymbol{\Omega}) = \frac{\mathrm{tr}(\mathbf{S}_n\boldsymbol{\Omega})}{||\boldsymbol{\Omega}||_F^2}\left(1 - \hat{\alpha}^*(\boldsymbol{\Omega})\right). \tag{9}$$

2.2 Test Statistics and Its Large-Dimensional Distribution Under the Null Hypothesis

Since the aim of the chapter is to derive a new statistical test for the hypotheses in (1) and, as such, the null distribution of the test statistic is needed for the decision rule, we consider the sample covariance matrix for the normalized sample $\widetilde{\mathbf{Y}}_n = \boldsymbol{\Sigma}_n^{-\frac{1}{2}}\mathbf{Y}_n$ defined by

$$\widetilde{\mathbf{S}}_n = \frac{1}{n}\widetilde{\mathbf{Y}}_n\widetilde{\mathbf{Y}}_n^T,$$

while the population covariance matrix corresponding to the sample $\widetilde{\mathbf{Y}}_n$ is the identity matrix under the null hypothesis in (1). For this reason, it is obvious to choose $(1/p)\mathbf{I}$ as the shrinkage target in the linear shrinkage estimator (3) for the population covariance matrix related to the normalized sample $\widetilde{\mathbf{Y}}_n$. The factor $1/p$ is needed due to the assumption of the bounded trace norm imposed on the shrinkage target. Moreover, it holds that

$$\alpha^* = \alpha^*\left(\frac{1}{p}\mathbf{I}\right) = 0,$$

under the null hypothesis in (1), which is consistently estimated by

$$\hat{\alpha}^* = 1 - \frac{\frac{p}{n}\left\|\widetilde{\mathbf{S}}_n\right\|_{tr}^2}{p\left\|\widetilde{\mathbf{S}}_n\right\|_F^2 - \left(\mathrm{tr}\left(\widetilde{\mathbf{S}}_n\right)\right)^2}. \tag{10}$$

Moreover, since $\beta^*(\boldsymbol{\Omega})$ and $\hat{\beta}^*(\boldsymbol{\Omega})$ in (6) and (8) are functions of $\alpha^*(\boldsymbol{\Omega})$ and $\hat{\alpha}^*(\boldsymbol{\Omega})$, respectively, it is sufficient to determine the large-dimensional asymptotic distribution of the latter to derive a statistical test on the structure of $\boldsymbol{\Sigma}_n$.

The eigenvalues of the sample covariance matrix are a central object in large-dimensional statistics. Let $\{\lambda_1, \cdots, \lambda_p\}$ be the set of eigenvalues of $\widetilde{\mathbf{S}}_n$. For the sake of notation we omit the subscript n in the notations of eigenvalues. It can be seen in (10) that $\hat{\alpha}^*$ is completely determined by the squared Frobenius norm $||\widetilde{\mathbf{S}}_n||_F^2$ and the trace $tr(\widetilde{\mathbf{S}}_n)$ which are functions of $\{\lambda_1, \cdots, \lambda_p\}$. Namely, it holds that

$$||\widetilde{\mathbf{S}}_n||_F^2 = \sum_{i=1}^{p} \lambda_i^2$$

and

$$\mathrm{tr}(\widetilde{\mathbf{S}}_n) = \sum_{i=1}^{p} \lambda_i.$$

The joint large-dimensional asymptotic distribution of $||\widetilde{\mathbf{S}}_n||_F^2$ and $\mathrm{tr}(\widetilde{\mathbf{S}}_n)$ is derived in Lemma 2.2 from Wang and Yao (2013). We formulate this result as Lemma 1 below.

Lemma 1 (Lemma 2.2 in Wang and Yao (2013)) *Assume that conditions **(A1)–(A2)** hold. Then, under the null hypothesis in (1), we get that*

$$\begin{pmatrix} \sum_{i=1}^{p} \lambda_i^2 - p\,(1 + c_n) \\ \sum_{i=1}^{p} \lambda_i - p \end{pmatrix} \xrightarrow{D} N\left(\boldsymbol{\mu}_1, \mathbf{V}_1\right), \tag{11}$$

where

$$\boldsymbol{\mu}_1 = \begin{pmatrix} (\kappa - 1 + \beta)c \\ 0 \end{pmatrix} \tag{12}$$

and

$$\mathbf{V}_1 = \begin{pmatrix} 2\kappa c^2 + 4(\kappa + \beta)\left(c + 2c^2 + c^3\right) & 2(\kappa + \beta)\left(c + c^2\right) \\ 2(\kappa + \beta)\left(c + c^2\right) & (\kappa + \beta)c \end{pmatrix}. \tag{13}$$

Since $\hat{\alpha}^*$ is only a function of $||\widetilde{\mathbf{S}}_n||_F^2$ and $\mathrm{tr}(\widetilde{\mathbf{S}}_n)$ whose joint large-dimensional asymptotic distribution is given in Lemma 1, then the large-dimensional asymptotic distribution of $\hat{\alpha}^*$ can be deduced by applying the delta method, which is Theorem 1 from Doob (1935). For the completeness of the presentation we present this theorem as Lemma 2.

Lemma 2 (Theorem 1 in Doob (1935)) *Let* $\mathbf{X} = (X_1, \ldots, X_k)^\top$ *be a random vector and* $g : \mathbb{R}^k \to \mathbb{R}^d$ *be a differentiable function with derivative* $\nabla g(\mathbf{a})$ *at* $\mathbf{a} \in \mathbb{R}^k$. *If we have for some* $b > 0$ *and* $p \to \infty$

$$p^b\{\mathbf{X} - \mathbf{a}\} \xrightarrow{D} \mathbf{Y},$$

then

$$p^b\{g(\mathbf{X}) - g(\mathbf{a})\} \xrightarrow{D} [\nabla g(\mathbf{a})]^T \mathbf{Y}.$$

The delta method says essentially that if a random vector converges to a multivariate normal distribution in the limit, then differentiable functions of that random vector are also normally distributed. The application of the delta method to $\hat{\alpha}^*$ which is a differentiable function of $||\widetilde{\mathbf{S}}_n||_F^2$ and $\mathrm{tr}(\widetilde{\mathbf{S}}_n)$ leads to the following result:

Theorem 1 *Assume that conditions* ***(A1)–(A2)*** *hold. Then, under the null hypothesis in* (1) *we get*

$$T = p\hat{\alpha}^* \xrightarrow{D} N(\mu, \sigma^2), \tag{14}$$

with

$$\mu = \kappa - 1 + \beta \quad \textit{and} \quad \sigma^2 = 2\kappa. \tag{15}$$

Proof The application of Lemma 1 leads to

$$\begin{pmatrix} \|\mathbf{S}_n\|_F^2 - p(1+c) \\ \operatorname{tr}(\mathbf{S}_n) - p \end{pmatrix} = p \begin{pmatrix} \frac{1}{p}\sum_{i=1}^{p}\lambda_i^2 - (1+c) \\ \frac{1}{p}\sum_{i=1}^{p}\lambda_i - 1 \end{pmatrix} \xrightarrow{D} \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix},$$

with $\mathbf{u} \sim N(\boldsymbol{\mu}_1, \mathbf{V}_1)$ where $\boldsymbol{\mu}_1$ and $\mathbf{V}_1$ are defined in (12) and (13), respectively.

Define $b = 1$, $\mathbf{a} = [1 + c\ 1]^T$, and

$$g(\mathbf{t}) = g(t_1, t_2) = 1 - \frac{ct_2^2}{t_1 - t_2^2}.$$

Then,

$$g\left(\frac{1}{p}\|\mathbf{S}_n\|_F^2, \left(\frac{1}{p}\operatorname{tr}(\mathbf{S}_n)\right)^2\right) = 1 - \frac{c \cdot \left(\frac{1}{p}\operatorname{tr}(\mathbf{S}_n)\right)^2}{\frac{1}{p}\|\mathbf{S}_n\|_F^2 - \left(\frac{1}{p}\operatorname{tr}(\mathbf{S}_n)\right)^2}$$

$$= 1 - \frac{c \cdot \operatorname{tr}(\mathbf{S}_n)^2}{p\|\mathbf{S}_n\|_F^2 - (\operatorname{tr}(\mathbf{S}_n))^2} = \hat{\alpha}^*$$

and

$$g(\mathbf{a}) = g(1+c, 1) = 1 - \frac{c}{1+c-1} = 0.$$

Moreover, we get

$$\frac{\partial g}{\partial t_1} = \frac{\partial}{\partial t_1}\left\{1 - \frac{ct_2^2}{t_1 - t_2^2}\right\} = \frac{ct_2^2}{\left(t_1 - t_2^2\right)^2},$$

$$\frac{\partial g}{\partial t_2} = \frac{\partial}{\partial t_2}\left\{1 - \frac{ct_2^2}{t_1 - t_2^2}\right\} = -\frac{2ct_1t_2}{\left(t_1 - t_2^2\right)^2},$$

and, hence,

$$\nabla g(\mathbf{a}) = \begin{pmatrix} \frac{c(1^2)}{(1+c-(1^2))^2} \\ -\frac{2c(1+c)\cdot 1}{(1+c-(1^2))^2} \end{pmatrix} = \begin{pmatrix} \frac{1}{c} \\ -\frac{2(1+c)}{c} \end{pmatrix}.$$

The delta method of Lemma 2 now implies that

$$p\hat{\alpha}^* \xrightarrow{D} \left[\frac{1}{c}, -\frac{2(1+c)}{c}\right]\mathbf{u} = \frac{1}{c}u_1 - \frac{2(1+c)}{c}u_2,$$

which is a normal distribution with mean

$$\begin{aligned}\mu &= \mathbb{E}\left(\frac{1}{c}u_1 - \frac{2(1+c)}{c}u_2\right)\\ &= \frac{1}{c}(\kappa - 1 + \beta)c - \frac{2(1+c)}{c}\cdot 0 = \kappa - 1 + \beta\end{aligned}$$

and variance

$$\begin{aligned}\sigma^2 &= \mathrm{Var}\left(\frac{1}{c}u_1 - \frac{2(1+c)}{c}u_2\right)\\ &= \frac{1}{c^2}\mathrm{Var}\,(u_1) + \left(-\frac{2(1+c)}{c}\right)^2 \mathrm{Var}\,(u_2) - 2\cdot\frac{1}{c}\frac{2(1+c)}{c}\mathrm{Cov}\,(u_1, u_2)\\ &= \frac{1}{c^2}\left(4(\kappa+\beta)\left(c^3 + 2c^2 + c\right) + 2\kappa c^2\right) + \left(\frac{2(1+c)}{c}\right)^2(\kappa+\beta)(c)\\ &\quad - 2\frac{1}{c}\frac{2(1+c)}{c}2(\kappa+\beta)\left(c^2 + c\right) = 2\kappa.\end{aligned}$$

This concludes the proof. □

The result of Theorem 1 is also present in Versteegh (2020) and Nilsson (2021), while we prove it in another way. It also should be noted that the limiting distribution in (14) is the same as the one obtained for the statistic of the corrected John test derived by Wang and Yao (2013).

To visualize the results presented in Theorem 1, the histograms of the centralized random variable $W = (T - \mu)/\sigma$ are depicted in Fig. 1 for $p = 128$ and $n = 256$ together with the standard normal distribution, the large-dimensional asymptotic distribution of $W = (T - \mu)/\sigma$. The left-hand-side plot corresponds

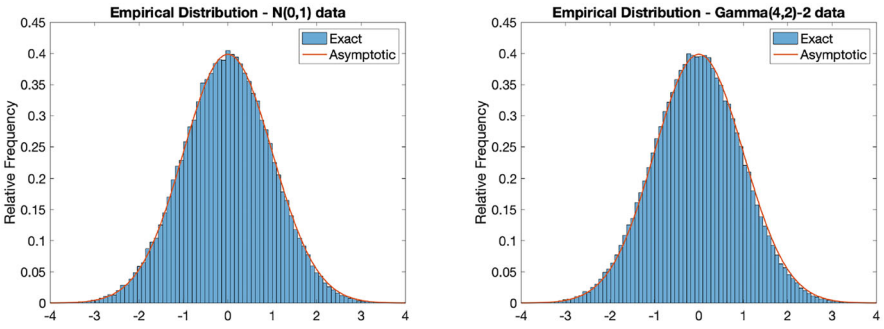

Fig. 1 Histograms together with the large-dimensional asymptotic density of the centralized random variable $W = (T - \mu)/\sigma$ calculated for random samples drawn from N(0,1) (left) and $Gamma(4, 2) - 2$ (right) distributions for $p = 128$ and $n = 256$

to the case when samples from the standard normal distribution are drawn, while the $Gamma(4, 2) - 2$ distribution is used in the right-hand-side plot. For the second figure the $Gamma(4, 2) - 2$ distribution is chosen because this gives $\beta = 3/2$ instead of $\beta = 0$ (for the standard normal distribution), although it has still zero mean and unit variance. The figures are created by using 10^5 independent replications. We observe in both plots of Fig. 1 that both histograms are very well approximated by the density function of the standard normal distribution. Therefore, we can conclude that the large-dimensional asymptotic distribution derived in Theorem 1 provides a good approximation already for moderate sample sizes. More importantly, W is a ready-to-use test statistic in the large-dimensional case that is based on the linear shrinkage estimator.

3 Simulation Study

In this section the linear shrinkage (LS) test will be compared with some well-known tests in large-dimensional statistics, such as the corrected likelihood ratio test (CLRT) and the corrected John test (CJ) both derived by Wang and Yao (2013).

Before we start off with the simulation study, it should be noted that the CLRT and CJ tests are one-tailed tests and that the LS test is a two-tailed one. It is known that one-tailed tests can be transformed to two-tailed tests without any changes. This will also be done in this simulation study to compare all the tests equally. Furthermore, it should be noted that the CJ and the LS test statistics have the same limiting distribution. Moreover, one should bear in mind that the limiting distribution of the CLRT test statistic depends on c. In particular, it depends on the $\log(1 - c)$, so it is expected that this test will break down when c increases to 1 and will not work when $c > 1$.

The three tests are compared with each other in terms of the empirical size and the empirical power. The calculation of the empirical size and empirical power is performed similarly. Recall that the size of a test is equal to the probability of rejecting the null hypothesis when it is true, while the power of a test is equal to the probability of rejecting the null hypothesis when the alternative is true. In the simulation study both probabilities are approximated by their empirical counterparts, which are computed in the following way:

(i) Draw a sample from the data-generating model.
(ii) Calculate the sample covariance matrix from the generated data.
(iii) Compute the value of the test statistic.
(iv) Using the decision rule of the test and the computed value of the test statistic, make a decision about the rejection of the null hypothesis.
(v) Repeat steps (i)–(iv) many times and compute the relative frequency of rejecting the null hypothesis.

If the data-generating model corresponds to the null hypothesis, the algorithm will produce the empirical size of the test; otherwise we get the empirical power.

If the number of repetitions is relatively large, following the law of large numbers, the empirical size and the empirical power will provide a good approximation of the true size and the true power.

3.1 Empirical Size Comparison

In this section we compare the three tests in terms of their size properties. The empirical sizes of the CLRT, CJ, and LS tests performed at 5% significance level are computed by using 10000 independent repetitions. In each repetitions the samples are drawn from the standard normal distribution and the $Gamma(4, 2) - 2$ distribution as described at the end of Sect. 2. The results of the simulation study are depicted in Table 1.

It can be seen in Table 1 that all the empirical sizes computed for samples generated from the standard normal distribution are close to the desired rejection level $\alpha = 0.05$. This conclusion holds for almost all values of p and n considered in the simulation study. As such, since all empirical sizes are close to the rejection level α, it does not really matter which combination of (p, n) to take in the empirical power comparison. This is because the three tests possess similar size properties and a fair comparison can be made in terms of power.

Unfortunately, this is not the case for the empirical sizes computed for the samples generated from the $Gamma(4, 2) - 2$ distribution. It can be seen in Table 1 that the empirical sizes of the LS test behaves quite well. However, for the CJ test this only holds for higher combinations of (p, n). It looks like that the empirical

Table 1 Empirical sizes of the CLRT, CJ, and LS tests at 5% significance level based on 10000 independent repetitions. The samples are drawn from N(0,1) (left) and $Gamma(4, 2) - 2$ (right) distributions for several values of p and n

(p, n)	CLRT	CJ	LS	(p, n)	CLRT	CJ	LS
(8, 128)	0.0565	0.0581	0.0661	(8,128)	0.2518	0.1178	0.0808
(16,128)	0.0539	0.0552	0.0479	(16,128)	0.2619	0.0911	0.0513
(32,128)	0.0518	0.0525	0.0432	(32,128)	0.2588	0.0750	0.0468
(64,128)	0.0536	0.0538	0.0479	(64,128)	0.2197	0.0645	0.0460
(96,128)	0.0547	0.0540	0.0484	(96,128)	0.1643	0.0537	0.0423
(112,128)	0.0538	0.0553	0.0516	(112,128)	0.1329	0.0601	0.0514
(120,128)	0.0522	0.0524	0.0485	(120,128)	0.1105	0.0598	0.0515
(16,256)	0.0544	0.0531	0.0473	(16,256)	0.2777	0.0861	0.0531
(32,256)	0.0519	0.0502	0.0433	(32,256)	0.2849	0.0723	0.0471
(64,256)	0.0499	0.0499	0.0437	(64,256)	0.2654	0.0625	0.0467
(128,256)	0.0516	0.0541	0.0504	(128,256)	0.2252	0.0591	0.0513
(192,256)	0.0542	0.0503	0.0488	(192,256)	0.1695	0.0572	0.0513
(224,256)	0.0505	0.0512	0.0495	(224,256)	0.1384	0.0554	0.0510
(240,256)	0.0517	0.0513	0.0480	(240,256)	0.1164	0.0547	0.0490

sizes of the CJ test approaches α from above. This means that when (p, n) is low, the empirical distribution has heavier tails than it should be. On the other side, if both p and n increase, then the corresponding large-dimensional asymptotic distribution is becoming a better approximation. Thus, the CJ test relies more on the limiting aspect in this case. The empirical sizes for the CLRT test are behaving quite pure for every combination of (p, n). They are approximately five times larger than the desired significance level $\alpha = 0.05$. As p and n increase, the performance is becoming better although still we have all empirical sizes being larger than 0.1. From this observation it can be concluded that when the data are drawn from the $Gamma(4, 2) - 2$ distribution, then the large-dimensional asymptotic distribution of the CLRT test does not provide a good approximation. Therefore, it will be difficult to make a fair empirical power comparison when the data are taken from the $Gamma(4, 2) - 2$ distribution because not all the tests will have the same size properties. So, the empirical power comparison will only be based on the standard normal distribution.

3.2 Empirical Power Comparison

In this subsection the empirical powers for the three tests will be compared. The comparison will be performed only for the samples drawn from the standard normal distribution because the empirical sizes in the case of the $Gamma(4, 2) - 2$ distribution are not all close to the desired significance level $\alpha = 0.05$ and it will lead to unfair comparison. In this simulation study we calculate the empirical powers based on the Bernoulli experiment as described at the beginning of Sect. 3. The following three types of the covariance matrices will be considered under the alternative hypothesis:

(1) H_1: compound symmetry relation
(2) H_1: autoregressive relation
(3) H_1: heteroscedasticity relation

The dimensions used in the comparison are $(p, n) = (32, 128)$, $(p, n) = (64, 128)$, $(p, n) = (96, 128)$, and $(p, n) = (120, 128)$. This results in $c = 1/4$, $c = 1/2$, $c = 3/4$, and $c = 15/16$, respectively. All computations of the empirical powers are based on 1000 repetitions.

3.2.1 Compound Symmetry

The first alternative hypothesis that will be used to make a power comparison is a compound symmetry relation. The compound symmetry means that every variable of the underlying data has variance equal to 1 and covariance equal to $Cov(y_i, y_j) = \rho$ for every $i \neq j$. This means that for $\rho \neq 0$ the underlying variables of the data are correlated and thus dependent. The compound symmetry alternative can be

represented as a linear combination of the identity matrix and a matrix of all ones. So, for $\rho \in (0, 1)$, the covariance matrix under the alternative hypotheses is defined as

$$\boldsymbol{\Sigma}_{n,\rho} = (1-\rho)\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & 1 \end{bmatrix} + \rho \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & & \vdots \\ \vdots & & \ddots & \\ 1 & \cdots & & 1 \end{bmatrix} = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & & \vdots \\ \vdots & & \ddots & \\ \rho & \cdots & & 1 \end{bmatrix}.$$

In the simulation ρ runs from 0 to 1. So as ρ increases, the alternative hypothesis $\boldsymbol{\Sigma}_{n,\rho}$ becomes less like the identity matrix $\mathbf{I}$ or the null hypothesis. To compare the different empirical powers for each test, a power plot is used. The power plot will be constructed as follows: the Bernoulli experiment will be executed for each ρ separately, and because each ρ gives a different alternative hypothesis, different empirical powers are obtained for each ρ. Plotting the calculated empirical power against the corresponding ρ results in the required power plot. The quicker a test reaches the power of 1, the better the test is, since the power is the probability that a false null hypothesis is correctly rejected. The results of the simulation study are presented in Fig. 2.

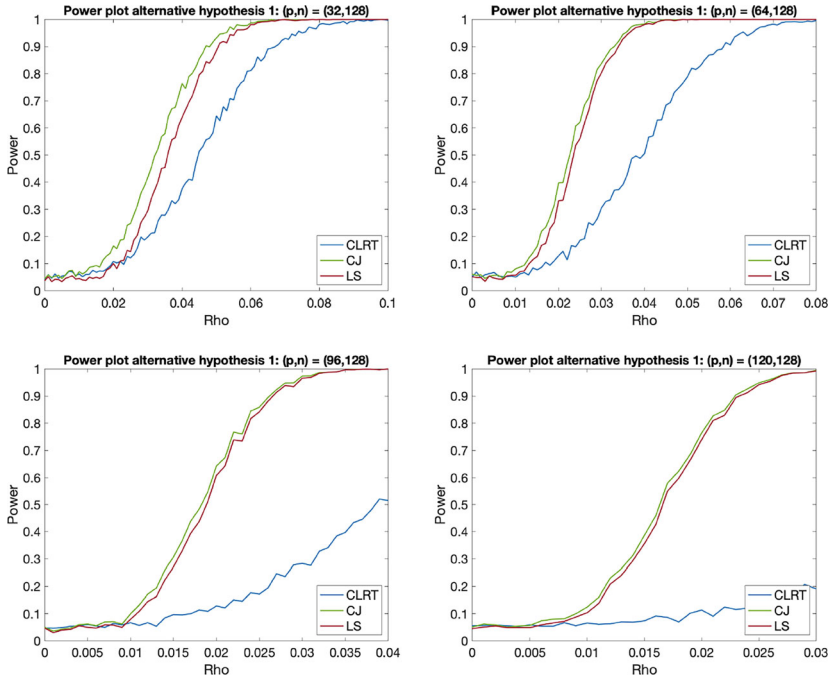

Fig. 2 Empirical powers of the CLRT, CJ, and LS tests for the testing hypotheses (1) with $\boldsymbol{\Sigma}_0 = \mathbf{I}$ and $\boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}_{n,\rho}$ where $\rho \in (0, 1)$. The computations are based on 1000 independent repetitions

In Fig. 2 it can be seen how the three tests perform in terms of the empirical power for different combinations of c. For $\rho \geq 0.08$ all tests have power close 1 and as expected the CJ test and the LS test behave nearly the same. This is due to the fact that they have the same limiting distribution. Most noticeable in Fig. 2 is that the CLRT test breaks down when p is getting closer to n. This is again what is expected since the limiting distribution depends on $\log(1-c)$. Overall, the CJ and the LS tests perform best and they are the first to reach a power of 1 for every combination of (p, n).

3.2.2 Autoregressive Relation

The second alternative hypothesis is the autoregressive relation. The autoregressive relation is based on an autoregressive model, which is a popular type of univariate time series. The autoregressive model specifies that the output variable depends linearly on its own previous values and on a stochastic error term. Under the autoregressive relation, the covariance matrix under the alternative hypothesis is specified by

$$\boldsymbol{\Sigma}_{n,\delta} = \begin{bmatrix} 1 & \delta & \delta^2 & \dots & \delta^{p-1} \\ \delta & 1 & \delta & \dots & \delta^{p-2} \\ \delta^2 & \delta & \ddots & & \vdots \\ \vdots & & & \ddots & \delta \\ \delta^{p-1} & \delta^{p-2} & \dots & \delta & 1 \end{bmatrix}$$

for $\delta \in (-1, 1)$. The simulation study is organized in the same way as for the previous alternative hypothesis. As δ goes away from 0 in both directions, this could be seen as moving away from the null hypothesis $H_0 : \boldsymbol{\Sigma}_n = \mathbf{I}$ because the alternative hypothesis matrix becomes less like the population covariance matrix. Then for every δ the Bernoulli experiment is carried out and the empirical power is computed. The results of the simulation study are presented in Fig. 3.

It can be seen in Fig. 3 that for $p = 32$ and $p = 64$ the CJ, LS, and CLRT tests perform quite the same. Still the CJ test performs best but the other two are not far behind. Then, when p gets larger, the CJ and LS tests are still outperforming the CLRT test, whose power becomes worse as p increases. This is in line with the observations from the previous simulation for the compound symmetry alternative.

3.2.3 Heteroscedasticity Relation

The third alternative hypothesis corresponds to the case when a fixed ratio r of the variables has a variance equal to $1 + \gamma$, while the variance of the rest variables is one. In econometrics such a relation is called heteroscedasticity. For any $r \in (0, 1)$

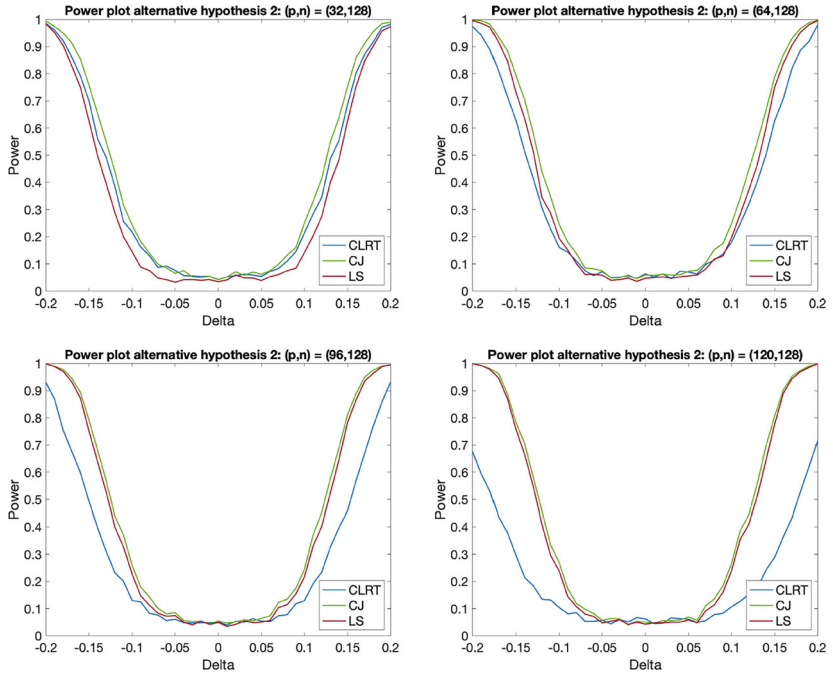

Fig. 3 Empirical powers of the CLRT, CJ, and LS tests for the testing hypotheses (1) with $\boldsymbol{\Sigma}_0 = \mathbf{I}$ and $\boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}_{n,\delta}$ where $\delta \in (-1, 1)$. The computations are based on 1000 independent repetitions

and $\gamma > -1$, we define

$$\boldsymbol{\Sigma}_{n,r,\gamma} = \begin{bmatrix} 1 & 0 & & \cdots & 0 \\ 0 & \ddots & & & \vdots \\ & & 1+\gamma & & \\ \vdots & & & \ddots & \\ 0 & \cdots & & & 1+\gamma \end{bmatrix},$$

which presents the covariance matrix under the alternative hypothesis used in the third simulation study. If it happens that $r \cdot p$ is not a whole number, it will be rounded down. In the simulation study γ will run from -1 to 1. This can again be seen as departing from the null hypotheses $H_0 : \boldsymbol{\Sigma}_n = \mathbf{I}$ when γ goes away from 0 in both directions. For every $\gamma \in (-1, 1)$ we will compute the empirical powers which are depicted for $r = 1/2$ in Fig. 4, for $r = 1/4$ in Fig. 5, and for $r = 3/4$ in Fig. 6.

It can be seen in Fig. 4 that the CJ, LS, and CLRT tests are again quite comparable for small values of c. However, when p increases, the CLRT gets worse and worse for the same reason as in the previous simulations. Moreover, it should be noted that

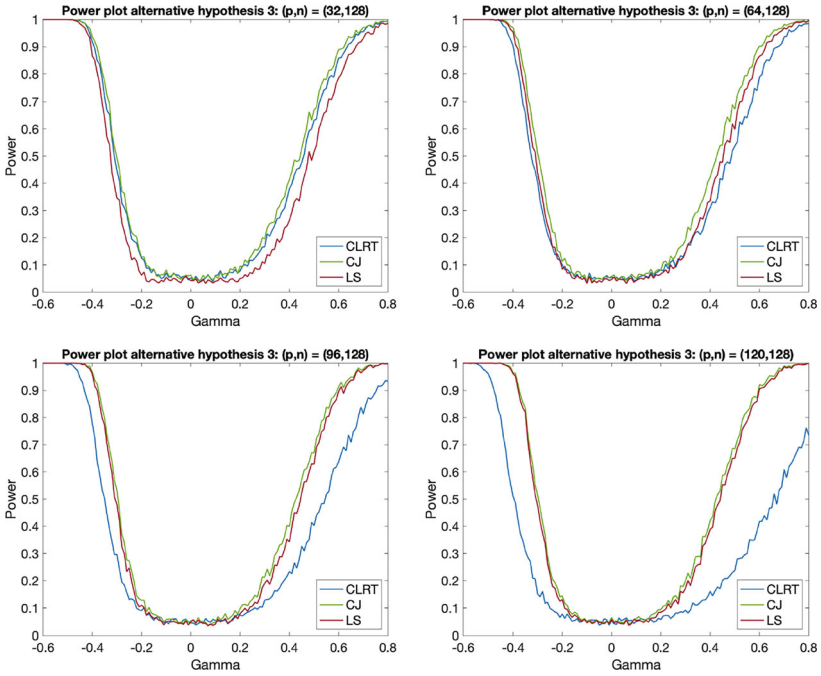

Fig. 4 Empirical powers of the CLRT, CJ, and LS tests for the testing hypotheses (1) with $\boldsymbol{\Sigma}_0 = \mathbf{I}$ and $\boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}_{n,r,\gamma}$ where $\gamma \in (-1, 1)$ and $r = 1/2$. The computations are based on 1000 independent repetitions

the empirical powers of all the tests are not symmetric around zero. The powers of the three tests increase much faster for negative values of γ than for positive values.

In Figs. 5 and 6 it can be seen that all tree tests perform better when r decreases, especially when γ is positive. This behavior can be explained by the fact that the null hypothesis, which is actually tested, is whether the population covariance matrix is equal to a multiple of the identity matrix. This means that for $r = 1/2$ the alternative hypothesis is furthest away from the null hypothesis, and we observe the highest powers in this case. Therefore, it can be concluded that the CJ, LS, and CLRT tests are invariant under multiples of the identity matrix what is expected from the expressions of their test statistics.

4 Summary

In many statistical applications one would like to perform hypothesis tests on the structure of large-dimensional covariance matrices. In particular, sphericity testing and testing for certain dependence structure of the covariance matrix are important

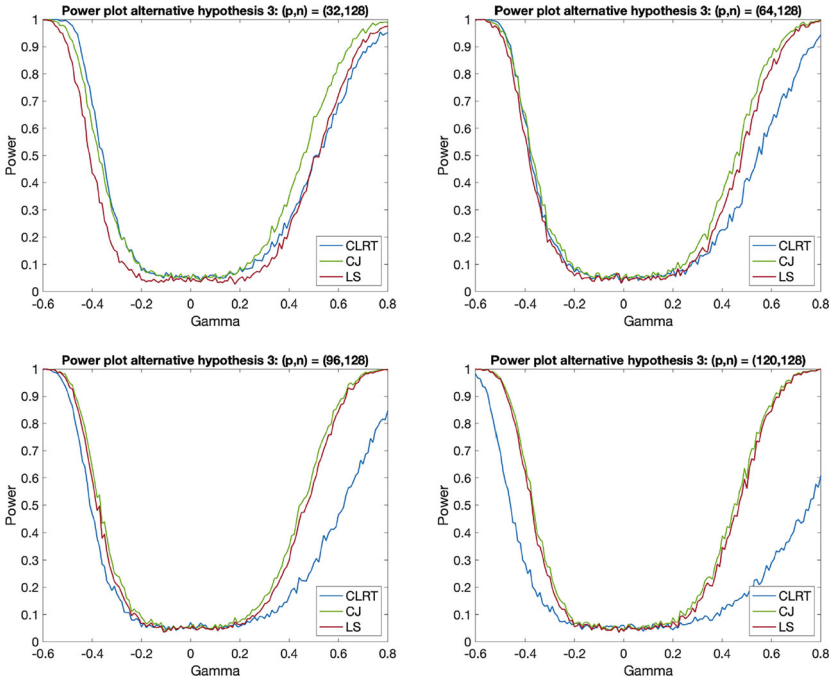

Fig. 5 Empirical powers of the CLRT, CJ, and LS tests for the testing hypotheses (1) with $\boldsymbol{\Sigma}_0 = \mathbf{I}$ and $\boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}_{n,r,\gamma}$ where $\gamma \in (-1, 1)$ and $r = 1/4$. The computations are based on 1000 independent repetitions

problems in economics and finance. Therefore, this chapter presents a new approach to construct such a hypothesis test in the large-dimensional framework.

The new test statistic is based on the linear shrinkage estimator and on the shrinkage intensities used in its construction. In the derivation of the large-dimensional limiting distribution of the test statistics, the asymptotic properties of linear spectral statistics are used. Even though linear shrinkage test statistic is different from the corrected John test statistic, they still have the same limiting distribution. The construction provides some new inferential procedures for large-dimensional data analysis using a connection between estimators and test statistics.

The theoretical results are illustrated by means of a simulation study. The proposed new test is compared with the corrected John test and the corrected likelihood ratio test. It was found that the linear shrinkage test behaves nearly the same as the corrected John test in terms of the empirical power. The differences in the powers of these two tests are explained by the fact that the linear shrinkage test depends more on the limiting aspect than the corrected John test does. Moreover, both the linear shrinkage test and the corrected John test outperform the corrected likelihood ratio test.

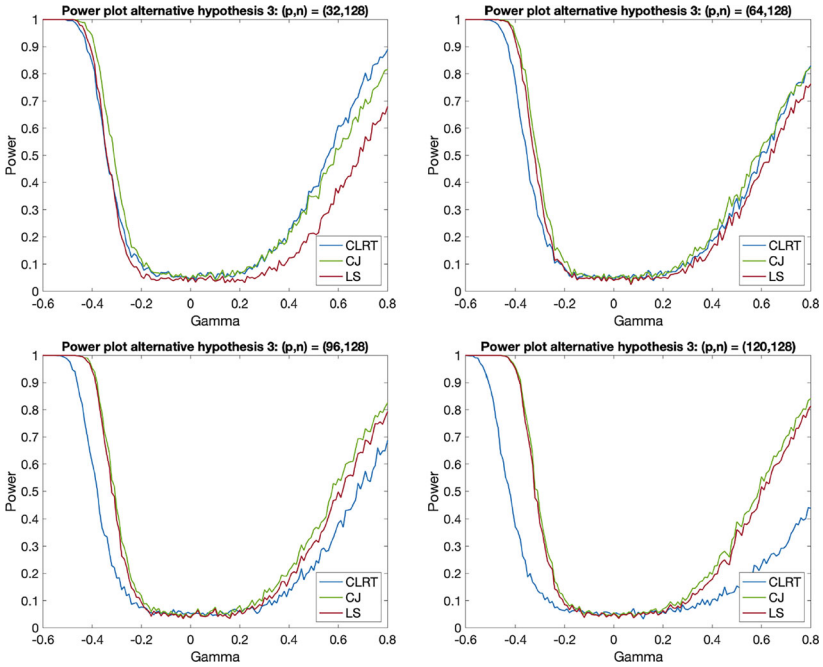

Fig. 6 Empirical powers of the CLRT, CJ, and LS tests for the testing hypotheses (1) with $\boldsymbol{\Sigma}_0 = \mathbf{I}$ and $\boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}_{n,r,\gamma}$ where $\gamma \in (-1, 1)$ and $r = 3/4$. The computations are based on 1000 independent repetitions

Acknowledgments The authors would like to thank Professor Yarema Okhrin and an anonymous reviewer for their constructive comments that improved the quality of this chapter.

References

Anderson, T. W. (1984). *An introduction to multivariate statistical analysis*. Wiley.

Bai, Z., & Silverstein, J. W. (2010). *Spectral analysis of large dimensional random matrices* (Vol. 20). Springer.

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

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

Doob, J. L. (1935). The limiting distributions of certain statistics. *The Annals of Mathematical Statistics, 6*(3), 160–169.

John, S. (1971). Some optimal multivariate tests. *Biometrika, 58*(1), 123–127.

Ledoit, O., & Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *The Annals of Statistics, 30*(4), 1081–1102.

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis, 88*(2), 365–411.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance, 7*(1), 77–91.
Nilsson, M. (2021). *A shrinkage test for large-dimensional covariance matrix*. Master thesis, Stockholm University.
Paul, D., & Aue, A. (2014). Random matrix theory in statistics: A review. *Journal of Statistical Planning and Inference, 150*, 1–29.
Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the third Berkeley symposium on mathematical statistics and probability: Contributions to the theory of statistics* (Vol. 1, pp. 197–206). University of California Press.
Versteegh, W. (2020). *Central limit theorems for linear spectral statistics of large regularized covariance matrices*. Bachelor thesis, Delft University of Technology.
Wang, Q., & Yao, J. (2013). On the sphericity test with large-dimensional observations. *Electronic Journal of Statistics, 7*, 2164–2192.
Yao, J., Zheng, S., & Bai, Z. (2015). *Sample covariance matrices and high-dimensional data analysis*. Cambridge: Cambridge University Press.