Nonlinear shrinkage test on a large-dimensional covariance matrix

Bodnar, Taras; Parolya, Nestor; Veldman, Frederik

DOI
10.1111/stan.12348

Publication date
2025

Document Version
Final published version

Published in
Statistica Neerlandica

Citation (APA)
Bodnar, T., Parolya, N., & Veldman, F. (2025). Nonlinear shrinkage test on a large-dimensional covariance matrix. *Statistica Neerlandica*, *79*(1), Article e12348. https://doi.org/10.1111/stan.12348

Received: 4 June 2024 | Accepted: 14 June 2024

DOI: 10.1111/stan.12348

ORIGINAL ARTICLE

WILEY

Nonlinear shrinkage test on a large-dimensional covariance matrix

Taras Bodnar[1] | **Nestor Parolya**[2] | **Frederik Veldman**[2]

[1]Department of Management and Engineering, Linköping University, Linköping, Sweden

[2]Department of Applied Mathematics, Delft University of Technology, Delft, The Netherlands

Correspondence
Taras Bodnar, Department of Management and Engineering, Linköping University, 581 83 Linköping, Sweden.
Email: taras.bodnar@liu.se

This paper is concerned with deriving a new test on a covariance matrix which is based on its nonlinear shrinkage estimator. The distribution of the test statistic is deduced under the null hypothesis in the large-dimensional setting, that is, when $p/n \to c \in (0, +\infty)$ with p variables and n samples both tending to infinity. The theoretical results are illustrated by means of an extensive simulation study where the new nonlinear shrinkage-based test is compared with existing approaches, in particular with the commonly used corrected likelihood ratio test, the corrected John test, and the test based on the linear shrinkage approach. It is demonstrated that the new nonlinear shrinkage test possesses better power properties under heteroscedastic alternative.

KEYWORDS
large-dimensional asymptotics, large-dimensional covariance matrix, linear spectral statistics, nonlinear shrinkage, random matrix theory, shrinkage test

1 | INTRODUCTION

Testing the structure of a covariance matrix is an important problem in multivariate statistics with many applications across different fields of science such as finance, environmetrics, signal processing, and wireless communications, among others (see, e.g., Anderson, 1984; Cai & Jiang, 2011;

Taras Bodnar, Nestor Parolya, and Frederik Veldman contributed equally to this study.

Gupta & Bodnar, 2014; John, 1971; Gupta & Xu, 2006; Nagao, 1973). The topic becomes very challenging under the large-dimensional setting when the dimension of the data-generating model is comparable to the sample size or even it is larger than the sample size. This is because fixed-dimension assumptions do not yield proper approximations of asymptotic distributions, which are better deduced when considering dimensions that grow with the sample size. This led to a new area in asymptotic statistics, the so-called large-dimensional asymptotic regime (see, Bai & Silverstein, 2010). Although several approaches have been suggested, the derivation of tests on a large-dimensional covariance matrix is still a hot topic in statistical literature with plenty of posvsible applications (cf. Bodnar, Dette, & Parolya, 2019; Chen, Zhang, & Zhong, 2010; Fisher, Sun, & Gallagher, 2010; Ledoit & Wolf, 2002; Srivastava, 2005; Wang & Yao, 2013). Recently, several large-dimensional asymptotic tests on functions of the elements of a large-dimensional covariance matrix were developed which have a direct application in portfolio theory (see, e.g., Bodnar, Dmytriv, Parolya, & Schmid, 2019; Bodnar, Dmytriv, Okhrin, Parolya, & Schmid, 2021).

Shrinkage estimation of the mean vector and covariance matrix presents another rapidly growing line of research in statistics with many applications. While Stein (1956), James and Stein (1961), Berger, Bock, Brown, Casella, and Gleser (1977), Gleser (1986), Chételat and Wells (2012), Wang, Tong, Cao, and Miao (2014), Bodnar, Okhrin, and Parolya (2019) proposed shrinkage-based estimators for the mean vector, Dey and Srinivasan (1985), Kubokawa and Srivastava (2008), Ledoit and Wolf (2012), Bodnar, Gupta, and Parolya (2014, 2016) developed several shrinkage estimators for the covariance matrix. The shrinkage estimators and their asymptotic properties under the large-dimensional asymptotic regime for the functions involving both the mean vector and covariance matrix were derived in Bodnar, Parolya, and Schmid (2018), Bodnar, Okhrin, and Parolya (2023) and implemented to the optimal portfolio choice problems. The recent developments in the field are reviewed in Bodnar, Bodnar, and Parolya (2022) among others.

While Bodnar, Dmytriv, et al. (2019), Bodnar et al. (2021) constructed a test for the weights of large-dimensional optimal portfolio by using the shrinkage approaches, Versteegh (2020), Nilsson (2021), Bodnar, Parolya, and Veldman (2024) derived tests on the large-dimensional covariance matrix using its linear shrinkage estimator. In particular, it was shown that the test statistic of the linear shrinkage test on the large-dimensional covariance matrix is a function of the large-dimensional John test (see, e.g., Bodnar et al., 2024) and, consequently, it possesses the optimality properties of the latter test.

Ledoit and Wolf (2012) proposed the application of the nonlinear shrinkage estimator of the covariance matrix as a generalization of its linear shrinkage estimator. Moreover, the properties of the nonlinear estimators of the covariance matrix were established in Dey and Srinivasan (1985), Ledoit and Wolf (2012), among others and it was argued that it possesses better asymptotic distributional properties in comparison to the linear shrinkage estimator, especially due to its generality.

In this paper, we contribute to the existent literature by deriving a new test on the large-dimensional covariance matrix which is based on the nonlinear shrinkage approach. The new test, based on the generality of the nonlinear shrinkage estimator of the covariance matrix, will feature a closed-form expression for the test statistic, with its large-dimensional asymptotic distribution deduced under the null hypothesis. Within an extensive simulation study, we will compare the new approach with the existing tests, namely with the test based on the linear shrinkage method, the corrected John (CJ) test, and the likelihood ratio test.

The rest of the paper is organized in the following way. In Section 2, the set of assumptions is discussed that will be maintained throughout this paper as well as some of the relevant results from random matrix theory in the large-dimensional asymptotic framework are reviewed. In

Section 3, a new test on a large-dimensional covariance matrix is introduced, which is based on the nonlinear shrinkage estimator, and its limiting distribution is derived, which is the main theoretical contribution of the paper. Final remarks are provided in Section 4, while the proofs of theoretical results are postponed to the Appendix. The finite-sample performance of the derived theoretical results is investigated through an extensive simulation study in Appendix A, where the new test is compared with the test based on the linear shrinkage approach as well as with the corrected likelihood ratio (CLRT) test and the CJ test, both derived by Wang and Yao (2013). The comparison of the tests is performed in terms of the empirical rejection rate and the empirical power.

2 | PRELIMINARY RESULTS

Let $\mathbf{y}_1, \dots, \mathbf{y}_n$ be a sample from a p-dimensional distribution and let $\mathbf{Y}_n = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)$ denote the observation matrix which is assumed to follow the stochastic model given by

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

where $\mathbf{\Sigma}_n$ is the $p \times p$ dimensional population covariance matrix. Throughout the paper, we work in the large-dimensional setting, that is, when $\frac{p}{n} \to c \in (0, +\infty)$ as $n \to \infty$, and the following assumptions are imposed on the data-generating model (1):

- **(A1)** The population covariance matrix $\mathbf{\Sigma}_n$ is a nonrandom positive definite matrix.
- **(A2)** The matrix $\mathbf{X}_n$ consists of independent and identically distributed (i.i.d.) random variables with mean zero, unit variance, and finite fourth moment equal to $\mathbb{E}[|x_{i,j}|^4] = \beta + 1 + \kappa < \infty$, where $\kappa = 2$ in the case of real variables and $\kappa = 1$ in the case of complex variables. Thus, β plays the role of the excess kurtosis, for example, the fourth moment minus three in the case of real variables. Also, it holds that $\mathbb{E}[x_{i,j}^2] = 0$ in the case of complex variables.

We note that $\mathbf{Y}_n$ is only observable, while the aim is to derive a large-dimensional test on the structure of the population covariance matrix $\mathbf{\Sigma}_n$. Assumptions (A1) and (A2) are needed for derivation of the asymptotic distribution under the null hypothesis, that is, $\mathbf{\Sigma}_n = \mathbf{\Sigma}_{0,n}$. Note that they are not sufficient for analysis under the alternative since one has to put some conditions on the spectra of $\mathbf{\Sigma}_n$. Nevertheless, they do not impose any specific distributional assumption on the data-generating model. Only the existence of the fourth moments is required. Furthermore, it is assumed that the columns of the data matrix $\mathbf{Y}_n$ are independent, while the dependence between its rows is captured by $\mathbf{\Sigma}_n$. Finally, we establish the theoretical results in the case of real variables, that is, when $\kappa = 2$. However, to keep the results as general as possible we derive them as a function of κ.

The eigenvalues of the sample covariance matrix are the central object in large-dimensional statistics. Let $\{\lambda_{n,1}, \dots, \lambda_{n,p}\}$ be the set of eigenvalues corresponding to the set of eigenvectors $\{\mathbf{u}_{n,1}, \cdots, \mathbf{u}_{n,p}\}$ of the sample covariance matrix defined by[1]

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

[1]The findings discussed in this paper remain valid even when dealing with a non-zero mean vector. The only adjustment necessary is to use the centered version of the sample covariance matrix, which is obtained by subtracting the sample mean from each observation before calculating the covariance, and replacing n by $n-1$ in the expression of asymptotic mean and variance due to the "substitution principle" presented in Zheng, Bai, and Yao (2015).

For the sake of notation simplicity, the first subscript n in the set of eigenvalues and eigenvectors will be omitted. The empirical spectral distribution (ESD) of the sample eigenvalues is defined as

$$F_n(\lambda) = \frac{1}{p}\sum_{i=1}^{p} \mathbb{1}_{[\lambda_i,+\infty)}(\lambda), \quad \lambda \in R,$$

where $\mathbb{1}_{\mathcal{A}}(.)$ is the indicator function of set $\mathcal{A}$.

Consider a sequence of empirical spectral distributions of sample covariance matrices, that is, $\{F_n(\lambda)\}_{n\geq 1}$. If this sequence converges to a measure $F(\lambda)$, then $F(\lambda)$ is called its limiting spectral distribution (LSD) which characterizes the asymptotic behavior of the eigenvalues of the sample covariance matrix. One of the main results in large-dimensional statistics states that the LSD of the sample eigenvalues converges to the standard Marchenko–Pastur distribution (Marchenko & Pastur, 1967). Marchenko and Pastur were the first to discover this property, which has later been extended in several directions such as Theorem 2.9 in Yao, Zheng, and Bai (2015). This theorem states that if $\mathbf{\Sigma}_n$ is the identity matrix and the entries $\{x_{i,j}\}$ of $\mathbf{X}_n$ are i.i.d. complex random variables with mean zero and variance one, then almost surely F_n converges to the standard Marchenko–Pastur distribution F_c (standard M-P law) as $\frac{p}{n} \to c \in (0, +\infty)$ for $n \to \infty$. The standard Marchenko-Pastur distribution F_c with index c has the density expressed as

$$p_c(x) = \begin{cases} \frac{1}{2\pi cx}\sqrt{(b-x)(x-a)} & a \leq x \leq b \\ 0 & \text{otherwise,} \end{cases}$$

with an additional point mass of value $(1 - \frac{1}{c})$ at the origin if $c > 1$, where $a = (1 - \sqrt{c})^2$ and $b = (1 + \sqrt{c})^2$.

Linear spectral statistic (LSS) is another important object in large-dimensional statistics widely used in the hypothesis test. For a specific function φ, it is given by

$$T_n = \frac{1}{p}\sum_{i=1}^{p}\varphi(\lambda_i) = \int \varphi(\lambda_i)dF_n(x) =: F_n(\varphi),$$

where the specific choice of the function φ depends on the testing problem to be studied. Note that by Theorem 2.9 in Yao et al. (2015), $F_n \to F_c$ almost surely. As such, one may expect that the random process $G_n(x) = a_n(F_n(x) - F_c(x))$ converges to some limiting process for an appropriate normalizing sequence a_n. Unfortunately, that is not generally possible, as discussed in Bai and Silverstein (2004). However, it may still be possible that for $a_n = p$ the sequence of random variables $Z_n = p\int \varphi(\lambda_i)d(F_n(x) - F_c(x))$ may converge to some limit law for a suitable regular class of functions φ. The random variables Z_n can also be written as,

$$Z_n = p\int \varphi(x)(dF_n(x) - dF_c(x)) = p\left\{\int \varphi(x)dF_n(x) - \int \varphi(x)dF_c(x)\right\}$$
$$= p\left\{\frac{1}{p}\sum_{i=1}^{p}\varphi(\lambda_i) - \int \varphi(x)dF_c(x)\right\}.$$

Thus, to find the distribution of a LSS for a particular class of functions φ, the fluctuations of the LSS around its limit under the null hypothesis need to be investigated. Fortunately, the central limit theorem (CLT) for LSSs offers a solution. A convenient expression can be found in Theorem 3.4 from Yao et al. (2015). This theorem states that for a specific class of function the fluctuations of a LSS around its limit can be described by a normal distribution. This is one of the main tools used to construct a statistical test in the large-dimensional setting.

Even though the eigenvalues of a matrix are continuous functions of its entries, when the dimension of a matrix is larger than four, these functions have no closed-form expressions anymore. To study their properties, the Stieltjes transform method can be used. For a finite measure μ, it is defined by

$$m_{\mu}(z) = \int \frac{1}{x-z}\mu(dx), \qquad \forall z \in \mathbb{C}^{+} = \{z \in \mathbb{C} : \Im(z) > 0\}.$$

One of the main properties of the Stieltjes transform is that it characterizes the vague convergence of finite measures. This is a key tool in studying empirical spectral distributions of random matrices. It is also summarized in Theorem 2.7 from Yao et al. (2015) which states that a sequence of probability measures $\{\mu_n\}$ converges vaguely to some positive measure μ if and only if the sequence of their Stieltjes transforms $\{m_{\mu_n}\}$ converges to m_{μ} on $\mathbb{C}^{+}$. Moreover, it can be shown that the Stieltjes transform of the empirical spectral distribution of the sample covariance matrix S_n is equal to

$$m_{F_n}(z) = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{\lambda_i - z} = \frac{1}{p}tr[(S_n - zI)^{-1}], \quad \forall z \in \mathbb{C}^{+}.$$

From Theorem 2.9 in Yao et al. (2015), it is known that the sequence of the ESDs $F_n(\lambda)$ of the sample covariance matrix converges almost surely to a nonrandom limit $F(\lambda)$. As such, the sequence of the Stieltjes transforms of the ESDs of the sample covariance matrix should also converge by Theorem 2.7 in Yao et al. (2015). This is one of the main results of Marchenko and Pastur (1967). The most convenient expression for this limit is found in (Silverstein & Choi, 1995), which is given by

$$m_F(z) = \int_{-\infty}^{+\infty} \frac{1}{\tau[1 - c - czm_F(z)] - z} dH(\tau), \quad z \in \mathbb{C}^{+}, \tag{2}$$

where $H(\tau)$ is the LSDs of the population eigenvalues. Moreover, Silverstein and Choi (1995) showed that for all $z \in \mathbb{C}^{+}$ the following limit exists

$$\lim_{z \to \lambda} m_F(z) = \breve{m}_F(\lambda), \quad \forall \lambda \in \mathbb{R} \setminus \{0\}. \tag{3}$$

The above two results are the main reasons, why the Stieltjes transform is very important in statistics. Equation (3) appears in the nonlinear shrinkage estimator and Equation (2) is needed to investigate if it is possible to construct a statistical test by using a nonlinear shrinkage estimator of the population covariance matrix.

3 | CLT FOR THE NONLINEAR SHRINKAGE ESTIMATOR

The main idea of introducing the nonlinear shrinkage method is to improve the linear shrinkage estimator by tackling the higher-order effects. The simplicity of the linear shrinkage approach, which only takes the first-order effect into account, may result in ignoring important information present in the higher moments. As such, a nonlinear shrinkage estimation of the covariance matrix was introduced in the statistical literature by Dey and Srinivasan (1985) and Ledoit and Wolf (2012).

Under the absence of specific information about the population covariance matrix, it is reasonable to consider those estimators, which are invariant under rotations of the observed data. In Perlman (2007), the rotation-invariant estimator $\mathbf{U}_n\mathbf{T}_n\mathbf{U}_n^T$ for $\mathbf{\Sigma}_n$ was considered where $\mathbf{T}_n = \text{diag}(\tau_1, \dots, \tau_p)$ is a diagonal matrix with true population eigenvalues and $\mathbf{U}_n$ is the matrix whose ith column is the sample eigenvector $\mathbf{u}_i$. This class of estimators is employed in constructing a nonlinear shrinkage estimator.

The objective is to find the estimator of $\mathbf{\Sigma}_n$ closest to the population covariance matrix. To quantify the word "closest," the quadratic Frobenius norm is used defined by $||\mathbf{A}||_F^2 = tr(\mathbf{A}\mathbf{A}^T)$ for a matrix $\mathbf{A}$. To find the estimator closest to the population covariance matrix $\mathbf{\Sigma}_n$, the following minimization problem needs to be solved

$$\min_{\mathbf{D}_n} ||\mathbf{U}_n\mathbf{D}_n\mathbf{U}_n^T - \mathbf{\Sigma}_n||_F^2, \tag{4}$$

where the minimum in (4) is taken over all diagonal matrices $\mathbf{D}_n = \text{diag}(d_1, \dots, d_p)$. Elementary matrix algebra shows that the optimal solution of (4) is given by

$$\widetilde{\mathbf{D}}_n = \text{diag}(\tilde{d}_1, \dots, \tilde{d}_p) \text{ with } \tilde{d}_i = \mathbf{u}_i^T\mathbf{\Sigma}_n\mathbf{u}_i \text{ for all } i \in \{1, \dots p\}. \tag{5}$$

The interpretation of $\tilde{d}_i$ is that it catches how the ith sample eigenvector $\mathbf{u}_i$ relates to the population covariance matrix $\mathbf{\Sigma}_n$. As a result, the finite-sample optimal estimator is given by $\mathbf{\Sigma}_n^* = \mathbf{U}_n\widetilde{\mathbf{D}}_n\mathbf{U}_n^T$ where $\widetilde{\mathbf{D}}_n$ is given in (5). However, it is not possible to calculate $\tilde{d}_i$, $i \in \{1, \dots, p\}$ explicitly because it depends on the nonobservable population covariance matrix $\mathbf{\Sigma}_n$. Therefore, it is important to get as close to $\mathbf{\Sigma}^*$ as possible by characterizing the asymptotic behavior of $\tilde{d}_i$, $i \in \{1, \dots, p\}$. To do this Ledoit and Péché (2011) introduced a nondecreasing function defined by

$$\Delta_p(\lambda) = \frac{1}{p}\sum_{i=1}^{p} \tilde{d}_i \mathbb{1}_{[\lambda_i, +\infty)}(\lambda) \quad \text{for} \quad \lambda \in \mathbb{R}. \tag{6}$$

Moreover, it was shown that $\Delta_p(\lambda)$ converges almost surely to a nonrandom quantity expressed as

$$\delta(\lambda) = \frac{\lambda}{|1 - c - c\lambda\breve{m}_F(\lambda)|^2}, \tag{7}$$

where $\breve{m}_F(\lambda)$ is given in (3).

Using this result, Ledoit and Wolf (2012) showed that the asymptotic quantity corresponding to $\tilde{d}_i$ is $\delta(\lambda_i)$. This result leads to the introduction of a nonlinear shrinkage estimator of $\mathbf{\Sigma}_n$ given by

$$\widehat{\mathbf{\Sigma}}_n = \mathbf{U}_n\widehat{\mathbf{D}}_n\mathbf{U}_n^T \quad \text{with} \quad \widehat{\mathbf{D}}_n = \text{Diag}(\delta(\lambda_1), \dots, \delta(\lambda_p)), \tag{8}$$

where λ_i, $i \in \{1, \dots, p\}$, are the eigenvalues of the sample covariance matrix $\mathbf{S}_n$ and $\mathbf{U}_n$ is the matrix whose ith column is the sample eigenvector $\mathbf{u}_i$ corresponding to the eigenvalue λ_i. Note that $\widehat{\boldsymbol{\Sigma}}_n$ is a nonlinear shrinkage estimator because the eigenvalues of $\widehat{\mathbf{D}}_n$ are obtained by applying the nonlinear shrinking function $\delta(\lambda)$ form (7) to every sample eigenvalue λ_i. The obtained nonlinear shrinkage estimator is a so-called oracle estimator since it depends on the limiting distribution of the sample eigenvalues and not on the observed one.

One could wonder why a test on the structure of a covariance matrix should be based on a (non)linear shrinkage estimator of the covariance matrix rather than on the sample covariance matrix, which is a problem that has extensively been studied before. The motivation behind the application of a shrinkage approach is to construct a statistical test with better finite-sample performance in terms of (i) controlling the size of the test (or the null-rejection probability) and (ii) achieving higher power.

Next, we investigate which function φ should be used in the definition of LSS, when a test on a covariance matrix is derived by using a nonlinear shrinkage estimator. The testing hypotheses are given by

$$H_0 : \boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}_{0,n} \quad \text{against} \quad H_1 : \boldsymbol{\Sigma}_n \neq \boldsymbol{\Sigma}_{0,n}, \tag{9}$$

for a positive definite matrix $\boldsymbol{\Sigma}_{0,n}$. For the derivation of φ, the expression of $\delta(\lambda)$ is utilized as given in (7).

Since the target covariance matrix $\boldsymbol{\Sigma}_{0,n}$ is the true covariance matrix of $\mathbf{y}_i$ under the null hypothesis in (9), we normalize the original data matrix and define $\widetilde{\mathbf{Y}}_n = \boldsymbol{\Sigma}_{0,n}^{-1/2}\mathbf{Y}_n$, using which the sample estimator for the covariance matrix is constructed. Furthermore, testing the null hypothesis in (9) is equivalent to verify whether the population covariance matrix of the normalized sample is the identity matrix under the null hypothesis. This allows us, without loss of generality, to apply the presented previously results with $\boldsymbol{\Sigma}_n$ replaced by the identity matrix and note that the original data matrix is normalized as in the definition of $\widetilde{\mathbf{Y}}_n$. The testing problem (9) in the case of $\widetilde{\mathbf{Y}}_n$ is given by

$$H_0 : \boldsymbol{\Sigma}_{n;\widetilde{\mathbf{Y}}_n} = \mathbf{I} \quad \text{against} \quad H_1 : \boldsymbol{\Sigma}_{n;\widetilde{\mathbf{Y}}_n} \neq \mathbf{I}, \tag{10}$$

where $\mathbf{I}$ is the identity matrix and $\boldsymbol{\Sigma}_{n;\widetilde{\mathbf{Y}}_n}$ denotes the population covariance matrix in the case of $\widetilde{\mathbf{Y}}_n$.

Therefore, for the normalized sample, the underlying distribution of the population eigenvalues $H(\tau)$ jumps to 1 at $\tau = 1$. Hence, under the null hypothesis $m_F(z)$ satisfies the following equation

$$m_F(z) = \int_{-\infty}^{+\infty} \frac{1}{\tau[1-c-czm_F(z)]-z} dH(\tau) = \frac{1}{1-c-czm_F(z)-z}, \quad z \in \mathbb{C}^+,$$

or, equivalently,

$$m_F(z)^2 cz + m_F(z)(c+z-1) + 1 = 0.$$

Applying the quadratic formula gives the following solutions for $m_F(z)$ for all $z \in \mathbb{C}^+$

$$m_F(z) = \frac{1-c-z \pm \sqrt{(b-z)(a-z)}}{2cz}, \tag{11}$$

where $a = (1 - \sqrt{c})^2$ and $b = (1 + \sqrt{c})^2$ are the boundaries of the support of the sample eigenvalues.

Since $z \in \mathbb{C}^+$ is a complex number, a branch cut for the square root has to be chosen. Without loss of generality, it is possible to take the principal branch. Moreover, since the Stieltjes transform is defined on the upper half complex plane its imaginary part should stay positive as well, which is only possible if one takes a "+" in formula (11), see Lemma 3.3.1 in Bai and Silverstein (2004) for details. Hence, the square root is continuous and it is possible to take the limit $z \to \lambda$:

$$\breve{m}_F(\lambda) = \lim_{z\to\lambda} m_F(z) = \lim_{z\to\lambda} \frac{1 - c - z + \sqrt{(b - z)(a - z)}}{2cz}$$
$$= \frac{1 - c - \lambda + \sqrt{(b - \lambda)(a - \lambda)}}{2c\lambda} = \frac{1 - c - \lambda + \sqrt{(\lambda - 1 - c)^2 - 4c}}{2c\lambda}.$$

Substituting the above expression for $\breve{m}_F(\lambda)$ into (7), the following function is obtained

$$\varphi(\lambda) = \frac{4\lambda}{|c - 1 - \lambda + \sqrt{(\lambda - 1 - c)^2 - 4c}|^2}. \tag{12}$$

Note that the function $\varphi(\cdot)$ is just a simplified version of the function $\delta(\cdot)$ from (7) under the null hypothesis and thus can be used in the construction of LSS. Moreover, by closer look at (12) the function $\varphi(\lambda)$ is not defined at point zero for $c < 1$ but $\varphi(0) = 0$ if $c > 1$. Thus, the domain of $\varphi(\cdot)$ can be chosen as $R \setminus \{0\}$ for $c < 1$ and the whole R in case $c > 1$. In Theorem 1 we present some properties of $\varphi(\lambda)$ in case $c < 1$ but note that very similar formula is true for $c > 1$ by a simple modification at point zero.

Theorem 1. *Let $\varphi : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ be defined as in (12) and let $f : (a, b) \to \mathbb{R}$ be a linear function with $(a, b) = ((1 - \sqrt{c})^2, (1 + \sqrt{c})^2)$, the support of the Marchenko–Pastur distribution, and let $c < 1$. Then, under the null hypothesis in (10), it holds that*

$$\varphi(f(\lambda_i)) = \begin{cases} 1 & f(\lambda_i) \in (a, b) \\ \frac{4f(\lambda_i)}{(c-1-f(\lambda_i)+\sqrt{(f(\lambda_i)-1-c)^2-4c})^2} & else \end{cases},$$

where λ_i are the eigenvalues of the sample covariance matrix.

The results of the theorem state that choosing the identity transformation $f(x) = x$ would lead to a degenerate statistic. Therefore, the definition of $\varphi(\lambda)$ in (12) should be adjusted in such a way, that the resulting test statistic will have a nondegenerate asymptotic distribution. This leads to the definition of a modification of $\varphi(\lambda)$ expressed as

$$\varphi_\varepsilon(\lambda) = \frac{4\lambda}{|(c - 1 - \lambda) + \frac{1}{2}\sqrt{(\lambda - 1 - c)^2 - 4c}|^2 + 4\varepsilon} = \frac{\lambda}{\lambda + \varepsilon}, \tag{13}$$

where the last equality follows from the proof of Theorem 1.

The application of the transformation (13), which is obtained by utilizing the properties of the nonlinear shrinkage estimator for the covariance matrix, the following LSS is obtained

$$T_\varepsilon = \sum_{i=1}^{p} \varphi_\varepsilon(\lambda_i) = \sum_{i=1}^{p} \frac{\lambda_i}{\lambda_i + \varepsilon}. \tag{14}$$

This is an LSS with a nondegenerate asymptotic distribution since $\varphi_1(\lambda)$ has only one singularity at $\lambda = -\varepsilon$ and it is analytic on the support of the Marchenko–Pastur distribution of λ. Therefore, it is possible to apply the CLT for LSS to T_ε. It should be noted that for $\varepsilon = 1$ the statistic T_ε is similar to the Bartlett-Nanda-Pillai (BNP) trace test statistic proposed by Pillai (1955), whose asymptotic distribution for a large-dimensional Fisher matrix is derived by Bodnar, Dette, and Parolya (2019). As such, the LSS in (14) generalizes the BNP trace test statistic.

The CLT for LSSs (see, Theorem 3.4 in Yao et al., 2015) states that the fluctuation of corrected T_ε around its limit is normally distributed with a specific mean μ and variance σ^2. In Theorem 2 the analytical expressions of the correction term, the mean, and the variance are present.

Theorem 2. *Assume that conditions (A1) and (A2) hold and let $\varepsilon > 0$ be arbitrary. Then, under the null hypothesis in (10), it holds that*

$$W = \sum_{i=1}^{p} \frac{\lambda_i}{\lambda_i + \varepsilon} + p\frac{A}{\sqrt{c}} \xrightarrow{D} N(\mu, \sigma^2),$$

for $p/n \to c > 0$ as $n \to \infty$ where

$$\mu = -(\kappa - 1)\frac{\varepsilon}{\sqrt{c}(A - B)(B^2 - 1)} - \beta\frac{\varepsilon A^3}{\sqrt{c}(A^2 - 1)}, \quad \sigma^2 = \kappa\frac{\varepsilon^2}{c(A - B)^4} + \beta\frac{\varepsilon^2 A^4}{c(A^2 - 1)^2}, \tag{15}$$

and

$$A = \frac{-c - \varepsilon - 1 + \sqrt{(c + \varepsilon + 1)^2 - 4c}}{2\sqrt{c}}, B = \frac{-c - \varepsilon - 1 - \sqrt{(c + \varepsilon + 1)^2 - 4c}}{2\sqrt{c}}.$$

From the definition of A and B, we directly get that $AB = 1$. Moreover, $A \neq B$, since $(c + \varepsilon + 1)^2 - 4c = (c - 1)^2 + \varepsilon^2 + 2\varepsilon(1 + c) > 0$ for $\varepsilon > 0$. Finally, it holds that $B < -1 < A < 0$.

To study the finite sample performance of the asymptotic results derived in Theorem 2, we consider the normalized version of T_ε statistics expressed as

$$Z = \frac{W - \mu}{\sqrt{\sigma^2}},$$

where μ and σ^2 are given in (15). Then, by Theorem 2, $Z \to N(0, 1)$ in distribution. Figure 1 depicts the histograms of the random variable Z calculated with $p = 128$, $n = 256$, and $\varepsilon = 1$ when the observation matrix is obtained by generating random variables from the standard normal distribution (left-hand side plot) and from the Gamma(4, 2) − 2 distribution (right-hand side plot). For the second figure Gamma(4, 2) − 2 distributed data are chosen because this gives $\beta = 3/2$ instead of $\beta = 0$, corresponding to the standard normal distribution. Moreover, it still has zero mean and unit variance. The results in the two plots are obtained for 100.000 independent replications.

Note that both empirical distribution functions in Figure 1 are well approximated by the standard normal distribution. This result holds for the moderate sample size $n = 256$. Moreover, the

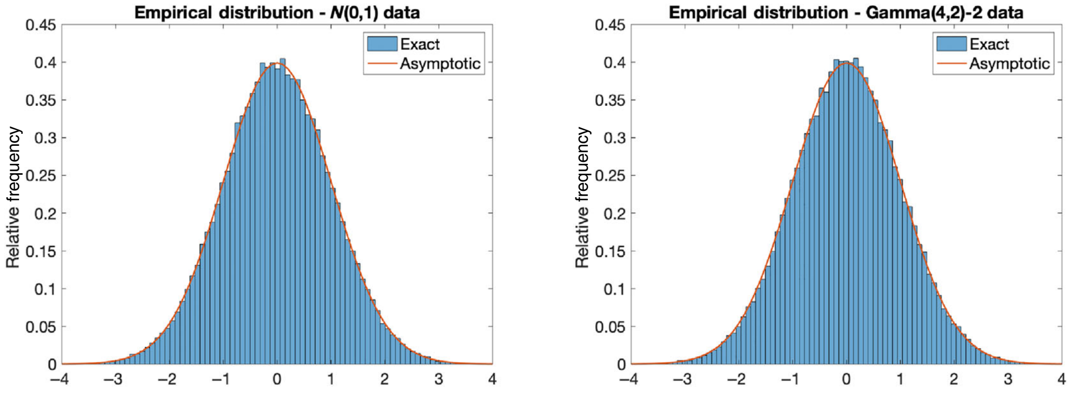

FIGURE 1 Empirical distribution functions for centralized random variable W calculated with $p = 128$, $n = 256$, $\epsilon = 1$ and based on the standard normal distribution (left) and the Gamma(4, 2) − 2 distribution (right).

random variable W is a ready-to-use statistic in the large-dimensional case, based on the nonlinear shrinkage estimator. We refer to the new test based on the statistic W as the nonlinear shrinkage test or for short the NLS-ϵ test. In Appendix A the new test will be compared with existing benchmark approaches within an extensive simulation study. Finally, we note that $\epsilon = 1$ was chosen arbitrarily in Figure 1. However, in the general case, ϵ needs to be chosen such that the power of the test is maximal. Determining the appropriate value of ϵ for different scenarios is challenging because it largely depends on the specific alternative hypothesis being considered. In future research, we will address this question by investigating the asymptotic power of the NLS test under various conditions.

4 | SUMMARY

Testing the structure of the covariance matrix is a challenging statistical problem with many potential applications in different fields of science, such as finance, economics, environmetrics, medical imaging analysis, signal processing, and wireless communications. The problem becomes even more difficult when the dimension of the covariance matrix is large, that is, when one should opt for the large-dimensional setting in the derivation of the asymptotic distribution of the test statistic. Even though several large-dimensional tests exist in the statistical literature, the research in this direction is still ongoing.

We contribute to the existing literature by extending the large-dimensional test based on the shrinkage approach to the test deduced from the nonlinear shrinkage estimator of the covariance matrix. The asymptotic distribution of the test statistic is derived under the null hypothesis by applying the theory developed for LSSs. The properties of the new approach are investigated within an extensive simulation study. It is concluded that the suggested test based on the nonlinear shrinkage approach outperforms the benchmark testing strategies for some alternative hypotheses, while the large-dimensional John test and the test derived from the linear shrinkage approach perform better for other alternative hypotheses.

DATA AVAILABILITY STATEMENT

Research data are not shared.

ORCID

Taras Bodnar https://orcid.org/0000-0001-7855-8221
Nestor Parolya https://orcid.org/0000-0003-2147-2288
Frederik Veldman https://orcid.org/0009-0005-1124-120X

REFERENCES

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

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

Bai, Z. D., & Silverstein, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *The Annals of Probability*, *32*(1A), 553–605.

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

Bodnar, O., Bodnar, T., & Parolya, N. (2022). Recent advances in shrinkage-based high-dimensional inference. *Journal of Multivariate Analysis*, *188*, 104826.

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

Bodnar, T., Dmytriv, S., Okhrin, Y., Parolya, N., & Schmid, W. (2021). Statistical inference for the expected utility portfolio in high dimensions. *IEEE Transactions on Signal Processing*, *69*, 1–14.

Bodnar, T., Dmytriv, S., Parolya, N., & Schmid, W. (2019). Tests for the weights of the global minimum variance portfolio in a high-dimensional setting. *IEEE Transactions on Signal Processing*, *67*(17), 4479–4493.

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., Gupta, A. K., & Parolya, N. (2016). Direct shrinkage estimation of large dimensional precision matrix. *Journal of Multivariate Analysis*, *146*, 223–236.

Bodnar, T., Okhrin, O., & Parolya, N. (2019). Optimal shrinkage estimator for high-dimensional mean vector. *Journal of Multivariate Analysis*, *170*, 63–79.

Bodnar, T., Okhrin, Y., & Parolya, N. (2023). Optimal shrinkage-based portfolio selection in high dimensions. *Journal of Business & Economic Statistics*, *41*(1), 140–156.

Bodnar, T., Parolya, N., & 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., & Veldman, F. (2024). *Linear shrinkage-based hypothesis test for large dimensional covariance matrix*. In S. Knoth, Y. Okhrin, & P. Otto (Eds.), *Advanced statistical methods in process monitoring, finance, and environmental science - essays in honour of Wolfgang Schmid*. Cham: Springer To appear.

Cai, T. T., & Jiang, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *The Annals of Statistics*, *39*(3), 1496–1525.

Chen, S. X., Zhang, L. X., & Zhong, P. S. (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association*, *105*(490), 810–819.

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

Dey, D. K., & Srinivasan, C. (1985). Estimation of a covariance matrix under Stein's loss. *The Annals of Statistics*, *13*, 1581–1591.

Fisher, T. J., Sun, X., & Gallagher, C. M. (2010). A new test for sphericity of the covariance matrix for high dimensional data. *Journal of Multivariate Analysis*, *101*(10), 2554–2570.

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

Gupta, A. K., & Bodnar, T. (2014). An exact test about the covariance matrix. *Journal of Multivariate Analysis*, *125*, 176–189.

Gupta, A. K., & Xu, J. (2006). On some tests of the covariance matrix under general conditions. *Annals of the Institute of Statistical Mathematics*, *58*, 101–114.

James, W., & Stein, C. (1961). Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, *1*, 361–379.

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

Kubokawa, T., & 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.

Ledoit, O., & Péché, S. (2011). Eigenvectors of some large sample covariance matrix ensembles. *Probability Theory and Related Fields*, *151*(1), 233–264.

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. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, *40*(2), 1024–1060.

Marchenko, V., & Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik*, *114*(4), 507–536.

Nagao, H. (1973). On some test criteria for covariance matrix. *The Annals of Statistics*, *1*(4), 700–709.

Nilsson, M. (2021). *A shrinkage test for large-dimensional covariance matrix*.

Perlman, M. D. (2007). *STAT 542: Multivarite statistical analysis*. Washington: Seatle.

Pillai, K. C. S. (1955). Some new test criteria in multivariate analysis. *Annals of Mathematical Statistics*, *26*(1), 117–121.

Silverstein, J. W., & Choi, S. I. (1995). Analysis of the limiting spectral distribution of large dimensional random matrices. *Journal of Multivariate Analysis*, *54*(2), 295–309.

Srivastava, M. S. (2005). Some tests concerning the covariance matrix in high dimensional data. *Journal of the Japan Statistical Society*, *35*(2), 251–272.

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, volume 1: Contributions to the theory of statistics* (pp. 197–206). Berkeley, CA: University of California Press. https://projecteuclid.org/euclid.bsmsp/1200501656

Versteegh, W. (2020). Central limit theorems for linear spectral statistics of large regularized covariance matrices. *arXiv:1912.06357v1*.

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

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, England: Cambridge University Press.

Zheng, S., Bai, Z., & Yao, J. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *The Annals of Statistics*, *43*(2), 546–591. https://doi.org/10.1214/14-AOS1292

How to cite this article: Bodnar, T., Parolya, N., & Veldman, F. (2025). Nonlinear shrinkage test on a large-dimensional covariance matrix. *Statistica Neerlandica*, *79*(1), e12348. https://doi.org/10.1111/stan.12348

APPENDIX A. SIMULATION STUDY

In this section, the nonlinear shrinkage test (NLS-ε) is compared with other tests in the large-dimensional framework. The considered benchmark approaches are the CLRT, the CJ test, both derived by Wang and Yao (2013), and the linear shrinkage (LS) test introduced in Bodnar et al. (2024). The CLRT and the CJ approaches present one-sided tests, while the NLS-ε and the LS methods are two-sided tests. Note that a one-sided test can be adapted to a two-sided test by reallocating the significance level equally across both tails of the distribution. This transformation will also be applied in this simulation to ensure all tests are compared on the same basis. Moreover, note that the CJ and the LS test statistics have the same limiting distribution and that the limiting distributions of the CLRT and the NLS-ε test statistics depend on c. In particular, the CLRT depends on the $\log(1-c)$. Therefore, this test breaks down when c increases to 1 and will not work when $c > 1$.

The null hypothesis that will be tested is given in (10) and we will work with the transformed sample in this section. During the comparison, we also look for which ε the NLS-ε test has the highest power using grid search for ε. The comparison is performed in terms of the empirical null-rejection probabilities and the empirical power.

A.1 Empirical rejection rates comparison

Before the tests are compared using their powers, their empirical rejection probabilities are compared. If a test statistic's exact distribution closely approximates its limiting distribution, then the empirical rejection rates should align closely with the desired significance level, α. Therefore, the closer the empirical rejection rate is to a fixed significance level α, the better approximation is. Also to make a fair comparison later on with the empirical powers, the empirical rejection rates of the tests should all be close to the significance level.

Without loss of generality, we set $\alpha = 0.05$ as the desired significance level and calculate the empirical rejection rates of the tests by using 10,000 independent repetitions with the results depicted in Tables A1 and A2. The empirical rejection rates in Table A1 are computed by drawing samples from the standard normal distribution, while the values in Table A2 are obtained by generating the elements of the observation matrix from the Gamma(4, 2) − 2 distribution as described in Section 3.

TABLE A1 Empirical rejection rates at 5% significance level when the elements of the observation matrix are drawn from the standard normal distribution.

(p, n)	CLRT	CJ	LS	NLS-10	NLS-1.5	NLS-1	NLS-0.5	NLS-0.1
(8,128)	0.0565	0.0581	0.0661	0.0480	0.0487	0.0495	0.0496	0.0523
(16,128)	0.0539	0.0552	0.0479	0.0451	0.0460	0.0463	0.0452	0.0475
(32,128)	0.0518	0.0525	0.0432	0.0458	0.0460	0.0469	0.0484	0.0512
(64,128)	0.0536	0.0538	0.0479	0.0503	0.0491	0.0483	0.0504	0.0520
(96,128)	0.0547	0.0540	0.0484	0.0440	0.0502	0.0500	0.0513	0.0527
(112,128)	0.0538	0.0553	0.0516	0.0556	0.0539	0.0531	0.0514	0.0499
(120,128)	0.0522	0.0524	0.0485	0.0477	0.0484	0.0479	0.0482	0.0478
(16,256)	0.0544	0.0531	0.0473	0.0449	0.0452	0.0458	0.0463	0.0477
(32,256)	0.0519	0.0502	0.0433	0.0512	0.0516	0.0517	0.0496	0.0492
(64,256)	0.0499	0.0499	0.0437	0.0500	0.0502	0.0492	0.0499	0.0498
(128,256)	0.0516	0.0541	0.0504	0.0514	0.0517	0.0509	0.0511	0.0498
(192,256)	0.0542	0.0503	0.0488	0.0535	0.0519	0.0505	0.0509	0.0496
(224,256)	0.0505	0.0512	0.0495	0.0503	0.0495	0.0502	0.0519	0.0511
(240,256)	0.0517	0.0513	0.0480	0.0460	0.0469	0.0472	0.0488	0.0499

Abbreviations: CJ, corrected John test; CLRT, corrected likelihood ratio test.

TABLE A2 Empirical rejection rates at 5% significance level when the elements of the observation matrix are drawn from the Gamma(4, 2) − 2 distribution.

(p,n)	CLRT	CJ	LS	NLS-10	NLS-1.5	NLS-1	NLS-0.5	NLS-0.1
(8,128)	0.2518	0.1178	0.0808	0.0480	0.0463	0.0460	0.0456	0.0479
(16,128)	0.2619	0.0911	0.0513	0.0469	0.0438	0.0436	0.0440	0.0485
(32,128)	0.2588	0.0750	0.0468	0.0498	0.0472	0.0460	0.0466	0.0496
(64,128)	0.2197	0.0645	0.0460	0.0485	0.0458	0.0470	0.0459	0.0492
(96,128)	0.1643	0.0537	0.0423	0.0489	0.0474	0.0459	0.0456	0.0464
(112,128)	0.1329	0.0601	0.0514	0.0511	0.0451	0.0444	0.0454	0.0452
(120,128)	0.1105	0.0598	0.0515	0.0482	0.0462	0.0462	0.0466	0.0489
(16,256)	0.2777	0.0861	0.0531	0.0495	0.0468	0.0470	0.0472	0.0488
(32,256)	0.2849	0.0723	0.0471	0.0485	0.0471	0.0471	0.0479	0.0499
(64,256)	0.2654	0.0625	0.0467	0.0488	0.0489	0.0499	0.0511	0.0514
(128,256)	0.2252	0.0591	0.0513	0.0532	0.0509	0.0521	0.0519	0.0511
(192,256)	0.1695	0.0572	0.0513	0.0508	0.0489	0.0485	0.0477	0.0500
(224,256)	0.1384	0.0554	0.0510	0.0519	0.0510	0.0503	0.0532	0.0537
(240,256)	0.1164	0.0547	0.0490	0.0505	0.0503	0.0489	0.0499	0.0502

Abbreviations: CJ, corrected John test; CLRT, corrected likelihood ratio test.

It can be seen in Table A1 that all the empirical rejection rates based on data generated from the standard normal distribution are close to the target significance level $\alpha = 0.05$. This means that the exact distributions of the test statistics are well approximated by their limiting distributions derived in the large-dimensional setting. Unfortunately, this is not the case when the empirical rejection rates are computed using data generated from the Gamma(4, 2) − 2 distribution.

In Table A2, we observe that the empirical rejection rates of the NLS-ε and LS tests behave quite well. For lower combinations of (p, n) the empirical rejection rates of the NLS-ε test seem a little low but overall they are close to the target significance level α. As such, it can be concluded that for these combinations of (p, n), the exact distributions of the test statistics of the NLS-ε and the LS tests are close to their large-dimensional asymptotic distributions when the data are drawn from the Gamma(4, 2) − 2 distribution. However, for the CJ test, this only holds when both p and n are relatively large. It looks like the empirical rejection rates of the CJ test approach α from above. This means that if p and n are relatively low, then the exact distribution has heavier tails than it should be. However, when p and n increase, then the sampling distribution of the CJ test statistic is getting closer to its limiting distribution. Thus, the CJ test relies more on the limiting aspect in this case. The empirical rejection rates for the CLRT behave quite poor for every combination of (p, n). They are approximately two times as large as they should be when p and n are large and the results are even worse for smaller values of p and n. From this observation, it can be concluded that when the data are generated from the Gamma(4, 2) − 2 distribution, then the exact distribution of the CLRT test statistic is not close to its limiting distribution. Therefore, it will be difficult to make a fair comparison based on the empirical power for the Gamma(4, 2) − 2 distribution, because not all tests will have the same starting point. So, the empirical power comparison will only be based on the data generated from the standard normal distribution.

A.2 Empirical power comparison

In this subsection, the empirical powers for all tests will be compared, including the NLS-ε test for different values of ε. The comparison will be based on the data drawn from the standard normal distribution. In this simulation study, the comparison is based on the increasing distance between the null hypothesis and a particular alternative hypothesis. Three alternative hypotheses are considered

1. H_1: Compound symmetry relation,
2. H_1: Autoregressive relation,
3. H_1: Heteroscedasticity.

The dimensions that will be 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 $c = 15/16$, repressively. The calculation of the empirical powers is based on 1,000 independent repetitions.

A.2.1 Compound symmetry relation

The first alternative hypothesis that will be used to make a power comparison is the compound symmetry relation. In this case, the covariance matrix under the alternative hypothesis is given by

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

The coefficient ρ runs from 0 to 1 in the simulation study. When $\rho = 0$, the $\boldsymbol{\Sigma}_{n,\rho}$ is the identity matrix and the null hypothesis is true. As ρ increases, the covariance matrix $\boldsymbol{\Sigma}_{n,\rho}$ becomes less like the identity matrix, that is, the distance from the null hypothesis to the alternative hypothesis increases. After some investigations, it seems like the most important ε's under the compound symmetry alternative for the NLS-ε test are $\varepsilon = 1.5$, $\varepsilon = 1$ and $\varepsilon = 0.5$. We use these values in the comparison study.

In Figure A1, it can be seen how the tests perform in terms of the empirical power for different combinations of $\frac{p}{n}$. For $\rho > 0.15$, all tests have power 1 and, as expected, the CJ and the LS tests behave nearly the same. Most noticeable in Figure A1 is the observation that when p increases, then the NLS-ε performs better and the CLRT becomes worse. For $(p, n) = (120{,}128)$ the NLS-ε test outperforms the CLRT test. Also, the CLRT test breaks down when p is getting closer to n. Overall, the CJ and the LS tests perform best because they are the first to reach a power of 1 for every combination of (p, n).

Now focusing only on the NLS-ε test, it can be seen in Figure A1 that when p increases the performance of the NLS-ε test changes as well. For small p it seems that the NLS-0.5 performs best but when $p = 96$ and $p = 120$ the NLS-1 test performs best with minimum difference. Therefore, the optimal ε for the NLS-ε test depends on the combination of (p, n) under the compound symmetry alternative.

A.2.1 Autoregressive relation

The second alternative hypothesis is the autoregressive relation, based on an autoregressive model. The autoregressive model states that the output variable depends linearly on its previous

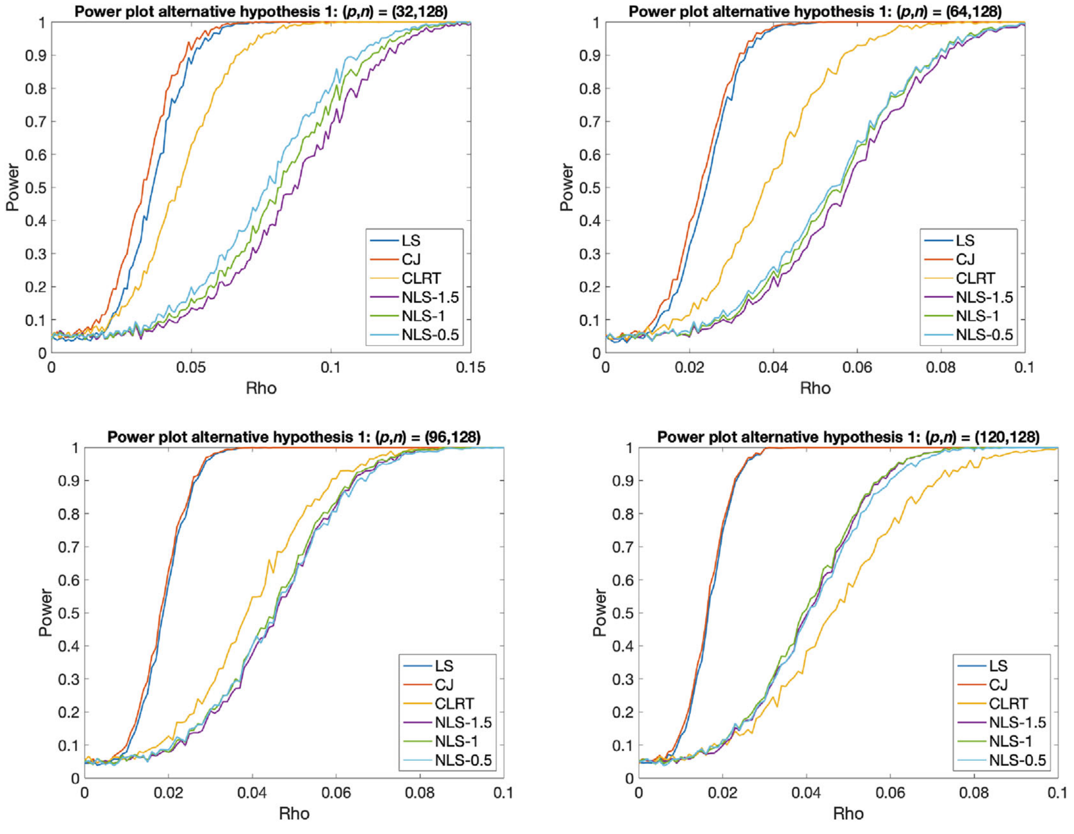

FIGURE A1 Empirical powers under alternative hypothesis 1 for $\rho \in (0, 1)$.

values and an error term. In the autoregressive alternative hypothesis, the entries of the matrix $\boldsymbol{\Sigma}_{n,\Delta}$ also depend recursively on each other. For $\Delta \in \mathbb{R}$, the entry on the ith row and the jth column of the alternative hypothesis matrix is $\Delta^{|i-j|}$. The autoregressive alternative hypothesis is then given by

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

In the simulation study, $\Delta \in (-1, 1)$ is chosen, because this corresponds to a stationary autoregressive model. The simulation goes 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 in (10) because the covariance matrix becomes less like the identity matrix. After some pre-analysis the most important ε's to consider for NLS-ε test are $\varepsilon = 1$, $\varepsilon = 0.5$, and $\varepsilon = 0.1$. Note that not all of these ε's are the same as in the simulation for the first alternative hypothesis.

It can be seen in Figure A2 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. The NLS-ε test performs the worst for p small. Then, when p gets bigger the CJ and LS tests are still performing

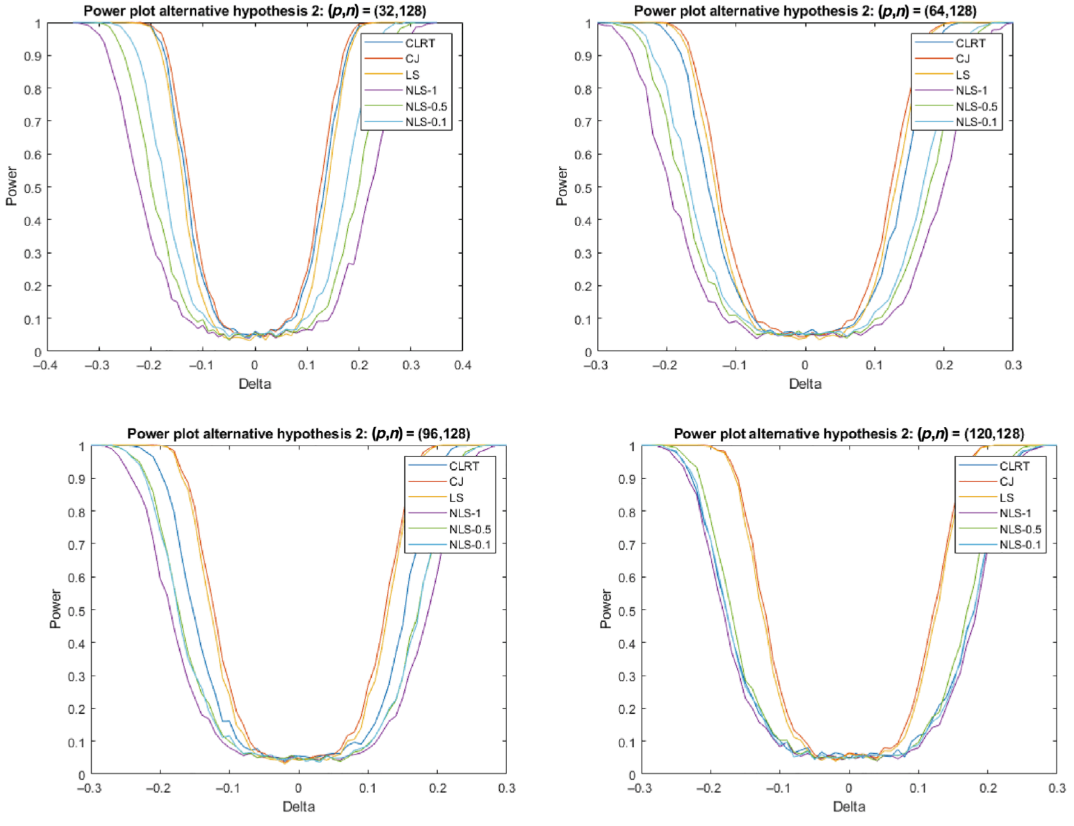

FIGURE A2 Empirical powers under alternative hypothesis 2 for $\Delta \in (-1, 1)$.

best, but the CLRT test is performing worse and worse. For $p = 120$ the NLS-ε test outperforms the CLRT test when $\varepsilon = 0.5$. This is in line with the observation from the previous simulation for the compound symmetry alternative, only now with different ε.

Focusing on the NLS-ε test, it can be seen in Figure A2 that for $p = 32$ and $p = 64$, the NLS-0.1 test performs best. When p gets larger, the NLS-0.5 test takes the lead. Again the the optimal ε varies with different values of p compared to n and it is in line with the observation from the simulation study performed under the compound symmetry alternative.

A.2.2 Heteroscedasticity

The third alternative hypothesis that will be considered is the alternative hypothesis where a fixed ratio r of the variables has a variance not equal to 1, but equal to $1 + \gamma$. In econometrics, this is also called a heteroscedasticity alternative. For any $r \in (0, 1)$ and $\gamma \in \mathbb{R}$, the covariance matrix under the third alternative hypothesis is defined as

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

If $r = 1/2$, then half of the variables have variances equal to $1+\gamma$. If $r \cdot p$ is not a whole number, it will be rounded down. In the same way, as in the previous simulation, γ will run from -1 to 1. This can again be seen as departing from the null hypothesis in (10) when γ goes away from 0 in both directions because the covariance matrix $\mathbf{\Sigma}_{n,r,\gamma}$ under the alternative hypothesis becomes less like the identity matrix. After some pre-analysis of the NLS-ϵ test under the third alternative hypothesis, we find that the most important ϵ's to consider are $\epsilon = 10$, $\epsilon = 1$ and $\epsilon = 0.1$.

Figure A3 shows that the NLS-ϵ test performs by far the best. The NLS-ϵ test reaches the power of 1 much faster than the other competitors. For low c the CJ, LS, and CLRT tests are again quite comparable. However, when p increases the CLRT is getting worse and worse for the same reason as in the previous simulations. As a result, it can be concluded that the NLS-ϵ test performs best. Furthermore, it can be seen that the NLS-ϵ tests are symmetric around zero but the other tests are not because the powers of the other tests increase much faster for negative values of γ than for positive values.

In Figures A4 and A5, it can be seen that the NLS-ϵ test works better when r increases from 0 to 1. The other tests only perform better when r increases to $1/2$ only because when $r > 1/2$ the power decreases again, especially when γ is positive. This behavior is explained by the fact that the distributions of the test statistics of the CJ, LS, and CLRT tests remain the same when the covariance matrix is multiplied by a constant. This means that for $r = 1/2$ the alternative hypothesis is furthest away from the null hypothesis and should give the highest powers. Therefore, it can be concluded that the CJ, LS, and CLRT tests are invariant under multiples of the identity

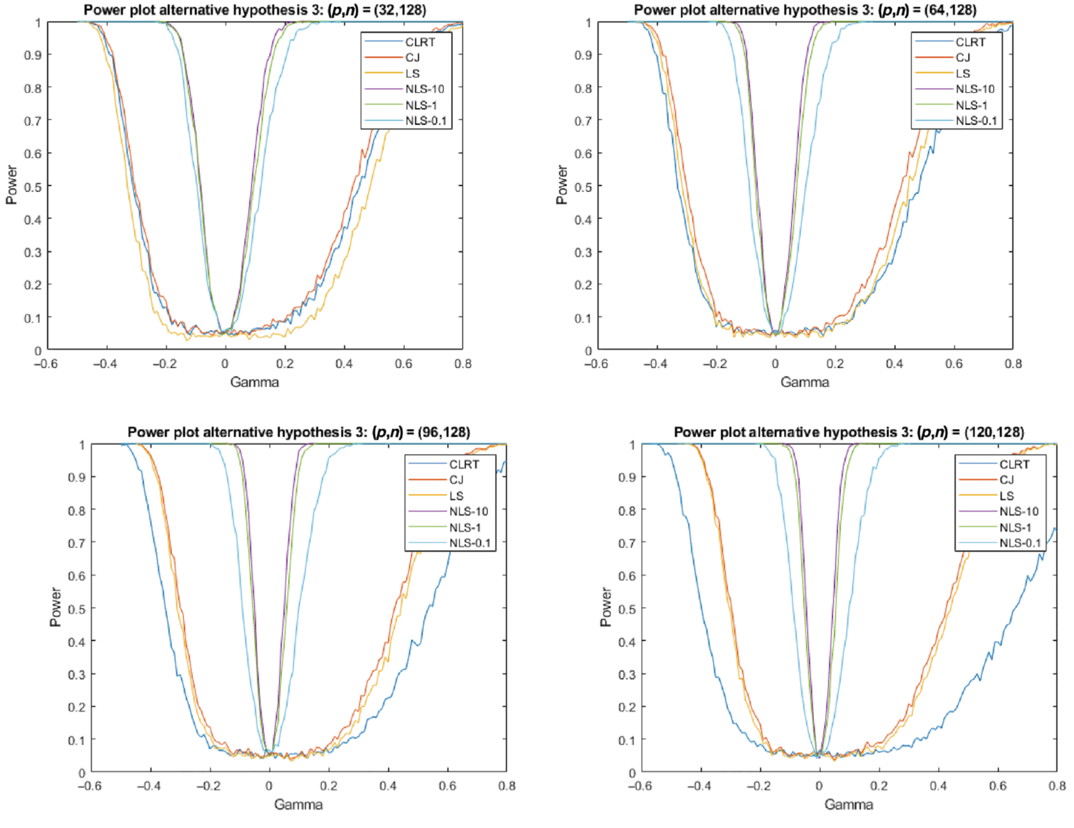

FIGURE A3 Empirical powers under alternative hypothesis 3 for $\gamma \in (-1, 1)$ and $r = 1/2$.

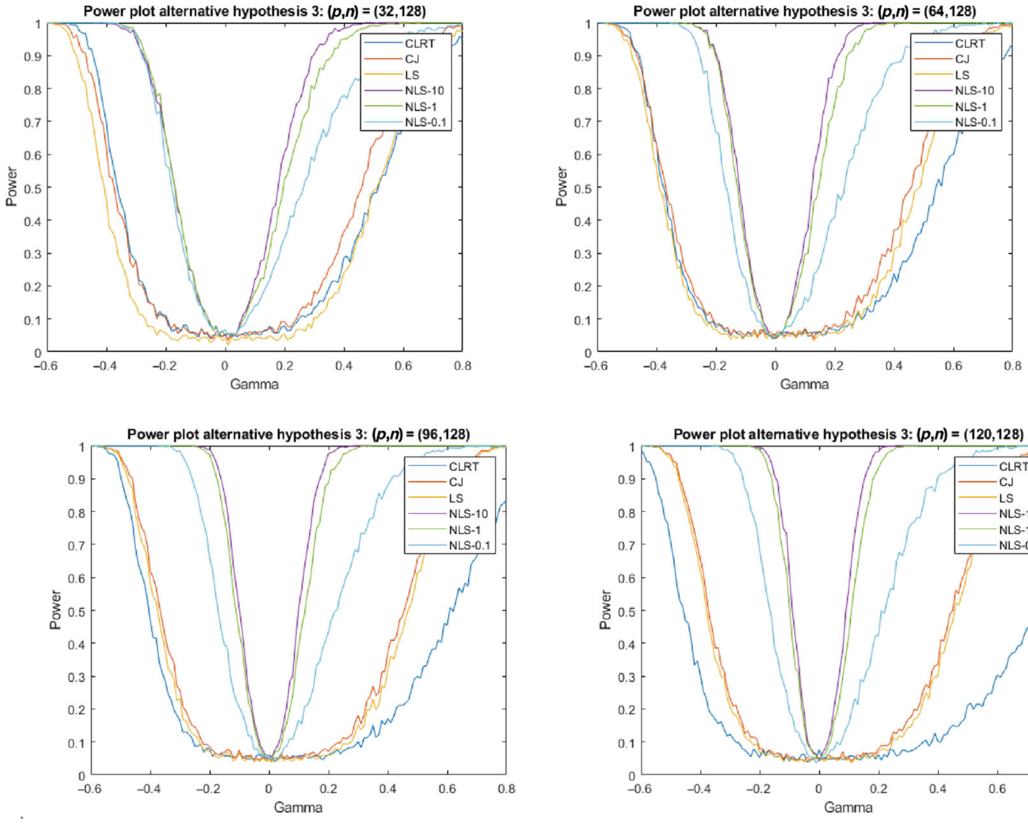

FIGURE A4 Empirical powers under alternative hypothesis 3 for $\gamma \in (-1, 1)$ and $r = 1/4$.

matrix which was already expected. In contrast, the NLS-ε is not invariant under the multiplication of the identity matrix by a constant, and thus the test gets more powerful when r increases from 0 to 1.

Finally, it can be seen in Figures A3–A5 that for all p's the NLS-10 test performs best and NLS-0.1 the worst. This is not in line with the findings from the previous alternative hypotheses where the optimal ε depends on p.

A.3 Combining compound symmetry and heteroscedasticity

It is observed in the previous subsections that the CJ and LS tests are the best-performing tests under the compound symmetry alternative. However, the NLS-ε test outperforms the others under the heteroscedasticity alternative. This makes it interesting to combine the two alternative hypotheses, and to examine which combinations of the variables (ρ, γ) will lead to the best test. It is expected that the performance of the tests depends on the tradeoff between the two variables, since for $(\rho, 0)$ the CJ and LS tests perform better and for $(0, \gamma)$ the NLS-ε test is the best approach. For any $r \in (0, 1)$, $\rho \in (0, 1)$ and $\gamma \in (-1, 1)$, the covariance matrix under the fourth alternative hypothesis is defined as

$$\mathbf{\Sigma}_{n,\rho,\gamma,r} = \mathbf{\Sigma}_{n,\gamma,r}^{1/2} \cdot \mathbf{\Sigma}_{n,\rho} \cdot \mathbf{\Sigma}_{n,\gamma,r}^{1/2}.$$

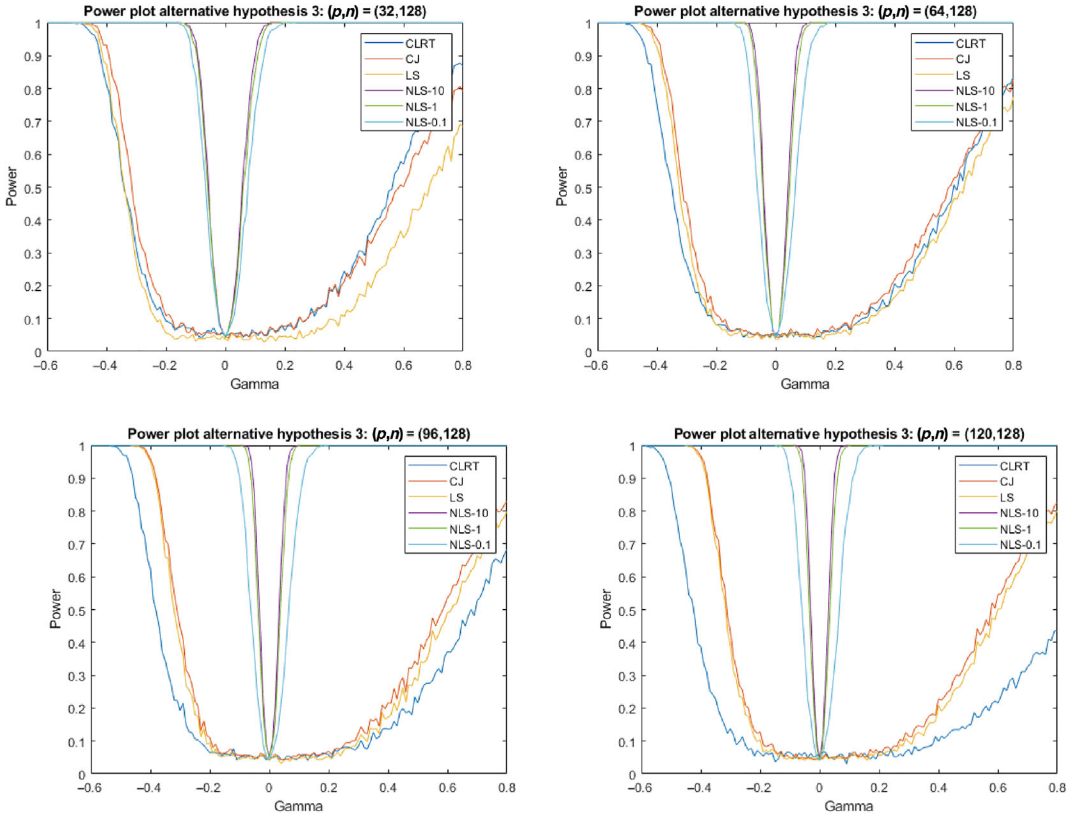

FIGURE A5 Empirical powers under alternative hypothesis 3 for $\gamma \in (-1, 1)$ and $r = 3/4$.

The fourth alternative hypothesis means that for $\rho \neq 0$ there is a positive correlation between the variables and a fixed ratio r of these variables have variance other than one, namely $1 + \gamma$. The fourth alternative depends now on three parameters, ρ, r, and γ. Therefore, as in the previous simulations, ρ will run from 0 to 1, γ from -1 to 1, and $r = 1/2$ is taken arbitrarily. This can again be seen as departing from the null hypothesis when ρ increases from 0 and γ moving away from 0 in both directions. This is because the covariance matrix $\mathbf{\Sigma}_{n,\rho,\gamma,r}$ becomes less like the identity matrix under the alternative hypothesis. The simulation will give different empirical powers for each combination of (ρ, γ). Plotting ρ and γ against the obtained empirical powers will then result in a three-dimensional power plot to compare the tests. The NLS-ε test that will be chosen is the NLS-1 test because it performs over both alternative hypotheses on average the best. Differently from the previous simulations, this simulation will only be performed for $p = 64$.

In Figure A6, the empirical powers under the fourth alternative hypothesis are plotted. It can be noted that the NLS-1 test does not seem to gain any power when γ is close to zero and ρ runs from 0 to 1. Moreover, from this figure, it is not immediately clear whether the empirical powers behave just as in Figure A1 and Figure A3, that is, when one of the variables ρ or γ is close to zero. Therefore, in Figure A7a the empirical powers are plotted for $\gamma \in (0, 0.01)$ and in Figure A7b for $\rho \in (0, 0.01)$, that is, for γ and ρ small. It can be seen in these figures that the tests behave just as expected. The only difference is that the lines are changed for planes.

Next, we investigate what happens when the variables ρ and γ are not close to zero, that is, when the deviation from the null hypothesis increases in both directions. The empirical powers

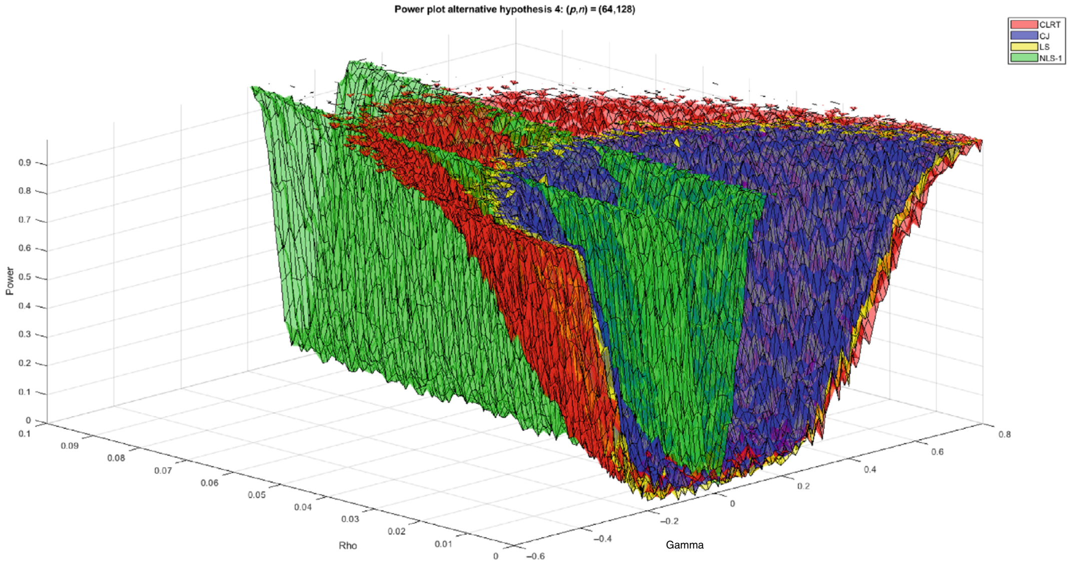

FIGURE A6 Empirical powers under alternative hypothesis 4 for $p = 64$, $\rho \in (0, 1)$, $\gamma \in (-1, 1)$ and $r = 1/2$.

of the CJ, LS, and CLRT tests increase, as expected, following the results presented in Figures A1 and A3. However, the NLS-1 test shows strange behavior. The power does not increase in the ρ direction when γ moves away from zero. Therefore, we focus on this property of the NLS-1 test in the following.

In Figure A8 the empirical power of the NLS-1 test is plotted for both ρ and γ small. From this figure, it can be concluded that when ρ or γ is equal to zero the test behaves as expected. However, when ρ and γ are lying in a particular region, the empirical power is very low. From Figure A8 it can be deduced that if γ is approximately equal to $\frac{3}{2} \cdot \rho$, the NLS-1 test will not increase in power.

APPENDIX B. AUXILIARY RESULTS

The following auxiliary results are Theorem 3.4, Proposition 3.6, and Proposition 2.10 from Yao et al. (2015). Theorem 3.4 is the CLT for LSSs. This theorem states that the fluctuations of a LSS around its limit are normally distributed under some conditions and for a specific class of functions. Proposition 3.6 helps reduce the difficulty of the calculations, while Proposition 2.10 gives a way to calculate the limit of a LSS. The three results are summarized below.

Lemma 1. *(CLT for LSSs) Let $f_1, \dots, f_k$ be functions analytic on an open region containing the support of F_c. Then under Assumptions (A1) and (A2) the random vector $\{X_n(f_1), \dots, X_n(f_k)\}$ where*

$$X_n(f) = p\{F^{S_n}(f) - F_{c_n}(f)\},$$

converges weakly for $p/n \to c > 0$ as $n \to \infty$ to a Gaussian vector $\{X(f_1), \dots, X(f_k)\}$ with mean function and covariance function expressed as

$$\mathbb{E}[X_f] = (\kappa - 1)I_1(f) + \beta I_2(f), \qquad \text{cov}(X_f, X_g) = \kappa J_1(f, g) + \beta J_2(f, g),$$

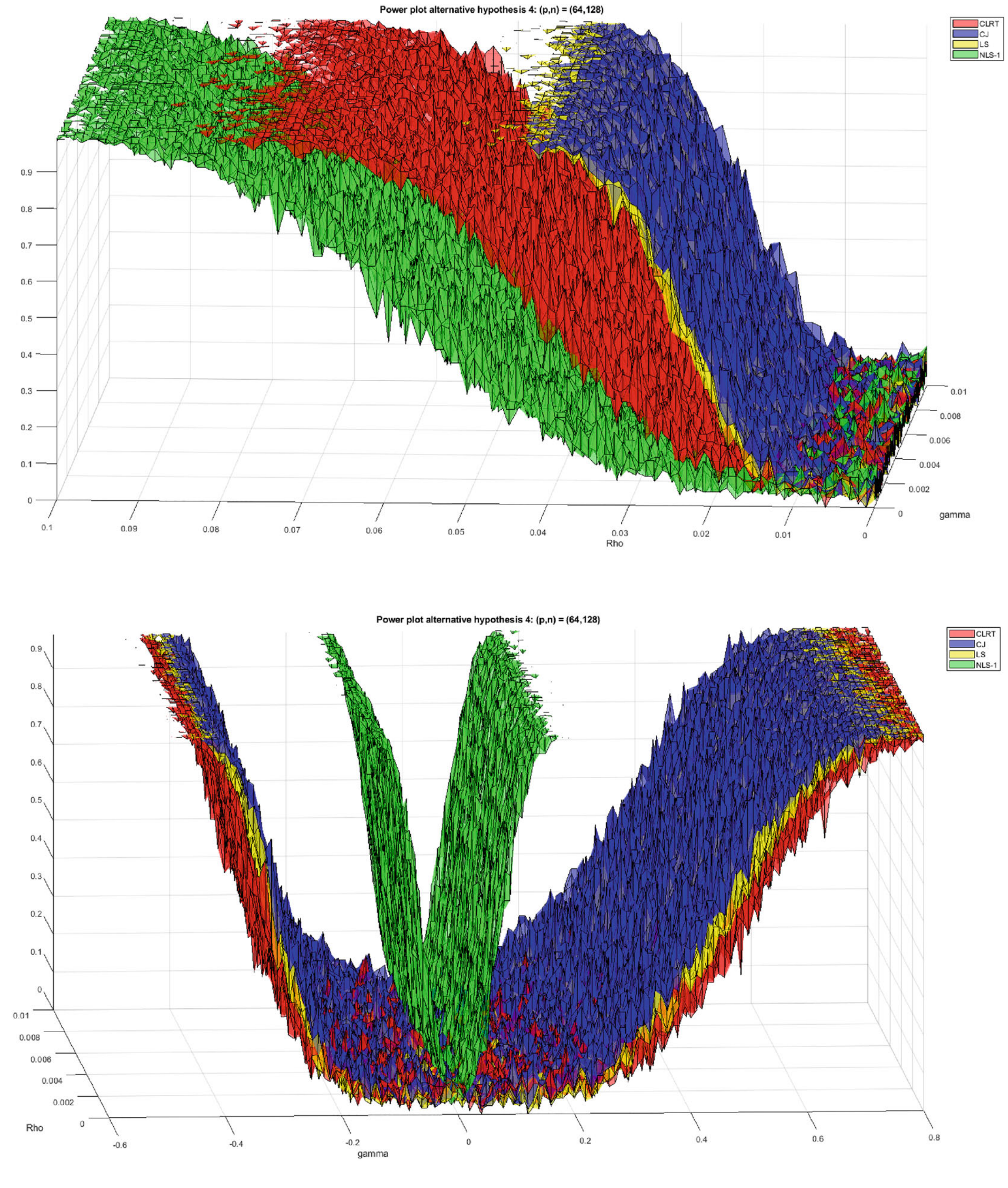

FIGURE A7 Empirical powers under alternative hypothesis 4 for $p = 64$, $r = 1/2$, $\rho \in (0, 0.1)$ and $\gamma \in (0, 0.01)$ (upper plot) and $\rho \in (0, 0.01)$ and $\gamma \in (-0.6, 0.8)$ (lower plot).

where

$$I_1(f) = \frac{1}{2\pi i} \oint \frac{c\{\underline{s}/(1+\underline{s})\}^3(z) f(z)}{[1 - c\{\underline{s}/(1+\underline{s})\}^2]^2} dz, \qquad I_2(f) = \frac{1}{2\pi i} \oint \frac{c\{\underline{s}/(1+\underline{s})\}^3(z) f(z)}{1 - c\{\underline{s}/(1+\underline{s})\}^2} dz,$$

$$J_1(f, g) = \frac{1}{4\pi^2} \oint \oint \frac{f(z_1) g(z_2)}{(m(z_1) - m(z_2))^2} m'(z_1) m'(z_2) dz_1 dz_2,$$

$$J_2(f, g) = \frac{-1}{4\pi^2} \oint f(z_1) \frac{\partial}{\partial z_1} \left\{ \frac{\underline{s}}{1+\underline{s}}(z_1) \right\} dz_1 \cdot \oint g(z_2) \frac{\partial}{\partial z_2} \left\{ \frac{\underline{s}}{1+\underline{s}}(z_2) \right\} dz_2,$$

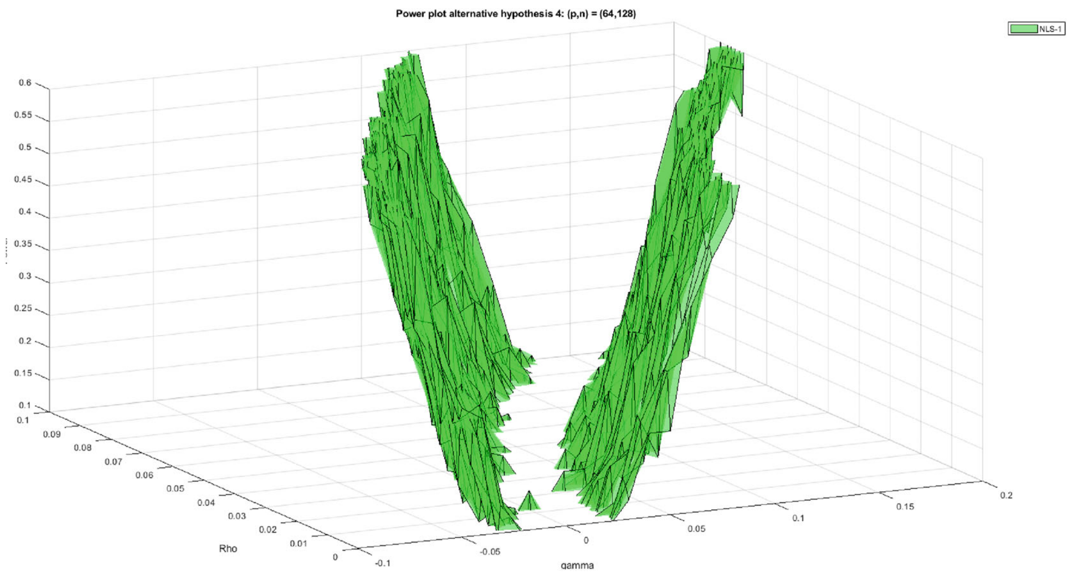

FIGURE A8 Empirical power of the NLS-1 test under alternative hypothesis 4 for $p = 64$ and $r = 1/2$.

the integrals are taken along contours (nonoverlapping in J_1) enclosing the support of F_c and $\underline{s}$ satisfies the equation

$$z = -\frac{1}{\underline{s}} + \frac{c}{1+\underline{s}}, \quad z \in \mathrm{C}^+.$$

As can be seen above, calculations of difficult line integrals are needed in the CLT. Fortunately, these calculations can significantly be simplified using the following results.

Lemma 2. *Let $\gamma := \{z \in C : |z| = 1\}$, $\gamma_1 := \{z_1 \in C : |z_1| = 1\}$, and $\gamma_2 := \{z_2 \in C : |z_2| = 1\}$. Then, the limiting parameters in Lemma 1 are expressed as follows*

$$I_1(f) = \lim_{r \downarrow 1} I_1(f, r), \qquad I_2(f) = \frac{1}{2\pi i} \oint_\gamma f(|1+hz|^2) \frac{1}{z^3} dz,$$

$$J_1(f,g) = \lim_{r \downarrow 1} J_1(f,g,r), \qquad J_2(f,g) = -\frac{1}{4\pi^2} \oint_\gamma \frac{f(|1+hz_1|^2)}{z_1^2} dz_1 \cdot \oint_\gamma \frac{g(|1+hz_2|^2)}{z_2^2} dz_2,$$

with

$$I_1(f,r) = \frac{1}{2\pi i} \oint_\gamma f(|1+hz|^2) \left[\frac{z}{z^2 - r^{-2}} - \frac{1}{z} \right] dz,$$

$$J_1(f,g,r) = -\frac{1}{4\pi^2} \oint_{\gamma_1} \oint_{\gamma_2} \frac{f(|1+hz_1|^2) g(|1+hz_2|^2)}{(z_1 - r z_2)^2} dz_1 dz_2,$$

where $h = \sqrt{c}$.

To calculate the limit of a LSS, the following result is useful:

Lemma 3. *Let* $\gamma := \{z \in \mathbb{C} : |z| = 1\}$. *For the standard Marchenko–Pastur distribution* F_c *with index* $c > 0$ *and* $\sigma^2 = 1$ *and for all functions* f *analytic on a domain containing the support interval* $[a, b] = [(1 - \sqrt{c})^2, (1 + \sqrt{c})^2]$, *it holds that*

$$\int f(x)dF_c(x) = -\frac{1}{4\pi i}\oint_\gamma \frac{f(|1+\sqrt{c}z|^2)(1-z^2)^2}{z^2(1+\sqrt{c}z)(z+\sqrt{c})}dz.$$

APPENDIX C. PROOFS OF THEOREMS 1 AND 2

Proof of Theorem 1. Let $f : (a, b) \to \mathbb{R}$ be a linear function with $(a, b) = ((1 - \sqrt{c})^2, (1 + \sqrt{c})^2)$, the support of the Marchenko–Pastur distribution, and let $f(\lambda_i) \in (a, b)$. Equation (12) can be written as

$$\varphi(\lambda_i) = \frac{4\lambda_i}{|c - 1 - \lambda_i + \sqrt{(\lambda_i - 1 - c)^2 - 4c}|^2} = \frac{4\lambda_i}{|c - 1 - \lambda_i + \sqrt{(a - \lambda_i)(b - \lambda_i)}|^2}.$$

Then,

$$(\varphi \circ f)(\lambda_i) = \varphi(f(\lambda_i)) = \frac{4f(\lambda_i)}{|c - 1 - f(\lambda_i) + \sqrt{(a - f(\lambda_i))(b - f(\lambda_i))}|^2}.$$

Now since $a < f(\lambda_i) < b$, the expression in the square root is negative and real. Therefore, it is possible to write this as

$$\sqrt{(b - f(\lambda_i)(a - f(\lambda_i)} = i\sqrt{(b - f(\lambda_i)(f(\lambda_i) - a)},$$

where the principal branch for the square root is used. Hence,

$$\varphi(f(\lambda_i)) = \frac{4f(\lambda_i)}{|c - 1 - f(\lambda_i) + i\sqrt{(b - f(\lambda_i)(f(\lambda_i) - a)}|^2}.$$

Let $z = c - 1 - f(\lambda_i) + i\sqrt{(b - f(\lambda_i)(f(\lambda_i) - a)}$. Then

$$\begin{aligned}
\mathrm{Re}(z)^2 &= (c - 1 - f(\lambda_i))^2 = c^2 - 2cf(\lambda_i) + f(\lambda_i)^2 - 2c + 2f(\lambda_i) + 1,\\
\mathrm{Im}(z)^2 &= \left(\sqrt{(b - f(\lambda_i)(f(\lambda_i - a)}\right)^2 = \left(\sqrt{(-1)((f(\lambda_i) - 1 - c)^2 - 4c)}\right)^2\\
&= -c^2 + 2cf(\lambda_i) - f(\lambda_i)^2 + 2c + 2f(\lambda_i) - 1.
\end{aligned}$$

As such, $|z|^2 = \mathrm{Re}(z)^2 + \mathrm{Im}(z)^2 = 4f(\lambda_i)$ and, consequently,

$$\varphi(f(\lambda_i)) = \frac{4f(\lambda_i)}{|c - 1 - f(\lambda_i) + \sqrt{(f(\lambda_i) - 1 - c)^2 - 4c}|^2} = \frac{4f(\lambda_i)}{4f(\lambda_i)} = 1.$$

Now, let $f(\lambda_i) \notin (a, b)$, i.e., $f(\lambda_i) < a$ or $f(\lambda_i) > b$. Then,

$$(\varphi \circ f)(\lambda_i) = \varphi(f(\lambda_i)) = \frac{4f(\lambda_i)}{|c - 1 - f(\lambda_i) + \sqrt{(a - f(\lambda_i))(b - f(\lambda_i))}|^2}.$$

Since $f(\lambda_i) < a$ or $f(\lambda_i) > b$, then the expression in the square root is always positive. Combining the result with the previous finding, we get the statement of the theorem. ■

Proof of Theorem 2. Consider the LSS $T_\varepsilon = \sum_{i=1}^{p} \frac{\lambda_i}{\lambda_i + \varepsilon}$ for $\varepsilon > 0$ with $\varphi(\lambda) = \frac{\lambda}{\lambda + \varepsilon}$. By Lemma 1, we get under $H_0 : \Sigma_n = I$ that

$$p\{F^{S_n}(\varphi) - F_{c_n}(\varphi)\},$$

converges weakly to a Gaussian distribution with mean and variance given by

$$\mu = (\kappa - 1)I_1(\varphi) + \beta I_2(\varphi), \qquad \sigma^2 = \kappa J_1(\varphi, \varphi) + \beta J_2(\varphi, \varphi), \tag{C1}$$

where the closed-form expressions of the mean μ, the variance σ^2 and the correction factor $F_{c_n}(\varphi)$ are given below in Lemmas 4,5, and 6, respectively. ■

Before starting the proofs of the lemmas, we note that

$$|1 + hz|^2 = (1 + hz)\overline{(1 + hz)} = (1 + hz)(1 + \frac{h}{z}) = \frac{1}{z}(z + h)(1 + zh),$$

where $z \in \mathbb{C}$ runs over the unit circle with complex conjugate $\bar{z} = \frac{1}{z}$.

Lemma 4. *Under the statement of Theorem 2, it holds that*

$$\mu = -(\kappa - 1)\frac{\varepsilon}{\sqrt{c}(A - B)(B^2 - 1)} - \beta \frac{\varepsilon A^3}{\sqrt{c}(A^2 - 1)}.$$

Proof. By Lemma 1, it holds that $\mu = (\kappa - 1)I_1(\varphi) + \beta I_2(\varphi)$. Next, we use Lemma 2 to compute $I_1(\varphi)$ and $I_2(\varphi)$. Let $\gamma := \{z \in C : |z| = 1\}$. By Lemma 2 we get that

$$\begin{aligned}
I_1(\varphi) &= \lim_{r \downarrow 1} \frac{1}{2\pi i} \oint_\gamma \frac{|1 + hz|^2}{|1 + hz|^2 + \varepsilon} \left[\frac{z}{z^2 - r^{-2}} - \frac{1}{z}\right] dz \\
&= \lim_{r \downarrow 1} \frac{1}{2\pi i} \oint_\gamma \frac{\frac{1}{z}(z + h)(1 + zh)}{\frac{1}{z}(z + h)(1 + zh) + \varepsilon} \frac{r^{-2}}{z(z^2 - r^{-2})} dz \\
&= \lim_{r \downarrow 1} \frac{1}{hr^2} \frac{1}{2\pi i} \oint_\gamma \frac{(z + h)(1 + zh)}{(z - A)(z - B)(z - \frac{1}{r})(z + \frac{1}{r})z} dz,
\end{aligned}$$

where

$$A \frac{-h^2 - \varepsilon - 1 + \sqrt{(h^2 + \varepsilon + 1)^2 - 4h^2}}{2h}, B \frac{-h^2 - \varepsilon - 1 - \sqrt{(h^2 + \varepsilon + 1)^2 - 4h^2}}{2h}.$$

For notation reasons the above defined A and B as a function of $h = \sqrt{c}$ will be used throughout the appendix. Moreover, from the definition of A and B, we get that $AB = 1$

and $B < -1 < A < 0$. As such, A lies inside γ and B lies outside γ. Moreover, since $r > 1$ we have that $\pm\frac{1}{r}$ lie inside γ. Therefore, the function inside the contour integral has four simple poles inside γ, $z = 0$, $z = \frac{1}{r}$, $-\frac{1}{r}$ and $z = A$. To evaluate the contour integral we need to calculate the residues of these poles. It holds that

$$Res(0) = \lim_{z\to\frac{1}{r}} z\frac{(z+h)(1+zh)}{z(z-A)(z-B)(z-\frac{1}{r})(z+\frac{1}{r})} = \frac{-hr^2}{AB} = -hr^2,$$

$$Res(1/r) = \lim_{z\to\frac{1}{r}}(z-\frac{1}{r})\frac{(z+h)(1+zh)}{z(z-A)(z-B)(z-\frac{1}{r})(z+\frac{1}{r})} = \frac{r^2(hr+1)(r+h)}{2(Ar-1)(Br-1)},$$

$$Res(-1/r) = \lim_{z\to-\frac{1}{r}}(z+\frac{1}{r})\frac{(z+h)(1+zh)}{z(z-A)(z-B)(z-\frac{1}{r})(z+\frac{1}{r})} = \frac{r^2(hr-1)(r-h)}{2(Ar+1)(Br+1)},$$

$$Res(A) = \lim_{z\to A}(z-A)\frac{(z+h)(1+zh)}{z(z-A)(z-B)(z-\frac{1}{r})(z+\frac{1}{r})} = \frac{r^2(A+h)(1+Ah)}{A(A-B)(r^2A^2-1)}.$$

Using Cauchy's Residue Theorem and the facts that $AB = 1$ and

$$(A+h)(1+Ah) = A+h+A^2h+Ah^2 = A(1+h^2+(A+B)h) = -\varepsilon A,$$

we find that

$$I_1(\varphi) = \lim_{r\downarrow1}\frac{1}{2\pi i}\left[2\pi i(Res(A)+Res(1/r)+Res(-1/r)+Res(0))\right]$$
$$= \lim_{r\downarrow1}\left[\frac{-\varepsilon}{h(A-B)(r^2A^2-1)} + \frac{(hr+1)(r+h)}{2h(Ar-1)(Br-1)} + \frac{(hr-1)(r-h)}{2h(Ar+1)(Br+1)} - 1\right]$$
$$= \lim_{r\downarrow1}\left[\frac{-\varepsilon}{h(A-B)(r^2A^2-1)} - \frac{\varepsilon r^2(A+B)}{h(r^2A^2-1)(r^2B^2-1)}\right] = \frac{-\varepsilon}{h(A-B)(B^2-1)}.$$

Next, we will calculate $I_2(\varphi)$. By Lemma 2 it holds that

$$I_2(\varphi) = \frac{1}{2\pi i}\oint_\gamma \frac{|1+hz|^2}{|1+hz|^2+\varepsilon}\frac{1}{z^3}dz = \frac{1}{2h\pi i}\oint_\gamma \frac{(z+h)(1+zh)}{(z-A)(z-B)z^3}dz.$$

The function inside the contour integral has one simple pole $z = A$ and one pole $z = 0$ of order 3. The residues of these poles are given by

$$Res(A) = \lim_{z\to A}(z-A)\frac{(z+h)(1+zh)}{(z-A)(z-B)z^3} = \frac{(A+h)(1+Ah)}{A^2(A^2-1)} = \frac{-\varepsilon}{A(A^2-1)},$$

and

$$Res(0) = \frac{1}{2}\lim_{z\to0}\frac{d^2}{dz^2}(z-0)^3\frac{(z+h)(1+zh)}{(z-A)(z-B)z^3} = \frac{1}{2}\lim_{z\to0}\frac{d^2}{dz^2}\frac{(z+h)(1+zh)}{(z-A)(z-B)}$$
$$= \frac{h}{AB} + (1+h^2)\left(\frac{1}{A^2B}+\frac{1}{AB^2}\right) + h\left(\frac{1}{A^3B}+\frac{1}{A^2B^2}+\frac{1}{AB^3}\right)$$
$$= h + (1+h^2)(A+B) + h(A^2+B^2+1) = (A+B)(1+h^2+h(A+B)) = -\varepsilon(A+B).$$

Therefore, by Cauchy's Residue Theorem it holds that

$$I_2(\varphi) = \lim_{r\downarrow 1} \frac{1}{2h\pi i}\left[2\pi i(Res(A) + Res(0))\right] = \frac{-\varepsilon}{h}\left(A + B + \frac{1}{A(A^2-1)}\right) = \frac{-\varepsilon A^3}{h(A^2-1)}.$$

Combining $I_1(\varphi)$ and $I_2(\varphi)$ and using $h = \sqrt{c}$, we get the statement of the lemma. ■

Lemma 5. *Under the statement of Theorem 2, it holds that*

$$\sigma^2 = \kappa \frac{\varepsilon^2}{c(A-B)^4} + \beta \frac{\varepsilon^2 A^4}{c(A^2-1)^2}.$$

Proof. By Lemma 1 it holds that $\sigma^2 = \kappa J_1(\varphi,\varphi) + \beta J_2(\varphi,\varphi)$, where Lemma 2 is used to derive $J_1(\varphi,\varphi)$ and $J_2(\varphi,\varphi)$.

Let $\gamma_1 := \{z_1 \in C : |z_1| = 1\}$ and $\gamma_2 := \{z_2 \in C : |z_2| = 1\}$. Then by Lemma 2 we have that

$$\begin{aligned}
J_1(\varphi,\varphi) &= \lim_{r\downarrow 1} J_1(\varphi,\varphi,r) = \lim_{r\downarrow 1} -\frac{1}{4\pi^2}\oint_{\gamma_1}\oint_{\gamma_2} \frac{\varphi(|1+hz_1|^2)\varphi(|1+hz_2|^2)}{(z_1 - rz_2)^2} dz_2 dz_1 \\
&= \lim_{r\downarrow 1} \frac{1}{2\pi i}\oint_{\gamma_1} \varphi(|1+hz_1|^2)\frac{1}{2\pi i}\oint_{\gamma_2} \frac{|1+hz_2|^2}{|1+hz_2|^2+\varepsilon}\frac{1}{(z_1 - rz_2)^2} dz_2 dz_1 \\
&= \lim_{r\downarrow 1} \frac{1}{2\pi i}\oint_{\gamma_1} \varphi(|1+hz_1|^2)\frac{1}{2\pi i}\oint_{\gamma_2} \frac{(z_2+h)(1+z_2h)}{h(z_2-A)(z_2-B)(z_1-rz_2)^2} dz_2 dz_1.
\end{aligned}$$

The inner contour integral has two poles inside γ_2, a simple pole $z_2 = A$ and a pole $z_2 = \frac{z_1}{r}$ of order 2, since $\frac{z_1}{r}$ lies inside γ_2 for fixed z_1 such that $|z_1| = 1$ and $r > 1$. To calculate the inner contour integral, we compute the residues of the poles expressed as

$$Res(A) = \lim_{z_2\to A}(z_2 - A)\frac{(z_2+h)(1+z_2h)}{h(z_2-A)(z_2-B)(z_1-rz_2)^2} = \frac{A(A+h)(1+Ah)}{h(A^2-1)(z_1-rA)^2},$$

and

$$\begin{aligned}
Res(z_1/r) &= \lim_{z_2\to \frac{z_1}{r}} \frac{d}{dz_2}(z_1 - rz_2)^2\frac{(z_2+h)(1+z_2h)}{h(z_2-A)(z_2-B)(z_1-rz_2)^2} = \lim_{z_2\to \frac{z_1}{r}} \frac{d}{dz_2}\frac{(z_2+h)(1+z_2h)}{h(z_2-A)(z_2-B)} \\
&= \lim_{z_2\to \frac{z_1}{r}} \frac{(z_2-A)(z_2-B)(h^2+1+2hz_2) - (z_2+h)(1+z_2h)(2z_2-A-B)}{h(z_2-A)^2(z_2-B)^2} \\
&= \frac{r(z_1-Ar)(z_1-Br)(h^2r+r+2hz_1) - r(hr+z_1)(r+hz_1)(2z_1-Br-Ar)}{h(z_1-rA)^2(z_1-rB)^2} \\
&= \frac{-r((A+B)h+1+h^2)(z_1^2-r^2)}{h(z_1-rA)^2(z_1-rB)^2} = \frac{\varepsilon r(z_1^2-r^2)}{h(z_1-rA)^2(z_1-rB)^2},
\end{aligned}$$

where we use that $(A+B)h + 1 + h^2 = -\varepsilon$. By Cauchy's Residue Theorem we get that

$$\begin{aligned}
J_1(\varphi,\varphi) &= \lim_{r\downarrow 1} \frac{1}{2\pi i}\oint_{\gamma_1} \varphi(|1+hz_1|^2)\frac{1}{2\pi i}[2\pi i \cdot (Res(A) + Res(z_1/r)]dz_1 \\
&= \lim_{r\downarrow 1} \frac{1}{2\pi i}\oint_{\gamma_1} \frac{(z_1+h)(1+z_1h)}{h(z_1-A)(z_1-B)}\left[\frac{A(A+h)(1+Ah)}{h(A^2-1)(z_1-rA)^2} + \frac{\varepsilon r(z_1^2-r^2)}{h(z_1-rA)^2(z_1-rB)^2}\right] dz_1
\end{aligned}$$

$$= \lim_{r\downarrow 1}[\frac{-\varepsilon A^2}{h^2(A^2-1)}\frac{1}{2\pi i}\oint_{\gamma_1}\frac{(z_1+h)(1+z_1h)}{(z_1-A)(z_1-B)(z_1-rA)^2}dz_1 + \frac{\varepsilon r}{h^2}\frac{1}{2\pi i}\oint_{\gamma_1}\frac{(z_1+h)(1+z_1h)(z_1^2-r^2)}{(z_1-A)(z_1-B)(z_1-rA)^2(z_1-rB)^2}dz_1] = \lim_{r\downarrow 1}\left[\frac{-\varepsilon A^2}{h^2(A^2-1)}F + \frac{\varepsilon r}{h^2}W\right].$$

The function inside the contour integral F has two poles inside $\{z \in \mathbb{C} : |z_1| = 1\}$, a simple pole $z_1 = A$ and a pole of order 2, $z_1 = rA$. It holds that

$$\begin{aligned}
Res(A) &= \lim_{z_1\to A}(z_1-A)\frac{(z_1+h)(1+z_1h)}{(z_1-A)(z_1-B)(z_1-rA)^2} = \lim_{z_1\to A}\frac{(z_1+h)(1+z_1h)}{(z_1-B)(z_1-rA)^2} \\
&= \frac{A(A+h)(1+Ah)}{(A^2-1)(A-rA)^2} = \frac{-\varepsilon A^2}{(A^2-1)(A-rA)^2}, \\
Res(rA) &= \lim_{z_1\to rA}\frac{d}{dz_1}(z_1-rA)^2\frac{(z_1+h)(1+z_1h)}{(z_1-A)(z_1-B)(z_1-rA)^2} = \lim_{z_1\to rA}\frac{d}{dz_1}\frac{(z_1+h)(1+z_1h)}{(z_1-A)(z_1-B)} \\
&= \lim_{z_1\to rA}\frac{(z_1-A)(z_1-B)(h^2+1+2hz_1)-(z_1+h)(1+z_1h)(2z_1-A-B)}{(z_1-A)^2(z_1-B)^2} \\
&= \frac{(rA-A)(rA-B)(h^2+1+2hrA)-(rA+h)(1+rAh)(2rA-A-B)}{(rA-A)^2(rA-B)^2} \\
&= -\frac{A(1+hA)(h+A)(r^2A^2-1)}{(rA-A)^2(rA^2-1)^2} = \frac{\varepsilon A^2(r^2A^2-1)}{(rA-A)^2(rA^2-1)^2},
\end{aligned}$$

where we use that $(A+h)(1+Ah) = -\varepsilon A$. Then, by Cauchy's Residue Theorem the first integral is equal to

$$\begin{aligned}
F &= \frac{1}{2\pi i}[2\pi i(Res(A)+Res(Ar))] = -\frac{\varepsilon A^2}{(A^2-1)(A-rA)^2} + \frac{\varepsilon A^2(r^2A^2-1)}{(rA-A)^2(rA^2-1)^2} \\
&= \frac{-\varepsilon A^2}{(A^2-1)(rA^2-1)^2}.
\end{aligned}$$

Next, we compute the second integral W. The function inside the integral has two poles inside $\{z \in C : |z_1| = 1\}$, a simple pole $z_1 = A$ and a pole $z_1 = rA$ of order 2. It holds that

$$\begin{aligned}
Res(A) &= \lim_{z_1\to A}(z_1-A)\frac{(z_1+h)(1+z_1h)(z_1^2-r^2)}{(z_1-A)(z_1-B)(z_1-rA)^2(z_1-rB)^2} = \frac{(A+h)(1+Ah)(A^2-r^2)}{(A-B)(A-rA)^2(A-rB)^2} \\
&= \frac{-\varepsilon A^2(A^2-r^2)}{(A^2-1)(1-r)^2(A^2-r)^2}, \\
Res(rA) &= \lim_{z_1\to rA}\frac{d}{dz_1}\frac{(z_1+h)(1+z_1h)(z_1^2-r^2)}{(z_1-A)(z_1-B)(z_1-rB)^2} \\
&= \frac{(1+hrA)(r^2A^2-r^2)+h(rA+h)(r^2A^2-r^2)+2rA(rA+h)(1+hrA)}{(rA-A)(rA-B)(rA-rB)^2} \\
&\quad - \frac{(rA+h)(1+hrA)(r^2A^2-r^2)}{(rA-A)^2(rA-B)(rA-rB)^2} - \frac{(rA+h)(1+hrA)(r^2A^2-r^2)}{(rA-A)(rA-B)^2(rA-rB)^2} \\
&\quad - \frac{2(rA+h)(1+hrA)(r^2A^2-r^2)}{(rA-A)(rA-B)(rA-rB)^3},
\end{aligned}$$

$$= \frac{-(A+h)(1+hrA)(r^2A^2-r^2)}{(rA-A)^2(rA-B)(rA-rB)^2} - \frac{(A+h)(rA+h)(r^2A^2-r^2)}{A(rA-A)(rA-B)^2(rA-rB)^2}$$
$$= \frac{-(A+h)(r^2A^2-r^2)(1+Ah)(r^2A^2-1)}{A(rA-A)^2(rA-B)^2(rA-rB)^2}$$
$$= \frac{\varepsilon A^2(r^2A^2-1)}{(r-1)^2(rA^2-1)^2(A^2-1)}.$$

Then by Cauchy's Residue Theorem the second integral is equal to

$$W = \frac{1}{2\pi i}[2\pi i(Res(A)+Res(Ar))] = \frac{-\varepsilon A^2(A^2-r^2)}{(A^2-1)(1-r)^2(A^2-r)^2} + \frac{\varepsilon A^2(r^2A^2-1)}{(r-1)^2(rA^2-1)^2(A^2-1)}$$
$$= \frac{\varepsilon A^4(A^2+1)(r^2-1)}{(A^2-r)^2(rA^2-1)^2(A^2-1)}.$$

Hence, combining the two integrals we find $J_1(\varphi,\varphi) = \lim_{r\downarrow 1} J_1(\varphi,\varphi,r) = \frac{\varepsilon^2}{h^2(A-B)^4}$. Finally, we evaluate $J_2(\varphi,\varphi)$. By Lemma 2, with $\gamma_1 := \{z_1 \in \mathbb{C} : |z_1| = 1\}$ and $\gamma_2 := \{z_2 \in \mathbb{C} : |z_2| = 1\}$, we get

$$J_2(\varphi,\varphi) = \frac{1}{2\pi i}\oint_{\gamma_1} \frac{f(|1+hz_1|^2)}{z_1^2}dz_1 \cdot \frac{1}{2\pi i}\oint_{\gamma_2} \frac{g(|1+hz_2|^2)}{z_2^2}dz_2$$
$$= \frac{1}{2\pi i}\oint_{\gamma_1} \frac{(z_1+h)(1+z_1h)}{((z_1+h)(1+z_1h)+z_1\varepsilon)z_1^2}dz_1 \cdot \frac{1}{2\pi i}\oint_{\gamma_2} \frac{(z_2+h)(1+z_2h)}{((z_2+h)(1+z_2h)+z_2\varepsilon)z_2^2}dz_2$$
$$= \left(\frac{1}{2\pi i}\oint_{\gamma_1} \frac{(z_1+h)(1+z_1h)}{h(z_1-A)(z_1-B)z_1^2}dz_1\right)^2 = \widetilde{W}^2.$$

The function inside the contour integral $\widetilde{W}$ has two poles inside γ_1, one simple pole $z_1 = A$ and one pole $z_1 = 0$ of order two. We get

$$Res(A) = \lim_{z_1\to A}(z_1-A)\frac{(z_1+h)(1+z_1h)}{h(z_1-A)(z_1-B)z_1^2} = \lim_{z_1\to A}\frac{(z_1+h)(1+z_1h)}{h(z_1-B)z_1^2}$$
$$= \frac{(A+h)(1+Ah)}{h(A-B)A^2} = \frac{(A+h)(1+Ah)}{h(A^2-1)A} = \frac{-\varepsilon}{h(A^2-1)},$$
$$Res(0) = \lim_{z_1\to 0}\frac{d}{dz_1}(z_1-0)^2\frac{(z_1+h)(1+z_1h)}{h(z_1-A)(z_1-B)z_1^2} = \lim_{z_1\to 0}\frac{d}{dz_1}\frac{(z_1+h)(1+z_1h)}{h(z_1-A)(z_1-B)}$$
$$= \lim_{z_1\to 0}\frac{(z_1-A)(z_1-B)(h^2+1+2hz_1)-(z_1+h)(1+z_1h)(2z_1-A-B)}{h(z_1-A)^2(z_1-B)^2}$$
$$= \frac{h^2+1+h(A+B)}{h} = -\frac{\varepsilon}{h}.$$

Hence, Cauchy's Residue Theorem yields $J_2(\varphi,\varphi) = \frac{\varepsilon^2A^4}{h^2(A^2-1)^2}$. Combining the limiting parameters $J_1(\varphi,\varphi)$ and $J_2(\varphi,\varphi)$ and using $h = \sqrt{c}$, we get the statement of the lemma. ■

Lemma 6. *Under the statement of Theorem 2, it holds that* $U = -\frac{A}{h}$.

Proof. To compute the centering factor $F_c(\varphi)$ Lemma 3 is used. Let $\gamma := \{z \in \mathbb{C} : |z| = 1\}$. Then,

$$
\begin{aligned}
U &= -\frac{1}{4\pi i}\oint_\gamma \left(\frac{|1+hz|^2}{|1+hz|^2+\varepsilon}\right)\frac{(1-z^2)^2}{z^2(1+hz)(z+h)}dz \\
&= -\frac{1}{4\pi i}\oint_\gamma \frac{(z+h)(1+zh)(1-z^2)^2}{((z+h)(1+zh)+z\varepsilon)z^2(1+hz)(z+h)}dz \\
&= -\frac{1}{4\pi i}\oint_\gamma \frac{(1-z^2)^2}{((z+h)(1+zh)+z\varepsilon)z^2}dz = -\frac{1}{4\pi i}\oint_\gamma \frac{(1-z^2)^2}{h(z-A)(z-B)z^2}dz.
\end{aligned}
$$

The function inside the contour integral has two poles inside $\gamma := \{z \in C : |z| = 1\}$, a simple pole $z = A$ and a pole $z = 0$ of order 2. To calculate the integral we need to calculate the residues of these poles, expressed as

$$
\begin{aligned}
Res(0) &= \lim_{z\to 0}\frac{d}{dz}(z^2-0)\frac{(1-z^2)^2}{h(z-A)(z-B)z^2} = \lim_{z\to 0}\frac{d}{dz}\frac{(1-z^2)^2}{h(z-A)(z-B)} \\
&= \lim_{z\to 0}\frac{(z^2-1)(4ABz-3Az^2-3Bz^2+2z^3-A-B+z)}{h(A-z)^2(B-z)^2} = \frac{A+B}{hA^2B^2} = \frac{A^2+1}{hA}, \\
Res(A) &= \lim_{z\to A}(z-A)\frac{(1-z^2)^2}{h(z-A)(z-B)z^2} = \frac{(A^2-1)^2}{h(A^2-1)A} = \frac{A^2-1}{hA}.
\end{aligned}
$$

Then, by Cauchy's Residue Theorem and the equality $h = \sqrt{c}$, we get the statement of the lemma. ∎