CLT for Nonlinear Shrinkage Estimators of Large Covariance Matrices

by

F.J.G. Veldman

to obtain the degree of Bachelor of Science
at the Delft University of Technology,
to be defended publicly on July 5, 2022

An electronic version of this thesis is available at http://repository.tudelft.nl/.

ŤUDelft

Abstract

This thesis is concerned with finding the asymptotic distributions of linear spectral statistics of the nonlinear shrinkage estimator for large covariance matrices derived by Ledoit and Wolf (2012). It provides some new inferential procedures for large-dimensional data and shed some light on the power of the new statistical tests on the structure of the large covariance matrices. After a short review of the relevant theory, two linear spectral statistics are proposed which are deduced from the nonlinear shrinkage estimator where for one of these linear spectral statistics its limiting distribution is derived. This results in a ready to use sphericity test statistic in the large-dimensional framework and is one of the main results of this thesis. The sphericity test corresponding to this new test statistic is called the nonlinear shrinkage test (NLS-ϵ). The theoretical results are illustrated by means of a simulation study where the new nonlinear shrinkage test is compared with already existing tests, in particular the commonly used corrected likelihood ratio test and the corrected John's test. It is demonstrated that the new nonlinear shrinkage test is most powerful under a non homogeneous variance alternative where it outperforms well known sphericity tests. Moreover, it is observed that the new nonlinear shrinkage test encounters some problems when different alternatives are combined.

Preface

This research has been conducted under supervision of Dr. N. Parolya on behalf of the department of Statistics of the faculty EEMCS at the University of Technology Delft.

I would like to express my special thanks to my supervisor Dr. N. Parolya for this guidance during this project. He was always there to assist and was always open for questions. I also would like to thank Prof. dr. F.H.J. Redig for taking a seat in my graduation committee.

F.J.G. Veldman
Delft, July 2022

Contents

1

Introduction

Due to the rapid advancement of modern technology there is now more data available than ever before. Large amounts of data appear in many fields, for example in finance, where online data from markets around the word are processed on a scale that is not imaginable. Or in genetics, where in recent years it has become possible to register the expression of several thousands of genes from a single tissue, and much more. To handle large amounts of data one always relied on the large sample theory until very recently. The large sample theory means that when given a sample $(\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x_n})$ of random observations of dimensions p, the sample size n could tend to infinity but p needed to say fixed. This led to many problems when analysing large-dimensional data where the dimension p of the data is not as small as several tens. It has been pointed out by numerous authors 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 to let the dimension go to infinity as well. This led to a new area in asymptotic statistics where the dimension p is no longer fixed, as in the large sample theory, but tend to infinity together with the sample size n. This framework is called large-dimensional asymptotic.

Many tools that used to work in large sample theory did not work properly anymore in the large-dimensional asymptotic framework and needed to be altered. Moreover, new ones have been developed and are still being developed. So also many multivariate statistical tests. Multivariate statistical tests are key in the analysis of data with multiple dimensions and because of this widely used in many fields. Therefore, this thesis is concerned with constructing a new multivariate statistical hypothesis test in the large-dimensional asymptotic framework. Many random processes can be modelled using a stochastic model of the form $Y = \Sigma_n^{1/2} X_n$. In this stochastic model, Σ_n is the population or true covariance matrix and explains the variability and the relation between different variables in theoretical sense. The matrix X_n represents the randomness of the model. For numerous application one would like to know if the variables in the observed data set can be assumed to be independent of each other with a certain accuracy. This corresponds to testing whether the true or population covariance matrix is equal or not the the identity matrix when only the data matrix Y is observed. This type of multivariate statistical hypothesis test is called a sphericity test and this is the one that will be constructed.

This new sphericity test in the large-dimensional framework will be based on the nonlinear shrinkage estimator derived by Ledoit and Wolf (2012). It has been pointed out that the nonlinear shrinkage estimator is in many cases a better estimator for the population covariance matrix than for example the linear shrinkage estimator which is derived by Ledoit and Wolf (2004). This makes it interesting to develop a test based on the nonlinear shrinkage estimator and investigate whether it also performs better than the test that is based on the linear shrinkage estimator which is derived by Versteegh (2020). Furthermore, it will also be interesting to explore how this new test behaves compared to well known and established sphericity tests in a simulation study.

Outline

A central object in higher dimensional statistics is the sample covariance matrix S_n. Almost all statistical methods in multivariate analysis rely on this sample covariance matrix: multivariate regressions, one-sample or two-sample hypothesis testing, and much more. In the case of the one-sample or two-sample hypothesis testing, many statistics are functions of the eigenvalues of the sample covariance matrix. Therefore, this thesis starts with literature review of the sample covariance matrix and its important properties. In addition, there will be looked at other estimators such as the linear shrinkage estimator and the nonlinear shrinkage estimator on which the new sphericity test will be based. After the review of the estimators some of the well established sphericity tests will discussed. These test will be used in the simulation study. Then when all relevant literature is reviewed and discussed, the construction of the new sphericity test can be started. As said earlier, many statistics are functions of eigenvalues of the sample covariance matrix. This will also be the case for this new test statistic. The function that will be used in the test statistic is derived from the nonlinear shrinkage function proposed by Ledoit and Wolf (2012). To construct a test out of this new test statistic the limiting distribution will be calculated. This will be done using the central limit theorem for linear spectral statistics and the found limiting distribution for this new test statistics will be one of the main results of this thesis.

After this theoretical part there will be a simulation study to asses the performance of the new test and to compare it with the already existing sphericity tests. The comparison will be made using the following indicators: the empirical size, the empirical power and receiver operating characteristic curves or also know as ROC curves. In the conclusion the main results of this thesis will be highlighted and in the discussion the potential flaws will be discussed as well as some recommendations will be given for further research. Now it is time to start with reviewing some of the important topics in random matrix theory.

2

Estimators for Large Covariance Matrices

In this chapter three estimators for large covariance matrices will be considered; the sample covariance matrix, the linear shrinkage estimator and the nonlinear shrinkage estimator. In particular, there will we looked at some of the properties as well as some of the problems occurring when using the sample covariance matrix in large-dimensional statistics. Furthermore, the linear shrinkage and nonlinear shrinkage estimators will be constructed. Before diving into this chapter the large-dimensional asymptotic framework need to be specified: let n denote the sample size and $p = p(n)$ the number of variables, with $\frac{p}{n} \to c \in (0,1)$ as $n \to \infty$.

2.1. Sample Covariance Matrix

A central object in higher dimensional statistics is the sample covariance matrix. Almost all statistical methods in multivariate analysis rely on this sample covariance matrix: multivariate regressions, one-sample or two-sample hypothesis testing, and much more. The sample covariance matrix S_n is an estimator for the true or population covariance matrix Σ_n of a given data set. The true covariance matrix is a matrix containing the variance of every variable and explains the relation between different variables in theoretical sense. The sample covariance matrix that will be used throughout this thesis is defined as follows:

Definition 2.1.1. Given a p x n data matrix $Y = (\mathbf{y_1}, \ldots, \mathbf{y_n})$, with each $\mathbf{y_i}$ a data vector of dimension p, the *unbiased sample covariance matrix S_n* is the matrix:

$$S_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{y_i}\mathbf{y_i}^T = \frac{1}{n} Y Y^T$$

In this definition there is assumed that the mean is known and equal to zero. This will also be one of the assumptions for the data that will be used. To study some of the key properties of the sample covariance matrix a few more definitions are needed, starting with the empirical spectral distribution (ESD). The empirical spectral distribution is defined as:

Definition 2.1.2. Let S be a p x p matrix with eigenvalues $\{\lambda_1, \ldots, \lambda_p\}$. *The Empirical Spectral Distribution F_S* of the matrix S is:

$$F^S = \frac{1}{p} \sum_{i=1}^{p} \delta_{\lambda_i}$$

where δ_{λ_i} is denoted as the Dirac mass placed at the eigenvalue λ_i.

The next definition that is needed is the one of the limiting spectral distribution (LSD):

Definition 2.1.3. Let $\{S_n\}_{n\geq 1}$ be a sequence of p x p matrices. If the sequence of corresponding empirical spectral distributions $\{F^{S_n}\}_{n\geq 1}$ vaguely converges to a measure F, then F is called the *limiting spectral distribution* (LSD) of the sequence of matrices $\{S_n\}$.

The last definition that is needed to study the properties of the sample covariance matrix is that of the Marchenko-Pastur distribution or law. This distribution is defined as follows:

Definition 2.1.4. The *standard Marchenko-Pastur distribution F_c* (M-P law) with index c has the density function

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

with an additional point mass of value $(1-\frac{1}{c})$ at the origin if $c > 1$, where $a = (1-\sqrt{c})^2$ and $b = (1+\sqrt{c})^2$. Here, the constant c is the dimension to sample size ratio index. Moreover, there is assumed that the variance σ^2 is equal to 1.

Three standard M-P density functions for $c \in \{\frac{1}{8}, \frac{1}{4}, \frac{1}{2}\}$ are displayed in Figure 2.1
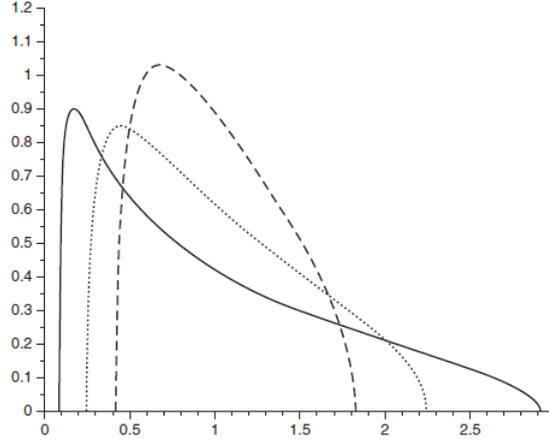
Figure 2.1: Density plots of the Marcenko-Pastur distributions with indexes $c = \frac{1}{8}$ (dashed line), $c = \frac{1}{4}$ (dotted line) and $c = \frac{1}{2}$ (solid line). Note. Figure taken from the book *Large Sample Covariance Matrices and High-Dimensional Data Analysis* (p.11), By Yao, J., Zheng, S., Bai, Z. D., (2015), Cambridge University Press.

Using the above definitions one of the main properties of the sample covariance matrix in large-dimensional statistics can be presented. This property states that the limiting spectral distribution of the eigenvalues of the sample covariance matrix is equal to the standard Marchenko-Pastur distribution. Marchenko and Pastur (1967) first discovered this and has been extended in several directions such as Theorem 2.9 of the book Yao et al. (2015),*Large Sample Covariance Matrices and High-Dimensional Data Analysis*:

Theorem 2.1.1. *Suppose that the entries $\{x_{i,j}\}$ of the matrix X are i.i.d. complex random variables with mean zero and variance one, and $\frac{p}{n} \to c \in (0, \infty)$. Then almost surely, F^{S_n} converges to the standard M-P law F_c*

It is understood that if $n \to \infty$, the sample covariance matrix converges to the population covariance matrix which is equal to the identity matrix when all entries of the data matrix X are i.i.d with mean zero en variance 1. Therefore, since the eigenvalues are continuous functions of matrix entries, the sample eigenvalues of S_n should converge to 1 almost surly. However, in the large-dimensional case when $p \to \infty$ as well, Theorem 2.1.1 states that the sample eigenvalues obey the M-P law. Thus that the eigenvalues of the sample covariance matrix are spread out over the interval $[a, b] = [(1 - \sqrt{c})^2, (1 + \sqrt{c})^2$ and will therefore not converge to 1. So the eigenvalues of S_n will not be a consistent estimator for the true eigenvalues of the population covariance matrix. In conclusion, using the sample covariance matrix directly in the large-dimensional case will lead to some serious errors. That is why there is need for different estimators in the large-dimensional case.

2.2. Linear Shrinkage Estimator

In this section the Linear Shrinkage Estimator will be presented. This is an estimator which estimates the true covariance matrix in the large-dimensional case. This estimator is based on the work of Ledoit and Wolf (2004) and extended by Bodnar et al. (2014). The general linear shrinkage estimator Σ_{GLSE} is of the form:

$$\Sigma_{GLSE} = \alpha_n S_n + \beta_n \Sigma_0$$

Where Σ_0 is a symmetric positive definite matrix, bounded in trace and can be seen as a prior belief of the true covariance matrix. The parameters α_n and β_n are called the shrinkage intensities because they basically

shrink the matrices which they are multiplied with. Thus the Σ_{GLSE} is essentially a linear combination between the sample covariance matrix and the prior belief of the true covariance matrix. Using the loss function $L_f^2 = ||\Sigma_{GLSE} - \Sigma_n||_F^2$, where $||.||_F^2$ is the squared Frobenius norm, the asymptotic optimal shrinkage estimators $\hat{\alpha}^*$ and $\hat{\beta}^*$ can be found. This loss function L_f^2 measures the "distance" between the estimator Σ_{GLSE} and the true covariance matrix Σ_n. For the estimator to be working properly, this "distance", thus the loss function, should be as small as possible. Versteegh (2020) minimised this loss function L_f^2 and found that the asymptotic optimal shrinkage estimators are equal to

$$\hat{\alpha}^* = 1 - \frac{\frac{1}{n} tr(S_n)^2 ||\Sigma_0||_F^2}{||S_n||_F^2 ||\Sigma_0||_F^2 - (tr(S_n\Sigma_n))^2}$$

$$\hat{\beta}^* = \frac{tr(S_n\Sigma_n)}{||\Sigma_0||_F^2}(1 - \hat{\alpha}^*)$$

These parameters will consistently estimate the optimal asymptotic value of the general linear shrinkage estimator in the limit $\frac{p}{n} \to c$. Note that these parameters depend on a prior belief of Σ_0. However, using these values in the general linear shrinkage estimator, the true covariance matrix Σ_n can be estimated in the large-dimensional case. In addition, Versteegh (2020) uses these values to construct a new sphericity test. This will be covered in a later chapter.

2.3. Construction of the Nonlinear Shrinkage Estimator

This chapter is concerned with the construction of a new type of estimator for large-dimensional covariance matrices derived by Ledoit and Wolf (2012); the nonlinear shrinkage estimator. This estimator is in some way an extension of the linear shrinkage estimator. Where the linear shrinkage estimator shrinks the sample covariance matrix in a linear way, the nonlinear shrinkage estimator does it in a nonlinear way. To build the nonlinear shrinkage estimator some assumptions of the underlying stochastic model are needed, where the stochastic model is given by $Y = \Sigma^{\frac{1}{2}} X_n$.

- **(A1)** The population covariance matrix Σ_n is a non-random p-dimensional positive definite matrix.

- **(A2)** Let X_n be an n × p matrix of real independent and identically distributed (i.i.d.) random variables with mean zero and unit variance. One only observes Y where $Y = \Sigma^{\frac{1}{2}} X_n$.

- **(A3)** Let $\tau_{n,1}, \ldots, \tau_{n,p}$ denote the eigenvalues of the population or true covariance matrix Σ_n. The empirical spectral distribution of the population eigenvalues $H_n(\tau)$ converges almost surly (a.s.) to a non-random limit $H(\tau)$ at every point of continuity.

- **(A4)** $Supp(H)$, the support of H, is the union of a finite number of closed intervals, bounded away from zero and infinity. Furthermore, there exists a compact interval in $(0, +\infty)$ that contains $Supp(H_n)$ for all n large enough.

Before the nonlinear shrinkage estimator can be constructed one last mathematical tool is needed; the Stieltjes transform method. As said earlier, the eigenvalues of a matrix are continuous function of entries of the matrix. Now when the dimension of a matrix is larger than four, these functions do not have a closed form any more. To still study their properties the Stieltje transform method can be used and that is why the Stieltjes transform method is widely used in the literature of large-dimensional statistics. Why the Stieltjes transform method is of importance in this particular case will be made clear in a minute. First the definition of the Stieltjes transform is presented.

Definition 2.3.1. Let μ be a finite measure on the real line. *The Stieltjes transform* of the measure μ with $z \in \mathbb{C}^+$ where $\mathbb{C}^+ = \{z \in \mathbb{C} : \text{Im}(z) > 0\}$ is defined as:

$$m_\mu(z) = \int \frac{1}{x - z} \mu(dx)$$

One of the main properties of the Stieltjes transform is that it characterises the vague convergence of finite measures. This is a key tool in studying empirical spectral distributions (ESDs) of random matrices. This is summarized in Theorem 2.7 from Yao et al. (2015) which is given by:

Theorem 2.3.1. *A sequence $\{\mu_n\}$ of probability measures on $\mathbb{R}$ converges vaguely to some positive measure μ if and only if their Stieltjes transforms $\{m_{\mu_n}\}$ converges to m_μ on $\mathbb{C}^+$.*

It can be shown that the Stieljes transform of the empirical spectral distribution of the sample covariance matrix S_n is equal to:

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

From Theorem 2.1.1 it is known that the ESD $F^{S_n}(\lambda)$ of the sample covariance matrix converges almost surely to some nonrandom limit $F(\lambda)$. Now by Theorem 2.3.1 also the Stieltjes transform of the ESD of the covariance matrix should convergence. This one of the main results of Marchenko and Pastur (1967). The most convenient expression for this limit is found in Silverstein and Choi (1995), which is given by:

$$m_F(z) = \int_{-\infty}^{+\infty} \frac{1}{\tau[1 - c - czm_F(z)] - z} dH(\tau) \tag{2.1}$$

Moreover,Silverstein and Choi (1995) showed that:

$$\lim_{z \in \mathbb{C}^+ \to \lambda} m_F(z) = \breve{m}_F(\lambda), \quad \forall \lambda \in \mathbb{R} \backslash \{0\} \tag{2.2}$$

exist. The above two results is the main reason why the Stieltjes transform is of such importance because equation (2.2) appears in the nonlinear shrinkage estimator and equation (2.1) is needed to investigate if it is possible to construct statistical test from the nonlinear shrinkage estimator. Now it is possible to proceed with the construction of the nonlinear shrinkage estimator. In the absence of specific information about the population or true covariance matrix, it seems reasonable to only consider estimators which are invariant under rotations of the observed data. In Perlman (2007) it is mentioned that every rotation-invariant estimator for Σ_n is of the form:

$$U_n D_n U_n^T$$

where $D_n = Diag(d_1, \dots, d_p)$ is a diagonal matrix and where U_n is the matrix whose ith column is the sample eigenvector $\mathbf{u_i}$. This is the class of estimators that will be considered. The objective is to find an estimator that is closest to the population matrix. To quantify the word 'closest' the Frobenius norm is used, which is defined as: $||A||_F = \sqrt{tr(AA^T)}$. Now to find the estimator that is closed to population matrix Σ_n the following minimization problem needs to be solved: $\min_{D_n} ||U_n D_n U_n^T - \Sigma_n||_F$. Elementary matrix algebra shows that the optimal solution is equal to:

$$\widetilde{D}_n = Diag(\tilde{d}_1, \dots, \tilde{d}_p), \quad \text{where} \quad \forall i \in \{1, \dots p\} \quad \tilde{d}_i = \mathbf{u_i}^T \Sigma_n \mathbf{u_i} \tag{2.3}$$

the interpretation of $\tilde{d}_i$ is that it catches how the ith sample eigenvector $\mathbf{u_i}$ relates to the population covariance matrix Σ_n. As a result, the finite sample optimal estimator is given by $\Sigma_n^* = U_n \widetilde{D}_n U_n^T$, where $\tilde{D}_n$ is given by equation (2.3). However, it is not possible to calculate the explicit because it depends on the non-observable population covariance matrix Σ_n. Therefore, it is important to get as close to Σ^* as possible by characterising the asymptotic behaviour of $\tilde{d}_i$ for every $i \in \{1, \dots, p\}$. To do this Ledoit and Peche (2011) introduce a new object which is a non-decreasing function defined by:

$$\Delta_p(x) = \frac{1}{p} \sum_{i=1}^{p} \tilde{d}_i \mathbb{1}_{[\lambda_i, +\infty)}(x) \quad \forall x \in \mathbb{R} \tag{2.4}$$

This function is used in their Theorem 4 which is one of the main results they present. It says that Equation (2.4) converges almost surely to a non-random quantity. This theorem is defined as follows:

Theorem 2.3.2. *Assume that conditions (A1)-(A4) hold and let Δ_p be defined as in Equation (2.3). There exist a nonrandom function Δ defined over $\mathbb{R}$ such that $\Delta_p(x)$ converges a.s to $\Delta(x)$ for all $x \in \mathbb{R} \backslash \{0\}$. If in addition $c < 1$, then Δ can be expressed as: $\forall x \in \mathbb{R}, \quad \Delta(x) = \int_{-\infty}^{x} \delta(\lambda) dF(\lambda)$, where $\forall \lambda > 0$*

$$\delta(\lambda) = \frac{\lambda}{|1 - c - c\lambda \breve{m}_F(\lambda)|^2} \tag{2.5}$$

where $\breve{m}_F(\lambda)$ is given by equation (2.2).

Using this theorem Ledoit and Peche ([2011]) show using that the asymptotic quantity corresponding to $\tilde{d}_i$ is $\delta(\lambda)$. From which they deduce the nonlinear shrinkage estimator:

$$\widehat{\Sigma}_n = U_n \widehat{D}_n U_n^T \quad \text{where} \quad \widehat{D}_n = Diag(\delta(\lambda_1), \ldots, \delta(\lambda_p)) \tag{2.6}$$

where λ_i for all $i \in \{1, \ldots, p\}$ are the eigenvalues of the sample covariance matrix and where U_n is the matrix whose ith column is the sample eigenvector $\mathbf{u_i}$ corresponding to the eigenvalue λ_i. Note that $\widehat{\Sigma}_n$ is a nonlinear shrinkage estimator because the eigenvalues of $\widehat{D}_n$ are obtained by applying the nonlinear shrinking function $\delta(\lambda)$ form Equation (2.5) to every sample eigenvalue λ_i. The obtained nonlinear shrinkage estimator is what is called a oracle estimator. That means in this case that it depends on the 'limiting' distribution of the sample eigenvalues and not the observed one. The nonlinear shrinkage function $\delta(\lambda)$ from Equation (2.5) will later be used to construct a new statistical test based on the nonlinear shrinkage estimator.

3

Sphericity Tests in Large-dimensional Asymptotics

One of the main goals of this thesis is to construct a sphericity test from the nonlinear shrinkage estimator. Before constructing this sphericity test three other sphericity tests are introduced;, The corrected likelihood ratio test (CLRT), the corrected John's test (CJ) and the linear shrinkage test (LS). A sphericity test is a statistical hypothesis test which test the following hypothesis:

$$H_0 : \Sigma_n = \sigma^2 I$$
$$H_1 : \Sigma_n \neq \sigma^2 I$$

In words, a sphericity test, tests the null hypothesis whether the true covariance matrix is a multiple of the identity matrix. This corresponds with testing if the variables of the data are independent and have variance σ^2. The test that will be considered are independent of the variance and therefore without loss of generality it is possible to assume that $\sigma^2 = 1$. When performing a hypothesis test a test statistic is needed. In multivariate statistics many statistics are function of eigenvalues $\{\lambda_i\}$ of the sample covariance matrix S_n. Such a statistic is called a linear spectral statistic (LSS) and is defined by:

$$T_n = \sum_{i=1}^{p} \varphi(\lambda_i) = \int \varphi(\lambda_i) dF^{S_n}(x) =: F^{S_n}(\varphi)$$

for a specific function φ. In the end of this chapter the three sphericity tests will be compered based on the findings of Versteegh (2020) who did an extensive comparison between the three.

3.1. Corrected Likelihood Ratio Test (CLRT)

The likelihood ratio test is denoted in the literature as one of the three classical test in finite dimensional statistics. Moreover, by the Neyman-Pearson lemma from the paper Neyman and Pearson (1933) it is also the most powerful one. The likelihood ratio test statistic from the the book *An Introduction in Multivariate Statistics*, Anderson (1984) is given by:

$$L_n = \left[\frac{(\lambda_1 \cdots \lambda_p)^{\frac{1}{p}}}{\frac{1}{p}(\lambda_1 + \cdots + \lambda_p)} \right]^{\frac{pn}{2}} \tag{3.1}$$

Where $\lambda_1, \cdots, \lambda_p$ are the eigenvalues of the sample covariance matrix S_n. This is the test statistic which is used in the finite dimensional case. For the test also to work in the the large-dimensional case the test statistic needed to be altered and the limiting distribution needed to be derived again. This is done by Wang and Yao (2013) and summarized in the following theorem:

Theorem 3.1.1. *Let $\Lambda_n = -\frac{2}{n} Log(L_n)$ be the test statistic, with L_n equal to Equation (3.1). Assume that the realisations $\{x_{i,j}\}$ are independent and identically distributed, satisfying $\mathbb{E}[x_{i,j}] = 0, \mathbb{E}[|x_{i,j}|^2] = 1, \mathbb{E}[|x_{i,j}|^4] < \infty$. Then under $H_0 : \Sigma_n = I$*

$$\Lambda_n + (p - n)Log(1 - \frac{p}{n}) - p \to N\left(-\frac{\kappa - 1}{2} Log(1 - c) + \frac{1}{2}\beta c, -\kappa Log(1 - c) - \kappa c \right)$$

in distribution and where $\kappa = 2$ *if the data is real and* $\kappa = 1$ *if the data is complex. Furthermore,* $\beta = \mathbb{E}[|x_{i,j}|^4] - 1 - \kappa$

This is the corrected likelihood ratio test statistics and its limiting distribution in the large-dimensional case under the null hypothesis $H_0 : \Sigma_n = I$. The test associated with this test statistics is the corrected likelihood ratio test (CLRT). One should notice that this test is indeed independent of σ but depends on the $\mathrm{Log}(1 - c)$, in both the expectations as well as in the variance. This means that when c is close to 1, thus p close to n, the logarithm will go to minus infinity and the variance will blow out proportion. Therefore it is expected that the CLRT will break down when p is close to n, and that the test will not work when p is greater than n.

3.2. Corrected John's Test (CJ)

The second test that will be considered is the corrected John's Test (CJ). The original test was proposed in the paper of John (1971) and has been modified by Wang and Yao (2013) to work in the large-dimensional case. This test is a rotation invariant test and considered as one of the most powerful ones. That means that if the matrix X is rotated by some orthogonal matrix Q, the test still works. The test is then computed from a new sample covariance matrix given by:

$$S_n = \frac{1}{n} Y Y^T = \frac{1}{n} \Sigma^{\frac{1}{2}} QX (\Sigma^{\frac{1}{2}} QX)^T$$

The data in this thesis will not be rotated, this means that Q can be taken equal to the identity matrix. The test statistic that originally was proposed in John (1971) is of the form:

$$T = \frac{p^2 n}{2} tr\{S_n (tr(S_n))^{-1} - I p^{-1}\}^2 \tag{3.2}$$

When p is fixed and $n \to \infty$, under the null hypothesis, is holds that $T \to \chi_f^2$ in distribution, a Chi-squared distribution with $f = \frac{1}{2} p(p + 1) - 1$ degrees of freedom. This is referred as the John's test. For the John's test to work in the large-dimensional case Wang and Yao (2013) proposed a new test statistic $U = \frac{2}{pn} T$. Using this new test statistic they proved the following theorem:

Theorem 3.2.1. *Assume that the entries* $\{x_{i,j}\}$ *are i.i.d. , satisfying* $\mathbb{E}[x_{i,j}] = 0, \mathbb{E}[|x_{i,j}|^2] = 1, \mathbb{E}[|x_{i,j}|^4] < \infty$, *and let* $U = \frac{2}{pn} T$ *be the test statistic with T equal to Equation (3.2) . Then under* H_0 *and when* $p \to \infty, n \to \infty, \frac{p}{n} \to c \in (0, \infty)$,

$$nU - p \to N(\kappa - 1 + \beta, 2\kappa)$$

in distribution and where $\kappa = 2$ *if the data is real and* $\kappa = 1$ *if the data is complex and where* $\beta = \mathbb{E}[|x_{i,j}|^4] - 1\kappa$

It is important to note, that the limiting distribution is independent from c unlike the CLRT.

3.3. Linear Shrinkage Test (LS)

The third test that will be considered is the test resulting from the central limit theorem of the linear shrinkage estimator, which will be stated in a later section. This test is proposed by Versteegh (2020) and it depends on the limiting distribution of the optimal shrinkage intensity $\hat{\alpha}^*$. One of the main results of Versteegh (2020) is his Theorem 3.3.1

Theorem 3.3.1. *Let X be a p x n data matrix consisting of i.i.d random variables with mean zero and variance 1, and let* $Y = \Sigma_n^{\frac{1}{2}} X$. *Let* $S_n = \frac{1}{n} Y Y^T$ *be its sample covariance matrix with eigenvalues* $\lambda_1, \cdots, \lambda_p$. *Let* $(p, n) \to \infty$ *and* $\frac{p}{n} \to c > 0$. *Then under* H_0

$$p\hat{\alpha}^* \to N(\kappa - 1 + \beta, 2\kappa)$$

in distribution. Where $\hat{\alpha}^*$ *is the optimal consistent estimator of the form:*

$$\hat{\alpha}^* = 1 - \frac{\frac{1}{n} tr(S_n)^2 ||\Sigma_0||_F^2}{||S_n||_F^2 ||\Sigma_0||_F^2 - (tr(S_n \Sigma_n))^2}$$

This means that $p\hat{\alpha}^*$ is a ready to use statistic for a sphericity test in the large-dimensional case. It should be noted that the limiting distributions of the LS test and the CJ test are the same. This means that in the limit they should perform equally but it could be, since the test statistics are not the same, that the one of the test outperforms the other in a earlier stage.

3.4. Comparison

It is not only interesting to know what kind of sphericity test there exist in large-dimensional statistics but also how well they perform compared to each other. Therefore, the performance of the three sphericity test mentioned earlier in this chapter will be compared; the corrected likelihood ratio test (CLRT), the corrected John's test (CJ) and the test based on the linear shrinkage estimator (LS). One way to quantify the performance of a test is to use its power. The power of a test is defined as:

$$\text{Power } = P(\text{reject } H_0 | H_1 \text{ is true})$$

that is, the probability that the null hypothesis is rejected correctly. So the more powerful a test is, the quicker it rejects a false hypothesis. An other way to quantify the performance of a test is to look at the size of a test. The size of a test is defined as:

$$\text{Size } = P(\text{reject } H_0 | H_0 \text{ is true})$$

that is, the probability that the null hypothesis is rejected falsely . If a statistic is close to its limiting distribution, then the size of a test should be close to a rejection level α under H_0. Versteegh (2020) did an extensive comparison between these three test using three different methods; empirical size comparison, empirical power comparison and a comparison using ROC curves. ROC curves or receiver operating characteristic curves are curves that compare how the true positive rate changes when the false positive rate is varied. It can also been seen as a plot of the power as function of the size of a test.

Since Versteegh (2020) already did an extensive comparison, only his observations and conclusions will be presented in this section. He observed that CLRT test performs worse or sometimes equal to the CJ and LS test and that the CJ test slightly outperforms the LS test, especially when the alternative hypothesis is close to the null hypothesis. Moreover, he observed that the CJ test has a small head start in power compared to the LS test and that the size of the LS test is small. He argues these two observations have one explanation that is that the LS test need the large-dimensional aspect more than the CLRT and the CJ test. He concludes that the CJ test performs best for the reason that the LS never really catches up because of the difference in power in starts with. Moreover, he concludes that the LS test still mostly outperforms the CLRT test and that CLRT test performs decent in a low dimension, but disappoints when p increases.

In a later chapter the new constructed sphericity test from the nonlinear shrinkage estimator will be evaluated in a simulation study in the same manner as Versteegh (2020) did. In addition, the new constructed test will be compared with the the already established sphericity tests, mentioned in this chapter.

4

CLT for Nonlinear Shrinkage Estimator

As have been pointed out in the previous chapter, one of the main goals of this thesis is to construct a new sphericity test from the nonlinear shrinkage estimator constructed in section 2.3. Recall that in multivariate statistics many statistics are function of eigenvalues $\{\lambda_i\}$ of the sample covariance matrix S_n. Such a statistic is called a linear spectral statistic (LSS) and is defined by

$$T_n = \sum_{i=1}^{p} \varphi(\lambda_i) = \int \varphi(\lambda_i) \, dF^{S_n}(x) =: F^{S_n}(\varphi)$$

for a specific function φ. One of the goals of is chapter is to deduce this specific function φ from the nonlinear shrinkage estimator such that it can be used for sphericity testing. Remember that by Theorem 2.1.1 the linear spectral statistic $F^{S_n}(\varphi)$ converges to $\int \varphi(\lambda) \, dF_c(\lambda)$, where F_c is the standard M-P law with index c. To actually construct a sphericity test from this particular LSS it is necessary to investigate the fluctuations around its limit under the null hypothesis. That is,

$$p\Big\{\frac{1}{p}\sum_{i=1}^{p} \varphi(\lambda_i) - \int \varphi(x) \, dF_c(x)\Big\} = p\Big\{\int \varphi(x) \, dF^{S_n}(x) - \int \varphi(x) \, dF_c(x)\Big\}$$

$$= p \int \varphi(x)(dF^{S_n}(x) - dF_c(x))$$

To investigate these fluctuations the central limit theorem for linear spectral statistics comes in handy, which is Theorem 3.4 from the book of Yao et al. (2015). This theorem says that the fluctuations of the LSS around its limit are normally distributed.

4.1. Preliminary Results

Before moving on, some machinery is needed, such as Theorem 3.4, Proposition 3.6 and Proposition 2.10 from the book of Yao et al. (2015). Theorem 3.4 is the central limit theorem (CLT) for linear spectral statistics, Proposition 3.6 helps reducing the difficulty of the calculations and Proposition 2.10 gives a way to calculate the limit of the linear spectral statistic. The three results are given below.

Theorem 4.1.1. *(CLT for Linear Spectral Statistics) Assume that the variables $\{x_{i,j}\}$ of the data matrix $X = (\mathbf{x_1}, \ldots, \mathbf{x_n})$ are independent and identically distributed satisfying $\mathbb{E}[x_{i,j}] = 0$, $\mathbb{E}[|x_{i,j}|^2] = 1$, $\mathbb{E}[|x_{i,j}|^4] = \beta + 1 + \kappa < \infty$, where $\kappa = 2$ in case of real variables and $\kappa = 1$ in case of complex variables, also $\mathbb{E}[x_{i,j}^2] = 0$ in case of complex variables. Assume, moreover,*

$$p \to \infty, \quad n \to \infty, \quad \frac{p}{n} \to c > 0$$

Let $f_1, \ldots, f_k$ be functions analytic on an open region containing the support of F_c. The random vector $\{X_n(f_1), \ldots, X_n(f_k)\}$ where

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

13

converges weakly to a Gaussian vector $X_{f_1}, \ldots, X_{f_k}$ with mean function and covariance function:

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

where

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

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

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

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

where the integrals are along contours (non-overlapping in J_1) enclosing the support of F_c

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

Proposition 4.1.1. *The limiting parameters in Theorem 4.1.1 can be expressed as follows:*

$$I_1(f) = \lim_{r \downarrow 1} I_1(f, r)$$

$$I_2(f) = \frac{1}{2\pi i} \oint_{|\xi|=1} f(|1 + h\xi|^2) \frac{1}{\xi^3} \, d\xi$$

$$J_1(f, g) = \lim_{r \downarrow 1} J_1(f, g, r)$$

$$J_2(f, g) = -\frac{1}{4\pi^2} \oint_{|\xi|=1} \frac{f(|1 + h\xi_1|^2)}{\xi_1^2} \, d\xi_1 \cdot \oint_{|\xi|=1} \frac{g(|1 + h\xi_2|^2)}{\xi_2^2} \, d\xi_2$$

with

$$I_1(f, r) = \frac{1}{2\pi i} \oint_{|\xi|=1} f(|1 + h\xi|^2) \left[\frac{\xi}{\xi^2 - r^{-2}} - \frac{1}{\xi} \right] d\xi$$

$$J_1(f, g, r) = -\frac{1}{4\pi^2} \oint_{|\xi_1|=1} \oint_{|\xi_2|=1} \frac{f(|1 + h\xi_1|^2) g(|1 + h\xi_2|^2)}{(\xi_1 - r\xi_2)^2} \, d\xi_1 d\xi_2$$

To calculate the limit of a linear spectral statistic the following proposition is useful:

Proposition 4.1.2. *For the standard Marchenko-Pastur distribution F_c with index $c > 0$ and $\sigma^2 = 1$, it holds for all functions f analytic on a domain containing the support interval $[a, b] = [(1 - \sqrt{c})^2, (1 + \sqrt{c})^2]$,*

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

4.2. Finding φ for the Linear Spectral Statistic

Now having acquired all knowledge and tools necessary, it is possible to investigate which function φ to use in a LSS. Since the goal of this thesis is to construct a sphericity test from the nonlinear shrinkage estimator the function φ should be derived from Equation (2.5) from Theorem 2.3.2. So lets investigate what would happen to Equation (2.5) under the null hypothesis $H_0 : \Sigma_n = I$. Remember that Equation (2.5) is given by

$$\delta(\lambda) = \frac{\lambda}{|1 - c - c\lambda \breve{m}_F(\lambda)|^2}, \quad \text{with} \quad \lim_{z \to \lambda} m_F(z) = \breve{m}_F(\lambda)$$

Computing Equation (2.5) under the null hypothesis requires the computation of $\breve{m}_F(\lambda) = \lim_{z \to \lambda} m_F(z)$ under the null hypothesis. Recall from Equation (2.1) that $m_F(z)$ satisfies:

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

Under the the null hypothesis $H_0 : \Sigma_n = I$, all eigenvalues are equal to 1 with multiplicity p. Therefore the underlying distribution of the true eigenvalues $H(\tau)$ jumps to 1 at $\tau = 1$. Hence under the null hypothesis $m_F(z)$ satisfies:

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

Rewriting the above expression gives a quadratic equation in $m_F(z)$:

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

Applying the quadratic formula gives the following solutions for $m_F(z)$:

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

where $a = (1 - \sqrt{c})^2$ and $b = (1 + \sqrt{c})^2$ are the boundaries of the support of the eigenvalues. Now since $z \in \mathbb{C}^+$ is a complex number, a branch cut for the square root has to be chosen. It is possible without loss of generality to chose the principal branch. Then since the eigenvalues are strictly positive ($c < 1$) the expression under the square root can not equal 0. Hence the square root is continuous and it is possible to take the limit $z \to \lambda$:

$$\begin{aligned}
\breve{m}_F(\lambda) &= \lim_{z \to \lambda} m_F(z) \\
&= \lim_{z \to \lambda} \frac{1 - c - z \pm \sqrt{(b - z)(a - z)}}{2cz} \\
&= \frac{1 - c - \lambda \pm \sqrt{(b - \lambda)(a - \lambda)}}{2c\lambda} \\
&= \frac{1 - c - \lambda \pm \sqrt{(\lambda - 1 - c)^2 - 4c}}{2c\lambda}
\end{aligned}$$

Substituting the above expression for $\breve{m}_F(\lambda)$ into Equation (2.5), the following function is obtained:

$$\varphi(\lambda) = \frac{4\lambda}{|c - 1 - \lambda \pm \sqrt{(\lambda - 1 - c)^2 - 4c}|^2} \tag{4.1}$$

This is Equation (2.5) under the null hypothesis and could be used in a LSS. However, Versteegh (2020) shows in his Theorem 5.2.1 that this is not convenient. This result is stated in the following theorem:

Theorem 4.2.1. *Let $\delta(\lambda_i)$ be the transformed eigenvalues as in Theorem 2.3.2, with $\lambda_i \in (a, b) = ((1 - \sqrt{c})^2, (1 + \sqrt{c})^2)$, the support of the Marchenko-Pastur distribution. Then under $H_0 : \Sigma_n = I$, $\delta(\lambda_i) = \varphi(\lambda_i) = 1$*

This theorem says essentially that under H_0 the eigenvalues are mapped into non-random quantities ($\varphi(\lambda_i) = 1, \forall i \in [1, p]$). So when using this function in a LSS, say $T = \frac{1}{p} \sum_{i=1}^{p} \varphi(\lambda_i)$, then under H_0, $T = \frac{1}{p} \cdot p \cdot 1 = 1$. This is a degenerate statistic, and using this would always result in power 0 or 1 and size 0 or 1. Therefore, the function $\varphi(\lambda)$ need to be adjusted to be non degenerate under H_0.

There are now multiple ways to proceed. The first option is to simply add an $\epsilon > 0$ to the denominator of Equation (4.1). This could be interesting to explore because by adding an $\epsilon > 0$ to the denominator it is possible that some of the key properties of this function will remain. First the equation has to be rewritten:

$$\begin{aligned}
\varphi_1(\lambda) &= \frac{4\lambda}{|c - 1 - \lambda \pm \sqrt{(\lambda - 1 - c)^2 - 4c}|^2} \\
&= \frac{\lambda}{|\frac{1}{2}(c - 1 - \lambda) \pm \frac{1}{2}\sqrt{(\lambda - 1 - c)^2 - 4c}|^2}
\end{aligned}$$

Then adding $\epsilon > 0$ to the denominator and using that $|\frac{1}{2}(c - 1 - \lambda) \pm \frac{1}{2}\sqrt{(\lambda - 1 - c)^2 - 4c}|^2 = \lambda$ for all $\lambda \in (a, b) = ((1 - \sqrt{c})^2, (1 + \sqrt{c})^2$, which is a result from the proof of Theorem 5.2.1 of Versteegh (2020), results in the following function:

$$\begin{aligned}
\varphi_1(\lambda) &= \frac{\lambda}{|\frac{1}{2}(c - 1 - \lambda) \pm \frac{1}{2}\sqrt{(\lambda - 1 - c)^2 - 4c}|^2 + \epsilon} \\
&= \frac{\lambda}{\lambda + \epsilon}
\end{aligned}$$

This function can now be used in the following LSS:

$$T_1 = \sum_{i=1}^{p} \varphi_1(\lambda_i) = \sum_{i=1}^{p} \frac{\lambda_i}{\lambda_i + \epsilon} \tag{4.2}$$

This is a non-degenerate linear spectral statistic and since $\varphi_1(\lambda)$ only has one signularity for $\lambda = -\epsilon$, the function is analytic on the support of the Marchenko-Pastur distribution of λ. Therefore, it is possible to apply the CLT for linear spectral statistics to T_1. It should be noted that for $\epsilon = 1$ this LSS is similar to the Bartlett-Nanda-Pillai (BNP) trace test statistic originally proposed by Pillai (1955). Where many year later its CLT for Fisher matrices is derived by Bodnar et al. (2019). So the linear spectral statistic of Equation (4.2) is actually a more general version of the BNP trace test statistic.

The second option is to first transform the eigenvalues with a linear function $f(\lambda)$ and then apply equation (4.1) to the transformed eigenvalues, that is $(\varphi \circ f)(\lambda) = \varphi(f(\lambda))$. Functions $f(\lambda)$ that intuitively make sense to use are, $f_\epsilon(\lambda) = \lambda + \epsilon$ for $\epsilon > 0$ or $f_\alpha(\lambda) = \alpha\lambda + (1 - \alpha)$ for $\alpha \in (0, 1)$. The second example is essentially a linear shrinkage function. However, by the following theorem proceeding in this way will not always be helpful:

Theorem 4.2.2. *Let $\varphi(x)$ be as in equation (4.1) and $f(\lambda_i)$ a linear function, with $\lambda_i \in (a, b) = ((1 - \sqrt{c})^2, (1 + \sqrt{c})^2)$, the support of the Marchenko-Pastur distribution. Then under $H_0 : \Sigma_n = I$,*

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

Proof. let $f(\lambda_i)$ a linear function with, $\lambda_i \in (a, b) = ((1 - \sqrt{c})^2, (1 + \sqrt{c})^2)$ and also $f(\lambda_i) \in (a, b)$. Recall that equation (4.1) can also be written as

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

where $(a, b) = ((1 - \sqrt{c})^2, (1 + \sqrt{c})^2)$, the support of the Marchenko-Pastur distribution. then

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

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

$$\sqrt{(b - f(\lambda_i)(a - f(\lambda_i))} = \sqrt{(-1)(b - f(\lambda_i)(f(\lambda_i) - a)} = i\sqrt{(b - f(\lambda_i)(f(\lambda_i) - a)}$$

Again the principal branch for the square root is used. Inserting this gives

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

For any complex number z it holds that $|z|^2 = \text{Re}(z)^2 + \text{Im}(z)^2$. Where in this case $\text{Re}(z) = c - 1 - f(\lambda_i)$ and $\text{Im}(z) = \sqrt{(b - f(\lambda_i)(f(\lambda_i - a)}$. Working this out gives

$$\begin{aligned} \text{Re}(z)^2 &= (c - 1 - f(\lambda_i))^2 \\ &= c^2 - 2cf(\lambda_i) + f(\lambda_i)^2 - 2c + 2f(\lambda_i)^2 + 1 \end{aligned}$$

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

Taking these together gives then

$$
\begin{aligned}
|z|^2 &= \text{Re}(z)^2 + \text{Im}(z)^2 \\
&= (c^2 - 2cf(\lambda_i) + f(\lambda_i)^2 - 2c + 2f(\lambda_i)^2 + 1) \\
&\quad + (-c^2 + 2cf(\lambda_i) - f(\lambda_i)^2 + 2c + 2f(\lambda_i) - 1) \\
&= 4f(\lambda_i)
\end{aligned}
$$

Hence

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

Now let $f(\lambda_i)$ a linear function with, $\lambda_i \in (a, b) = ((1 - \sqrt{c})^2, (1 + \sqrt{c})^2)$ but $f(\lambda_i) \notin (a, b)$. That means that $f(\lambda_i) < a$ or $f(\lambda_i) > b$. Then again

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

Now since $f(\lambda_i) < a$ or $f(\lambda_i) > b$, the expression in the square root is always positive. Therefore the expression inside the absolute value is real and it is possible to make use of the definition of absolute value for real values, $|x| = \sqrt{x^2}$. Hence

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

Combining the previous results gives the required equation. $\square$

When investigating $f_\epsilon(\lambda)$ more closely it can be found that, for all $\lambda_i \in (a, b) = ((1 - \sqrt{c})^2, (1 + \sqrt{c})^2)$ and $\epsilon > 4\sqrt{c}$, $f_\epsilon(\lambda) \notin (a, b)$. This means that when using $f_\epsilon(\lambda)$ with $\epsilon > 4\sqrt{c}$ to transform the eigenvalues and then using $\varphi(f(\lambda_\epsilon))$ in a LSS, the resulting statistic will be non degenerate. Therefore the second LSS that is proposed is

$$
T_2 = \sum_{i=1}^{p} \varphi_2(f_\epsilon(\lambda_i)) = \sum_{i=1}^{p} \frac{4(\lambda_i + \epsilon)}{(c - 1 - (\lambda_i + \epsilon) + \sqrt{((\lambda_i + \epsilon) - 1 - c)^2 - 4c})^2}
$$

The function $\varphi_(f_\epsilon(\lambda))$ only has one singularity in $\lambda = 0$ when $c \in (0, 1)$ and is therefore analytic on the support of the Marchenko-Pastur distribution of λ. So it is possible to apply the CLT for linear spectral statistics to T_2.

On the other hand, when investigating $f_\alpha(\lambda)$ more closely. It can be found that for all $\lambda_i \in (a, b) = ((1 - \sqrt{c})^2, (1 + \sqrt{c})^2)$, there exist no $\alpha \in (0, 1)$ such that $f_\alpha(\lambda) \notin (a, b)$. This means that is it not possible to find a $\alpha \in (0, 1)$ such that the resulting statistic is non degenerate. Therefore, it does not make sense to use $f_\alpha(\lambda)$ in a LSS.

In conclusion, there are two LSS to which the CLT can be applied to. These are

$$
T_1 = \sum_{i=1}^{p} \frac{\lambda_i}{\lambda_i + \epsilon} \quad \epsilon > 0 \tag{4.3}
$$

$$
T_2 = \sum_{i=1}^{p} \frac{4(\lambda_i + \epsilon)}{(c - 1 - (\lambda_i + \epsilon) + \sqrt{((\lambda_i + \epsilon) - 1 - c)^2 - 4c})^2} \quad \epsilon > 4\sqrt{c} \tag{4.4}
$$

The CLT from Theorem 4.1.1 will only be applied to LSS T_1. The second LSS T_2 can be used in further research because it requires to calculate the poles of the function $\varphi_2(f_\epsilon(|1 + \sqrt{c}z|^2))$, which are non trivial to find.

4.3. CLT for Linear Spectral Statistics
In this section the main result of this chapter will be presented; the central limit theorem for the linear spectral statistics T_1, which is Equation (4.3) found in the previous section. To calculate the CLT for this LSS, Theorem 4.1.1, Proposition 4.1.1 and Proposition 4.1.2 are used.

Theorem 4.3.1. *let X be a p x n data matrix consisting of independent identically distributed random variables with mean 0 and variance 1, and let $Y = \Sigma^{\frac{1}{2}} X$. Let $S_n = \frac{1}{n} Y Y^T$ be its sample covariance matrix with eigenvalues $\lambda_1, \cdots, \lambda_p$. Let $(p, n) \to \infty$, $\frac{p}{n} \to c \in (0, 1)$ and $\epsilon > 0$. Then under $H_0; \Sigma_n = I$*

$$\sum_{i=1}^{p} \frac{\lambda_i}{\lambda_i + \epsilon} + p \cdot \frac{A^2 B^2 - 2B^2 + 1}{2\sqrt{c}(A - B)B^2} \to N(\mu, \sigma^2)$$

in distribution, where

$$\mu = (\kappa - 1)\left[\frac{(\sqrt{c} + B)(B\sqrt{c} + 1)}{\sqrt{c}(A - B)(B^2 - 1)B} \right] + \beta\left[\frac{(\sqrt{c} + B)(B\sqrt{c} + 1)}{B^3 \sqrt{c}(A - B)} \right] \tag{4.5}$$

$$\sigma^2 = \kappa\left[\frac{(B + \sqrt{c})(B\sqrt{c} + 1)(\sqrt{c} + A)(A\sqrt{c} + 1)}{(A - B)^4 c} \right] + \beta\left[\frac{(\sqrt{c} + A)(A\sqrt{c} + 1)}{\sqrt{c}A^2(A - B)} - \frac{\epsilon}{\sqrt{c}} \right]^2 \tag{4.6}$$

and

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

The proof of this theorem can be found in the Appendix A.1. To demonstrate the limiting behaviour of the fluctuations of Equation 4.3 around its limit, define the random variable W

$$W = \sum_{i=1}^{p} \frac{\lambda_i}{\lambda_i + \epsilon} + p \cdot \frac{A^2 B^2 - 2B^2 + 1}{\sqrt{c}2(A - B)B^2} \tag{4.7}$$

then by theorem 4.3.1, $W \to N(\mu, \sigma^2)$ in distribution with μ and σ^2 defined by Equations (4.5) and (4.6). Now when W is properly centralized, that is

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

then $Z \to N(0, 1)$ in distribution. This means that the random variable Z converges to a standard normal in distribution. To visualize this, in Figure 4.1(a) the empirical distribution function of the random variable Z is calculated with $p = 128$, $n = 256$, $\epsilon = 1$ and with standard normal distributed data. In Figure 4.1(b) the empirical distribution with $p = 128$, $n = 256$, $\epsilon = 1$ but now with $Gamma(4, 2) - 2$ data is calculated. For the second figure a $Gamma(4, 2) - 2$ distribution for the data is chosen because this gives $\beta = 3/2$ instead of $\beta = 0$ (for standard normal data) but has still zero mean and unit variance. To obtain a proper empirical distribution function, 100.000 replications are used.

(a) $n = 256$, $p = 128$ and $\epsilon = 1$ (b) $n = 256$, $p = 128$ and $\epsilon = 1$
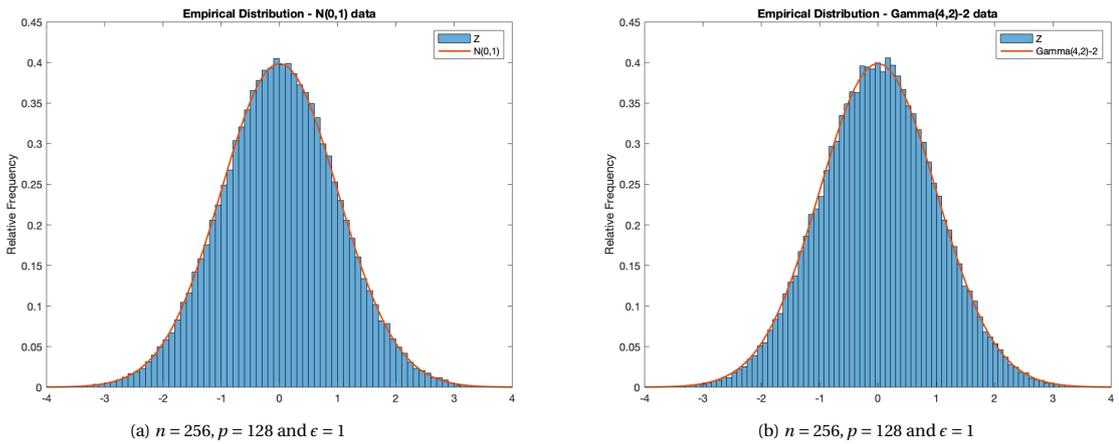
Figure 4.1: Empirical distribution functions for centralized random variable W given by Equation (4.7)

Note that the empirical distribution functions in Figure 4.1(a) and 4.1(b) are approximately standard normal distributions. Therefore it is save to conclude that the calculated limiting distribution in Theorem 4.3.1 is correct. More importantly, the random variable W of Equation (4.7) is a ready to use spericity test statistic in the

large-dimensional case that is based on the nonlinear shrinkage estimator. The new sphericity test that uses this new test statistic W is called the nonlinear shirnkage test or for short the NLS-ϵ test. In the next chapter this new spericity test will be compared with already existing tests in a simulation study

In Figure 4.1 $\epsilon = 1$ was chosen arbitrarily. However in the general case ϵ need to be chosen such that the power of the test statistic is maximal. This will be done in the next section.

4.4. Maximizing Power of the New NLS-ϵ Test

In this section, the optimal test statistic W from Equation (4.7) is calculated. This will be done by calculating $\epsilon > 0$ such that the power of the NLS-ϵ test is maximal. Recall that the power of the test is defined by $P(\text{reject } H_0 | H_1 \text{ is true})$. That is the probability that H_0 is rejected given that H_1 is true or just the probability that H_0 is rejected correctly. So to find the optimal value for ϵ of the new NLS-ϵ test, there need to be conditioned on the alternative hypothesis H_1. Given this new assumption the following function needs to be maximized over $\epsilon > 0$.

$$\begin{aligned} \text{Power} &= P(\text{reject } H_0 | H_1 \text{ is true}) \\ &= P(|T_1| > t | H_1) \\ &= P(T_1 > t | H_1) + P(T_1 < -t | H_1) \end{aligned}$$

Define $g_\epsilon(t) = P(T_1 > t | H_1) + P(T_1 < -t | H_1)$. Then maximizing the function $g_\epsilon(t)$ with respect to ϵ is not possible because it does not have the same distribution as derived in Theorem 4.3.1. This is because Theorem 4.3.1 holds under H_0 and it was just mentioned that to find the ϵ for which te power is maximum, there need to be conditioned on a alternative hypothesis H_1. Therefore before maximizing $g_\epsilon(t)$ with respect to ϵ the distribution of $g_\epsilon(t)$ need to be found under H_1. Unfortunately, this can not be done with the CLT of Theorem 4.1.1. However, it can be done with Theorem 3.9 of Yao et al. (2015). This theorem gives the CLT for linear spectral statistics of random Fisher matrices but this theorem is based on a different kind of stochastic model and it is not as general as one would like it to be. Because one of the assumptions is that $\mathbb{E}[x_{i,j}^4] = 3$. In other words, this means that the data need to come from a standard normal distribution. This is a quid restrictive assumption. To find a more general CLT the paper of Najim and Yao (2016) need to be consulted. Najim and Yao (2016) derived a general CLT which is far more general then the ones discussed before can therefore be used to find this optimal ϵ. The downside is that this CLT is very tough to work with and it outside the scope of a bachelor thesis. So it will not be included in this work.

An other approach to this problem is to find the optimal value for ϵ numerically. This will be done in the next chapter simultaneously with comparing the new NLS-ϵ test with other sphericity test for large-dimensional statistics.

5

Simulation Study

In this section the optimal value(s) for ϵ will be found for which the new nonlinear shrinkage sphericity test (NLS-ϵ), derived in Theorem 4.3.1, has the highest power. Furthermore, the NLS-ϵ test will be compered with other already existing test mentioned earlier in chapter 3. Finding the optimal value for ϵ and comparing the different test will actually be done simultaneously. How this will work will be discussed in a minute. First the setting of the simulations study will be specified.

5.1. Setting of the Simulations

Before diving into the simulations, the setting need to be specified. Remember that all the results of this thesis are based on the large-dimensional asymptotic framework or also known as the large-dimensional statistics framework. That is, n denotes the sample size and p the number of variables, with $\frac{p}{n} \to c \in (0, 1)$ as $(p, n) \to \infty$. Moreover, some assumptions for the data are needed. The data comes from the stochastic model $Y = \Sigma^{\frac{1}{2}} X_n$ with the following assumptions

- **(B1)** The population covariance matrix Σ_n is a non-random p-dimensional positive definite matrix.

- **(B2)** Let X_n be an n × p matrix of real independent and identically distributed (i.i.d.) random variables with mean zero and unit variance. One only observes Y where $Y = \Sigma^{\frac{1}{2}} X_n$.

- **(B3)** Let $\tau_{n,1}, \ldots, \tau_{n,p}$ denote the eigenvalues of the population matrix Σ_n. The empirical spectral distribution (ESD) of the population eigenvalues $H_n(\tau)$ converges a.s. to a nonrandom limit $H(\tau)$ at every point of continuity.

Then the tests with corresponding test statistics that will be compared are:

- **Corrected Likelihood Ratio Test (CLRT)**
  For $c \in (0, 1)$

$$T_1 = -\frac{2}{n} \cdot \mathrm{Log}\left[\left(\frac{(\lambda_1 \cdots \lambda_p)^{\frac{1}{p}}}{\frac{1}{p}(\lambda_1 + \cdots + \lambda_p)}\right)^{\frac{pn}{2}}\right] + (p-n)\mathrm{Log}(1-\frac{p}{n}) - p \to N\left(-\frac{\kappa-1}{2}\mathrm{Log}(1-c) + \frac{1}{2}\beta c, -\kappa\mathrm{Log}(1-c) - \kappa c\right)$$

- **Corrected John's Test (CJ)**

$$T_2 = \frac{2}{p}\left(\frac{p^2 n}{2} tr\{S_n(tr(S_n))^{-1} - Ip^{-1}\}^2\right) - p \to N(\kappa - 1 + \beta, 2\kappa)$$

- **Linear Shrinkage Test (LS)**

$$T_3 = p\left(1 - \frac{\frac{1}{n}tr(S_n)^2 ||\Sigma_0||_F^2}{||S_n||_F^2 ||\Sigma_0||_F^2 - (tr(S_n\Sigma_n))^2}\right) \to N(\kappa - 1 + \beta, 2\kappa)$$

- **Nonlinear Shrinkage Test (NLS-ϵ)**

  For $c \in (0,1)$ and $\epsilon > 0$

  $$T_{4,\epsilon} = \sum_{i=1}^{p} \frac{\lambda_i}{\lambda_i + \epsilon} + p \cdot \frac{A^2 B^2 - 2B^2 + 1}{2\sqrt{c}(A - B)B^2} \to N(\mu, \sigma^2)$$

  where

  $$\mu = (\kappa - 1)\left[\frac{(\sqrt{c} + B)(B\sqrt{c} + 1)}{\sqrt{c}(A - B)(B^2 - 1)B}\right] + \beta\left[\frac{(\sqrt{c} + B)(B\sqrt{c} + 1)}{B^3\sqrt{c}(A - B)}\right]$$

  $$\sigma^2 = \kappa\left[\frac{(B + \sqrt{c})(B\sqrt{c} + 1)(\sqrt{c} + A)(A\sqrt{c} + 1)}{(A - B)^4 c}\right] + \beta\left[\frac{(\sqrt{c} + A)(A\sqrt{c} + 1)}{\sqrt{c}A^2(A - B)} - \frac{\epsilon}{\sqrt{c}}\right]^2$$

  and

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

In the above test statistics the κ stands for whether the data is real or not. In this simulation study the data will always be real and therefore $\kappa = 2$. The β in the test statistics is equal to $\beta = \mathbb{E}[|x_{i,j}|^4] - \kappa - 1$ and will therefore vary because the data will be taken from a standard normal distribution and a $Gamma(4,2) - 2$ distribution. This will result in a β of repressively $\beta = 0$ and $\beta = 1,5$.

It should be noted that the CLRT, CJ and LS tests are all one-tailed tests, because for the parameters that are used in this simulation the test statistics will always be positive. However, the NLS-ϵ test is a two-tailed test, this is because for the parameters that are used in this simulation the test statistic will give positive as well as negative values. It is known that one-tailed tests can be transformed to two-tailed tests without changing anything. This will also be done is this simulation to compare all the test equally. Moreover, it should be noted that the CJ and the LS test statistics have the same limiting distribution and that the limiting distributions of CLRT and the NLS-ϵ test statistics depend on c. In particular, the CLRT depends on the $\log(1 - c)$, so it is expected that this test will breakdown when c increases to 1 and will not work when $c > 1$.

The null-hypothesis that will be tested is as always $H_0 : \Sigma_n = I$, where I is the identity matrix. During the comparison between the four different test, there will also be looked at for which ϵ, the NLS-ϵ test has the highest power. So comparing the different tests and finding the optimal ϵ will be done simultaneously. To compare the different test, there will be made use of the following indicators: the empirical size, the empirical power and ROC curves.

The calculation of the empirical size and empirical power go in similar manner. Recall that the size of a test is equal to $P(\text{reject } H_0 | H_0 \text{ is true})$ and the power of a test is equal to $P(\text{reject } H_0 | H_1 \text{ is true})$. For a general two-tailed test statistic T with known distribution $T \sim N(\mu, \sigma^2)$, the null hypothesis is rejected whenever the centralized statistic $Z = \frac{T - \mu}{\sigma} < -w(\alpha/2)$ or $Z = \frac{T - \mu}{\sigma} > w(\alpha/2)$. The value $w(\alpha/2)$ is the value the test statistic would need to to exceed and depend on the distribution of the test statistic and the required accuracy. Then proceeding in the following manner:

1. Calculating the sample covariance matrix S_n from the generated data

2. Calculating the test statistics

3. Check whether the centralized test statistics exceed a prespecified $w(\alpha/2)$.

gives a Bernoulli experiment. Depending on which hypothesis one conditions, this experiment either has success probability $p_0 = P(\text{reject } H_0 | H_0 \text{ is true})$ or $p_1 = P(\text{reject } H_0 | H_1 \text{ is true})$. By definitions of the size and power of a statistical test, $p_0 = $ Size and $p_1 = $ Power. Then by the weak law of large numbers, when the number of trails is large enough and assuming H_0 is true, one has:

$$\frac{\text{Number of times } H_0 \text{ is rejected}}{n} \to p_0 = \text{ Size} \quad \text{(In probability)}$$

In the same way, assuming a particular H_1 is true, one has:

$$\frac{\text{Number of times } H_0 \text{ is rejected}}{n} \to p_1 = \text{ Power} \quad \text{(In probability)}$$

This is how the empirical size and empirical power is calculated. The next section starts with the calculation of the empirical size.

5.2. Empirical Size Comparison

Before the tests will be compared using their powers, the empirical sizes will be compared. For a statistic to be close to its limiting distribution, the empirical size should be close to a particular rejection level α. Therefore, the closer the size is to a fixed rejection level α, the better it is. Also to make a fair comparison later on with the empirical powers, the tests should all be close to the rejection level. Now if Z is standard normally distributed random variable and the significance level is equal to α, then the null hypothesis is rejected whenever $Z < -w(\alpha/2)$ or $Z > w(\alpha/2)$. Without loss of generality it is possible to choose the rejection level equal to $\alpha = 0.05$. This will lead to the rejection of the null hypothesis whenever $Z < -w(0.025) = -1.960$ or $Z > w(0.025) = 1.960$. This is a commonly chosen rejected level because for a two-tailed test it represents the 0.250 and the 0.975 percentile of the standard normal distribution. Simulating the previous described Bernoulli experiment for the sizes 10000 times, give the sizes in the tables below. The sizes in table 5.1 are based on real standard normal data and the sizes in table 5.2 are based on real $Gamma(4,2) - 2$ data.

(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

Table 5.1: Empirical sizes at 5% significance level based on 10000 independent realisations of real $N(0,1)$ random variables

(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

Table 5.2: Empirical sizes at 5% significance level based on 10000 independent realisations of real $Gamma(4,2) - 2$ random variables

It can be seen in Table 5.1 that all the empirical sizes based on standard normal data are close to the rejection level $\alpha = 0.05$. This means that all test statistics are close to their limiting distribution. Moreover, 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 starting point will approximate be the same and a fair comparison can be made. Unfortunately this is not the case for the empirical sizes based on $Gamma(4,2) - 2$ data. It can be seen in Table 5.2 that the empirical sizes of the NLS-ϵ and LS tests behave quite well. For lower combinations of (p, n) the empirical sizes of the NLS-ϵ test seem a little low but overall they are close to the rejection level α. So it can be concluded that for these combinations of (p, n), the NLS-ϵ test and the LS test are close to their limiting distributions when the data is based on a $Gamma(4,2) - 2$ distribution. However for the CJ test this only holds for higher combinations of (p, n). It looks like that the empirical sizes of the CJ test approaches α from above. This means that when (p, n) is low, the empirical distribution has bigger tails than it should be but when (p, n) is increasing the empirical distribution is getting closer to its limiting distribution. Thus in this case the CJ test relies more on the limiting aspect. The empirical sizes for the CLRT test are behaving quite pore for every combination of (p, n). They are approximately half of what its should be and its getting worse when (p, n) is increasing. From this observation it can be concluded that when the data is based on a $Gamma(4,2) - 2$ distribution the CLRT test is not close to its limiting distribution at all. Therefore, it will be difficult to make a fair empirical power comparison when the data is taken from the $Gamma(4,2) - 2$ distribution because not all test will have the same starting point. So the empirical power comparison will only be based on standard normal data.

5.3. Empirical Power Comparison

In this section the empirical powers for all test will be compared, including the NLS-ϵ test for different values of ϵ. The comparison will only be based on standard normal data because the empirical sizes based on $Gamma(4,2) - 2$ data are not all close to the rejection level α and will lead to unfair comparisons. In this simulation study the empirical power will be based on the increasing distance between the null hypothesis $H_0 : \Sigma_n = I$ and a particular alternative hypothesis $H_1 : \Sigma_1 \neq I$. To actually calculate the empirical powers the Bernoulli experiment from Section 5.1 will be used again. Only this time there will be conditioned on a particular alternative hypotheses. The three alternative hypothesis, or in other words, the three ways to increase the distance, are

1. H_1: Equicorrelation relation

2. H_1: Autoregressive relation

3. H_1: Fixed ratio of variables have variance $\neq 1$

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 repressively $c = 1/4$, $c = 1/2$, $c = 3/4$ and $c = 15/16$. $n = 256$ will not be used and all comparisons are based on 1000 repetitions because of computational reasons. For demonstrating purposes, we assume these parameters to be large enough such that the limiting distributions of the statistics are present.

5.3.1. Equicorrelation Relation

The first alternative hypothesis that will be used to make a power comparison is an equicorrelation relation. Equicorrelation 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 equicorrelation 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 equicorrelation alternative hypotheses is defined as

$$\Sigma_{n,\rho} = (1 - \rho)I + \rho \cdot ones(p, p)$$

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

In the simulation ρ runs from 0 to 1. So as ρ increases, the alternative hypothesis $\Sigma_{n,\rho}$ becomes less like the identity matrix I or the null hypothesis. This is also meant by increasing the distance from the null hypothesis

to the alternative 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 form Section 5.1 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 ρ against the obtained power gives the required power plot. The quicker a test reached a power of 1 the better the test is because remember, the power is the probability that a falls null hypothesis is correctly rejected. So the quicker the better. After some investigations it seems like the most interesting ϵ's under the equicorrelation alternative for the NLS-ϵ test are $\epsilon = 1.5, \epsilon = 1$ and $\epsilon = 0.5$. Doing the simulation gives the following power plots.

(a) $p = 32$ (b) $p = 64$

(c) $p = 96$ (d) $p = 120$
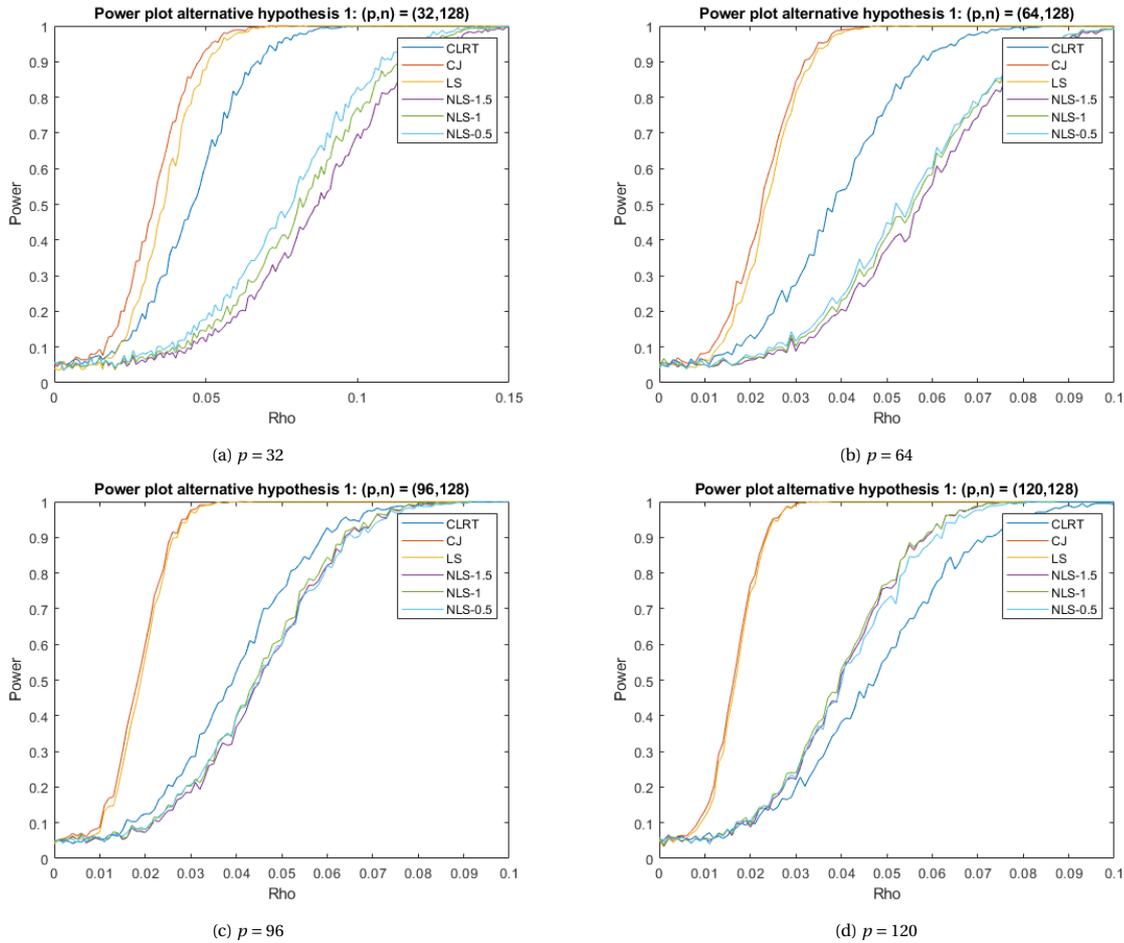
Figure 5.1: Empirical powers under alternative hypothesis 1 with 1000 replications and $\rho \in (0, 1)$

In Figure 5.1 it can be seen how the tests perform in terms of power for different combinations of $\frac{p}{n} = c$. For $\rho < 0.15$ all test have power 1 and as expected the CJ and the LS test behave nearly the same. This is due to the fact that they have the same limiting distribution. Most noticeable in Figure 5.1 is that when p increases, the NLS-ϵ performs better and the CLRT performs worse. For $(p, n) = (120, 128)$ the NLS-ϵ test even outperforms the CLRT test. The reason why the performs of both tests changes when c changes is that, as said earlier, the limiting distributions of these test depend on c. 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 5.1 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. The increase of the optimal ϵ could be to compensate for some numerical issues when p increases to n. So for the equicorrelation alternative the optimal ϵ for the NLS-ϵ test depend on the combination of (p, n).

5.3.2. Autoregressive Relation

The second alternative hypothesis is the autoregressive relation. This autoregressive relation is based on a autoregressive model, which is a type of random process that is used to describe time-varying processes. The autoregressive model specifies that the output variable depends linearly on its own previous values and on a stochastic term. In the autoregressive alternative hypothesis the entries of the matrix $\Sigma_{n,\delta}$ also depend recursive on each other. For $\delta \in \mathbb{R}$, define the entry on the $i-th$ row and the $j-th$ column of the alternative hypothesis matrix as $\delta^{|i-j|}$. The autoregressive alternative hypothesis is then defined as

$$\Sigma_{n,\delta} = \begin{bmatrix} 1 & \delta & \delta^2 & \cdots & \delta^{p-1} \\ \delta & 1 & \delta & \cdots & \delta^{p-2} \\ \delta^2 & \delta & \ddots & & \vdots \\ \vdots & & & \ddots & \delta \\ \delta^{p-1} & \delta^{p-2} & \cdots & \delta & 1 \end{bmatrix}$$

Any $\delta \in \mathbb{R}$ could be picked but in this simulation $\delta \in (-1,1)$ is chosen, because this corresponds with a stationary autoregressive model. The simulation goes in similar way as in the previous section. δ runs from -1 to 1. As δ goes away from 0 in both directions, this could be seen as moving away form the null hypothesis $H_0 : \Sigma_n = I$ because the alternative hypothesis matrix becomes less like the true covariance matrix. Then for every delta the Bernoulli experiment form Section 5.1 is carried out. Varying δ will give different empirical powers for each δ. Plotting δ against the obtained empirical powers will then give the required power plot to compare the different tests. The faster a test has power 1 the better the test is performing, as already explained in the previous subsection. Again after some inspection the most interesting ϵ's to consider for NLS-ϵ test are $\epsilon = 1$, $\epsilon = 0.5$ and $\epsilon = 0.1$. Note that not all of these ϵ's are the same as in the simulation for the first alternative hypothesis. Doing the simulations give the following power plots:

(a) $p = 32$

(b) $p = 64$

(c) $p = 96$

(d) $p = 120$
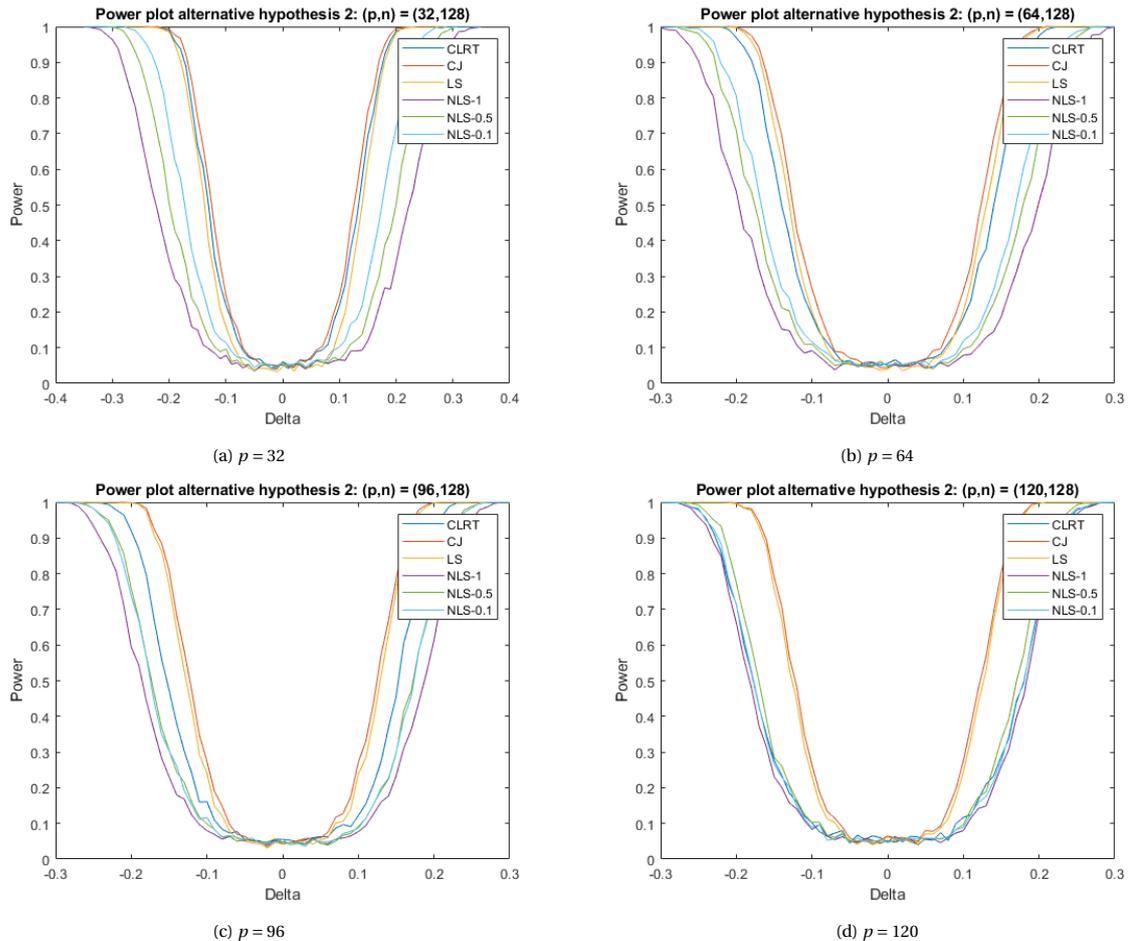
Figure 5.2: Empirical powers under alternative hypothesis 2 with 1000 replications and $\delta \in (-1,1)$

In can be seen in Figure 5.2 that for $p = 32$ and $p = 64$ the CJ, LS and CLRT 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 test are still performing best but the CLRT test is performing worse and worse. For $p = 120$ the NLS-ϵ test even outperforms the CLRT test when $\epsilon = 0.5$. This is in line with the observations from the previous simulation for the equicorrelation alternative, only now with different ϵ.

Then specifically focusing on the NLS-ϵ test, it can be seen in Figure 5.2 that for $p = 32$ and $p = 64$ the NLS-0.1 test performs best. While when p gets bigger, the NLS-0.5 test takes the lead. Again the the optimal ϵ varies with different values of p compared with n and is in line with the observations from the simulation under the equicorrelation alternative.

5.3.3. Fixed ratio with variance other than one

The third and last alternative hypothesis that will be considered is the alternative hypothesis where a fixed ratio r of the variables have a variance not equal to 1, but equal to $1 + \gamma$. For any $r \in (0, 1)$ and $\gamma \in \mathbb{R}$, the third alternative hypotheses is defined as

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

Suppose $r = 1/2$, this means that half of the variables p have variance equal to $1 + \gamma$. If it happens that $r \cdot p$ is not a whole number, it will be rounded down. In the same way as in the previous simulation γ will run form -1 to 1. This can again be seen as departing from the null hypotheses $H_0 : \Sigma_n = I$ when γ goes away from 0 in both directions. Because the alternative hypohtesis matrix $\Sigma_{n,r,\gamma}$ becomes less like the true covariance matrix Σ_n. Then for every γ the Bernoulli experiment form Section 5.1 is executed. This will give different empirical powers for each γ. Plotting γ against the obtained empirical powers will then give the required power plot to compare the tests. For demonstrating purposes only the power plots for $r = 1/2$ will be extensively analysed. The power plots for $r = 1/4$ and $r = 3/4$ can be found in Appendix B.1. After some pre-analysis of the NLS-ϵ test under this alternative hypothesis the most interesting ϵ's to consider are $\epsilon = 10$, $\epsilon = 1$ and $\epsilon = 0.1$. Doing the simulation with 1000 repetitions gives the power plots for $r = 1/2$.

It can be seen in Figure 5.3 on the next page that the NLS-ϵ test performs by far the best. The NLS-ϵ test reaches a power of 1 much faster than the other tests. For low c the CJ, LS and CLRT test are again quite comparable. However, when p increases the CLRT is getting worse and worse for the same reason as in the previous simulations. So it can be concluded that the NLS-ϵ tests performs best under the fixed ratio with variance other than 1 alternative. Furthermore, it can be seen that the NLS-ϵ tests are symmetric around zero but the other test not because the power of the other tests increase must faster for negative values of γ than for positive values.

It was mentioned that only the power plots of $r = 1/2$ would be analysed extensively but there are some thing that should be noted. In Figures B.1 and B.2 it can be seen that the NLS-ϵ test works better when r increases form 0 to 1. While the other tests only perform better when r increases to $1/2$ because when $r > 1/2$ the power decreases again, specially when γ is positive. This behaviour can be explained because the null hypothesis that is actually tested is whether the true 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 alternative 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 matrix what already was expected. The NLS-ϵ is not invariant under multiples of the identity matrix because when r increases form 0 to 1, thus moving away from the null hypothesis when no multiples of the identity matrix are allowed, the test gets only more powerful.

Now diving deeper into the NLS-ϵ test, it can be seen in Figure 5.3 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. Because for the previous alternative hypotheses the optimal ϵ was depended on p while in for fixed ratio with variance other than 1 alternative, its seems the bigger ϵ is the better.

(a) $p = 32$ (b) $p = 64$

(c) $p = 96$ (d) $p = 120$
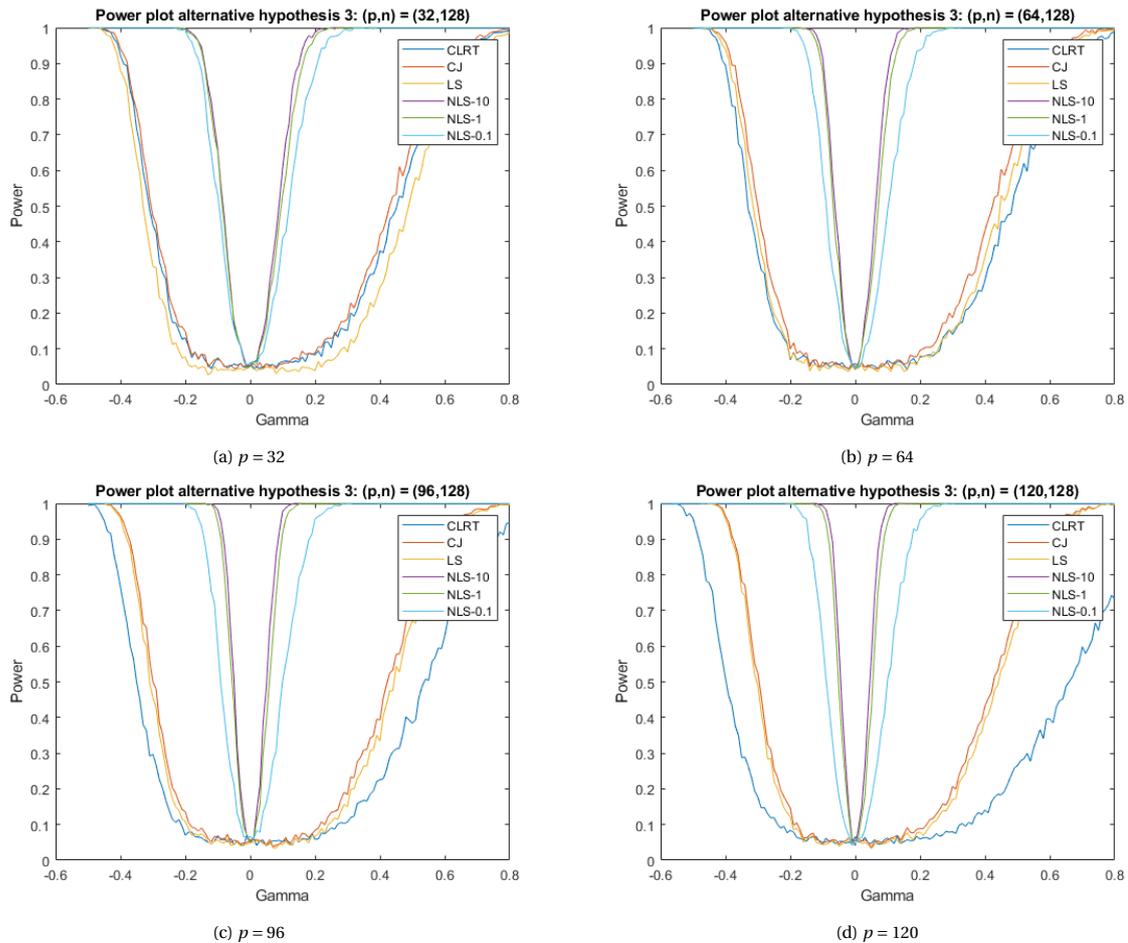
Figure 5.3: Empirical powers under alternative hypothesis 3 with 1000 replications, $\gamma \in (-1, 1)$ and $r = 1/2$

5.4. ROC Curves

In the previous section it is observed that the powers of the NLS-ϵ test are close to together for different ϵ's as well as the powers of the LS and CJ tests. When the powers are close it is more difficult to compare them and come to a well considered conclusion which test is better. Therefore to compare different tests and assess there quality when their powers are close to each other, receiver operating characteristic (ROC) curves are used. ROC curves are plots that compare how the true positive rate, or in other words the power, changes as the desired false positive rate is varied. It is a type of power plot, but instead of varying the underlying distance variable as in the previous section, the significance level is varied for which one rejects the null hypothesis. Since the CJ and the LS test perform much better or much worse than the NLS-ϵ test, it is much more interesting to have a closer look at the NLS-ϵ test for different ϵ's. Furthermore, because the NLS-ϵ test performs sometimes the same or even better than the CLRT test, it is interesting to include this test as well. Therefore, only the the CLRT and the NLS-ϵ test for different ϵ's will be compared using the ROC curves in this section.

To assess the quality of the tests using ROC curves, there will again be conditioned on one of the three alternative hypothesis used in the previous section. However, as said earlier, not the distance variable will be varied but the significance level α for which one rejects the alternative hypothesis. This means that, depending on which alternative hypothesis there will be conditioned, a value of either ρ, δ and γ need to be chosen for which the powers are close. However, it is also necessary to choose these values for the distance variables such that they are as far way as possible from the null hypothesis because otherwise the tests has not gained any power yet. So the distance values need to be chosen such that the powers of the tests are close but in the same time are far away enough from the null hypothesis.

In the empirical power plots of the previous section the significance level α was fixed. Now it will run from 0 to 1. Since the significance level is changing, the value $w(\alpha/2)$ the centralized test statistic need to exceed to reject the null hypothesis changes as well. The value $w(\alpha/2)$ can be computed using the following relation:

$$1 - \frac{\alpha}{2} = \frac{1}{2\sqrt{\pi}} \int_0^{w(\alpha/2)} e^{\frac{-t^2}{2}} \, dt$$

$$1 - \frac{\alpha}{2} = \frac{1}{\sqrt{\pi}} \int_0^{w(\alpha/2)/\sqrt{2}} e^{-t^2} \, dt$$

$$1 - \frac{\alpha}{2} = \frac{1}{2} + \frac{1}{2}\text{erf}(w(\alpha/2)/\sqrt{2})$$

$$w(\alpha/2) = \sqrt{2} \cdot \text{erfinv}(1 - \alpha) \tag{5.1}$$

The error function $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} \, dt$ and its inverse $\text{erfinv}(x)$ are used because they can easily be computed numerically. If $\alpha = 0.05$ is inserted into Equation (5.1), one would find $w(0.025) = 1.960$. Now if $\alpha = 0$ is inserted into Equation (5.1), one would find $w(0) = \infty$. If $\alpha = 0$, it is possible to think of this as no outcome is significant because the test statistic can never be greater than infinity and the null hypothesis is never rejected no matter what hypothesis is true. Therefore for $\alpha = 0$, the power is equal to 0 as well. Now if α grows to one, depending on the test, there will be different but increasing powers. Then when $\alpha = 1$, which corresponds to the belief that every outcome of the test is significant, there will always be rejected. This is because $w(1) = -\infty$ and the test statistic is always greater than minus infinity. This will give a power of 1.

The comparison using ROC curves work in the following way: α runs from 0 to 1 and using Equation (5.1) the value $w(\alpha/2)$ which the test statistic need to exceed to reject the null hypothesis will be calculated. Then for every rejection level $w(\alpha/2)$ the empirical power is calculated. The test which gains more power for the same significance level can be considered as a more powerfully test. There is also a straight line included in the ROC plots. This line will be referred to as the standard line and the better the test is, the further away it is from this line. 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 repressively $c = 1/4$, $c = 1/2$, $c = 3/4$ and $c = 15/16$. $n = 256$ will not be used and all comparisons are based on 1000 repetitions because of computational reasons. As in the previous section there will be started with comparing the ROC curves for the equicorrelation alternative hypothesis.

5.4.1. ROC Curves for Equicorrelation

Before the comparison of the ROC curves for the equicorrelation can be made, a suitable value for the distance variable ρ need to be chosen. In Figure 5.1 it can be seen that for $\rho = 0.03$ the empirical powers are relatively close for all combinations of (p,n). Therefore the ROC curves will be calculated using $\rho = 0.03$ but other values for ρ may suffice as well. The ϵ's that will be used are the same as used in the power plots for the equicorrelation alternative, these are $\epsilon = 1.5$, $\epsilon = 1$ and $\epsilon = 0.5$.

The calculated ROC curves can be found in Figure 5.4. In this figure it can be seen that for every value of ϵ, the ROC curves of the NLS-ϵ tests are still close together. However for small p, $\epsilon = 0.5$ seems to be optimal because the ROC curve for the NLS-0.5 is furthest away from the standard line compared to the other NLS-ϵ tests, but for $p = 120$ $\epsilon = 1$ seems to work better. Still the margins are very small. An other thing that should be noted in Figure 5.4 is that the NLS-ϵ test is performing way better when p increases and that the performance of the CLRT is not really changing when p changes. That the NLS-ϵ performs better when p increases compared to n is in line with the observations made from Figure 5.1, which are the power plots of the equicorrelation alternative.

(a) $p = 32$

(b) $p = 64$

(c) $p = 96$

(d) $p = 120$
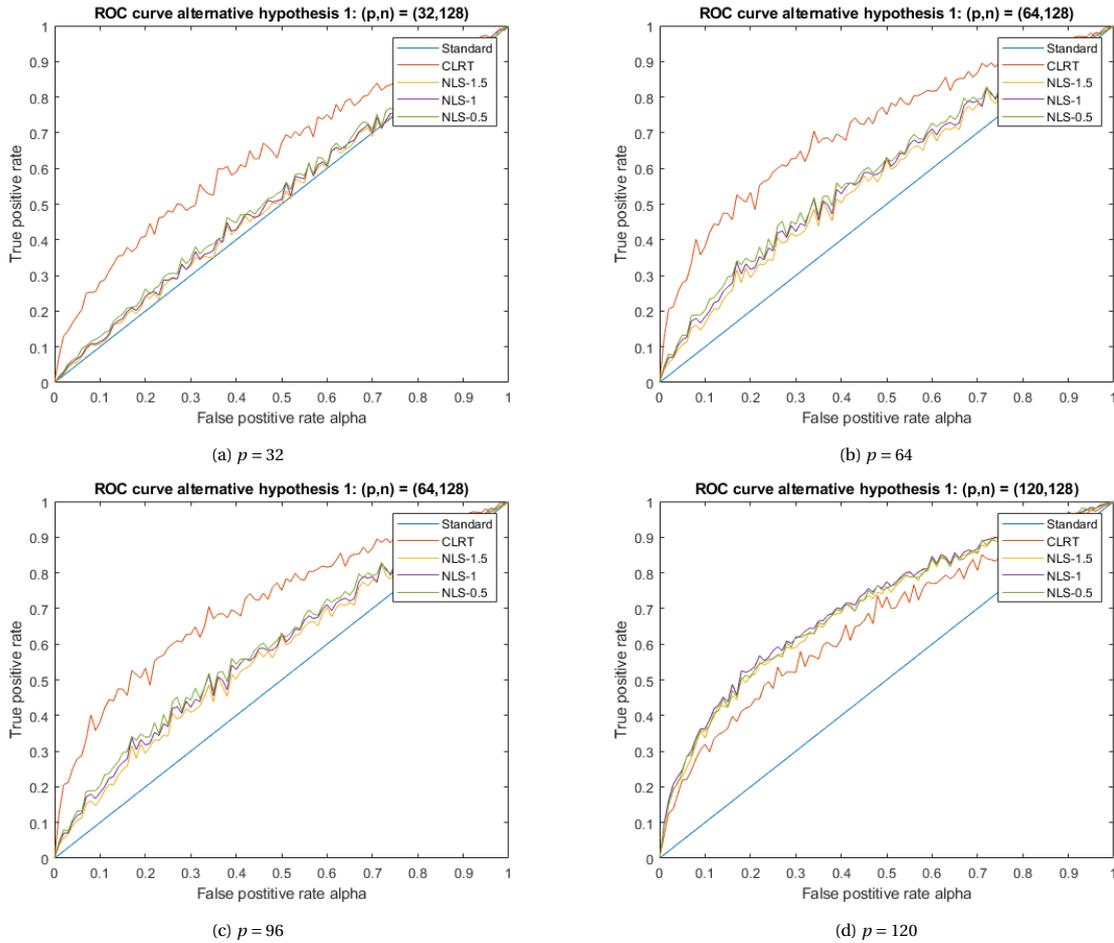
Figure 5.4: ROC plots under alternative hypothesis 1 with 1000 replications and $\rho = 0.03$

5.4.2. ROC Curves for Autoregressive Relation

The second alternative hypothesis for which the ROC curves will be calculated is the autoregressive alternative. In the same way as in the previous subsection a suitable value for $\delta \neq 0$ need to be chosen. In Figure 5.2 it can be seen that for $\delta = 0.12$ the empirical powers are relatively close for all combinations of (p, n). Therefore, the simulation will be done with $\delta = 0.12$. The same ϵ's that were used in the power plots for this alternative hypothesis will be used again. These are, $\epsilon = 1$, $\epsilon = 0.5$ and $\epsilon = 0.1$. The ROC curves with $\delta = 0.12$ with 1000 receptions can be found in Figure 5.5.

In Figure 5.5 something different is happening from the equicorrelation alternative. The CRLT is for small p already quite power full but decreases in power when p increases. The NLS-ϵ tests behave in opposite direction, these tests again increase in power when p increases compared to n. An other thing that stands out is that when p increases, the NLS-ϵ tests seem to move toward each other. Overall, for small p the NLS-0.1 test is the second most power full test after the CLRT test, but for $p = 120$ the NLS-0.5 test performs best in terms of power.

(a) $p = 32$

(b) $p = 64$

(c) $p = 96$

(d) $p = 120$
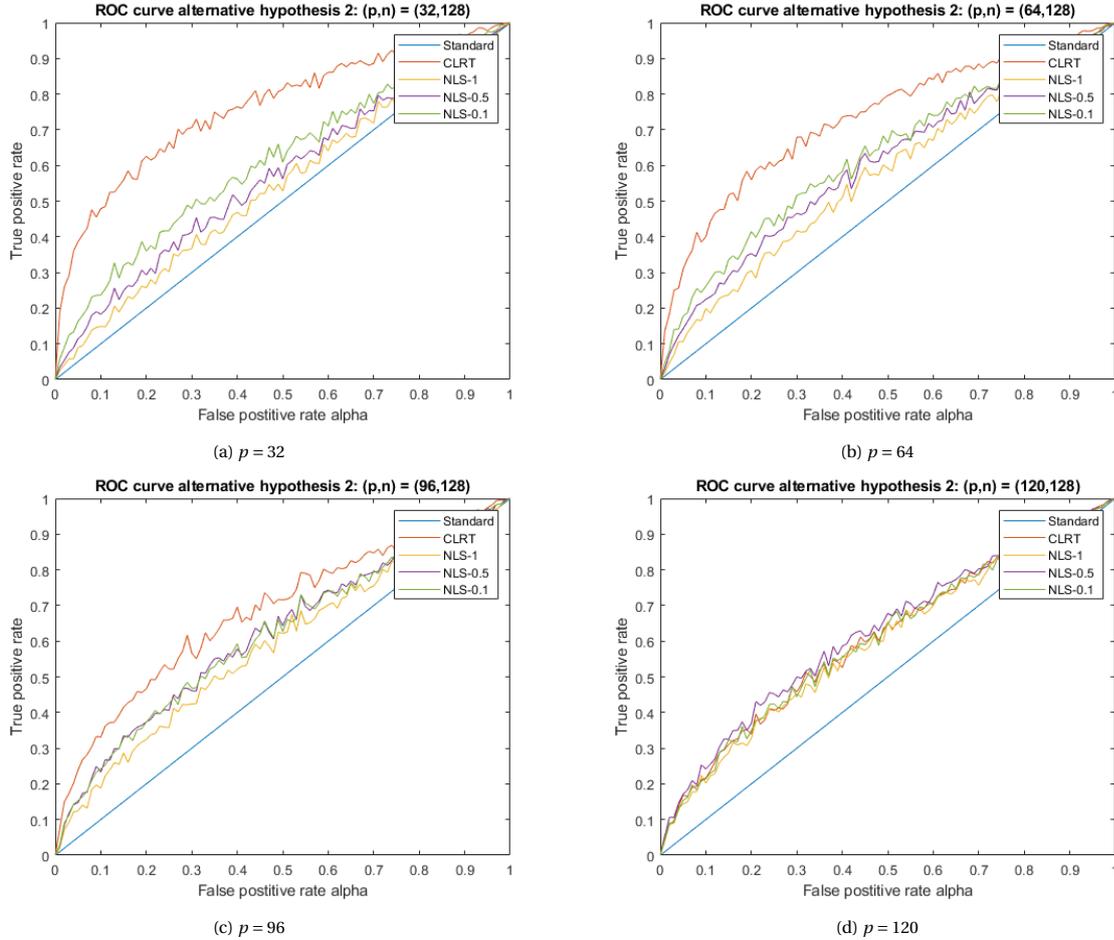
Figure 5.5: ROC plots under alternative hypothesis 2 with 1000 replications and $\delta = 0.12$

5.4.3. ROC Curves for Fixed Ratio With Variance Other Than One

Now it is time to have a look at the ROC curves under the last alternative hypothesis: the fixed ratio with variance other than 1 alternative. In the same way as with the two previous alternative hypotheses, a suitable γ needs to be chosen. In Figure 5.3 it can be seen that for $\gamma = 0.05$ the empirical powers are relatively close for all combinations of (p, n). Therefore, the simulation will be done with $\gamma = 0.05$. The same ϵ's that were used in the power plots for this alternative hypothesis will be chosen again. These are: $\epsilon = 10$, $\epsilon = 1$ and $\epsilon = 0.1$. In this subsection only the ROC plots for $r = 1/2$ will be examined. The ROC curves for $r = 1/4$ and $r = 3/4$ can be found in the Appendix B.2. The ROC curves with $= 0.05$, $r = 1/2$ and 1000 receptions can be found in Figure 5.6.

It can be seen in Figure 5.6 that the NLS-ϵ tests already have high powers in the beginning and that the CLRT does not leave the standard line for this particular γ. This is also what is expected when looking at Figure 5.3 from Section 5.3.3 because it takes a while before the CLRT test gains some power. What is different from observations made for the first and second alternative hypothesis, is that the performance of the NLS-ϵ tests do not depend on p. It can be seen that the NLS-10 test is the best performing test for every p. This is inline with the observation made in Figure 5.3 in section 5.3.3 where the power plots of this alternative where examined.

The ROC curves for $r = 1/4$ and $r = 3/4$ can be found in Figures B.3 and B.4 in Appendix B.2. These ROC curves are similar with the ones found in Figure 5.6. The only thing that is different is that when r increases the NLS-ϵ are more curved outwards. This is again inline with the observations made in Figure 5.3 in section 5.3.3, where the power plots for this alternative are analysed. Furthermore, for the chosen γ the CLRT has not gained any power yet therefore it is close to the standard line.

(a) $p = 32$

(b) $p = 64$

(c) $p = 96$

(d) $p = 120$
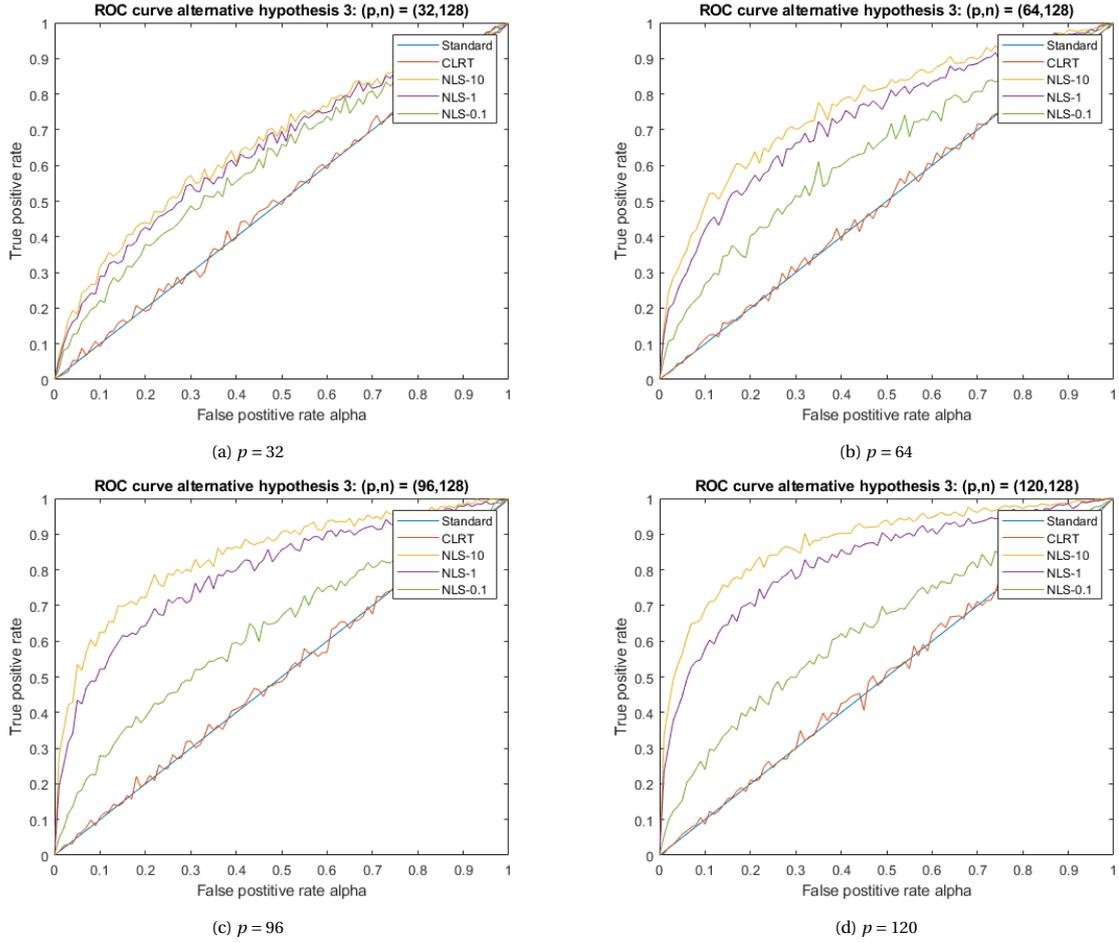
Figure 5.6: ROC plots under alternative hypothesis 3 with 1000 replications, $\gamma = 0.05$ and $r = 1/2$

5.5. Combining Equicorrelation and Variance Other Than One

It is observed in the previous sections that the CJ and LS tests where the best performing test under the equicorrelation alternative but under the fixed ratio other than 1 alternative, the NLS-ϵ test was outperforming the others. This makes it interesting to combine the two alternative hypothesis and examine which combinations of the distance variables (ρ, γ) will lead to the best test. It is expected that the performance of the tests depend on a trade of between the two distance variables since for $(\rho, 0)$ the CJ and LS test are better performing and for $(0, \gamma)$ the NLS-ϵ test is better performing. For any $r \in (0,1)$, $\rho \in (0,1)$ and $\gamma \in \mathbb{R}$, the fourth alternative hypotheses is defined as

$$\Sigma_{n,\rho,\gamma,r} = \Sigma_{n,\gamma,r}^{1/2} \cdot \Sigma_{n,\rho} \cdot \Sigma_{n,\gamma,r}^{1/2}$$

$$= \begin{bmatrix} 1 & 0 & & \cdots & 0 \\ 0 & \ddots & & & \vdots \\ & & 1+\gamma & & \\ \vdots & & & \ddots & \\ 0 & \cdots & & & 1+\gamma \end{bmatrix}^{1/2} \cdot \begin{bmatrix} 1 & \rho & & \cdots & \rho \\ \rho & \ddots & & & \vdots \\ & & 1 & & \\ \vdots & & & \ddots & \\ \rho & \cdots & & & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & & \cdots & 0 \\ 0 & \ddots & & & \vdots \\ & & 1+\gamma & & \\ \vdots & & & \ddots & \\ 0 & \cdots & & & 1+\gamma \end{bmatrix}^{1/2}$$

$$= \begin{bmatrix} 1 & \rho & & \cdots & \rho \\ \rho & \ddots & & & \vdots \\ & & 1+\gamma & & \\ \vdots & & & \ddots & \rho \\ \rho & \cdots & & \rho & 1+\gamma \end{bmatrix}$$

This fourth alternative hypothesis means that for $\rho \neq 0$ there is correlation between the variables in the data and a fixed ratio r of these variables have variance other than one, namely $1 + \gamma$. This fourth alternative depends now on three parameters: ρ, r and γ. Therefore, just as in the previous simulations, ρ will run from 0 to 1, γ from -1 to 1 and $r \in (0, 1)$ 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 alternative hypothesis $\Sigma_{n,\rho,\gamma,r}$ becomes less like the true covariance matrix which is just the identity matrix. Then for every combination of ρ and γ the Bernoulli experiment form Section 5.1 is executed. This will give different empirical powers for each combination of (ρ, γ). Plotting ρ and γ against the obtained empirical powers will then give a 3D power plot to compare the tests. The empirical power will only be calculated for one NLS-ϵ test otherwise the plots will be too full and not useful for analysis. The NLS-ϵ test that will be chosen is the NLS-1 test because this test performed over both alternative hypotheses on average the best. Different from the previous simulations, this simulation will only be done for $p = 64$ and with 100 repetitions because of computational reasons. Furthermore, for demonstrating purposes only the power plots for $r = 1/2$ will be extensively analysed. The power plots for $r = 1/4$, $r = 3/4$ and $r = 1$ can be found in Appendix B.3.
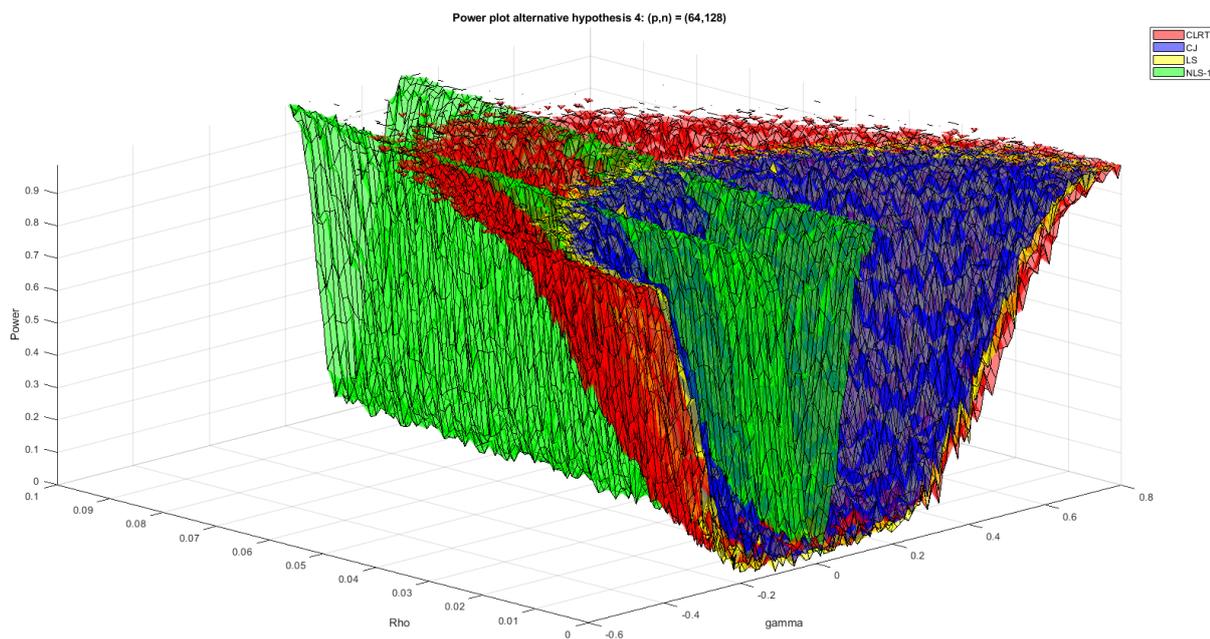
Figure 5.7: 3D power plot under alternative hypothesis 4 with 100 replications, $p = 64$, $\rho \in (0, 1)$, $\gamma \in (-1, 1)$ and $r = 1/2$

In Figure 5.7 the empirical powers for the fourth alternative hypothesis are plotted. One thing that immediately stands out is that the NLS-1 test does not seem to gain any power when γ is close the zero and ρ runs form 0 to 1. Moreover, from this figure is is not immediately clear whether the empirical powers behave just as in Figure 5.1(c) in Subsection 5.3.1 and Figure 5.3(c) in Subsection 5.3.3: the empirical powers when one of the distance variables ρ or γ is equal to zero. Therefore in Figure 5.8(a) the empirical power is plotted for $\gamma \in (0, 0.01)$ and in Figure 5.8(b) for $\rho \in (0, 0.01)$, thus for ρ and γ small. In can be seen in these figures that the tests behave just as expected. The only difference is that the lines are changed for planes.

(a) $\gamma \in (0, 0.01)$

(b) $\rho \in (0, 0.01)$
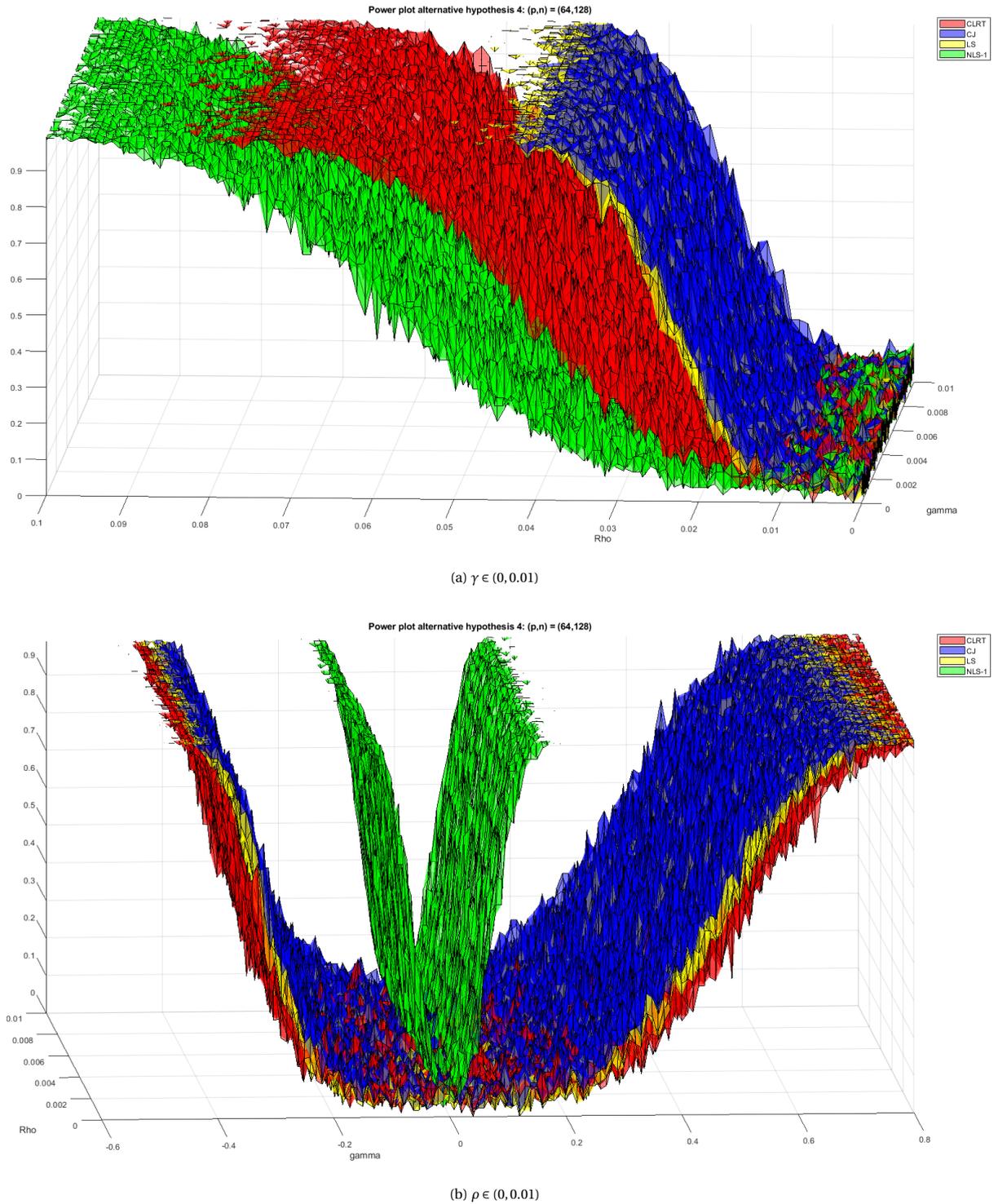
Figure 5.8: 3D power plot under alternative hypothesis 4 with 100 replications, $p = 64$ and $r = 1/2$

Now lets investigate what happens when the distance variables ρ and γ are not close to zero. Thus when the distance from the null hypothesis increases in both directions. For the CJ, LS and CLRT tests no strange things are happening there. The empirical powers of these test increase just as one would expected based on the observations made in in Figure 5.1(c) of Subsection 5.3.1 and Figure 5.3(c) of Subsection 5.3.3. However, the NLS-1 test behaves strange. The power does not increase in the ρ direction when γ moves a way from zero. Therefore, lets focus now only on the NLS-1 test. In Figure 5.9(a) the empirical power of the NLS-1 test

is plotted for both ρ and γ small. From this figure it can be seen again that when ρ or γ is equal to zero the test behaves normal but when ρ and γ are lying in a particular region, the empirical power stays very low. This "no power phenomenon" is better displayed in Figure 5.9(b), where the empirical power is plotted between 0.1 and 0.9. From this figure 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.
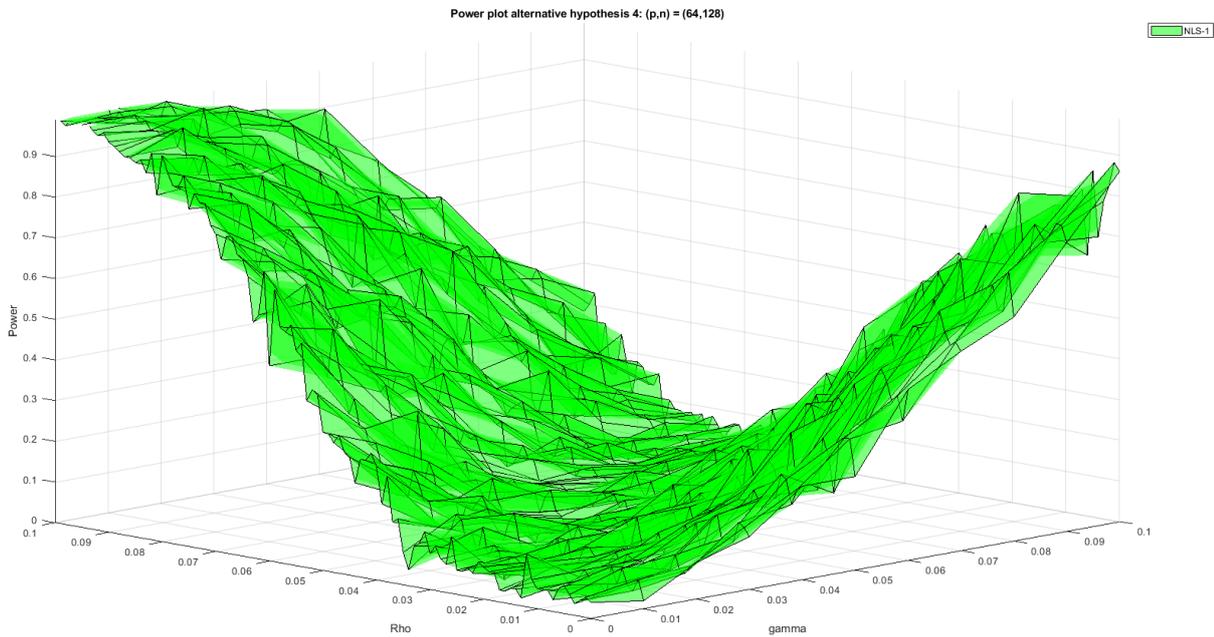
(a) $\rho \in (0,0.1)$ and $\gamma \in (0,0.1)$
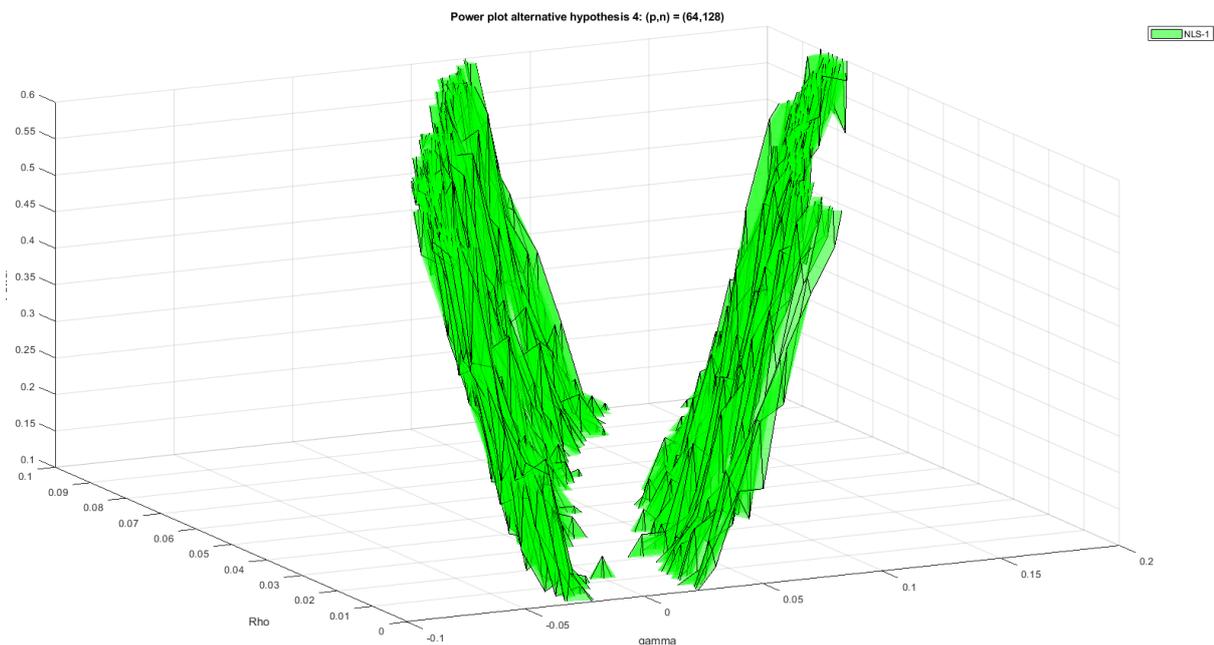
(b) $\rho \in (0,0.1)$, $\gamma \in (-0.1,0.2)$ and empirical power $\in (0.1,0.9)$

Figure 5.9: 3D power plot of the NLS-1 test under alternative hypothesis 4 with 100 replications, $p = 64$ and $r = 1/2$

To find out why this no power phenomenon for the NLS-1 test is happening it is useful to examine the eigenvalues of the sample covariance matrix calculated from the stochastic model $Y = \Sigma_{\rho,r,\gamma}^{1/2} X$ of this simulation. This is because the eigenvalues of a sample covariance matrix contain a lot of information, so when they behave out of ordinary, this could be fatal for a test. For example, when there are a lot of eigenvalues very close to zero then no information is left and this could cause problems. Or on the other hand, some test might have

problems when there are eigenvalues which are very large. These very large eigenvalues are called spikes. As mentioned earlier, when looking at 5.9(b) the no power phenomenon for the NLS-1 test only occurs when γ is approximately equal to $\frac{3}{2} \cdot \rho$ this line will be referred to as the no power line. A 3D histogram of the eigenvalues is made in the ρ direction with $\gamma = 0$ and 3D histogram in the γ direction for $\rho = 0$. This, to investigate how the "normal" situation looks like. These 3D histograms can be found in Figures 5.10(a) and 5.10(b). In addition, to investigate the "no power phenomenon", a 3D histogram is made in the ρ direction with $\gamma = 0.1$ and a 3D histogram in the γ direction with $\rho = 0.05$. These ρ and γ are chosen such that the plots cross the no power line. These last two 3D histograms can be found in repressively Figures 5.11(a) and 5.11(b). All the 3D histograms of the eigenvalues of the sample covariance matrix are plotted using 1000 replications of the stochastic model $Y = \Sigma_{\rho,r,\gamma}^{1/2} X$ for the parameters mentioned earlier.
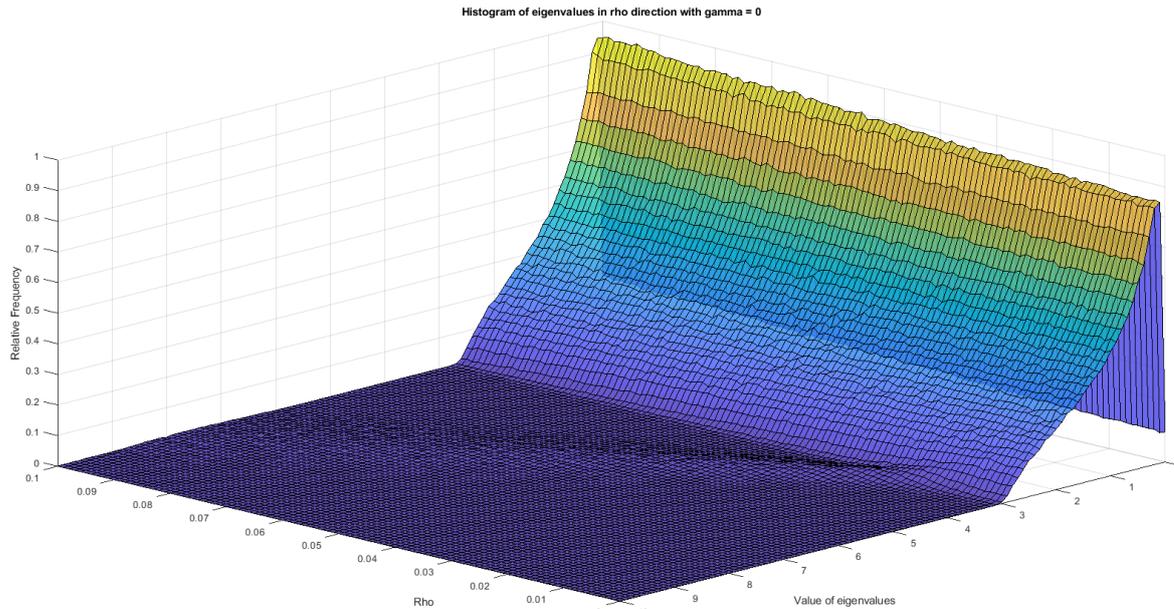
(a) In ρ direction with $\gamma = 0$
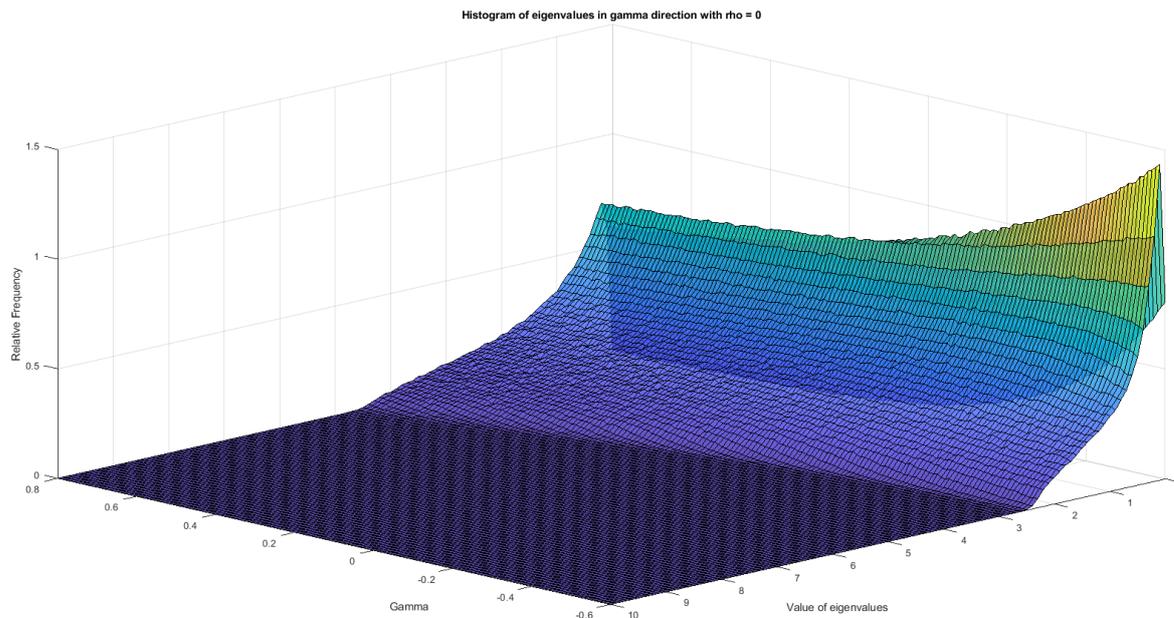
(b) In γ direction with $\rho = 0$

Figure 5.10: 3D histogram plot of the eigenvalues of the sample covariance matrix with 1000 replications

In Figure 5.10(a) it can be seen that when ρ increases from 0 to 1, from $\rho = 0.02$ onwards, there are starting to arise eigenvalues which are much larger than they used to be. Moreover, right from the beginning the number of small eigenvalues increases as well, only this increase is very small. Then in Figure 5.10(b), which is the histogram in the γ direction for $\rho = 0$, it can be seen that when γ decreases there becoming more and more small eigenvalues. It looks like the eigenvalues are all shifted towards zero in a rapid tempo. The eigenvalues in Figures 5.10(a) and 5.10(b) are calculated in the "normal" situation. From this normal situation it is known that all the tests behave properly. Now lets have a look if there is a change when the distance variables ρ and γ cross the no power line.

(a) In ρ direction with $\gamma = 0.1$

(b) In γ direction with $\rho = 0.05$
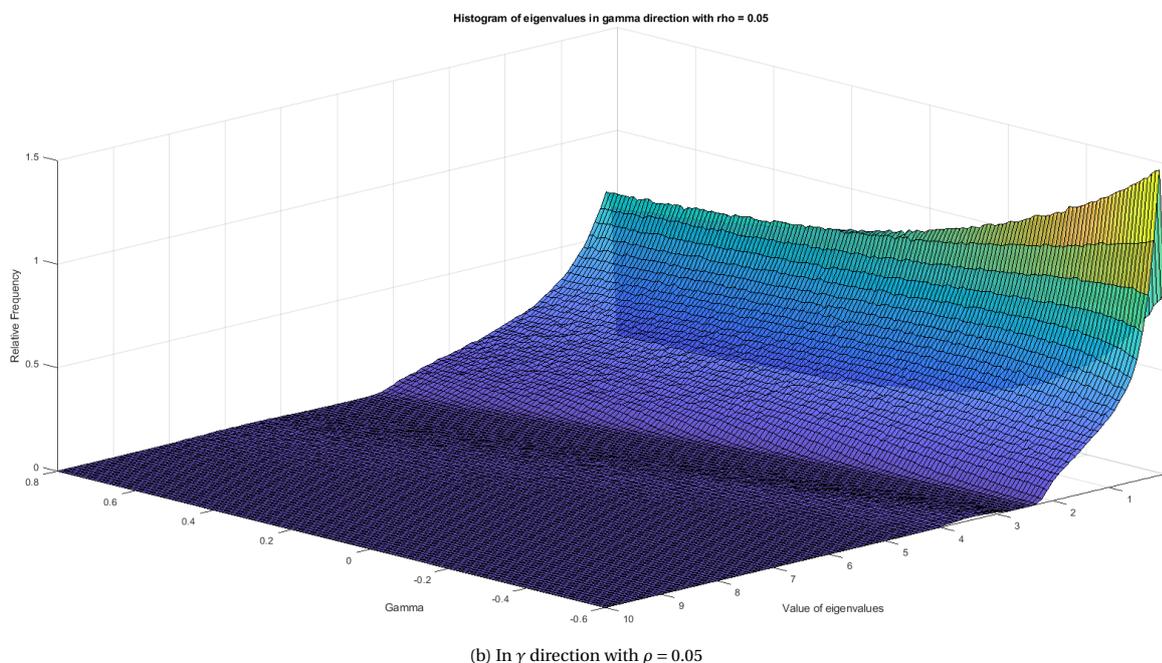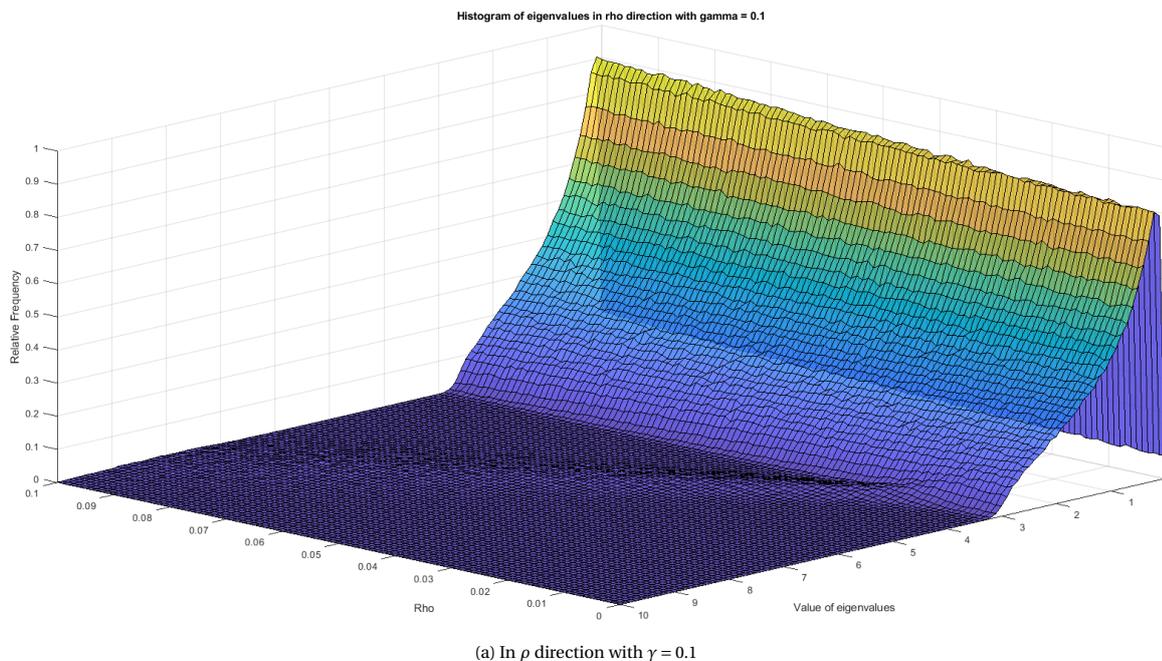
Figure 5.11: 3D histogram plot of the eigenvalues of the sample covariance matrix with 1000 replications

In Figure 5.11(a) and 5.11(b) the eigenvalues are calculated when the distance variables ρ and γ cross the no power line. This happens when $\gamma = 0.1$ and ρ increases from 0 to 1 and for $\rho = 0.05$ and γ increases from -0.4 to 0.8. These 3D histograms look very similar to the ones in the normal situation. The only difference that can be detected is that in Figure 5.11(b) there is a small line with spiked eigenvalues. This line with spiked eigenvalues is very small but it is there. Unfortunately, from these observations it is hard to conclude what exactly causes the no power phenomenon. This means there are still some unanswered questions left: could spiked eigenvalues cause the no power phenomenon? or is it due to some very small eigenvalues after all? or maybe because of something else what is not discussed yet?

5.6. Concluding The Simulation Study

In this section the simulation study will be reviewed and some conclusions will be made about the performance of different tests. The simulation study started with the empirical size comparison. In this section it is observed that the empirical sizes for all the tests are close to the rejection level α when the data is based on standard normal realisation. This means that in can be concluded that the statistics are close tho their limiting distributions for the given parameters. Unfortunately, when the data is based on the $Gamma(4,2) - 2$ distribution this is not the case. This is because the empirical sizes for CLRT are way too small compared to the rejection level α. For the CJ test this is the other way around. They start way above the rejection level α but come closer when p increases. The NLS-ϵ and LS tests do the best job in terms of empirical sizes. Also for data based on the $Gamma(4,2) - 2$ distribution they are close to the rejection level α. So from this observations it is concluded that the the statistics of the NLS-ϵ, LS tests are close to their limiting distributions for all combinations of p and n and the CJ test only when p gets closer to n when the data is based on the $Gamma(4,2) - 2$ distribution.

The second indicator that is used to compare the tests' performance is their power. This is done using power plots and ROC curves based on three different alternative hypothesis: the equicorrelation alternative, the autoregressive relation alternative and the fixed ratio with variance other than 1 alternative. It is observed that for the first two alternative hypothesis the LS and CJ tests performed the best and that the NLS-ϵ test performs the worse. Only when p is sufficiently large to compared to n it outperforms the CLRT. This is because of two reasons: the first is that the CLRT test breaks down when c increases to 1, the second reason is that the NLS-ϵ test performs just better when p increases to n. Now zooming in on the NLS-ϵ test: $\epsilon = 0.5$ works best for the equicorrelation alternative and $\epsilon = 0.1$ works best for the autoregressive relation alternative. However, when p gets close to n the the optimal ϵ increases a little. This could to compensate for some numerical issues when p gets close to n. For the third alternative something different is happening. For this alternative the roles have been reversed and the NLS-ϵ test performs by far the best. The power of the NLS-ϵ test increases even more when a larger ϵ is chosen. This is different than for the first two alternatives where the optimal ϵ is dependent on p. Moreover, when the ratio r increases the NLS-ϵ tests also increase in power. The CJ, LS and CLRT behave very different. First of all, these test are not symmetric around $\gamma = 0$, where the NLS-ϵ tests are. They increase much faster in power for negative values of γ than for positive. secondly, the CJ, LS and CLRT test only increase in power when r increases to $1/2$ but when $r > 1/2$ they decrease again. This is due to the fact that the CJ, LS and CLRT tests are invariant under multiples of the identity matrix where the NLS-ϵ test is not.

The simulation study is closed with a power comparison of the the first and third alternative hypotheses combined. This is interesting because the for the first alternative the LS and CJ test are performing best and for the third alternative the NLS-ϵ test is performing best. The empirical powers of CJ, LS and CLRT tests are performing as one would expect when considering the empirical powers of the alternative hypotheses separate but the NLS-ϵ test not. This test does not gain power in the ρ direction when $\gamma = \frac{3}{2}\rho$, which is referred to as the no power line. To investigate what causes this no power phenomenon the eigenvalues of the sample covariance matrix are examined for different combinations of the distance variables ρ and γ. Some spiked eigenvalues as well as some very small eigenvalues are detected. Unfortunately, from these observations nothing could be concluded yet. Further research should be conducted to provide an answer for what causes the no power phenomenon.

6

Conclusion

In this thesis a new sphericity test in the large-dimensional framework is constructed which is based on the nonlinear shrinkage estimator derived by Ledoit and Wolf (2012). It has been pointed out by Ledoit and Wolf (2012) that the nonlinear shrinkage estimator is in many cases a better estimator for the true covariance matrix than for example the linear shrinkage estimator. Now since a sphericity test, in fact tests whether the true covariance matrix is equal or not to the identity matrix, the nonlinear shrinkage estimator it is a good starting point to construct a sphericity test from. The nonlinear shrinkage function which defines the nonlinear shrinkage estimator could not be used directly in a linear spectral statistic. Its form under the null hypothesis $H_0 : \Sigma_n = I$ had to be found and it needed to be altered such that it gives a non degenerate statistic. This has resulted in two linear spectral statistics, which are Equation (4.3) and (4.4), for which the central limit theorem for linear spectral statistics could be applied to.

For only the linear spectral statistics from Equation (4.3) the limiting distribution is derived. This led to one of the main results of this thesis, which is Theorem 4.3.1. This theorem states that the fluctuations of the linear spectral statistic from Equation (4.3) are normally distributed. This theorem led to the introduction of a new sphericity test: the NLS-ϵ test. The test statistic corresponding to this new test is Equation (4.7). The NLS-ϵ test is actually the more general version of the Bartlett-Nanda-Pillai (BNP) trace test originally proposed by Pillai (1955). This is because instead of only using $\epsilon = 1$ in case of the BNP trace test, every $\epsilon > 0$ could be used in the NLS-ϵ test. The construction of the NLS-ϵ test concluded the theoretical part of this thesis because unfortunately optimising the NLS-ϵ test with respect to ϵ under a general alternative hypotheses is outside the scoop a bachelor thesis.

In the last chapter a simulation study is carried out to compare the corrected John's test (CJ), the linear shrinkage test (LS), the corrected likelihood ratio test (CLRT) and the the nonlinear shrinkage test for different $\epsilon > 0$ (NLS-ϵ) in terms of size and power. Moreover, the simulation study is used to find the optimal values for ϵ for which the NLS-ϵ test has the highest power. To asses the different tests the empirical size, empirical power and receiver operating characteristic (ROC) curves are used. From the comparison made using the empirical size it is concluded that for standard normal data the statistics are all close to their limiting distribution but for data taken from the $Gamma(4,2) - 2$ distribution the empirical sizes of the CLRT and the CJ test are repressively too small and too big compared to the prespecified rejection level $\alpha = 0.05$.

To compare the powers, 3 alternative hypothesis are used. These are: the equicorrelation alternative, the autoregressive alternative and the fixed ratio variance other than one. From the power comparison it is concluded that the CJ and LS tests are performing the best for the equicorrelation and autoregrssive alternative but that the NLS-ϵ performs best for the fixed ratio variance other than one alternative for every ratio r. For this third alternative the power of the NLS-ϵ test could be increased even more when one increases ϵ from one onwards. In conclusion, the new NLS-ϵ is most appropriate for testing a fixed ratio variance is other than one alternative or an all variance are other than one alternative. For these alternatives the NLS-ϵ has by far the highest power compared to other test considered in this thesis.

The simulation study has ended with a power comparison of the equicorrelation and fixed ratio variance other than one alternatives combined. Comparing the powers for this fourth and last alternative is interesting because for each alternative an other test comes out best. The CJ, LS and CLRT behave as one would expect considering their behaviour for each alternative separate. However, this is not the case for the NLS-ϵ test. It is observed that for when the distance variable γ is close but not equal to zero, the NLS-ϵ test does not gain any power in the ρ direction. This phenomenon is referred to as the no power phenomenon and is still inexplicable.

7

Discussion

During the theoretical part as well as during the simulations study some assumptions has been simplified and some research questions could not be answered. These will be discussed in this section. First the simplifications will be considered where after the unanswered research questions will be discussed and some recommendations for further research will be given. The first simplification that has been made is in the case of ratio between the sample size n and the dimension p in the large-dimensional asymptotic framework. This is because the case when the dimension p is larger than the sample size n, that is when $\frac{p}{n} = c > 1$, has not been considered during the derivations. The reason for this was to simplify certain mathematical results as well for computational reasons during the simulation study. As a result also for the the construction of the nonlinear shrinkage test (NLS-ϵ), because this test is initially constructed for $c \in (0, 1)$. However, during the simulation study it turned out that the NLS-ϵ also holds whenever $c \geq 1$. The reason for this is that the central limit theorem from Theorem 4.1.1, that was used to find the limiting distribution of the NLS-ϵ test statistic, holds for all $c > 0$. Moreover, for $c \geq 1$ the poles in the proof of Theorem 4.3.1 have stayed the same and did not switched. As a result, the NLS-ϵ also works for all $c > 0$.

The second simplification that has been made is that the empirical power comparison in the simulation study is only based on standard normal distributed data. The case when the data is coming from a $Gamma(4, 2) - 2$ distribution has not been considered. The reason for this is that the empirical sizes, in case the data comes from a $Gamma(4, 2) - 2$ distribution, are not all close together. Therefore, it is difficult to make a fair comparison. However, for the power comparison using receiver operating characteristic (ROC) curves this does not matter. Since not the distance variable but the rejection level α varies in these curves, all the powers start at zero. This means that a fair comparison can be made using ROC curves. An other thing that should be noted is that the empirical sizes of the linear shrinkage test (LS) based on standard normal data, and the corrected likelihood ratio test (CLRT), the corrected John's test (CJ) and the linear shrinkage test (LS) based on $Gamma(4, 2) - 2$ data, deviate from the once found by Versteegh (2020). The reason for this is not really clear but it could be lying in the fact that all one-tailed hypothesis test are transformed into two-tailed test and the CJ and LS test are one of them.

Recommendations for Further Research

The simulation study closes with the general question: what causes the no power phenomenon of the NLS-ϵ test under the equicorrelation and fixed ratio variance other than one alternative? Multiple suggestion have been given such as, because of spiked eigenvalues or because of a large number of very small eigenvalues of the sample covariance matrix. Unfortunately, based on the observations made in section 5.6 no answer could be given. The no power phenomenon was initially not expected because the NLS-ϵ test works fine for the equicorrelation alternative. Therefore, because this phenomenon so unexpected, finding the reason why could be an interesting for further research. Not only because it provides an answer to the unsolved problem but also because it sheds some light on the structure of large-dimensional covariance matrices, which are very important in multivariate analysis. Moreover, it could give inside in what kind of structures of a linear spectral statistic one should avoid in constructing a multivariate statistical test in the large-dimensional framework.

The final question that has not been answered is why the NLS-ϵ test has so much power under the fixed variance other than one alternative. This question has not been posed during the simulation study but is very interesting for two reasons. The first reason is that the NLS-ϵ test outperforms the CJ test, which is a well established and powerful test. The second reason is that it will provide some deep understanding in the structures of linear spectral statistics and again give more inside in the structures of large-dimensional covariance matrices.

A

Appendix A: Proofs

A.1. Proof Limiting Distribution T_1

In this appendix Theorem 4.3.1 will be proved. That is the limiting distribution of the linear spectral statistic $T_1 = \sum_{i=1}^{p} \frac{\lambda_i}{\lambda_i + \epsilon}$ for $\epsilon > 0$. Before starting this proof, a handy remark is presented:

Remark. *let $z \in \mathbb{C}$ run over the unit circle counterclockwise once, with complex conjugate $\bar{z} = \frac{1}{z}$. Then*

$$
\begin{aligned}
|1 + hz|^2 &= (1 + hz)(\overline{1 + hz}) \\
&= (1 + hz)(1 + \frac{h}{z}) \\
&= 1 + \frac{h}{z} + hz + h^2 \\
&= (z + h)(\frac{1}{z} + h) \\
&= \frac{1}{z}(z + h)(1 + zh)
\end{aligned}
$$

Now it is possible to start the proof of theorem 4.3.1.

Proof. Consider the linear spectral statistic $T_1 = \sum_{i=1}^{p} \frac{\lambda_i}{\lambda_i + \epsilon}$ for $\epsilon > 0$ and with $\varphi(\lambda) = \frac{\lambda}{\lambda + \epsilon}$. We will derive its limiting distribution. Assume that the variables $\{x_{i,j}\}$ of the matrix $X = (\mathbf{x_1}, \dots, \mathbf{x_n})$ are independent and identically distributed satisfying $\mathbb{E}[x_{i,j}] = 0$, $\mathbb{E}[|x_{i,j}|^2] = 1$, $\mathbb{E}[|x_{i,j}|^4] = \beta + 1 + \kappa < \infty$, and in case of complex variables, $\mathbb{E}[x_{i,j}^2] = 0$. Assume, moreover,

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

By theorem 4.1.1, which is the central limit theorem for linear spectral statistics, we know that under $H_0 : \Sigma_n = I$,

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

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

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

using proposition 4.1.1 we can derive the limiting parameters.

Mean: μ

First of all we start by calculating the limiting parameters $I_1(\varphi)$ and $I_2(\varphi)$ in the mean. Let $\gamma := |z| = 1$ and $\xi = z$. By Proposition 4.1.1 we have that

$$
\begin{aligned}
I_1(\varphi) = \lim_{r\downarrow 1} I_1(\varphi, r) &= \lim_{r\downarrow 1} \frac{1}{2\pi i} \oint_\gamma f(|1+hz|^2) \left[\frac{z}{z^2 - r^{-2}} - \frac{1}{z} \right] dz \\
&= \lim_{r\downarrow 1} \frac{1}{2\pi i} \oint_\gamma \frac{|1+hz|^2}{|1+hz|^2 + \epsilon} \left[\frac{z}{z^2 - r^{-2}} - \frac{1}{z} \right] dz \\
&= \lim_{r\downarrow 1} \frac{1}{2\pi i} \oint_\gamma \frac{|1+hz|^2}{|1+hz|^2 + \epsilon} \frac{z}{z^2 - r^{-2}} dz - \lim_{r\downarrow 1} \frac{1}{2\pi i} \oint_\gamma \frac{|1+hz|^2}{|1+hz|^2 + \epsilon} \frac{1}{z} dz \\
&= W - F
\end{aligned}
$$

Then

$$
\begin{aligned}
W = \lim_{r\downarrow 1} \frac{1}{2\pi i} \oint_\gamma \frac{|1+hz|^2}{|1+hz|^2 + \epsilon} \frac{z}{z^2 - r^{-2}} dz &= \lim_{r\downarrow 1} \frac{1}{2\pi i} \oint_\gamma \frac{\frac{1}{z}(z+h)(1+zh)}{\frac{1}{z}(z+h)(1+zh) + \epsilon} \frac{z}{z^2 - r^{-2}} dz \\
= \lim_{r\downarrow 1} \frac{1}{2\pi i} \oint_\gamma \frac{(z+h)(1+zh)}{(\frac{1}{z}(z+h)(1+zh) + \epsilon)(z^2 - r^{-2})} dz &= \lim_{r\downarrow 1} \frac{1}{2\pi i} \oint_\gamma \frac{z(z+h)(1+zh)}{((z+h)(1+zh) + z\epsilon)(z^2 - r^{-2})} dz \\
= \lim_{r\downarrow 1} \frac{1}{2\pi i} \oint_\gamma \frac{z(z+h)(1+zh)}{h(z-A)(z-B)(z-\frac{1}{r})(z+\frac{1}{r})} dz &
\end{aligned}
$$

where

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

For notation reasons the above defined A and B will be used trough out this proof. For $c \in (0, 1)$ and for $\epsilon > 0$ we find that A lies inside γ and B outside γ also since $r > 1$ we have that $z = \pm \frac{1}{r}$ lie inside γ. Therefore, the function inside the contour integral has 3 simple poles inside γ, $z = A$, $z = \frac{1}{r}$ and $-\frac{1}{r}$. To calculate this contour integral we need to calculate the residues of these poles.

$$
\begin{aligned}
Res(A) = \lim_{z\to A} (z - A) \frac{z(z+h)(1+zh)}{h(z-A)(z-B)(z-\frac{1}{r})(z+\frac{1}{r})} &= \lim_{z\to A} \frac{z(z+h)(1+zh)}{h(z-B)(z-\frac{1}{r})(z+\frac{1}{r})} \\
&= \frac{A(A+h)(1+Ah)}{h(A-B)(A-\frac{1}{r})(A+\frac{1}{r})} \\
Res(1/r) = \lim_{z\to \frac{1}{r}} (z - \frac{1}{r}) \frac{z(z+h)(1+zh)}{h(z-A)(z-B)(z-\frac{1}{r})(z+\frac{1}{r})} &= \lim_{z\to \frac{1}{r}} \frac{z(z+h)(1+zh)}{h(z-A)(z-B)(z+\frac{1}{r})} \\
&= \frac{(hr+1)(r+h)}{2h(Ar-1)(Br-1)} \\
Res(-1/r) = \lim_{z\to -\frac{1}{r}} (z + \frac{1}{r}) \frac{z(z+h)(1+zh)}{h(z-A)(z-B)(z-\frac{1}{r})(z+\frac{1}{r})} &= \lim_{z\to -\frac{1}{r}} \frac{z(z+h)(1+zh)}{h(z-A)(z-B)(z-\frac{1}{r})} \\
&= -\frac{(hr-1)(-r+h)}{2h(Ar+1)(Br+1)}
\end{aligned}
$$

Then by Cauchy's Residue Theorem we find that

$$
\begin{aligned}
W &= \lim_{r\downarrow 1} \frac{1}{2\pi i} \left[2\pi i (Res(A) + Res(1/r) + Res(-1/r)) \right] \\
&= \lim_{r\downarrow 1} \left[\frac{A(A+h)(1+Ah)}{h(A-B)(A-\frac{1}{r})(A+\frac{1}{r})} + \frac{(hr+1)(r+h)}{2h(Ar-1)(Br-1)} - \frac{(hr-1)(-r+h)}{2h(Ar+1)(Br+1)} \right] \\
&= \frac{(Ah + h^2 + 1)B^2 + 2Bh - Ah}{h(A-B)(B^2 - 1)}
\end{aligned}
$$

Now lets continue with the second integral F

$$F = \lim_{r \downarrow 1} \frac{1}{2\pi i} \oint_\gamma \frac{|1+hz|^2}{|1+hz|^2+\epsilon} \frac{1}{z} dz = \lim_{r \downarrow 1} \frac{1}{2\pi i} \oint_\gamma \frac{\frac{1}{z}(z+h)(1+zh)}{\frac{1}{z}(z+h)(1+zh)+\epsilon} \frac{1}{z} dz$$

$$= \lim_{r \downarrow 1} \frac{1}{2\pi i} \oint_\gamma \frac{(z+h)(1+zh)}{((z+h)(1+zh)+z\epsilon)z} dz = \lim_{r \downarrow 1} \frac{1}{2\pi i} \oint_\gamma \frac{(z+h)(1+zh)}{h(z-A)(z-B)z} dz$$

The function inside the above contour integral has two simple poles, $z = A$ and $z = 0$. To calculate F, we need to calculate the residues.

$$Res(A) = \lim_{z \to A} (z-A) \frac{(z+h)(1+zh)}{h(z-A)(z-B)z} = \lim_{z \to A} \frac{(z+h)(1+zh)}{h(z-A)(z-B)z}$$

$$= \frac{(A+h)(1+Ah)}{h(A-B)A}$$

$$Res(0) = \lim_{z \to 0}(z-0) \frac{(z+h)(1+zh)}{h(z-A)(z-B)z} = \lim_{z \to 0} \frac{(z+h)(1+zh)}{h(z-A)(z-B)}$$

$$= \frac{1}{AB}$$

Then by Cauchy's Residue Theorem we find that

$$F = \lim_{r \downarrow 1} \frac{1}{2\pi i} \left[2\pi i (Res(A) + Res(0)) \right]$$

$$= \frac{(A+h)(1+Ah)}{h(A-B)A} + \frac{1}{AB}$$

Then combining the above results gives $I_1(\varphi)$

$$I_1(\varphi) = W - F = \frac{(Ah+h^2+1)B^2+2Bh-Ah}{h(A-B)(B^2-1)} - \frac{(A+h)(1+Ah)}{h(A-B)A} - + \frac{1}{AB}$$

$$= \frac{(h+B)(Bh+1)}{h(A-B)(B^2-1)B}$$

Now we will calculate the second limiting parameter $I_2(\varphi)$. By proposition 4.1.1 this is given by

$$I_2(\varphi) = \frac{1}{2\pi i} \oint_\gamma f(|1+hz|^2) \frac{1}{z^3} dz = \frac{1}{2\pi i} \oint_\gamma \frac{|1+hz|^2}{|1+hz|^2+\epsilon} \frac{1}{z^3} dz$$

$$= \frac{1}{2\pi i} \oint_\gamma \frac{\frac{1}{z}(z+h)(1+zh)}{\frac{1}{z}(z+h)(1+zh)+\epsilon} \frac{1}{z^3} dz = \frac{1}{2\pi i} \oint_\gamma \frac{(z+h)(1+zh)}{((z+h)(1+zh)+z\epsilon)z^3} dz$$

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

The function inside the contour integral has one simple pole $z = A$ and one pole $z = 0$ of order 3. To calculate the this contour integral we need to calculate the residues of these poles.

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

$$= \frac{(A+h)(1+Ah)}{h(A-B)A^3}$$

$$Res(0) = \frac{1}{2} \lim_{z \to 0} \frac{d^2}{dz^2} (z-0)^3 \frac{(z+h)(1+zh)}{h(z-A)(z-B)z^3} = \frac{1}{2} \lim_{z \to 0} \frac{d^2}{dz^2} \frac{(z+h)(1+zh)}{h(z-A)(z-B)}$$

$$= \frac{(h+B)(Bh+1)A^2 + ((h^2+1)B+h)BA + B^2h}{A^3B^3h}$$

Therefore by Cauchy's Residue Theorem we find that

$$
\begin{aligned}
I_2(\varphi) &= \frac{1}{2\pi i}\left[2\pi i\,(Res(A) + Res(0))\right] \\
&= \frac{(A+h)(1+Ah)}{h(A-B)A^3} + \frac{(h+B)(Bh+1)A^2 + ((h^2+1)B+h)BA + B^2h}{A^3 B^3 h} \\
&= \frac{(h+B)(Bh+1)}{B^3 h(A-B)}
\end{aligned}
$$

Combining $I_1(\varphi)$ and $I_2(\varphi)$ and letting $h = \sqrt{c}$, we find the mean

$$
\begin{aligned}
\mu &= (\kappa - 1)I_1(\varphi) + \beta I_2(\varphi) \\
&= (\kappa - 1)\left[\frac{(\sqrt{c}+B)(B\sqrt{c}+1)}{\sqrt{c}(A-B)(B^2-1)B}\right] + \beta\left[\frac{(\sqrt{c}+B)(B\sqrt{c}+1)}{B^3\sqrt{c}(A-B)}\right]
\end{aligned}
$$

Variance: σ^2

Secondly we compute the limiting parameters $J_1(\varphi,\varphi)$ and $J_2(\varphi,\varphi)$ in the variance. Let $\xi = z$, $\gamma_1 := |z_1| = 1$, $\gamma_2 := |z_2| = 1$ and $\varphi_1 = \varphi_2$ (because are computing the variance not the covariance). Moreover, to compute $J_1(\varphi,\varphi)$ we assume that we can change the order of integration. Then by Proposition 4.1.1 we have that

$$
\begin{aligned}
J_1(\varphi_1,\varphi_2) = J_1(\varphi,\varphi) = \lim_{r\downarrow 1} J_1(\varphi,\varphi,r) &= \lim_{r\downarrow 1} -\frac{1}{4\pi^2}\oint_{\gamma_1}\oint_{\gamma_2}\frac{\varphi(|1+hz_1|^2)\varphi(|1+hz_2|^2)}{(z_1 - rz_2)^2}dz_1\,dz_2 \\
&= \lim_{r\downarrow 1} J_1(\varphi,\varphi,r) = \lim_{r\downarrow 1} -\frac{1}{4\pi^2}\oint_{\gamma_1}\oint_{\gamma_2}\frac{\varphi(|1+hz_1|^2)\varphi(|1+hz_2|^2)}{(z_1 - rz_2)^2}dz_2\,dz_1 \\
&= \lim_{r\downarrow 1}\frac{1}{2\pi i}\oint_{\gamma_1}\varphi(|1+hz_1|^2)\frac{1}{2\pi i}\oint_{\gamma_2}\frac{\varphi(|1+hz_2|^2)}{(z_1 - rz_2)^2}dz_2\,dz_1 \\
&= \lim_{r\downarrow 1}\frac{1}{2\pi i}\oint_{\gamma_1}\varphi(|1+hz_1|^2)\frac{1}{2\pi i}\oint_{\gamma_2}\frac{|1+hz_2|^2}{|1+h_2|^2+\epsilon}\frac{1}{(z_1-rz_2)^2}dz_2\,dz_1 \\
&= \lim_{r\downarrow 1}\frac{1}{2\pi i}\oint_{\gamma_1}\varphi(|1+hz_1|^2)\frac{1}{2\pi i}\oint_{\gamma_2}\frac{\frac{1}{z_2}(z_2+h)(1+z_2h)}{\frac{1}{z_2}(z_2+h)(1+z_2h)+\epsilon}\frac{1}{(z_1-rz_2)^2}dz_2\,dz_1 \\
&= \lim_{r\downarrow 1}\frac{1}{2\pi i}\oint_{\gamma_1}\varphi(|1+hz_1|^2)\frac{1}{2\pi i}\oint_{\gamma_2}\frac{(z_2+h)(1+z_2h)}{((z_2+h)(1+z_2h)+z_2\epsilon)(z_1-rz_2)^2}dz_2\,dz_1 \\
&= \lim_{r\downarrow 1}\frac{1}{2\pi i}\oint_{\gamma_1}\varphi(|1+hz_1|^2)\frac{1}{2\pi i}\oint_{\gamma_2}\frac{(z_2+h)(1+z_2h)}{h(z_2-A)(z_2-B)(z_1-rz_2)^2}dz_2\,dz_1
\end{aligned}
$$

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. Because for fixed $|z_1| = 1$ and $r > 1$, $\frac{z_1}{r}$ lies inside γ_2. To calculate the inner contour integral, we need to calculate the residues of the poles.

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

By Cauchy's Residue Theorem we find that

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

We will calculate the three contour integrals separately. Lets calculate the first integral F.

$$
\begin{aligned}
F &= \frac{1}{2\pi i}\oint_{\gamma_1}\frac{(z_1+h)(1+z_1h)(A+h)(1+Ah)}{h^2(z_1-A)(z_1-B)(A-B)(z_1-rA)^2}dz_1\\
&= \frac{(A+h)(1+Ah)}{h^2(A-B)}\frac{1}{2\pi i}\oint_{\gamma_1}\frac{(z_1+h)(1+z_1h)}{(z_1-A)(z_1-B)(z_1-rA)^2}dz_1
\end{aligned}
$$

The function inside the contour integral has 2 poles inside $|z_1|=1$, a simple pole $z_1=A$ and a pole of order 2, $z_1=rA$. To calculate the integral we need to calculate the residues of these poles.

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

Then by Cauchy's residue theorem we have that the first integral is equal to

$$
\begin{aligned}
F &= \frac{(A+h)(1+Ah)}{h^2(A-B)}\frac{1}{2\pi i}[2\pi i(Res(A)+Res(Ar))]\\
&= \left(\frac{(A+h)(1+Ah)}{h(A-B)(A-rA)}\right)^2\\
&\quad +\frac{(A+h)(1+Ah)}{h^2(A-B)}\cdot\frac{(rA-A)(rA-B)(h^2+1+2hrA)-rA+h)(1+rAh)(2rA-A-B)}{(rA-A)^2(rA-B)^2}\\
&= \frac{(h+B)(Bh+1)(h+A)(Ah+1)}{(Ar-B)^2(A-B)^2h^2}
\end{aligned}
$$

Now lets continue with the second integral W.

$$
W = \frac{1}{2\pi i}\oint_{\gamma_1}\frac{r(z_1+h)(1+z_1h)(h^2r+r+2hz_1)}{h^2(z_1-A)(z_1-B)(z_1-rA)(z_1-rB)}dz_1
$$

We find that the function inside the integral has two poles inside $|z_1| = 1$, a simple pole $z_1 = A$ and a simple pole $z_1 = rA$. To calculate the integral we need to calculate the residues of these poles.

$$Res(A) = \lim_{z_1 \to A} (z_1 - A) \frac{r(z_1 + h)(1 + z_1 h)(h^2 r + r + 2h z_1)}{h^2(z_1 - A)(z_1 - B)(z_1 - rA)(z_1 - rB)}$$

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

$$Res(rA) = \lim_{z_1 \to rA} (z_1 - rA) \frac{r(z_1 + h)(1 + z_1 h)(h^2 r + r + 2h z_1)}{h^2(z_1 - A)(z_1 - B)(z_1 - rA)(z_1 - rB)}$$

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

Then by Cauchy's residue theorem the second integral is equal to

$$W = \frac{1}{2\pi i}[2\pi i(Res(A) + Res(Ar))]$$

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

Now lets calculate the last integral G.

$$G = \frac{1}{2\pi i} \oint_{\gamma_1} \frac{r(z_1 + h)(1 + z_1 h)(hr + z_1)(r + h z_1)(2z_1 - Br - Ar)}{h^2(z_1 - A)(z_1 - B)(z_1 - rA)^2(z_1 - rB)^2} dz_1$$

The function inside the integral has two poles inside $|z_1 = 1$. A simple pole $z_1 = A$ and a pole $z_1 = rA$ of order 2. To calculate the integral we need to calculate the residues of these poles.

$$Res(A) = \lim_{z_1 \to A} (z_1 - A) \frac{r(z_1 + h)(1 + z_1 h)(hr + z_1)(r + h z_1)(2z_1 - Br - Ar)}{h^2(z_1 - A)(z_1 - B)(z_1 - rA)^2(z_1 - rB)^2}$$

$$= \frac{r(A + h)(1 + Ah)(hr + A)(r + hA)(2A - Br - Ar)}{h^2(A - B)(A - rA)^2(A - rB)^2}$$

$$Res(rA) = \lim_{z_1 \to rA} \frac{d}{dz_1} (z_1 - rA)^2 \frac{r(z_1 + h)(1 + z_1 h)(hr + z_1)(r + h z_1)(2z_1 - Br - Ar)}{h^2(z_1 - A)(z_1 - B)(z_1 - rA)^2(z_1 - rB)^2}$$

$$= \lim_{z_1 \to rA} \frac{d}{dz_1} \frac{r(z_1 + h)(1 + z_1 h)(hr + z_1)(r + h z_1)(2z_1 - Br - Ar)}{h^2(z_1 - A)(z_1 - B)(z_1 - rB)^2}$$

$$= \text{ For full expression use Maple}$$

Then by Cauchy's residue theorem we find

$$G = \frac{1}{2i}[2\pi i(Res(A) + Res(rA))] = \text{ For full expression use Maple}$$

Now combining the three integrals we find

$$J_1(\varphi, \varphi) = \lim_{r \downarrow 1} J_1(\varphi, \varphi, r) = \lim_{r \downarrow 1}[F + W - G]$$

$$= \lim_{r \downarrow 1} F + \lim_{r \downarrow 1} W - \lim_{r \downarrow 1} G \quad \text{(since the limits are bounded)}$$

$$= \frac{(h + B)(Bh + 1)(h + A)(Ah + 1)}{(Ar - B)^2(A - B)^2 h^2}$$

$$+ \frac{(-A - B)h^2 + (-6AB - 2)h^2 + (2A^3 - 6A^2 B - 4A - 4B)h^2 + (-6AB - 2)h - A - B}{(A - B)^3 h^2}$$

$$- \frac{(-A - B)h^2 + (-6AB - 2)h^2 + (2A^3 - 6A^2 B - 4A - 4B)h^2 + (-6AB - 2)h - A - B}{(A - B)^3 h^2}$$

$$= \frac{(h + B)(Bh + 1)(h + A)(Ah + 1)}{(Ar - B)^2(A - B)^2 h^2}$$

Now we will calculate the last limiting parameter $J_2(\varepsilon_1, \varepsilon_2)$. Again since we are calculating the variance we have that $\varphi_1 = \varphi_2$. By proposition 4.1.1, with $\xi = z$, $\gamma_1 := |z_1| = 1$ and $\gamma_2 := |z_2| = 1$, we have

$$
\begin{aligned}
J_2(\varphi_1, \varphi_2) = J_2(\varphi, \varphi) &= \frac{1}{2\pi i} \oint_{\gamma_1} \frac{f(|1 + h z_1|^2)}{z_1^2} dz_1 \cdot \frac{1}{2\pi i} \oint_{\gamma_2} \frac{g(|1 + h z_2|^2)}{z_2^2} dz_2 \\
&= \frac{1}{2\pi i} \oint_{\gamma_1} \frac{|1 + h z_1|^2}{|1 + z h_1|^2 + \epsilon} \frac{1}{z_1^2} dz_1 \cdot \frac{1}{2\pi i} \oint_{\gamma_2} \frac{|1 + h z_2|^2}{|1 + z h_2|^2 + \epsilon} \frac{1}{z_2^2} dz_2 \\
&= \frac{1}{2\pi i} \oint_{\gamma_1} \frac{\frac{1}{z_1}(z_1 + h)(1 + z_1 h)}{\frac{1}{z_1}((z_1 + h)(1 + z_1 h) + \epsilon) z_1^2} dz_1 \cdot \frac{1}{2\pi i} \oint_{\gamma_2} \frac{\frac{1}{z_2}(z_2 + h)(1 + z_2 h)}{\frac{1}{z_2}((z_2 + h)(1 + z_2 h) + \epsilon) z_2^2} dz_2 \\
&= \frac{1}{2\pi i} \oint_{\gamma_1} \frac{(z_1 + h)(1 + z_1 h)}{((z_1 + h)(1 + z_1 h) + z_1 \epsilon) z_1^2} dz_1 \cdot \frac{1}{2\pi i} \oint_{\gamma_2} \frac{(z_2 + h)(1 + z_2 h)}{((z_2 + h)(1 + z_2 h) + z_2 \epsilon) z_2^2} dz_2 \\
&= W \cdot F
\end{aligned}
$$

Note that solving the first integral is equivalent to solving the second. We will first solve the first one.

$$
W = \frac{1}{2\pi i} \oint_{\gamma_1} \frac{(z_1 + h)(1 + z_1 h)}{((z_1 + h)(1 + z_1 h) + z_1 \epsilon) z_1^2} dz_1 = \frac{1}{2\pi i} \oint_{\gamma_1} \frac{(z_1 + h)(1 + z_1 h)}{h(z_1 - A)(z_1 - B) z_1^2} dz_1
$$

The function inside the contour integral has two poles inside γ_1, one simple pole $z_1 = A$ and one pole $z_1 = 0$ of order 2. To calculate the integral we need to calculate the the residues of the poles.

$$
\begin{aligned}
Res(A) &= \lim_{z_1 \to A} (z_1 - A) \frac{(z_1 + h)(1 + z_1 h)}{h(z_1 - A)(z_1 - B) z_1^2} = \lim_{z_1 \to A} \frac{(z_1 + h)(1 + z_1 h)}{h(z_1 - B) z_1^2} \\
&= \frac{(A + h)(1 + A h)}{h(A - B) A^2} \\
Res(0) &= \lim_{z_1 \to 0} \frac{d}{d z_1} (z_1 - 0)^2 \frac{(z_1 + h)(1 + z_1 h)}{h(z_1 - A)(z_1 - B) z_1^2} = \lim_{z_1 \to 0} \frac{d}{d z_1} \frac{(z_1 + h)(1 + z_1 h)}{(h + z_1)(1 + h z_1) + \epsilon z} \\
&= \lim_{z_1 \to 0} \frac{\epsilon h(z_1^2 - 1)}{(h^2 z_1 + h z_1^2 + h + z)^2} \\
&= -\frac{\epsilon}{h}
\end{aligned}
$$

By Cauchy's Residue Theorem we find that

$$
\begin{aligned}
W &= \frac{1}{2\pi i} \left[2\pi i (Res(A) + Res(0)) \right] \\
&= \frac{(A + h)(1 + A h)}{h(A - B) A^2} + \frac{AB h^2 + AB + hA + Bh}{h A^2 B^2} \\
&= \frac{(h + B)(Bh + 1)}{h B^2 (A - B)}
\end{aligned}
$$

Hence

$$
\begin{aligned}
J_2(\varphi_1, \varphi_2) = J_2(\varphi, \varphi) &= W \cdot F = W \cdot W \\
&= \left[\frac{(h + A)(A h + 1)}{h A^2 (A - B)} - -\frac{\epsilon}{h} \right]^2
\end{aligned}
$$

Combining the limiting parameters $J_1(\varphi, \varphi)$ and $J_2(\varphi, \varphi)$ and letting $h = \sqrt{c}$, we find the variance

$$
\begin{aligned}
\sigma^2 &= \kappa J_1(\varphi, \varphi) + \beta J_2(\varphi, \varphi) \\
&= \kappa \left[\frac{(\sqrt{c} + B)(B\sqrt{c} + 1)(\sqrt{c} + A)(A\sqrt{c} + 1)}{(Ar - B)^2 (A - B)^2 c} \right] + \beta \left[\frac{(\sqrt{c} + A)(A\sqrt{c} + 1)}{\sqrt{c} A^2 (A - B)} - -\frac{\epsilon}{\sqrt{c}} \right]^2
\end{aligned}
$$

Centering factor $F_c(\varphi)$

It rest us to compute the centering factor $F_c(\varphi)$. To calculate this centering factor Proposition 4.1.2 is used. That is,

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

The function inside the contour integral has two poles inside $\gamma := |z| = 1$, 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.

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

Then by Cauchy's Residue Theorem and letting $h = \sqrt{c}$, we find that

$$
\begin{aligned}
\int \varphi(x) dF_c(x) &= -\frac{1}{4\pi i} [2\pi i (Res(A) + Res(0))] \\
&= -\frac{1}{2} \left[\frac{(1-A^2)}{h(A-B)A^2} + \frac{A+B}{hA^2B^2} \right] \\
&= -\frac{A^2B^2 - 2B^2 + 1}{2\sqrt{c}(A-B)B^2}
\end{aligned}
$$

This concludes the full proof of theorem 4.3.1. □

B

Appendix B: Additional Figures

B.1. Additional Power Plots: Alternative Hypothesis 3

(a) $p = 32$

(b) $p = 64$

(c) $p = 96$

(d) $p = 120$
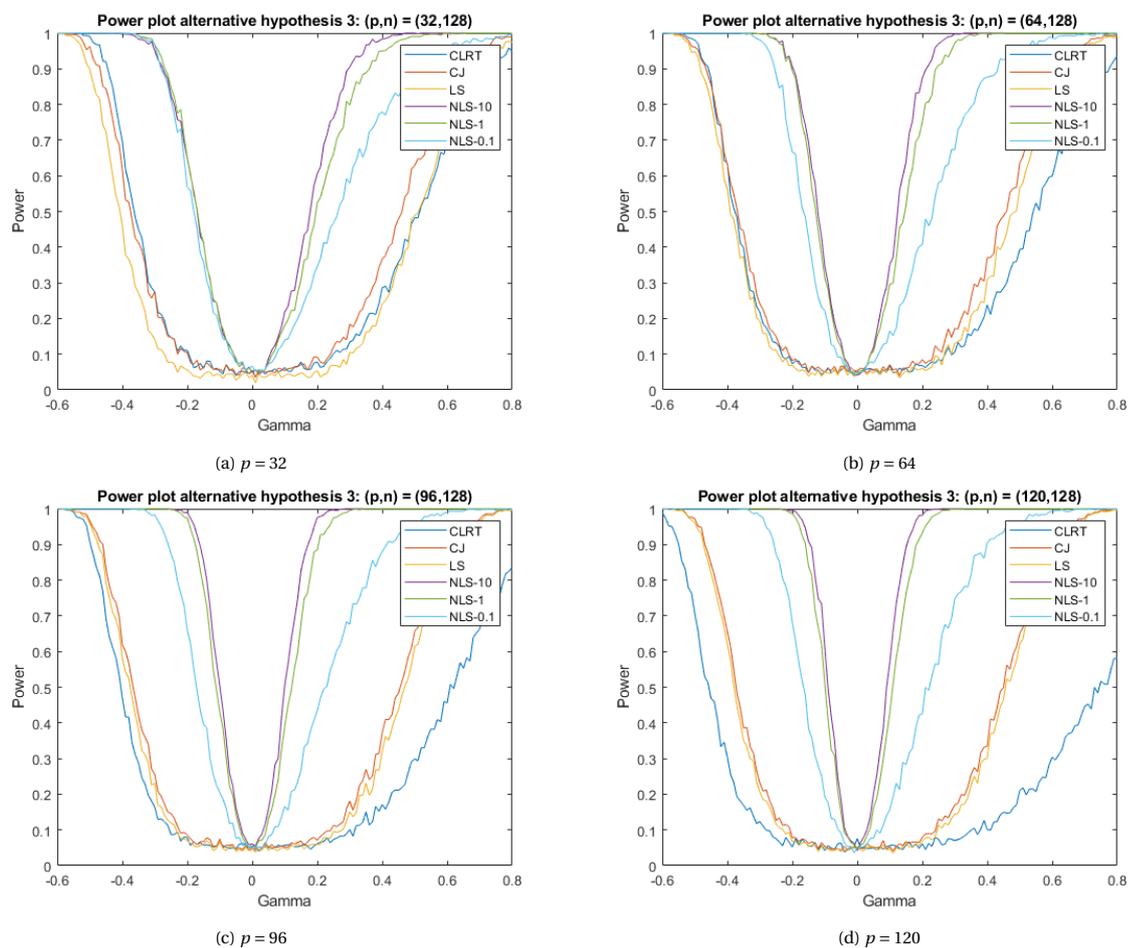
Figure B.1: Empirical powers under alternative hypothesis 3 with 1000 replications, $\gamma \in (-1, 1)$ and $r = 1/4$

(a) $p = 32$

(b) $p = 64$

(c) $p = 96$

(d) $p = 120$
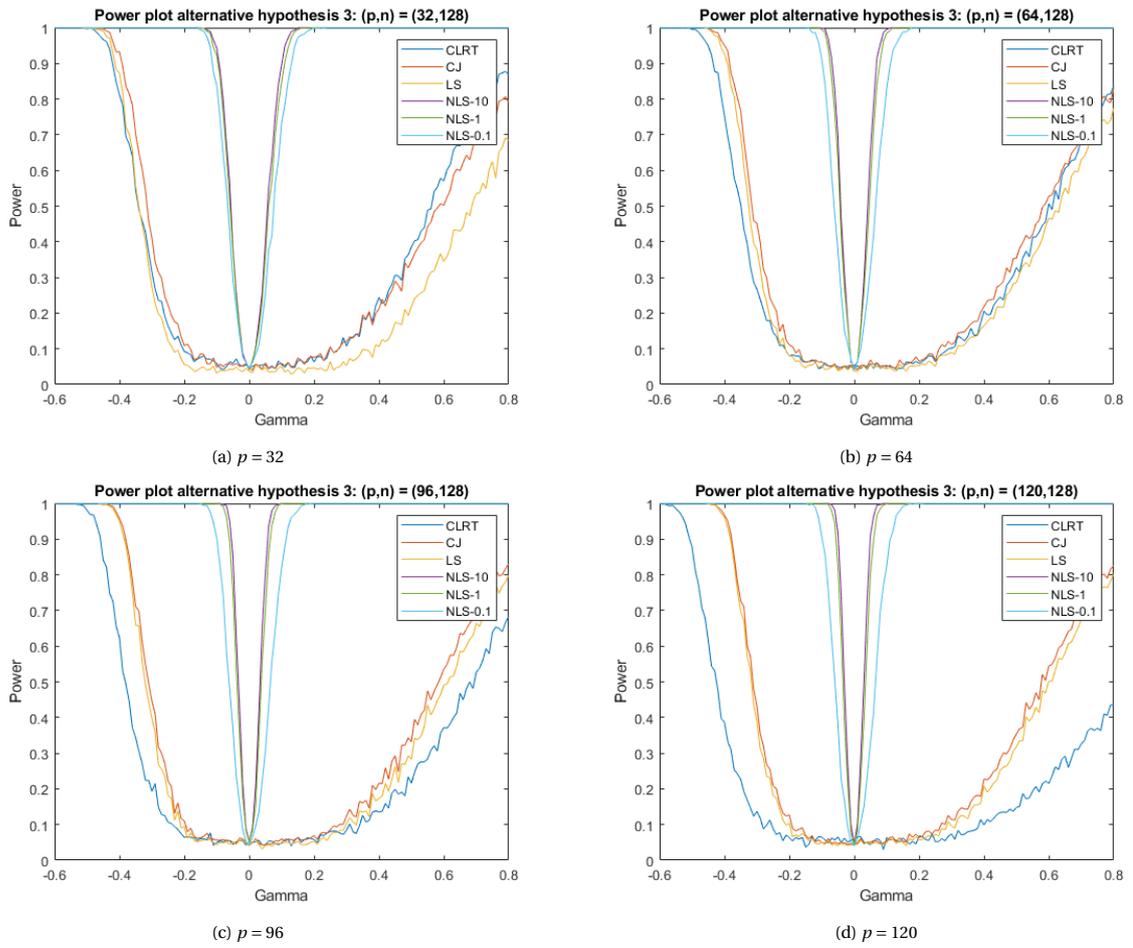
Figure B.2: Empirical powers under alternative hypothesis 3 with 1000 replications, $\gamma \in (-1, 1)$ and $r = 3/4$

B.2. Additional ROC Curves: Alternative Hypothesis 3
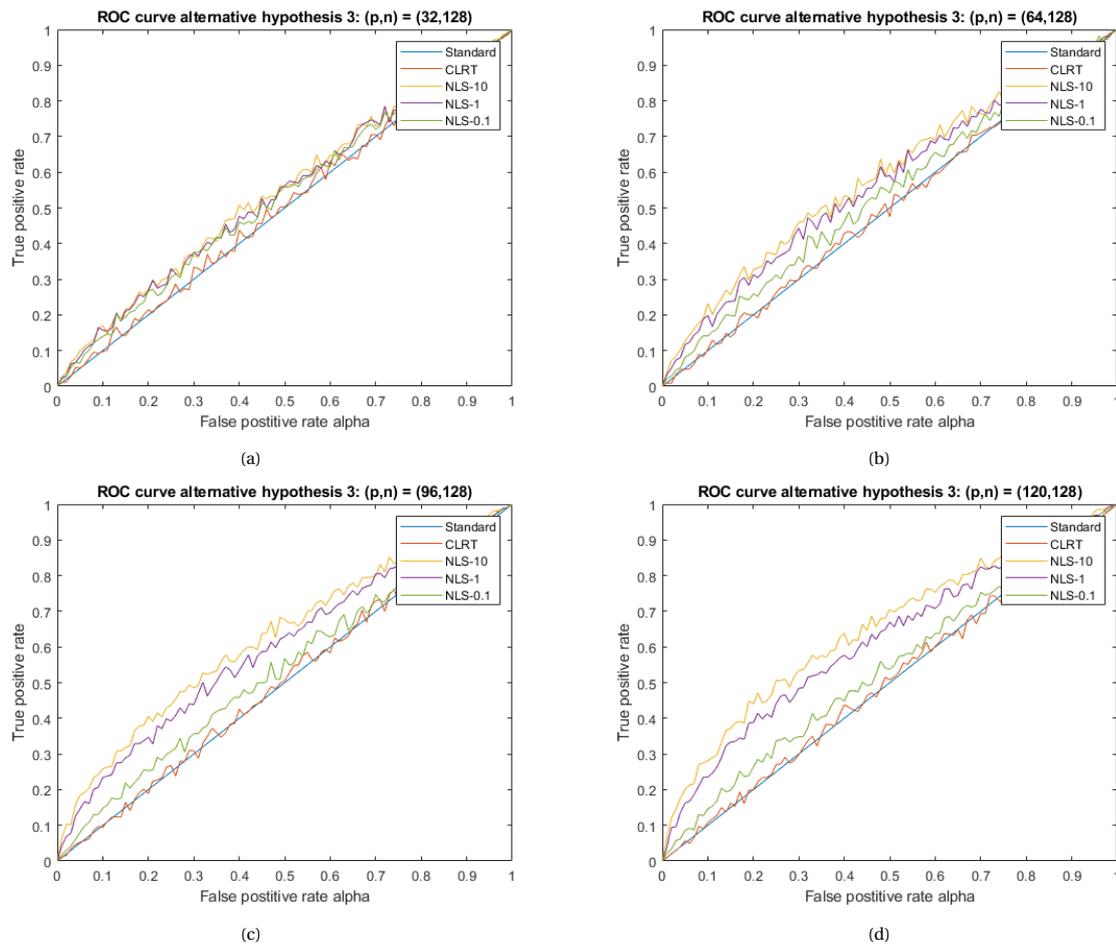
Figure B.3: ROC curves under alternative hypothesis 3 with 1000 replications, $\gamma = 0.05$ and $r = 1/4$

(a) $p = 32$

(b) $p = 64$

(c) $p = 96$

(d) $p = 120$
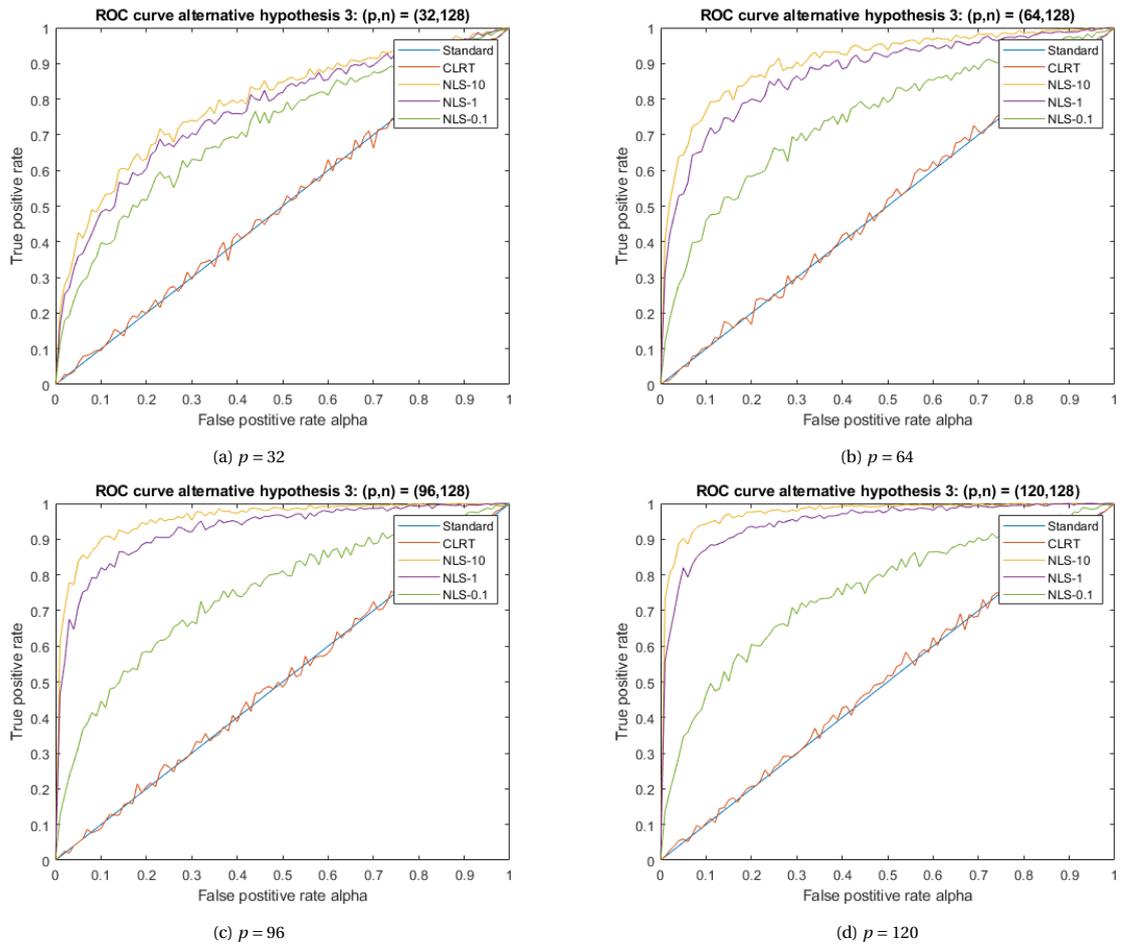
Figure B.4: ROC curves under alternative hypothesis 3 with 1000 replications, $\gamma = 0.05$ and $r = 3/4$

B.3. Additional 3D Power Plots: Alternative Hypothesis 4
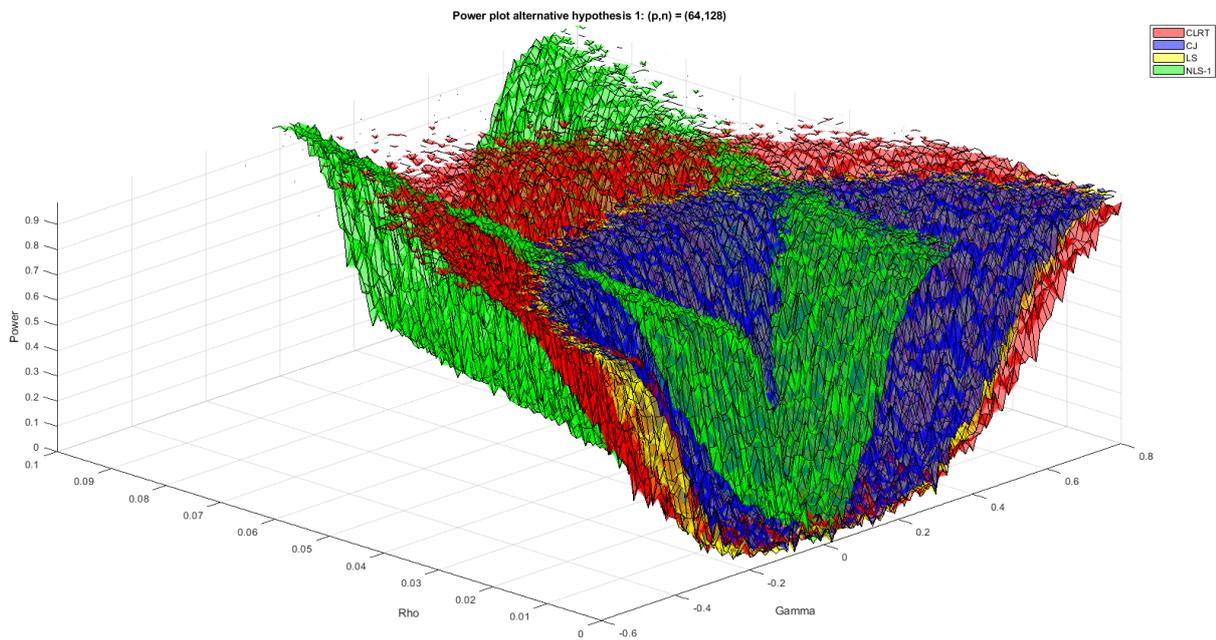
Figure B.5: 3D power plot under alternative hypothesis 4 with 100 replications, $p = 64$, $\rho \in (0,1)$, $\gamma \in (-1,1)$ and $r = 1/4$
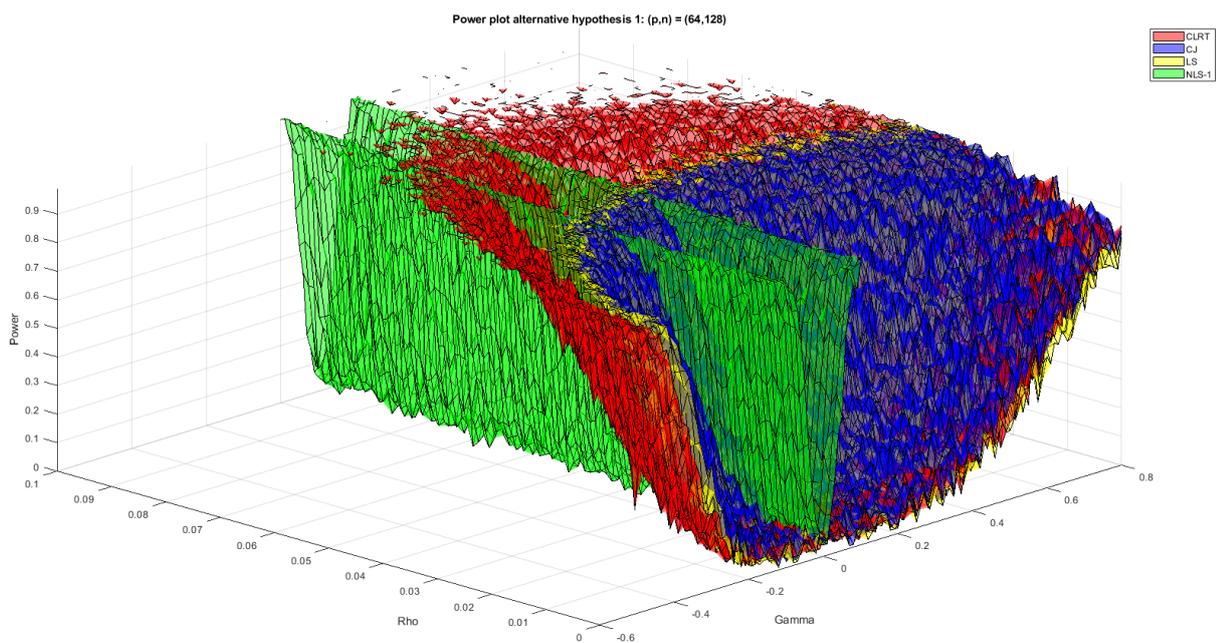
Figure B.6: 3D power plot under alternative hypothesis 4 with 100 replications, $p = 64$, $\rho \in (0,1)$, $\gamma \in (-1,1)$ and $r = 3/4$
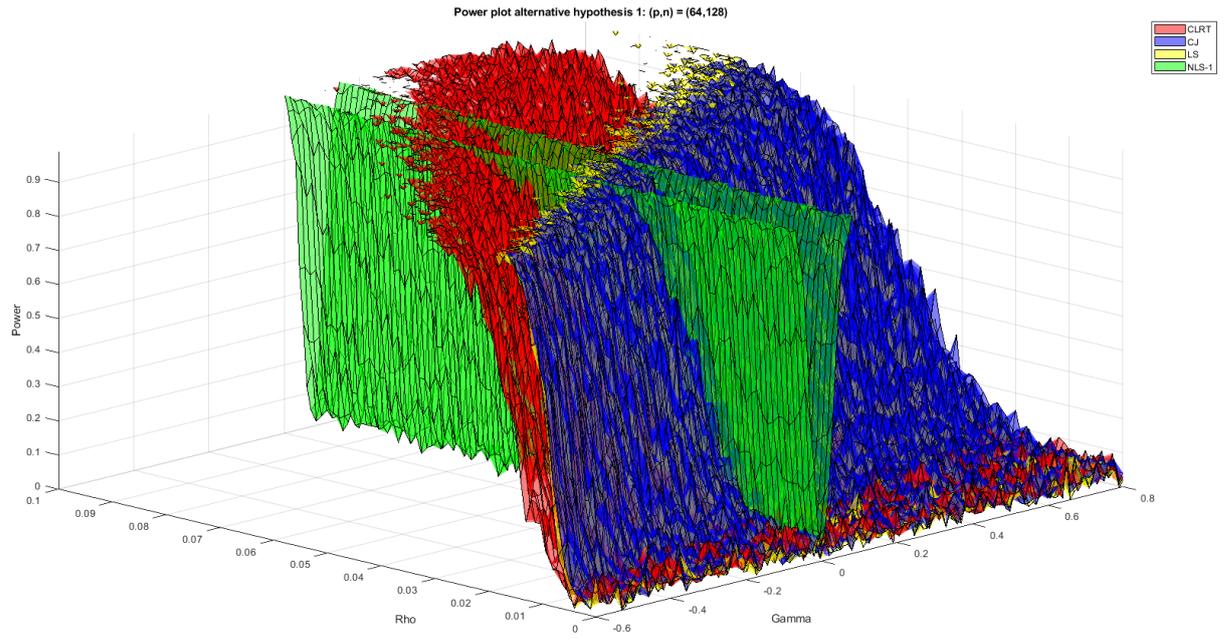
Figure B.7: 3D power plot under alternative hypothesis 4 with 100 replications, $p = 64$, $\rho \in (0, 1)$, $\gamma \in (-1, 1)$ and $r = 1$

C

Appendix C: Matlab Codes

```matlab
1  clearvars;
2
3  M = 10000;
4
5
6  n = 256;
7  p = 180;
8  c = p/n;
9  e = 1;
10 h = sqrt(c);
11
12 fourth_moment = 3; %4.5 for gamma(2,4)-2 data
13 kappa = 2;
14 beta = fourth_moment-1-kappa;
15
16 A = ((-c-e-1)+sqrt((c+e+1)^2-4*c))/(2*sqrt(c));
17 B = ((-c-e-1)-sqrt((c+e+1)^2-4*c))/(2*sqrt(c));
18 mu = (kappa-1)*(((h+B)*(B*h+1))/(h*(A-B)*(B*B-1)*B))+beta*(h+B)*(B*h+1)/(B^3*h*(A-B));
19 sigma = kappa*((h+B)*(h*B+1)*(h+A)*(h*A+1))/((A-B)^4*h^2)+beta*((h+A)*(A*h+1)/(A^2*h*(A
       -B))-e/h)^2;
20 limiting_distribution = -((A*A*B*B-2*B*B+1)/(h*2*(A-B)*B*B));
21
22 Y = zeros(M,1);
23
24 tic
25 for j = 1:M
26     Z = randn(p,n);
27     %Z = gamrnd(4,1/2,p,n)-2;
28     S = 1/n*(Z*transpose(Z));
29     X = trace(S*(S+e*eye(p))^(-1));
30
31     Y(j) = ((X-p*limiting_distribution)-mu)/sqrt(sigma);
32 end
33 toc
34
35 nbins = 92;
36 hist = histogram(Y,nbins, 'Normalization','pdf','DisplayName', "Z")
37 hold on
38 y = -4:0.1:4;
39 mu = 0;
40 sigma = 1;
41 f = exp(-(y-mu).^2./(2*sigma^2))./(sigma*sqrt(2*pi));
42 plot(y,f,'LineWidth',1.5,'DisplayName', "N(0,1)")
43 ylabel("Relative Frequency")
44 lgd = legend;
45 title("Empirical Distribution - N(0,1) data")
46 xlim([-4 4])
47 ylim([0 0.45])
48
49 disp(mean(Y))
```

```matlab
50    disp(var(Y))
```

Listing C.1: Empirical distribution functions

```matlab
1   clearvars;
2
3   M = 10000;
4
5   n = 128; %256
6   p_totaal = [8, 16, 32, 64, 96, 112, 120]; %128
7   %p_totaal = [16, 32, 64, 128, 192, 224, 240]; %256
8   e = [10, 1.5, 1, 0.5, 0.1];
9
10  fouth_moment = 4.5; %4.5 for gamma(2,4)-2
11  kappa = 2 ;
12  beta = fouth_moment - 1 - kappa;
13
14
15  CLRT = zeros(length(p_totaal),1);
16  CJ = zeros(length(p_totaal),1);
17  LS = zeros(length(p_totaal),1);
18  NLS = zeros(length(p_totaal),length(e));
19
20
21  tic
22  for j = 1 : length(p_totaal)
23      p = p_totaal(j);
24      c= p/n ;
25      h = sqrt(c);
26      Sigma_0 = eye(p);
27      countLS = 0;
28      countCLRT = 0;
29      countCJ = 0;
30      countNLS = zeros(length(e),1);
31      for i = 1 :M
32          Z = randn(p,n);
33          %Z = gamrnd(4,1/2,p,n)-2;
34          S = 1/n*(Z*transpose(Z));
35
36          %%statistics
37
38          %CLRT
39          T_1 = -p*(1/p*log(det(S))-log(trace(S)/p));
40          if ((T_1+(p-n)*log(1-c)-p)-(-kappa+1)/2*log(1-c)+1/2*beta*c)/sqrt(-kappa*log(1-c)-kappa*c) > 1.960 || ((T_1+(p-n)*log(1-c)-p)-(-kappa+1)/2*log(1-c)+1/2*beta*c)/sqrt(-kappa*log(1-c)-kappa*c) < -1.960
41              countCLRT = countCLRT+1;
42          end
43          %CJ TEST
44          T2 =(p^2*n/2)*trace((S/trace(S)-eye(p)/p)^2);
45          U = (2/(n*p))*T2;
46          if ((n*U-p)-(kappa-1+beta))/sqrt(2*kappa) > 1.960 || ((n*U-p)-(kappa-1+beta))/sqrt(2*kappa) < -1.960
47              countCJ = countCJ + 1 ;
48          end
49          %LS
50          T_3 = 1-(1/n*trace(S)^2*Frob(Sigma_0)^2)/(Frob(S)^2*Frob(Sigma_0)^2-trace(S*Sigma_0)^2);
51          if (p*T_3-(kappa-1+beta))/sqrt(2*kappa) > 1.960 || (p*T_3-(kappa-1+beta))/sqrt(2*kappa)< -1.960
52              countLS = countLS +1;
53          end
54          %NLS
55          for k = 1 : length(e)
56              e_fun = e(k);
57              A = ((-c-e_fun-1)+sqrt((c+e_fun+1)^2-4*c))/(2*sqrt(c));
58              B = ((-c-e_fun-1)-sqrt((c+e_fun+1)^2-4*c))/(2*sqrt(c));
59              mu = (kappa-1)*(((h+B)*(B*h+1))/(h*(A-B)*(B*B-1)*B))+beta*(h+B)*(B*h+1)/(B^3*h*(A-B));
60              sigma = kappa*((h+B)*(h*B+1)*(h+A)*(h*A+1))/((A-B)^4*h^2)+beta*((h+A)*(A*h+1)/(A^2*h*(A-B))-e_fun/h)^2;
```

```
61            limiting_distribution = -((A*A*B*B-2*B*B+1)/(h*2*(A-B)*B*B));
62            T_NLS =  trace(S*(S+e_fun*eye(p))^(-1)) - p*limiting_distribution;
63            if (T_NLS-mu)/sqrt(sigma) > 1.960 || (T_NLS-mu)/sqrt(sigma) < -1.960
64                countNLS(k) = countNLS(k) + 1;
65            end
66        end
67      end
68 CLRT(j) = round(countCLRT/M,4);
69 CJ(j) = round(countCJ/M,4);
70 LS(j) = round(countLS/M,4);
71 NLS(j,:) = round(countNLS/M,4);
72 end
73 toc
74
75 c_table = transpose(["(8,128)", "(16,128)", "(32,128)", "(64,128)", "(96,128)",
       "(112,128)", "(120,128)"]); %128
76 %c_table = transpose(["(16,256)", "(32,256)", "(64,256)", "(128,256)", "(192,256)",
       "(224,256)", "(240,256)"]);
77 T = table(c_table, CLRT, CJ, LS, NLS(:,1), NLS(:,2), NLS(:,3),NLS(:,4), NLS(:,5));
```

Listing C.2: Empirical sizes

```
1 clearvars;
2
3 M = 1000;
4
5 n = 128;
6 p = 64;
7 c= p/n ;
8 h = sqrt(c);
9 e = [1.5, 1, 0.5]; %H1
10 %e = [1, 0.5, 0.1]; %H2
11 %e = [10, 1, 1/2]; %H3
12
13 rho = 0: 0.001 :0.1;
14 w = 1.6449;
15
16 fouth_moment = 3; %4.5 for gamma(2,4)-2
17 kappa = 2 ;
18 beta = fouth_moment - 1 - kappa;
19 Sigma_0 = eye(p);
20
21 CLRT = zeros(length(rho),1);
22 CJ = zeros(length(rho),1);
23 LS = zeros(length(rho),1);
24 NLS = zeros(length(rho),length(e));
25
26 tic
27 for j = 1 : length(rho)
28     rh = rho(j);
29     %Alternative 1
30     Sigma = (1-rh)*eye(p)+rh*ones(p);
31
32     %Alternative 2
33     %Sigma = zeros(p);
34     %d=rh;
35     %for a = 1:p
36         %for b = 1:p
37             %Sigma(a,b) = d^(abs(a-b));
38         %end
39     %end
40
41     %gamma = rh;
42     %r = 1/2;
43     %lendiag = floor(r*p);
44     %Sigma = diag([ones(1,p-lendiag),(1+gamma)*ones(1,lendiag )]);
45
46     countLS = 0;
47     countCLRT = 0;
48     countCJ = 0;
49     countNLS = zeros(length(e),1);
```

```matlab
50      for i = 1 :M
51          X = randn(p,n);
52          %X = gamrnd(4,1/2,p,n)-2;
53          Y = Sigma^(1/2)*X;
54          S = 1/n*(Y*transpose(Y));
55
56          %%statistics
57
58          %CLRT
59          T_1 = -p*(1/p*log(det(S))-log(trace(S)/p));
60          if ((T_1+(p-n)*log(1-c)-p)-(-kappa+1)/2*log(1-c)+1/2*beta*c)/sqrt(-kappa*log(1-
    c)-kappa*c) > 1.960 || ((T_1+(p-n)*log(1-c)-p)-(-kappa+1)/2*log(1-c)+1/2*beta*c)/
    sqrt(-kappa*log(1-c)-kappa*c) < -1.960
61              countCLRT = countCLRT+1;
62          end
63
64          %CJ TEST
65          T2 =(p^2*n/2)*trace((S/trace(S)-eye(p)/p)^2);
66          U = (2/(n*p))*T2;
67          if ((n*U-p)-(kappa-1+beta))/sqrt(2*kappa) > 1.960 || ((n*U-p)-(kappa-1+beta))/
    sqrt(2*kappa) < -1.960
68              countCJ = countCJ + 1 ;
69          end
70          %LS
71          T_3 = 1-(1/n*trace(S)^2*Frob(Sigma_0)^2)/(Frob(S)^2*Frob(Sigma_0)^2-trace(S*
    Sigma_0)^2);
72          if  (p*T_3-(kappa-1+beta))/sqrt(2*kappa) > 1.960 || (p*T_3-(kappa-1+beta))/sqrt
    (2*kappa)< -1.960
73              countLS = countLS +1;
74          end
75
76          %NLS
77          for k = 1 : length(e)
78              e_fun = e(k);
79              A = ((-c-e_fun-1)+sqrt((c+e_fun+1)^2-4*c))/(2*sqrt(c));
80              B = ((-c-e_fun-1)-sqrt((c+e_fun+1)^2-4*c))/(2*sqrt(c));
81              mu = (kappa-1)*(((h+B)*(B*h+1))/(h*(A-B)*(B*B-1)*B))+beta*(h+B)*(B*h+1)/(B
    ^3*h*(A-B));
82              sigma = kappa*((h+B)*(h*B+1)*(h+A)*(h*A+1))/((A-B)^4*h^2)+beta*((h+A)*(A*h
    +1)/(A^2*h*(A-B))-e_fun/h)^2;
83              limiting_distribution = -((A*A*B*B-2*B*B+1)/(h*2*(A-B)*B*B));
84              T_NLS =  trace(S*(S+e_fun*eye(p))^(-1)) - p*limiting_distribution;
85              if (T_NLS-mu)/sqrt(sigma) > 1.960 || (T_NLS-mu)/sqrt(sigma) < -1.960
86                  countNLS(k) = countNLS(k) + 1;
87              end
88          end
89      end
90 CLRT(j) = countCLRT/M;
91 CJ(j) = countCJ/M;
92 LS(j) = countLS/M;
93 NLS(j,:) = countNLS/M;
94 end
95 toc
96
97 plot(rho,CLRT,'DisplayName', "CLRT")
98 hold on
99 plot(rho,CJ,'DisplayName', "CJ")
100 plot(rho,LS,'DisplayName', "LS")
101 for t = 1 : length(e)
102     plot(rho, NLS(:,t),'DisplayName', "NLS-"+num2str(e(t)))
103 end
104 hold off
105 xlabel("Rho")
106 ylabel("Power")
107 title("Power plot alternative hypothesis 1: (p,n) = ("+num2str(p)+","+num2str(n)+")")
108 lgd = legend;
```

Listing C.3: Empirical Power Plots

```matlab
1 clearvars;
2
```

```matlab
3   M = 1000;
4
5   n = 128;
6   p = 120;
7   c= p/n ;
8   h = sqrt(c);
9   %e = [1.5, 1, 0.5]; %H1
10  %e = [1, 0.5, 0.1]; %H2
11  e = [10, 1, 1/2]; %H3
12
13  rho = 0.03;
14  delta = 0.12;
15  gamma = 0.05;
16  %w = 1.960; %1.645
17
18  alpha = 0: 0.01 :1;
19  x = sqrt(2)*erfinv(1-alpha);
20
21  fouth_moment = 3; %4.5 for gamma(2,4)-2
22  kappa = 2 ;
23  beta = fouth_moment - 1 - kappa;
24  Sigma_0 = eye(p);
25
26  CLRT = zeros(length(x),1);
27  CJ = zeros(length(x),1);
28  LS = zeros(length(x),1);
29  NLS = zeros(length(x),length(e));
30
31  %Alternative 1
32  %Sigma = (1-rho)*eye(p)+rho*ones(p);
33
34  %Alternative 2
35  %Sigma = zeros(p);
36  %d=delta;
37  %for a = 1:p
38      %for b = 1:p
39          %Sigma(a,b) = d^(abs(a-b));
40      %end
41  %end
42
43  %Alternative 3
44  r = 1/2;
45  lendiag = floor(r*p);
46  Sigma = diag([ones(1,p-lendiag),(1+gamma)*ones(1,lendiag )]);
47
48
49  tic
50  for j = 1 : length(x)
51
52      w = x(j);
53
54      countLS = 0;
55      countCLRT = 0;
56      countCJ = 0;
57      countNLS = zeros(length(e),1);
58      for i = 1 :M
59          X = randn(p,n);
60          Y = Sigma^(1/2)*X;
61          S = 1/n*(Y*transpose(Y));
62
63  %%statistics
64
65          %CLRT
66          T_1 = -p*(1/p*log(det(S))-log(trace(S)/p));
67          if ((T_1+(p-n)*log(1-c)-p)-(-kappa+1)/2*log(1-c)+1/2*beta*c)/sqrt(-kappa*log(1-
    c)-kappa*c) > w || ((T_1+(p-n)*log(1-c)-p)-(-kappa+1)/2*log(1-c)+1/2*beta*c)/sqrt(-
    kappa*log(1-c)-kappa*c) < -w
68              countCLRT = countCLRT+1;
69          end
70
71          %CJ TEST
```

```
72        %T2 =(p^2*n/2)*trace((S/trace(S)-eye(p)/p)^2);
73        %U = (2/(n*p))*T2;
74        %if (n*U-p) > w*sqrt(2*kappa)+(kappa-1+beta)
75        %if ((n*U-p)-(kappa-1+beta))/sqrt(2*kappa) > w || ((n*U-p)-(kappa-1+beta))/sqrt
   (2*kappa) < -w
76            %countCJ = countCJ + 1 ;
77        %end
78
79        %LS
80        %T_3 = 1-(1/n*trace(S)^2*Frob(Sigma_0)^2)/(Frob(S)^2*Frob(Sigma_0)^2-trace(S*
   Sigma_0)^2);
81        %if p*T_3 > w*sqrt(2*kappa)+(kappa-1+beta)
82        %if  (p*T_3-(kappa-1+beta))/sqrt(2*kappa) > w || (p*T_3-(kappa-1+beta))/sqrt(2*
   kappa)< -w
83            %countLS = countLS +1;
84        %end
85
86        %NLS
87        for k = 1 : length(e)
88            e_fun = e(k);
89            A = ((-c-e_fun-1)+sqrt((c+e_fun+1)^2-4*c))/(2*sqrt(c));
90            B = ((-c-e_fun-1)-sqrt((c+e_fun+1)^2-4*c))/(2*sqrt(c));
91            mu = (kappa-1)*(((h+B)*(B*h+1))/(h*(A-B)*(B*B-1)*B))+beta*(h+B)*(B*h+1)/(B
   ^3*h*(A-B));
92            sigma = kappa*((h+B)*(h*B+1)*(h+A)*(h*A+1))/((A-B)^4*h^2)+beta*((h+A)*(A*h
   +1)/(A^2*h*(A-B))-e_fun/h)^2;
93            limiting_distribution = -((A*A*B*B-2*B*B+1)/(h*2*(A-B)*B*B));
94            T_NLS =  trace(S*(S+e_fun*eye(p))^(-1)) - p*limiting_distribution;
95            if (T_NLS-mu)/sqrt(sigma) > w || (T_NLS-mu)/sqrt(sigma) < -w
96                countNLS(k) = countNLS(k) + 1;
97            end
98        end
99    end
100 CLRT(j) = countCLRT/M;
101 CJ(j) = countCJ/M;
102 LS(j) = countLS/M;
103 NLS(j,:) = countNLS/M;
104 end
105 toc
106
107 plot(alpha,alpha,'DisplayName', "Standard")
108 hold on
109 plot(alpha,CLRT,'DisplayName', "CLRT")
110 %plot(alpha,CJ,'DisplayName', "CJ")
111 %plot(alpha,LS,'DisplayName', "LS")
112 for t = 1 : length(e)
113     plot(alpha, NLS(:,t),'DisplayName', "NLS-"+num2str(e(t)))
114 end
115 hold off
116 xlabel("False postitive rate alpha")
117 ylabel("True positive rate")
118 title("ROC curve alternative hypothesis 3: (p,n) = ("+num2str(p)+","+num2str(n)+")")
119 lgd = legend;
```

Listing C.4: ROC Curves

```
1 clearvars;
2
3 M = 100;
4
5 n = 128;
6 p = 64;
7 c = p/n;
8 h = sqrt(c);
9 %e = [1.5, 1, 0.5]; %H1
10 %e = [1, 0.5, 0.1]; %H2
11 %e = [10, 1, 1/2]; %H3
12 e = 1;
13
14 rho = 0: 0.0001 :001;
15 gamma = -0.8: 0.01 :0.6;
```

```matlab
16  w = 1.6449;
17
18  fouth_moment = 3; %4.5 for gamma(2,4)-2
19  kappa = 2 ;
20  beta = fouth_moment - 1 - kappa;
21  Sigma_0 = eye(p);
22
23  CLRT = zeros(length(rho),length(gamma));
24  CJ = zeros(length(rho),length(gamma));
25  LS = zeros(length(rho),length(gamma));
26  NLS = zeros(length(rho),length(gamma));
27
28  tic
29  for j = 1 : length(rho)
30      for i = 1 : length(gamma)
31          rh = rho(j);
32          %Alternative 1
33          Sigma_1 = (1-rh)*eye(p)+rh*ones(p);
34
35          %Alternative 3
36          ga = gamma(i);
37          r = 1/2;
38          lendiag = floor(r*p);
39          Sigma_3 = diag([ones(1,p-lendiag),(1+ga)*ones(1,lendiag)]);
40
41
42          %Alternative 4
43          Sigma = Sigma_3^(1/2)*Sigma_1*Sigma_3^(1/2);
44
45          countLS = 0;
46          countCLRT = 0;
47          countCJ = 0;
48          countNLS = 0;
49          for k = 1 :M
50              X = randn(p,n);
51              %X = gamrnd(4,1/2,p,n)-2;
52              Y = Sigma^(1/2)*X;
53              S = 1/n*(Y*transpose(Y));
54
55              %%statistics
56
57              %CLRT
58              T_1 = -p*(1/p*log(det(S))-log(trace(S)/p));
59              if ((T_1+(p-n)*log(1-c)-p)-(-kappa+1)/2*log(1-c)+1/2*beta*c)/sqrt(-kappa*
      log(1-c)-kappa*c) > 1.960 || ((T_1+(p-n)*log(1-c)-p)-(-kappa+1)/2*log(1-c)+1/2*beta
      *c)/sqrt(-kappa*log(1-c)-kappa*c) < -1.960
60                  countCLRT = countCLRT+1;
61              end
62              %CJ TEST
63              T2 =(p^2*n/2)*trace((S/trace(S)-eye(p)/p)^2);
64              U = (2/(n*p))*T2;
65              if ((n*U-p)-(kappa-1+beta))/sqrt(2*kappa) > 1.960 || ((n*U-p)-(kappa-1+beta
      ))/sqrt(2*kappa) < -1.960
66                  countCJ = countCJ + 1 ;
67              end
68              %LS
69              T_3 = 1-(1/n*trace(S)^2*Frob(Sigma_0)^2)/(Frob(S)^2*Frob(Sigma_0)^2-trace(S
      *Sigma_0)^2);
70              if (p*T_3-(kappa-1+beta))/sqrt(2*kappa) > 1.960 || (p*T_3-(kappa-1+beta))/
      sqrt(2*kappa)< -1.960
71                  countLS = countLS +1;
72              end
73              %NLS
74              e_fun = e;
75              A = ((-c-e_fun-1)+sqrt((c+e_fun+1)^2-4*c))/(2*sqrt(c));
76              B = ((-c-e_fun-1)-sqrt((c+e_fun+1)^2-4*c))/(2*sqrt(c));
77              mu = (kappa-1)*(((h+B)*(B*h+1))/(h*(A-B)*(B*B-1)*B))+beta*(h+B)*(B*h+1)/(B
      ^3*h*(A-B));
78              sigma = kappa*((h+B)*(h*B+1)*(h+A)*(h*A+1))/((A-B)^4*h^2)+beta*((h+A)*(A*h
      +1)/(A^2*h*(A-B))-e_fun/h)^2;
79              limiting_distribution = -((A*A*B*B-2*B*B+1)/(h*2*(A-B)*B*B));
```

```matlab
80              T_NLS =  trace(S*(S+e_fun*eye(p))^(-1)) - p*limiting_distribution;
81              if (T_NLS-mu)/sqrt(sigma) > 1.960 || (T_NLS-mu)/sqrt(sigma) < -1.960
82                  countNLS = countNLS + 1;
83              end
84          end
85      CLRT(j,i) = countCLRT/M;
86      CJ(j,i) = countCJ/M;
87      LS(j,i) = countLS/M;
88      NLS(j,i) = countNLS/M;
89      end
90 end
91 toc
92
93 surf(gamma,rho,CLRT,'DisplayName', "CLRT",'FaceColor', 'r', 'FaceAlpha',0.5)
94 hold on
95 surf(gamma,rho, CJ,'DisplayName', "CJ", 'FaceColor', 'b', 'FaceAlpha',0.5)
96 surf(gamma,rho, LS,'DisplayName', "LS", 'FaceColor', 'y', 'FaceAlpha',0.5)
97 surf(gamma,rho, NLS,'DisplayName', "NLS-"+num2str(e), 'FaceColor', 'g', 'FaceAlpha'
       ,0.5)
98 hold off
99 ylabel("Rho")
100 xlabel("gamma")
101 zlabel("Power")
102 zlim([0 0.99])
103 %xlim([0 0.01])
104 %ylim([0 0.01])
105 title("Power plot alternative hypothesis 4: (p,n) = ("+num2str(p)+","+num2str(n)+")")
106 lgd = legend;
```

Listing C.5: 3D Empirical Power Plots

```matlab
1 clearvars;
2
3 M = 1000;
4
5 n = 128;
6 p = 64;
7 c = p/n;
8
9
10 rho = 0: 0.001 :0.1;
11 %gamma = 0.1; %0.135
12 %rho = 0;% 0.07
13
14 eigv = 0:0.1:11.9;
15 K = zeros(length(eigv),length(rho));
16 tic
17 for j = 1 : length(rho)
18     %Alternative 1
19     rh = rho(j);
20     Sigma_1 = (1-rh)*eye(p)+rh*ones(p);
21
22     %Alternative 3
23     ga = 0; %rho(j);
24     r = 1/2;
25     lendiag = floor(r*p);
26     Sigma_3 = diag([ones(1,p-lendiag),(1+ga)*ones(1,lendiag)]);
27
28     %Alternative 4
29     Sigma = Sigma_3^(1/2)*Sigma_1*Sigma_3^(1/2);
30
31     Z = zeros(M,p);
32     for i = 1:M
33         X = randn(p,n);
34         Y = Sigma^(1/2)*X;
35         S = 1/n*(Y*transpose(Y));
36         l = eig(S);
37         Z(i,:) = l;
38     end
39     eigenvalues = Z(:);
```

```matlab
40      histog = histogram(eigenvalues, 'binWidth',0.1,'BinLimits', [0 12], 'normalization'
        , 'pdf').Values;
41      K(:,j) = histog;
42  end
43  toc
44
45
46  surf(rho,eigv,K,'DisplayName', "Eigenvalues",'FaceAlpha',0.8)
47  ylabel("Value of eigenvalues")
48  xlabel("rho")
49  zlabel("Relative Frequency")
50  title("Histogram of eigenvalues in rho direction")
```

Listing C.6: 3D Histogram Eigenvalues

Bibliography

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

Bodnar, T., Dette, H., & Parolya, N. (2019). Testing for independence of large dimensional vectors. *The Annals of Statistics*, *47*(5), 2977–3008. https://doi.org/10.48550/arXiv.1708.03964

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. https://doi.org/10.1016/j.jmva.2014.08.006

John, S. (1971). Some optimal multivariate tests. *Biometrika*, *58*(1), 123–127. https://doi.org/10.1093/biomet/58.1.123

Ledoit, O., & Peche, S. (2011). Eigenvectors of some large sample covariance matrix ensembles. *Probability Theory and Related Fields*, *151*, 233–264. https://doi.org/10.1007/s00440-010-0298-3

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*(2), 365–411. https://doi.org/10.1016/S0047-259X(03)00096-4

Ledoit, O., & Wolf, M. (2012). Nonlinear shrinkage estimator of large-dimensional covariance matrices. *The Annals of Statistics*, *40*(2), 1024–1060. https://doi.org/10.1214/12-AOS989

Marchenko, V. A., & Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Math. USSR-sb*, *1*, 457–483.

Najim, J., & Yao, J. (2016). Gaussian fluctuations for linear spectral statistics of large random covariance matrices. *The Annals of Applied Probability*, *26*(3), 1837–1887. https://doi.org/10.1214/15-AAP1135

Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *The Royal Society*, *231*, 694–706. https://doi.org/10.1098/rsta.1933.0009

Perlman, M. D. (2007). *Stat 542: Multivarite statistical analysis* (On-Line Class Notes). Univ. Washington. Seatle, Washington.

Pillai, K. C. S. (1955). Some new test criteria in multivariate analysis. *Annals of Mathematical Statistics*, *26*(1), 117–121. https://doi.org/10.1214/aoms/1177728599

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. https://doi.org/10.1006/jmva.1995.1058

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. *Electron. J. Statist.*, *7*, 2164–2192. https://doi.org/10.1214/13-EJS842

Yao, J., Zheng, S., & Bai, Z. D. (2015). *Large sample covariance matrices and high-dimensional data analysis*. Cambridge University Press.