

UC Davis

UC Davis Electronic Theses and Dissertations

Title

Goodness-of-Fit Testing and Bootstrap Methods for High-Dimensional Elliptical Distributions

Permalink

<https://escholarship.org/uc/item/02v5019z>

ISBN

9798193262619

Author

Wang, Siyao

Publication Date

2026-09-08

Peer reviewed|Thesis/dissertation

Goodness-of-Fit Testing and Bootstrap Methods for High-Dimensional Elliptical Distributions

By

SIYAO WANG

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Miles E. Lopes, Chair

Xiucui Ding

Debashis Paul

Committee in Charge

2026

Contents

Abstract	iv
Acknowledgments	v
Chapter 1. Introduction	1
Chapter 2. Testing Elliptical Models in High Dimensions	6
2.1. Introduction	6
2.2. Method	9
2.3. Theory	11
2.4. Numerical experiments	15
Chapter 3. A Bootstrap Method for Spectral Statistics in High-Dimensional Elliptical Models	22
3.1. Introduction	22
3.2. Method	25
3.3. Theoretical results	29
3.4. Numerical results	37
3.5. Conclusion	50
Appendix A. Supplementary Material for Chapter 2	52
A.1. Preliminaries	52
A.2. Asymptotic normality of T_n	53
A.3. Ratio consistency of $\widehat{\sigma}_n^2$	86
A.4. Background results	94
A.5. Additional numerical experiments	97
Appendix B. Supplementary Material for Chapter 3	98
B.1. Proof of Theorem 3.3.1	98

B.2. Proof of Theorem 3.3.2	107
B.3. Proof of Theorem 3.3.3	109
B.4. Background results	115
B.5. Discussion of examples in Section 3.3.1	116
Bibliography	120

Abstract

High-dimensional statistical inference has become a fundamental topic in modern statistics, particularly in settings where the number of features is comparable to or even exceeds the sample size. Within this context, elliptical distributions are widely used as a flexible alternative to the multivariate normal distribution. Although recent research has substantially expanded the theoretical foundations of elliptical models in high dimensions, important methodological gaps remain. This thesis addresses two of them, by developing a goodness-of-fit procedure for validating the elliptical assumption and a bootstrap-based inference framework for spectral statistics.

The first part of this thesis introduces a novel goodness-of-fit test to verify the elliptical assumption in high dimensions. We propose a test statistic based on the asymptotic equivalence of two distinct kurtosis estimators. This method avoids the estimation of an inverse covariance matrix, which prevents classical goodness-of-fit tests for elliptical models from being applicable in high-dimensional settings. Importantly, the asymptotic validity of the proposed test statistic is established under a general setting that requires *no structural assumptions* on the population covariance matrix.

The second part introduces a parametric bootstrap methodology specifically designed for spectral statistics in high-dimensional elliptical models. Recognizing that the nonparametric bootstrap can be inconsistent for spectral statistics in this regime, we propose a bootstrap procedure that explicitly incorporates the elliptical structure of the data. We prove that this method consistently approximates the limiting distributions of linear spectral statistics by correctly capturing the fluctuations of the sample eigenvalues using recent advancements in high-dimensional central limit theorems. Extensive empirical results demonstrate the effectiveness of the method.

Acknowledgments

I would like to thank my advisor, Professor Miles E. Lopes, for his patience and support throughout this process. I am especially grateful that he always encouraged me to explore new directions and backed me in whatever decisions I made. I also want to thank the members of my defense and qualifying exam committees for their time and feedback. Finally, my deepest thanks to my family and friends for always standing by me.

CHAPTER 1

Introduction

The classical theory of multivariate statistics is built on the low-dimensional regime in which the dimension is fixed and the sample size tends to infinity (Anderson, 2003; Johnson et al., 2002; Mardia et al., 2024; Muirhead, 1982). In this setting, the Law of Large Numbers and the Central Limit Theorem provide the asymptotic guarantees for parameter estimation and hypothesis testing. For parameter estimation, classical procedures including maximum likelihood and method of moments yield estimators that are consistent and asymptotically normal under standard regularity conditions, such as the sample covariance matrix as a consistent estimator of the population covariance (Anderson, 2003; Muirhead, 1982). For hypothesis testing, two complementary lines of work have developed. The first is goodness-of-fit testing, which validates distributional assumptions. Examples include the Shapiro-Wilk and Jarque-Bera tests for normality (Jarque and Bera, 1980; Shapiro and Wilk, 1965), the Pearson Chi-Square test and the Kolmogorov-Smirnov test for general goodness-of-fit (Massey Jr, 1951; Slakter, 1965), and Mardia's tests for multivariate skewness and kurtosis (Mardia, 1970). The second is hypothesis testing for specific parameters of some distribution, such as Hotelling's T^2 for testing the mean vector and likelihood ratio tests for the covariance matrix (Anderson, 2003; Muirhead, 1982).

These asymptotic results are useful, but they can be analytically intractable in practice since the limiting distribution may depend on unknown population parameters or have complex forms. In such cases, resampling methods provide a computational alternative. The most prominent variant is the nonparametric bootstrap (Efron, 1979), which approximates the sampling distribution of complex statistics by repeatedly resampling from the empirical distribution without relying on parametric assumptions. In low-dimensional settings, the bootstrap is known to be consistent and to often achieve higher-order accuracy compared to normal approximations (Hall, 2013; Shao and Tu, 2012). Beyond the nonparametric approach, parametric and wild bootstrap variants have also been developed and are widely applied in regression and time-series analysis (Efron, 1982, 2012).

While classical methods provide a comprehensive framework for the low-dimensional regime, modern applications in genomics, finance, and signal processing frequently generate datasets where the dimension is comparable to, or exceeds, the sample size. In this high-dimensional regime, the theoretical guarantees of classical statistics often fail to hold. In covariance estimation, when the dimension exceeds the sample size, the sample covariance matrix becomes singular and cannot be inverted. Many classical methods that rely on the inverse sample covariance matrix, including Hotelling’s T^2 (Bai and Saranadasa, 1996) and Fisher’s Linear Discriminant Analysis (Bickel and Levina, 2004), therefore become invalid. Even when the dimension does not exceed the sample size, the high dimensionality itself distorts the eigenvalues of the sample covariance matrix. Marcenko and Pastur (1967) showed that the empirical eigenvalues of the sample covariance matrix do not converge to the population eigenvalues, but instead follow the Marchenko-Pastur law. The bias is systematic: the largest sample eigenvalues are biased upward (Geman, 1980), while the smallest are biased downward (Bai et al., 1993). Therefore, classical inference procedures based on sample eigenvalues, including likelihood ratio tests (Johnstone, 2001) and Principal Component Analysis (PCA) (Johnstone and Lu, 2009), become inconsistent. Standard resampling techniques encounter the same difficulty. The nonparametric bootstrap often fails to estimate the distribution of statistics based on sample eigenvalues in high dimensions (El Karoui and Purdom, 2016, 2019), since it resamples from the empirical spectral distribution and therefore inherits the same Marchenko-Pastur bias rather than correcting for the fluctuations induced by high dimensionality.

This “curse of dimensionality” (Donoho et al., 2000) has motivated substantial theoretical developments such as the Tracy-Widom limit for the largest eigenvalue (Johnstone, 2001) and central limit theorems for functions of eigenvalues (Bai and Silverstein, 2004; Pan and Zhou, 2008; Zheng, 2012) in random matrix theory. In addition, many classical inference procedures have been extended to the high-dimensional setting, including robust hypothesis testing procedures for mean vectors and covariance matrices (Bai and Saranadasa, 1996; Chen and Qin, 2010; Srivastava, 2005), goodness-of-fit tests (Bing and Latremouille, 2025; Chen and Xia, 2023; Cui and Zhang, 2026), and specialized bootstrap schemes tailored to mitigate high-dimensional bias (Chernozhukov et al., 2013; Lopes et al., 2023). Despite this progress, much of this work is based on the multivariate normal assumption. This assumption simplifies the analysis but imposes restrictive structural constraints, such as exponential tail decay and asymptotic independence, that often fail to hold for

real-world data in finance (Cont, 2001; Mandelbrot et al., 1963), genomics (Purdom et al., 2005), and environmental science (Davison and Huser, 2015).

As an alternative to the Gaussian framework, independent component models have received considerable attention (Bai and Silverstein, 2010; Ledoit and Wolf, 2017; Ledoux, 2001; Lytova and Pastur, 2009; Najim and Yao, 2016; Pan and Zhou, 2008; Wang and Yao, 2013; Zheng, 2012). In this model, the data vector is given by $\mathbf{x} = \Sigma^{1/2}\mathbf{z}$, where the components of $\mathbf{z} = (z_1, \dots, z_p)^\top$ are independent, identically distributed, and centered. Unlike the Gaussian model, this framework can capture heavy-tailed marginal behavior through the choice of z_i . However, a critical limitation of this framework in high dimensions is the absence of tail dependence. By concentration of measure, the Euclidean norm of a high-dimensional vector with independent components concentrates sharply around its expectation (Ledoux, 2001; Vershynin, 2018), so that extreme events in such models occur independently across orthogonal directions. By contrast, real-world systems often exhibit systemic risk, in which extreme events co-occur, as observed in market-wide financial crashes and biological shocks (Embrechts et al., 2002; McNeil et al., 2015).

The family of elliptical distributions addresses these limitations, providing a natural framework that accommodates both heavy tails and tail dependence. A random vector follows an elliptical distribution if it can be represented in the form $\mathbf{x} = \xi \Sigma^{1/2} \mathbf{u}$, where $\mathbf{u}$ is uniformly distributed on the unit sphere, and ξ is a non-negative scalar random variable independent of $\mathbf{u}$ (Cambanis et al., 1981; Muirhead, 1982). Elliptical distributions generalize the multivariate normal distribution while preserving key properties such as symmetry and linear invariance. In addition, the radial variable ξ controls the tail behavior and induces tail dependence among the components of $\mathbf{x}$. These properties have made elliptical models a standard choice in many applied domains. In finance, they capture the heavy tails commonly observed in asset returns (Gupta et al., 2013; McNeil et al., 2015). In signal processing, they model impulsive noise and radar clutter (Hung et al., 2022; Li and Mandic, 2022; Mahmood and Chitre, 2015; Ollila et al., 2012). Beyond these applications, many statistical methodologies have been extended and developed for the elliptical setting. In the low-dimensional regime, classical procedures have been extended to robust estimation (Maronna, 1976; Tyler, 1987), hypothesis testing (Anderson et al., 1986; Beran, 1979; Hallin and Paindaveine, 2006; Kariya, 1981; Koltchinskii and Sakhanenko, 2000; Osiewalski and Steel, 1993; Schott, 2002; Tang and Li, 2023), multivariate analysis (Croux and Haesbroeck, 2000; Hubert et al., 2005; McLachlan, 2005),

dimension reduction (Li, 1991), and regression (Kano et al., 1993; Zellner, 1976). The statistical foundations of elliptical models in high-dimensional settings are also rapidly being established, covering robust covariance estimation (Chen et al., 2011; Goes et al., 2020; Ollila and Tyler, 2014; Sun et al., 2016), testing and inference (Dörnemann, 2026; Han and Liu, 2018; Shi et al., 2024), regression (Chen et al., 2022; Han and Liu, 2013; Lin et al., 2019) and spectral statistics (Ding and Xie, 2023; El Karoui, 2009; Hu et al., 2019a,b; Jun et al., 2022; Yang et al., 2021; Zhang et al., 2022). Nevertheless, significant challenges remain in the high-dimensional regime. This thesis addresses two such challenges by developing theoretical and computational tools for high-dimensional inference under elliptical distributions, supported by theoretical analysis and empirical results.

First, testing the validity of the elliptical assumption is a crucial prerequisite for downstream inference. Classical tests break down in this regime due to the singularity issues discussed above. To resolve these limitations, we propose a novel testing procedure based on the asymptotic equivalence of distinct kurtosis estimates. This approach exploits the fact that for an elliptical distribution, estimators derived from different spectral moments converge to a unique kurtosis parameter, whereas any significant discrepancy between them indicates a departure from ellipticity. While comparing estimators is a well-established concept in low-dimensional settings, the theoretical analysis in the high-dimensional regime is fundamentally different and more challenging, requiring new tools to handle the bias inherent in higher-order sample moments. We derive the asymptotic distribution of our test statistic, providing a robust diagnostic tool that does not require matrix inversion.

Second, under the assumption of ellipticity, there is a need for reliable inferential methods for spectral statistics of the form $T_n = \psi(\lambda_1(\hat{\Sigma}_n), \dots, \lambda_p(\hat{\Sigma}_n))$, where $\lambda_i(\hat{\Sigma}_n)$ are the sorted eigenvalues of the sample covariance matrix and ψ is a generic real-valued function. Such statistics are central to many classical methods in multivariate analysis (Muirhead, 1982). While the limiting distributions of some of these statistics have been analyzed (Ding and Xie, 2023; Hu et al., 2019a; Jun et al., 2022; Zhang et al., 2022), the resulting asymptotic formulas typically involve complex integrals and Stieltjes transforms that are often analytically intractable or computationally unstable. Furthermore, limiting laws for many specific spectral functions remain underived. To circumvent these issues, and given that the standard nonparametric bootstrap is generally inconsistent in the high-dimensional regime, we propose a parametric bootstrap procedure that leverages the validated

elliptical structure. We show that this method accurately reproduces the fluctuation patterns of eigenvalues that the naive bootstrap fails to capture. Rigorous theoretical guarantees are provided for a broad class of spectral functions ψ , and the result is a versatile and stable inference tool for high-dimensional spectral analysis.

CHAPTER 2

Testing Elliptical Models in High Dimensions

2.1. Introduction

Based on the extensive use of elliptical models, the problem of validating them with goodness-of-fit tests has become a long-standing topic of research (e.g. Albisetti et al., 2020; Babić et al., 2021; Beran, 1979; Ducharme and de Micheaux, 2020; Huffer and Park, 2007; Koltchinskii and Sakhanenko, 2000; Manzotti et al., 2002; Schott, 2002; Tang and Li, 2023). However, this literature is generally confined to low-dimensional settings—often due to the use of the inverse sample covariance matrix in the construction of goodness-of-fit tests. Likewise, several authors have recently drawn attention to the need for new approaches to goodness-of-fit testing that can handle elliptical and related models in high dimensions (Chen and Xia, 2023; Lin et al., 2023; Schkoda and Drton, 2025; Tang and Li, 2023). Yet, to the best of our knowledge, there are no established goodness-of-fit tests for elliptical models that are theoretically supported in high dimensions. Indeed, even in the simpler case of normal models, theoretical guarantees for goodness-of-fit tests in high dimensions are at the cutting edge of research (Chen and Xia, 2023).

In the current paper, we seek to rectify this situation by developing a new goodness-of-fit test for elliptical models and establishing its asymptotic validity when the dimension and sample size diverge proportionally. To outline our formulation of the problem, we deal with a set of centered i.i.d. observations $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$ drawn from an unknown distribution $\mathbb{P}$, and the goal is to use the observations to test

$$(2.1) \quad H_0 : \mathbb{P} \in \mathcal{P} \quad \text{versus} \quad H_1 : \mathbb{P} \notin \mathcal{P},$$

where $\mathcal{P}$ is a class of elliptical distributions. To account for the role of high dimensionality and allow p to increase as a function of n , the problem (2.1) should be understood more precisely as representing a sequence of hypothesis testing problems implicitly indexed by n , in which $\mathbb{P}$ and $\mathcal{P}$ both depend on n . The exact definition of the sequence of classes $\mathcal{P}$ will be given later in

Assumption 1, but in brief, the associated elliptical distributions will only be subject to some simple moment conditions and will be allowed to have *unrestricted covariance matrices*.

The idea of the proposed test is based on two special properties of the moment structure of elliptical distributions. First, if $\mathbf{x}_1 = (x_{11}, \dots, x_{1p})$ is drawn from an elliptical distribution such that each covariate is non-degenerate and has finite kurtosis, then all of the covariates must have the same kurtosis, denoted by κ . In particular, when $\mathbf{x}_1$ is centered, this means $\kappa = \mathbb{E}(x_{1j}^4)/(\mathbb{E}(x_{1j}^2))^2$ for all $j = 1, \dots, p$, as recorded in Lemma D.1. Second, it also turns out that κ can be independently calculated in a “coordinate-free” way (based on Lemma D.1), which leads to the following system of p simultaneous equations that hold for any centered elliptical distribution with non-degenerate covariates and $\kappa < \infty$,

$$(2.2) \quad \kappa = \frac{\mathbb{E}(x_{1j}^4)}{(\mathbb{E}(x_{1j}^2))^2} = \frac{3(\text{var}(\|\mathbf{x}_1\|_2^2) + \text{tr}(\Sigma)^2)}{2\text{tr}(\Sigma^2) + \text{tr}(\Sigma)^2} \quad \text{for all } j = 1, \dots, p,$$

where Σ denotes the covariance matrix of $\mathbf{x}_1$. The equations above will lead us to construct two distinct kurtosis estimates, denoted $\tilde{\kappa}$ and $\check{\kappa}$, and our strategy is to “play them off of each other” in designing a goodness-of-fit test. That is, our proposed test statistic will measure the discrepancy between $\tilde{\kappa}$ and $\check{\kappa}$, and the null hypothesis will be rejected when this discrepancy is sufficiently large. So, in essence, the proposed test seeks to detect violations of equations (2.2), and this approach is well suited to the high-dimensional setting because larger numbers of covariates provide more opportunities to detect violations. Furthermore, *the test can even detect non-elliptical distributions when all of the covariates have the same kurtosis*. This is because one of the kurtosis estimates is based on a structural property that generally does not hold for non-elliptical distributions, and our simulation results for Model (4) in Section 2.4.2 demonstrate the power of our test in such cases.

From a theoretical standpoint, our main result (Theorem 2.3.1) is a central limit theorem for the proposed test statistic under the null hypothesis. One of the main virtues of this result is that it allows the population covariance matrix to vary in an unrestricted manner as n and p diverge. An interesting theoretical consequence of this flexibility is that the difference $\tilde{\kappa} - \check{\kappa}$ may fluctuate over many different scales, depending on the correlation structure of the data. This is in contrast to more conventional high-dimensional central limit theorems, in which an unnormalized statistic often has fluctuations of the same order (such as $1/\sqrt{n}$ or $1/p$) for all of the allowed choices of the population covariance matrix. Accordingly, constructing an accurate estimate for the standard

deviation of $\tilde{\kappa} - \check{\kappa}$ is especially crucial in our method, and proving the consistency of this estimate is a substantial undertaking (see Appendix A.3).

A key technical obstacle in proving Theorem 2.3.1 is bounding sums that have the form

$$\sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(x_{1j_1}^{k_1} x_{1j_2}^{k_2} x_{1j_3}^{k_3} x_{1j_4}^{k_4})$$

for exponents $k_1, \dots, k_4 \in \{0, 2, 4\}$. In order to analyze such sums, we develop a systematic approach based on *Isserlis' theorem* (Janson, 1997), also known as Wick's theorem, which is a powerful tool for calculating mixed moments of normal random vectors that is based on the combinatorial framework of pair partitions. Because the statement of Isserlis' theorem is only applicable to normal random vectors, an important merit of our analysis is that we are able to successfully adapt it to the setting of elliptical random vectors. Furthermore, even in the special case of normal data, Isserlis' theorem produces a formula for each mixed moment $\mathbb{E}(x_{1j_1}^{k_1} x_{1j_2}^{k_2} x_{1j_3}^{k_3} x_{1j_4}^{k_4})$ that may involve an extremely large number of terms. To deal with this issue, we introduce an equivalence relation among the terms that allows us to decompose the moment calculations across a tractable number of equivalence classes. (See the proof of Proposition A.2.3 in Appendix A.2.)

With regard to numerical performance, we show through simulations that the proposed test has an empirical level that nearly matches the nominal level when the data are generated from elliptical distributions in 60 distinct parameter settings. Also, the test is able to reject the null hypothesis even when the data-generating distribution differs from elliptical distributions by small perturbations. For comparison, we show that the proposed test has substantially higher power than a recently developed normality test (Chen and Xia, 2023), which seems to be the only such test whose asymptotic validity has been established in high dimensions. (Such a comparison of power is justified because both tests are intended to reject in the presence of non-elliptical data.) Lastly, we present some examples involving natural datasets, illustrating that for commonly used significance levels, the proposed test is able to reject, whereas the normality test is not. Thus, our approach is of independent interest in connection with the topic of high-dimensional normality testing.

Notation and conventions. For any real number $q \geq 1$, the entrywise ℓ_q norm of a real matrix A is $\|A\|_q = (\sum_{i,j} |A_{ij}|^q)^{1/q}$, and the L^q norm of a scalar random variable ζ is $\|\zeta\|_{L^q} = (\mathbb{E}(|\zeta|^q))^{1/q}$. For two sequences of real numbers $\{a_n\}$ and $\{b_n\}$, the relation $a_n = o(b_n)$ means $a_n/b_n \rightarrow 0$ as

$n \rightarrow \infty$. When $\{a_n\}$ and $\{b_n\}$ are non-negative, we write $a_n \lesssim b_n$ if there is a constant $C > 0$ not depending on n such that $a_n \leq Cb_n$ holds for all large n . When both of the relations $a_n \lesssim b_n$ and $b_n \lesssim a_n$ hold, we write $a_n \asymp b_n$. The maximum and minimum of a pair of real numbers a and b are sometimes written as $\max\{a, b\} = a \vee b$ and $\min\{a, b\} = a \wedge b$. Convergence in probability and convergence in distribution are respectively denoted by $\xrightarrow{\mathbb{P}}$ and $\Rightarrow$. Lastly, certain formulas in our testing procedure involve fractions of random variables in which the denominator may vanish. To make the procedure well defined in any finite-sample situation, a fraction with a zero denominator is arbitrarily defined to be equal to one. This convention is of minor importance, because it is only relevant for events that occur with asymptotically negligible probability, and hence, does not affect the limiting null distribution of the proposed test statistic.

2.2. Method

Recall from our discussion of (2.2) that the proposed test seeks to detect violations of the null hypothesis by measuring the discrepancy between two distinct estimates $\tilde{\kappa}$ and $\check{\kappa}$ for the kurtosis parameter κ . These two estimates will be constructed using separate halves of the data, $\mathbf{x}_1 \dots, \mathbf{x}_{n/2}$ and $\mathbf{x}_{n/2+1} \dots, \mathbf{x}_n$, and we assume henceforth that n is even for simplicity. Since the left side of (2.2) asserts that κ is equal to the entrywise kurtosis $\mathbb{E}(x_{1j}^4)/\mathbb{E}(x_{1j}^2)^2$ for all $j = 1, \dots, p$, it is natural to construct the first estimate $\tilde{\kappa}$ by averaging the entrywise kurtosis estimates $\tilde{\kappa}_j = \frac{1}{n/2} \sum_{i=1}^{n/2} x_{ij}^4 / (\frac{1}{n/2} \sum_{i=1}^{n/2} x_{ij}^2)^2$ over $j = 1, \dots, p$, which gives

$$\tilde{\kappa} = \frac{1}{p} \sum_{j=1}^p \tilde{\kappa}_j.$$

To define the second estimate $\check{\kappa}$, it is helpful to introduce the parameters $\varsigma^2 = \text{var}(\|\mathbf{x}_1\|_2^2)$ and $\nu_k = \text{tr}(\Sigma^k)$ for $k \geq 1$, so that the right side of (2.2) can be rewritten as

$$(2.3) \quad \kappa = \frac{3(\varsigma^2 + \nu_1^2)}{\nu_1^2 + 2\nu_2}.$$

From this equation, we construct $\check{\kappa}$ by substituting in estimates of ς^2 , ν_1 and ν_2 . For estimating the latter two parameters, let $\check{\Sigma} = \frac{1}{n/2} \sum_{i=n/2+1}^n \mathbf{x}_i \mathbf{x}_i^\top$, as well as $\check{\nu}_k = \text{tr}(\check{\Sigma}^k)$ for $k \geq 1$. Specifically, we estimate ν_1 using $\check{\nu}_1$, and we estimate ν_2 using $\check{\nu}_2 - \frac{1}{n/2} \check{\nu}_1^2$, which is a bias-corrected version of $\check{\nu}_2$. Next, we estimate ς^2 using the sample variance of the squared norms $\|\mathbf{x}_i\|_2^2$ for $i = n/2 + 1, \dots, n$,

which is

$$\zeta^2 = \frac{1}{2 \binom{n/2}{2}} \sum_{n/2 < i < i' \leq n} (\|\mathbf{x}_i\|_2^2 - \|\mathbf{x}_{i'}\|_2^2)^2.$$

Hence, the second estimate of κ is defined by

$$\check{\kappa} = \frac{3(\zeta^2 + \check{\nu}_1^2)}{\check{\nu}_1^2 + 2(\check{\nu}_2 - \frac{2}{n}\check{\nu}_1^2)}.$$

To compare $\tilde{\kappa}$ and $\check{\kappa}$, it will be technically convenient to work with a shifted and rescaled version of the difference $\tilde{\kappa} - \check{\kappa}$, denoted by

$$T_n = \sqrt{\frac{pn}{2}} \left(\frac{\tilde{\kappa} - \check{\kappa}}{3} + \frac{4}{n} \right),$$

where the term $4/n$ merely ensures that T_n is asymptotically centered. Finally, the statistic T_n must be normalized by an estimate of its standard deviation, which is denoted by $\hat{\sigma}_n$ and will be defined below in Section 2.2.1. So altogether, the proposed test statistic is $T_n/\hat{\sigma}_n$, and our main theoretical result in Theorem 2.3.1 will show that under the null hypothesis, it satisfies $T_n/\hat{\sigma}_n \Rightarrow N(0, 1)$ as $n \rightarrow \infty$ with $p \asymp n$. Accordingly, the proposed testing procedure rejects at a nominal level of $\alpha \in (0, 1)$ if the event

$$R_n(\alpha) = \left\{ |T_n| > \hat{\sigma}_n \Phi^{-1}(1 - \alpha/2) \right\}$$

occurs, where Φ is the standard normal distribution function.

2.2.1. Estimating the variance of T_n . We will show in Proposition A.2.1 of Appendix A.2 that the variance of T_n is asymptotically equivalent to a parameter $\sigma_n^2 = \sigma_{n,1}^2 + \sigma_{n,2}^2$ defined by

$$(2.4) \quad \sigma_{n,1}^2 = \frac{8\|R\|_4^4}{3p} \quad \text{and} \quad \sigma_{n,2}^2 = \frac{8p(2\nu_4 + \nu_2^2)}{(\nu_1^2 + 2\nu_2)^2},$$

where R is the correlation matrix of $\mathbf{x}_1$. To write down an estimate for σ_n^2 , let $\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top$, as well as $\hat{\nu}_k = \text{tr}(\hat{\Sigma}^k)$ for $k \geq 1$, and $\hat{R}_{jj'} = \hat{\Sigma}_{jj'} / (\hat{\Sigma}_{jj} \hat{\Sigma}_{j'j'})^{1/2}$ for $1 \leq j, j' \leq p$. Then, we define $\hat{\sigma}_n^2 = \hat{\sigma}_{n,1}^2 + \hat{\sigma}_{n,2}^2$ according to

$$(2.5) \quad \hat{\sigma}_{n,1}^2 = \frac{8}{3p} \left(\|\hat{R}\|_4^4 - \hat{c} \right) \quad \text{and} \quad \hat{\sigma}_{n,2}^2 = \frac{8p(2\hat{\nu}_4 + (\hat{\nu}_2 - \frac{1}{n}\hat{\nu}_1^2)^2)}{(\hat{\nu}_1^2 + 2(\hat{\nu}_2 - \frac{1}{n}\hat{\nu}_1^2))^2},$$

where $\hat{c}$ is a correction term whose definition requires some additional notation: For any real number x and positive number t , define the threshold function $\llbracket x \rrbracket_t = \text{sgn}(x)(|x| \wedge t)$. Also, for any fixed symmetric matrix $M \in \mathbb{R}^{p \times p}$ and integer $k \geq 1$, define the function $g_k(M) = \mathbb{E}((\mathbf{z}^\top M \mathbf{z})^k)$, where $\mathbf{z}$ is a standard normal vector in $\mathbb{R}^p$. (Explicit formulas for this function are provided in Lemma D.5.) In this notation, the correction term $\hat{c}$ is given by

$$(2.6) \quad \hat{c} = \hat{\beta} \|\hat{R}\|_4^4 - 3 \llbracket 1 - \hat{\beta} + \hat{\gamma} \rrbracket_{t_p} \|\hat{R}\|_2^2,$$

where $t_p = p^{-3/4} \log(p)$, and

$$(2.7) \quad \hat{\beta} = 1 - \frac{\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|_2^8}{g_4(\hat{\Sigma}) - \frac{12}{n}(\hat{\nu}_1^4 + 2\hat{\nu}_1^2\hat{\nu}_2 - \frac{1}{n}\hat{\nu}_1^4)}$$

$$(2.8) \quad \hat{\gamma} = 1 + \frac{\hat{\zeta}^2 - 2(\hat{\nu}_2 - \frac{1}{n}\hat{\nu}_1^2)}{g_2(\hat{\Sigma}) - \frac{2}{n}\hat{\nu}_1^2} - \frac{\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|_2^6}{\frac{1}{2}g_3(\hat{\Sigma}) - \frac{3}{n}\hat{\nu}_1^3}.$$

The definition of the correction term in (2.6) is motivated by an asymptotic formula for $\sigma_{n,1}^2$ that results from (A.28) and (A.29) in the proof of Lemma B.4. In particular, it can be shown that the term $3 \llbracket 1 - \hat{\beta} + \hat{\gamma} \rrbracket_{t_p} \|\hat{R}\|_2^2$ is asymptotically negligible, but it turns out that retaining this term is beneficial at a finite-sample level. Finally, when interpreting the formulas above, recall that in the unlikely event where the denominator of a fraction vanishes, we arbitrarily define the fraction to be equal to one.

2.3. Theory

Our analysis will be done with a sequence of models implicitly embedded in a triangular array, where the rows are indexed by n , and all parameters may vary with n except when stated otherwise. The following assumption specifies the sequence of null hypotheses in (2.1) corresponding to different values of n .

ASSUMPTION 1. *As $n \rightarrow \infty$, the dimension p grows such that $p \asymp n$. For each $i = 1, \dots, n$, the observation $\mathbf{x}_i \in \mathbb{R}^p$ has the form*

$$\mathbf{x}_i = \xi_i \Sigma^{1/2} \mathbf{u}_i,$$

where $\Sigma \in \mathbb{R}^{p \times p}$ is deterministic and positive semidefinite with positive diagonal entries, and $(\mathbf{u}_1, \xi_1), \dots, (\mathbf{u}_n, \xi_n)$ are i.i.d. random vectors in $\mathbb{R}^p \times [0, \infty)$ satisfying the following conditions:

The vector $\mathbf{u}_1$ is drawn from the uniform distribution on the unit sphere of $\mathbb{R}^p$, and is independent of ξ_1 . In addition, the random variable ξ_1^2 satisfies $\mathbb{E}(\xi_1^2) = p$, as well as the conditions

$$(2.9) \quad \text{var}\left(\frac{\xi_1^2 - p}{\sqrt{p}}\right) = \tau + o(1) \quad \text{and} \quad \left\|\frac{\xi_1^2 - p}{\sqrt{p}}\right\|_{L^8} = o(p^{1/4})$$

as $n \rightarrow \infty$, for some constant $\tau \geq 0$ that is fixed with respect to n .

Remarks. Perhaps the most striking feature of Assumption 1 is that the population covariance matrix Σ is unrestricted. (The condition that Σ has positive diagonal entries does not impose any real restriction, because if $\Sigma_{jj} = 0$, then the j th covariate is zero almost surely and is hence irrelevant.) Also, the model only places conditions on a finite number of moments. Similar moment assumptions have been considered elsewhere, such as in the papers (Hu et al., 2019a,c), which analyze eigenvalue statistics of sample covariance matrices in high-dimensional elliptical models. In essence, the purpose of the two conditions in (2.9) is to make it possible to precisely determine the magnitudes of higher-order remainder terms in the proofs of certain CLTs underlying Theorem 2.3.1 below. It should be noted too that the second condition in (2.9) allows for the eighth moment of $(\xi_1^2 - p)/\sqrt{p}$ to diverge with the dimension, i.e. $\|(\xi_1^2 - p)/\sqrt{p}\|_{L^8} \rightarrow \infty$ as $p \rightarrow \infty$. In fact, this condition can be weakened to allow for even faster divergence as $p \rightarrow \infty$, but it would involve a somewhat more complicated statement of Assumption 1, and so we have retained the version above to keep the statement as concise as possible.

Examples. Below, we provide an assortment of examples for the distribution of ξ_1^2 that conform with Assumption 1. In each case, the distribution is parameterized so that the basic normalization condition $\mathbb{E}(\xi_1^2) = p$ holds.

- Chi-Squared on p degrees of freedom,
- Poisson(p),
- $(1 - \tau)$ Negative-Binomial($p, 1 - \tau$), for any fixed $\tau \in (0, 1)$,
- $(p + 2b)$ Beta($p/2, b$), for any fixed $b > 0$,
- Gamma($p/\tau, 1/\tau$), for any fixed $\tau > 0$,
- Beta-Prime $\left(\frac{p(1+p+\tau)}{\tau}, \frac{1+p+2\tau}{\tau}\right)$, for any fixed $\tau > 0$,
- Log-Normal $\left(\log(p) - \frac{1}{2} \log\left(1 + \frac{\tau}{p}\right), \log\left(1 + \frac{\tau}{p}\right)\right)$, for any fixed $\tau > 0$.

Several of these examples for ξ_1^2 can be subsumed within a more general non-parametric class of distributions that satisfy Assumption 1. This class corresponds to choices of ξ_1^2 that can be represented in the form $\xi_1^2 = \sum_{j=1}^p \eta_j$ for some independent random variables $\eta_1, \dots, \eta_p \geq 0$ satisfying

$$\frac{1}{p} \sum_{j=1}^p \mathbb{E}(\eta_j) = 1, \quad \frac{1}{p} \sum_{j=1}^p \text{var}(\eta_j) = \tau + o(1), \quad \text{and} \quad \max_{1 \leq j \leq p} \mathbb{E}(\eta_j^8) = o(p^5),$$

again with $\tau \geq 0$ being a constant not depending on n . (This can be verified using Rosenthal's inequality, as given in Lemma D.7 of the supplement.) It is also notable that the moments $\mathbb{E}(\eta_j^8)$ are allowed to grow quite rapidly as $p \rightarrow \infty$.

The following theorem is our main theoretical result.

THEOREM 2.3.1. *If Assumption 1 holds, then as $n \rightarrow \infty$ we have*

$$\frac{T_n}{\hat{\sigma}_n} \Rightarrow N(0, 1),$$

and hence, for any fixed $\alpha \in (0, 1)$, the level of the test satisfies

$$\mathbb{P}(\mathbf{R}_n(\alpha)) \rightarrow \alpha.$$

Remarks. At a high level, the proof of Theorem 2.3.1 involves several main parts. The first two are given in Propositions A.2.1 and A.3.1 of the supplement, which respectively show that $T_n/\sigma_n \Rightarrow N(0, 1)$ and $\hat{\sigma}_n/\sigma_n \xrightarrow{\mathbb{P}} 1$ as $n \rightarrow \infty$. The problem of showing that T_n/σ_n is asymptotically normal is then divided into two more substantial parts, corresponding to separate CLTs for standardized versions of the estimates $\tilde{\kappa}$ and $\tilde{\kappa}$, which utilize different sets of theoretical tools.

The CLT for $\tilde{\kappa}$ involves analyzing a sum of i.i.d. random variables $\sum_{i=1}^{n/2} w_i$, where $w_i = \frac{1}{\sqrt{p}} \sum_{j=1}^p \frac{1}{3}(x_{ij}^4 - \mathbb{E}(x_{ij}^4)) - 2(x_{ij}^2 - \mathbb{E}(x_{ij}^2))$. We establish asymptotic normality by showing that a Lyapunov condition of the form $\mathbb{E}(w_1^4) = o(n \text{var}(w_1)^2)$ holds, which in turn, requires bounding sums of the form $\sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(x_{1j_1}^{k_1} x_{1j_2}^{k_2} x_{1j_3}^{k_3} x_{1j_4}^{k_4})$ for exponents $k_1, \dots, k_4 \in \{0, 2, 4\}$. Although the random vector $\mathbf{x}_1$ is not necessarily normal under the null hypothesis, we are able to control its mixed moments by adapting Isserlis' theorem (Janson, 1997), which states that the following formula holds for any dimension d and any centered normal vector $(g_1, \dots, g_d)$,

$$(2.10) \quad \mathbb{E}(g_1 \cdots g_d) = \sum_{\pi \in \mathcal{P}_d} \prod_{\{j, j'\} \in \pi} \mathbb{E}(g_j g_{j'}),$$

where $\mathcal{P}_d$ is the collection of all partitions π of $\{1, \dots, d\}$ into distinct pairs $\{j, j'\}$. Here, it is important to note that even if the vector $\mathbf{x}_1$ was assumed to be normal, a brute-force application of Isserlis' theorem would not provide a workable solution to establishing the Lyapunov condition, because of the enormous number of terms on the right side of (2.10). For instance, in the special case when $\mathbf{x}_1$ is normal and $k_1 = \dots = k_4 = 4$, there are more than 2 million terms in the formula for $\mathbb{E}(x_{1j_1}^{k_1} x_{1j_2}^{k_2} x_{1j_3}^{k_3} x_{1j_4}^{k_4})$, since $d = \sum_{l=1}^4 k_l$ and $\text{card}(\mathcal{P}_{16}) = 15 \cdot 13 \cdot 11 \cdot \dots \cdot 1$. Nevertheless, we are able to overcome this difficulty by taking advantage of symmetries among the partitions. To explain the how this works, let $t_\pi(j_1, j_2, j_3, j_4)$ denote the term corresponding to the partition π in the formula for $\mathbb{E}(x_{1j_1}^4 x_{1j_2}^4 x_{1j_3}^4 x_{1j_4}^4)$ so that interchanging the order of summation gives

$$\sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(x_{1j_1}^4 x_{1j_2}^4 x_{1j_3}^4 x_{1j_4}^4) = \sum_{\pi \in \mathcal{P}_{16}} \sum_{j_1, j_2, j_3, j_4=1}^p t_\pi(j_1, j_2, j_3, j_4).$$

If we define an equivalence relation on $\mathcal{P}_{16}$ by saying that π and π' are equivalent if $t_\pi(j_1, j_2, j_3, j_4)$ and $t_{\pi'}(j_1, j_2, j_3, j_4)$ produce the same value when summed over all j_1, j_2, j_3, j_4 , then remarkably, it turns out that there is only a handful of equivalence classes. Hence, the sum over $\mathcal{P}_{16}$ that contains more than 2 million terms can be reduced to a tractable calculation, making it possible to establish the asymptotic normality of $\tilde{\kappa}$. (Strictly speaking, the technical aspects of this strategy are implemented somewhat differently in the proof of Proposition A.2.3, and there are additional considerations required to find the sizes of the different equivalence classes, but this explanation conveys the main idea.)

The asymptotic normality of the second estimate $\tilde{\kappa}$ is proven in an entirely different way, by adapting classical results on U-statistics to the high-dimensional setting. For this purpose, Propositions A.2.4 and A.2.5 show that the behavior of $\tilde{\kappa}$ is largely determined by the statistic $U = \xi^2 - 2(\tilde{\nu}_2 - \frac{2}{n}\tilde{\nu}_1^2)$, which is shown to be a U-statistic in the proof of Lemma B.8. There are two aspects of the classical theory of U-statistics that break down if they are used to prove a CLT for U in our setup. The first is in showing that U is asymptotically equivalent to its Hájek projection $\hat{U}$, and the second is in showing that $\hat{U}$ is asymptotically normal. In essence, the issue in the high-dimensional setting is that we must account for the fact that $\text{var}(U)$ and $\text{var}(\hat{U})$ may vary in their orders of magnitude as $p \rightarrow \infty$, and this is accomplished in Lemmas B.8-B.11 through a delicate analysis of the covariances of random quadratic forms.

2.4. Numerical experiments

In this section, we study the level and power of the proposed test under a substantial range of simulation settings. In addition, we offer comparisons with a high-dimensional normality test recently developed in (Chen and Xia, 2023), as well as some examples involving natural datasets.

2.4.1. Assessment of level.

Design of experiments. For the simulations under the null hypothesis, we generated data from elliptical models based on five choices of the distribution of the squared random variable ξ_1^2 , and four choices of the covariance matrix Σ . The choices for ξ_1^2 satisfy $\mathbb{E}(\xi_1^2) = p$ and are given by:

- (i). Chi-Squared distribution with p degrees of freedom,
- (ii). Beta-Prime $\left(\frac{p(1+p+3)}{3}, \frac{1+p+6}{3}\right)$,
- (iii). $(p+4)\text{Beta}(p/2, 2)$,
- (iv). Gamma($p/5, 1/5$),
- (v). $\frac{1}{p+1}\text{Gamma}(p, 1)^2$.

The choices for the covariance matrix Σ are:

- (1). (Spiked eigenvalues with generic eigenvectors). The eigenvalues of Σ are $\lambda_j(\Sigma) = 5$ for $j = 1, \dots, 5$ and $\lambda_j(\Sigma) = 1$ for $j = 6, \dots, p$. The $p \times p$ matrix of eigenvectors of Σ is drawn from the uniform distribution on orthogonal matrices.
- (2). (Toeplitz). The matrix Σ has entries of the form $\Sigma_{ij} = \rho^{|i-j|}$ with $\rho = 0.1$.
- (3). (Decaying eigenvalues with generic eigenvectors). The eigenvalues of Σ are $\lambda_j(\Sigma) = j^{-1/4}$, and the eigenvectors are the same as in case (1).
- (4). The covariance matrix Σ is the identity matrix.

The dimension was selected from $p \in \{200, 400, 600\}$, and the sample size was chosen to be $n = 400$. In each possible setting of the tuple (ξ_1^2, Σ, n, p) , we generated 10000 datasets, and we applied the proposed test at a 5% nominal level to each dataset. Hence, in total, the experiments presented here cover 60 different settings of the null hypothesis. Furthermore, in Appendix A.5, we present additional results in cases where the dimension is even larger, for values of the ratio p/n as large as 10.

Discussion of results. In Table 2.1, we report the rejection rate (i.e. empirical level) based on the 10000 trials under each elliptical distribution. For example, the entry in the lower left corner

of Table 2.1 shows that in settings (v) and (1) with $p/n = 0.5$, the null hypothesis was rejected in 4.52% of the trials.

TABLE 2.1. Results on empirical level for the proposed test (5% nominal level).

	$p/n = 0.5$				$p/n = 1$				$p/n = 1.5$			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
(i)	4.06	4.05	3.94	4.00	3.92	4.05	4.12	4.04	3.91	4.05	3.91	3.99
(ii)	4.26	3.96	3.96	4.01	4.04	4.04	4.16	3.93	3.78	3.79	4.00	3.82
(iii)	4.20	4.16	4.31	4.17	4.49	4.23	4.38	4.32	4.11	4.16	3.96	4.17
(iv)	5.27	4.79	4.78	4.61	4.19	4.25	4.28	4.28	4.11	4.15	3.92	4.14
(v)	4.52	4.43	4.71	4.25	4.03	4.10	4.22	4.11	3.95	4.12	3.99	3.98

In all settings, the empirical level is within approximately 1% of the nominal level. It is also notable that the difference between the empirical and nominal levels is reliably in the conservative direction, which is often preferable to a situation where the direction of disagreement is unpredictable or anti-conservative. Lastly, a comparison of the empirical level of our method and that of (Chen and Xia, 2023) is implicitly contained in the rejection rate curves at $h = 0$ in Figures 2.1-2.6, and this comparison is discussed below.

2.4.2. Assessment of power.

Design of experiments. To study the power of the proposed test, we generated data under the alternative hypothesis using models of the form $\mathbf{x}_1 = \Sigma^{1/2} \mathbf{s}_1$, where $\mathbf{s}_1 \in \mathbb{R}^p$ is a random vector whose entries are i.i.d random variables satisfying $\mathbb{E}(s_{11}) = 0$ and $\mathbb{E}(s_{11}^2) = 1$. For the choice of Σ , we used the same menu of options (1)-(4) described in Section 2.4.1. The “distance” between the null and alternative hypotheses is controlled through a parameter $h \in [0, 1]$, such that the entries of $\mathbf{s}_1$ are distributed according to

$$s_{11} = \sqrt{1-h} z_{11} + \sqrt{h} y_{11},$$

where z_{11} is a standard normal random variable and y_{11} is a standardized (a) Laplace or (b) Beta random variable. In detail, the choices of y_{11} correspond to

$$\begin{aligned} \text{(a). } y_{11} &\sim \frac{\text{Laplace}(0,1)}{\sqrt{2}}, \\ \text{(b). } y_{11} &\sim \frac{\text{Beta}(2,3/2)-4/7}{\sqrt{8/147}}. \end{aligned}$$

Note that if $h = 0$, then $\mathbf{x}_1$ has a $N(0, \Sigma)$ distribution, which is elliptical, and so the null hypothesis holds. By contrast, if $0 < h \leq 1$, then the alternative hypothesis holds. Another reason for generating the data in this manner is that it allows for convenient comparisons of both level and power with respect to the high-dimensional normality test proposed in (Chen and Xia, 2023). Such a comparison is legitimate, because both tests are intended to not reject the null hypothesis when $h = 0$, and reject it when $h > 0$.

As in Section 2.4.1, the dimension was selected from $p \in \{200, 400, 600\}$ and the sample size was chosen to be $n = 400$. For each possible choice of (y_{11}, Σ, n, p) , we generated 1000 datasets corresponding to several values of h in a grid within $[0, 1]$, and we performed both the proposed test and the normality test at a 5% nominal level on every dataset. (The normality test also involves two parameters, denoted L and B , which we set to the values $L = 1$ and $B = 500$ suggested in (Chen and Xia, 2023).) Finally, at each setting of $(y_{11}, \Sigma, n, p, h)$, we recorded the rejection rate of each test over the 1000 datasets, allowing the results to be visualized as curves depending on h , such that their heights at $h = 0$ represent the empirical level, and their heights at $h \in (0, 1]$ represent the empirical power.

Discussion of results. The results are displayed in Figures 2.1-2.6, which contain 24 panels corresponding to the different settings of (y_{11}, Σ, n, p) . In each panel, the rejection rates of the proposed test and the normality test are respectively plotted with blue and red curves. Overall, the most apparent difference between the tests is that they tend to produce curves with different shapes. The power of the proposed test has the intuitive property that it increases toward 1 as h increases, whereas the power of the normality test stays relatively flat as a function of h . (Recall that larger values of h correspond to data-generating distributions that are farther from being normal.) To fully appreciate the favorable power characteristics of the proposed test, it is worth emphasizing that normal models correspond to a strict subclass of elliptical models. So, in a heuristic sense, normality is an easier hypothesis to reject—and yet, the proposed test still manages to outperform.

In addition to the power of the tests, it is important to compare how well they respect the nominal level of 5%, which is marked with a dashed black line in every panel. In all settings, the empirical level of the proposed test at $h = 0$ is close to the desired value of 5%, whereas the empirical level of the normality is often substantially inflated, and in several cases it is nearly 20%. Due to this issue, the fact that the red curves are sometimes higher than the blue curves for small

values of h is not a disadvantage of the proposed test. In other words, the higher values of the red curves for small h occur simply because those curves start too high when $h = 0$.

Remark on covariates with equal kurtosis. Apart from the comparisons of power and level just discussed, a distinct point that deserves special emphasis is the power of the proposed test to detect non-elliptical distributions whose covariates have the same kurtosis. Considering that the test is motivated by the identity (2.2), which implies that all covariates in an elliptical model have the same kurtosis, it might seem surprising that the test can still be powerful against alternatives for which all covariates have the same kurtosis. Nevertheless, the results for Model (4) in Figures 2.1-2.6 show that the test achieves substantially higher power in such cases compared to the normality test. This behavior becomes more understandable once it is recognized that the kurtosis estimate $\hat{\kappa}$ used to construct the proposed test is designed based on a distinct structural property of elliptical models. Hence, it is less surprising that $\hat{\kappa}$ disagrees with the other kurtosis estimate $\tilde{\kappa}$ when data are generated from non-elliptical distributions whose covariates have the same kurtosis.

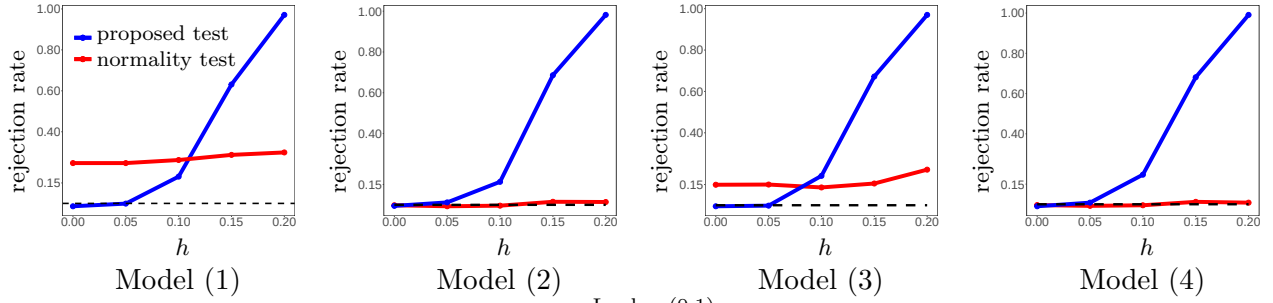


FIGURE 2.1. Results for $y_{11} \sim \frac{\text{Laplace}(0,1)}{\sqrt{2}}$ when $p/n = 0.5$ and $p = 200$.

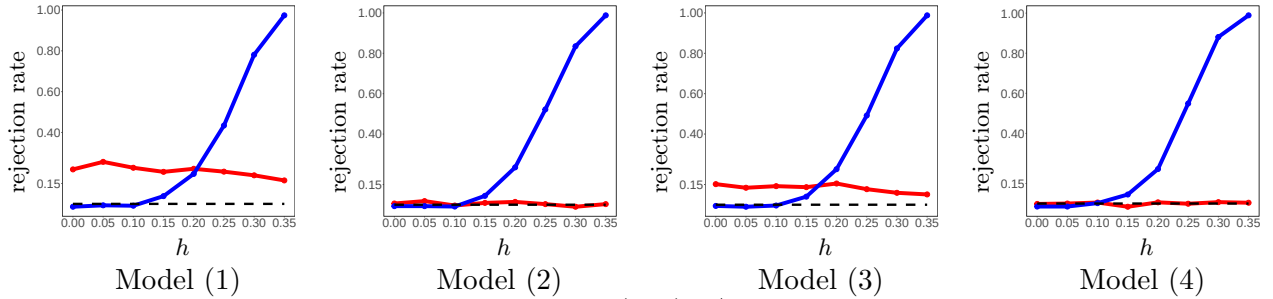
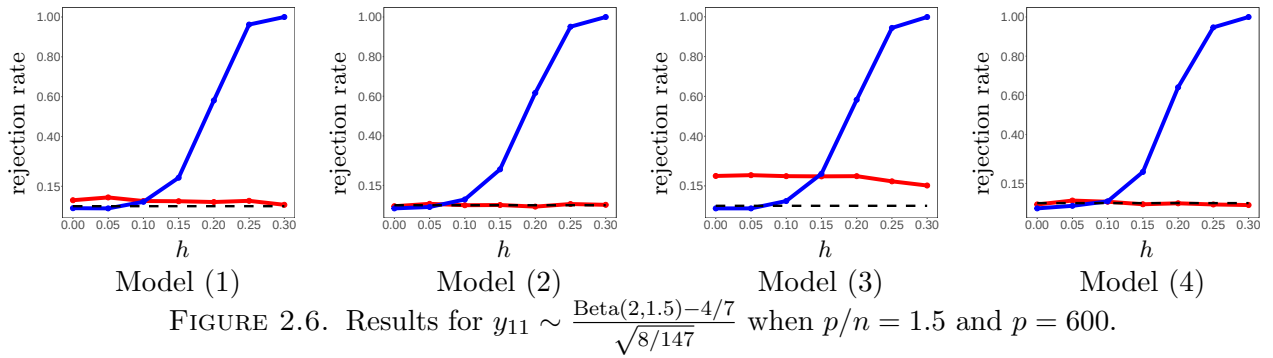
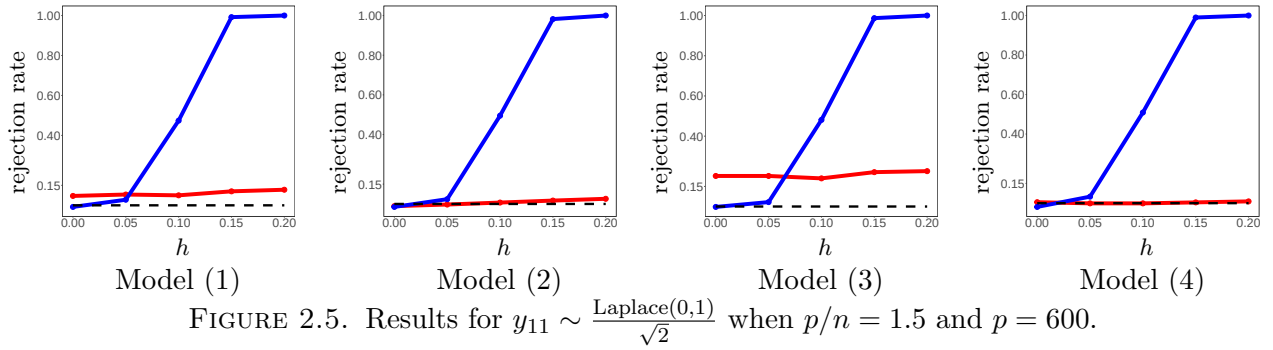
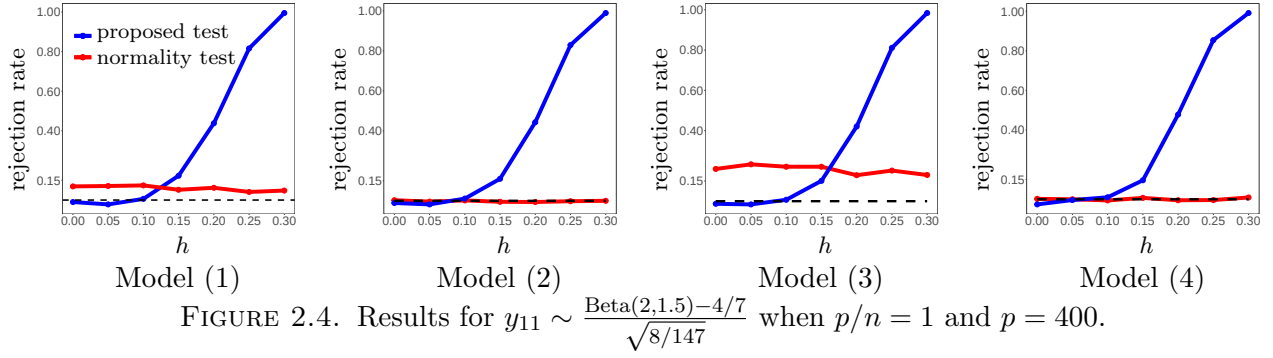
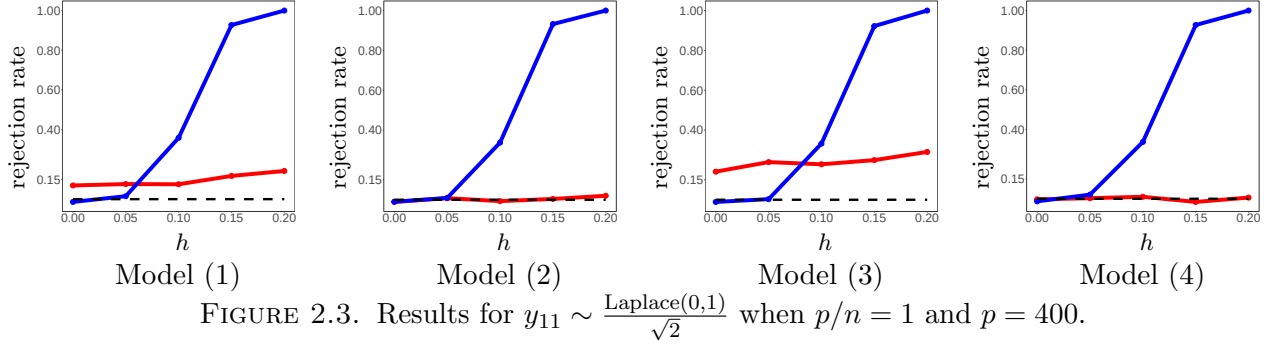


FIGURE 2.2. Results for $y_{11} \sim \frac{\text{Beta}(2,1.5)-4/7}{\sqrt{8/147}}$ when $p/n = 0.5$ and $p = 200$.



2.4.3. Stock market data. Elliptical distributions are commonly used in finance to model investment returns (Gupta et al., 2013; McNeil et al., 2015). Here, we apply the proposed test as well as the normality test from (Chen and Xia, 2023) to a collection of stock market return datasets with dimensions ranging from 100 to 480. The datasets were constructed from the monthly log returns of 480 stocks in the S&P 500 that had complete histories between July 2012 and June 2022, based on the records available at <https://finance.yahoo.com/lookup>. Hence, the initial “full” dataset may be regarded as a matrix of size 120×480 , based on $n = 120$ monthly observations in $p = 480$ dimensions. Next, we labeled the stocks according to the ranks of their prices at the beginning of July 2012, with the first stock having the highest price at that time. Lastly, for each number $d \in \{100, 110, 120, \dots, 480\}$, we applied both the proposed test and the normality test to the $n \times d$ data matrix corresponding to the first d columns of the full $n \times p$ matrix.

In Figure 2.7, we plot the resulting p-values as a function of d , with the proposed test shown in blue, and the normality test shown in red. For all choices of d , the proposed test produced p-values that are effectively 0. On the other hand, none of the p-values of the normality test fall below 0.05, and its p-values corresponding to $d \geq 160$ are all equal to 1. Thus, the results for this example conform with the power advantage of the proposed test that was demonstrated earlier using simulated data.

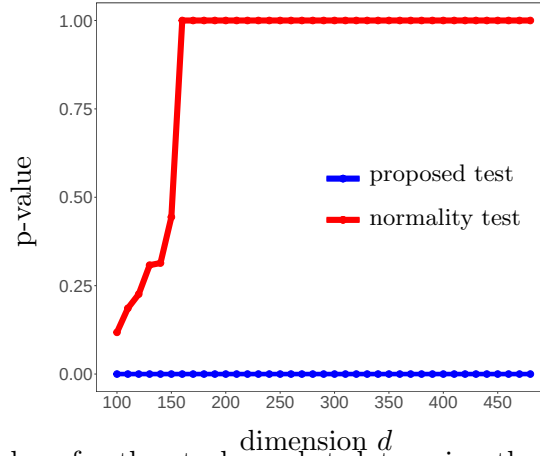


FIGURE 2.7. p-values for the stock market data using the proposed test and the normality test.

2.4.4. Breast cancer data. For our second example, we look at a breast cancer dataset that was studied in the influential paper van De Vijver et al. (2002) and is available in the R package `cancerdata` (Budczies and Kosztyla, 2022). The observations correspond to $n = 295$ tumor

samples with $p = 24481$ gene expression measures. We excluded the gene expression measures with missing values, resulting in $p = 22223$ gene expression measures. After centering these observations with the sample mean, we constructed 300 datasets in the following way. For each number $d \in \{200, 300, 500\}$, we randomly sampled d columns without replacement from the full $n \times p$ matrix and repeated this 100 times. So, in other words, for each choice of d , we generated 100 datasets of size $n \times d$ corresponding to random subsets of genes.

We applied the proposed test and the normality test to all the randomly sampled datasets and recorded the p-values. Then, for each choice of d , we computed the median of the 100 associated p-values for each test. The proposed test produced median p-values that were effectively 0 for each choice of d , whereas the normality test produced median p-values equal to 0.656, 1, and 1 in the respective cases of $d = 200, 300, 500$. Once again, the results of this example reflect the power advantage of the proposed test that was observed in the experiments with simulated data.

CHAPTER 3

A Bootstrap Method for Spectral Statistics in High-Dimensional Elliptical Models

3.1. Introduction

The analysis of spectral statistics of sample covariance matrices is a major research area within multivariate analysis, random matrix theory, high-dimensional statistics, and related fields (Bai and Silverstein, 2010; Potters and Bouchaud, 2020; Vershynin, 2018; Yao et al., 2015). If $\mathbf{x}_1, \dots, \mathbf{x}_n$ are centered i.i.d. observations in $\mathbb{R}^p$ with a sample covariance matrix denoted by

$$(3.1) \quad \hat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top,$$

then we say that a random variable T_n is a spectral statistic if it has the form $T_n = \psi(\lambda_1(\hat{\Sigma}_n), \dots, \lambda_p(\hat{\Sigma}_n))$, where $\lambda_1(\hat{\Sigma}_n) \geq \dots \geq \lambda_p(\hat{\Sigma}_n)$ are the sorted eigenvalues of $\hat{\Sigma}_n$, and ψ is a generic real-valued function. Over the past two decades, there has been a tremendous growth of interest in spectral statistics in high-dimensional settings where $p = p(n)$ grows so that p/n converges to a positive constant as $n \rightarrow \infty$. Likewise, statistical tools for approximating the distributions of spectral statistics have been applied to high-dimensional data in a broad range of domains, such as electrical engineering, finance, and biology (Akemann et al., 2011; Couillet and Debbah, 2011; Edelman and Wang, 2013; Patterson et al., 2006).

Although the research on spectral statistics has dealt with many different statistical models, two of the most influential ones have been *elliptical* and *independent components* (IC) models. To be specific, we say that the random vector $\mathbf{x}_1$ follows an elliptical model if it can be represented in the form

$$(3.2) \quad \mathbf{x}_1 = \xi_1 \Sigma_n^{1/2} \mathbf{u}_1$$

where $\Sigma_n = \mathbb{E}(\mathbf{x}_1 \mathbf{x}_1^\top)$, and $(\mathbf{u}_1, \xi_1) \in \mathbb{R}^p \times [0, \infty)$ is a random vector such that $\mathbf{u}_1$ and ξ_1 are independent, with $\mathbf{u}_1$ being uniformly distributed on the unit sphere of $\mathbb{R}^p$. Alternatively, we say that $\mathbf{x}_1$ follows an IC model if

$$(3.3) \quad \mathbf{x}_1 = \Sigma_n^{1/2} \mathbf{z}_1,$$

where $\mathbf{z}_1 \in \mathbb{R}^p$ is a random vector whose entries are centered and independent.

At first sight, these two models may seem to be very similar, but this outward appearance conceals some crucial differences in modelling capabilities. In particular, it should be stressed that the entries of the random vector $\xi_1 \mathbf{u}_1$ in an elliptical model are correlated, which contrasts with the independence of the entries of $\mathbf{z}_1$ in an IC model. Also, since the scalar random variable ξ_1 is shared across all entries of $\mathbf{x}_1$ in an elliptical model, this enhances the ability to capture scenarios where the magnitudes of all entries of $\mathbf{x}_1$ move in the same direction simultaneously. This is a key effect in some application domains, such as in finance, where the entries of $\mathbf{x}_1$ correspond to stock prices that can fall in tandem during a sharp market downturn. Additional background on related merits of elliptical models can be found in (El Karoui, 2009; Gupta et al., 2013; McNeil et al., 2015). More generally, the multivariate analysis literature has placed a longstanding emphasis on the benefits of elliptical models in fitting various types of non-Gaussian data (Anderson, 2003; Fang, 2018; Härdle and Simar, 2019).

However, looking beyond the points just mentioned, IC models have played a more dominant role than elliptical models in the literature on spectral statistics in high dimensions. Consequently, the established body of high-dimensional limit theory is much less complete for elliptical models. Indeed, the challenge of extending results from IC models to elliptical ones has become a prominent topic of ongoing research, which has led to important advances in the limit theory for spectral statistics (e.g. Hu et al., 2019a,c; Jun et al., 2022; Li et al., 2022; Li and Yao, 2018; Yang et al., 2021; Zhang et al., 2022). As a matter of historical context, it is also worth bearing in mind that for some spectral statistics, it took many years for such extensions to be established.

From the standpoint of statistical methodology, a corresponding set of gaps exists between elliptical and IC models. These gaps are especially apparent in the current state of bootstrap methods for high-dimensional data. In particular, it is known from (Lopes et al., 2019) that a form of parametric bootstrapping can successfully approximate the distributions of spectral statistics in

IC models, whereas very little is known for elliptical models. Accordingly, our primary goal in the current paper is to resolve this issue by developing a parametric bootstrap method that is both theoretically and empirically effective in high-dimensional elliptical models.

With regard to theory, we will focus on the class of *linear spectral statistics*, which have the form

$$(3.4) \quad T_n(f) = \frac{1}{p} \sum_{j=1}^p f(\lambda_j(\hat{\Sigma}_n)),$$

for a suitable real-valued function f . Beginning with the pathbreaking works (Bai and Silverstein, 2004; Jonsson, 1982) that established the earliest versions of the central limit theorem for linear spectral statistics in high dimensions, these statistics have been a perennial focus of research. Their importance is underscored by the fact that they appear frequently throughout multivariate analysis, with some of the most well-known examples being $\frac{1}{p}\text{tr}(\hat{\Sigma}_n)$, $\frac{1}{p}\text{tr}(\hat{\Sigma}_n^2)$, and $\frac{1}{p}\log \det(\hat{\Sigma}_n)$, among various other classical statistics for testing hypotheses (Yao et al., 2015).

Motivated by these considerations, our main theoretical result (Theorem 3.3.2) shows that the proposed bootstrap method consistently approximates the distributions of linear spectral statistics when the underlying data are elliptical and p/n converges to a positive constant as $n \rightarrow \infty$. The proof substantially leverages recent progress on the central limit theorem for linear spectral statistics in elliptical models due to (Hu et al., 2019a). Also, an intermediate step in the proof (Lemma A.4) shows that the well-known eigenvalue estimation method QuEST is consistent in elliptical models—which may be of independent interest, since it seems that QuEST’s consistency has not previously been reported outside of IC models (Ledoit and Wolf, 2015). Moreover, Section 3.3.4 develops an application of Theorem 3.3.2 where we establish the asymptotic validity of inference procedures related to the *stable rank* parameter $r_n = \text{tr}(\Sigma_n)^2 / \text{tr}(\Sigma_n^2)$ of the population covariance matrix Σ_n .

To address the empirical performance of the proposed method, Section 3.4 presents numerical results for a wide variety of model settings and statistics. Most notably, these results show encouraging performance for *both linear and nonlinear spectral statistics*. (We regard any function of $(\lambda_1(\hat{\Sigma}_n), \dots, \lambda_p(\hat{\Sigma}_n))$ that is not of the form (3.4) as a nonlinear spectral statistic.) To put this point into perspective, it is important to highlight the fact that asymptotic formulas for the distributions of nonlinear spectral statistics are typically developed on a case-by-case basis, and are relatively scarce in comparison to those for linear spectral statistics. Even when such formulas

are available, they may be very different from those for linear spectral statistics—as they may require the estimation of different model parameters, or the implementation of different algorithms for numerical evaluation. On the other hand, the bootstrap approach may be considered more user-friendly, since it can be applied to different types of spectral statistics in an automatic and unified way. Similarly, the bootstrap approach can provide the user with the freedom to easily explore statistics that depend on several linear spectral statistics in a complicated manner (which would otherwise require intricate delta-method calculations), or statistics for which formulas may not be available at all.

Notation and terminology. For a random object Y , the expression $\mathcal{L}(Y)$ denotes its distribution, and $\mathcal{L}(Y|X)$ denotes its conditional distribution given the observations $\mathbf{x}_1, \dots, \mathbf{x}_n$. Similarly, we use $|X$ when referring to probabilities, expectations, and variances that are conditional on the observations. The symbols $\xrightarrow{\mathbb{P}}$ and $\Rightarrow$ respectively denote convergence in probability and convergence in distribution. For a set $A \subset \mathbb{R}^k$ and a number $\delta > 0$, the outer δ -neighborhood of A is defined as $A^\delta = \{\mathbf{a}' \in \mathbb{R}^k \mid \inf_{\mathbf{a} \in A} \|\mathbf{a}' - \mathbf{a}\|_2 \leq \delta\}$, where $\|\cdot\|_2$ is the Euclidean norm. If $\mathbf{v}$ and $\mathbf{w}$ are random vectors in $\mathbb{R}^k$, then the Lévy-Prohorov metric $d_{\text{LP}}(\mathcal{L}(\mathbf{v}), \mathcal{L}(\mathbf{w}))$ between their distributions is defined as the infimum over all numbers $\delta > 0$ such that the inequality $\mathbb{P}(\mathbf{v} \in A) \leq \mathbb{P}(\mathbf{w} \in A^\delta) + \delta$ holds for all Borel sets $A \subset \mathbb{R}^k$. If $\nu, \nu_1, \nu_2, \dots$ is a sequence of random probability distributions on $\mathbb{R}^k$, then the expression $\nu_n \xrightarrow{\mathbb{P}} \nu$ means that the sequence of scalar random variables $d_{\text{LP}}(\nu_n, \nu)$ converges to 0 in probability as $n \rightarrow \infty$. For two sequences of non-negative real numbers $\{a_n\}$ and $\{b_n\}$, we write $a_n \lesssim b_n$ if there is a constant $C > 0$ not depending on n such that $a_n \leq Cb_n$ holds for all large n . When both of the relations $a_n \lesssim b_n$ and $b_n \lesssim a_n$ hold, we write $a_n \asymp b_n$. The relation $a_n = o(b_n)$ means $a_n/b_n \rightarrow 0$ as $n \rightarrow \infty$, and the relation $a_n = \mathcal{O}(b_n)$ is equivalent to $a_n \lesssim b_n$. The $k \times k$ identity matrix is denoted as I_k , and the indicator function for a condition $\dots$ is denoted as $1\{\dots\}$. Lastly, we use $\mathbb{C}^+$ to refer to the set of complex numbers with a positive imaginary part.

3.2. Method

Conceptually, the proposed method is motivated by the fact that the standard nonparametric bootstrap, based on sampling with replacement, often performs poorly when it is applied naively

to high-dimensional data, unless special low-dimensional structure is available. (For additional background, see the papers (El Karoui and Purdom, 2019; Lopes, 2022b; Lopes et al., 2023; Yao and Lopes, 2023), as well as the numerical results presented here at the end of Section 3.4.2.) This general difficulty can be understood by noting that sampling with replacement implicitly relies on the empirical distribution of the data as a substitute for the true data-generating distribution. In other words, the nonparametric bootstrap attempts to approximate a p -dimensional distribution in a fully non-parametric way, which can be challenging for even moderately large values of p . For this reason, alternative bootstrap methods that sample from parametric distributions have been advocated to improve upon the nonparametric bootstrap in high dimensions (e.g. Lopes et al., 2019; Mammen, 1993; Zhao and Candès, 2022), and this is the viewpoint that we pursue here.

3.2.1. Bootstrap algorithm. At an algorithmic level, the proposed method is built on top of two estimators. The first is an estimator $\hat{\zeta}_n^2$ for the variance parameter $\zeta_n^2 = \text{var}(\xi_1^2)$. The second is an estimator $\tilde{\Sigma}_n$ for the diagonal matrix of population eigenvalues $\text{diag}(\lambda_1(\Sigma_n), \dots, \lambda_p(\Sigma_n))$. Once these two estimators have been assembled, the method generates bootstrap data from an elliptical model parameterized in terms of $\hat{\zeta}_n^2$ and $\tilde{\Sigma}_n$. More specifically, the i th sample of each bootstrap dataset $\mathbf{x}_1^*, \dots, \mathbf{x}_n^*$ is generated to be of the form

$$\mathbf{x}_i^* = \xi_i^* \tilde{\Sigma}_n^{1/2} \mathbf{u}_i^*,$$

where ξ_i^* is a non-negative random variable satisfying $\mathbb{E}((\xi_i^*)^2|X) = p$ and $\text{var}((\xi_i^*)^2|X) = \hat{\zeta}_n^2$, and $\mathbf{u}_i^* \in \mathbb{R}^p$ is drawn uniformly from the unit sphere, independently of ξ_i^* . Then, a single bootstrap sample of a generic spectral statistic $T_n = \psi(\lambda_1(\hat{\Sigma}_n), \dots, \lambda_p(\hat{\Sigma}_n))$ is computed as $T_n^* = \psi(\lambda_1(\hat{\Sigma}_n^*), \dots, \lambda_p(\hat{\Sigma}_n^*))$, where $\hat{\Sigma}_n^*$ is the sample covariance matrix of the bootstrap data.

To emphasize the modular role that $\hat{\zeta}_n^2$ and $\tilde{\Sigma}_n$ play in the bootstrap sampling process, we will provide the details for their construction later in Sections 3.2.2 and 3.2.3. With these points understood, the following algorithm shows that the method is very easy to implement.

ALGORITHM 1 (Bootstrap for spectral statistics).

Input: The number of bootstrap replicates B , as well as $\tilde{\Sigma}_n$ and $\hat{\zeta}_n^2$.

For: $b = 1, \dots, B$ **do in parallel**

1. Generate independent random variables $g_1^*, \dots, g_n^*$ from a Gamma distribution with mean p and variance $\hat{\varsigma}_n^2$, and then put $\xi_i^* = \sqrt{g_i^*}$ for $i = 1, \dots, n$.
2. Generate independent random vectors $\mathbf{u}_1^*, \dots, \mathbf{u}_n^* \in \mathbb{R}^p$ from the uniform distribution on the unit sphere.
3. Compute $\mathbf{x}_i^* = \xi_i^* \tilde{\Sigma}_n^{1/2} \mathbf{u}_i^*$ for $i = 1, \dots, n$, and form $\hat{\Sigma}_n^* = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^* (\mathbf{x}_i^*)^\top$.
4. Compute the spectral statistic $T_{n,b}^* = \psi(\lambda_1(\hat{\Sigma}_n^*), \dots, \lambda_p(\hat{\Sigma}_n^*))$.

end for

Return: The empirical distribution of $T_{n,1}^*, \dots, T_{n,B}^*$.

Remarks. One basic but valuable feature of the algorithm is that it can be applied with equal ease to both linear and nonlinear spectral statistics. To comment on some more technical aspects of the algorithm, the Gamma distribution is used in step 1 because it offers a convenient way to generate non-negative random variables whose means and variances can be matched to any pair of positive numbers. (If the event $\hat{\varsigma}_n^2 = 0$ happens to occur, then the Gamma distribution in step 1 is interpreted as the point mass at p , so that $\xi_i^* = \sqrt{p}$ for all $i = 1, \dots, n$.) Nevertheless, the choice of the Gamma distribution for generating $g_1^*, \dots, g_n^*$ is not required. Any other family of distributions on $[0, \infty)$ parameterized by means μ and variances σ^2 , say $\{G_{\mu, \sigma^2}\}$, will be compatible with our bootstrap consistency result in Theorem 3.3.2 if it satisfies the following two conditions: First, the pair (μ, σ^2) can be set to (p, v) for any integer $p \geq 1$ and real number $v > 0$. Second, there exists some fixed $\varepsilon > 0$ such that the quantity $\int |\frac{t-p}{\sqrt{p}}|^{4+\varepsilon} dG_{p, \sigma^2}(t)$ remains bounded when p diverges and σ^2/p converges to a finite limit.

3.2.2. Variance estimation. The estimation of the parameter $\varsigma_n^2 = \text{var}(\xi_1^2)$ is complicated by the fact that the random variables $\xi_1, \dots, \xi_n$ are not directly observable in an elliptical model. It is possible to overcome this challenge with the following estimating equation, which can be derived

from an explicit formula for $\text{var}(\|\mathbf{x}_1\|_2^2)$ that is given in Lemma D.1,

$$(3.5) \quad \varsigma_n^2 = p(p+2) \frac{\text{var}(\|\mathbf{x}_1\|_2^2) - 2\text{tr}(\Sigma_n^2)}{\text{tr}(\Sigma_n)^2 + 2\text{tr}(\Sigma_n^2)} + 2p.$$

Based on this equation, our approach is to separately estimate each of the three moment parameters on the right hand side, denoted as

$$\alpha_n = \text{tr}(\Sigma_n^2)$$

$$\beta_n = \text{var}(\|\mathbf{x}_1\|_2^2)$$

$$\gamma_n = \text{tr}(\Sigma_n)^2.$$

These parameters have the advantage that they can be estimated in a more direct manner, due to their simpler relations with the observations and the matrix Σ_n . Specifically, we use estimates defined according to

$$\begin{aligned} \hat{\alpha}_n &= \text{tr}(\hat{\Sigma}_n^2) - \frac{1}{n} \text{tr}(\hat{\Sigma}_n)^2, \\ \hat{\beta}_n &= \frac{1}{n-1} \sum_{i=1}^n \left(\|\mathbf{x}_i\|_2^2 - \frac{1}{n} \sum_{i'=1}^n \|\mathbf{x}_{i'}\|_2^2 \right)^2, \\ \hat{\gamma}_n &= \text{tr}(\hat{\Sigma}_n)^2. \end{aligned}$$

Substituting these estimates into (3.5) yields our proposed estimate for ς_n^2

$$(3.6) \quad \hat{\varsigma}_n^2 = \left(p(p+2) \frac{\hat{\beta}_n - 2\hat{\alpha}_n}{\hat{\gamma}_n + 2\hat{\alpha}_n} + 2p \right)_+$$

where $x_+ = \max\{x, 0\}$ denotes the non-negative part of any real number x . The consistency of this estimate will be established in Theorem 3.3.1, which shows $\frac{1}{p}(\hat{\varsigma}_n^2 - \varsigma_n^2) \rightarrow 0$ in probability as $n \rightarrow \infty$.

3.2.3. Spectrum estimation. The problem of estimating the eigenvalues of a population covariance matrix has attracted long-term interest in the high-dimensional statistics literature, and many different estimation methods have been proposed (e.g. Bai et al., 2010; El Karoui, 2008; Kong and Valiant, 2017; Ledoit and Wolf, 2015; Mestre, 2008). In order to estimate $\lambda_1(\Sigma_n), \dots, \lambda_p(\Sigma_n)$ in our current setting, we modify the method of QuEST (Ledoit and Wolf, 2015), which has become widely used in recent years. This choice has the benefit of making all aspects of our proposed

method easy to implement, because QuEST is supported by a turnkey software package (Ledoit and Wolf, 2017).

We denote the estimates produced by QuEST as $\hat{\lambda}_{Q,1} \geq \dots \geq \hat{\lambda}_{Q,p}$, and we will use modified versions of them defined by

$$(3.7) \quad \tilde{\lambda}_j = \min(\hat{\lambda}_{Q,j}, \hat{b}_n),$$

for $j = 1, \dots, p$, where we let $\hat{b}_n = \lambda_1(\hat{\Sigma}_n) + 1$. The modification is done for theoretical reasons, to ensure that $\tilde{\lambda}_1, \dots, \tilde{\lambda}_p$ are asymptotically bounded, which follows from Lemma A.5. In addition, we define the diagonal $p \times p$ matrix associated with these estimates as

$$(3.8) \quad \tilde{\Sigma}_n = \text{diag}(\tilde{\lambda}_1, \dots, \tilde{\lambda}_p).$$

Later, in Theorem 3.3.1, we will show that the estimates $\tilde{\lambda}_1, \dots, \tilde{\lambda}_p$ are consistent, in the sense that their empirical distribution converges weakly in probability to the correct limit as $n \rightarrow \infty$.

3.3. Theoretical results

In this section, we present three theoretical guarantees for the proposed method. Theorem 3.3.1 establishes appropriate notions of consistency for each of the estimators $\hat{\zeta}_n^2$ and $\tilde{\Sigma}_n$. Second, our main result in Theorem 3.3.2 shows that the bootstrap samples generated in Algorithm 1 consistently approximate the distributions of linear spectral statistics. Lastly, Theorem 3.3.3 demonstrates the asymptotic validity of bootstrap-based inference procedures involving nonlinear spectral statistics.

3.3.1. Setup. All of our theoretical analysis is framed in terms of a sequence of models indexed by n , so that all model parameters are allowed to vary with n , except when stated otherwise. The details of our model assumptions are given below in Assumptions 2 and 3.

ASSUMPTION 2 (Data generating model). *As $n \rightarrow \infty$, the dimension p grows so that the ratio $c_n = p/n$ satisfies $c_n \rightarrow c$ for some positive constant c different from 1. For each $i = 1, \dots, n$, the observation $\mathbf{x}_i \in \mathbb{R}^p$ can be represented as*

$$(3.9) \quad \mathbf{x}_i = \xi_i \Sigma_n^{1/2} \mathbf{u}_i,$$

where $\Sigma_n \in \mathbb{R}^{p \times p}$ is a deterministic non-zero positive semidefinite matrix, and $(\mathbf{u}_1, \xi_1), \dots, (\mathbf{u}_n, \xi_n)$ are i.i.d. random vectors in $\mathbb{R}^p \times [0, \infty)$ satisfying the following conditions: The vector $\mathbf{u}_1$ is drawn from the uniform distribution on the unit sphere of $\mathbb{R}^p$, and is independent of ξ_1 . In addition, as $n \rightarrow \infty$, the random variable ξ_1^2 satisfies $\mathbb{E}(\xi_1^2) = p$, as well as the conditions

$$(3.10) \quad \text{var} \left(\frac{\xi_1^2 - p}{\sqrt{p}} \right) = \tau + o(1) \quad \text{and} \quad \mathbb{E} \left| \frac{\xi_1^2 - p}{\sqrt{p}} \right|^{4+\varepsilon} \lesssim 1,$$

for some fixed constants $\tau \geq 0$ and $\varepsilon > 0$ that do not depend on n .

Remarks. With regard to the limiting value c for the ratio p/n , the single case of $c = 1$ is excluded so that we may employ certain facts about the QuEST method that were established in Ledoit and Wolf (2015). From a practical standpoint, our proposed method can still be used effectively when $p/n = 1$, as shown in Section 3.4. To address the moment conditions on ξ_1^2 , a similar set of conditions was used in (Hu et al., 2019a) to establish a high-dimensional central limit theorem for linear spectral statistics. However, our condition involving a $4 + \varepsilon$ moment for $(\xi_1^2 - p)/\sqrt{p}$ replaces a corresponding $2 + \varepsilon$ moment condition in that work. The extra bit of integrability is used here to show that the estimators $\hat{\zeta}_n^2$ and $\tilde{\Sigma}_n$ have suitable asymptotic properties for ensuring bootstrap consistency.

Examples. To illustrate that the conditions in (3.10) cover a substantial range of situations, it is possible to provide quite a few explicit examples of distributions for ξ_1^2 that are conforming:

- (1) Chi-Squared distribution with p degrees of freedom
- (2) Poisson(p)
- (3) $(1 - \tau)\text{Negative-Binomial}(p, 1 - \tau)$, for any $\tau \in (0, 1)$
- (4) Gamma($p/\tau, 1/\tau$), for any $\tau > 0$
- (5) Beta-Prime $\left(\frac{p(1+p+\tau)}{\tau}, \frac{1+p+2\tau}{\tau}\right)$, for any $\tau > 0$
- (6) Log-Normal $\left(\log(p) - \frac{1}{2} \log\left(1 + \frac{\tau}{p}\right), \log\left(1 + \frac{\tau}{p}\right)\right)$, for any $\tau > 0$

(7) $(p + 2\beta) \text{Beta}(p/2, \beta)$, for any $\beta > 0$

It is also possible to give a more abstract class of examples that subsumes some of the previous ones as special cases. In detail, the conditions in (3.10) will hold for any $\tau \geq 0$ if $\xi_1^2 = \sum_{j=1}^p z_{1j}^2$ for some independent random variables $z_{11}, \dots, z_{1p}$ satisfying

$$(3.11) \quad \frac{1}{p} \sum_{j=1}^p \mathbb{E}(z_{1j}^2) = 1, \quad \frac{1}{p} \sum_{j=1}^p \text{var}(z_{1j}^2) \rightarrow \tau, \quad \text{and} \quad \max_{1 \leq j \leq p} \mathbb{E}|z_{1j}|^{8+2\varepsilon} \lesssim p^{1+\frac{\varepsilon}{4}}.$$

Further details for checking the validity of the previous examples, as well as explicit parameterizations, are provided in Appendix B.5.

In addition to Assumption 2, we need one more assumption dealing with the spectrum of the population covariance matrix Σ_n . To state this assumption, let H_n denote the empirical distribution function associated with $\lambda_1(\Sigma_n), \dots, \lambda_p(\Sigma_n)$, which is defined for any $t \in \mathbb{R}$ according to

$$(3.12) \quad H_n(t) = \frac{1}{p} \sum_{j=1}^p 1\{\lambda_j(\Sigma_n) \leq t\}.$$

ASSUMPTION 3 (Spectral structure). *There is a limiting spectral distribution H such that as $n \rightarrow \infty$,*

$$(3.13) \quad H_n \Rightarrow H,$$

where the support of H is a finite union of closed intervals, bounded away from 0 and ∞ . Furthermore, there is a fixed compact interval in $(0, \infty)$ containing the support of H_n for all large n .

Remarks. Variations of these conditions on population eigenvalues are commonly used throughout random matrix theory. This particular set of conditions was used in (Ledoit and Wolf, 2015) to establish theoretical guarantees for the QuEST estimation method in the context of IC models.

3.3.2. Consistency of estimators. Here, we establish the consistency of the estimators $\hat{\zeta}_n^2$ and $\tilde{\Sigma}_n$, defined in (3.6) and (3.8). The appropriate notion of consistency for $\tilde{\Sigma}_n$ is stated in terms of its empirical spectral distribution function, which is defined for any $t \in \mathbb{R}$ as

$$(3.14) \quad \tilde{H}_n(t) = \frac{1}{p} \sum_{j=1}^p 1\{\tilde{\lambda}_j \leq t\}.$$

THEOREM 3.3.1. *Under Assumptions 2 and 3, the following limits hold as $n \rightarrow \infty$,*

$$(3.15) \quad \frac{1}{p}(\zeta_n^2 - \varsigma_n^2) \xrightarrow{\mathbb{P}} 0,$$

$$(3.16) \quad \tilde{H}_n \xrightarrow{\mathbb{P}} H.$$

Remarks. The limits (3.15) and (3.16) are proved in Appendices B.1.1 and B.1.2 respectively. Although these limits can be stated in a succinct form, quite a few details are involved in their proofs. For instance, the analysis of $\hat{\zeta}_n^2$ is based on extensive calculations with polynomial functions of the quadratic forms $\|\mathbf{x}_i\|_2^2$ and $\mathbf{x}_i^\top \mathbf{x}_j$ with $i \neq j$, as well as associated mixed moments. The consistency of $\tilde{H}_n$ is also notable because it requires showing the consistency of QuEST in elliptical models, and it seems that the consistency of QuEST has not previously been reported outside of IC models.

3.3.3. Consistency of bootstrap. To develop our main result on the consistency of the proposed bootstrap method, it is necessary to recall some background facts and introduce several pieces of notation.

Under Assumptions 2 and 3, it is known from (Bai and Zhou, 2008, Theorem 1.1) that an extended version of the classical Marčenko-Pastur Theorem holds for the empirical spectral distribution function $\hat{H}_n(t) = \frac{1}{p} \sum_{j=1}^p 1\{\lambda_j(\hat{\Sigma}_n) \leq t\}$. Namely, there is a probability distribution $\Psi(H, c)$ on $[0, \infty)$, depending only on H and c , such that the weak limit $\hat{H}_n \Rightarrow \Psi(H, c)$ occurs almost surely. In this statement, we may regard $\Psi(\cdot, \cdot)$ as a map that takes a distribution H' on $[0, \infty)$ and a number $c' > 0$ as input, and returns a new distribution $\Psi(H', c')$ on $[0, \infty)$ whose Stieltjes transform $m_{H', c'}(z) = \int \frac{1}{\lambda - z} d(\Psi(H', c'))(\lambda)$ solves the Marčenko-Pastur equation (3.17) below. That is, for any $z \in \mathbb{C}^+$, the number $m_{H', c'} = m_{H', c'}(z)$ is the unique solution to the equation

$$(3.17) \quad m_{H', c'} = \int \frac{1}{t(1 - c' - c'zm_{H', c'}) - z} dH'(t)$$

within the set $\{m \in \mathbb{C} \mid (c' - 1)/z + c'm \in \mathbb{C}^+\}$.

The map $\Psi(\cdot, \cdot)$ is relevant to our purposes here, because it determines a centering parameter that is commonly used in limit theorems for linear spectral statistics. In detail, if H_{n, c_n} denotes a shorthand for the probability distribution $\Psi(H_n, c_n)$, and if f is a real-valued function defined on

the support of H_{n,c_n} , then the associated centering parameter is defined as

$$(3.18) \quad \vartheta_n(f) = \int f(t) dH_{n,c_n}(t).$$

Similarly, let $\tilde{H}_{n,c_n} = \Psi(\tilde{H}_n, c_n)$ and $\tilde{\vartheta}_n(f) = \int f(t) d\tilde{H}_{n,c_n}(t)$. Also, in order to simplify notation for handling the joint distribution of several linear spectral statistics arising from a fixed set of functions $\mathbf{f} = (f_1, \dots, f_k)$, we write $T_n(\mathbf{f}) = (T_n(f_1), \dots, T_n(f_k))$, and likewise for $T_{n,1}^*(\mathbf{f})$, $\vartheta_n(\mathbf{f})$, and $\tilde{\vartheta}_n(\mathbf{f})$.

As one more preparatory item, recall from page 25 that d_{LP} denotes the Lévy-Prohorov metric for comparing distributions on $\mathbb{R}^k$. This metric is a standard choice for formulating bootstrap consistency results, as it has the fundamental property of metrizing weak convergence.

THEOREM 3.3.2. *Suppose Assumptions 2 and 3 hold, and let $\mathbf{f} = (f_1, \dots, f_k)$ be a fixed set of real-valued functions that are analytic on an open subset of $\mathbb{R}$ containing $[0, \infty)$. Under these conditions, if $T_{n,1}^*(\mathbf{f})$ is generated as in Algorithm 1 using the estimators $\hat{\zeta}_n^2$ and $\hat{\Sigma}_n$ defined by (3.6) and (3.8), then the following limit holds as $n \rightarrow \infty$*

$$(3.19) \quad d_{\text{LP}}\left(\mathcal{L}(p\{T_n(\mathbf{f}) - \vartheta_n(\mathbf{f})\}), \mathcal{L}(p\{T_{n,1}^*(\mathbf{f}) - \tilde{\vartheta}_n(\mathbf{f})\} | X)\right) \xrightarrow{\mathbb{P}} 0.$$

Remarks. The proof is given in Appendix B.2, and makes key use of a recently developed central limit theorem for linear spectral statistics due to (Hu et al., 2019a). Regarding other aspects of the theorem, there are two points to discuss. First, the assumption that the functions $f_1, \dots, f_k$ are defined on an open set containing $[0, \infty)$ has been made for technical simplicity. In the setting where $p/n \rightarrow c > 1$, this assumption is minor because $f_1, \dots, f_k$ must be defined at 0 due to the singularity of $\hat{\Sigma}_n$. Nevertheless, if $p/n \rightarrow c \in (0, 1)$, then it is possible to show that a corresponding version of the theorem holds for analytic functions that are not defined at 0, and our numerical results confirm that the proposed method can successfully handle such cases. Second, the quantities $\vartheta_n(\mathbf{f})$ and $\tilde{\vartheta}_n(\mathbf{f})$ are only introduced so that the distributions appearing in (3.19) have non-trivial weak limits, which facilitates the proof. Indeed, the proposed method can be applied without requiring any particular type of centering.

3.3.4. Application to inference on stable rank with guarantees. When dealing with high-dimensional covariance matrices, it is often of interest to have a measure of the number of

“dominant eigenvalues”. One such measure is the *stable rank*, defined as

$$(3.20) \quad r_n = \frac{\text{tr}(\Sigma_n)^2}{\text{tr}(\Sigma_n^2)},$$

which arises naturally in a plethora of situations (e.g. Bai and Saranadasa, 1996; Lopes, 2016; Lopes and Yao, 2022; Marcus et al., 2022; Srivastava, 2005; Tang and Nehorai, 2012). Whenever Σ_n is non-zero, this parameter satisfies $1 \leq r_n \leq \text{rank}(\Sigma_n)$, and the equality $r_n = p$ holds if and only if Σ_n is proportional to the identity matrix.

In this subsection, we illustrate how the proposed bootstrap method can be applied to solve some inference problems involving the parameter r_n . Our first example shows how to construct a confidence interval for r_n , and our subsequent examples deal with testing procedures related to r_n . Later, in Theorem 3.3.3, we establish the theoretical validity of the methods used in these examples—showing that the confidence interval has asymptotically exact coverage, and that the relevant testing procedures maintain asymptotic control of their levels.

3.3.4.1. *Confidence interval for stable rank.* Our confidence interval for r_n is constructed using the estimator

$$(3.21) \quad \hat{r}_n = \frac{\text{tr}(\hat{\Sigma}_n)^2}{\text{tr}(\hat{\Sigma}_n^2) - \hat{\Delta}_n},$$

where we define

$$\begin{aligned} \hat{\Delta}_n = & \frac{\text{tr}(\hat{\Sigma}_n^2)}{n} \left[\frac{2(n-1)}{n} \frac{p^2 + \zeta_n^2}{p(p+2)} - 1 \right] \\ & + \frac{\text{tr}(\hat{\Sigma}_n)^2}{n} \left[\frac{n+1}{n} + \frac{n-1}{n} \frac{\zeta_n^2 - 2p}{p(p+2)} - \frac{2(n-1)}{n^2} \frac{p^2 + \zeta_n^2}{p(p+2)} \right], \end{aligned}$$

and we set $\hat{r}_n$ equal to n in the exceptional case that its denominator is 0. It should be noted that $\hat{r}_n$ differs from the naive plug-in rule $\text{tr}(\hat{\Sigma}_n)^2/\text{tr}(\hat{\Sigma}_n^2)$, since the extra term $\hat{\Delta}_n$ in the denominator serves as a bias correction.

To proceed, let $q_{1-\alpha}$ denote $(1-\alpha)$ -quantile of the random variable $(\hat{r}_n - r_n)/p$ for any fixed $\alpha \in (0, 1)$, and consider the interval $\mathcal{I}_n = [\hat{r}_n - pq_{1-\alpha/2}, \hat{r}_n - pq_{\alpha/2}]$. Whenever the distribution of $(\hat{r}_n - r_n)/p$ is continuous, this interval satisfies

$$(3.22) \quad \mathbb{P}(r_n \in \mathcal{I}_n) = 1 - \alpha.$$

However, the quantiles $q_{\alpha/2}$ and $q_{1-\alpha/2}$ are unknown, and so they must be estimated. This can be accomplished by generating bootstrap samples of the following form in Algorithm 1,

$$T_{n,1}^* = \frac{1}{p} \left(\frac{\text{tr}(\hat{\Sigma}_n^*)^2}{\text{tr}((\hat{\Sigma}_n^*)^2) - \hat{\Delta}_n^*} - \frac{\text{tr}(\tilde{\Sigma}_n)^2}{\text{tr}(\tilde{\Sigma}_n^2)} \right),$$

where $\hat{\Delta}_n^*$ is defined by modifying the previous formula for $\hat{\Delta}_n$ so that $\hat{\Sigma}_n$ and $\hat{\zeta}_n^2$ are replaced by versions computed with the bootstrap data $\mathbf{x}_1^*, \dots, \mathbf{x}_n^*$. Also, we define $T_{n,1}^*$ to be 0 in the exceptional case of a denominator being equal to 0. Letting $\hat{q}_{\alpha/2}$ and $\hat{q}_{1-\alpha/2}$ denote the respective $\alpha/2$ and $(1 - \alpha/2)$ -quantiles of $\mathcal{L}(T_{n,1}^*|X)$, the proposed confidence interval is defined as

$$(3.23) \quad \hat{\mathcal{I}}_n = \left[\hat{r}_n - p\hat{q}_{1-\alpha/2}, \hat{r}_n - p\hat{q}_{\alpha/2} \right].$$

Below, Theorem 3.3.3 shows that as $n \rightarrow \infty$, the coverage probability $\mathbb{P}(r_n \in \hat{\mathcal{I}}_n)$ converges to $1 - \alpha$, as desired.

3.3.4.2. Hypotheses related to stable rank.

Screening data for PCA. Consider a scenario where a collection of different datasets are to be screened for further investigation by principal components analysis (PCA). In this situation, the datasets that should be discarded are the ones that cannot be well summarized by a moderate number of principal components. To put this in more quantitative terms, a dataset may be considered unsuitable for PCA if the stable rank r_n exceeds a certain fraction of the full dimension p . That is, if $\frac{r_n}{p} > \epsilon_0$ for some fixed reference value $\epsilon_0 \in (0, 1)$. On the other hand, if $\frac{r_n}{p} \leq \epsilon_0$, then the dataset may be retained. This leads to considering the hypothesis testing problem

$$(3.24) \quad H_{0,n} : \frac{r_n}{p} \leq \epsilon_0 \quad \text{vs.} \quad H_{1,n} : \frac{r_n}{p} > \epsilon_0.$$

To develop a testing procedure, we may again consider the $(1 - \alpha)$ -quantile $q_{1-\alpha}$ of the random variable $(\hat{r}_n - r_n)/p$. This quantile serves as a conceptual basis for a rejection criterion, because it satisfies the following inequality under the null hypothesis

$$(3.25) \quad \mathbb{P}\left(\frac{\hat{r}_n}{p} - \epsilon_0 > q_{1-\alpha}\right) \leq \alpha.$$

In other words, if $q_{1-\alpha}$ were known, then a level- α testing procedure would result from using $\frac{1}{p}\hat{r}_n - \epsilon_0 > q_{1-\alpha}$ as a rejection criterion. Accordingly, a bootstrap-based version of this procedure

rejects the null hypothesis when $\frac{1}{p}\widehat{r}_n - \epsilon_0 > \widehat{q}_{1-\alpha}$, with the quantile estimate $\widehat{q}_{1-\alpha}$ defined as before.

Testing for sphericity. One more example of a testing problem related to r_n is that of testing for sphericity,

$$(3.26) \quad H'_{0,n} : \Sigma_n = \sigma^2 I \text{ for some } \sigma^2 > 0 \quad \text{vs.} \quad H'_{1,n} : \Sigma_n \neq \sigma^2 I \text{ for any } \sigma^2 > 0.$$

The connection to the parameter r_n arises from the fact that (3.26) is equivalent to the problem $H'_{0,n} : \frac{r_n}{p} = 1$ vs. $H'_{1,n} : \frac{r_n}{p} < 1$. This observation was used in the influential paper (Srivastava, 2005) to develop a formula-based sphericity test. By analogy with our discussion of the problem (3.24), this observation can also be used to develop a bootstrap-based sphericity test. However, there is one point of distinction in the current situation, which is that the eigenvalues of Σ_n no longer need to be estimated. The reason is that under $H'_{0,n}$, the scale-invariance of the statistic $(\widehat{r}_n - r_n)/p$ causes it to behave as if $\Sigma_n = I_p$. Consequently, to estimate the quantiles of the null distribution of $(\widehat{r}_n - r_n)/p$, Algorithm 1 can be run using $\widetilde{\Sigma}_n = I_p$. To summarize, if we let $\widehat{q}'_\alpha$ denote the resulting estimate for the α -quantile of the null distribution of $(\widehat{r}_n - r_n)/p$, then the rejection criterion is $\frac{\widehat{r}_n}{p} - 1 \leq \widehat{q}'_\alpha$.

The following result establishes the theoretical validity of the procedures discussed in this subsection.

THEOREM 3.3.3. *If Assumptions 2 and 3 hold, then as $n \rightarrow \infty$,*

$$(3.27) \quad \mathbb{P}(r_n \in \widehat{\mathcal{I}}_n) \rightarrow 1 - \alpha.$$

Furthermore, if $H_{0,n}$ holds for all large n , then

$$(3.28) \quad \mathbb{P}\left(\frac{\widehat{r}_n}{p} - \epsilon_0 > \widehat{q}_{1-\alpha}\right) \leq \alpha + o(1),$$

or if $H'_{0,n}$ holds for all large n , then

$$(3.29) \quad \mathbb{P}\left(\frac{\widehat{r}_n}{p} - 1 \leq \widehat{q}'_\alpha\right) \rightarrow \alpha.$$

Remarks. This result provides a notable complement to Theorem 3.3.2, because it demonstrates that bootstrap consistency can be established in tasks that are based on a nonlinear spectral statistic, namely $\widehat{r}_n - r_n$. The proof of this result is given in Appendix B.3, where it can be seen

that the limiting distribution of $\widehat{r}_n - r_n$ has a very complicated dependence on the moments of H , due to the correlation between $\text{tr}(\widehat{\Sigma}_n)$ and $\text{tr}(\widehat{\Sigma}_n^2)$. In this way, the proof illustrates the utility of the bootstrap, since the bootstrap enables the user to completely bypass such complexity.

3.4. Numerical results

This section explores the empirical performance of the proposed bootstrap method in three different ways. Sections 3.4.2 and 3.4.3 deal with bootstrap approximations for linear and nonlinear spectral statistics, while Section 3.4.4 looks at procedures for doing inference on the stable rank parameter r_n .

3.4.1. Parameter settings. In our simulations, we generated data from elliptical models that were parameterized as follows. The random variable ξ_1^2 was generated using four choices of distributions:

- (i). Chi-Squared distribution with p degrees of freedom
- (ii). Beta-Prime $\left(\frac{p(1+p+8)}{8}, \frac{1+p+16}{8}\right)$,
- (iii). $(p+4)\text{Beta}(p/2, 2)$,
- (iv). $\frac{9p}{10}\text{F}(p, 20)$,

where $\text{F}(d_1, d_2)$ denotes an F-distribution with d_1 and d_2 degrees of freedom. Note that cases (i), (iii), (iv), correspond respectively to a multivariate Gaussian distribution, a multivariate Pearson type II distribution, and a multivariate t-distribution with 20 degrees of freedom. Also, the numerical values appearing in (i)-(iv) were chosen to ensure the normalization condition $\mathbb{E}(\xi_1^2) = p$.

The population covariance matrix Σ_n was selected from five options:

1. The eigenvalues of Σ_n are $\lambda_1(\Sigma_n) = \dots = \lambda_5(\Sigma_n) = \frac{4}{3}$ and $\lambda_j(\Sigma_n) = 1$ for $j \in \{6, \dots, p\}$.
The $p \times p$ matrix of eigenvectors of Σ_n is generated from the uniform distribution on orthogonal matrices.
2. The eigenvalues of Σ_n are $\lambda_j(\Sigma_n) = \exp(-j/4)$ for $j \in \{1, \dots, 20\}$, and $\lambda_{20}(\Sigma_n) = \dots = \lambda_p(\Sigma_n)$. The eigenvectors are the same as in case 1.

3. The matrix Σ_n has entries of the form $(\Sigma_n)_{ij} = (\frac{1}{10})^{|i-j|} + 1\{i = j\}$.
4. The matrix Σ_n has entries of the form $(\Sigma_n)_{ij} = (\frac{1}{10})1\{i \neq j\} + 1\{i = j\}$.
5. The eigenvalues of Σ_n are $\lambda_1(\Sigma_n) = 5$ and $\lambda_j(\Sigma_n) = 1$ for $j \in \{2, \dots, p\}$. The eigenvectors are the same as in case 1.

3.4.2. Linear spectral statistics. Our experiments for linear spectral statistics were based on the task of using the proposed bootstrap to estimate three parameters of $\mathcal{L}(p(T_n(f) - \vartheta_n(f)))$: the mean, standard deviation, and 95th percentile. We considered two choices for the function f , namely $f(x) = x^2$ and $f(x) = x - \log(x) - 1$. In the first case, we selected the ratio p/n so that $p/n \in \{0.5, 1, 1.5\}$, and in the second case we used $p/n \in \{0.3, 0.5, 0.7\}$.

Design of experiments. For each possible choice of $(\xi_1, \Sigma_n, p/n)$, we generated 5000 realizations of the dataset $\mathbf{x}_1, \dots, \mathbf{x}_n$, with $n = 400$. These datasets allowed us to compute 5000 realizations of the statistic $p(T_n(f) - \vartheta_n(f))$, and we treated the empirical mean, standard deviation, and 95th percentile of these 5000 realizations as ground truth for our parameters of interest. In Tables 3.1 and 3.2, the ground truth values are reported in the first row of numbers corresponding to each choice of Σ_n .

With regard to the bootstrap, we ran Algorithm 1 on the first 500 datasets corresponding to each parameter setting. Also, we generated $B = 250$ bootstrap samples of the form $p(T_{n,1}^*(f) - \tilde{\vartheta}_n(f))$ during every run. As a result of these runs, we obtained 500 different bootstrap estimates of the mean, standard deviation, and 95th percentile of $\mathcal{L}(p(T_n(f) - \vartheta_n(f)))$. In Tables 3.1 and 3.2, we report the empirical mean and standard deviation (in parenthesis) of these 500 estimates in the second row of numbers corresponding to each choice of Σ_n .

One more detail to mention is related to the computation of $\vartheta_n(f)$ and $\tilde{\vartheta}_n(f)$. For each parameter setting, we approximated $\vartheta_n(f)$ as follows. We averaged 30 realizations of

$$\frac{1}{40p} \sum_{j=1}^{40p} f\left(\lambda_j\left(\frac{1}{40n} \sum_{i=1}^{40n} \xi_i^2 \Sigma_n^{1/2} \mathbf{u}_i \mathbf{u}_i^\top \Sigma_n^{1/2}\right)\right),$$

where $\Sigma_n = I_{40} \otimes \Sigma_n$ is of size $40p \times 40p$, each $\mathbf{u}_i$ was drawn from the uniform distribution on the unit sphere of $\mathbb{R}^{40p}$, and each ξ_i^2 was generated as in (i)-(iv), but with $40p$ replacing p . For the bootstrap

samples, we computed one realization of the statistic $\frac{1}{40p} \sum_{j=1}^{40p} f(\lambda_j(\frac{1}{40n} \sum_{i=1}^{40n} (\xi_i^*)^2 \tilde{\Sigma}_n^{1/2} \mathbf{u}_i^* (\mathbf{u}_i^*)^\top \tilde{\Sigma}_n^{1/2}))$ to approximate $\tilde{\vartheta}_n(f)$ during every run of Algorithm 1, where $\tilde{\Sigma}_n = I_{40} \otimes \tilde{\Sigma}_n$ is of size $40p \times 40p$, each $\mathbf{u}_i^*$ was drawn from the uniform distribution on the unit sphere of $\mathbb{R}^{40p}$, and each $(\xi_i^*)^2$ was drawn from a Gamma distribution with mean $40p$ and variance $40\zeta_n^2$.

Comments on results. It is easiest to explain the format of the tables with an example: The two entries in the upper right corner of Table 3.1 show that in settings (i) and 1 with $p/n = 1.5$, the 95th percentile of $\mathcal{L}(p(T_n(f) - \vartheta_n(f)))$ with $f(x) = x^2$ is equal to 16.48, and the bootstrap estimate for the 95th percentile has a mean (standard deviation) of 16.73 (1.48). Table 3.2 presents results for $f(x) = \log(x) - x - 1$ in the same format.

In most settings, the bootstrap estimates perform well, with their bias and standard deviation being small in proportion to the parameter being estimated. However, there are some specific parameter settings that require more attention. These settings involve choice 4 for Σ_n , which is an equi-correlation matrix, and choice (iv) for ξ_1 , which induces a multivariate t-distribution with 20 degrees of freedom. Notably, these choices correspond to settings that violate Assumptions 2 and 3 of our theoretical results. In the case of the equi-correlation matrix, the bootstrap approximations for $f(x) = x^2$ are less accurate in comparison to other choices of Σ_n , due to increased variance. By contrast, if $f(x) = \log(x) - x - 1$, then the bootstrap approximations have similar accuracy across all choices of Σ_n while holding other parameters fixed.

In the case of the multivariate t-distribution, the bootstrap is able to accurately estimate the standard deviation of $\mathcal{L}(p(T_n(f) - \vartheta_n(f)))$ for both choices of f , but difficulties arise in estimating the mean and 95th percentile. To understand these mixed results, it is important to recognize the standard deviation does not depend on the centering parameter $\vartheta_n(f)$, whereas the mean and 95th percentile do. Also, the choice of $\vartheta_n(f)$ as a centering parameter is based on the CLT for linear spectral statistics established in (Hu et al., 2019a), and the assumptions underlying that result are violated by the multivariate t-distribution.

TABLE 3.1. Results for $p(T_n(f) - \vartheta_n(f))$ with $f(x) = x^2$

ξ_1	Σ_n	$p/n = 0.5$			$p/n = 1$			$p/n = 1.5$		
		mean	sd	95th	mean	sd	95th	mean	sd	95th
(i)	1	0.51	3.31	6.02	0.85	6.11	10.87	1.44	9.15	16.48
		0.48(0.24)	3.27(0.19)	5.85(0.53)	0.96(0.44)	6.09(0.35)	11.02(1)	1.47(0.65)	9.25(0.51)	16.73(1.48)
	2	0	0.11	0.19	0.01	0.11	0.19	0	0.11	0.2
		0(0.01)	0.11(0.01)	0.19(0.03)	0(0.01)	0.11(0.01)	0.19(0.02)	0(0.01)	0.11(0.01)	0.2(0.03)
	3	2.16	12.88	23.12	4.02	24.12	44.03	5.8	36.13	65.72
		1.9(0.88)	12.85(0.73)	23.2(2.07)	3.94(1.75)	24.27(1.35)	43.95(3.82)	5.8(2.77)	37.08(2.2)	67.09(6.1)
	4	0.64	63.35	110.6	2.48	241.3	420	14.11	540	952.2
		1.53(4.7)	62.9(9.6)	111.2(19.68)	4.95(17.04)	245(34.43)	433(70.15)	11.44(36.41)	541.9(78.44)	958.3(161.6)
	5	0.55	5.03	9.1	1.03	7.49	13.4	1.57	10.42	18.58
		0.55(0.34)	5.1(0.46)	9.15(1.05)	1.04(0.53)	7.41(0.48)	13.4(1.22)	1.55(0.74)	10.35(0.65)	18.61(1.7)
	1	3.57	6.35	14.09	6.69	11.93	26.9	10.25	18.09	39.92
		3.47(0.59)	6.36(0.45)	14.03(1.29)	6.83(1)	11.85(0.76)	26.51(2.32)	10.26(1.53)	17.99(1.11)	39.92(3.25)
	2	0.01	0.12	0.21	0	0.11	0.2	0	0.11	0.19
		0.01(0.01)	0.12(0.01)	0.2(0.03)	0.01(0.01)	0.11(0.01)	0.2(0.03)	0(0.01)	0.11(0.01)	0.2(0.03)
	3	14.16	25.27	56	27.59	47.57	105.8	41.49	71.66	159.7
		13.67(2.24)	25.12(1.76)	55.35(5.1)	26.94(3.72)	47.09(2.97)	104.9(8.33)	40.39(6.13)	71.59(4.5)	158.6(13.38)
	4	2.26	65.49	114.9	9.27	246.8	435.7	23.51	539.9	970.1
		4.7(4.65)	66.35(9.67)	119.7(20.26)	9.81(16.29)	248.2(38.15)	441.7(75.8)	19.95(37.73)	548.5(83.74)	971.1(172)
	5	3.75	7.75	16.66	7.21	13.01	28.58	10.46	18.87	42.02
		3.59(0.65)	7.86(0.62)	16.69(1.58)	7.03(1.05)	12.87(0.84)	28.29(2.43)	10.4(1.51)	18.83(1.1)	41.45(3.22)
(ii)	1	-0.46	1.13	1.44	-0.95	2.13	2.61	-1.47	3.12	3.71
		-0.47(0.08)	1.14(0.07)	1.4(0.17)	-0.96(0.14)	2.15(0.11)	2.58(0.33)	-1.43(0.21)	3.17(0.15)	3.81(0.45)
	2	0.01	0.11	0.19	0	0.11	0.19	0	0.11	0.19
		0(0.01)	0.11(0.01)	0.19(0.03)	0(0.01)	0.11(0.01)	0.19(0.02)	0(0.01)	0.11(0.01)	0.19(0.03)
	3	-1.93	4.49	5.41	-3.89	8.66	10.62	-5.62	13	15.33
		-1.85(0.33)	4.56(0.25)	5.68(0.69)	-3.78(0.57)	8.69(0.48)	10.57(1.26)	-5.7(0.9)	12.87(0.68)	15.52(1.88)
	4	1.61	63.23	110	6.1	242	420.7	-6.78	541.8	944.3
		0.75(4.21)	63.54(9.08)	111.2(18.37)	3.06(16.56)	242.5(37.18)	423.6(74.62)	5.52(35.54)	530.5(80.96)	928.1(161.8)
	5	-0.52	3.78	5.98	-0.9	4.45	6.74	-1.3	5.35	7.75
		-0.48(0.25)	3.83(0.51)	6.14(1.02)	-0.94(0.32)	4.53(0.55)	6.81(1.17)	-1.4(0.36)	5.38(0.53)	7.67(1.15)
	1	0.67	12.57	21.94	1.19	33.98	58.49	0.77	64.69	114.1
		13.02(2.03)	12.28(1.26)	33.54(4.18)	50.46(7.66)	33.11(3.74)	105.7(13.79)	111.2(15.79)	62.29(6.76)	215.5(26.71)
	2	0.01	0.13	0.24	0.01	0.14	0.25	0	0.14	0.23
		0.01(0.01)	0.13(0.02)	0.24(0.04)	0.02(0.01)	0.14(0.02)	0.25(0.04)	0.02(0.01)	0.14(0.02)	0.25(0.04)
	3	0.68	48.68	82.62	1.3	136.4	236.7	15.26	260	457.1
		51.21(8.4)	48.62(5.03)	132.7(17.04)	198.5(28.4)	131.5(13.71)	418.8(51.54)	442.9(64.24)	250.9(28.24)	861.1(108)
	4	2.75	73.19	135.4	17.25	281.2	506.3	27.98	603.8	1097
		14.03(5.46)	73.41(12.32)	142.2(25.81)	54.28(20.28)	277.7(44.26)	538.9(90.87)	123.8(46.62)	622.4(97.37)	1217(206.5)
	5	0.42	14.02	23.86	2.62	35.23	62.08	-0.69	66.26	110.7
		13.2(2.08)	13.82(1.41)	36.31(4.4)	50.71(7.33)	34.45(3.39)	108.3(12.79)	111.5(15.1)	64.04(6.65)	219(26.3)
(iv)	1	0.67	12.57	21.94	1.19	33.98	58.49	0.77	64.69	114.1
		13.02(2.03)	12.28(1.26)	33.54(4.18)	50.46(7.66)	33.11(3.74)	105.7(13.79)	111.2(15.79)	62.29(6.76)	215.5(26.71)
	2	0.01	0.13	0.24	0.01	0.14	0.25	0	0.14	0.23
		0.01(0.01)	0.13(0.02)	0.24(0.04)	0.02(0.01)	0.14(0.02)	0.25(0.04)	0.02(0.01)	0.14(0.02)	0.25(0.04)
	3	0.68	48.68	82.62	1.3	136.4	236.7	15.26	260	457.1
		51.21(8.4)	48.62(5.03)	132.7(17.04)	198.5(28.4)	131.5(13.71)	418.8(51.54)	442.9(64.24)	250.9(28.24)	861.1(108)
	4	2.75	73.19	135.4	17.25	281.2	506.3	27.98	603.8	1097
		14.03(5.46)	73.41(12.32)	142.2(25.81)	54.28(20.28)	277.7(44.26)	538.9(90.87)	123.8(46.62)	622.4(97.37)	1217(206.5)
	5	0.42	14.02	23.86	2.62	35.23	62.08	-0.69	66.26	110.7
		13.2(2.08)	13.82(1.41)	36.31(4.4)	50.71(7.33)	34.45(3.39)	108.3(12.79)	111.5(15.1)	64.04(6.65)	219(26.3)

TABLE 3.2. Results for $p(T_n(f) - \vartheta_n(f))$ with $f(x) = x - \log(x) - 1$

ξ_1	Σ_n	$p/n = 0.3$			$p/n = 0.5$			$p/n = 0.7$		
		mean	sd	95th	mean	sd	95th	mean	sd	95th
(i)	1	0.18	0.35	0.77	0.34	0.64	1.4	0.6	1	2.24
		0.17(0.03)	0.35(0.02)	0.74(0.05)	0.34(0.06)	0.63(0.03)	1.37(0.09)	0.58(0.09)	1.01(0.05)	2.25(0.14)
	2	0.18	0.83	1.55	0.36	1.16	2.25	0.59	1.53	3.1
		0.19(0.22)	0.79(0.26)	1.49(0.65)	0.36(0.45)	1.11(0.37)	2.19(1.05)	0.62(0.7)	1.51(0.44)	3.1(1.42)
	3	0.18	0.86	1.62	0.35	1.19	2.27	0.55	1.58	3.12
		0.17(0.06)	0.86(0.05)	1.59(0.14)	0.33(0.09)	1.19(0.06)	2.3(0.19)	0.58(0.12)	1.56(0.08)	3.17(0.25)
	4	0.16	0.91	1.66	0.32	1.53	2.92	0.57	2.21	4.29
		0.17(0.09)	0.91(0.07)	1.71(0.18)	0.33(0.18)	1.54(0.12)	2.91(0.31)	0.58(0.31)	2.21(0.16)	4.29(0.49)
	5	0.18	0.44	0.91	0.32	0.7	1.46	0.56	1.04	2.28
		0.17(0.04)	0.45(0.03)	0.92(0.07)	0.34(0.06)	0.69(0.03)	1.48(0.11)	0.59(0.09)	1.05(0.05)	2.31(0.15)
	1	1	0.37	1.62	1.74	0.65	2.79	2.56	1.02	4.25
		1.04(0.11)	0.37(0.02)	1.66(0.12)	1.79(0.17)	0.66(0.03)	2.87(0.19)	2.64(0.22)	1.04(0.05)	4.35(0.25)
	2	1.04	1.52	3.54	1.74	2.02	5.06	2.54	2.49	6.54
		1.05(0.4)	1.51(0.25)	3.54(0.81)	1.77(0.7)	2.01(0.35)	5.09(1.27)	2.54(0.96)	2.46(0.41)	6.59(1.63)
	3	1.02	1.63	3.78	1.78	2.15	5.34	2.57	2.63	6.94
		1.04(0.16)	1.6(0.12)	3.68(0.36)	1.8(0.22)	2.11(0.14)	5.31(0.45)	2.61(0.28)	2.59(0.16)	6.88(0.53)
	4	1.02	0.94	2.59	1.69	1.57	4.26	2.58	2.22	6.3
		1.04(0.16)	0.94(0.07)	2.63(0.24)	1.8(0.31)	1.58(0.12)	4.45(0.43)	2.63(0.49)	2.25(0.17)	6.4(0.63)
	5	1.02	0.47	1.8	1.78	0.71	2.96	2.56	1.06	4.31
		1.05(0.13)	0.48(0.03)	1.84(0.16)	1.8(0.17)	0.72(0.04)	2.98(0.2)	2.64(0.22)	1.08(0.05)	4.41(0.27)
(iii)	1	-0.11	0.34	0.45	-0.13	0.62	0.9	-0.14	1	1.49
		-0.11(0.02)	0.34(0.02)	0.45(0.05)	-0.14(0.04)	0.62(0.03)	0.88(0.08)	-0.09(0.07)	1.01(0.04)	1.57(0.14)
	2	-0.12	0.36	0.49	-0.14	0.64	0.93	-0.08	1.01	1.57
		-0.04(0.12)	0.47(0.18)	0.74(0.43)	0.05(0.29)	0.85(0.28)	1.45(0.74)	0.18(0.45)	1.22(0.31)	2.19(0.95)
	3	-0.11	0.38	0.52	-0.14	0.65	0.93	-0.07	1.03	1.66
		-0.11(0.03)	0.38(0.02)	0.51(0.05)	-0.14(0.05)	0.65(0.03)	0.93(0.09)	-0.09(0.07)	1.03(0.05)	1.6(0.14)
	4	-0.09	0.91	1.43	-0.12	1.55	2.47	-0.14	2.21	3.64
		-0.09(0.07)	0.9(0.08)	1.42(0.17)	-0.1(0.13)	1.54(0.11)	2.48(0.28)	0.01(0.21)	2.21(0.16)	3.7(0.4)
	5	-0.11	0.43	0.62	-0.15	0.67	0.94	-0.15	1.06	1.57
		-0.11(0.03)	0.44(0.03)	0.62(0.07)	-0.14(0.05)	0.68(0.03)	0.99(0.09)	-0.09(0.07)	1.04(0.05)	1.64(0.15)
	1	0.19	0.45	0.95	0.37	0.86	1.81	0.68	1.44	3.15
		2.35(0.28)	0.44(0.03)	3.07(0.31)	6.45(0.79)	0.84(0.06)	7.85(0.87)	12.53(1.51)	1.41(0.1)	14.87(1.64)
	2	0.15	2.15	3.75	0.26	3.41	5.85	0.46	4.77	8.24
		2.37(0.81)	2.2(0.31)	6.01(1.31)	6.37(1.52)	3.64(0.43)	12.36(2.23)	12.6(2.65)	5.16(0.56)	21.06(3.51)
	3	0.13	2.42	4.3	0.22	3.96	6.84	0.57	5.75	10.49
		2.37(0.34)	2.32(0.21)	6.25(0.7)	6.45(0.86)	3.8(0.35)	12.82(1.43)	12.67(1.58)	5.31(0.5)	21.49(2.33)
	4	0.19	1.01	1.89	0.4	1.7	3.29	0.5	2.53	4.75
		2.39(0.35)	0.99(0.09)	4.06(0.44)	6.46(0.93)	1.72(0.15)	9.37(1.08)	12.44(1.78)	2.52(0.2)	16.64(1.97)
	5	0.19	0.54	1.1	0.35	0.93	1.92	0.62	1.49	3.12
		2.36(0.3)	0.54(0.04)	3.25(0.35)	6.35(0.78)	0.9(0.07)	7.86(0.87)	12.55(1.55)	1.46(0.1)	14.97(1.7)

Breakdown of nonparametric bootstrap. To highlight one of the motivations for our parametric bootstrap method, we close this subsection with some numerical results exhibiting the breakdown

TABLE 3.3. Estimation of standard deviation of $p(T_n(f) - \vartheta_n(f))$ with the non-parametric bootstrap

p/n	ξ_1	Σ_n	$f(x) = x^2$	$f(x) = x - \log(x) - 1$
0.5	(i)	1	3.31 9.1(0.45)	0.64 6.08(0.27)
0.5	(i)	2	0.11 0.11(0.01)	1.16 6.11(0.28)
0.5	(i)	3	12.88 35.72(1.83)	1.19 6.23(0.31)
0.5	(i)	4	63.35 66.65(10.63)	1.53 6.25(0.3)
0.5	(i)	5	5.03 10.35(0.63)	0.7 6.11(0.27)

of the standard nonparametric bootstrap, based on sampling with replacement. For the sake of brevity, we focus only on the task of estimating the standard deviation of $p(T_n(f) - \vartheta_n(f))$ in a subset of the previous parameter settings, corresponding to $p/n = 1/2$, and $(\xi_n, \Sigma_n) \in \{(i)\} \times \{1, 2, 3, 4, 5\}$. The standard deviation is of particular interest, because it clarifies that the breakdown does not depend on how the statistic is centered.

The results are presented in Table 3.3, showing that the nonparametric bootstrap tends to overestimate the true standard deviation, often by a factor of 2 or more. For comparison, the corresponding estimates obtained from the proposed bootstrap method are much more accurate, as can be seen from Tables 3.1 and 3.2. In addition, the nonparametric bootstrap can have difficulties with nonlinear spectral statistics in settings like those considered here. Numerical results along these lines can be found in the paper (El Karoui and Purdom, 2019).

3.4.3. Nonlinear spectral statistics. This subsection looks at how well the proposed bootstrap handles nonlinear spectral statistics. The statistics under consideration here are the largest sample eigenvalue $\lambda_1(\widehat{\Sigma}_n)$, and the leading eigengap $\lambda_1(\widehat{\Sigma}_n) - \lambda_2(\widehat{\Sigma}_n)$. The underlying experiments for these statistics were designed in the same manner as in Section 3.4.2, and Tables 3.4 and 3.5 display the results in the same format as Tables 3.1 and 3.2. To a large extent, the favorable patterns that were noted in the results for linear spectral statistics are actually enhanced in the results for nonlinear spectral statistics—in the sense that the bias and standard deviations of the bootstrap estimates are generally smaller here than before. Also, in contrast to linear spectral statistics, the

results for nonlinear spectral statistics are relatively unaffected by choice 4 for Σ_n . Lastly, under choice (iv) for ξ_1 , the accuracy for nonlinear spectral statistics is reduced in comparison to other choices of ξ_1 . Nevertheless, the reduction in accuracy under choice (iv) is less pronounced here than it was in the context of linear spectral statistics. This makes sense in light of the fact that the statistics $\lambda_1(\widehat{\Sigma}_n)$ and $\lambda_1(\widehat{\Sigma}_n) - \lambda_2(\widehat{\Sigma}_n)$ do not involve the centering parameter $\vartheta_n(f)$ that was used for linear spectral statistics. (Recall from Section 3.4.2 that the reduced accuracy for linear spectral statistics under choice (iv) appeared to be related to $\vartheta_n(f)$.)

3.4.4. Inference on stable rank. Confidence interval. Table 3.6 presents numerical results on the width and coverage probability of the bootstrap confidence interval $\widehat{\mathcal{I}}_n$ (defined in (3.23)) for the stable rank parameter r_n . The results were computed using experiments based on the same design as in Section 3.4.2, with every interval having a nominal coverage probability of 95%. Due to the fact that the width of the interval scales with the value of r_n , we report the width as a percentage of r_n . To illustrate a particular example, the upper right corner of Table 3.6 shows that in settings 1 and (i) with $p/n = 1.5$, the average width of the interval over repeated experiments is 1.97% of r_n , with a standard deviation of 0.12%. Under choices 1, 3, and 5 for Σ_n , the width is typically quite small as a percentage of r_n . By contrast, the percentage is larger in cases 2 and 4, which seems to occur because r_n is smaller in these cases compared to 1, 3, and 5. Lastly, to consider coverage probability, the table shows good general agreement with the nominal level. Indeed, among the 60 distinct settings, there are only a few where the coverage probability differs from the nominal level by more than 2%.

TABLE 3.4. Results for $\lambda_1(\widehat{\Sigma}_n)$

ξ_1	Σ_n	$p/n = 0.5$			$p/n = 1$			$p/n = 1.5$		
		mean	sd	95th	mean	sd	95th	mean	sd	95th
(i)	1	2.9	0.06	3	3.96	0.06	4.07	4.9	0.06	5.01
		2.93(0.05)	0.06(0.01)	3.04(0.07)	3.99(0.04)	0.06(0.01)	4.1(0.06)	4.93(0.04)	0.07(0.01)	5.05(0.06)
	2	0.8	0.05	0.89	0.8	0.05	0.89	0.81	0.05	0.9
		0.8(0.05)	0.05(0.01)	0.89(0.06)	0.8(0.05)	0.05(0.01)	0.89(0.06)	0.81(0.05)	0.05(0.01)	0.9(0.06)
	3	5.76	0.11	5.95	7.91	0.12	8.12	9.81	0.13	10.03
		5.81(0.08)	0.12(0.02)	6.02(0.11)	7.98(0.09)	0.13(0.02)	8.21(0.13)	9.87(0.1)	0.14(0.03)	10.12(0.14)
	4	21.35	1.47	23.81	41.79	2.87	46.62	62.34	4.31	69.6
		21.21(1.5)	1.47(0.13)	23.68(1.71)	41.8(2.73)	2.9(0.23)	46.7(3.09)	62.34(4.19)	4.3(0.35)	69.62(4.73)
	5	5.62	0.35	6.21	6.25	0.34	6.81	6.88	0.33	7.44
		5.61(0.33)	0.35(0.03)	6.2(0.38)	6.22(0.33)	0.34(0.03)	6.79(0.37)	6.87(0.33)	0.33(0.03)	7.43(0.38)
	1	2.96	0.07	3.08	4.02	0.07	4.14	4.96	0.07	5.08
		3.03(0.07)	0.07(0.01)	3.15(0.09)	4.09(0.07)	0.07(0.01)	4.22(0.09)	5.04(0.06)	0.08(0.01)	5.17(0.08)
	2	0.8	0.06	0.89	0.8	0.05	0.89	0.8	0.05	0.9
		0.8(0.05)	0.05(0.01)	0.89(0.06)	0.8(0.06)	0.05(0.01)	0.89(0.06)	0.81(0.06)	0.05(0.01)	0.9(0.06)
	3	5.88	0.13	6.11	8.04	0.14	8.26	9.93	0.14	10.16
		6.03(0.13)	0.14(0.02)	6.27(0.16)	8.2(0.12)	0.15(0.02)	8.47(0.15)	10.1(0.12)	0.16(0.02)	10.37(0.16)
	4	21.34	1.51	23.86	41.8	2.92	46.72	62.32	4.3	69.65
		21.44(1.48)	1.51(0.13)	23.99(1.69)	41.87(3)	2.92(0.25)	46.79(3.38)	62.37(4.37)	4.34(0.38)	69.67(4.99)
	5	5.65	0.35	6.24	6.28	0.35	6.86	6.9	0.34	7.48
		5.67(0.36)	0.36(0.03)	6.27(0.41)	6.31(0.35)	0.35(0.03)	6.9(0.4)	6.91(0.35)	0.34(0.03)	7.48(0.4)
(ii)	1	2.88	0.05	2.97	3.94	0.06	4.04	4.88	0.06	4.99
		2.89(0.03)	0.06(0.01)	2.98(0.05)	3.95(0.03)	0.06(0.01)	4.06(0.05)	4.89(0.03)	0.06(0.01)	5(0.05)
	2	0.8	0.05	0.89	0.8	0.05	0.89	0.8	0.05	0.9
		0.8(0.05)	0.05(0.01)	0.89(0.06)	0.8(0.05)	0.05(0.01)	0.89(0.06)	0.8(0.05)	0.05(0.01)	0.89(0.06)
	3	5.72	0.1	5.9	7.88	0.12	8.08	9.77	0.12	9.99
		5.74(0.07)	0.11(0.02)	5.94(0.11)	7.9(0.08)	0.12(0.02)	8.12(0.13)	9.79(0.08)	0.13(0.02)	10.01(0.11)
	4	21.39	1.47	23.84	41.83	2.88	46.61	62.14	4.34	69.51
		21.39(1.38)	1.47(0.12)	23.87(1.56)	41.74(2.95)	2.87(0.25)	46.57(3.35)	61.71(4.39)	4.25(0.37)	68.87(4.96)
	5	5.61	0.34	6.17	6.24	0.34	6.82	6.87	0.34	7.44
		5.59(0.34)	0.34(0.03)	6.17(0.38)	6.23(0.35)	0.34(0.03)	6.8(0.39)	6.88(0.34)	0.34(0.03)	7.45(0.39)
	1	3.24	0.18	3.55	4.77	0.42	5.48	6.27	0.67	7.51
		3.39(0.22)	0.12(0.03)	3.59(0.27)	5(0.46)	0.18(0.05)	5.31(0.54)	6.59(0.64)	0.25(0.07)	7.03(0.75)
	2	0.8	0.06	0.9	0.8	0.06	0.91	0.81	0.06	0.91
		0.81(0.06)	0.06(0.01)	0.91(0.07)	0.81(0.06)	0.06(0.01)	0.91(0.07)	0.81(0.06)	0.06(0.01)	0.91(0.07)
	3	6.42	0.36	7.03	9.51	0.81	10.97	12.52	1.33	15.13
		6.77(0.4)	0.24(0.06)	7.18(0.49)	10.17(0.81)	0.37(0.1)	10.82(0.97)	13.55(1.35)	0.54(0.16)	14.48(1.6)
	4	21.45	1.62	24.28	42.01	3.23	47.49	62.59	4.68	70.57
		21.52(1.71)	1.62(0.16)	24.26(1.96)	42.08(3.22)	3.17(0.29)	47.44(3.66)	63.05(4.62)	4.76(0.43)	71.15(5.29)
	5	5.71	0.39	6.37	6.48	0.4	7.18	7.33	0.51	8.14
		5.79(0.39)	0.39(0.04)	6.45(0.45)	6.66(0.42)	0.39(0.04)	7.33(0.49)	7.69(0.45)	0.39(0.05)	8.37(0.53)

TABLE 3.5. Results for $\lambda_1(\hat{\Sigma}_n) - \lambda_2(\hat{\Sigma}_n)$

ξ_1	Σ_n	$p/n = 0.5$			$p/n = 1$			$p/n = 1.5$		
		mean	sd	95th	mean	sd	95th	mean	sd	95th
(i)	1	0.09	0.05	0.18	0.1	0.05	0.19	0.1	0.06	0.21
		0.1(0.02)	0.06(0.01)	0.2(0.05)	0.1(0.02)	0.06(0.01)	0.21(0.04)	0.11(0.02)	0.06(0.01)	0.22(0.04)
	2	0.18	0.07	0.3	0.18	0.07	0.29	0.18	0.07	0.29
		0.19(0.06)	0.06(0.01)	0.29(0.07)	0.18(0.06)	0.06(0.01)	0.29(0.07)	0.19(0.06)	0.06(0.01)	0.3(0.07)
	3	0.18	0.1	0.35	0.19	0.1	0.38	0.2	0.11	0.41
		0.2(0.04)	0.11(0.02)	0.4(0.08)	0.22(0.05)	0.12(0.03)	0.44(0.1)	0.23(0.05)	0.13(0.03)	0.47(0.11)
	4	18.78	1.47	21.26	38.26	2.87	43.07	57.95	4.31	65.22
		18.62(1.5)	1.47(0.13)	21.09(1.71)	38.24(2.74)	2.9(0.23)	43.15(3.09)	57.94(4.19)	4.3(0.35)	65.22(4.73)
	5	2.78	0.35	3.38	2.32	0.35	2.9	2.01	0.34	2.58
		2.75(0.33)	0.35(0.03)	3.34(0.38)	2.28(0.33)	0.34(0.03)	2.85(0.37)	1.98(0.33)	0.34(0.03)	2.55(0.38)
	1	0.1	0.05	0.2	0.1	0.05	0.2	0.11	0.06	0.21
		0.11(0.02)	0.06(0.01)	0.22(0.05)	0.11(0.02)	0.06(0.01)	0.23(0.05)	0.12(0.02)	0.07(0.01)	0.24(0.04)
	2	0.18	0.07	0.29	0.18	0.07	0.29	0.18	0.07	0.29
		0.18(0.06)	0.06(0.01)	0.29(0.07)	0.19(0.06)	0.06(0.01)	0.29(0.07)	0.19(0.06)	0.06(0.01)	0.3(0.07)
	3	0.19	0.1	0.38	0.2	0.11	0.41	0.21	0.11	0.42
		0.22(0.04)	0.12(0.02)	0.44(0.09)	0.23(0.05)	0.13(0.03)	0.48(0.1)	0.24(0.04)	0.13(0.03)	0.49(0.1)
(ii)	4	18.72	1.5	21.24	38.21	2.92	43.11	57.88	4.3	65.19
		18.76(1.48)	1.51(0.13)	21.3(1.68)	38.22(2.99)	2.92(0.25)	43.13(3.38)	57.87(4.37)	4.34(0.38)	65.17(4.98)
	5	2.74	0.36	3.34	2.29	0.35	2.88	1.97	0.35	2.55
		2.68(0.36)	0.36(0.03)	3.29(0.41)	2.24(0.35)	0.35(0.03)	2.84(0.4)	1.89(0.35)	0.34(0.03)	2.47(0.4)
	1	0.09	0.05	0.17	0.09	0.05	0.19	0.1	0.06	0.21
		0.09(0.02)	0.05(0.01)	0.19(0.04)	0.1(0.02)	0.05(0.01)	0.2(0.03)	0.11(0.01)	0.06(0.01)	0.21(0.03)
	2	0.18	0.07	0.29	0.18	0.07	0.29	0.18	0.07	0.29
		0.19(0.06)	0.06(0.01)	0.3(0.07)	0.18(0.06)	0.06(0.01)	0.29(0.07)	0.18(0.06)	0.06(0.01)	0.29(0.07)
	3	0.17	0.09	0.34	0.19	0.1	0.38	0.2	0.11	0.41
		0.19(0.04)	0.1(0.02)	0.38(0.08)	0.21(0.05)	0.11(0.03)	0.42(0.1)	0.18(0.06)	0.06(0.01)	0.29(0.07)
	4	18.84	1.48	21.29	38.31	2.88	43.1	57.77	4.34	65.18
		18.83(1.38)	1.47(0.12)	21.32(1.57)	38.21(2.95)	2.88(0.25)	43.05(3.35)	57.32(4.39)	4.25(0.37)	64.49(4.96)
	5	2.78	0.34	3.35	2.34	0.35	2.92	2.02	0.34	2.6
		2.75(0.34)	0.35(0.03)	3.34(0.38)	2.31(0.35)	0.34(0.03)	2.9(0.39)	2.03(0.34)	0.34(0.03)	2.6(0.39)
(iii)	1	0.17	0.14	0.39	0.31	0.34	0.91	0.5	0.54	1.5
		0.17(0.08)	0.09(0.03)	0.34(0.14)	0.27(0.22)	0.14(0.06)	0.53(0.32)	0.4(0.33)	0.21(0.09)	0.78(0.47)
	2	0.19	0.07	0.3	0.18	0.07	0.3	0.19	0.07	0.3
		0.19(0.06)	0.07(0.01)	0.31(0.07)	0.19(0.06)	0.07(0.01)	0.3(0.08)	0.19(0.06)	0.07(0.01)	0.3(0.07)
	3	0.32	0.28	0.75	0.62	0.64	1.73	0.99	1.06	2.97
		0.35(0.17)	0.19(0.07)	0.69(0.28)	0.6(0.44)	0.3(0.13)	1.15(0.64)	0.98(0.87)	0.45(0.21)	1.8(1.19)
	4	18.59	1.61	21.41	37.75	3.22	43.23	57.01	4.69	65
		18.45(1.69)	1.61(0.16)	21.18(1.93)	37.46(3.21)	3.16(0.29)	42.78(3.62)	56.94(4.6)	4.74(0.43)	64.99(5.25)
	5	2.54	0.41	3.21	1.79	0.47	2.53	1.24	0.49	2.03
		2.4(0.39)	0.39(0.04)	3.06(0.45)	1.58(0.41)	0.4(0.05)	2.27(0.47)	1.09(0.37)	0.4(0.06)	1.78(0.45)
(iv)	1	0.17	0.14	0.39	0.31	0.34	0.91	0.5	0.54	1.5
		0.17(0.08)	0.09(0.03)	0.34(0.14)	0.27(0.22)	0.14(0.06)	0.53(0.32)	0.4(0.33)	0.21(0.09)	0.78(0.47)
	2	0.19	0.07	0.3	0.18	0.07	0.3	0.19	0.07	0.3
		0.19(0.06)	0.07(0.01)	0.31(0.07)	0.19(0.06)	0.07(0.01)	0.3(0.08)	0.19(0.06)	0.07(0.01)	0.3(0.07)
	3	0.32	0.28	0.75	0.62	0.64	1.73	0.99	1.06	2.97
		0.35(0.17)	0.19(0.07)	0.69(0.28)	0.6(0.44)	0.3(0.13)	1.15(0.64)	0.98(0.87)	0.45(0.21)	1.8(1.19)
	4	18.59	1.61	21.41	37.75	3.22	43.23	57.01	4.69	65
		18.45(1.69)	1.61(0.16)	21.18(1.93)	37.46(3.21)	3.16(0.29)	42.78(3.62)	56.94(4.6)	4.74(0.43)	64.99(5.25)
	5	2.54	0.41	3.21	1.79	0.47	2.53	1.24	0.49	2.03
		2.4(0.39)	0.39(0.04)	3.06(0.45)	1.58(0.41)	0.4(0.05)	2.27(0.47)	1.09(0.37)	0.4(0.06)	1.78(0.45)

TABLE 3.6. Results for the stable rank confidence interval (95% nominal coverage probability)

ξ_1	Σ_n	$p/n = 0.5$		$p/n = 1$		$p/n = 1.5$	
		width/ r_n (%)	coverage (%)	width/ r_n (%)	coverage (%)	width/ r_n (%)	coverage (%)
(i)	1	2.00(0.12)	95.20	1.96(0.12)	95.40	1.97(0.12)	93.80
	2	14.53(1.19)	95.40	16.99(1.32)	91.60	18.7(1.34)	95.40
	3	2.00(0.12)	95.00	1.98(0.12)	97.40	1.97(0.12)	95.40
	4	34.68(2.87)	94.00	41.55(4.30)	94.00	43.82(5.33)	93.80
	5	5.18(0.65)	94.80	3.24(0.35)	95.20	2.64(0.24)	94.60
(ii)	1	2.07(0.12)	94.00	2.01(0.13)	92.80	1.98(0.12)	93.80
	2	14.68(1.17)	94.60	17.10(1.34)	95.00	18.75(1.42)	93.20
	3	2.08(0.13)	95.20	2.02(0.12)	95.20	2.00(0.13)	93.40
	4	34.86(2.76)	93.40	41.47(4.09)	94.80	44.99(5.34)	94.00
	5	5.12(0.63)	94.00	3.22(0.33)	93.40	2.63(0.24)	94.00
(iii)	1	1.96(0.12)	94.40	1.96(0.11)	96.00	1.96(0.11)	95.80
	2	14.45(1.11)	93.40	17.01(1.27)	94.60	18.57(1.42)	94.00
	3	1.97(0.12)	95.00	1.96(0.12)	93.20	1.96(0.12)	93.00
	4	34.71(2.67)	93.80	41.49(4.38)	92.80	44.37(5.21)	92.40
	5	5.17(0.64)	94.20	3.22(0.33)	95.80	2.61(0.24)	95.20
(iv)	1	2.38(0.21)	95.00	2.32(0.19)	94.80	2.31(0.25)	93.80
	2	15.31(1.19)	97.40	17.86(1.42)	93.80	19.53(1.49)	94.00
	3	2.4(0.20)	96.00	2.37(0.24)	95.00	2.33(0.27)	94.20
	4	35.43(2.90)	93.60	42.01(4.33)	93.60	44.59(5.06)	93.40
	5	5.20(0.64)	93.20	3.26(0.33)	93.00	2.73(0.29)	93.40

Stable rank test. Here, we discuss numerical results for an instance of the testing problem (3.24) given by

$$(3.30) \quad H_{0,n} : \frac{r_n}{p} \leq 0.1 \quad \text{vs.} \quad H_{1,n} : \frac{r_n}{p} > 0.1.$$

As a way to unify the study of level and power, we modified the experiments from Section 3.4.2 as follows. We rescaled the leading 15 eigenvalues in setting (1) to tune the ratio r_n/p within the grid $\{0.0980, 0.0985, \dots, 0.1045, 0.1050\}$. More precisely, the eigenvalues of Σ_n were taken to be of the form $\lambda_j(\Sigma_n) = 4s/3$ for $j \in \{1, \dots, 5\}$, $\lambda_j(\Sigma_n) = s$ for $j \in \{6, \dots, 15\}$ and $\lambda_j(\Sigma_n) = 1$ for $j \in \{16, \dots, p\}$, with different values of s being chosen to produce values of r_n/p matching the stated gridpoints. Hence, gridpoints less than 0.1 correspond to $H_{n,0}$, and gridpoints larger than 0.1 correspond to $H_{1,n}$.

At each gridpoint, we performed experiments based on the design of those in Section 3.4.2, allowing for p/n to take the values 0.5, 1.0, 1.5, and allowing the distribution of ξ_1 to be of the

types (i), (ii), (iii), and (iv). For each such setting, we applied the relevant bootstrap test from Section 3.3.4.2 at a 5% nominal level to 500 datasets, and then we recorded the rejection rate over the 500 trials. Figures 3.1, 3.2, 3.3, and 3.4 display the results by plotting the rejection rate as a function of the ratio r_n/p . The separate figures correspond to choices of the distribution of ξ_1 , and within each figure, three colored curves correspond to choices of p/n , as indicated in the legend. In addition, all the plots include a dashed horizontal line to signify the 5% nominal level.

An important feature that is shared by all the curves is that they stay below the nominal 5% level for essentially every value of $r_n/p < 0.1$, which corroborates our theoretical bound (3.28) in Theorem 3.3.3. Furthermore, when $r_n/p = 0.1$, the curves are mostly quite close to 5%, demonstrating that the testing procedure is well calibrated. For values of $r_n/p > 0.1$, the procedure exhibits substantial power, with the curve corresponding to $p/n = 0.5$ achieving approximately 100% power at $r_n/p = 0.105$ for every choice of ξ_1 . In the cases of $p/n \in \{1.0, 1.5\}$, the procedure still retains power, but with some diminution, as might be anticipated in settings of higher dimension.

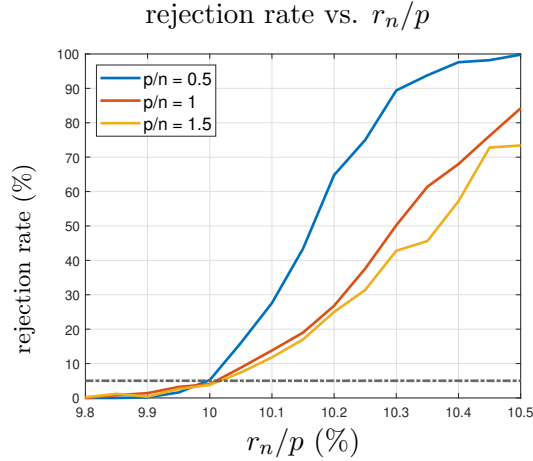


FIGURE 3.1. Results for the testing problem (3.30) when ξ_1^2 follows a Chi-Squared distribution with p degrees of freedom.

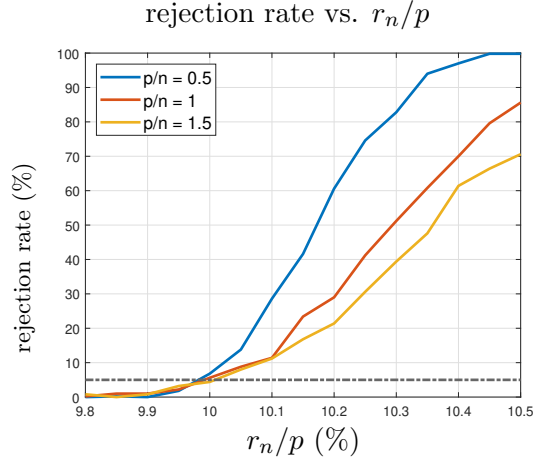


FIGURE 3.2. Results for the testing problem (3.30) when ξ_1^2 follows a Beta-Prime $\left(\frac{p(1+p+8)}{8}, \frac{1+p+16}{8}\right)$ distribution.

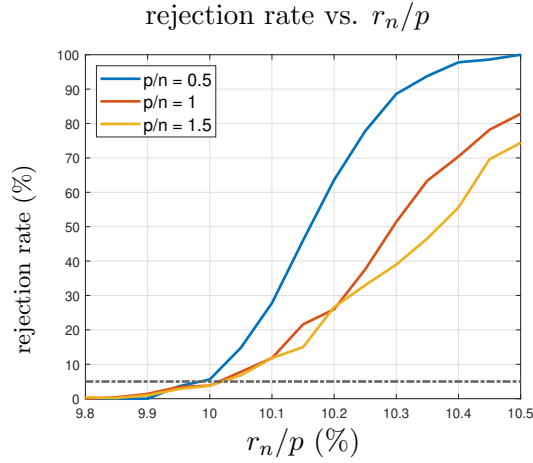


FIGURE 3.3. Results for the testing problem (3.30) when ξ_1^2 follows a $(p + 4)\text{Beta}(p/2, 2)$ distribution.

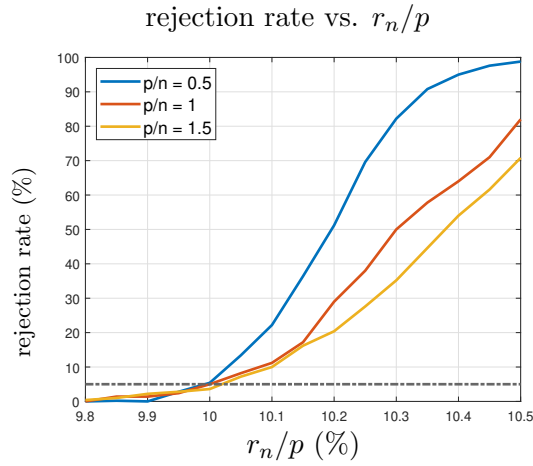


FIGURE 3.4. Results for the testing problem (3.30) when ξ_1^2 follows a $\frac{9p}{10}F(p, 20)$ distribution.

3.4.5. Sphericity test. Let $T_{\text{sr}} = \hat{r}_n/p - 1$ be a shorthand for the statistic that was introduced in Section 3.3.4.2 for testing sphericity. We now provide numerical comparisons with three other testing procedures based on linear spectral statistics. To define these other procedures, let

$$(3.31) \quad \hat{\mathcal{S}}_n = \frac{p}{n} \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\|\mathbf{x}_i\|_2^2}$$

denote the sample covariance matrix of rescaled observations, and consider the following three statistics,

$$(3.32) \quad T_1 = \frac{1}{p} \text{tr}(\hat{\mathcal{S}}_n^2) - c_n - 1$$

$$(3.33) \quad T_2 = \frac{1}{p} \text{tr}(\hat{\mathcal{S}}_n^4) - \frac{4c_n}{p} \text{tr}(\hat{\mathcal{S}}_n^3) - 2c_n \left(\frac{1}{p} \text{tr}(\hat{\mathcal{S}}_n^2) \right)^2 + \frac{10c_n^2}{p} \text{tr}(\hat{\mathcal{S}}_n^2) - 5c_n^3 - 1$$

$$(3.34) \quad T_3 = \max \left\{ \frac{nT_1 + 1}{2}, \frac{nT_2 + 6 - c_n}{\sqrt{8(18 + 12c_n + c_n^2)}} \right\}.$$

The testing procedures corresponding to these three statistics reject the sphericity hypothesis when the statistics take large values. The first two statistics can be attributed in part to the papers (Srivastava, 2005) and (Fisher et al., 2010), and the proposal of taking the maximum was made in (Tian et al., 2015). However, in all of these works, an ordinary sample covariance matrix was used in place of $\hat{\mathcal{S}}_n$. Variants of the definition of T_1 in (3.32) have been studied in (Paindaveine and Verdebout, 2016) and references therein, while the definitions T_2 and T_3 in (3.33) and (3.34) were proposed in (Hu et al., 2019a). The latter paper also derived the limiting null distributions of all three statistics in high-dimensional elliptical models.

Since numerical comparisons of the statistics T_1 , T_2 , and T_3 were given previously in (Hu et al., 2019a), our experiments here are designed using similar settings. Under the null hypothesis, we generated data from a standard multivariate normal distribution in 15 cases, corresponding to 3 choices of $p/n \in \{0.5, 1, 2\}$ and 5 choices of $n \in \{100, 200, 300, 400, 500\}$. For the statistics T_1 , T_2 , and T_3 , we used the analytical critical values derived previously in Hu et al. (2019a), and for the statistic T_{sr} , we determined its critical value using the proposed bootstrap with $B = 500$. For each setting under the null hypothesis, we generated 50000 datasets and calculated the empirical level of each test as the fraction of rejections among the 50000 trials. The results corresponding to a nominal level of 5% are displayed in Table 3.7, which shows that the empirical and nominal levels are in close agreement for all four statistics.

TABLE 3.7. Results on empirical level for sphericity tests (5% nominal level)

n	$p/n = 0.5$				$p/n = 1$				$p/n = 2$			
	T_1	T_2	T_3	T_{sr}	T_1	T_2	T_3	T_{sr}	T_1	T_2	T_3	T_{sr}
100	0.049	0.048	0.051	0.050	0.049	0.049	0.051	0.054	0.048	0.051	0.053	0.051
200	0.049	0.050	0.052	0.057	0.043	0.047	0.048	0.047	0.051	0.052	0.054	0.051
300	0.046	0.048	0.051	0.046	0.046	0.047	0.048	0.048	0.051	0.054	0.059	0.052
400	0.048	0.051	0.050	0.051	0.049	0.056	0.051	0.054	0.051	0.049	0.053	0.054
500	0.049	0.045	0.048	0.052	0.050	0.046	0.048	0.053	0.044	0.045	0.044	0.046

TABLE 3.8. Results on power for sphericity tests

n	$p/n = 0.5$				$p/n = 1$				$p/n = 2$			
	T_1	T_2	T_3	T_{sr}	T_1	T_2	T_3	T_{sr}	T_1	T_2	T_3	T_{sr}
100	0.190	0.165	0.185	0.217	0.201	0.154	0.190	0.211	0.208	0.136	0.194	0.207
200	0.497	0.391	0.466	0.502	0.513	0.350	0.473	0.503	0.503	0.271	0.465	0.520
300	0.793	0.660	0.766	0.797	0.801	0.589	0.763	0.812	0.793	0.458	0.752	0.804
400	0.952	0.863	0.941	0.952	0.951	0.790	0.933	0.954	0.958	0.670	0.941	0.949
500	0.993	0.958	0.990	0.994	0.993	0.922	0.989	0.993	0.993	0.816	0.987	0.995

Regarding the alternative hypothesis, we retained all the settings described above, except that we replaced the null covariance matrix $\Sigma_n = I_p$ with a diagonal spiked covariance matrix such that $\lambda_j(\Sigma_n) = 1.3$ for all $1 \leq j \leq p/2$, and $\lambda_j(\Sigma_n) = 1$ for all other j . This choice has the benefit that it creates variation in the numerical values of power, so that they are not too concentrated near 1. Similar alternatives were also used for the experiments in (Hu et al., 2019a). The results are presented in Table 3.8, which is organized in the same format as Table 3.7. In each setting, the power of the statistic T_{sr} approximately matches the highest power achieved among T_1, T_2 , and T_3 .

3.5. Conclusion

Up to now, high-dimensional elliptical models have generally fallen outside the scope of existing bootstrap methods for spectral statistics. In the current paper, we have addressed this problem by showing how a parametric bootstrap approach that is specialized to IC models (Lopes et al., 2019) can be extended to elliptical models in high dimensions. In addition, we have shown that the new method is supported by two types of theoretical guarantees in the elliptical setting: First, the method consistently approximates the distributions of linear spectral statistics (Theorem 3.3.2). Second, the method can be applied to a nonlinear combination of these statistics to construct

asymptotically valid confidence intervals and hypothesis tests (Theorem 3.3.3). From a practical perspective, a valuable property of the method is its user-friendliness, since it can be applied to generic spectral statistics in an automatic way. In particular, this provides the user with the flexibility to easily explore spectral statistics whose asymptotic distributions are analytically inconvenient or unknown. With regard to empirical performance, we have presented extensive simulation results, showing that with few exceptions, the method accurately approximates the distributions of both linear and nonlinear spectral statistics across many settings and inference tasks. An interesting question for future work is to determine if a consistent parametric bootstrap method for spectral statistics can be developed within more general models that unify IC and elliptical models.

APPENDIX A

Supplementary Material for Chapter 2

In Appendices A.1-A.4, we prove our main result (Theorem 3.3.2), which follows from Proposition 1 (Appendix A.2) and Proposition 6 (Appendix A.3) as explained in the main text. In Appendix A.1, we introduce some notation and frequently used facts. In Appendix A.2, we establish the asymptotic normality of T_n by analyzing $\tilde{\kappa}$ and $\tilde{\kappa}$ in Appendices A.2.1 and A.2.2 respectively. The ratio-consistency of $\hat{\sigma}_n^2$ is shown in Appendix A.3. Theoretical background results are given in Appendix A.4. Lastly, additional empirical results are given in Appendix A.5.

A.1. Preliminaries

For two sequences of random variables $\{y_n\}$ and $\{v_n\}$, the relation $y_n = o_{\mathbb{P}}(v_n)$ means that $y_n/v_n \xrightarrow{\mathbb{P}} 0$, and the relation $y_n = \mathcal{O}_{\mathbb{P}}(v_n)$ means that for all $\epsilon > 0$, there is a constant M not depending on n such that $\sup_{n \geq 1} \mathbb{P}(|y_n/v_n| > M) < \epsilon$. The $k \times k$ identity matrix is denoted as I_k , and the indicator function for a condition $\cdots$ is denoted as $1\{\cdots\}$. The j th canonical basis vector in an Euclidean space is denoted as $\mathbf{e}_j$, with the dimension being understood from context. For two matrices A and B of the same size, their Hadamard product $A \circ B$ is the matrix of that size whose ij entry is given by $(A \circ B)_{ij} = A_{ij}B_{ij}$, and $A^{\circ k}$ denotes the k -fold Hadamard product of A with itself. The stable rank of the covariance matrix Σ is defined as $\mathbf{r}(\Sigma) = \frac{\text{tr}(\Sigma)^2}{\text{tr}(\Sigma^2)}$. The correlation matrix associated to Σ is denoted by R .

Next, let $\mathbf{z}_1, \dots, \mathbf{z}_n$, denote i.i.d. standard normal vectors in $\mathbb{R}^p$ independent of $\xi_1, \dots, \xi_n$, and for each $i = 1, \dots, n$ let

$$(A.1) \quad \boldsymbol{\zeta}_i = \Sigma^{1/2} \mathbf{z}_i.$$

This definition allows us to express the i th observation $\mathbf{x}_i$ as

$$(A.2) \quad \mathbf{x}_i = \frac{\xi_i}{\|\mathbf{z}_i\|_2} \Sigma^{1/2} \mathbf{z}_i = \frac{\xi_i}{\|\mathbf{z}_i\|_2} \boldsymbol{\zeta}_i$$

for $i = 1, \dots, n$. For any $k \geq 0$, we define the parameter

$$(A.3) \quad r_k = \frac{\mathbb{E}(\|\mathbf{x}_1\|_2^{2k})}{\mathbb{E}((\mathbf{z}_1^\top \Sigma \mathbf{z}_1)^k)},$$

which can also be expressed as

$$(A.4) \quad r_k = \frac{\mathbb{E}(\xi_1^{2k})}{\mathbb{E}(\|\mathbf{z}_1\|_2^{2k})}.$$

In particular, we will often use the fact that Assumption 1 implies

$$(A.5) \quad r_2 = 1 + \frac{\tau - 2}{p} + o\left(\frac{1}{p}\right).$$

By noting that $\zeta_1 \sim N(0, \Sigma)$ and $\zeta_{1j} \sim N(0, \Sigma_{jj})$, we have the following moment formula for any integer $k \geq 0$ and $j = 1, \dots, p$,

$$\begin{aligned} \mathbb{E}(x_{1j}^{2k}) &= \mathbb{E}\left(\frac{\xi_1^{2k}}{\|\mathbf{z}_1\|_2^{2k}} \zeta_{1j}^{2k}\right) \\ &= \frac{\mathbb{E}(\xi_1^{2k})}{\mathbb{E}(\|\mathbf{z}_1\|_2^{2k})} \mathbb{E}(\zeta_{1j}^{2k}) \\ &= (2k - 1)!! r_k \Sigma_{jj}^k, \end{aligned}$$

where $m!! = m(m - 2)(m - 4) \cdots 1$ for an odd integer m . In the special case when the diagonal entries of Σ are equal to 1, we will often use the fact that

$$(A.6) \quad \max_{1 \leq j \leq p} \mathbb{E}(x_{1j}^{2k}) \lesssim 1$$

holds for $k = 1, \dots, 8$, which is established based on Lemma B.7.

A.2. Asymptotic normality of T_n

The following proposition is a counterpart to Theorem 3.3.2 in which $\widehat{\sigma}_n$ is replaced by σ_n .

PROPOSITION A.2.1. *If Assumption 1 holds, then as $n \rightarrow \infty$,*

$$(A.7) \quad \frac{T_n}{\sigma_n} \Rightarrow N(0, 1).$$

PROOF. We decompose the statistic T_n as

$$\begin{aligned} T_n &= \sqrt{\frac{pn}{2}} \left(\frac{\tilde{\kappa} - \kappa}{3} + \frac{4}{n} \right) - \sqrt{\frac{pn}{2}} \left(\frac{\tilde{\kappa} - \kappa}{3} \right) \\ &=: T_{n,1} - T_{n,2}. \end{aligned}$$

Also, the formula for κ given in (2.3) allows us to write

$$(A.8) \quad T_{n,1} = \sqrt{\frac{pn}{2}} \left(\frac{\tilde{\kappa}}{3} - 1 \right) - \sqrt{\frac{pn}{2}} \left(\frac{\kappa}{3} - 1 \right) + 2\sqrt{\frac{2p}{n}}$$

$$(A.9) \quad T_{n,2} = \sqrt{\frac{pn}{2}} \left(\frac{\zeta^2 - 2(\tilde{\nu}_2 - \frac{2}{n}\tilde{\nu}_1^2)}{\tilde{\nu}_1^2 + 2(\tilde{\nu}_2 - \frac{2}{n}\tilde{\nu}_1^2)} - \frac{\zeta^2 - 2\nu_2}{\nu_1^2 + 2\nu_2} \right).$$

We analyze the asymptotic properties of $T_{n,1}$ and $T_{n,2}$ separately. Proposition A.2.2 shows that as $n \rightarrow \infty$,

$$(A.10) \quad \frac{T_{n,1}}{\sigma_{n,1}} \Rightarrow N(0, 1).$$

Regarding $T_{n,2}$, Proposition A.2.4 shows that if $\mathbf{r}(\Sigma) \lesssim \sqrt{p}$, then as $n \rightarrow \infty$,

$$\frac{T_{n,2}}{\sigma_{n,2}} \Rightarrow N(0, 1).$$

Since $T_{n,1}$ and $T_{n,2}$ are constructed from independent sets of observations, it follows that $(\frac{1}{\sigma_{n,1}}T_{n,1}, \frac{1}{\sigma_{n,2}}T_{n,2})$ converges weakly to a standard bivariate normal distribution, which implies the stated result (A.7) when $\mathbf{r}(\Sigma) \lesssim \sqrt{p}$.

When $\mathbf{r}(\Sigma)/\sqrt{p} \rightarrow \infty$, Proposition A.2.5 shows that

$$(A.11) \quad \frac{T_{n,2}}{\sigma_n} \xrightarrow{\mathbb{P}} 0.$$

Also, when $\mathbf{r}(\Sigma)/\sqrt{p} \rightarrow \infty$, it can be checked that $\frac{\sigma_{n,1}^2}{\sigma_n^2} \rightarrow 1$. Therefore, the limits (A.10) and (A.11) imply (A.7) because

$$\frac{T_{n,1} - T_{n,2}}{\sigma_n} = \frac{T_{n,1}}{\sigma_{n,1}} + o_{\mathbb{P}}(1) \Rightarrow N(0, 1).$$

□

A.2.1. Asymptotic normality of $T_{n,1}$.

PROPOSITION A.2.2. *If Assumption 1 holds, then as $n \rightarrow \infty$,*

$$\frac{T_{n,1}}{\sigma_{n,1}} \Rightarrow N(0, 1).$$

PROOF. Recall from equation (A.8) that $T_{n,1}$ may be expressed as

$$T_{n,1} = \sqrt{\frac{pn}{2}} \left(\frac{\tilde{\kappa}}{3} - 1 \right) - \sqrt{\frac{pn}{2}} \left(\frac{\kappa}{3} - 1 \right) + 2\sqrt{\frac{2p}{n}}.$$

We begin by analyzing the first term on the right side. Since each $\tilde{\kappa}_j$ in the definition of $\tilde{\kappa}$ is invariant to rescaling the coordinates of each observation, we may assume that all the diagonal entries of Σ are equal to 1. Let $\tilde{\Sigma}$ denote the sample covariance matrix of the first $n/2$ observations so that

$$\tilde{\Sigma}_{jj} = \frac{1}{n/2} \sum_{i=1}^{n/2} x_{ij}^2.$$

Also, for each $j = 1, \dots, p$, define the event

$$(A.12) \quad A_j = \{ |\tilde{\Sigma}_{jj} - 1| \leq \epsilon_n \},$$

where we put $\epsilon_n = n^{-1/20}$. It is shown in Lemma B.1 that

$$(A.13) \quad \sqrt{\frac{pn}{2}} \left(\frac{\tilde{\kappa}}{3} - 1 \right) = \sqrt{\frac{n}{2p}} \sum_{j=1}^p \left(\frac{\tilde{\kappa}_j}{3} - 1 \right) 1\{A_j\} + o_{\mathbb{P}}(1).$$

Next, define the random variables

$$(A.14) \quad \ell_j = \left(\frac{1/3}{n/2} \sum_{i=1}^{n/2} (x_{ij}^4 - \mathbb{E}(x_{1j}^4)) \right) - 2(\tilde{\Sigma}_{jj} - 1)$$

$$(A.15) \quad \Delta_j = -(\tilde{\Sigma}_{jj} - 1)^2 + \frac{\mathbb{E}(x_{ij}^4)}{3} - 1$$

$$(A.16) \quad \varepsilon_j = \left(\frac{1\{A_j\}}{\tilde{\Sigma}_{jj}^2} - 1 \right) 1\{\tilde{\Sigma}_{jj} \neq 0\},$$

which satisfy the algebraic relation

$$\left(\frac{\tilde{\kappa}_j}{3} - 1 \right) 1\{A_j\} = (1 + \varepsilon_j)(\ell_j + \Delta_j).$$

It follows from Lemma B.2 that

$$\sqrt{\frac{n}{2p\sigma_{n,1}^2}} \sum_{j=1}^p \varepsilon_j(\ell_j + \Delta_j) = o_{\mathbb{P}}(1),$$

and so after combining with (A.13), we have

$$\sqrt{\frac{pn}{2\sigma_{n,1}^2}} \left(\frac{\tilde{\kappa}}{3} - 1 \right) = \sqrt{\frac{n}{2p\sigma_{n,1}^2}} \sum_{j=1}^p (\ell_j + \Delta_j) + o_{\mathbb{P}}(1).$$

In addition, Lemma B.5 shows that

$$\sqrt{\frac{n}{2p\sigma_{n,1}^2}} \sum_{j=1}^p \left(\Delta_j - \left[\frac{\kappa}{3} - 1 - \frac{4}{n} \right] \right) \xrightarrow{\mathbb{P}} 0,$$

which implies that the normalized statistic $T_{n,1}/\sigma_{n,1}$ satisfies

$$(A.17) \quad \frac{T_{n,1}}{\sigma_{n,1}} = \sqrt{\frac{n}{2p\sigma_{n,1}^2}} \sum_{j=1}^p \ell_j + o_{\mathbb{P}}(1).$$

Finally, Lemma B.4 shows that

$$\sqrt{\frac{n}{2p\sigma_{n,1}^2}} \sum_{j=1}^p \ell_j \Rightarrow N(0, 1),$$

and combining this with (A.17) completes the proof. $\square$

A.2.1.1. *Negligible terms associated with $T_{n,1}$.* Since the proof of Proposition A.2.2 is based on a reduction to the case where all the diagonal entries of Σ are 1, this condition will be assumed in all of the supporting results for Proposition A.2.2 in this and the next subsection. Also, for the following lemma, recall that the event A_j is defined in equation (A.12).

LEMMA B.1. *If Assumption 1 holds and all the diagonal entries of Σ are 1, then*

$$\sum_{j=1}^p \left(\frac{\tilde{\kappa}_j}{3} - 1 \right) 1\{A_j^c\} = o_{\mathbb{P}}(1).$$

PROOF. For any $j = 1, \dots, p$, it can be checked that the inequalities $1 \leq \tilde{\kappa}_j \leq n$ hold almost surely, yielding

$$\sum_{j=1}^p \left| \frac{\tilde{\kappa}_j}{3} - 1 \right| 1\{A_j^c\} \leq n \sum_{j=1}^p 1\{A_j^c\}.$$

Hence, it is enough to show $\sum_{j=1}^p \mathbb{P}(A_j^c) = o(\frac{1}{n})$. We first apply Chebyshev's inequality to each $\mathbb{P}(A_j^c)$, which leads to

$$\mathbb{P}(A_j^c) \leq n^{2/5} \left\| \frac{2}{n} \sum_{i=1}^{n/2} (x_{ij}^2 - 1) \right\|_{L^8}^8.$$

The L^8 norm of a sum of centered independent random variables can be bounded with Rosenthal's inequality (Lemma D.7) to obtain

$$\begin{aligned} \mathbb{P}(A_j^c) &\lesssim n^{2/5} \max \left\{ \left(\frac{2}{n} \text{var}(x_{1j}^2) \right)^4, \left(\frac{2}{n} \right)^7 \|x_{1j}^2 - 1\|_{L^8}^8 \right\} \\ (A.18) \quad &\lesssim n^{-18/5}, \end{aligned}$$

where the last step relies on the condition $\Sigma_{jj} = 1$ for all $j = 1, \dots, p$, as well as the moment bounds given in (A.6). Thus, we conclude that $\sum_{j=1}^p \mathbb{P}(A_j^c) = o(\frac{1}{n})$, as needed. $\square$

For the next lemma, recall that the random variables ℓ_j , Δ_j , and ε_j are defined in equations (A.14), (A.15) and (A.16).

LEMMA B.2. *If Assumption 1 holds and the diagonal entries of Σ are equal to 1, then as $n \rightarrow \infty$,*

$$\frac{1}{\sigma_{n,1}} \sum_{j=1}^p \varepsilon_j (\ell_j + \Delta_j) \xrightarrow{\mathbb{P}} 0.$$

PROOF. For any $a > 0$, consider the following algebraic identity, which arises from a Taylor expansion of the inverse square function at $a = 1$,

$$\frac{1}{a^2} - 1 = -2(a - 1) + \left(\frac{2}{a} + \frac{1}{a^2} \right) (a - 1)^2.$$

Letting $\tilde{\Sigma}_{jj}$ play the role of a leads to the following almost-sure expansion for ε_j ,

$$\varepsilon_j = 1\{A_j\} \left(-2(\tilde{\Sigma}_{jj} - 1) + \left(\frac{2}{\tilde{\Sigma}_{jj}} + \frac{1}{\tilde{\Sigma}_{jj}^2} \right) (\tilde{\Sigma}_{jj} - 1)^2 \right) - 1\{A_j^c\} 1\{\tilde{\Sigma}_{jj} \neq 0\}.$$

Since the key term in this expansion is $-2(\tilde{\Sigma}_{jj} - 1)$, we define the random variable η_j to be the remainder that satisfies

$$(A.19) \quad \varepsilon_j = -2(\tilde{\Sigma}_{jj} - 1) + \eta_j,$$

and therefore, the sum in the stated result may be expressed as

$$\frac{1}{\sigma_{n,1}} \sum_{j=1}^p \varepsilon_j(\ell_j + \Delta_j) = -\frac{2}{\sigma_{n,1}} \sum_{j=1}^p (\tilde{\Sigma}_{jj} - 1)(\ell_j + \Delta_j) + \frac{1}{\sigma_{n,1}} \sum_{j=1}^p \eta_j(\ell_j + \Delta_j).$$

The problem of showing that the first sum on the right is $o_{\mathbb{P}}(1)$ is quite involved, and so we defer the details to Lemma B.3. The rest of this proof will show that $\sum_{j=1}^p \eta_j(\ell_j + \Delta_j)$ is $o_{\mathbb{P}}(1)$, which is sufficient because $\sigma_{n,1}^2 \gtrsim 1$. To do this, we start from the observation that the L^1 norm of this sum is upper bounded by $\sum_{j=1}^p \|\eta_j\|_{L^2} \|\ell_j + \Delta_j\|_{L^2}$. Next, from the definition of η_j in (A.19), it follows that

$$\|\eta_j\|_{L^2} \lesssim \|\tilde{\Sigma}_{jj} - 1\|_{L^4}^2 + (\mathbb{P}(A_j^c))^{1/4} \|\tilde{\Sigma}_{jj} - 1\|_{L^4} + (\mathbb{P}(A_j^c))^{1/2}.$$

From (A.18), the order of the third term on the right is $(\mathbb{P}(A_j^c))^{1/2} \lesssim n^{-9/5}$. With regard to the first and second terms, Rosenthal's inequality (recorded in Lemma D.7) gives

$$\begin{aligned} \|\tilde{\Sigma}_{jj} - 1\|_{L^4} &= \left\| \frac{2}{n} \sum_{i=1}^{n/2} (x_{ij}^2 - 1) \right\|_{L^4} \\ (A.20) \quad &\lesssim \max \left\{ \left(\frac{2}{n} \text{var}(x_{1j}^2) \right)^{1/2}, \left(\frac{2}{n} \right)^{3/4} \|x_{1j}^2 - 1\|_{L^4} \right\} \\ &\lesssim \frac{1}{\sqrt{n}}, \end{aligned}$$

where we have used the condition that the diagonal entries of Σ are equal to 1, as well as moment bounds in (A.6). It follows that $\|\eta_j\|_{L^2} \lesssim \frac{1}{n}$.

Turning to $\|\ell_j + \Delta_j\|_{L^2}$, the definitions of ℓ_j and Δ_j given in (A.14) and (A.15), as well as the triangle inequality for the L^2 norm yield

$$\|\ell_j + \Delta_j\|_{L^2} \leq \sqrt{\frac{2}{n} \text{var}\left(\frac{1}{3}x_{1j}^4 - 2x_{1j}^2\right)} + \|\tilde{\Sigma}_{jj} - 1\|_{L^4}^2 + |r_2 - 1|.$$

The first term on the right can be handled with (A.6), which shows that if the diagonal entries of Σ are equal to 1, then $\text{var}\left(\frac{1}{3}x_{1j}^4 - 2x_{1j}^2\right) \lesssim 1$. Combining with equations (A.20) and (A.5) yields $\|\ell_j + \Delta_j\|_{L^2} \lesssim \frac{1}{\sqrt{n}}$. Finally, the previous bounds on $\|\eta_j\|_{L^2}$ and $\|\ell_j + \Delta_j\|_{L^2}$ show that the norm $\left\| \sum_{j=1}^p \eta_j(\ell_j + \Delta_j) \right\|_{L^1}$ is of order $\frac{1}{\sqrt{n}}$, which completes the proof. $\square$

LEMMA B.3. *If Assumption 1 holds and the diagonal entries of Σ are equal to 1, then as $n \rightarrow \infty$*

$$(A.21) \quad \frac{1}{\sigma_{n,1}} \sum_{j=1}^p (\tilde{\Sigma}_{jj} - 1)(\ell_j + \Delta_j) \xrightarrow{\mathbb{P}} 0.$$

PROOF. It suffices to show that the expectation and variance of the left side of (A.21) are both $o(1)$. To handle the expectation, observe that

$$(\tilde{\Sigma}_{jj} - 1)(\ell_j + \Delta_j) = \left(\frac{2}{n} \sum_{i=1}^{n/2} (x_{ij}^2 - 1) \right) \left(\frac{2}{3n} \sum_{i=1}^{n/2} x_{ij}^4 - \left(\frac{2}{n} \sum_{i=1}^{n/2} x_{ij}^2 \right)^2 \right),$$

and so

$$\begin{aligned} \mathbb{E}\left((\tilde{\Sigma}_{jj} - 1)(\ell_j + \Delta_j)\right) &= \frac{1}{3(n/2)^2} \sum_{1 \leq i_1, i_2 \leq n/2} \mathbb{E}(x_{i_1 j}^2 x_{i_2 j}^4) - \frac{1}{3(n/2)} \sum_{i_1=1}^{n/2} \mathbb{E}(x_{i_1 j}^4) \\ &\quad - \frac{1}{(n/2)^3} \sum_{1 \leq i_1, i_2, i_3 \leq n/2} \mathbb{E}(x_{i_1 j}^2 x_{i_2 j}^2 x_{i_3 j}^2) + \frac{1}{(n/2)^2} \sum_{1 \leq i_1, i_2 \leq n/2} \mathbb{E}(x_{i_1 j}^2 x_{i_2 j}^2). \end{aligned}$$

When the diagonal entries of Σ are equal to 1, the expressions for $\mathbf{x}_i$ and r_k in (A.2) and (A.4) imply $\mathbb{E}(x_{1j}^4) = 3r_2$, $\mathbb{E}(x_{1j}^6) = 15r_3$, as well as

$$\begin{aligned} \sum_{1 \leq i_1, i_2 \leq n/2} \mathbb{E}(x_{i_1 j}^2 x_{i_2 j}^4) &= \frac{n}{2} \mathbb{E}(x_{1j}^6) + \frac{n}{2} \left(\frac{n}{2} - 1 \right) \mathbb{E}(x_{1j}^4) \mathbb{E}(x_{1j}^2) \\ &= \frac{15n}{2} r_3 + \frac{3n}{2} \left(\frac{n}{2} - 1 \right) r_2, \\ \sum_{1 \leq i_1, i_2, i_3 \leq n/2} \mathbb{E}(x_{i_1 j}^2 x_{i_2 j}^2 x_{i_3 j}^2) &= \frac{n}{2} \mathbb{E}(x_{1j}^6) + \frac{3n}{2} \left(\frac{n}{2} - 1 \right) \mathbb{E}(x_{1j}^4) \mathbb{E}(x_{1j}^2) + \frac{n}{2} \left(\frac{n}{2} - 1 \right) \left(\frac{n}{2} - 2 \right) \mathbb{E}(x_{1j}^2)^3 \\ &= \frac{15n}{2} r_3 + \frac{9n}{2} \left(\frac{n}{2} - 1 \right) r_2 + \frac{n}{2} \left(\frac{n}{2} - 1 \right) \left(\frac{n}{2} - 2 \right), \\ \sum_{1 \leq i_1, i_2 \leq n/2} \mathbb{E}(x_{i_1 j}^2 x_{i_2 j}^2) &= \frac{n}{2} \mathbb{E}(x_{1j}^4) + \frac{n}{2} \left(\frac{n}{2} - 1 \right) \mathbb{E}(x_{1j}^2)^2 \\ &= \frac{3n}{2} r_2 + \frac{n}{2} \left(\frac{n}{2} - 1 \right). \end{aligned}$$

Combining with Lemma B.7, which shows that r_2 and r_3 are both equal to $1 + o(p^{-3/4})$, we have

$$\begin{aligned} \mathbb{E}\left((\tilde{\Sigma}_{jj} - 1)(\ell_j + \Delta_j)\right) &= \left(r_2 + \frac{5r_3 - r_2}{n/2} \right) - r_2 - \left(1 + \frac{9r_2 - 3}{n/2} + \frac{15r_3 - 9r_2 + 2}{(n/2)^2} \right) + \left(1 + \frac{3r_2 - 1}{n/2} \right) \\ &= o\left(\frac{1}{n}\right). \end{aligned}$$

So, by summing over j and noting that $\sigma_{n,1}^2 \gtrsim 1$, we conclude

$$\frac{1}{\sigma_{n,1}} \sum_{j=1}^p \mathbb{E}\left((\tilde{\Sigma}_{jj} - 1)(\ell_j + \Delta_j)\right) = o(1).$$

Next, we show that the variance of the left side of (A.21) is $o(1)$. For this purpose, we begin by noting that

$$\sum_{j=1}^p (\tilde{\Sigma}_{jj} - 1)(\ell_j + \Delta_j) = \sum_{j=1}^p (\tilde{\Sigma}_{jj} - 1) \left(\frac{2}{3n} \sum_{i=1}^{n/2} (x_{ij}^4 - \mathbb{E}(x_{1j}^4)) - 2(\tilde{\Sigma}_{jj} - 1) - (\tilde{\Sigma}_{jj} - 1)^2 + r_2 - 1 \right).$$

Due to the basic inequality $\text{var}(\sum_{l=1}^4 V_l) \leq 4 \sum_{l=1}^4 \text{var}(V_l)$ that holds for generic random variables $V_1, \dots, V_4$, as well as the relation $\sigma_1^2 = \frac{8}{3p} \|R\|_4^4$, it is sufficient to establish the following four limits,

$$(A.22) \quad \text{var} \left(\sum_{j=1}^p (\tilde{\Sigma}_{jj} - 1) \left(\frac{1}{3(n/2)} \sum_{i=1}^{n/2} (x_{ij}^4 - \mathbb{E}(x_{1j}^4)) \right) \right) = o(\|R\|_4^4/p),$$

$$(A.23) \quad \text{var} \left(\sum_{j=1}^p (\tilde{\Sigma}_{jj} - 1)^2 \right) = o(\|R\|_4^4/p),$$

$$(A.24) \quad \text{var} \left(\sum_{j=1}^p (\tilde{\Sigma}_{jj} - 1)^3 \right) = o(\|R\|_4^4/p),$$

$$(A.25) \quad (r_2 - 1)^2 \text{var} \left(\sum_{j=1}^p (\tilde{\Sigma}_{jj} - 1) \right) = o(\|R\|_4^4/p).$$

Proof of (A.22). The definition of $\tilde{\Sigma}_{jj}$ implies that the left hand side of (A.22) is equal to

$$(A.26) \quad \frac{1}{9(n/2)^4} \sum_{1 \leq i_1, \dots, i_4 \leq n/2} \sum_{j,k=1}^p \text{cov} \left((x_{i_1 j}^4 - \mathbb{E}(x_{1j}^4))(x_{i_2 j}^2 - 1), (x_{i_3 k}^4 - \mathbb{E}(x_{1k}^4))(x_{i_4 k}^2 - 1) \right).$$

The covariance terms in (A.26) are non-zero only when $|\{i_1, i_2\} \cap \{i_3, i_4\}|$ is equal to 1 or 2. When $|\{i_1, i_2\} \cap \{i_3, i_4\}| = 2$, there are two possible types of tuples $(i_1, \dots, i_4)$, corresponding to $i_1 = i_3 \neq i_2 = i_4$ and $i_1 = i_4 \neq i_2 = i_3$, with the number of tuples of each type being $\mathcal{O}(n^2)$. When the covariance terms corresponding to the first type are summed over (j, k) , Lemmas D.6 and B.7 show that

$$\begin{aligned} \sum_{j,k=1}^p |\text{cov}(x_{1j}^4, x_{1k}^4) \text{cov}(x_{1j}^2, x_{1k}^2)| &\lesssim p^2 |(r_4 - r_2^2)(r_2 - 1)| + \|R\|_2^2 \\ &= o(p\|R\|_4^4), \end{aligned}$$

where the last step use the general fact that $\|R\|_2^2 \lesssim p^{1/2} \|R\|_4^4$. Likewise, these lemmas show that when the covariance terms corresponding to the second type of tuples $(i_1, \dots, i_4)$ are summed over

(j, k) we have

$$\begin{aligned} \sum_{j,k=1}^p |\text{cov}(x_{1j}^4, x_{1k}^2) \text{cov}(x_{1j}^2, x_{1k}^4)| &\lesssim p^2(r_3 - r_2)^2 + \|R\|_2^2 \\ &= o(p\|R\|_4^4). \end{aligned}$$

When $|\{i_1, i_2\} \cap \{i_3, i_4\}| = 1$, there are three possible types of tuples $(i_1, \dots, i_4)$, in which there are 1, 2, or 3 distinct indices. If there are 2 or 3 distinct indices among $(i_1, \dots, i_4)$, then the associated covariance terms in (A.26) are zero. With regard to tuples $(i_1, \dots, i_4)$ in which all indices are equal, there are $\mathcal{O}(n)$ such tuples. Also, equation (A.6) implies that the associated covariance terms in (A.26) are $\mathcal{O}(1)$, uniformly over (j, k) . So, the sum of these covariance terms over (j, k) is $\mathcal{O}(p^2)$. Combining the last several steps shows that the total sum in (A.26) is $\mathcal{O}(n^{-4}[n^2 o(p\|R\|_4^4) + np^2]) = o(\|R\|_4^4/p)$, which proves (A.22).

Proof of (A.23). We proceed analogously to the work above. The left hand side of (A.23) is equal to

$$(A.27) \quad \frac{1}{(n/2)^4} \sum_{1 \leq i_1, i_2, i_3, i_4 \leq n/2} \sum_{j,k=1}^p \text{cov}\left((x_{i_1 j}^2 - 1)(x_{i_2 j}^2 - 1), (x_{i_3 k}^2 - 1)(x_{i_4 k}^2 - 1)\right).$$

As before, it is enough to restrict our attention to the tuples $(i_1, \dots, i_4)$ for which the cardinality $|\{i_1, i_2\} \cap \{i_3, i_4\}|$ is equal to 1 or 2. When $|\{i_1, i_2\} \cap \{i_3, i_4\}| = 2$, it follows from equation (A.5) and Lemma D.6 that the sum of the corresponding covariance terms over (j, k) is equal to

$$\begin{aligned} \sum_{j,k=1}^p \left(\text{cov}(x_{1j}^2, x_{1k}^2) \right)^2 &\lesssim p^2(r_2 - 1)^2 + \|R\|_2^2 \\ &= o(p\|R\|_4^4). \end{aligned}$$

When $|\{i_1, i_2\} \cap \{i_3, i_4\}| = 1$, the associated tuples $(i_1, \dots, i_4)$ may have 1, 2, or 3 distinct indices. If there are 2 or 3 distinct indices, then the corresponding covariance terms in (A.27) are zero, and so it is enough to consider the tuples for which all indices are equal. Due to equation (A.6), the covariance terms in (A.27) corresponding to these tuples are $\mathcal{O}(1)$, uniformly over (j, k) , and so the sum of these terms over (j, k) is $\mathcal{O}(p^2)$. Altogether, it follows that the total sum in (A.27) is $\mathcal{O}(n^{-4}[n^2 o(p\|R\|_4^4) + np^2]) = o(\|R\|_4^4/p)$, which proves (A.23).

Proof of (A.24). Using the triangle inequality for the L^2 norm, followed by Rosenthal's inequality (Lemma D.7) and equation (A.6), the left hand side of (A.24) can be bounded as

$$\begin{aligned} \text{var} \left(\sum_{j=1}^p (\tilde{\Sigma}_{jj} - 1)^3 \right) &\leq \left(\sum_{j=1}^p \|\tilde{\Sigma}_{jj} - 1\|_{L^6}^3 \right)^2 \\ &\lesssim p^2 \left(\max_{1 \leq j \leq p} \max \left\{ \left(\frac{1}{n/2} \text{var}(x_{1j}^2) \right)^{1/2}, \left(\frac{1}{n/2} \right)^{-5/6} \|x_{1j}^2 - 1\|_{L^6} \right\} \right)^6 \\ &\lesssim \frac{1}{n}, \end{aligned}$$

which implies (A.24).

Proof of (A.25). Similarly to the previous argument, the left hand side of and (A.25) can be bounded as

$$\begin{aligned} (r_2 - 1)^2 \text{var} \left(\sum_{j=1}^p (\tilde{\Sigma}_{jj} - 1) \right) &\lesssim \frac{1}{p^2} \left(\sum_{j=1}^p \|\tilde{\Sigma}_{jj} - 1\|_{L^2} \right)^2 \\ &= \frac{1}{p^2} \left(\sum_{j=1}^p \sqrt{\frac{2}{n} \text{var}(x_{1j}^2)} \right)^2 \\ &\lesssim \frac{1}{n}, \end{aligned}$$

which implies (A.25) and completes the proof of the lemma. $\square$

A.2.1.2. Main terms of $T_{n,1}$.

LEMMA B.4. *If Assumption 1 holds and the diagonal entries of Σ are equal to 1, then as $n \rightarrow \infty$,*

$$\sqrt{\frac{n}{2p\sigma_{n,1}^2}} \sum_{j=1}^p \ell_j \Rightarrow N(0, 1).$$

PROOF. We begin the proof by expressing $\sum_{j=1}^p \ell_j$ in terms of a sum of centered i.i.d. random variables $w_1, \dots, w_n$ defined according to

$$w_i = \frac{1}{\sqrt{p}} \sum_{j=1}^p \frac{1}{3} (x_{ij}^4 - \mathbb{E}(x_{1j}^4)) - 2(x_{ij}^2 - 1).$$

In this notation, we have

$$\sqrt{\frac{n}{2p}} \sum_{j=1}^p \ell_j = \sqrt{\frac{2}{n}} \sum_{i=1}^{n/2} w_i.$$

To determine the variance of w_1 , a direct calculation gives

$$\text{var}(w_1) = \frac{1}{p} \sum_{j,k=1}^p \left(\frac{1}{9} \text{cov}(x_{1j}^4, x_{1k}^4) + 4 \text{cov}(x_{1j}^2, x_{1k}^2) - \frac{2}{3} (\text{cov}(x_{1k}^4, x_{1j}^2) + \text{cov}(x_{1k}^2, x_{1j}^4)) \right).$$

Each of the covariances appearing on the right is calculated explicitly in Lemma D.6, which leads to

$$(A.28) \quad \text{var}(w_1) = \frac{1}{p} \left(\frac{8r_4}{3} \|R\|_4^4 + 8(r_4 + r_2 - 2r_3) \|R\|_2^2 \right) + p(r_4 - 4r_3 - r_2^2 + 8r_2 - 4).$$

Furthermore, Lemma B.7 shows that $r_k = 1 + o(p^{-3/4})$ for $k = 1, \dots, 4$, as well as

$$p(r_4 - 4r_3 - r_2^2 + 8r_2 - 4) = o(1).$$

So, combining with the general fact that $\|R\|_2^2 \lesssim p^{1/2} \|R\|_4^4$, and recalling that $\sigma_{n,1}^2 = \frac{8}{3p} \|R\|_4^4$, we have

$$(A.29) \quad \frac{\text{var}(w_1)}{\sigma_{n,1}^2} = 1 + o(1).$$

To complete the proof, we apply the Lindeberg CLT for triangular arrays given in (van der Vaart, 2000, Prop. 2.27). To verify the Lindeberg condition, it is sufficient to show that the following Lyapunov-type condition holds,

$$(A.30) \quad \mathbb{E}(w_1^4) = o(n\sigma_{n,1}^4).$$

Verifying this condition is by far the most challenging aspect of the proof and depends on delicate cancellations of high-order terms in an expansion of the fourth moment of w_1 . To proceed, define the random variable

$$(A.31) \quad y_j = \frac{1}{3} x_{1j}^4 - 2x_{1j}^2$$

for each $j = 1, \dots, p$. In this notation, we have

$$\begin{aligned}
p^2 \mathbb{E}(w_1^4) &= \left\| \sum_{j=1}^p (y_j + 2 - r_2) \right\|_{L^4}^4 \\
&= \sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(y_{j_1} y_{j_2} y_{j_3} y_{j_4}) + 4(2 - r_2)p \sum_{j_1, j_2, j_3=1}^p \mathbb{E}(y_{j_1} y_{j_2} y_{j_3}) \\
&\quad + 6(2 - r_2)^2 p^2 \sum_{j_1, j_2=1}^p \mathbb{E}(y_{j_1} y_{j_2}) - 3(2 - r_2)^4 p^4.
\end{aligned}$$

The last term on the right side can be simplified with respect to its dependence on r_2 by using the expansion in (A.5), which yields

$$-3(2 - r_2)^4 p^4 = (4r_2 - 7)p^4 + 8(\tau - 2)p^3 + o(p^3).$$

We will use Isserlis' theorem (Janson, 1997, Theorem 1.28) to evaluate the three sums in the previous formula for $p^2 \mathbb{E}(w_1^4)$, which is a substantial undertaking that we defer to Proposition A.2.3. This proposition establishes the following three equations

$$\begin{aligned}
\sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(y_{j_1} y_{j_2} y_{j_3} y_{j_4}) &= p^4(r_8 - 8r_7 + 24r_6 - 32r_5 + 16r_4) + 16p^2 \|R\|_4^4 + o(p \|R\|_4^8), \\
4(2 - r_2)p \sum_{j_1, j_2, j_3=1}^p \mathbb{E}(y_{j_1} y_{j_2} y_{j_3}) &= p^4(4r_6 - 24r_5 + 48r_4 - 32r_3) + 4(\tau - 2)p^3 - 32p^2 \|R\|_4^4 + o(p \|R\|_4^8), \\
6(2 - r_2)^2 p^2 \sum_{j_1, j_2=1}^p \mathbb{E}(y_{j_1} y_{j_2}) &= p^4(6r_4 - 24r_3 + 24r_2) - 12(\tau - 2)p^3 + 16p^2 \|R\|_4^4 + o(p \|R\|_4^8).
\end{aligned}$$

Based on the relations among $r_1, \dots, r_8$ established in Lemma B.7, the last several equations can be combined to obtain

$$p^2 \mathbb{E}(w_1^4) = o(p \|R\|_4^8),$$

which implies the desired condition in equation (A.30), since $\sigma_{n,1}^2 = \frac{8}{3p} \|R\|_4^4$. $\square$

LEMMA B.5. *If Assumption 1 holds and the diagonal entries of Σ are 1, then*

$$\frac{1}{\sigma_{n,1}} \sum_{j=1}^p \left(\Delta_j - \left[\frac{\kappa}{3} - 1 - \frac{4}{n} \right] \right) \xrightarrow{\mathbb{P}} 0.$$

PROOF. First, define the random variable

$$m_n = \sum_{j=1}^p \left(\frac{2}{n} \sum_{i=1}^{n/2} (x_{ij}^2 - 1) \right)^2,$$

and observe that

$$\sum_{j=1}^p \left(\Delta_j - \left[\frac{\kappa}{3} - 1 - \frac{4}{n} \right] \right) = \frac{4p}{n} - m_n.$$

Since $\sigma_{n,1}^2 \gtrsim 1$, it is enough to show that $\mathbb{E}(m_n) = 4p/n + o(1)$ and $\text{var}(m_n) = o(\sigma_{n,1}^2)$. Noting that $\sigma_{n,1}^2 = (8/3)\|R\|_4^4/p$, the condition $\text{var}(m_n) = o(\sigma_{n,1}^2)$ follows directly from (A.23). To calculate the expectation of m_n , note that the relations $\mathbb{E}(x_{1j}^2) = 1$ and $\text{var}(x_{1j}^2) = 3r_2 - 1$ hold for all $j = 1, \dots, p$ when the diagonal entries of Σ are 1, and equation (A.5) implies $3r_2 - 1 = 2 + o(1)$. Therefore, we have

$$\begin{aligned} \mathbb{E}(m_n) &= \frac{2}{n} \sum_{j=1}^p \text{var}(x_{1j}^2) \\ &= \frac{4p}{n} + o(1), \end{aligned}$$

which completes the proof. $\square$

PROPOSITION A.2.3. *Suppose that Assumption 1 holds and the diagonal entries of Σ are equal to 1. Let $y_j = \frac{1}{3}x_{1j}^4 - 2x_{1j}^2$ for $j = 1, \dots, p$, then the following equations hold,*

$$(A.32) \quad \sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(y_{j_1} y_{j_2} y_{j_3} y_{j_4}) = p^4(r_8 - 8r_7 + 24r_6 - 32r_5 + 16r_4) + 16p^2\|R\|_4^4 + o(p\|R\|_4^8),$$

$$(A.33) \quad (2 - r_2)p \sum_{j_1, j_2, j_3=1}^p \mathbb{E}(y_{j_1} y_{j_2} y_{j_3}) = p^4(r_6 - 6r_5 + 12r_4 - 8r_3) + (\tau - 2)p^3 - 8p^2\|R\|_4^4 + o(p\|R\|_4^8),$$

$$(A.34) \quad (2 - r_2)^2 p^2 \sum_{j_1, j_2=1}^p \mathbb{E}(y_{j_1} y_{j_2}) = p^4(r_4 - 4r_3 + 4r_2) - 2(\tau - 2)p^3 + \frac{8}{3}p^2\|R\|_4^4 + o(p\|R\|_4^8).$$

Proof of Proposition A.2.3, equation (A.32). The definition of y_j leads to the following algebraic relation

$$\begin{aligned}
\sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(y_{j_1} y_{j_2} y_{j_3} y_{j_4}) &= \frac{1}{3^4} \sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(x_{1j_1}^4 x_{1j_2}^4 x_{1j_3}^4 x_{1j_4}^4) \\
&\quad - \frac{4 \cdot 2}{3^3} \sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(x_{1j_1}^4 x_{1j_2}^4 x_{1j_3}^4 x_{1j_4}^2) \\
&\quad + \frac{6 \cdot 2^2}{3^2} \sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(x_{1j_1}^4 x_{1j_2}^4 x_{1j_3}^2 x_{1j_4}^2) \\
&\quad - \frac{4 \cdot 2^3}{3} \sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(x_{1j_1}^4 x_{1j_2}^2 x_{1j_3}^2 x_{1j_4}^2) \\
&\quad + 2^4 \sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(x_{1j_1}^2 x_{1j_2}^2 x_{1j_3}^2 x_{1j_4}^2).
\end{aligned} \tag{A.35}$$

The proof consists in developing a way to systematically calculate the sums on the right side of (A.35). When the diagonal entries of Σ are 1, the random vector $\zeta_1 = \Sigma^{1/2} \mathbf{z}_1$ defined in (A.1) is equivalent to $\zeta_1 = R^{1/2} \mathbf{z}_1$, where $\mathbf{z}_1$ is a standard normal vector independent of ξ_1 . Likewise, we may write $\mathbf{x}_1 = \frac{\xi_1}{\|\mathbf{z}_1\|_2} \zeta_1$, and so for any positive integers $k_1, \dots, k_4$ we have

$$\mathbb{E}(x_{1j_1}^{k_1} x_{1j_2}^{k_2} x_{1j_3}^{k_3} x_{1j_4}^{k_4}) = r_{\frac{k_1+k_2+k_3+k_4}{2}} \mathbb{E}(\zeta_{1j_1}^{k_1} \zeta_{1j_2}^{k_2} \zeta_{1j_3}^{k_3} \zeta_{1j_4}^{k_4}).$$

To compute the expectation on the right, we will apply Isserlis' theorem (Janson, 1997, Theorem 1.28), which states that the following relation holds for any centered d -dimensional normal vector $(g_1, \dots, g_d)$,

$$\mathbb{E}(g_1 \cdots g_d) = \sum_{\pi \in \mathcal{P}_d} \prod_{\{j, j'\} \in \pi} \mathbb{E}(g_j g_{j'}), \tag{A.36}$$

where $\mathcal{P}_d$ is the collection of all partitions π of $\{1, \dots, d\}$ into distinct pairs $\{j, j'\}$. (Such partitions are also called “pair partitions”.) More specifically, we will apply Isserlis' theorem in the case when $d = \sum_{l=1}^4 k_l$ and

$$(g_1, \dots, g_d) = (\underbrace{\zeta_{1j_1}, \dots, \zeta_{1j_1}}_{k_1}, \underbrace{\zeta_{1j_2}, \dots, \zeta_{1j_2}}_{k_2}, \underbrace{\zeta_{1j_3}, \dots, \zeta_{1j_3}}_{k_3}, \underbrace{\zeta_{1j_4}, \dots, \zeta_{1j_4}}_{k_4}). \tag{A.37}$$

Fortunately, in this context, if a factor $\mathbb{E}(g_j g_{j'})$ does not correspond to one of the diagonal entries of R , then it can only take one of six possible forms: R_{j_1, j_2} , R_{j_1, j_3} , R_{j_1, j_4} , R_{j_2, j_3} , R_{j_2, j_4} , R_{j_3, j_4} .

Therefore, Isserlis' theorem gives

$$(A.38) \quad \mathbb{E}(x_{1j_1}^{k_1} x_{1j_2}^{k_2} x_{1j_3}^{k_2} x_{1j_4}^{k_4}) = r_{\frac{k_1+k_2+k_3+k_4}{2}} \sum_{\mathbf{e} \in \mathcal{E}} c_{\mathbf{e}} R_{j_1, j_2}^{e_1} R_{j_1, j_3}^{e_2} R_{j_1, j_4}^{e_3} R_{j_2, j_3}^{e_4} R_{j_2, j_4}^{e_5} R_{j_3, j_4}^{e_6},$$

where $\mathcal{E}$ is the set of all possible tuples of exponents $\mathbf{e} = (e_1, \dots, e_6)$ for the six factors $R_{j_1, j_2}, \dots, R_{j_3, j_4}$ and $c_{\mathbf{e}}$ is the number of pair partitions of $\{1, \dots, k_1 + k_2 + k_3 + k_4\}$ corresponding to $\mathbf{e}$ in Isserlis' theorem.

To facilitate the summation over $\mathcal{E}$, we define an equivalence relation on its elements. We say that $\mathbf{e}$ is equivalent to another tuple $\mathbf{e}' = (e'_1, \dots, e'_6)$ if the product $R_{j_1, j_2}^{e_1} \dots R_{j_3, j_4}^{e_6}$ can be obtained from $R_{j_1, j_2}^{e'_1} \dots R_{j_3, j_4}^{e'_6}$ by permuting $j_1, \dots, j_4$. Hence, if $\mathbf{e}$ and $\mathbf{e}'$ are equivalent, then their corresponding terms in (A.38) will produce the same value when they are summed over $j_1, \dots, j_4$. Also, let $[\mathbf{e}]$ denote the equivalence class of $\mathbf{e}$, and define

$$C([\mathbf{e}]) = \sum_{\mathbf{e} \in [\mathbf{e}]} c_{\mathbf{e}}$$

as well as

$$s([\mathbf{e}]) = \sum_{j_1, j_2, j_3, j_4=1}^p R_{j_1, j_2}^{e_1} R_{j_1, j_3}^{e_2} R_{j_1, j_4}^{e_3} R_{j_2, j_3}^{e_4} R_{j_2, j_4}^{e_5} R_{j_3, j_4}^{e_6}.$$

Then, the following formula results from summing (A.38) over $j_1, \dots, j_4$,

$$\sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(x_{1j_1}^{k_1} x_{1j_2}^{k_2} x_{1j_3}^{k_2} x_{1j_4}^{k_4}) = r_{\frac{k_1+k_2+k_3+k_4}{2}} \sum_{[\mathbf{e}] \in [\mathcal{E}]} C([\mathbf{e}]) s([\mathbf{e}]),$$

where $[\mathcal{E}]$ denotes the collection of equivalence classes.

To choose a unique representative for each equivalence class, we use the largest element with respect to the lexicographical ordering on tuples. A complete system of representatives is displayed in the second column of Table A.1. (More precisely, since $\mathcal{E}$ depends on (k_1, k_2, k_3, k_4) , each possible choice of (k_1, k_2, k_3, k_4) induces a separate system of representatives. For a column of Table A.1 that is headed by a particular choice of (k_1, k_2, k_3, k_4) , the representatives correspond to the values of $\mathbf{e}$ associated with the non-empty entries in that column.)

For the equivalence class $[\mathbf{e}]$ represented by the tuple $\mathbf{e}$ in the i^{th} row of Table A.1, we define the shorthand

$$(A.39) \quad s_i = s([\mathbf{e}]).$$

In the case when $(k_1, k_2, k_3, k_4) = (4, 4, 4, 4)$, Lemma B.6 shows that the sums s_i for $i = 7, 8, 10, 11, 12, 13, 15, 17, 19, 20$, are of order $o(p\|R\|_4^8)$. As will be seen below, it follows that these sums play a negligible role in the calculation of (A.32). Hence, it is not necessary to calculate the associated coefficients $C([\mathbf{e}])$, and these are marked by $\mathbf{x}$ in Table A.1.

To illustrate how the other coefficients $C([\mathbf{e}])$ can be calculated, we give two detailed examples in the case when $(k_1, k_2, k_3, k_4) = (4, 4, 4, 4)$. First, the choice $\mathbf{e} = (0, 0, 0, 0, 0, 0)$ corresponds to terms in (A.38) that are equal to 1, which arise from partitioning the 16 indices in (A.37) as

$$\{j_1, j_1\}, \{j_1, j_1\}, \{j_2, j_2\}, \{j_2, j_2\}, \{j_3, j_3\}, \{j_3, j_3\}, \{j_4, j_4\}, \{j_4, j_4\}.$$

(Note that each pair $\{j_l, j_l\}$ corresponds to a factor $R_{j_l, j_l} = 1$.) The number of such partitions is $c_{\mathbf{e}} = 3^4$, because there are $k_l = 4$ copies of each “type” j_l that must be partitioned into pairs, and any set of 4 elements has exactly 3 pair partitions. Furthermore, the equivalence class of $\mathbf{e} = (0, 0, 0, 0, 0, 0)$ is a singleton, and so $C([\mathbf{e}]) = c_{\mathbf{e}} = 81$.

As a second example of how to calculate $C([\mathbf{e}])$, consider the choice $\mathbf{e} = (1, 1, 0, 0, 1, 1)$, whose equivalence class is equal to $[\mathbf{e}] = \{(1, 1, 0, 0, 1, 1), (1, 0, 1, 1, 0, 1), (0, 1, 1, 1, 1, 0)\}$. This choice of $\mathbf{e}$ corresponds to terms in (A.38) of the form $R_{j_1, j_2} R_{j_1, j_3} R_{j_2, j_4} R_{j_3, j_4}$, which arise from partitioning the 16 indices in (A.37) as

$$\{j_1, j_1\}, \{j_2, j_2\}, \{j_3, j_3\}, \{j_4, j_4\}, \{j_1, j_2\}, \{j_1, j_3\}, \{j_2, j_4\}, \{j_3, j_4\}.$$

(Note that the first four pairs above lead to unit factors, $R_{j_l, j_l} = 1$ for $l = 1, \dots, 4$, and so these pairs are “invisible” in the product $R_{j_1, j_2} R_{j_1, j_3} R_{j_2, j_4} R_{j_3, j_4}$.) The number of ways to form such a partition is $c_{\mathbf{e}} = \binom{4}{2} \binom{4}{2} \binom{4}{2} \binom{4}{2} 2^4$, and the equivalence class $[\mathbf{e}]$ has 3 elements, so we conclude that $C([\mathbf{e}]) = 3c_{\mathbf{e}} = 62208$. Following similar reasoning, we can calculate the coefficients $C([\mathbf{e}])$ associated with the other choices of $\mathbf{e}$, as given in Table A.1.

Using the coefficients $C([\mathbf{e}])$ from the table yields the equations below. (Note that the first equation has coefficients from the column headed by $(4, 4, 4, 4)$, the second equation has coefficients

e		(k_1, k_2, k_3, k_4)				
		(4, 4, 4, 4)	(4, 4, 4, 2)	(4, 4, 2, 2)	(4, 2, 2, 2)	(2, 2, 2, 2)
1	(0, 0, 0, 0, 0, 0)	81	27	9	3	1
2	(1, 1, 0, 0, 1, 1)	62208	10368	1728	288	48
3	(1, 1, 0, 1, 0, 0)	20736	4320	864	168	32
4	(2, 0, 0, 0, 0, 0)	3888	972	234	54	12
5	(2, 0, 0, 0, 0, 2)	15552	2592	432	72	12
6	(2, 1, 1, 0, 0, 1)	248832	31104	3456	288	
7	(2, 1, 1, 1, 1, 0)	✖	✖	✖		
8	(2, 1, 1, 1, 1, 2)	✖				
9	(2, 2, 0, 0, 0, 0)	31104	5184	720	72	
10	(2, 2, 0, 0, 2, 0)	✖	✖	✖		
11	(2, 2, 0, 0, 2, 2)	✖				
12	(2, 2, 0, 1, 1, 1)	✖	✖			
13	(2, 2, 0, 2, 0, 0)	✖	✖			
14	(3, 1, 0, 0, 1, 1)	165888	13824	768		
15	(3, 1, 0, 0, 1, 3)	✖				
16	(3, 1, 0, 1, 0, 0)	41472	5184	384		
17	(3, 1, 0, 1, 0, 2)	✖	✖			
18	(4, 0, 0, 0, 0, 0)	1296	216	24		
19	(4, 0, 0, 0, 0, 2)	✖	✖	✖		
20	(4, 0, 0, 0, 0, 4)	✖				

TABLE A.1. Coefficient values $C([\mathbf{e}])$. For a column headed by a particular choice of (k_1, k_2, k_3, k_4) , the non-empty entries in that column represent the values of $C([\mathbf{e}])$. If an entry in, say, the i^{th} row is marked with **x**, then the associated sum s_i defined in (A.39) is of negligible size, and so the value of the coefficient is not needed.

from the column headed by (4, 4, 4, 2), and so on.)

$$\begin{aligned}
\frac{1}{3^4} \sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(x_{1j_1}^4 x_{1j_2}^4 x_{1j_3}^4 x_{1j_4}^4) &= \frac{r_8}{3^4} (1296 s_{18} + 41472 s_{16} + 165888 s_{14} + 31104 s_9 + 248832 s_6 \\
&\quad + 15552 s_5 + 3888 s_4 + 20736 s_3 + 62208 s_2 + 81 s_1) + o(p \| R \|_4^8) \\
\frac{4 \cdot 2}{3^3} \sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(x_{1j_1}^4 x_{1j_2}^4 x_{1j_3}^4 x_{1j_4}^2) &= \frac{4 \cdot 2}{3^3} r_7 (216 s_{18} + 5184 s_{16} + 13824 s_{14} + 5184 s_9 + 31104 s_6 \\
&\quad + 2592 s_5 + 972 s_4 + 4320 s_3 + 10368 s_2 + 27 s_1) + o(p \| R \|_4^8) \\
\frac{6 \cdot 2^2}{3^2} \sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(x_{1j_1}^4 x_{1j_2}^4 x_{1j_3}^2 x_{1j_4}^2) &= \frac{6 \cdot 2^2}{3^2} r_6 (24 s_{18} + 384 s_{16} + 768 s_{14} + 720 s_9 + 3456 s_6 + 432 s_5 \\
&\quad + 234 s_4 + 864 s_3 + 1728 s_2 + 9 s_1) + o(p \| R \|_4^8) \\
\frac{4 \cdot 2^3}{3} \sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(x_{1j_1}^4 x_{1j_2}^2 x_{1j_3}^2 x_{1j_4}^2) &= \frac{4 \cdot 2^3}{3} r_5 (72 s_9 + 288 s_6 + 72 s_5 + 54 s_4 + 168 s_3 + 288 s_2 + 3 s_1) \\
&\quad + o(p \| R \|_4^8) \\
2^4 \sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(x_{1j_1}^2 x_{1j_2}^2 x_{1j_3}^2 x_{1j_4}^2) &= 2^4 r_4 (12 s_5 + 12 s_4 + 32 s_3 + 48 s_2 + s_1) + o(p \| R \|_4^8).
\end{aligned}$$

Applying these results to (A.35), it follows from algebraic simplification that

$$\begin{aligned}
\sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(y_{j_1} y_{j_2} y_{j_3} y_{j_4}) &= 16(r_8 - 4r_7 + 4r_6) s_{18} + 512(r_8 - 3r_7 + 2r_6) s_{16} + 2048(r_8 - 2r_7 + r_6) s_{14} \\
&+ 384(r_8 - 4r_7 + 5r_6 - 2r_5) s_9 + 3072(r_8 - 3r_7 + 3r_6 - r_5) s_6 \\
&+ 192(r_8 - 4r_7 + 6r_6 - 4r_5 + r_4) s_5 + 48(r_8 - 6r_7 + 13r_6 - 12r_5 + 4r_4) s_4 \\
&+ 256(r_8 - 5r_7 + 9r_6 - 7r_5 + 2r_4) s_3 + 768(r_8 - 4r_7 + 6r_6 - 4r_5 + r_4) s_2 \\
&+ (r_8 - 8r_7 + 24r_6 - 32r_5 + 16r_4) s_1 + o(p\|R\|_4^8).
\end{aligned}$$

Lemma B.7 gives $r_k = 1 + o(p^{-3/4})$ for $k = 1, \dots, 8$ and Lemma B.6 shows that all of the s_i appearing in the equation displayed above, except s_1 and s_{18} , are $\mathcal{O}(p^{13/8}\|R\|_4^8)$. Also, straightforward calculations show that $s_1 = p^4$ and $s_{18} = p^2\|R\|_4^4$. Accounting for these facts in the equation displayed above leads to (A.32), as needed. $\square$

Proof of Proposition A.2.3, equation (A.33). We apply the approach developed above. In particular, we start by observing that

$$\begin{aligned}
p \sum_{j_1, j_2, j_3=1}^p \mathbb{E}(y_{j_1} y_{j_2} y_{j_3}) &= \frac{1}{3^3} \sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(x_{1j_1}^4 x_{1j_2}^4 x_{1j_3}^4) \\
&- \frac{3 \cdot 2}{3^2} \sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(x_{1j_1}^4 x_{1j_2}^4 x_{1j_3}^2) \\
&+ \frac{3 \cdot 2^2}{3} \sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(x_{1j_1}^4 x_{1j_2}^2 x_{1j_3}^2) \\
&- 2^3 \sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(x_{1j_1}^2 x_{1j_2}^2 x_{1j_3}^2).
\end{aligned} \tag{A.40}$$

We need to analyze $\mathbb{E}(x_{1j_1}^{k_1} x_{1j_2}^{k_2} x_{1j_3}^{k_3} x_{1j_4}^{k_4})$ when the integers $k_1 \geq k_2 \geq k_3$ reside in $\{2, 4\}$ and $k_4 = 0$. In this case, the calculation for (A.33) may be completed using the information provided in Table A.2, just as Table A.1 was used to complete the calculation for (A.32).

e	(k_1, k_2, k_3, k_4)			
	$(4, 4, 4, 0)$	$(4, 4, 2, 0)$	$(4, 2, 2, 0)$	$(2, 2, 2, 0)$
1 $(0, 0, 0, 0, 0, 0)$	27	9	3	1
3 $(1, 1, 0, 1, 0, 0)$	1728	288	48	8
4 $(2, 0, 0, 0, 0, 0)$	648	144	30	6
9 $(2, 2, 0, 0, 0, 0)$	2592	288	24	
13 $(2, 2, 0, 2, 0, 0)$	x			
16 $(3, 1, 0, 1, 0, 0)$	3456	192		
18 $(4, 0, 0, 0, 0, 0)$	216	24		

TABLE A.2. Coefficient values $C([\mathbf{e}])$. The table is organized analogously to Table A.1.

Using the coefficients $C([\mathbf{e}])$ from the table then yields

$$\begin{aligned}
\frac{1}{3^3} \sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(x_{1j_1}^4 x_{1j_2}^4 x_{1j_3}^4) &= \frac{r_6}{3^3} (216s_{18} + 3456s_{16} + 2592s_9 + 648s_4 + 1728s_3 \\
&\quad + 27s_1) + o(p\|R\|_4^8) \\
\frac{3 \cdot 2}{3^2} \sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(x_{1j_1}^4 x_{1j_2}^4 x_{1j_3}^2) &= \frac{3 \cdot 2}{3^2} r_5 (24s_{18} + 192s_{16} + 288s_9 + 144s_4 + 288s_3 \\
&\quad + 9s_1) + o(p\|R\|_4^8) \\
\frac{3 \cdot 2^2}{3} \sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(x_{1j_1}^4 x_{1j_2}^2 x_{1j_3}^2) &= \frac{3 \cdot 2^2}{3} r_4 (24s_9 + 30s_4 + 48s_3 + 3s_1) + o(p\|R\|_4^8) \\
2^3 \sum_{j_1, j_2, j_3, j_4=1}^p \mathbb{E}(x_{1j_1}^2 x_{1j_2}^2 x_{1j_3}^2) &= 2^3 r_3 (6s_4 + 8s_3 + s_1) + o(p\|R\|_4^8).
\end{aligned}$$

Applying these equations to (A.40) yields

$$\begin{aligned}
p \sum_{j_1, j_2, j_3=1}^p \mathbb{E}(y_{j_1} y_{j_2} y_{j_3}) &= 8(r_6 - 2r_5)s_{18} + 128(r_6 - r_5)s_{16} + 96(r_6 - 2r_5 + r_4)s_9 \\
&\quad + 24(r_6 - 4r_5 + 5r_4 - 2r_3)s_4 + 64(r_6 - 3r_5 + 3r_4 - r_3)s_3 \\
&\quad + (r_6 - 6r_5 + 12r_4 - 8r_3)s_1 + o(p\|R\|_4^8) \\
&= -8p^2\|R\|_4^4 + (r_6 - 6r_5 + 12r_4 - 8r_3)p^4 + o(p\|R\|_4^8),
\end{aligned}$$

where the last step is based on $s_1 = p^4$, $s_{18} = p^2\|R\|_4^4$, and the formulas for $r_3, \dots, r_6$ and $s_3, \dots, s_{16}$ given respectively in Lemmas B.7 and B.6. Next, we multiply both sides of the last equation by

$2 - r_2$. Noting the relation in (A.5) gives $2 - r_2 = 1 - \frac{\tau-2}{p} + o(\frac{1}{p})$, we have

$$(2 - r_2)p \sum_{j_1, j_2, j_3=1}^p \mathbb{E}(y_{j_1} y_{j_2} y_{j_3}) = -8p^2 \|R\|_4^4 + (r_6 - 6r_5 + 12r_4 - 8r_3)(p^4 - (\tau - 2)p^3 + o(p^3)) + o(p\|R\|_4^8).$$

Finally, the middle term on the right side can be handled by Lemma B.7, which ensures $r_6 - 6r_5 + 12r_4 - 8r_3 = -1 + o(p^{-3/4})$ and leads to (A.33). $\square$

Proof of Proposition A.2.3, equation (A.34). The definition of y_j gives

$$\sum_{j_1, j_2=1}^p \mathbb{E}(y_{j_1} y_{j_2}) = \frac{1}{3^2} \sum_{j_1, j_2=1}^p \mathbb{E}(x_{1j_1}^4 x_{1j_2}^4) - \frac{2 \cdot 2}{3} \sum_{j_1, j_2=1}^p \mathbb{E}(x_{1j_1}^4 x_{1j_2}^2) + 2^2 \sum_{j_1, j_2=1}^p \mathbb{E}(x_{1j_1}^2 x_{1j_2}^2).$$

The proof of Lemma D.6 shows that when the diagonal entries of Σ are equal to 1, the following equations hold

$$\begin{aligned} \frac{1}{3^2} \sum_{j_1, j_2=1}^p \mathbb{E}(x_{1j_1}^4 x_{1j_2}^4) &= r_4 \left(p^2 + 8\|R\|_2^2 + \frac{8}{3}\|R\|_4^4 \right) \\ \frac{2 \cdot 2}{3} \sum_{j_1, j_2=1}^p \mathbb{E}(x_{1j_1}^4 x_{1j_2}^2) &= r_3 \left(4p^2 + 16\|R\|_2^2 \right) \\ 2^2 \sum_{j_1, j_2=1}^p \mathbb{E}(x_{1j_1}^2 x_{1j_2}^2) &= r_2 \left(4p^2 + 8\|R\|_2^2 \right). \end{aligned}$$

Combining these calculations, we have

$$\begin{aligned} p^2 \sum_{j_1, j_2=1}^p \mathbb{E}(y_{j_1} y_{j_2}) &= (r_4 - 4r_3 + 4r_2)p^4 + 8(r_4 - 2r_3 + r_2)p^2\|R\|_2^2 + \frac{8r_4}{3}p^2\|R\|_4^4 \\ &= (r_4 - 4r_3 + 4r_2)p^4 + \frac{8}{3}p^2\|R\|_4^4 + o(p\|R\|_4^8), \end{aligned}$$

where the last step uses Lemma B.7 and the fact that $\|R\|_2^2 \lesssim p^{1/2}\|R\|_4^4$. Combining the relation $r_4 - 4r_3 + 4r_2 = 1 + o(p^{-3/4})$ given by Lemma B.7 with $(2 - r_2)^2 = 1 - \frac{2(\tau-2)}{p} + o(\frac{1}{p})$ implied by Assumption 1, we have

$$\begin{aligned} (2 - r_2)^2 p^2 \sum_{j_1, j_2=1}^p \mathbb{E}(y_{j_1} y_{j_2}) &= (r_4 - 4r_3 + 4r_2)(p^4 - 2(\tau - 2)p^3 + o(p^3)) + \frac{8}{3}p^2\|R\|_4^4 + o(p\|R\|_4^8) \\ &= p^4(r_4 - 4r_3 + 4r_2) - 2(\tau - 2)p^3 + \frac{8}{3}p^2\|R\|_4^4 + o(p\|R\|_4^8), \end{aligned}$$

which proves (A.34). $\square$

LEMMA B.6. *The sums s_i defined in (A.39) satisfy*

$$(A.41a) \quad s_i = \begin{cases} o(p\|R\|_4^8) & \text{for } i = 7, 8, 10, 11, 12, 13, 15, 17, 19, 20 \\ \mathcal{O}(p^{13/8}\|R\|_4^8) & \text{for } i = 2, 3, 4, 5, 6, 9, 14, 16. \end{cases}$$

PROOF. We begin by analyzing the sums s_i in the first case (A.41a). For these calculations, we will often use the fact that $\frac{\|R\|_2}{\|R\|_4^2} = \mathcal{O}(p^{1/4})$, which holds for a generic correlation matrix R .

- $i = 19, 20$: The sums s_i are bounded by

$$\sum_{j_1, j_2, j_3, j_4=1}^p R_{j_1, j_2}^4 R_{j_3, j_4}^2 = \|R\|_4^4 \|R\|_2^2 = o(p\|R\|_4^8).$$

- $i = 17$: Due to the basic inequality $|ab| \leq \frac{1}{2}(a^2 + b^2)$, the absolute value $|s_i|$ is bounded by

$$(A.42) \quad \begin{aligned} \sum_{j_1, j_2, j_3, j_4=1}^p \left| R_{j_1, j_2}^3 R_{j_1, j_3} R_{j_2, j_3} R_{j_3, j_4}^2 \right| &\leq \sum_{j_1, j_2, j_3, j_4=1}^p \frac{1}{2} R_{j_1, j_2}^2 R_{j_3, j_4}^2 (R_{j_1, j_3}^2 + R_{j_2, j_3}^2) \\ &= \mathbf{1}^\top R^{\circ 2} R^{\circ 2} R^{\circ 2} \mathbf{1} \\ &\leq p \text{tr}(R^{\circ 2} R^{\circ 2} R^{\circ 2}) \\ &\leq p \text{tr}(R^{\circ 2} R^{\circ 2})^{3/2} \\ &= p(\|R\|_4^4)^{3/2} \\ &= o(p\|R\|_4^8), \end{aligned}$$

where $\mathbf{1}$ denotes the p -dimensional vector of ones, and the third inequality is based on the fact that the Schur product theorem guarantees that $R^{\circ k}$ is positive semidefinite for any integer $k \geq 1$.

- $i = 15$: The absolute value $|s_i|$ is equal to

$$\begin{aligned}
\left| \sum_{j_1, j_2, j_3, j_4=1}^p R_{j_1, j_2}^3 R_{j_1, j_3} R_{j_2, j_4} R_{j_3, j_4}^3 \right| &= |\text{tr}(R^{\circ 3} R R^{\circ 3} R)| \\
&\leq \|R^{\circ 3} R\|_2^2 \\
&\leq \|R\|_2^2 \|R^{\circ 3}\|_2^2 \\
&\leq \|R\|_2^2 \|R\|_4^4 \\
&= o(p\|R\|_4^8),
\end{aligned}$$

where the step going from the third to the fourth line uses the fact that $R_{jk}^6 \leq R_{jk}^4$ for all j, k .

- $i = 13$: The sum s_i is equal to

$$\sum_{j_1, j_2, j_3, j_4=1}^p R_{j_1, j_2}^2 R_{j_1, j_3}^2 R_{j_2, j_3}^2 = p \text{tr}(R^{\circ 2} R^{\circ 2} R^{\circ 2}) = o(p\|R\|_4^8),$$

where the last step uses the calculation in (A.42).

- $i = 12$: The absolute value $|s_i|$ is equal to

$$\begin{aligned}
\left| \sum_{j_1, j_2, j_3, j_4=1}^p R_{j_1, j_2}^2 R_{j_1, j_3}^2 R_{j_2, j_3} R_{j_2, j_4} R_{j_3, j_4} \right| &= |\text{tr}(R^{\circ 2} (R \circ R^2) R^{\circ 2})| \\
&\leq \text{tr}(R^{\circ 2} R^{\circ 2}) \text{tr}(R \circ R^2) \\
&\leq \|R\|_4^4 \|R\|_2^2 \\
&= o(p\|R\|_4^8),
\end{aligned}$$

where the second inequality is based on the fact that the Schur product theorem guarantees that $R \circ R^2$ is positive semidefinite, and the second-to-last step uses $\text{tr}(R \circ R^2) = \text{tr}(R^2)$, which holds since R is a correlation matrix.

- $i = 11$: The sum s_i is equal to

$$\begin{aligned}
\sum_{j_1, j_2, j_3, j_4=1}^p R_{j_1, j_2}^2 R_{j_1, j_3}^2 R_{j_2, j_4}^2 R_{j_3, j_4}^2 &= \text{tr}(R^{\circ 2} R^{\circ 2} R^{\circ 2} R^{\circ 2}) \\
&\leq \text{tr}(R^{\circ 2} R^{\circ 2})^2 \\
&= \|R\|_4^8 \\
&= o(p\|R\|_4^8).
\end{aligned}$$

- $i = 10$: The calculation in (A.42) implies that the sum s_i is equal to

$$\sum_{j_1, j_2, j_3, j_4=1}^p R_{j_1, j_2}^2 R_{j_1, j_3}^2 R_{j_2, j_4}^2 = \mathbf{1}^\top R^{\circ 2} R^{\circ 2} R^{\circ 2} \mathbf{1} = o(p\|R\|_4^8).$$

- $i = 7, 8$: Using the basic inequality $|ab| \leq \frac{1}{2}(a^2 + b^2)$ twice, the absolute values $|s_i|$ are bounded by

$$\begin{aligned}
\sum_{j_1, j_2, j_3, j_4=1}^p \left| R_{j_1, j_2}^2 R_{j_1, j_3} R_{j_1, j_4} R_{j_2, j_3} R_{j_2, j_4} \right| &\leq \sum_{j_1, j_2, j_3, j_4=1}^p \frac{1}{4} R_{j_1, j_2}^2 (R_{j_1, j_3}^2 + R_{j_1, j_4}^2) (R_{j_2, j_3}^2 + R_{j_2, j_4}^2) \\
&\lesssim \sum_{j_1, j_2, j_3, j_4=1}^p R_{j_1, j_2}^2 R_{j_1, j_3}^2 R_{j_2, j_3}^2 \\
&\quad + \sum_{j_1, j_2, j_3, j_4=1}^p R_{j_1, j_2}^2 R_{j_1, j_3}^2 R_{j_2, j_4}^2 \\
&= o(p\|R\|_4^8),
\end{aligned}$$

where the last step uses the previous calculations for s_{10} and s_{13} .

In the remainder of the proof, we analyze the sums s_i in the second case (A.41b).

- $i = 2$: The sum s_i is equal to

$$\begin{aligned}
\sum_{j_1, j_2, j_3, j_4=1}^p R_{j_1, j_2} R_{j_1, j_3} R_{j_2, j_4} R_{j_3, j_4} &= \text{tr}(R^4) \\
&\leq \|R\|_2^4 \\
&= \mathcal{O}(p\|R\|_4^8).
\end{aligned}$$

- $i = 3$: The sum s_i is equal to

$$\begin{aligned}
\sum_{j_1, j_2, j_3, j_4=1}^p R_{j_1, j_2} R_{j_1, j_3} R_{j_2, j_3} &= \text{ptr}(R^3) \\
&\leq p \|R\|_2^3 \\
&= \mathcal{O}(p^{7/4} \|R\|_4^6) \\
&= \mathcal{O}(p^{5/4} \|R\|_4^8),
\end{aligned}$$

where the last step uses the inequality $\|R\|_4^2 \geq p^{1/2}$ that holds for a generic correlation matrix R .

- $i = 4, 5, 6, 9, 14, 16$: The absolute value $|s_i|$ is bounded by

$$\sum_{j_1, j_2, j_3, j_4=1}^p R_{j_1, j_2}^2 = p^2 \|R\|_2^2 = \mathcal{O}(p^{3/2} \|R\|_4^8).$$

□

LEMMA B.7. *If Assumption 1 holds, then for $k = 1, \dots, 8$,*

$$r_k = 1 + o(p^{-3/4}).$$

In addition, the following relations hold

$$(A.43) \quad r_4 - 4r_3 - r_2^2 + 8r_2 - 4 = \sum_{l=0}^4 (-1)^l \binom{4}{l} r_{l-1} + o\left(\frac{1}{p}\right)$$

$$(A.44) \quad \sum_{l=0}^4 (-1)^l \binom{4}{l} r_l = o\left(\frac{1}{p}\right)$$

$$(A.45) \quad \sum_{l=0}^8 (-1)^l \binom{8}{l} r_l = o\left(\frac{1}{p}\right).$$

PROOF. Assumption 1 implies $\mathbb{E}\left(\left(\frac{\xi_1^2}{p} - 1\right)^k\right) = o(p^{-3/4})$ for $k = 1, \dots, 8$, and so

$$\begin{aligned}
\frac{\mathbb{E}(\xi_1^{2k})}{p^k} &= \mathbb{E}\left(\left(\frac{\xi_1^2}{p} - 1 + 1\right)^k\right) \\
&= 1 + o(p^{-3/4}).
\end{aligned}$$

Also, the definition of r_k in (A.4) implies

$$\begin{aligned}
r_k &= \frac{\mathbb{E}(\xi_1^{2k})}{p^k} - \frac{\mathbb{E}(\xi_1^{2k}) \left(k(k-1)p^{k-2} + \mathcal{O}(p^{k-3}) \right)}{p^{k-1} \mathbb{E}(\|\mathbf{z}_1\|_2^{2k})} \\
&= \frac{\mathbb{E}(\xi_1^{2k})}{p^k} - \frac{k(k-1)}{p} + o\left(\frac{1}{p}\right) \\
&= 1 + o(p^{-3/4}).
\end{aligned}
\tag{A.46}$$

Towards proving (A.43), note that the relation $r_2 = 1 + \frac{\tau-2}{p} + o(\frac{1}{p})$ in (A.5) implies

$$r_2^2 = 2r_2 - 1 + o\left(\frac{1}{p}\right).$$

Combining $r_0 = r_1 = 1$ and the condition $\mathbb{E}\left(\left|\frac{\xi_1^2}{p} - 1\right|^4\right) = o(\frac{1}{p})$ implied by Assumption 1 with the second equality in (A.46), we have

$$\begin{aligned}
r_4 - 4r_3 - r_2^2 + 8r_2 - 4 &= r_4 - 4r_3 + 6r_2 - 4r_1 + r_0 + o\left(\frac{1}{p}\right) \\
&= \sum_{l=0}^4 (-1)^l \binom{4}{l} \frac{\mathbb{E}(\xi_1^{2l})}{p^l} + o\left(\frac{1}{p}\right) \\
&= \mathbb{E}\left(\left|\frac{\xi_1^2}{p} - 1\right|^4\right) + o\left(\frac{1}{p}\right) \\
&= o\left(\frac{1}{p}\right),
\end{aligned}$$

which proves (A.43) and (A.44). Lastly, to show (A.45), observe that the second equality in (A.46) implies

$$\begin{aligned}
\sum_{l=0}^8 (-1)^l \binom{8}{l} r_l &= \sum_{l=0}^8 (-1)^l \binom{8}{l} \frac{\mathbb{E}(\xi_1^{2l})}{p^l} + o\left(\frac{1}{p}\right) \\
&= \mathbb{E}\left(\left|\frac{\xi_1^2}{p} - 1\right|^8\right) + o\left(\frac{1}{p}\right) \\
&= o\left(\frac{1}{p}\right),
\end{aligned}$$

where the last step follows from Assumption 1. □

A.2.2. Asymptotic normality of $T_{n,2}$.

PROPOSITION A.2.4. *Suppose Assumption 1 holds. If $\mathbf{r}(\Sigma) \lesssim \sqrt{p}$, then as $n \rightarrow \infty$,*

$$\frac{T_{n,2}}{\sigma_{n,2}} \Rightarrow N(0, 1).$$

PROOF. Define the statistics

$$U = \zeta^2 - 2(\check{\nu}_2 - \frac{2}{n}\check{\nu}_1^2)$$

$$V = \check{\nu}_1^2 + 2(\check{\nu}_2 - \frac{2}{n}\check{\nu}_1^2),$$

which allow $T_{n,2}$ to be rewritten as

$$(A.47) \quad T_{n,2} = \sqrt{\frac{pn}{2}} \left(\frac{U}{V} - \frac{\zeta^2 - 2\nu_2}{\nu_1^2 + 2\nu_2} \right).$$

When $\mathbf{r}(\Sigma) \lesssim \sqrt{p}$, it will be shown in Lemma B.8 that the statistic U satisfies

$$(A.48) \quad \frac{\sqrt{np/2}(U - (\zeta^2 - 2\nu_2))}{\sigma_{n,2}(\nu_1^2 + 2\nu_2)} \Rightarrow N(0, 1)$$

as $n \rightarrow \infty$. Next, we establish a limit for the statistic V . The proofs of Lemmas A.1 and A.2 in (Wang and Lopes, 2023) show that if Assumption 1 holds, then as $n \rightarrow \infty$

$$(A.49) \quad \frac{\check{\nu}_2 - \frac{2}{n}\check{\nu}_1^2}{\nu_2} \xrightarrow{\mathbb{P}} 1 \quad \text{and} \quad \frac{\check{\nu}_1^2}{\nu_1^2} \xrightarrow{\mathbb{P}} 1,$$

and hence

$$(A.50) \quad \frac{V}{\nu_1^2 + 2\nu_2} \xrightarrow{\mathbb{P}} 1.$$

Now we will combine the previous limits for U and V to reach the stated result. Based on the representation of $T_{n,2}$ given in (A.47), the limit for V implies

$$\frac{T_{n,2}}{\sigma_{n,2}} = \sqrt{\frac{pn}{2}} \left(\frac{U}{\sigma_{n,2}(\nu_1^2 + 2\nu_2)(1 + o_{\mathbb{P}}(1))} - \frac{\zeta^2 - 2\nu_2}{\sigma_{n,2}(\nu_1^2 + 2\nu_2)} \right).$$

Due to the limit for U given in (A.48), the desired limit for $T_{n,2}$ will follow from Slutsky's lemma if we can show that

$$\frac{\sqrt{pn}U}{\sigma_{n,2}(\nu_1^2 + 2\nu_2)} = \mathcal{O}_{\mathbb{P}}(1).$$

Noting that the definition of $\sigma_{n,2}^2$ in (2.4) gives

$$(A.51) \quad \sqrt{16\nu_4 + 8\nu_2^2} = \frac{\sigma_{n,2}(\nu_1^2 + 2\nu_2)}{\sqrt{p}},$$

it follows from (A.48) that

$$\frac{\sqrt{pn}U}{\sigma_{n,2}(\nu_1^2 + 2\nu_2)} = \frac{\sqrt{n}(\zeta^2 - 2\nu_2)}{\sqrt{16\nu_4 + 8\nu_2^2}} + \mathcal{O}_{\mathbb{P}}(1).$$

So, to complete the proof, it is enough to show $\zeta^2 - 2\nu_2 = \mathcal{O}(\frac{\nu_2}{\sqrt{p}})$. To see this, observe that when $\mathbf{r}(\Sigma) \lesssim \sqrt{p}$, Lemma D.2 and equation (A.5) imply

$$\begin{aligned} \zeta^2 - 2\nu_2 &= \mathcal{O}\left(\frac{1}{p}\right)(\nu_1^2 + 2\nu_2) \\ &= \mathcal{O}\left(\frac{\nu_2}{\sqrt{p}}\right), \end{aligned}$$

as needed. □

LEMMA B.8. *Suppose Assumption 1 holds. If $\mathbf{r}(\Sigma) \lesssim \sqrt{p}$, then as $n \rightarrow \infty$,*

$$\frac{\sqrt{np/2}(U - (\zeta^2 - 2\nu_2))}{\sigma_{n,2}(\nu_1^2 + 2\nu_2)} \Rightarrow N(0, 1).$$

PROOF. To show that U has the structure of a U-statistic, define the kernel

$$(A.52) \quad h(\mathbf{x}_1, \mathbf{x}_2) = \left(\frac{1}{2} - \frac{2(n-2)}{n^2}\right) \left(\|\mathbf{x}_1\|_2^2 - \|\mathbf{x}_2\|_2^2\right)^2 - \frac{2(n-2)}{n} (\mathbf{x}_1^\top \mathbf{x}_2)^2,$$

which yields

$$U = \frac{1}{\binom{n/2}{2}} \sum_{n/2 < i < j \leq n} h(\mathbf{x}_i, \mathbf{x}_j).$$

Furthermore, by letting

$$(A.53) \quad \begin{aligned} \theta &= \mathbb{E}(h(\mathbf{x}_1, \mathbf{x}_2)) \\ h_1(\mathbf{x}_1) &= \mathbb{E}(h(\mathbf{x}_1, \mathbf{x}_2) | \mathbf{x}_1) - \theta, \end{aligned}$$

the Hájek projection of $U - \theta$ can be expressed as

$$\hat{U} = \frac{2}{n/2} \sum_{i=n/2+1}^n h_1(\mathbf{x}_i).$$

It is a classical fact (van der Vaart, 2000, Theorem 11.2) that if

$$(A.54) \quad \frac{\text{var}(\widehat{U})}{\text{var}(U)} \rightarrow 1$$

as $n \rightarrow \infty$, then

$$(A.55) \quad \frac{U - \theta}{\sqrt{\text{var}(U)}} = \frac{\widehat{U}}{\sqrt{\text{var}(\widehat{U})}} + o_{\mathbb{P}}(1).$$

This shows that if U and $\widehat{U}$ are suitably normalized, then the asymptotic normality of one implies that of the other. To check the condition (A.54), we first introduce the parameters

$$\psi_1 = \text{cov}(h(\mathbf{x}_1, \mathbf{x}_2), h(\mathbf{x}_1, \mathbf{x}_3))$$

$$\psi_2 = \text{var}(h(\mathbf{x}_1, \mathbf{x}_2)).$$

It is known from the theory of U-statistics (van der Vaart, 2000, Theorem 12.3) that the variances of $\widehat{U}$ and U have exact expressions in terms of ψ_1 and ψ_2 , namely

$$(A.56) \quad \text{var}(\widehat{U}) = \frac{4}{n/2} \psi_1$$

$$(A.57) \quad \text{var}(U) = \frac{2(n/2 - 2)}{\binom{n/2}{2}} \psi_1 + \frac{1}{\binom{n/2}{2}} \psi_2.$$

Therefore, the condition (A.54) will hold, if we can show

$$(A.58) \quad \psi_2 = o(n \psi_1).$$

The sizes of ψ_1 and ψ_2 can be compared with Lemmas B.10 and B.11, which show that under the condition $\mathbf{r}(\Sigma) \lesssim \sqrt{p}$ we have

$$(A.59) \quad \psi_2 \lesssim \text{tr}(\Sigma^2)^2$$

$$(A.60) \quad \psi_1 = 4\text{tr}(\Sigma^4) + 2\text{tr}(\Sigma^2)^2 + o(\text{tr}(\Sigma^2)^2).$$

From these relations, it is straightforward to check that (A.58) holds, and hence (A.55) holds as well.

Having shown (A.55), the proof will be complete if we can establish two more items, which are

$$(A.61) \quad \frac{\widehat{U}}{\sqrt{\text{var}(\widehat{U})}} \Rightarrow N(0, 1)$$

and

$$(A.62) \quad \frac{U - \theta}{\sqrt{\text{var}(U)}} = \frac{\sqrt{np/2}(U - (\varsigma^2 - 2\nu_2))}{\sigma_{n,2}(\nu_1^2 + 2\nu_2)} + o_{\mathbb{P}}(1).$$

Since $\widehat{U}$ is a sum of centered independent random variables, the limit (A.61) can be established by checking the following Lyapunov-type condition, which implies the Lindeberg central limit theorem in the setting of triangular arrays (van der Vaart, 2000, Proposition 2.27) (and so is applicable in the high-dimensional setting),

$$(A.63) \quad \frac{\mathbb{E}(h_1(\mathbf{x}_1)^4)}{n \text{var}(h_1(\mathbf{x}_1))^2} \rightarrow 0.$$

This will be established in Lemma B.9.

With regard to (A.62), it follows from equations (A.54), (A.56), and (A.60) that the standard deviation of U satisfies

$$\frac{\sqrt{\text{var}(U)}}{\sqrt{\frac{1}{n/2}(16\nu_4 + 8\nu_2^2)}} \rightarrow 1.$$

Consequently, equation (A.51) gives

$$(A.64) \quad \frac{\sqrt{\text{var}(U)}}{\sqrt{\frac{2}{np}\sigma_{n,2}(\nu_1^2 + 2\nu_2)}} \rightarrow 1.$$

Finally, to consider the mean parameter θ , it can be checked using Lemma D.2 as well as equation (A.5) that

$$(A.65) \quad \begin{aligned} \theta &= \left(1 - \frac{4(n-2)}{n^2}\right)\varsigma^2 - \frac{2(n-2)}{n}\nu_2 \\ &= \varsigma^2 - 2\nu_2 + \mathcal{O}\left(\frac{\nu_2}{n}\right). \end{aligned}$$

Combining this with (A.64) implies (A.62), which completes the proof. $\square$

LEMMA B.9. *Suppose Assumption 1 holds. If $\mathbf{r}(\Sigma) \lesssim \sqrt{p}$, then $h_1(\mathbf{x}_1)$ defined in (A.53) satisfies*

$$(A.66) \quad \mathbb{E}(h_1(\mathbf{x}_1)^4) = o(n \text{var}(h_1(\mathbf{x}_1))^2).$$

PROOF. From the definition of $h_1(\mathbf{x}_1)$, we have

$$h_1(\mathbf{x}_1) = \left(\frac{1}{2} - \frac{2(n-2)}{n^2} \right) \left((\mathbf{x}_1^\top \mathbf{x}_1 - \text{tr}(\Sigma))^2 + (1 - r_2) \text{tr}(\Sigma)^2 - 2r_2 \text{tr}(\Sigma^2) \right) - \frac{2(n-2)}{n} \left(\mathbf{x}_1^\top \Sigma \mathbf{x}_1 - \text{tr}(\Sigma^2) \right).$$

The variance of $h_1(\mathbf{x}_1)$ can be computed via equations (A.56) and (A.60) to obtain

$$\text{var}(h_1(\mathbf{x}_1)) = 4\text{tr}(\Sigma^4) + 2\text{tr}(\Sigma^2)^2 + o(\text{tr}(\Sigma^2)^2).$$

So, to complete the proof, it is sufficient to show that

$$\mathbb{E}(h_1(\mathbf{x}_1)^4) = o(n \text{tr}(\Sigma^2)^4).$$

Due to equation (A.5), we have $|1 - r_2| \text{tr}(\Sigma)^2 \lesssim \text{tr}(\Sigma^2)$, and so

$$(A.67) \quad \mathbb{E}(h_1(\mathbf{x}_1)^4) \lesssim \|\mathbf{x}_1^\top \mathbf{x}_1 - \text{tr}(\Sigma)\|_{L^8}^8 + \text{tr}(\Sigma^2)^4 + \|\mathbf{x}_1^\top \Sigma \mathbf{x}_1 - \text{tr}(\Sigma^2)\|_{L^4}^4.$$

Lemma D.4 implies

$$\begin{aligned} \|\mathbf{x}_1^\top \Sigma \mathbf{x}_1 - \text{tr}(\Sigma^2)\|_{L^4}^4 &\lesssim \text{tr}(\Sigma^4)^2 + o\left(\frac{1}{p}\right) \text{tr}(\Sigma^2)^4 \\ &= o(n \text{tr}(\Sigma^2)^4). \end{aligned}$$

In addition, when $\mathbf{r}(\Sigma) \lesssim \sqrt{p}$, Lemma D.4 gives

$$\|\mathbf{x}_1^\top \mathbf{x}_1 - \text{tr}(\Sigma)\|_{L^8}^8 = o(n \text{tr}(\Sigma^2)^4).$$

Applying the above results to (A.67) completes the proof. $\square$

PROPOSITION A.2.5. *Suppose Assumption 1 holds. If $\mathbf{r}(\Sigma)/\sqrt{p} \rightarrow \infty$ as $n \rightarrow \infty$, then*

$$\frac{T_{n,2}}{\sigma_n} \xrightarrow{\mathbb{P}} 0.$$

PROOF. The definition of $T_{n,2}$ in (A.47) gives

$$(A.68) \quad \frac{T_{n,2}}{\sigma_n} = \frac{\sqrt{np/2}U}{\sigma_n(\nu_1^2 + 2\nu_2)} \left(\frac{\nu_1^2 + 2\nu_2}{V} - 1 \right) + \frac{\sqrt{np/2}(U - (\varsigma^2 - 2\nu_2))}{\sigma_n(\nu_1^2 + 2\nu_2)}.$$

To address the second term, due to the fact $\sigma_n^2 \geq \sigma_{n,1}^2 \gtrsim 1$, we will show

$$(A.69) \quad \frac{\sqrt{np/2}}{\nu_1^2 + 2\nu_2} (U - (\varsigma^2 - 2\nu_2)) \xrightarrow{\mathbb{P}} 0$$

by showing that the mean and variance of the left side are both $o(1)$. When $\mathbf{r}(\Sigma)/\sqrt{p} \rightarrow \infty$, equation (A.65) gives

$$\begin{aligned} \frac{p}{\nu_1^2 + 2\nu_2} \mathbb{E}(U - (\varsigma^2 - 2\nu_2)) &= \mathcal{O}\left(\frac{1}{\mathbf{r}(\Sigma)}\right) \\ &= o(1). \end{aligned}$$

Regarding the variance of U , we will show $\text{var}(\frac{p}{\nu_1^2 + 2\nu_2}U) = o(1)$. Due to equation (A.57), it is sufficient to show $\psi_1 = o(\frac{\nu_1^4}{p})$ and $\psi_2 = o(\nu_1^4)$, and these relations will be established in Lemmas B.10 and B.11. Consequently, the limit in (A.69) holds.

Next, to consider the first term in (A.68), it is known from (A.50) that $\frac{V}{\nu_1^2 + 2\nu_2} \xrightarrow{\mathbb{P}} 1$. Also, it is known from Lemma D.2, (A.5) and (A.69) that

$$\frac{\sqrt{np/2}U}{\sigma_n(\nu_1^2 + 2\nu_2)} = \mathcal{O}_{\mathbb{P}}(1),$$

which implies that the first term in (A.68) is $o_{\mathbb{P}}(1)$ and completes the proof. $\square$

LEMMA B.10. *Suppose that Assumption 1 holds. If $\mathbf{r}(\Sigma)/\sqrt{p} \rightarrow \infty$ as $n \rightarrow \infty$, then*

$$\psi_2 = o(\text{tr}(\Sigma)^4).$$

Alternatively, if $\mathbf{r}(\Sigma) \lesssim \sqrt{p}$, then

$$\psi_2 \lesssim \text{tr}(\Sigma^2)^2.$$

PROOF. Using the definition of $h(\mathbf{x}_1, \mathbf{x}_2)$ in (A.52), as well as the bound $\text{var}(V_1 + V_2) \leq 2\text{var}(V_1) + 2\text{var}(V_2)$ for generic random variables V_1 and V_2 , we have

$$(A.70) \quad \psi_2 \lesssim \text{var}((\mathbf{x}_1^\top \mathbf{x}_1 - \mathbf{x}_2^\top \mathbf{x}_2)^2) + \text{var}((\mathbf{x}_1^\top \mathbf{x}_2)^2).$$

The second term on the right side can be controlled with Lemma D.2 in Wang and Lopes (2023) and equation (A.5), which gives

$$\begin{aligned} \text{var}((\mathbf{x}_1^\top \mathbf{x}_2)^2) &= 3r_2^2(\text{tr}(\Sigma^2)^2 + 2\text{tr}(\Sigma^4)) - \text{tr}(\Sigma^2)^2 \\ &\lesssim \text{tr}(\Sigma^2)^2. \end{aligned}$$

For the first term on the right side of (A.70), we have

$$\text{var}((\mathbf{x}_1^\top \mathbf{x}_1 - \mathbf{x}_2^\top \mathbf{x}_2)^2) \lesssim \mathbb{E}((\mathbf{x}_1^\top \mathbf{x}_1 - \text{tr}(\Sigma))^4).$$

By expanding out the fourth moment directly, and using Lemmas D.5 and B.7 as well as the definition of r_k in (A.3), we have

$$\begin{aligned} \mathbb{E}\left(|\mathbf{x}_1^\top \mathbf{x}_1 - \text{tr}(\Sigma)|^4\right) &= 48r_4\text{tr}(\Sigma^4) + 12r_4\text{tr}(\Sigma^2)^2 - 32(r_3 - r_4)\text{tr}(\Sigma)\text{tr}(\Sigma^3) \\ &\quad + 12(r_2 - 2r_3 + r_4)\text{tr}(\Sigma)^2\text{tr}(\Sigma^2) - (3 - 6r_2 + 4r_3 - r_4)\text{tr}(\Sigma)^4 \\ &= 48\text{tr}(\Sigma^4) + 12\text{tr}(\Sigma^2)^2 + o(p^{-1})\text{tr}(\Sigma)^4 + o(p^{-3/4})\text{tr}(\Sigma)^2\text{tr}(\Sigma^2). \end{aligned} \tag{A.71}$$

Applying the last few steps to (A.70) yields

$$\psi_2 \lesssim \text{tr}(\Sigma^2)^2 + o(p^{-1})\text{tr}(\Sigma)^4 + o(p^{-3/4})\text{tr}(\Sigma)^2\text{tr}(\Sigma^2).$$

When $\mathbf{r}(\Sigma) \lesssim \sqrt{p}$, all the three terms on the right side are $\mathcal{O}(\text{tr}(\Sigma^2)^2)$. On the other hand, if $\mathbf{r}(\Sigma)/\sqrt{p} \rightarrow \infty$ then all three terms are $o(\text{tr}(\Sigma)^4)$, which completes the proof. $\square$

LEMMA B.11. *Suppose that Assumption 1 holds. If $\mathbf{r}(\Sigma)/\sqrt{p} \rightarrow \infty$ as $n \rightarrow \infty$, then*

$$\psi_1 = o\left(\frac{\text{tr}(\Sigma)^4}{p}\right).$$

Alternatively, if $\mathbf{r}(\Sigma) \lesssim \sqrt{p}$, then

$$\psi_1 = 4\text{tr}(\Sigma^4) + 2\text{tr}(\Sigma^2)^2 + o(\text{tr}(\Sigma^2)^2).$$

PROOF. Using the definition of $h(\mathbf{x}_1, \mathbf{x}_2)$ in (A.52) and noting that $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ are independent, we have

$$\begin{aligned} \psi_1 &= \left(\frac{1}{4} + \mathcal{O}\left(\frac{1}{n}\right)\right) \text{var}\left((\mathbf{x}_1^\top \mathbf{x}_1 - \text{tr}(\Sigma))^2\right) + \left(4 + \mathcal{O}\left(\frac{1}{n}\right)\right) \text{var}(\mathbf{x}_1^\top \Sigma \mathbf{x}_1) \\ &\quad - \left(2 + \mathcal{O}\left(\frac{1}{n}\right)\right) \text{cov}\left((\mathbf{x}_1^\top \mathbf{x}_1 - \text{tr}(\Sigma))^2, \mathbf{x}_1^\top \Sigma \mathbf{x}_1\right). \end{aligned} \tag{A.72}$$

The first term on the right side can be analyzed by (A.5), (A.71) and Lemma D.2, which imply

$$\text{var}\left((\mathbf{x}_1^\top \mathbf{x}_1 - \text{tr}(\Sigma))^2\right) = 48\text{tr}(\Sigma^4) + 8\text{tr}(\Sigma^2)^2 + o(p^{-1})\text{tr}(\Sigma)^4 + o(p^{-3/4})\text{tr}(\Sigma)^2\text{tr}(\Sigma^2).$$

Lemma D.2 and (A.5) also give

$$\text{cov}(\mathbf{x}_1^\top \mathbf{x}_1, \mathbf{x}_1^\top \Sigma \mathbf{x}_1) = 2\text{tr}(\Sigma^3) + \mathcal{O}(p^{-1})\text{tr}(\Sigma)\text{tr}(\Sigma^2).$$

Noting that

$$\text{cov}\left((\mathbf{x}_1^\top \mathbf{x}_1 - \text{tr}(\Sigma))^2, \mathbf{x}_1^\top \Sigma \mathbf{x}_1\right) = \text{cov}\left((\mathbf{x}_1^\top \mathbf{x}_1)^2, \mathbf{x}_1^\top \Sigma \mathbf{x}_1\right) - 2\text{tr}(\Sigma)\text{cov}(\mathbf{x}_1^\top \mathbf{x}_1, \mathbf{x}_1^\top \Sigma \mathbf{x}_1),$$

it remains to analyze $\text{cov}((\mathbf{x}_1^\top \mathbf{x}_1)^2, \mathbf{x}_1^\top \Sigma \mathbf{x}_1)$. To calculate the covariance term, we may assume without loss of generality that Σ is diagonal, since $\mathbf{x}_1$ follows an elliptical distribution. It follows from the expression for $\mathbf{x}_i$ in (A.2) that

$$(A.73) \quad \text{cov}\left((\mathbf{x}_1^\top \mathbf{x}_1)^2, \mathbf{x}_1^\top \Sigma \mathbf{x}_1\right) = \sum_{j,k,l=1}^p \lambda_l(\Sigma)^2 \lambda_j(\Sigma) \lambda_k(\Sigma) \omega_{j,k,l}$$

where we define

$$\omega_{j,k,l} = \text{cov}\left(\frac{\xi_1^4}{\|\mathbf{z}_1\|_2^4} z_{1j}^2 z_{1k}^2, \frac{\xi_1^2}{\|\mathbf{z}_1\|_2^2} z_{1l}^2\right).$$

This quantity can be directly calculated as

$$\omega_{j,k,l} = \begin{cases} r_3 - r_2 & \text{if } j \neq k \neq l \\ 3(r_3 - r_2) & \text{if } j = k \neq l \\ 3r_3 - r_2 & \text{if } j \neq k = l \text{ or } k \neq j = l \\ 15r_3 - 3r_2 & \text{if } j = k = l. \end{cases}$$

Applying this result to (A.73), we conclude that

$$\begin{aligned} \text{cov}\left((\mathbf{x}_1^\top \mathbf{x}_1)^2, \mathbf{x}_1^\top \Sigma \mathbf{x}_1\right) &= (15r_3 - 3r_2) \sum_{l=1}^p \lambda_l(\Sigma)^4 + 2(3r_3 - r_2) \sum_{l \neq j} \lambda_l(\Sigma)^3 \lambda_j(\Sigma) \\ &\quad + 3(r_3 - r_2) \sum_{l \neq j} \lambda_l(\Sigma)^2 \lambda_j(\Sigma)^2 + (r_3 - r_2) \sum_{j \neq k \neq l} \lambda_l(\Sigma)^2 \lambda_j(\Sigma) \lambda_k(\Sigma) \\ &= 8\text{tr}(\Sigma^4) + 4\text{tr}(\Sigma^3)\text{tr}(\Sigma) + o(p^{-3/4})\text{tr}(\Sigma)^2\text{tr}(\Sigma^2), \end{aligned}$$

where the last step is based on Lemma B.7. Applying the above results to (A.72) yields

$$\psi_1 = 4\text{tr}(\Sigma^4) + 2\text{tr}(\Sigma^2)^2 + o(p^{-1})\text{tr}(\Sigma)^4 + o(p^{-3/4})\text{tr}(\Sigma)^2\text{tr}(\Sigma^2).$$

When $\mathbf{r}(\Sigma)/\sqrt{p} \rightarrow \infty$, the terms on the right side are all $o(\text{tr}(\Sigma)^4)$, and when $\mathbf{r}(\Sigma) \lesssim \sqrt{p}$, we have $p^{-1}\text{tr}(\Sigma)^4 \lesssim \text{tr}(\Sigma^2)^2$ as well as $p^{-3/4}\text{tr}(\Sigma)^2 \lesssim \text{tr}(\Sigma^2)$, which completes the proof of the lemma. $\square$

A.3. Ratio consistency of $\hat{\sigma}_n^2$

In order to state and prove the next result, recall from equation (2.4) that the parameter $\sigma_n^2 = \sigma_{n,1}^2 + \sigma_{n,2}^2$ is defined by

$$\sigma_{n,1}^2 = \frac{8\|R\|_4^4}{3p} \quad \text{and} \quad \sigma_{n,2}^2 = \frac{8p(2\nu_4 + \nu_2^2)}{(\nu_1^2 + 2\nu_2)^2}.$$

In addition, recall from equation (2.5) that the estimate $\hat{\sigma}_n^2 = \hat{\sigma}_{n,1}^2 + \hat{\sigma}_{n,2}^2$ is defined by

$$\hat{\sigma}_{n,1}^2 = \frac{8}{3p} \left(\|\hat{R}\|_4^4 - \hat{\beta}\|\hat{R}\|_4^4 + 3\llbracket 1 - \hat{\beta} + \hat{\gamma} \rrbracket_{t_p} \|\hat{R}\|_2^2 \right) \quad \text{and} \quad \hat{\sigma}_{n,2}^2 = \frac{8p(2\hat{\nu}_4 + (\hat{\nu}_2 - \frac{1}{n}\hat{\nu}_1^2)^2)}{(\hat{\nu}_1^2 + 2(\hat{\nu}_2 - \frac{1}{n}\hat{\nu}_1^2))^2}.$$

PROPOSITION A.3.1. *If Assumption 1 holds, then as $n \rightarrow \infty$,*

$$\frac{\hat{\sigma}_n^2}{\sigma_n^2} \xrightarrow{\mathbb{P}} 1.$$

PROOF. It is sufficient to establish the limits $\hat{\sigma}_{n,1}^2/\sigma_{n,1}^2 \xrightarrow{\mathbb{P}} 1$ and $(\hat{\sigma}_{n,2}^2 - \sigma_{n,2}^2)/\sigma_n^2 \xrightarrow{\mathbb{P}} 0$, which is done in Lemmas C.1 and C.2 below. $\square$

LEMMA C.1. *If Assumption 1 holds, then as $n \rightarrow \infty$,*

$$\frac{\hat{\sigma}_{n,1}^2}{\sigma_{n,1}^2} \xrightarrow{\mathbb{P}} 1.$$

PROOF. Note that

$$\frac{\hat{\sigma}_{n,1}^2}{\sigma_{n,1}^2} - 1 = \frac{\|\hat{R}\|_4^4 - \|R\|_4^4}{\|R\|_4^4} - \frac{\hat{\beta}\|\hat{R}\|_4^4}{\|R\|_4^4} + \frac{3\|\hat{R}\|_2^2}{\|R\|_4^4} \llbracket 1 - \hat{\beta} + \hat{\gamma} \rrbracket_{t_p}.$$

Lemma C.3 shows that the first term on the right is $o_{\mathbb{P}}(1)$. Also, Lemmas C.3 and C.4 together show that the second term on the right is $o_{\mathbb{P}}(1)$. For the last term, combining $\llbracket 1 - \hat{\beta} + \hat{\gamma} \rrbracket_{t_p} = \mathcal{O}(\log(p)p^{-3/4})$ with Lemma C.3 gives

$$\begin{aligned} \llbracket 1 - \hat{\beta} + \hat{\gamma} \rrbracket_{t_p} \|\hat{R}\|_2^2 &= \mathcal{O}(\log(p)p^{-3/4}) \left(\|R\|_2^2 + \mathcal{O}_{\mathbb{P}}(\|R\|_2^2) \right) \\ &= o_{\mathbb{P}}(\|R\|_4^4), \end{aligned}$$

where the last step uses the inequality $\frac{\|R\|_2^2}{\|R\|_4^4} \leq p^{1/2}$ that holds for a generic correlation matrix R . $\square$

LEMMA C.2. *If Assumption 1 holds, then as $n \rightarrow \infty$,*

$$\frac{\hat{\sigma}_{n,2}^2 - \sigma_{n,2}^2}{\sigma_n^2} \xrightarrow{\mathbb{P}} 0.$$

PROOF. Observe that $(\hat{\sigma}_{n,2}^2 - \sigma_{n,2}^2)/\sigma_n^2$ can be decomposed as

$$\frac{\hat{\sigma}_{n,2}^2 - \sigma_{n,2}^2}{\sigma_n^2} = \frac{8p}{\sigma_n^2(\nu_1^2 + 2\nu_2)^2} \left(\frac{(\nu_1^2 + 2\nu_2)^2}{(\hat{\nu}_1^2 + 2(\hat{\nu}_2 - \frac{1}{n}\hat{\nu}_1^2))^2} \left(2\hat{\nu}_4 + (\hat{\nu}_2 - \frac{1}{n}\hat{\nu}_1^2)^2 \right) - (2\nu_4 + \nu_2^2) \right).$$

Also, we have

$$\begin{aligned} \frac{\sigma_n^2(\nu_1^2 + 2\nu_2)^2}{p} &= 8(2\nu_4 + \nu_2^2) + \frac{8\|R\|_4^4}{3p^2}(\nu_1^2 + 2\nu_2)^2 \\ &\gtrsim \nu_2^2 + \frac{1}{p}\nu_1^4. \end{aligned}$$

If we can establish the following two conditions

$$(A.74) \quad \frac{(\nu_1^2 + 2\nu_2)^2}{(\hat{\nu}_1^2 + 2(\hat{\nu}_2 - \frac{1}{n}\hat{\nu}_1^2))^2} = 1 + o_{\mathbb{P}}(1)$$

$$(A.75) \quad 2\hat{\nu}_4 + (\hat{\nu}_2 - \frac{1}{n}\hat{\nu}_1^2)^2 = 2\nu_4 + \nu_2^2 + o_{\mathbb{P}}(\nu_2^2 + \frac{1}{p}\nu_1^4),$$

then combining with $\nu_4 \lesssim \nu_2^2$ yields

$$\begin{aligned} \left| \frac{\hat{\sigma}_{n,2}^2 - \sigma_{n,2}^2}{\sigma_n^2} \right| &\lesssim \frac{1}{\nu_2^2 + \frac{1}{p}\nu_1^4} \left| (1 + o_{\mathbb{P}}(1)) \left(2\nu_4 + \nu_2^2 + o_{\mathbb{P}}(\nu_2^2 + \frac{1}{p}\nu_1^4) \right) - (2\nu_4 + \nu_2^2) \right| \\ &= o_{\mathbb{P}}(1). \end{aligned}$$

Therefore, it is enough to show (A.74) and (A.75). The first of these conditions follows from

$$(A.76) \quad \frac{\hat{\nu}_2 - \frac{1}{n}\hat{\nu}_1^2}{\nu_2} \xrightarrow{\mathbb{P}} 1 \quad \text{and} \quad \frac{\hat{\nu}_1^2}{\nu_1^2} \xrightarrow{\mathbb{P}} 1,$$

which can be proven in the same manner as (A.49). Lastly, the condition (A.75) is established in Lemma C.6. $\square$

A.3.1. Norms of the sample correlation matrix.

LEMMA C.3. *If Assumption 1 holds, then*

$$(A.77) \quad \|\widehat{R}\|_4^4 - \|R\|_4^4 = o_{\mathbb{P}}(\|R\|_4^4)$$

$$(A.78) \quad \|\widehat{R}\|_2^2 - \|R\|_2^2 = \mathcal{O}_{\mathbb{P}}(\|R\|_2^2).$$

PROOF. The proof of Lemma B.1 shows that $\sum_{j=1}^p \mathbb{P}(|\widehat{\Sigma}_{jj}/\Sigma_{jj} - 1| > n^{-1/20}) = o(1/p)$, and so

$$\mathbb{P}\left(\max_{1 \leq j \leq p} |\widehat{\Sigma}_{jj}/\Sigma_{jj} - 1| > n^{-1/20}\right) = o(1/p).$$

This implies

$$\left| \|\widehat{R}\|_4^4 - \|R\|_4^4 \right| \leq o_{\mathbb{P}}(1) \sum_{j,k=1}^p \frac{\widehat{\Sigma}_{jk}^4}{\Sigma_{jj}^2 \Sigma_{kk}^2} + \left| \sum_{j,k=1}^p \frac{\widehat{\Sigma}_{jk}^4}{\Sigma_{jj}^2 \Sigma_{kk}^2} - \|R\|_4^4 \right|$$

and

$$\left| \|\widehat{R}\|_2^2 - \|R\|_2^2 \right| \leq o_{\mathbb{P}}(1) \sum_{j,k=1}^p \frac{\widehat{\Sigma}_{jk}^2}{\Sigma_{jj} \Sigma_{kk}} + \left| \sum_{j,k=1}^p \frac{\widehat{\Sigma}_{jk}^2}{\Sigma_{jj} \Sigma_{kk}} - \|R\|_2^2 \right|.$$

As a consequence, we may work with $\widehat{\Sigma}$ in place of $\widehat{R}$ and assume that $\Sigma = R$.

To begin, we bound the L^1 error of $\|\widehat{\Sigma}\|_4^4$ by isolating its bias,

$$(A.79) \quad \left\| \|\widehat{\Sigma}\|_4^4 - \|\Sigma\|_4^4 \right\|_{L^1} \leq \left\| \|\widehat{\Sigma}\|_4^4 - \mathbb{E}(\|\widehat{\Sigma}\|_4^4) \right\|_{L^1} + \left| \mathbb{E}(\|\widehat{\Sigma}\|_4^4) - \|\Sigma\|_4^4 \right|.$$

To control the bias, we calculate the expectation of $\|\widehat{\Sigma}\|_4^4$ directly as

$$\begin{aligned} \mathbb{E}(\|\widehat{\Sigma}\|_4^4) &= \sum_{j,k=1}^p \mathbb{E}(\widehat{\Sigma}_{jk}^4) \\ &= \frac{1}{n^4} \sum_{j,k=1}^p \sum_{i_1, \dots, i_4=1}^n \mathbb{E}\left(\prod_{l=1}^4 x_{i_l j} x_{i_l k}\right). \end{aligned}$$

In the sum over i_1, i_2, i_3, i_4 (with j and k held fixed), the number of terms involving 4 distinct indices is $n(n-1)(n-2)(n-3)$, and the common expectation of these terms is Σ_{jk}^4 . The number of terms involving 3 distinct indices is $6n(n-1)(n-2)$, and by the proof of Lemma D.6 their common expectation is $r_2 \Sigma_{jk}^2 (1 + 2\Sigma_{jk}^2)$. The number of terms involving 1 or 2 distinct indices is $\mathcal{O}(n^2)$ and by (A.6) their expectations are $\mathcal{O}(1)$. Combining these observations with $r_2 \lesssim 1$ and the fact that

$\|\Sigma\|_2 \leq p^{1/4} \|\Sigma\|_4^2$ holds when $\Sigma = R$, we have

$$(A.80) \quad \begin{aligned} \mathbb{E}(\|\widehat{\Sigma}\|_4^4) - \|\Sigma\|_4^4 &= (1 + \mathcal{O}(\frac{1}{n}))\|\Sigma\|_4^4 + \mathcal{O}(\frac{1}{n})(\|\Sigma\|_2^2 + \|\Sigma\|_4^4) + \mathcal{O}(\frac{1}{n^2}) - \|\Sigma\|_4^4 \\ &= o(\|\Sigma\|_4^4). \end{aligned}$$

Hence, this handles the bias term in the bound (A.79). Next, we address the L^1 deviation term in that bound by introducing an independent copy of $\widehat{\Sigma}$, denoted by $\widehat{\Sigma}'$. Using Jensen's inequality, as well as the basic inequality $|a^4 - b^4| \leq 4|a - b|(a^3 + b^3)$ for non-negative a and b , we have

$$\begin{aligned} \mathbb{E} \left[\left| \|\widehat{\Sigma}\|_4^4 - \mathbb{E}(\|\widehat{\Sigma}\|_4^4) \right| \right] &\lesssim \mathbb{E} \left[\|\widehat{\Sigma} - \widehat{\Sigma}'\|_4 \left(\|\widehat{\Sigma}\|_4^3 + \|\widehat{\Sigma}'\|_4^3 \right) \right] \\ &\lesssim \left(\mathbb{E}[\|\widehat{\Sigma} - \widehat{\Sigma}'\|_4^4] \right)^{1/4} \left(\mathbb{E} \left[\left(\|\widehat{\Sigma}\|_4^3 + \|\widehat{\Sigma}'\|_4^3 \right)^{4/3} \right] \right)^{3/4}, \end{aligned}$$

where Hölder's inequality with exponents 4 and 4/3 has been used in the last step. By simplifying the last line with the basic inequality $(a^3 + b^3)^{4/3} \leq 2^{4/3}(a^4 + b^4)$ and then applying our calculation of the bias of $\|\widehat{\Sigma}\|_4^4$ in equation (A.80), we have

$$\begin{aligned} \mathbb{E} \left[\left| \|\widehat{\Sigma}\|_4^4 - \mathbb{E}(\|\widehat{\Sigma}\|_4^4) \right| \right] &\lesssim \left(\mathbb{E}[\|\widehat{\Sigma} - \widehat{\Sigma}'\|_4^4] \right)^{1/4} \left(\mathbb{E}[\|\widehat{\Sigma}\|_4^4] \right)^{3/4} \\ &\lesssim \left(\mathbb{E}[\|\widehat{\Sigma} - \widehat{\Sigma}'\|_4^4] \right)^{1/4} \|\Sigma\|_4^3. \end{aligned}$$

Hence, it remains to show $\mathbb{E}[\|\widehat{\Sigma} - \Sigma\|_4^4] = o(\|\Sigma\|_4^4)$. Letting $\mathbf{e}_j \in \mathbb{R}^p$ denote the j th canonical basis vector, Rosenthal's inequality (Lemma D.7) gives

$$\begin{aligned} \mathbb{E}[\|\widehat{\Sigma} - \Sigma\|_4^4] &= \sum_{j,k=1}^p \left\| \sum_{i=1}^n \frac{1}{n} (\mathbf{e}_j^\top \mathbf{x}_i \mathbf{x}_i^\top \mathbf{e}_k - \mathbf{e}_j^\top \Sigma \mathbf{e}_k) \right\|_{L^4}^4 \\ &\lesssim \sum_{j,k=1}^p \max \left\{ \left(\frac{1}{n} \text{var}(\mathbf{e}_j^\top \mathbf{x}_1 \mathbf{x}_1^\top \mathbf{e}_k) \right)^2, \frac{1}{n^3} \left\| \mathbf{e}_j^\top \mathbf{x}_1 \mathbf{x}_1^\top \mathbf{e}_k - \mathbf{e}_j^\top \Sigma \mathbf{e}_k \right\|_{L^4}^4 \right\}. \end{aligned}$$

Since we may assume that all the diagonal entries of Σ are equal to 1, it follows from Lemmas D.2, D.4 and B.7 that the variance and L^4 norm inside the maximum are both $\mathcal{O}(1)$. This leads to $\mathbb{E}[\|\widehat{\Sigma} - \Sigma\|_4^4] \lesssim 1 = o(\|\Sigma\|_4^4)$ and completes the proof of (A.77).

To prove (A.78), the L^1 triangle inequality gives

$$(A.81) \quad \left\| \|\widehat{\Sigma}\|_2^2 - \|\Sigma\|_2^2 \right\|_{L^1} \leq \left\| (\|\widehat{\Sigma}\|_2^2 - \frac{1}{n} \text{tr}(\widehat{\Sigma})^2) - \|\Sigma\|_2^2 \right\|_{L^1} + \left\| \frac{1}{n} \text{tr}(\widehat{\Sigma})^2 \right\|_{L^1}.$$

Lemma A.1 in (Wang and Lopes, 2023) shows that the first term on the right side of (A.81) is $o(\|\Sigma\|_2^2)$. For the second term in (A.81), Lemma D.2 gives

$$\begin{aligned}\|\text{tr}(\widehat{\Sigma})^2\|_{L^1} &= \frac{n-1}{n}\text{tr}(\Sigma)^2 + \frac{r_2}{n}\left(\text{tr}(\Sigma)^2 + 2\text{tr}(\Sigma^2)\right) \\ &\lesssim p\|\Sigma\|_2^2,\end{aligned}$$

where the last step uses Lemma B.7 and the condition that all diagonal entries of Σ are equal to 1. $\square$

A.3.2. The limit of $\widehat{\beta}$.

LEMMA C.4. *If Assumption 1 holds, then as $n \rightarrow \infty$,*

$$\widehat{\beta} \xrightarrow{\mathbb{P}} 0.$$

PROOF. Recall the definition of $\widehat{\beta}$ in equation (2.7) and let $\Delta(\widehat{\Sigma}) = -\frac{12}{n}(\widehat{\nu}_1^4 + 2\widehat{\nu}_1^2\widehat{\nu}_2 - \frac{1}{n}\widehat{\nu}_1^4)$. Since the quantity $r_4 = \frac{\mathbb{E}(\|\mathbf{x}_1\|_2^8)}{g_4(\Sigma)}$ satisfies $r_4 = 1 + o(1)$ by Lemma B.7, we have

$$\begin{aligned}\widehat{\beta} &= \frac{\mathbb{E}(\|\mathbf{x}_1\|_2^8)}{g_4(\Sigma)} - \frac{\frac{1}{n}\sum_{i=1}^n \|\mathbf{x}_i\|_2^8}{g_4(\widehat{\Sigma}) + \Delta(\widehat{\Sigma})} + o(1) \\ &= r_4 \left(1 - \frac{g_4(\Sigma)}{g_4(\widehat{\Sigma}) + \Delta(\widehat{\Sigma})}\right) + r_4 \frac{g_4(\Sigma)}{g_4(\widehat{\Sigma}) + \Delta(\widehat{\Sigma})} \left(1 - \frac{\frac{1}{n}\sum_{i=1}^n \|\mathbf{x}_i\|_2^8}{\mathbb{E}(\|\mathbf{x}_1\|_2^8)}\right) + o(1).\end{aligned}$$

Since Lemma C.5 shows $g_4(\Sigma)/(g_4(\widehat{\Sigma}) + \Delta(\widehat{\Sigma})) \xrightarrow{\mathbb{P}} 1$, it is sufficient to show $\frac{1}{n}\sum_{i=1}^n \|\mathbf{x}_i\|_2^8$ is a ratio-consistent estimator of $\mathbb{E}(\|\mathbf{x}_1\|_2^8)$. To this end, observe that

$$\begin{aligned}\left\|\frac{\frac{1}{n}\sum_{i=1}^n \|\mathbf{x}_i\|_2^8}{\mathbb{E}(\|\mathbf{x}_1\|_2^8)} - 1\right\|_{L^2}^2 &= \frac{\text{var}(\|\mathbf{x}_1\|_2^8)}{n(\mathbb{E}(\|\mathbf{x}_1\|_2^8))^2} \\ &\lesssim \frac{1}{n} \frac{r_8}{r_4^2} \frac{\mathbb{E}((\mathbf{z}_1^\top \Sigma \mathbf{z}_1)^8)}{(\mathbb{E}((\mathbf{z}_1^\top \Sigma \mathbf{z}_1)^4))^2}.\end{aligned}$$

Combining the fact $\mathbb{E}((\mathbf{z}_1^\top \Sigma \mathbf{z}_1)^4) \geq (\mathbb{E}(\mathbf{z}_1^\top \Sigma \mathbf{z}_1))^4 = \text{tr}(\Sigma)^4$ with $r_8/r_4^2 \lesssim 1$ shown in Lemma B.7 and $\mathbb{E}((\mathbf{z}_1^\top \Sigma \mathbf{z}_1)^8) \lesssim \text{tr}(\Sigma)^8$ given by Lemma D.3, we conclude $\frac{1}{n}\sum_{i=1}^n \|\mathbf{x}_i\|_2^8/\mathbb{E}(\|\mathbf{x}_1\|_2^8) \xrightarrow{\mathbb{P}} 1$, which completes the proof. $\square$

LEMMA C.5. *If Assumption 1 holds, then as $n \rightarrow \infty$,*

$$\frac{g_4(\widehat{\Sigma}) + \Delta(\widehat{\Sigma})}{g_4(\Sigma)} \xrightarrow{\mathbb{P}} 1.$$

PROOF. Noting that equation (A.76) and Lemma D.5 imply $\Delta(\widehat{\Sigma})/g_4(\Sigma) \xrightarrow{\mathbb{P}} 0$, it is enough to show $g_4(\widehat{\Sigma})$ is a ratio-consistent estimator of $g_4(\Sigma)$. Let $\mathbf{z} \in \mathbb{R}^p$ denote a standard normal vector independent of $\mathbf{x}_1, \dots, \mathbf{x}_n$, and observe that

$$g_4(\widehat{\Sigma}) - g_4(\Sigma) = \mathbb{E} \left[(\mathbf{z}^\top \widehat{\Sigma} \mathbf{z})^4 - (\mathbf{z}^\top \Sigma \mathbf{z})^4 \middle| \widehat{\Sigma} \right].$$

So, it is sufficient to show that

$$\left\| (\mathbf{z}^\top \widehat{\Sigma} \mathbf{z})^4 - (\mathbf{z}^\top \Sigma \mathbf{z})^4 \right\|_{L^1} = o(g_4(\Sigma)).$$

Using the basic inequality $|a^4 - b^4| \leq 4|a - b|\sqrt{a^6 + b^6}$ that holds for non-negative a and b , it follows from the Cauchy-Schwarz inequality that

$$(A.82) \quad \left\| (\mathbf{z}^\top \widehat{\Sigma} \mathbf{z})^4 - (\mathbf{z}^\top \Sigma \mathbf{z})^4 \right\|_{L^1} \lesssim \left\| \mathbf{z}^\top \widehat{\Sigma} \mathbf{z} - \mathbf{z}^\top \Sigma \mathbf{z} \right\|_{L^2} \left\| (\mathbf{z}^\top \widehat{\Sigma} \mathbf{z})^6 + (\mathbf{z}^\top \Sigma \mathbf{z})^6 \right\|_{L^1}^{1/2}.$$

To handle the second factor, Lemma D.3 implies

$$\left\| (\mathbf{z}^\top \widehat{\Sigma} \mathbf{z})^6 + (\mathbf{z}^\top \Sigma \mathbf{z})^6 \right\|_{L^1}^{1/2} \lesssim \left(\mathbb{E}(\text{tr}(\widehat{\Sigma})^6) + \text{tr}(\Sigma)^6 \right)^{1/2}.$$

By viewing $\text{tr}(\widehat{\Sigma})$ as the sample average $\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|_2^2$, Jensen's inequality implies

$$\begin{aligned} \mathbb{E}(\text{tr}(\widehat{\Sigma})^6) &\leq \frac{1}{n} \sum_{i=1}^n \mathbb{E}(\|\mathbf{x}_i\|_2^{12}) \\ &\lesssim \text{tr}(\Sigma)^6, \end{aligned}$$

where the last step follows from Lemma D.4. To handle the first factor on the right side of (A.82), observe that Lemma D.5 gives

$$\mathbb{E} \left[(\mathbf{z}^\top \widehat{\Sigma} \mathbf{z} - \mathbf{z}^\top \Sigma \mathbf{z})^2 \middle| \widehat{\Sigma} \right] = 2\|\widehat{\Sigma} - \Sigma\|_2^2 + (\text{tr}(\widehat{\Sigma}) - \text{tr}(\Sigma))^2.$$

Furthermore, Lemmas D.2 and B.7 imply

$$\begin{aligned}
\mathbb{E}(\|\widehat{\Sigma} - \Sigma\|_2^2) &= \frac{1}{n} \sum_{j,k=1}^p \text{var}(\mathbf{x}_1^\top \mathbf{e}_k \mathbf{e}_j^\top \mathbf{x}_1) \\
&\lesssim \frac{1}{n} \sum_{j,k=1}^p (\Sigma_{jk}^2 + \Sigma_{jj} \Sigma_{kk}) \\
&\lesssim \frac{1}{n} \text{tr}(\Sigma)^2
\end{aligned}$$

as well as

$$\mathbb{E}((\text{tr}(\widehat{\Sigma}) - \text{tr}(\Sigma))^2) = \frac{1}{n} \text{var}(\mathbf{x}_1^\top \mathbf{x}_1) \lesssim \frac{1}{n} \text{tr}(\Sigma)^2.$$

The last few steps imply

$$\left\| \mathbf{z}^\top \widehat{\Sigma} \mathbf{z} - \mathbf{z}^\top \Sigma \mathbf{z} \right\|_{L^2} \lesssim \frac{1}{\sqrt{n}} \text{tr}(\Sigma).$$

Applying the above results to (A.82) and observing $g_4(\Sigma) \geq \text{tr}(\Sigma)^4$, we have

$$\left\| (\mathbf{z}^\top \widehat{\Sigma} \mathbf{z})^4 - (\mathbf{z}^\top \Sigma \mathbf{z})^4 \right\|_{L^1} = \mathcal{O}(\frac{1}{\sqrt{n}} \text{tr}(\Sigma)^4) = o(g_4(\Sigma)),$$

which completes the proof. □

A.3.3. Consistency of $\widehat{\nu}_4$.

LEMMA C.6. *If Assumption 1 holds, then as $n \rightarrow \infty$,*

$$\frac{\widehat{\nu}_4 - \nu_4}{\nu_2^2 + \frac{1}{p} \nu_1^4} \xrightarrow{\mathbb{P}} 0.$$

PROOF. It is enough to show $\mathbb{E}(|\text{tr}(\widehat{\Sigma}^4) - \text{tr}(\Sigma^4)|) = o(\text{tr}(\Sigma^2)^2 + \frac{\text{tr}(\Sigma)^4}{p})$. First, observe that the Cauchy-Schwarz inequality implies

$$\begin{aligned}
|\text{tr}(\widehat{\Sigma}^4) - \text{tr}(\Sigma^4)|^2 &= \left| \sum_{j=1}^p (\lambda_j(\widehat{\Sigma})^4 - \lambda_j(\Sigma)^4) \right|^2 \\
&\leq \sum_{j=1}^p \left(\lambda_j(\widehat{\Sigma})^2 - \lambda_j(\Sigma)^2 \right)^2 \sum_{j=1}^p \left(\lambda_j(\widehat{\Sigma})^2 + \lambda_j(\Sigma)^2 \right)^2 \\
&\leq 2 \|\widehat{\Sigma}^2 - \Sigma^2\|_2^2 \left(\text{tr}(\widehat{\Sigma}^4) + \text{tr}(\Sigma^4) \right),
\end{aligned}$$

where Wielandt's inequality (Horn and Johnson, 2012, Corollary 7.3.5) has been used in the last step. Taking the square root of both sides, followed by the expectation, the Cauchy-Schwarz inequality implies

$$(A.83) \quad \begin{aligned} \mathbb{E}(|\text{tr}(\widehat{\Sigma}^4) - \text{tr}(\Sigma^4)|) &\lesssim \sqrt{\mathbb{E}(\|\widehat{\Sigma}^2 - \Sigma^2\|_2^2)} \sqrt{\mathbb{E}(\text{tr}(\widehat{\Sigma}^4)) + \text{tr}(\Sigma^4)} \\ &= \sqrt{\mathbb{E}(\text{tr}(\widehat{\Sigma}^4)) - 2\mathbb{E}(\text{tr}(\Sigma^2\widehat{\Sigma}^2)) + \text{tr}(\Sigma^4)} \sqrt{\mathbb{E}(\text{tr}(\widehat{\Sigma}^4)) + \text{tr}(\Sigma^4)}. \end{aligned}$$

In the remainder of the proof, we will separately analyze $\mathbb{E}(\text{tr}(\widehat{\Sigma}^4))$ and $\mathbb{E}(\text{tr}(\Sigma^2\widehat{\Sigma}^2))$. For the first of these quantities, note that

$$(A.84) \quad \mathbb{E}(\text{tr}(\widehat{\Sigma}^4)) = \frac{1}{n^4} \sum_{i_1, \dots, i_4=1}^n \mathbb{E} \left(\text{tr} \left(\prod_{l=1}^4 \mathbf{x}_{i_l} \mathbf{x}_{i_l}^\top \right) \right).$$

The number of terms involving 4 distinct indices is $n(n-1)(n-2)(n-3)$ and the corresponding expectation is $\text{tr}(\Sigma^4)$. For the terms in (A.84) that involve 3 distinct indices, there are two possible cases. One case occurs when $i_1 = i_2 \neq i_3 \neq i_4$. The number of corresponding terms is $\mathcal{O}(n^3)$, and by Lemmas D.2 and B.7, the expectation of each such term is $r_2(2\text{tr}(\Sigma^4) + \text{tr}(\Sigma^3)\text{tr}(\Sigma)) \lesssim \sqrt{p}\text{tr}(\Sigma^2)^2$. The second case occurs when $i_1 = i_3 \neq i_2 \neq i_4$. The number of corresponding terms is $\mathcal{O}(n^3)$, and by Lemmas D.2 and B.7, the expectation of each such term is $r_2(2\text{tr}(\Sigma^4) + \text{tr}(\Sigma^2)^2) \lesssim \text{tr}(\Sigma^2)^2$. The number of terms in (A.84) that involve 1 or 2 distinct indices is $\mathcal{O}(n^2)$, with their expectations being at most $\mathcal{O}(\text{tr}(\Sigma)^4)$. Consequently, we have

$$\mathbb{E}(\text{tr}(\widehat{\Sigma}^4)) = \text{tr}(\Sigma^4) + \mathcal{O}(\frac{\sqrt{p}}{n})\text{tr}(\Sigma^2)^2 + \mathcal{O}(\frac{1}{n^2})\text{tr}(\Sigma)^4.$$

In addition, Lemmas D.2 and B.7 give

$$\begin{aligned} \mathbb{E}(\text{tr}(\Sigma^2\widehat{\Sigma}^2)) &= \frac{1}{n^2} \sum_{i_1, i_2=1}^n \mathbb{E}(\mathbf{x}_{i_1}^\top \mathbf{x}_{i_2} \mathbf{x}_{i_2}^\top \Sigma^2 \mathbf{x}_{i_1}) \\ &= \frac{n-1}{n} \text{tr}(\Sigma^4) + \frac{r_2}{n} \left(\text{tr}(\Sigma^3)\text{tr}(\Sigma) + 2\text{tr}(\Sigma^4) \right) \\ &= \text{tr}(\Sigma^4) + \mathcal{O}(\frac{\sqrt{p}}{n})\text{tr}(\Sigma^2)^2. \end{aligned}$$

Applying the last two bounds to (A.83), it follows that

$$\begin{aligned}
\mathbb{E}(|\text{tr}(\widehat{\Sigma}^4) - \text{tr}(\Sigma^4)|) &\lesssim \left(\frac{\sqrt{p}}{n} \text{tr}(\Sigma^2)^2 + \frac{1}{n^2} \text{tr}(\Sigma)^4 \right)^{1/2} \left(\text{tr}(\Sigma^4) + \frac{\sqrt{p}}{n} \text{tr}(\Sigma^2)^2 + \frac{1}{n^2} \text{tr}(\Sigma)^4 \right)^{1/2} \\
&= o\left((\text{tr}(\Sigma^2)^2 + \frac{\text{tr}(\Sigma)^4}{p})^{1/2} \right) \cdot \mathcal{O}\left((\text{tr}(\Sigma^2)^2 + \frac{\text{tr}(\Sigma)^4}{p})^{1/2} \right) \\
&= o\left(\text{tr}(\Sigma^2)^2 + \frac{\text{tr}(\Sigma)^4}{p} \right),
\end{aligned}$$

which implies the stated result. $\square$

A.4. Background results

LEMMA D.1. *Let $\mathbf{x}_1 = \xi_1 \Sigma^{1/2} \mathbf{u}_1$, where $\Sigma \in \mathbb{R}^{p \times p}$ is deterministic and positive semidefinite with positive diagonal entries, $\mathbf{u}_1$ is a random vector drawn from the uniform distribution on the unit sphere of $\mathbb{R}^p$, and $\xi_1 \geq 0$ is a random variable independent of $\mathbf{u}_1$. In addition, suppose that $\mathbb{E}(x_{1j}^4) < \infty$ for all $j = 1, \dots, p$. Then, $\frac{\mathbb{E}(x_{11}^4)}{(\mathbb{E}(x_{11}^2))^2} = \dots = \frac{\mathbb{E}(x_{1p}^4)}{(\mathbb{E}(x_{1p}^2))^2}$.*

PROOF. Note that the rotation invariance of the uniform distribution on the unit sphere implies that $x_{1j} = \mathbf{e}_j^\top \Sigma^{1/2} \xi_1 \mathbf{u}_1$ is equal in distribution to $\|\Sigma^{1/2} \mathbf{e}_j\|_2 \xi_1 u_{11}$, where $\mathbf{e}_j$ denotes the j th standard basis vector (with 1 in the j th entry and 0 in all other entries). Combining with the fact that kurtosis is a scale-invariant parameter, it follows that the kurtosis of x_{1j} is equal to that of $\xi_1 u_{11}$ for all $j = 1, \dots, p$. $\square$

LEMMA D.2 (Hu et al. (2019a), Lemma A.1). *Let $\xi_1 \in \mathbb{R}$ and $\mathbf{u}_1 \in \mathbb{R}^p$ satisfy the conditions in Assumption 1, and fix any matrices $M, \tilde{M} \in \mathbb{R}^{p \times p}$. Then,*

$$\text{cov}\left(\xi_1^2 \mathbf{u}_1^\top M \mathbf{u}_1, \xi_1^2 \mathbf{u}_1^\top \tilde{M} \mathbf{u}_1\right) = \frac{\mathbb{E}(\xi_1^4)}{p(p+2)} \left(\text{tr}(M) \text{tr}(\tilde{M}) + \text{tr}(M \tilde{M}) + \text{tr}(M \tilde{M}^\top) \right) - \text{tr}(M) \text{tr}(\tilde{M}).$$

LEMMA D.3 (Bai and Silverstein (2010), Lemma B.26). *Let $\mathbf{z} \in \mathbb{R}^p$ be a standard normal random vector, and fix any matrix $M \in \mathbb{R}^{p \times p}$. Then, for any $q \geq 1$,*

$$\left\| \mathbf{z}^\top M \mathbf{z} - \text{tr}(M) \right\|_{L^q} \leq C_q \text{tr}(M M^\top)^{1/2},$$

where C_q is a constant depending on q only.

LEMMA D.4. *If Assumption 1 holds, then for any fixed matrix $M \in \mathbb{R}^{p \times p}$ and any $1 \leq q \leq 8$,*

$$\left\| \mathbf{x}_1^\top M \mathbf{x}_1 - \text{tr}(M\Sigma) \right\|_{L^q} \lesssim \sqrt{\text{tr}(M\Sigma M^\top \Sigma)} + o\left(\frac{1}{p^{1/4}}\right) \text{tr}(M\Sigma).$$

PROOF. By Lyapunov's inequality, it is enough to handle the case when $q = 8$. For any fixed matrix $M \in \mathbb{R}^{p \times p}$, the definition of $\mathbf{x}_1$ and Assumption 1 give

$$\begin{aligned} \left\| \mathbf{x}_1^\top M \mathbf{x}_1 - \text{tr}(M\Sigma) \right\|_{L^8}^8 &\lesssim \mathbb{E}(\xi_1^{16}) \left\| \mathbf{u}_1^\top \Sigma^{1/2} M \Sigma^{1/2} \mathbf{u}_1 - \frac{1}{p} \text{tr}(M\Sigma) \right\|_{L^8}^8 + \left\| \text{tr}(M\Sigma)^8 \left(\frac{1}{p} \xi_1^2 - 1 \right) \right\|_{L^8}^8 \\ &\lesssim p^8 \left\| \mathbf{u}_1^\top \Sigma^{1/2} M \Sigma^{1/2} \mathbf{u}_1 - \frac{1}{p} \text{tr}(\Sigma^{1/2} M \Sigma^{1/2}) \right\|_{L^8}^8 + o\left(\frac{1}{p^2}\right) \text{tr}(M\Sigma)^8. \end{aligned}$$

If we can show

$$(A.85) \quad p^8 \left\| \mathbf{u}_1^\top A \mathbf{u}_1 - \frac{1}{p} \text{tr}(A) \right\|_{L^8}^8 \lesssim \text{tr}(AA^\top)^4$$

holds for any matrix $A \in \mathbb{R}^{p \times p}$, then the sub-additivity of the eighth-root function implies the stated result. To show (A.85), observe that

$$p^8 \left\| \mathbf{u}_1^\top A \mathbf{u}_1 - \frac{1}{p} \text{tr}(A) \right\|_{L^8}^8 \lesssim \left\| p \mathbf{u}_1^\top A \mathbf{u}_1 - \mathbf{z}_1^\top A \mathbf{z}_1 \right\|_{L^8}^8 + \left\| \mathbf{z}_1^\top A \mathbf{z}_1 - \text{tr}(A) \right\|_{L^8}^8,$$

where $\mathbf{z}_1$ denotes a standard normal vector in $\mathbb{R}^p$ satisfying $\mathbf{u}_1 = \frac{\mathbf{z}_1}{\|\mathbf{z}_1\|_2}$. The second term of this bound is $\mathcal{O}(\text{tr}(AA^\top)^4)$ due to Lemma D.3, while the first term can be rewritten as

$$\left\| p \mathbf{u}_1^\top A \mathbf{u}_1 - \mathbf{z}_1^\top A \mathbf{z}_1 \right\|_{L^8}^8 = \left\| p - \|\mathbf{z}_1\|_2^2 \right\|_{L^8}^8 \left\| \mathbf{u}_1^\top A \mathbf{u}_1 \right\|_{L^8}^8.$$

Lemma D.3 shows that the first factor on the right is $\mathcal{O}(p^4)$, and combining with the lower bound $\mathbb{E}(\|\mathbf{z}_1\|_2^{16}) \geq (\mathbb{E}(\|\mathbf{z}_1\|_2^2))^8 \geq p^8$ implies that the second factor satisfies

$$\begin{aligned} \left\| \mathbf{u}_1^\top A \mathbf{u}_1 \right\|_{L^8}^8 &\lesssim \frac{\left\| \mathbf{z}_1^\top A \mathbf{z}_1 - \text{tr}(A) \right\|_{L^8}^8}{\mathbb{E}(\|\mathbf{z}_1\|_2^{16})} + \frac{\text{tr}(A)^8}{\mathbb{E}(\|\mathbf{z}_1\|_2^{16})} \\ &\lesssim \frac{\text{tr}(AA^\top)^4}{p^8} + \frac{\text{tr}(A)^8}{p^8}. \end{aligned}$$

This leads to (A.85), since $\text{tr}(A)^8 \leq p^4 \text{tr}(AA^\top)^4$. □

LEMMA D.5 (Magnus et al. (1978), Lemma 2.2). *If $\mathbf{z} \in \mathbb{R}^p$ is a standard normal vector and $M \in \mathbb{R}^{p \times p}$ is a symmetric matrix, then*

$$\mathbb{E}((\mathbf{z}^\top M \mathbf{z})^2) = 2\text{tr}(M^2) + \text{tr}(M)^2,$$

$$\mathbb{E}((\mathbf{z}^\top M \mathbf{z})^3) = 8\text{tr}(M^3) + 6\text{tr}(M^2)\text{tr}(M) + \text{tr}(M)^3,$$

$$\mathbb{E}((\mathbf{z}^\top M \mathbf{z})^4) = 48\text{tr}(M^4) + 32\text{tr}(M^3)\text{tr}(M) + 12\text{tr}(M^2)^2 + 12\text{tr}(M^2)\text{tr}(M)^2 + \text{tr}(M)^4.$$

LEMMA D.6. *If Assumption 1 holds, then for any $j, k \in \{1, \dots, p\}$*

$$\text{cov}(x_{1j}^4, x_{1k}^4) = r_4(72\Sigma_{jj}\Sigma_{kk}\Sigma_{jk}^2 + 24\Sigma_{jk}^4) + 9(r_4 - r_2^2)\Sigma_{jj}^2\Sigma_{kk}^2,$$

$$\text{cov}(x_{1j}^2, x_{1k}^2) = 2r_2\Sigma_{jk}^2 + (r_2 - 1)\Sigma_{jj}\Sigma_{kk},$$

$$\text{cov}(x_{1k}^2, x_{1j}^4) = 12r_3\Sigma_{jj}\Sigma_{jk}^2 + 3(r_3 - r_2)\Sigma_{jj}^2\Sigma_{kk}.$$

PROOF. Recall the relation $\mathbf{x}_1 = (\xi_1/\|\mathbf{z}_1\|_2)\boldsymbol{\zeta}_1$ from (A.2), where $\boldsymbol{\zeta}_1 \sim N(0, \Sigma)$. Isserlis' theorem (as presented in (A.36)) ensures that for any $j, k \in \{1, \dots, p\}$ we have

$$\begin{aligned} \text{cov}(x_{1j}^4, x_{1k}^4) &= \frac{\mathbb{E}(\xi_1^8)}{\mathbb{E}(\|\mathbf{z}_1\|_2^8)} \mathbb{E}(\zeta_{1j}^4 \zeta_{1k}^4) - \left(\frac{\mathbb{E}(\xi_1^4)}{\mathbb{E}(\|\mathbf{z}_1\|_2^4)} \right)^2 \mathbb{E}(\zeta_{1j}^4) \mathbb{E}(\zeta_{1k}^4) \\ &= r_4(9\Sigma_{jj}^2\Sigma_{kk}^2 + 72\Sigma_{jj}\Sigma_{kk}\Sigma_{jk}^2 + 24\Sigma_{jk}^4) - 9r_2^2\Sigma_{jj}^2\Sigma_{kk}^2, \end{aligned}$$

$$\begin{aligned} \text{cov}(x_{1j}^2, x_{1k}^2) &= \frac{\mathbb{E}(\xi_1^4)}{\mathbb{E}(\|\mathbf{z}_1\|_2^4)} \mathbb{E}(\zeta_{1j}^2 \zeta_{1k}^2) - \mathbb{E}(\zeta_{1j}^2) \mathbb{E}(\zeta_{1k}^2) \\ &= r_2(\Sigma_{jj}\Sigma_{kk} + 2\Sigma_{jk}^2) - \Sigma_{jj}\Sigma_{kk}, \end{aligned}$$

$$\begin{aligned} \text{cov}(x_{1k}^2, x_{1j}^4) &= \frac{\mathbb{E}(\xi_1^6)}{\mathbb{E}(\|\mathbf{z}_1\|_2^6)} \mathbb{E}(\zeta_{1j}^4 \zeta_{1k}^2) - \frac{\mathbb{E}(\xi_1^4)}{\mathbb{E}(\|\mathbf{z}_1\|_2^4)} \mathbb{E}(\zeta_{1j}^4) \mathbb{E}(\zeta_{1k}^2) \\ &= r_3(3\Sigma_{jj}^2\Sigma_{kk} + 12\Sigma_{jj}\Sigma_{jk}^2) - 3r_2\Sigma_{jj}^2\Sigma_{kk}. \end{aligned}$$

□

LEMMA D.7 (Johnson et al. (1985a), Rosenthal's inequality). *Fix $r > 2$ and let $v_1, \dots, v_n$ be centered independent random variables. Then, there is an absolute constant $c > 0$ such that*

$$\left\| \sum_{i=1}^n v_i \right\|_{L^r} \leq c \cdot r \cdot \max \left\{ \left\| \sum_{i=1}^n v_i \right\|_{L^2}, \left(\sum_{i=1}^n \|v_i\|_{L^r}^r \right)^{1/r} \right\}.$$

□

A.5. Additional numerical experiments

The following is an extension of Table 12.1 from Section 4.12.4.1, which reports the empirical level of the proposed test for larger values of the ratio $p/n \in \{3, 5, 10\}$.

TABLE A.3. Results on empirical level for the proposed test (5% nominal level).

	$p/n = 3$				$p/n = 5$				$p/n = 10$			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
(i)	4.26	4.34	4.00	4.29	3.89	3.84	3.71	3.83	3.89	3.86	3.80	3.81
(ii)	4.02	4.13	3.88	4.17	3.74	3.85	3.87	3.89	3.77	3.73	3.71	3.81
(iii)	4.43	4.51	4.22	4.47	4.12	4.11	3.89	4.06	4.10	4.00	3.92	4.05
(iv)	3.99	4.18	3.95	4.17	3.73	3.80	3.97	3.83	3.62	3.61	3.33	3.71
(v)	4.07	4.23	3.99	4.12	3.78	3.79	3.92	3.87	3.73	3.68	3.66	3.68

APPENDIX B

Supplementary Material for Chapter 3

In Appendices B.1, B.2, and B.3, we give the proofs of Theorems 3.3.1, 3.3.2, and 3.3.3 respectively. Background results are given in Appendix B.4, and additional details related to the examples in Section 3.3.1 are given in Appendix B.5.

B.1. Proof of Theorem 3.3.1

Our consistency guarantees for $\hat{\zeta}_n^2$ and $\tilde{H}_n$ are proven separately in the next two subsections.

B.1.1. Consistency of $\hat{\zeta}_n^2$: Proof of (3.15) in Theorem 3.3.1. Define the parameter

$$(B.1) \quad \tau_n = \frac{(p+2)(\beta_n - 2\alpha_n)}{\gamma_n + 2\alpha_n} + 2,$$

and the estimate

$$(B.2) \quad \hat{\tau}_n = \frac{(p+2)(\hat{\beta}_n - 2\hat{\alpha}_n)}{\hat{\gamma}_n + 2\hat{\alpha}_n} + 2.$$

Based on the definitions of ζ_n^2 and $\hat{\zeta}_n^2$, note that

$$\frac{\zeta_n^2}{p} = \tau_n \quad \text{and} \quad \frac{\hat{\zeta}_n^2}{p} = \max\{\hat{\tau}_n, 0\},$$

as well as the fact that the Assumption 2 implies $\tau_n \rightarrow \tau$ as $n \rightarrow \infty$. Since the function $x \mapsto \max\{x, 0\}$ is 1-Lipschitz and $\tau_n \geq 0$, it suffices to show $\hat{\tau}_n - \tau_n \xrightarrow{\mathbb{P}} 0$.

In Lemmas A.1, A.2, and A.3 given later in this subsection, the following three limits will be established,

$$(B.3) \quad \frac{\hat{\alpha}_n}{\alpha_n} \xrightarrow{\mathbb{P}} 1$$

$$(B.4) \quad \frac{\hat{\gamma}_n}{\gamma_n} \xrightarrow{\mathbb{P}} 1$$

$$(B.5) \quad (p+2) \frac{\hat{\beta}_n - \beta_n}{\gamma_n + 2\alpha_n} \xrightarrow{\mathbb{P}} 0.$$

Due to the ratio-consistency of $\hat{\alpha}_n$ and $\hat{\gamma}_n$, as well as the fact that $\frac{(p+2)\alpha_n}{\gamma_n+2\alpha_n} \asymp 1$ holds under Assumption 3, it is straightforward to check that

$$(B.6) \quad \frac{\gamma_n + 2\alpha_n}{\hat{\gamma}_n + 2\hat{\alpha}_n} \xrightarrow{\mathbb{P}} 1 \quad \text{and} \quad \frac{(p+2)\alpha_n}{\gamma_n + 2\alpha_n} - \frac{(p+2)\hat{\alpha}_n}{\hat{\gamma}_n + 2\hat{\alpha}_n} \xrightarrow{\mathbb{P}} 0.$$

Therefore,

$$(B.7) \quad \begin{aligned} \hat{\tau}_n - \tau_n &= \frac{(p+2)\hat{\beta}_n}{\gamma_n + 2\alpha_n} \left(\frac{\gamma_n + 2\alpha_n}{\hat{\gamma}_n + 2\hat{\alpha}_n} - 1 \right) + (p+2) \frac{\hat{\beta}_n - \beta_n}{\gamma_n + 2\alpha_n} + \left(2 \frac{(p+2)\alpha_n}{\gamma_n + 2\alpha_n} - 2 \frac{(p+2)\hat{\alpha}_n}{\hat{\gamma}_n + 2\hat{\alpha}_n} \right) \\ &= \frac{(p+2)\beta_n}{\gamma_n + 2\alpha_n} o_{\mathbb{P}}(1) + o_{\mathbb{P}}(1), \end{aligned}$$

where we have applied (B.5) twice in the second step. Under Assumption 2, we have $\mathbb{E}(\xi_1^4) = p^2 + \tau p + o(p)$ and so Lemma D.1 gives

$$\begin{aligned} \beta_n &= \frac{\mathbb{E}(\xi_1^4)}{p(p+2)} \left(\text{tr}(\Sigma_n)^2 + 2\text{tr}(\Sigma_n^2) \right) - \text{tr}(\Sigma_n)^2 \\ &= \frac{p^2 + \tau p + o(p)}{p(p+2)} \left(\text{tr}(\Sigma_n)^2 + 2\text{tr}(\Sigma_n^2) \right) - \text{tr}(\Sigma_n)^2 \\ &= \left(\frac{(\tau-2)p + o(p)}{p(p+2)} \right) \text{tr}(\Sigma_n)^2 + 2(1 + o(1)) \text{tr}(\Sigma_n^2) \\ &\lesssim p. \end{aligned}$$

Consequently, we have $\frac{(p+2)\beta_n}{\gamma_n + 2\alpha_n} \lesssim 1$, and applying this to (B.7) completes the proof of (3.15) in Theorem 3.3.1. $\square$

LEMMA A.1. *If Assumptions 2 and 3 hold, then as $n \rightarrow \infty$*

$$\frac{\hat{\alpha}_n}{\alpha_n} \xrightarrow{\mathbb{P}} 1.$$

PROOF. Recall that $\alpha_n = \text{tr}(\Sigma_n^2)$ and that $\hat{\alpha}_n = \text{tr}(\hat{\Sigma}_n^2) - \frac{1}{n}\text{tr}(\hat{\Sigma}_n)^2$. The two terms in the estimate can be expanded as

$$\begin{aligned} \text{tr}(\hat{\Sigma}_n^2) &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (\mathbf{x}_i^\top \mathbf{x}_j)^2 \\ \frac{1}{n} \text{tr}(\hat{\Sigma}_n)^2 &= \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n (\mathbf{x}_i^\top \mathbf{x}_i)(\mathbf{x}_j^\top \mathbf{x}_j), \end{aligned} \quad (\text{B.8})$$

which leads to the algebraic relation

$$\begin{aligned} \frac{\hat{\alpha}_n - \alpha_n}{\alpha_n} &= \left(\frac{1}{\alpha_n n^2} \sum_{i \neq j} (\mathbf{x}_i^\top \mathbf{x}_j)^2 - 1 \right) + \frac{1}{\alpha_n n^3} \sum_{i \neq j} \left((\mathbf{x}_i^\top \mathbf{x}_i)^2 - (\mathbf{x}_i^\top \mathbf{x}_i)(\mathbf{x}_j^\top \mathbf{x}_j) \right) \\ &=: A_n + B_n. \end{aligned} \quad (\text{B.9})$$

In the remainder of the proof, we will show that A_n and B_n are both $o_{\mathbb{P}}(1)$. We begin with the analysis of B_n , since it is simpler. Note that B_n is always non-negative, since it can be rewritten as

$$B_n = \frac{1}{\alpha_n n^3} \sum_{i > j} (\mathbf{x}_i^\top \mathbf{x}_i - \mathbf{x}_j^\top \mathbf{x}_j)^2, \quad (\text{B.10})$$

and so it suffices to show that $\mathbb{E}(B_n) = o(1)$. Furthermore, the expectation of B_n can be computed directly as

$$\begin{aligned} \mathbb{E}(B_n) &= \frac{1}{\alpha_n n^3} \sum_{i > j} 2 \text{var}(\mathbf{x}_i^\top \mathbf{x}_i) \\ &= \frac{n(n-1)}{\alpha_n n^3} \text{var}(\mathbf{x}_1^\top \mathbf{x}_1) \\ &= \frac{n(n-1)}{\alpha_n n^3} \left[\left(\frac{\mathbb{E}(\xi_1^4)}{p(p+2)} - 1 \right) \text{tr}(\Sigma_n)^2 + 2 \frac{\mathbb{E}(\xi_1^4)}{p(p+2)} \text{tr}(\Sigma_n^2) \right] \quad (\text{Lemma D.1}) \\ &= \frac{n(n-1)}{\alpha_n n^3} \left[\mathcal{O}\left(\frac{1}{p}\right) \text{tr}(\Sigma_n)^2 + 2 \left(1 + \mathcal{O}\left(\frac{1}{p}\right) \right) \text{tr}(\Sigma_n^2) \right] \quad (\text{Assumption 2}) \\ &\lesssim \frac{1}{n}, \end{aligned} \quad (\text{B.11})$$

where the last step uses $\frac{\text{tr}(\Sigma_n)^2}{\alpha_n} \leq p$. Thus, $\mathbb{E}(B_n) = o(1)$.

Now we handle the term A_n by showing that $\mathbb{E}(A_n^2) = o(1)$. It is helpful to start by noting that if $i \neq j$, then

$$(B.12) \quad \mathbb{E}((\mathbf{x}_i^\top \mathbf{x}_j)^2) = \text{tr}(\Sigma_n^2),$$

which can be checked by a direct calculation. (See the calculation in (B.83) for additional details.) Consequently, if we expand out the square A_n^2 and then take the expectation, it follows that

$$(B.13) \quad \mathbb{E}(A_n^2) = \frac{1}{\alpha_n^2 n^4} \mathbb{E} \left[\left(\sum_{i \neq j} (\mathbf{x}_i^\top \mathbf{x}_j)^2 \right)^2 \right] - \frac{2n(n-1)}{n^2} + 1.$$

Next, we compute the second moment on the right as

$$(B.14) \quad \mathbb{E} \left[\left(\sum_{i \neq j} (\mathbf{x}_i^\top \mathbf{x}_j)^2 \right)^2 \right] = \sum_{i \neq j} \sum_{l \neq k} \mathbb{E} \left[(\mathbf{x}_i^\top \mathbf{x}_j)^2 (\mathbf{x}_l^\top \mathbf{x}_k)^2 \right].$$

In Lemma D.2, it is shown that if (i, j, k, l) are four distinct indices, then

$$(B.15) \quad \mathbb{E}((\mathbf{x}_i^\top \mathbf{x}_j)^4) = 3 \left(\frac{\mathbb{E}(\xi_1^4)}{p(p+2)} \right)^2 (\alpha_n^2 + 2\text{tr}(\Sigma_n^4))$$

$$(B.16) \quad \mathbb{E}((\mathbf{x}_i^\top \mathbf{x}_j)^2 (\mathbf{x}_i^\top \mathbf{x}_k)^2) = \frac{\mathbb{E}(\xi_1^4)}{p(p+2)} (\alpha_n^2 + 2\text{tr}(\Sigma_n^4))$$

$$(B.17) \quad \mathbb{E}((\mathbf{x}_i^\top \mathbf{x}_j)^2 (\mathbf{x}_l^\top \mathbf{x}_k)^2) = \alpha_n^2.$$

Note that in the double sum (B.14), the numbers of terms involving 2, 3, and 4 distinct indices are respectively $\mathcal{O}(n^2)$, $\mathcal{O}(n^3)$ and $n^4 + \mathcal{O}(n^3)$. Applying these observations to (B.14), we have

$$\begin{aligned} \frac{1}{\alpha_n^2 n^4} \mathbb{E} \left[\left(\sum_{i \neq j} (\mathbf{x}_i^\top \mathbf{x}_j)^2 \right)^2 \right] &= \frac{\alpha_n^2}{\alpha_n^2 n^4} (n^4 + \mathcal{O}(n^3)) \\ &\quad + \frac{1}{\alpha_n^2 n^4} \left(\frac{\mathbb{E}(\xi_1^4)}{p(p+2)} (\alpha_n^2 + 2\text{tr}(\Sigma_n^4)) \right) \mathcal{O}(n^3) \\ &\quad + \frac{1}{\alpha_n^2 n^4} \left(3 \left(\frac{\mathbb{E}(\xi_1^4)}{p(p+2)} \right)^2 (\alpha_n^2 + 2\text{tr}(\Sigma_n^4)) \right) \mathcal{O}(n^2) \\ &= 1 + \mathcal{O}\left(\frac{1}{n}\right). \end{aligned}$$

Combining this with (B.13), we reach the following bound on $\mathbb{E}(A_n^2)$,

$$(B.18) \quad \begin{aligned} \mathbb{E}(A_n^2) &= 1 + \mathcal{O}\left(\frac{1}{n}\right) - \frac{2n(n-1)}{n^2} + 1 \\ &\lesssim \frac{1}{n}, \end{aligned}$$

which completes the proof. $\square$

LEMMA A.2. *If Assumptions 2 and 3 hold, then as $n \rightarrow \infty$*

$$\frac{\hat{\gamma}_n}{\gamma_n} \xrightarrow{\mathbb{P}} 1.$$

PROOF. Recall $\gamma_n = \text{tr}(\Sigma_n)^2$ and $\hat{\gamma}_n = \text{tr}(\hat{\Sigma}_n)^2$, and note that the algebraic identity $a^2 - b^2 = 2b(a - b) + (a - b)^2$ gives

$$\begin{aligned} \mathbb{E}|\hat{\gamma}_n - \gamma_n| &= \mathbb{E} \left| 2\text{tr}(\Sigma_n)(\text{tr}(\hat{\Sigma}_n) - \text{tr}(\Sigma_n)) + (\text{tr}(\hat{\Sigma}_n) - \text{tr}(\Sigma_n))^2 \right| \\ &\leq 2\text{tr}(\Sigma_n) (\text{var}(\text{tr}(\hat{\Sigma}_n)))^{1/2} + \text{var}(\text{tr}(\hat{\Sigma}_n)). \end{aligned}$$

Next, observe that $\text{var}(\text{tr}(\hat{\Sigma}_n)) = \frac{1}{n} \text{var}(\mathbf{x}_1^\top \mathbf{x}_1)$, and that the calculation in (B.11) shows $\text{var}(\mathbf{x}_1^\top \mathbf{x}_1) \lesssim p$. Combining this with the fact that $\gamma_n \asymp p^2$ under Assumption 3, we have

$$(B.19) \quad \frac{\mathbb{E}|\hat{\gamma}_n - \gamma_n|}{\gamma_n} \lesssim \frac{\text{tr}(\Sigma_n) + 1}{p^2}$$

which leads to the stated result. $\square$

LEMMA A.3. *If Assumptions 2 and 3 hold, then as $n \rightarrow \infty$,*

$$(B.20) \quad (p+2) \frac{\hat{\beta}_n - \beta_n}{\gamma_n + 2\alpha_n} \xrightarrow{\mathbb{P}} 0.$$

PROOF. Recall that $\beta_n = \text{var}(\|\mathbf{x}_1\|_2^2)$ and $\hat{\beta}_n = \frac{1}{n-1} \sum_{i=1}^n \left(\|\mathbf{x}_i\|_2^2 - \frac{1}{n} \sum_{i'=1}^n \|\mathbf{x}_{i'}\|_2^2 \right)^2$. It is clear that $\mathbb{E}(\hat{\beta}_n) = \beta_n$, and so it suffices to show

$$\frac{(p+2)^2}{(\gamma_n + 2\alpha_n)^2} \text{var}(\hat{\beta}_n) = o(1).$$

Since Assumption 3 implies $\gamma_n + 2\alpha_n \asymp p^2$, it remains to show $\text{var}(\hat{\beta}_n) = o(p^2)$.

Making use of a standard bound for the variance of a sample variance, we have

$$(B.21) \quad \text{var}(\hat{\beta}_n) \lesssim \frac{1}{n} \mathbb{E} \left| \|\mathbf{x}_1\|_2^2 - \text{tr}(\Sigma_n) \right|^4.$$

To bound the right side of (B.21), observe that

$$\begin{aligned} \mathbb{E} \left| \|\mathbf{x}_1\|_2^2 - \text{tr}(\Sigma_n) \right|^4 &= \mathbb{E} \left| \xi_1^2 \left(\mathbf{u}_1^\top \Sigma_n \mathbf{u}_1 - \frac{1}{p} \text{tr}(\Sigma_n) \right) + \left(\frac{1}{p} \xi_1^2 - 1 \right) \text{tr}(\Sigma_n) \right|^4 \\ &\lesssim \mathbb{E}(\xi_1^8) \cdot \mathbb{E} \left| \mathbf{u}_1^\top \Sigma_n \mathbf{u}_1 - \frac{1}{p} \text{tr}(\Sigma_n) \right|^4 + \text{tr}(\Sigma_n)^4 \cdot \mathbb{E} \left| \frac{1}{p} \xi_1^2 - 1 \right|^4. \end{aligned}$$

Since $\lambda_1(\Sigma_n) \lesssim 1$ holds under Assumption 3, it follows from Lemma D.3 that

$$\mathbb{E} \left| \mathbf{u}_1^\top \Sigma_n \mathbf{u}_1 - \frac{1}{p} \text{tr}(\Sigma_n) \right|^4 \lesssim \frac{1}{p^2}.$$

Also, Assumption 2 implies

$$(B.22) \quad \mathbb{E} \left| \frac{1}{p} \xi_1^2 - 1 \right|^4 \lesssim \frac{1}{p^2} \quad \text{and} \quad \mathbb{E}(\xi_1^8) \lesssim p^4.$$

Applying the last several observations to (B.21) implies $\text{var}(\hat{\beta}_n) \lesssim p$, which yields the stated result. $\square$

B.1.2. Consistency of $\tilde{H}_n$: Proof of (3.16) in Theorem 3.3.1. For each $t \in \mathbb{R}$, denote the empirical distribution function of the QuEST eigenvalue estimates as

$$\hat{H}_{Q,n}(t) = \frac{1}{p} \sum_{j=1}^p 1\{\hat{\lambda}_{Q,j} \leq t\}.$$

It follows from Lemma A.4 below that the limit

$$(B.23) \quad \hat{H}_{Q,n} \Rightarrow H$$

holds almost surely as $n \rightarrow \infty$. So, to prove (3.16), it is sufficient to show $\sup_{t \in \mathbb{R}} |\hat{H}_{Q,n}(t) - \tilde{H}_n(t)| \xrightarrow{\mathbb{P}} 0$.

Since $\hat{\lambda}_{Q,j}$ and $\tilde{\lambda}_j$ can only disagree when $\hat{\lambda}_{Q,j} > \hat{b}_n$, we have

$$\begin{aligned} \sup_{t \in \mathbb{R}} |\hat{H}_{Q,n}(t) - \tilde{H}_n(t)| &\leq \frac{1}{p} \sum_{j=1}^p 1\{\hat{\lambda}_{Q,j} \neq \tilde{\lambda}_j\} \\ (B.24) \quad &= \frac{1}{p} \sum_{j=1}^p 1\{\hat{\lambda}_{Q,j} > \hat{b}_n\} \\ &= 1 - \hat{H}_{Q,n}(\hat{b}_n). \end{aligned}$$

Let u denote the upper endpoint of the support of the distribution associated with H , and fix any $\epsilon \in (0, \frac{1}{4})$. Since $\hat{H}_{Q,n}$ is a non-decreasing function, we have

$$(B.25) \quad \sup_{t \in \mathbb{R}} |\hat{H}_{Q,n} - \tilde{H}_n(t)| \leq 1 - \hat{H}_{Q,n}(u + \epsilon) 1\{\hat{b}_n \geq u + \epsilon\}.$$

By the definition of u , the value $u + \epsilon$ is a continuity point of H with $H(u + \epsilon) = 1$, and so the limit (B.23) implies $\hat{H}_{Q,n}(u + \epsilon) = 1 + o_{\mathbb{P}}(1)$. Hence, in order to show that the right side of (B.25) converges to 0 in probability, it is enough to show that $\mathbb{P}(\hat{b}_n \geq u + \epsilon) \rightarrow 1$.

We will handle this remaining task by showing instead that $\mathbb{P}(\hat{b}_n < u + \epsilon) \rightarrow 0$. Recall that $\hat{b}_n = \lambda_1(\hat{\Sigma}_n) + 1$ and note that the limit $H_n \Rightarrow H$ implies that $\lambda_1(\Sigma_n) \geq u - \epsilon$ must hold for all large n . Therefore, we have

$$(B.26) \quad \mathbb{P}(\hat{b}_n < u + \epsilon) \leq \mathbb{P}(\lambda_1(\hat{\Sigma}_n) \leq \lambda_1(\Sigma_n) + 2\epsilon - 1)$$

for all large n . To derive an upper bound on the last probability, we will replace $\lambda_1(\hat{\Sigma}_n)$ with a smaller random variable, and then rearrange the event. Let $\mathbf{v}_1$ denote an eigenvector of Σ_n corresponding to $\lambda_1(\Sigma_n)$ with $\|\mathbf{v}_1\|_2 = 1$. Defining the random matrix $W_n = \frac{1}{n} \sum_{i=1}^n \xi_i^2 \mathbf{u}_i \mathbf{u}_i^\top$, the variational representation of $\lambda_1(\hat{\Sigma}_n)$ gives $\lambda_1(\hat{\Sigma}_n) \geq \lambda_1(\Sigma_n) \mathbf{v}_1^\top W_n \mathbf{v}_1$. Also note that our choice of ϵ ensures $2\epsilon - 1 \leq -1/2$. This yields the following bounds for all large n ,

$$\begin{aligned} \mathbb{P}(\hat{b}_n < u + \epsilon) &\leq \mathbb{P}(\lambda_1(\Sigma_n) |\mathbf{v}_1^\top W_n \mathbf{v}_1 - 1| \geq 1/2) \\ &\leq 4\lambda_1(\Sigma_n)^2 \text{var}(\mathbf{v}_1^\top W_n \mathbf{v}_1) \\ (B.27) \quad &= \frac{4\lambda_1(\Sigma_n)^2}{n} \text{var}(\xi_1^2 \mathbf{u}_1^\top (\mathbf{v}_1 \mathbf{v}_1^\top) \mathbf{u}_1) \\ &= \frac{4\lambda_1(\Sigma_n)^2}{n} \left(3 \frac{\mathbb{E}(\xi_1^4)}{p(p+2)} - 1 \right) \quad (\text{Lemma D.1}) \\ &\lesssim \frac{1}{n}, \end{aligned}$$

where the last step uses Assumptions 2 and 3. □

LEMMA A.4. *Suppose that Assumptions 2 and 3 hold. Then, the following limit holds almost surely as $n \rightarrow \infty$*

$$(B.28) \quad \hat{H}_{Q,n} \Rightarrow H.$$

PROOF. In the paper (Ledoit and Wolf, 2015, pp.381-382), the limit (B.28) is established in the context of an IC model where $p/n \rightarrow c \in (0, \infty) \setminus \{1\}$, and the eigenvalues of the population covariance matrix Σ_n satisfy Assumption 3. To adapt the proof from (Ledoit and Wolf, 2015) to our current setting, it is sufficient to show that the following two facts hold under Assumptions 2 and 3. First, there is a constant $C > 0$ such that the bound $\limsup_{n \rightarrow \infty} \lambda_1(\widehat{\Sigma}_n) \leq C$ holds almost surely, which we prove later using a truncation argument in Lemmas A.5 and A.6. Second, the random distribution function $\widehat{H}_n(t) = \frac{1}{p} \sum_{j=1}^p 1\{\lambda_j(\widehat{\Sigma}_n) \leq t\}$ satisfies $\widehat{H}_n \Rightarrow \Psi(H, c)$ almost surely, where we recall that the distribution $\Psi(H, c)$ is defined near equation (3.17). The validity of this second fact under our current assumptions is a consequence of Theorem 1.1 in (Bai and Zhou, 2008). $\square$

B.1.3. Boundedness of sample eigenvalues.

LEMMA A.5. *Under Assumptions 2 and 3, there is a constant $C > 0$ such that the bounds*

$$(B.29) \quad \limsup_{n \rightarrow \infty} \lambda_1(\widehat{\Sigma}_n) \leq C \quad \text{and} \quad \limsup_{n \rightarrow \infty} \lambda_1(\check{\Sigma}_n) \leq C$$

hold almost surely.

PROOF. Since the estimates $\check{\lambda}_j = \lambda_j(\check{\Sigma}_n)$ are bounded above by $\widehat{b}_n = \lambda_1(\widehat{\Sigma}_n) + 1$ for all $j \in \{1, \dots, p\}$, it is enough to focus on the first inequality in (B.29). Define a sequence of truncated random variables $\check{\xi}_i = \xi_i 1\{|\xi_i^2 - p| < \sqrt{pn}\}$ for $i = 1, \dots, n$, as well as the following truncated version of $\widehat{\Sigma}_n$,

$$(B.30) \quad \check{\Sigma}_n = \frac{1}{n} \sum_{i=1}^n \check{\xi}_i^2 \Sigma_n^{1/2} \mathbf{u}_i \mathbf{u}_i^\top \Sigma_n^{1/2}.$$

Lemma A.6 below shows that $\mathbb{P}(\widehat{\Sigma}_n \neq \check{\Sigma}_n \text{ i.o.}) = 0$. Consequently, it suffices to show there is a constant $C > 0$ such that $\limsup_{n \rightarrow \infty} \lambda_1(\check{\Sigma}_n) \leq C$ holds almost surely.

Since the vectors $\mathbf{u}_1, \dots, \mathbf{u}_n$ are uniformly distributed on the unit sphere of $\mathbb{R}^p$, we may express them as $\mathbf{u}_i = \mathbf{z}_i / \|\mathbf{z}_i\|_2$ for a sequence of i.i.d. standard Gaussian vectors $\mathbf{z}_1, \dots, \mathbf{z}_n$ in $\mathbb{R}^p$. This yields

$$(B.31) \quad \check{\Sigma}_n = \frac{1}{n} \sum_{i=1}^n \frac{\check{\xi}_i^2/p}{\|\mathbf{z}_i\|_2^2/p} \Sigma_n^{1/2} \mathbf{z}_i \mathbf{z}_i^\top \Sigma_n^{1/2}.$$

By construction, we have $\max_{1 \leq i \leq n} \check{\xi}_i^2/p \leq 1 + 1/\sqrt{c_n}$ for all $n \geq 1$. Also, using standard tail bounds for the χ_p^2 distribution and the Borel-Cantelli lemma, it is straightforward to show that $\liminf_{n \rightarrow \infty} \min_{1 \leq i \leq n} \|\mathbf{z}_i\|_2^2/p$ is at least $1/2$ almost surely. Taken together, these observations imply there is a constant $C > 0$ such that the bound

$$(B.32) \quad \lambda_1(\check{\Sigma}_n) \leq C \lambda_1(\Sigma_n) \lambda_1\left(\frac{1}{n} \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top\right)$$

holds almost surely for all large n . In addition, note that $\sup_{n \geq 1} \lambda_1(\Sigma_n) \lesssim 1$ holds by Assumption 3. Lastly, it is known from (Yin et al., 1988, Theorem 3.1) that the limit

$$\lim_{n \rightarrow \infty} \lambda_1\left(\frac{1}{n} \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top\right) = (1 + \sqrt{c})^2$$

holds almost surely, which completes the proof. $\square$

LEMMA A.6. *Suppose that Assumption 2 holds, and let $\check{\Sigma}_n$ be as defined in (B.30). Then,*

$$(B.33) \quad \mathbb{P}(\hat{\Sigma}_n \neq \check{\Sigma}_n \text{ i.o.}) = 0.$$

PROOF. We adapt a classical argument from (Yin et al., 1988). For a fixed number n , the matrices $\hat{\Sigma}_n$ and $\check{\Sigma}_n$ can only disagree if $|\xi_i^2 - p| \geq \sqrt{pn}$ for at least one $i = 1, \dots, n$, and so

$$(B.34) \quad \mathbb{P}(\hat{\Sigma}_n \neq \check{\Sigma}_n \text{ i.o.}) \leq \lim_{j \rightarrow \infty} \mathbb{P}\left(\bigcup_{n=j}^{\infty} \bigcup_{i=1}^n \{|\xi_i^2 - p| \geq \sqrt{pn}\}\right).$$

Next, we partition the values of n into the intervals $[2^{m-1}, 2^m), [2^m, 2^{m+1}), \dots$ and take a union bound across the intervals, yielding

$$(B.35) \quad \mathbb{P}(\hat{\Sigma}_n \neq \check{\Sigma}_n \text{ i.o.}) \leq \lim_{j \rightarrow \infty} \sum_{m=j}^{\infty} \mathbb{P}\left(\max_{\substack{1 \leq i \leq n \\ 2^{m-1} \leq n < 2^m}} \left|\frac{\xi_i^2 - p}{\sqrt{p}}\right| \geq \sqrt{2^{m-1}}\right).$$

For a generic sequence of random variables $Y_1, \dots, Y_N$ and number $q \geq 1$, recall the standard maximal inequality

$$(B.36) \quad \mathbb{E}\left|\max_{1 \leq i \leq N} Y_i\right|^q \leq N \max_{1 \leq i \leq N} \mathbb{E}|Y_i|^q.$$

Since the number of pairs (i, n) in the maximum in (B.35) is at most 2^{2m} , if we apply Chebyshev's inequality to (B.35) and use the condition $\mathbb{E}\left|\frac{\xi_1^2 - p}{\sqrt{p}}\right|^{4+\varepsilon} \lesssim 1$ from Assumption 2, then for each $m \geq 1$

we have

$$(B.37) \quad \mathbb{P} \left(\max_{\substack{1 \leq i \leq n \\ 2^{m-1} \leq n \leq 2^m}} \left| \frac{\xi_i^2 - p}{\sqrt{p}} \right| \geq \sqrt{2^{m-1}} \right) \lesssim \frac{2^{2m}}{(\sqrt{2^{m-1}})^{4+\varepsilon}} \\ \lesssim 2^{-m\varepsilon/2}.$$

Hence, we may insert this bound into (B.35) to conclude that

$$\mathbb{P}(\widehat{\Sigma}_n \neq \check{\Sigma}_n \text{ i.o.}) \lesssim \lim_{j \rightarrow \infty} \sum_{m=j}^{\infty} 2^{-m\varepsilon/2} = 0,$$

which completes the proof. $\square$

B.2. Proof of Theorem 3.3.2

Let $\mathbf{z} \in \mathbb{R}^k$ denote a Gaussian random vector to be described in a moment, and consider the triangle inequality

$$(B.38) \quad d_{\text{LP}} \left(\mathcal{L}(p\{T_n(\mathbf{f}) - \vartheta_n(\mathbf{f})\}), \mathcal{L}(p\{T_{n,1}^*(\mathbf{f}) - \tilde{\vartheta}_n(\mathbf{f})\} | X) \right) \leq \text{I}_n + \text{II}_n,$$

where we define

$$(B.39) \quad \text{I}_n = d_{\text{LP}} \left(\mathcal{L}(p\{T_n(\mathbf{f}) - \vartheta_n(\mathbf{f})\}), \mathcal{L}(\mathbf{z}) \right)$$

$$(B.40) \quad \text{II}_n = d_{\text{LP}} \left(\mathcal{L}(\mathbf{z}), \mathcal{L}(p\{T_{n,1}^*(\mathbf{f}) - \tilde{\vartheta}_n(\mathbf{f})\} | X) \right).$$

To handle the terms I_n and II_n , we will apply a central limit theorem for linear spectral statistics established in (Hu et al., 2019a), which relies on the following two conditions when $c_n \rightarrow c > 0$ as $n \rightarrow \infty$:

(a). The elliptical model in Assumption 2 holds with the conditions on ξ_1^2 being replaced by

$$(B.41) \quad \mathbb{E}(\xi_1^2) = p, \quad \text{var} \left(\frac{\xi_1^2 - p}{\sqrt{p}} \right) = \tau + o(1) \quad \text{and} \quad \mathbb{E} \left| \frac{\xi_1^2 - p}{\sqrt{p}} \right|^{2+\varepsilon} \lesssim 1,$$

for some constants $\tau \geq 0$ and $\varepsilon > 0$ that do not depend on n .

(b). There is a distribution H such that $H_n \Rightarrow H$ as $n \rightarrow \infty$, and $\lambda_1(\Sigma_n) \lesssim 1$.

Under these conditions, Theorem 2.2 in (Hu et al., 2019a) ensures there exists a Gaussian distribution $\mathcal{L}(\mathbf{z})$ depending only on $(\mathbf{f}, H, c, \tau)$ such that $\mathbf{I}_n \rightarrow 0$ as $n \rightarrow \infty$.

To finish the proof, we must show $\mathbf{II}_n \xrightarrow{\mathbb{P}} 0$. This can be done using the mentioned central limit theorem, but some extra considerations are involved, due to the fact that $\mathbf{II}_n$ is random. It is sufficient to show that for any subsequence $J \subset \{1, 2, \dots\}$, the limit $\mathbf{II}_n \rightarrow 0$ holds almost surely along a further subsequence of J . Since the bootstrap data are generated from an elliptical model that is parameterized in terms of $\mathcal{L}((\xi_1^*)^2|X)$ and $\tilde{\Sigma}_n$, this amounts to verifying that $\mathcal{L}((\xi_1^*)^2|X)$ and $\tilde{H}_n$ satisfy analogues of (a) and (b) almost surely along a subsequence of J .

To proceed, recall that Algorithm 1 is designed so that $\mathbb{E}((\xi_1^*)^2|X) = p$ and $\text{var}((\xi_1^*)^2|X) = \zeta_n^2$. Also, note that under Assumption 2, the parameter $\zeta_n^2 = \text{var}(\xi_1^2)$ satisfies $\zeta_n^2/p \rightarrow \tau$. Consequently, Theorem 3.3.1 implies $\hat{\zeta}_n^2/p = \tau + o_{\mathbb{P}}(1)$, and so there is a subsequence $J' \subset J$ along which the limit $\hat{\zeta}_n^2/p \rightarrow \tau$ holds almost surely. In other words, the limit

$$(B.42) \quad \text{var}\left(\frac{(\xi_1^*)^2 - p}{\sqrt{p}} \middle| X\right) \rightarrow \tau$$

holds almost surely along J' . Moreover, since $\mathcal{L}((\xi_1^*)^2|X)$ is a Gamma distribution with mean p and variance $\hat{\zeta}_n^2$, Lemma E.4 implies there is a constant $C > 0$ not depending on n such that the bound

$$(B.43) \quad \left(\mathbb{E}\left(\left|\frac{(\xi_1^*)^2 - p}{\sqrt{p}}\right|^6 \middle| X\right)\right)^{1/6} \leq C\left(\frac{\hat{\zeta}_n}{\sqrt{p}} + \frac{\hat{\zeta}_n^2}{p^{3/2}}\right)$$

holds almost surely. Therefore, the left side of (B.43) is bounded almost surely along J' .

With regard to the empirical spectral distribution $\tilde{H}_n$ associated with the matrix $\tilde{\Sigma}_n$, it satisfies the limit $\tilde{H}_n \Rightarrow H$ almost surely along a further subsequence $J'' \subset J'$, due to Theorem 3.3.1. In addition, Lemma A.5 ensures there is a constant $C > 0$ such that $\limsup_{n \rightarrow \infty} \lambda_1(\tilde{\Sigma}_n) \leq C$ almost surely. Altogether, it follows that $\mathcal{L}((\xi_1^*)^2|X)$ and $\tilde{H}_n$ simultaneously satisfy analogues of (a) and (b) almost surely along J'' , which implies $\mathbf{II}_n \xrightarrow{\mathbb{P}} 0$. $\square$

B.3. Proof of Theorem 3.3.3

To begin the proof, we need to introduce three auxiliary statistics defined by

$$(B.44) \quad \check{r}_n = \frac{\text{tr}(\widehat{\Sigma}_n)^2}{\text{tr}(\widehat{\Sigma}_n^2) - \frac{1}{n}\text{tr}(\widehat{\Sigma}_n)^2}$$

$$(B.45) \quad \check{r}_n^* = \frac{\text{tr}(\widehat{\Sigma}_n^*)^2}{\text{tr}((\widehat{\Sigma}_n^*)^2) - \frac{1}{n}\text{tr}(\widehat{\Sigma}_n^*)^2}$$

$$(B.46) \quad \tilde{r}_n = \frac{\text{tr}(\tilde{\Sigma}_n)^2}{\text{tr}(\tilde{\Sigma}_n^2)}.$$

Also, we define $\widehat{r}_n^*$ as the statistic obtained by applying the formula (3.21) for $\widehat{r}_n$ to the bootstrap data $\mathbf{x}_1^*, \dots, \mathbf{x}_n^*$. The primary task is to establish the following four limits, which are established later in this appendix in Propositions B.3.1 and B.3.2. Specifically, these results show that there exists a non-degenerate Gaussian random variable ζ and a constant a , such that as $n \rightarrow \infty$,

$$(B.47) \quad \mathcal{L}(\check{r}_n - r_n) \Rightarrow \mathcal{L}(\zeta),$$

$$(B.48) \quad \mathcal{L}(\check{r}_n^* - \tilde{r}_n | X) \xRightarrow{\mathbb{P}} \mathcal{L}(\zeta),$$

$$(B.49) \quad \mathcal{L}(\widehat{r}_n - \check{r}_n) \Rightarrow \mathcal{L}(a),$$

$$(B.50) \quad \mathcal{L}(\widehat{r}_n^* - \check{r}_n^* | X) \xRightarrow{\mathbb{P}} \mathcal{L}(a),$$

where $\mathcal{L}(a)$ denotes the point mass distribution at a . (In the current section, we sometimes use the notation for weak convergence in limits where convergence in probability holds, because it will help to clarify how $\widehat{r}_n^* - \check{r}_n^*$ can be analyzed in an analogous manner to $\widehat{r}_n - \check{r}_n$.) Using Slutsky's lemma, it follows from the limits (B.47) and (B.49) that

$$(B.51) \quad \mathcal{L}(\widehat{r}_n - r_n) \Rightarrow \mathcal{L}(\zeta + a).$$

Analogously, Slutsky's lemma can be applied in a conditional manner to the limits (B.48) and (B.50), yielding

$$(B.52) \quad \mathcal{L}(\widehat{r}_n^* - \tilde{r}_n | X) \xRightarrow{\mathbb{P}} \mathcal{L}(\zeta + a).$$

Since the limiting distribution $\mathcal{L}(\zeta + a)$ is continuous, Pólya's theorem implies

$$(B.53) \quad \sup_{t \in \mathbb{R}} \left| \mathbb{P}(\hat{r}_n - r_n \leq t) - \mathbb{P}(\hat{r}_n^* - \tilde{r}_n \leq t | X) \right| \xrightarrow{\mathbb{P}} 0.$$

Due to this uniform limit and the continuity of the distribution $\mathcal{L}(\zeta + a)$, standard arguments can be used to show that the quantiles of $\mathcal{L}(\hat{r}_n^* - \tilde{r}_n | X)$ are asymptotically equivalent to those of $\mathcal{L}(\hat{r}_n - r_n)$. (For example, see the proof of Lemma 10.4 in (Lopes, 2022a).) More precisely, if we note that the quantile estimate $\hat{q}_{1-\alpha}$ defined in Section 3.3.4 is the same as the $(1 - \alpha)$ -quantile of $\mathcal{L}(\frac{1}{p}(\hat{r}_n^* - \tilde{r}_n) | X)$, then as $n \rightarrow \infty$ we have the limit

$$(B.54) \quad \mathbb{P}\left(\frac{1}{p}(\hat{r}_n - r_n) > \hat{q}_{1-\alpha}\right) \rightarrow \alpha.$$

This limit directly implies the first two statements (3.27) and (3.28) in Theorem 3.3.3. Regarding the third statement (3.29), if $H'_{0,n}$ holds for all large n , then replacing $\tilde{\Sigma}_n = \text{diag}(\tilde{\lambda}_1, \dots, \tilde{\lambda}_p)$ with $\tilde{\Sigma}_n = I_p$ does not affect the reasoning leading up to (B.54), because this replacement does not affect the proofs of Propositions B.3.1 and B.3.2 given later. Consequently, if $H'_{0,n}$ holds for all large n , then we have $\mathbb{P}(\frac{1}{p}(\hat{r}_n - 1) \leq \tilde{q}'_\alpha) \rightarrow \alpha$ as $n \rightarrow \infty$, which completes the proof. $\square$

PROPOSITION B.3.1. *Under Assumptions 2 and 3, the limits (B.47) and (B.48) hold as $n \rightarrow \infty$.*

PROOF. The proof is based on viewing $\check{r}_n$ as a nonlinear function of $T_n(\mathbf{f})$, where the components of $\mathbf{f} = (f_1, f_2)$ are taken to be $f_1(x) = x$ and $f_2(x) = x^2$. In this case, the centering parameter $\vartheta_n(\mathbf{f})$ defined by (3.18) reduces to

$$(B.55) \quad \vartheta_n(\mathbf{f}) = \left(\frac{1}{p} \text{tr}(\Sigma_n), \frac{1}{p} \text{tr}(\Sigma_n^2) + \frac{1}{np} \text{tr}(\Sigma_n)^2 \right),$$

as recorded in Lemma 2.16 of (Yao et al., 2015). Consequently, by considering the function

$$(B.56) \quad g_n(x_1, x_2) = \frac{x_1^2}{x_2 - c_n x_1^2},$$

we have the key relation

$$(B.57) \quad \check{r}_n - r_n = p \{ g_n(T_n(\mathbf{f})) - g_n(\vartheta_n(\mathbf{f})) \}.$$

We can also develop a corresponding relation for the bootstrap statistic $\check{r}_n^* - \tilde{r}$. Due to Lemma 2.16 in (Yao et al., 2015) and our choice of $\mathbf{f}$, the definition of $\tilde{\vartheta}_n(\mathbf{f})$ below (3.18) implies

$$(B.58) \quad \tilde{\vartheta}_n(\mathbf{f}) = \left(\frac{1}{p} \text{tr}(\tilde{\Sigma}_n), \frac{1}{p} \text{tr}(\tilde{\Sigma}_n^2) + \frac{1}{np} \text{tr}(\tilde{\Sigma}_n)^2 \right).$$

Likewise, the bootstrap version of (B.57) is

$$(B.59) \quad \check{r}_n^* - \tilde{r}_n = p \{ g_n(T_n^*(\mathbf{f})) - g_n(\tilde{\vartheta}_n(\mathbf{f})) \}.$$

We will proceed by applying the delta method to the relations (B.57) and (B.59). For this purpose, note that the proof of Theorem 3.3.2 shows there exists a Gaussian random vector $\mathbf{z} \in \mathbb{R}^2$ such that as $n \rightarrow \infty$,

$$(B.60) \quad \mathcal{L}(p\{T_n(\mathbf{f}) - \vartheta_n(\mathbf{f})\}) \Rightarrow \mathcal{L}(\mathbf{z})$$

$$(B.61) \quad \mathcal{L}(p\{T_{n,1}^*(\mathbf{f}) - \tilde{\vartheta}_n(\mathbf{f})\} | X) \xRightarrow{\mathbb{P}} \mathcal{L}(\mathbf{z}).$$

Also, to introduce some further notation, we refer to the j th moment of the distribution H as

$$(B.62) \quad \phi_j = \int t^j dH(t),$$

and we use these moments to define the parameter

$$(B.63) \quad \vartheta(\mathbf{f}) = (\phi_1, \phi_2 + c\phi_1^2).$$

This parameter arises in the following limits as $n \rightarrow \infty$,

$$(B.64) \quad \vartheta_n(\mathbf{f}) \rightarrow \vartheta(\mathbf{f}) \quad \text{and} \quad \tilde{\vartheta}_n(\mathbf{f}) \xrightarrow{\mathbb{P}} \vartheta(\mathbf{f}).$$

To see why these limits hold, first note that by Assumption 3, there is a compact interval containing support of H_n for all large n , and by Lemma A.5, the same statement holds almost surely for $\tilde{H}_n$. Moreover, Assumption 3 and Theorem 3.3.1 ensure the limits $H_n \Rightarrow H$ and $\tilde{H}_n \xRightarrow{\mathbb{P}} H$, and so it follows that the moments of H_n converge to those of H , and the moments of $\tilde{H}_n$ converge in probability to those of H . Combining these facts with the formulas (B.55) and (B.58) yields the limits (B.64).

In light of the relations (B.57) and (B.59), and the limits (B.64), we will expand the function g_n around $\vartheta(\mathbf{f})$ and apply the delta method. This is justified because the gradient ∇g_n has the following continuity property. Namely, if we let $g(x_1, x_2) = \frac{x_1^2}{x_2 - cx_1^2}$, let $\mathcal{U}$ be a sufficiently small open neighborhood of $\vartheta(\mathbf{f})$ in $\mathbb{R}^2$, and let $\{(x_{1,n}, x_{2,n})\}$ be any sequence of points within $\mathcal{U}$ that converges to $\vartheta(\mathbf{f})$, then we have the limit

$$(B.65) \quad \nabla g_n(x_{1,n}, x_{2,n}) \rightarrow \nabla g(\vartheta(\mathbf{f})) = \left(\frac{2\phi_1}{\phi_2} + \frac{2c\phi_1^3}{\phi_2^2}, -\frac{\phi_1^2}{\phi_2} \right).$$

So, applying the delta method to the relations (B.57) and (B.59) and the weak limits (B.60) and (B.61) gives

$$(B.66) \quad \mathcal{L}(\check{r}_n - r_n) \Rightarrow \mathcal{L}(\nabla g(\vartheta(\mathbf{f}))^\top \mathbf{z})$$

$$(B.67) \quad \mathcal{L}(\check{r}_n^* - \tilde{r}_n | X) \xRightarrow{\mathbb{P}} \mathcal{L}(\nabla g(\vartheta(\mathbf{f}))^\top \mathbf{z}).$$

Now, it remains to show that the variance of the Gaussian random variable $\nabla g(\vartheta(\mathbf{f}))^\top \mathbf{z}$ is positive. Letting K denote the 2×2 covariance matrix of $\mathbf{z}$, it is shown below equation 2.10 in (Hu et al., 2019a) that the entries of K are given by

$$K_{11} = 2c\phi_2 + c(\tau - 2)\phi_1^2,$$

$$K_{12} = 4c\phi_3 + 4c^2\phi_1\phi_2 + 2c(\tau - 2)\phi_1(c\phi_1^2 + \phi_2)$$

$$K_{22} = 8c\phi_4 + 4c^2\phi_2^2 + 16c^2\phi_1\phi_3 + 8c^3\phi_1^2\phi_2 + 4c(\tau - 2)(c\phi_1^2 + \phi_2)^2.$$

Combining this with the formula for $\nabla g(\vartheta(\mathbf{f}))$ in (B.65), a direct but lengthy computation of the quadratic form $\nabla g(\vartheta(\mathbf{f}))^\top K \nabla g(\vartheta(\mathbf{f}))$ yields

$$(B.68) \quad \text{var}\left(\nabla g(\vartheta(\mathbf{f}))^\top \mathbf{z}\right) = \frac{4c\phi_1^2}{\phi_4^2} \left(c\phi_1^2\phi_2^2 + 2\phi_1^2\phi_4 + 2\phi_2^3 - 4\phi_1\phi_2\phi_3 \right).$$

To see that the variance is positive, it suffices to check that $2\phi_1^2\phi_4 + 2\phi_2^3 - 4\phi_1\phi_2\phi_3$ is non-negative. Rewriting this as $2(\phi_1\phi_4^{1/2} - \phi_2^{3/2})^2 + 4\phi_1\phi_2(\phi_4^{1/2}\phi_2^{1/2} - \phi_3)$, its non-negativity follows from the observation that $\phi_3 \leq \phi_4^{1/2}\phi_2^{1/2}$ is an instance of the Cauchy-Schwarz inequality. $\square$

PROPOSITION B.3.2. *Under Assumptions 2 and 3, the limits (B.49) and (B.50) hold as $n \rightarrow \infty$.*

PROOF. Here, we retain the definitions of f_1 and f_2 used in the proof of the previous proposition. The difference $\hat{r}_n - \check{r}_n$ can be written explicitly in terms of $T_n(f_1)$, $T_n(f_2)$, and $\hat{\Delta}_n$ (defined below (3.21)) as

$$(B.69) \quad \hat{r}_n - \check{r}_n = \frac{T_n(f_1)^2 (\hat{\Delta}_n - p c_n T_n(f_1)^2)}{(T_n(f_2) - c_n T_n(f_1)^2)(T_n(f_2) - \hat{\Delta}_n/p)}.$$

Furthermore, $\hat{\Delta}_n$ can be decomposed as

$$(B.70) \quad \hat{\Delta}_n = \hat{\kappa}_1 T_n(f_1)^2 + \hat{\kappa}_2 T_n(f_2),$$

where the random coefficients $\hat{\kappa}_1$ and $\hat{\kappa}_2$ are defined by

$$(B.71) \quad \hat{\kappa}_1 = c_n p \left[\frac{n+1}{n} + \frac{n-1}{n} \frac{\hat{\zeta}_n^2 - 2p}{p(p+2)} - \frac{2(n-1)}{n^2} \frac{p^2 + \hat{\zeta}_n^2}{p(p+2)} \right]$$

$$(B.72) \quad \hat{\kappa}_2 = c_n \left[\frac{2(n-1)}{n} \frac{p^2 + \hat{\zeta}_n^2}{p(p+2)} - 1 \right].$$

The last few displays show that in order to determine the limit of $\hat{r}_n - \check{r}_n$, it suffices to determine the limit of the triple $(T_n(f_1), T_n(f_2), \hat{\zeta}_n^2/p)$. For the random variables $T_n(f_1)$ and $T_n(f_2)$, we can apply the limits (B.60) and (B.64) as well the formula (B.63) to obtain

$$(B.73) \quad \mathcal{L}(T_n(f_1), T_n(f_2)) \Rightarrow \mathcal{L}(\phi_1, \phi_2 + c\phi_1^2).$$

In addition, the proof of the limit (3.15) in Theorem 3.3.1 shows that

$$(B.74) \quad \mathcal{L}(\frac{1}{p}\hat{\zeta}_n^2) \Rightarrow \mathcal{L}(\tau).$$

Combining the previous two displays with the formulas (B.69)-(B.72), a direct calculation leads to

$$(B.75) \quad \mathcal{L}(\hat{r}_n - \check{r}_n) \Rightarrow \mathcal{L}(a),$$

where

$$a = \frac{c\phi_1^2(\phi_2 + (\tau-2)\phi_1^2)}{\phi_2^2},$$

which proves (B.49).

Now we turn to the proof of (B.50). By analogy with the previous argument that led to (B.75), it is enough to show the following two limits hold as $n \rightarrow \infty$,

$$(B.76) \quad \mathcal{L}(T_n^*(f_1), T_n^*(f_2)|X) \xRightarrow{\mathbb{P}} \mathcal{L}(\phi_1, \phi_2 + c\phi_1^2)$$

$$(B.77) \quad \mathcal{L}(\frac{1}{p}(\hat{\zeta}_n^2)^*|X) \xRightarrow{\mathbb{P}} \mathcal{L}(\tau),$$

where $(\hat{\zeta}_n^2)^*$ is obtained by applying the formula (3.6) to the bootstrap data $\mathbf{x}_1^*, \dots, \mathbf{x}_n^*$. The first limit (B.76) is a consequence of two limits that were established in the proof of Theorem 3.3.2, which are that $\Pi_n \xrightarrow{\mathbb{P}} 0$ and $\tilde{v}_n(\mathbf{f}) \xrightarrow{\mathbb{P}} (\phi_1, \phi_2 + c\phi_1^2)$.

Regarding the second limit (B.77), note that it is equivalent to showing that for any subsequence $J \subset \{1, 2, \dots\}$, there is a further subsequence $J' \subset J$ such that $\mathcal{L}(\frac{1}{p}(\hat{\zeta}_n^2)^*|X) \Rightarrow \mathcal{L}(\tau)$ holds almost surely along J' . The latter statement can be proven by analogy with the limit $\mathcal{L}(\frac{1}{p}\hat{\zeta}_n^2) \Rightarrow \mathcal{L}(\tau)$, which follows from the proof of (3.15) in Theorem 3.3.1. To be more precise, this analogy can be justified as follows: The proof of (3.15) only relies on Assumption 2 and two other conditions, which are

$$(B.78) \quad \lambda_1(\Sigma_n) \lesssim 1, \quad \text{and} \quad (\frac{1}{p}\text{tr}(\Sigma_n), \frac{1}{p}\text{tr}(\Sigma_n^2)) \rightarrow (\phi_1, \phi_2).$$

Consequently, it is enough to check that bootstrap counterparts of these conditions hold almost surely along J' . First, the bootstrap counterpart of Assumption 2 was shown to hold almost surely along subsequences in the proof of Theorem 3.3.2. Second, the bootstrap counterpart of $\lambda_1(\Sigma_n) \lesssim 1$ is implied by Lemma A.5, which guarantees that there is a constant $C > 0$ such that $\limsup_{n \rightarrow \infty} \lambda_1(\tilde{\Sigma}_n) \leq C$ holds almost surely. Lastly, the bootstrap counterpart of the limit $(\frac{1}{p}\text{tr}(\Sigma_n), \frac{1}{p}\text{tr}(\Sigma_n^2)) \rightarrow (\phi_1, \phi_2)$ is handled by the fact that the moments of $\tilde{H}_n$ converge in probability to the moments of H , which was shown in the proof of Proposition B.3.1. This completes the proof. $\square$

B.4. Background results

LEMMA D.1 (Hu et al. (2019a), Lemma A.1). *Let $\xi_1 \in \mathbb{R}$ and $\mathbf{u}_1 \in \mathbb{R}^p$ satisfy the conditions in Assumption 2, and fix any symmetric matrix $M \in \mathbb{R}^{p \times p}$. Then,*

$$(B.79) \quad \text{var}(\xi_1^2 \mathbf{u}_1^\top M \mathbf{u}_1) = \frac{\mathbb{E}(\xi_1^4)}{p(p+2)} (\text{tr}(M)^2 + 2\text{tr}(M^2)) - \text{tr}(M)^2.$$

LEMMA D.2. *Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ satisfy the conditions in Assumption 2, and let i, j, k, l be four distinct indices in $\{1, \dots, n\}$. Then,*

$$(B.80) \quad \mathbb{E}((\mathbf{x}_i^\top \mathbf{x}_j)^4) = 3 \left(\frac{\mathbb{E}(\xi_1^4)}{p(p+2)} \right)^2 (\text{tr}(\Sigma_n^2)^2 + 2\text{tr}(\Sigma_n^4))$$

$$(B.81) \quad \mathbb{E}((\mathbf{x}_i^\top \mathbf{x}_j)^2 (\mathbf{x}_i^\top \mathbf{x}_k)^2) = \frac{\mathbb{E}(\xi_1^4)}{p(p+2)} (\text{tr}(\Sigma_n^2)^2 + 2\text{tr}(\Sigma_n^4))$$

$$(B.82) \quad \mathbb{E}((\mathbf{x}_i^\top \mathbf{x}_j)^2 (\mathbf{x}_l^\top \mathbf{x}_k)^2) = \text{tr}(\Sigma_n^2)^2.$$

PROOF. Since the observations are i.i.d. and centered, we have

$$(B.83) \quad \begin{aligned} \mathbb{E}((\mathbf{x}_i^\top \mathbf{x}_j)^2 (\mathbf{x}_l^\top \mathbf{x}_k)^2) &= (\text{var}(\mathbf{x}_i^\top \mathbf{x}_j))^2 \\ &= (\text{cov}(\mathbf{x}_i^\top \mathbf{x}_j, \mathbf{x}_i^\top \mathbf{x}_j))^2 \\ &= \left(\sum_{r=1}^p \sum_{s=1}^p \mathbb{E}(\mathbf{x}_{ir} \mathbf{x}_{jr} \mathbf{x}_{is} \mathbf{x}_{js}) \right)^2 \\ &= \left(\sum_{r=1}^p \sum_{s=1}^p ((\Sigma_n)_{rs})^2 \right)^2, \end{aligned}$$

which establishes (B.82). The two other assertions can be shown using conditional expectation.

For the statement (B.80), we use Lemma D.1 to obtain

$$\begin{aligned} \mathbb{E}((\mathbf{x}_i^\top \mathbf{x}_j)^4) &= \mathbb{E}(\mathbb{E}((\mathbf{x}_i^\top \mathbf{x}_j \mathbf{x}_j^\top \mathbf{x}_i)^2 | \mathbf{x}_j)) \\ &= \frac{\mathbb{E}(\xi_1^4)}{p(p+2)} \mathbb{E} \left(\left(\text{tr}(\Sigma_n^{1/2} \mathbf{x}_j \mathbf{x}_j^\top \Sigma_n^{1/2}) \right)^2 + 2\text{tr} \left((\Sigma_n^{1/2} \mathbf{x}_j \mathbf{x}_j^\top \Sigma_n^{1/2})^2 \right) \right) \\ &= 3 \frac{\mathbb{E}(\xi_1^4)}{p(p+2)} \mathbb{E}((\mathbf{x}_j^\top \Sigma_n \mathbf{x}_j)^2) \\ &= 3 \left(\frac{\mathbb{E}(\xi_1^4)}{p(p+2)} \right)^2 (\text{tr}(\Sigma_n^2)^2 + 2\text{tr}(\Sigma_n^4)). \end{aligned}$$

The argument for (B.81) is similar. $\square$

LEMMA D.3. *Let $\mathbf{u}_1 \in \mathbb{R}^p$ be a random vector that is uniformly distributed on the unit sphere, and let $M \in \mathbb{R}^{p \times p}$ be a non-random positive semidefinite matrix with $\lambda_1(M) \lesssim 1$. Then,*

$$(B.84) \quad \mathbb{E} \left| \mathbf{u}_1^\top M \mathbf{u}_1 - \frac{1}{p} \text{tr}(M) \right|^4 \lesssim \frac{1}{p^2}.$$

PROOF. Due to the orthogonal invariance of $\mathbf{u}_1$, we may work under the assumption that M is diagonal, i.e. $M = \text{diag}(\lambda_1(M), \dots, \lambda_p(M))$. Therefore, the quantity $\mathbb{E} \left| \mathbf{u}_1^\top M \mathbf{u}_1 - \frac{1}{p} \text{tr}(M) \right|^4$ is the same as

$$(B.85) \quad \sum_{j_1, j_2, j_3, j_4} \lambda_{j_1}(M) \lambda_{j_2}(M) \lambda_{j_3}(M) \lambda_{j_4}(M) \mathbb{E} \left(\prod_{l=1}^4 (\mathbf{u}_{1_{j_l}}^2 - \frac{1}{p}) \right).$$

Depending on the number of distinct indices among (j_1, j_2, j_3, j_4) , the following bounds can be obtained via direct calculation

$$(B.86) \quad \left| \mathbb{E} \left(\prod_{l=1}^4 (\mathbf{u}_{1_{j_l}}^2 - \frac{1}{p}) \right) \right| \lesssim \begin{cases} \frac{1}{p^6} & (4 \text{ distinct indices}) \\ \frac{1}{p^5} & (3 \text{ distinct indices}) \\ \frac{1}{p^4} & (1 \text{ or } 2 \text{ distinct indices}). \end{cases}$$

Since the eigenvalues of M are non-negative with $\lambda_1(M) \lesssim 1$, and since the numbers of terms in (B.85) involving k distinct indices is $\mathcal{O}(p^k)$, the stated result (B.84) is proved. $\square$

B.5. Discussion of examples in Section 3.3.1

This section provides detailed information related to examples of the random variable ξ_1^2 stated in Section 3.3.1. We give explicit parameterizations, and we check that the distributions satisfy the conditions in Assumption 2. The only three examples we do not individually cover are the Chi-Squared, Poisson, and Negative-Binomial distributions, because they can be decomposed into sums of independent random variables, and consequently, such examples are covered by the following lemma.

LEMMA E.1. Suppose $\xi_1^2 = \sum_{j=1}^p z_{1j}^2$ for some independent random variables $z_{11}, \dots, z_{1p}$ satisfying

$$(B.87) \quad \frac{1}{p} \sum_{j=1}^p \mathbb{E}(z_{1j}^2) = 1, \quad \frac{1}{p} \sum_{j=1}^p \text{var}(z_{1j}^2) = \tau + o(1) \quad \text{and} \quad \max_{1 \leq j \leq p} \mathbb{E}|z_{1j}|^{8+2\varepsilon} \lesssim p^{1+\frac{\varepsilon}{4}}$$

as $n \rightarrow \infty$, for some fixed constants $\tau \geq 0$ and $\varepsilon > 0$ not depending on n . Then, ξ_1^2 satisfies the conditions in Assumption 2.

PROOF. It is only necessary to show $\mathbb{E}|(\xi_1^2 - \mathbb{E}(\xi_1^2))/\sqrt{p}|^{4+\varepsilon} \lesssim 1$. Using Rosenthal's inequality (Johnson et al., 1985b) to bound the $L^{4+\varepsilon}$ norm of a sum of independent centered random variables, we have

$$(B.88) \quad \begin{aligned} \left(\mathbb{E}|\xi_1^2 - \mathbb{E}(\xi_1^2)|^{4+\varepsilon}\right)^{\frac{1}{4+\varepsilon}} &= \left(\mathbb{E}\left|\sum_{j=1}^p z_{1j}^2 - \mathbb{E}(z_{1j}^2)\right|^{4+\varepsilon}\right)^{\frac{1}{4+\varepsilon}} \\ &\lesssim \sqrt{\sum_{j=1}^p \text{var}(z_{1j}^2)} + p^{\frac{1}{4+\varepsilon}} \max_{1 \leq j \leq p} (\mathbb{E}|z_{1j}^2 - \mathbb{E}(z_{1j}^2)|^{4+\varepsilon})^{\frac{1}{4+\varepsilon}} \\ &\lesssim \sqrt{p} + p^{1/4} \max_{1 \leq j \leq p} (\mathbb{E}|z_{1j}|^{8+2\varepsilon})^{\frac{1}{4+\varepsilon}} \\ &\lesssim \sqrt{p}, \end{aligned}$$

which completes the proof. $\square$

Beta distribution The Beta(a, b) distribution with parameters $a, b > 0$ has a density function that is proportional to $x^{a-1}(1-x)^{b-1}$ for $x \in (0, 1)$.

LEMMA E.2. If $\beta > 0$ is fixed with respect to n and $\xi_1^2 \sim (p+2\beta)\text{Beta}(p/2, \beta)$, then ξ_1^2 satisfies the conditions in Assumption 2.

PROOF. The mean and variance are given by $\mathbb{E}(\xi_1^2) = p$ and $\text{var}(\xi_1^2) = \frac{4\beta p}{p+2\beta+2}$, and in particular we have $\text{var}(\xi_1^2)/p \rightarrow 0$ as $p \rightarrow \infty$. Based on Equation (25.14) in Johnson et al. (1995), the higher moments of ξ_1^2 are

$$\mathbb{E}(\xi_1^{2k}) = \prod_{j=0}^{k-1} \left(p + \frac{4j\beta}{p+2\beta+2j}\right).$$

Using the general relationship between central moments to ordinary moments

$$(B.89) \quad \mathbb{E}\left|\frac{\xi_1^2 - p}{\sqrt{p}}\right|^6 = \frac{1}{p^3} \sum_{j=0}^6 \binom{6}{j} (-1)^{6-j} \mathbb{E}(\xi_1^2)^{6-j} \mathbb{E}(\xi_1^{2j}),$$

it can be checked that that $\mathbb{E}|\frac{\xi_1^2 - p}{\sqrt{p}}|^6$ is a rational function of β that converges pointwise to 0 as $p \rightarrow \infty$. This implies the $(4 + \varepsilon)$ -moment condition in (3.10). $\square$

Beta-Prime distribution. A random variable W is said to follow a Beta-Prime(a, b) distribution with parameters $a, b > 0$ if it can be expressed as $W = \frac{U}{1-U}$ with $U \sim \text{Beta}(a, b)$.

LEMMA E.3. *If $\tau > 0$ and $\xi_1^2 \sim \text{Beta-Prime}\left(\frac{p(1+p+\tau)}{\tau}, \frac{1+p+2\tau}{\tau}\right)$, then ξ_1^2 satisfies the conditions in Assumption 2.*

PROOF. For any positive integer $k < (1 + p + 2\tau)/\tau$, the random variable ξ_1^2 satisfies the following moment formula (Siegrist, 2017, §5),

$$(B.90) \quad \mathbb{E}(\xi_1^{2k}) = \prod_{j=1}^k \frac{p(1+p+\tau) + (j-1)\tau}{1+p+(2-j)\tau}.$$

Since $2 < (1 + p + 2\tau)/\tau$, this gives $\mathbb{E}(\xi_1^2) = p$ and $\mathbb{E}(\xi_1^4) = p(p + \tau)$, which yield $\text{var}\left(\frac{\xi_1^2 - p}{\sqrt{p}}\right) = \tau$. Also, using the formula (B.89), it can be checked that $\mathbb{E}|\frac{\xi_1^2 - p}{\sqrt{p}}|^6$, viewed as a function of τ , converges pointwise to a polynomial function of τ as $p \rightarrow \infty$. This implies the $(4 + \varepsilon)$ -moment condition in (3.10). $\square$

Gamma distribution. For $\alpha, \beta > 0$, we parameterize the Gamma(α, β) distribution so that its density function is proportional to $x^{\alpha-1}e^{-\beta x}$ when $x \in (0, \infty)$.

LEMMA E.4. *If $\tau > 0$ and $\xi_1^2 \sim \text{Gamma}(p/\tau, 1/\tau)$, then ξ_1^2 satisfies the conditions in Assumption 2.*

PROOF. The conditions $\mathbb{E}(\xi_1^2) = p$ and $\text{var}\left(\frac{\xi_1^2 - p}{\sqrt{p}}\right) = \tau$ follow directly from the stated parameterization. Also, Theorem 2.3 in Boucheron et al. (2013) gives

$$(B.91) \quad (\mathbb{E}|\xi_1^2 - p|^6)^{1/6} \leq C(\sqrt{\tau p} + \tau)$$

for an absolute constant $C > 0$, which implies the $(4 + \varepsilon)$ -moment condition in (3.10). $\square$

Log-Normal distribution. A positive random variable Y is said to follow a Log-Normal(μ, σ^2) distribution if $\log(Y) \sim N(\mu, \sigma^2)$.

LEMMA E.5. *If $\tau > 0$ and $\xi_1^2 \sim \text{Log-Normal}\left(\log(p) - \frac{1}{2} \log\left(1 + \frac{\tau}{p}\right), \log\left(1 + \frac{\tau}{p}\right)\right)$, then ξ_1^2 satisfies the conditions in Assumption 2.*

PROOF. Equations (14.8a)-(14.8b) in Johnson et al. (1994) show that if $Y \sim \text{Log-Normal}(\mu, \sigma^2)$, then $\mathbb{E}(Y) = e^{\mu + \sigma^2/2}$ and $\text{var}(Y) = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1)$. Consequently, the stated choice of ξ_1^2 satisfies $\mathbb{E}(\xi_1^2) = p$ and $\text{var}(\xi_1^2)/p = \tau$. Equation (14.8c) in Johnson et al. (1994) shows that the central 6th moment of Y is

$$\mathbb{E}|Y - \mathbb{E}(Y)|^6 = e^{6\mu + 3\sigma^2} \sum_{j=0}^6 (-1)^j \binom{6}{j} e^{\sigma^2(6-j)(5-j)/2},$$

and a direct calculation reveals that $\mathbb{E}|\frac{\xi_1^2 - p}{\sqrt{p}}|^6$ is a polynomial function of τ whose coefficients are bounded as $p \rightarrow \infty$. This ensures that ξ_1 satisfies the $(4 + \varepsilon)$ -moment condition in (3.10).

Bibliography

- Akemann, G., Baik, J., and Di Francesco, P. (2011). *The Oxford Handbook of Random Matrix Theory*. Oxford.
- Albisetti, I., Balabdaoui, F., and Holzmann, H. (2020). Testing for spherical and elliptical symmetry. *Journal of Multivariate Analysis*, 180:104667.
- Anderson, T., Fang, K.-T., and Hsu, H. (1986). Maximum-likelihood estimates and likelihood-ratio criteria for multivariate elliptically contoured distributions. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, pages 55–59.
- Anderson, T. W. (2003). *Introduction to Multivariate Statistical Analysis*. Wiley.
- Babić, S., Gelgras, L., Hallin, M., and Ley, C. (2021). Optimal tests for elliptical symmetry: specified and unspecified location. *Bernoulli*, 27(4):2189–2216.
- Bai, Z., Chen, J., and Yao, J. (2010). On estimation of the population spectral distribution from a high-dimensional sample covariance matrix. *Australian & New Zealand Journal of Statistics*, 52(4):423–437.
- Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica*, pages 311–329.
- Bai, Z. and Silverstein, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Annals of Probability*, 32(1A):553–605.
- Bai, Z. and Silverstein, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*. Springer.
- Bai, Z. and Zhou, W. (2008). Large sample covariance matrices without independence structures in columns. *Statistica Sinica*, pages 425–442.
- Bai, Z.-D., Yin, Y.-Q., et al. (1993). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Ann. Probab*, 21(3):1275–1294.
- Beran, R. (1979). Testing for ellipsoidal symmetry of a multivariate density. *Annals of Statistics*, pages 150–162.

- Bickel, P. J. and Levina, E. (2004). Some theory for Fisher’s linear discriminant function, ‘naive Bayes’, and some alternatives when there are many more variables than observations. *Bernoulli*, 10(6):989 – 1010.
- Bing, X. and Latremouille, D. (2025). High-dimensional invariant tests of multivariate normality based on radial concentration. *arXiv:2504.09237*.
- Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford.
- Budczies, J. and Kosztyla, D. (2022). **cancerdata**: *Development and validation of diagnostic tests from high-dimensional molecular data: Datasets*. R package version 1.34.0.
- Cambanis, S., Huang, S., and Simons, G. (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, 11(3):368–385.
- Chen, H. and Xia, Y. (2023). A normality test for high-dimensional data based on the nearest neighbor approach. *Journal of the American Statistical Association*, 118(541):719–731.
- Chen, S. X. and Qin, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing.
- Chen, X., Zhang, J., and Zhou, W. (2022). High-dimensional elliptical sliced inverse regression in non-Gaussian distributions. *Journal of Business & Economic Statistics*, 40(3):1204–1215.
- Chen, Y., Wiesel, A., and Hero, A. O. (2011). Robust shrinkage estimation of high-dimensional covariance matrices. *IEEE Transactions on Signal Processing*, 59(9):4097–4107.
- Chernozhukov, V., Chetverikov, D., and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, pages 2786–2819.
- Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative finance*, 1(2):223.
- Couillet, R. and Debbah, M. (2011). *Random Matrix Methods for Wireless Communications*. Cambridge.
- Croux, C. and Haesbroeck, G. (2000). Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies. *Biometrika*, 87(3):603–618.

- Cui, H. and Zhang, L. (2026). Atm: An aggregation test of moments approach for assessing high-dimensional normality. *Scandinavian Journal of Statistics*, 53(1):532–553.
- Davison, A. C. and Huser, R. (2015). Statistics of extremes. *Annual Review of Statistics and its Application*, 2(1):203–235.
- Ding, X. and Xie, J. (2023). Tracy-Widom distribution for the edge eigenvalues of elliptical model. *arXiv:2304.07893*.
- Donoho, D. L. et al. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. *AMS math challenges lecture*, 1(2000):32.
- Dörnemann, N. (2026). Two-sample covariance inference in high-dimensional elliptical models. *Electronic Journal of Statistics*, 20(1):633–666.
- Ducharme, G. R. and de Micheaux, P. L. (2020). A goodness-of-fit test for elliptical distributions with diagnostic capabilities. *Journal of Multivariate Analysis*, 178:104602.
- Edelman, A. and Wang, Y. (2013). Random matrix theory and its innovative applications. In *Advances in Applied Mathematics, Modeling, and Computational Science*, pages 91–116. Springer.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1 – 26.
- Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. SIAM.
- Efron, B. (2012). Bayesian inference and the parametric bootstrap. *The annals of applied statistics*, 6(4):1971.
- El Karoui, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Annals of Statistics*, 36(6):2757–2790.
- El Karoui, N. (2009). Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond. *Annals of Applied Probability*, 19(6):2362–2405.
- El Karoui, N. and Purdom, E. (2016). The bootstrap, covariance matrices and PCA in moderate and high-dimensions. *arXiv preprint arXiv:1608.00948*.
- El Karoui, N. and Purdom, E. (2019). The non-parametric bootstrap and spectral analysis in moderate and high-dimension. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2115–2124. PMLR.

- Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond*, 1:176–223.
- Fang, K. W. (2018). *Symmetric Multivariate and Related Distributions*. CRC Press.
- Fisher, T. J., Sun, X., and Gallagher, C. M. (2010). A new test for sphericity of the covariance matrix for high dimensional data. *Journal of Multivariate Analysis*, 101(10):2554–2570.
- Geman, S. (1980). A limit theorem for the norm of random matrices. *The Annals of Probability*, 8(2):252–261.
- Goes, J., Lerman, G., and Nadler, B. (2020). Robust sparse covariance estimation by thresholding Tyler’s M-estimator. *Annals of Statistics*, 48(1):86–110.
- Gupta, A. K., Varga, T., and Bodnar, T. (2013). *Elliptically Contoured Models in Statistics and Portfolio Theory*. Springer.
- Hall, P. (2013). *The Bootstrap and Edgeworth Expansion*. Springer Science & Business Media.
- Hallin, M. and Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape. i. optimal rank-based tests for sphericity.
- Han, F. and Liu, H. (2013). Robust sparse principal component regression under the high dimensional elliptical model. *Advances in Neural Information Processing Systems*, 26.
- Han, F. and Liu, H. (2018). ECA: High-dimensional elliptical component analysis in non-Gaussian distributions. *Journal of the American Statistical Association*, 113(521):252–268.
- Härdle, W. K. and Simar, L. (2019). *Applied Multivariate Statistical Analysis*. Springer.
- Horn, R. A. and Johnson, C. R. (2012). *Matrix Analysis*. Cambridge.
- Hu, J., Li, W., Liu, Z., and Zhou, W. (2019a). High-dimensional covariance matrices in elliptical distributions with application to spherical test. *Annals of Statistics*, 47(1):527–555.
- Hu, J., Li, W., and Zhou, W. (2019b). Central limit theorem for mutual information of large MIMO systems with elliptically correlated channels. *IEEE Transactions on Information Theory*, 65(11):7168–7180.
- Hu, J., Li, W., and Zhou, W. (2019c). Central limit theorem for mutual information of large MIMO systems with elliptically correlated channels. *IEEE Transactions on Information Theory*, 65(11):7168–7180.
- Hubert, M., Rousseeuw, P. J., and Vanden Branden, K. (2005). Robpca: a new approach to robust principal component analysis. *Technometrics*, 47(1):64–79.

- Huffer, F. W. and Park, C. (2007). A test for elliptical symmetry. *Journal of Multivariate Analysis*, 98(2):256–281.
- Hung, H., Huang, S.-Y., and Eguchi, S. (2022). Robust self-tuning semiparametric PCA for contaminated elliptical distribution. *IEEE Transactions on Signal Processing*, 70:5885–5897.
- Janson, S. (1997). *Gaussian Hilbert Spaces*. Cambridge.
- Jarque, C. M. and Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters*, 6(3):255–259.
- Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions, Volume 1, (2nd ed.)*. Wiley.
- Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions, Volume 2, (2nd ed.)*. Wiley.
- Johnson, R. A., Wichern, D. W., et al. (2002). Applied multivariate statistical analysis.
- Johnson, W. B., Schechtman, G., and Zinn, J. (1985a). Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. *Annals of Probability*, pages 234–253.
- Johnson, W. B., Schechtman, G., and Zinn, J. (1985b). Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. *Annals of Probability*, pages 234–253.
- Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *The Annals of statistics*, 29(2):295–327.
- Johnstone, I. M. and Lu, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association*, 104(486):682–693.
- Jonsson, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix. *Journal of Multivariate Analysis*, 12(1):1–38.
- Jun, W., Jiahui, X., Long, Y., and Wang, Z. (2022). Tracy-Widom limit for the largest eigenvalue of high-dimensional covariance matrices in elliptical distributions. *Bernoulli*, 28(4):2941–2967.
- Kano, Y., Berkane, M., and Bentler, P. M. (1993). Statistical inference based on pseudo-maximum likelihood estimators in elliptical populations. *Journal of the American Statistical Association*, 88(421):135–143.
- Kariya, T. (1981). Robustness of multivariate tests. *The Annals of Statistics*, 9(6):1267–1275.

- Koltchinskii, V. and Sakhanenko, L. (2000). Testing for ellipsoidal symmetry of a multivariate distribution. In *High Dimensional Probability II*, pages 493–510. Springer.
- Kong, W. and Valiant, G. (2017). Spectrum estimation from samples. *Annals of Statistics*, 45(5):2218–2247.
- Ledoit, O. and Wolf, M. (2015). Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. *Journal of Multivariate Analysis*, 139:360–384.
- Ledoit, O. and Wolf, M. (2017). Numerical implementation of the QuEST function. *Computational Statistics & Data Analysis*, 115:199–223.
- Ledoux, M. (2001). *The Concentration of Measure Phenomenon*. Number 89. American Mathematical Soc.
- Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86(414):316–327.
- Li, S. and Mandic, D. (2022). Von Mises-Fisher elliptical distribution. *IEEE Transactions on Neural Networks and Learning Systems*.
- Li, W., Wang, Q., Yao, J., and Zhou, W. (2022). On eigenvalues of a high-dimensional spatial-sign covariance matrix. *Bernoulli*, 28(1):606–637.
- Li, W. and Yao, J. (2018). On structure testing for component covariance matrices of a high dimensional mixture. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(2):293–318.
- Lin, Q., Zhao, Z., and Liu, J. S. (2019). Sparse sliced inverse regression via lasso. *Journal of the American Statistical Association*, 114(528):1726–1739.
- Lin, X., Tian, F., and Huang, D. (2023). Hypothesis testing in Gaussian graphical models: Novel goodness-of-fit tests and conditional randomization tests. *arXiv:2312.01815*.
- Lopes, M. E. (2016). Unknown sparsity in compressed sensing: Denoising and inference. *IEEE Transactions on Information Theory*, 62(9):5145–5166.
- Lopes, M. E. (2022a). Central limit theorem and bootstrap approximation in high dimensions: Near $1/\sqrt{n}$ rates via implicit smoothing. *The Annals of Statistics*, 50(5):2492–2513.
- Lopes, M. E. (2022b). Improved rates of bootstrap approximation for the operator norm: A coordinate-free approach. *arXiv:2208.03050*.

- Lopes, M. E., Blandino, A., and Aue, A. (2019). Bootstrapping spectral statistics in high dimensions. *Biometrika*, 106(4):781–801.
- Lopes, M. E., Erichson, N. B., and Mahoney, M. W. (2023). Bootstrapping the operator norm in high dimensions: Error estimation for covariance matrices and sketching. *Bernoulli*, 29(1):428 – 450.
- Lopes, M. E. and Yao, J. (2022). A sharp lower-tail bound for Gaussian maxima with application to bootstrap methods in high dimensions. *Electronic Journal of Statistics*, 16(1):58–83.
- Lytova, A. and Pastur, L. (2009). Central limit theorem for linear eigenvalue statistics of the Wigner and the sample covariance random matrices. *Metrika*, 69(2):153–172.
- Magnus, J. R. et al. (1978). The moments of products of quadratic forms in normal variables. *Statistica Neerlandica*, 32(4):201–210.
- Mahmood, A. and Chitre, M. (2015). Modeling colored impulsive noise by Markov chains and alpha-stable processes. In *OCEANS 2015-Genova*, pages 1–7. IEEE.
- Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics*, 21(1):255–285.
- Mandelbrot, B. et al. (1963). The variation of certain speculative prices. *Journal of business*, 36(4):394.
- Manzotti, A., Pérez, F. J., and Quiroz, A. J. (2002). A statistic for testing the null hypothesis of elliptical symmetry. *Journal of Multivariate Analysis*, 81(2):274–285.
- Marcenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457.
- Marcus, A. W., Spielman, D. A., and Srivastava, N. (2022). Interlacing families III: Sharper restricted invertibility estimates. *Israel Journal of Mathematics*, 247(2):519–546.
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3):519–530.
- Mardia, K. V., Kent, J. T., and Taylor, C. C. (2024). *Multivariate Analysis*. John Wiley & Sons.
- Maronna, R. A. (1976). Robust M-estimators of multivariate location and scatter. *Annals of statistics*, pages 51–67.
- Massey Jr, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253):68–78.

- McLachlan, G. J. (2005). *Discriminant analysis and statistical pattern recognition*. John Wiley & Sons.
- McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton university press.
- Mestre, X. (2008). Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Transactions on Information Theory*, 54(11):5113–5129.
- Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. John Wiley & Sons.
- Najim, J. and Yao, J. (2016). Gaussian fluctuations for linear spectral statistics of large random covariance matrices. *Annals of Applied Probability*, 26(3):1837–1887.
- Ollila, E. and Tyler, D. E. (2014). Regularized M-estimators of scatter matrix. *IEEE Transactions on Signal Processing*, 62(22):6059–6070.
- Ollila, E., Tyler, D. E., Koivunen, V., and Poor, H. V. (2012). Complex elliptically symmetric distributions: Survey, new results and applications. *IEEE Transactions on signal processing*, 60(11):5597–5625.
- Osiewalski, J. and Steel, M. F. (1993). Robust bayesian inference in elliptical regression models. *Journal of Econometrics*, 57(1-3):345–363.
- Paindaveine, D. and Verdebout, T. (2016). On high-dimensional sign tests. *Bernoulli*, 22(3):1745 – 1769.
- Pan, G. and Zhou, W. (2008). Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. *Annals of Applied Probability*, 18(3):1232–1270.
- Patterson, N., Price, A. L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS genetics*, 2(12):e190.
- Potters, M. and Bouchaud, J.-P. (2020). *A First Course in Random Matrix Theory: For Physicists, Engineers and Data Scientists*. Cambridge.
- Purdom, E., Holmes, S. P., et al. (2005). Error distribution for gene expression data. *Statistical applications in genetics and molecular biology*, 4(1):1070.
- Schkoda, D. and Drton, M. (2025). Goodness-of-fit tests for linear non-Gaussian structural equation models. *Biometrika*, 112(4):asaf046.
- Schott, J. R. (2002). Testing for elliptical symmetry in covariance-matrix-based analyses. *Statistics & Probability Letters*, 60(4):395–404.

- Shao, J. and Tu, D. (2012). *The Jackknife and Bootstrap*. Springer Science & Business Media.
- Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4):591–611.
- Shi, G., Lin, N., and Zhang, B. (2024). Testing the mean vector for high-dimensional data. *Communications in Mathematics and Statistics*, pages 1–24.
- Siegrist, K. (2017). *Probability, Mathematical Statistics, and Stochastic Processes*. LibreTexts.
- Slakter, M. J. (1965). A comparison of the Pearson chi-square and Kolmogorov goodness-of-fit tests with respect to validity. *Journal of the American Statistical Association*, 60(311):854–858.
- Srivastava, M. S. (2005). Some tests concerning the covariance matrix in high dimensional data. *Journal of the Japan Statistical Society*, 35(2):251–272.
- Sun, Y., Babu, P., and Palomar, D. P. (2016). Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions. *IEEE Transactions on Signal Processing*, 64(14):3576–3590.
- Tang, G. and Nehorai, A. (2012). The stability of low-rank matrix reconstruction: a constrained singular value view. *IEEE Transactions on Information Theory*, 58(9):6079–6092.
- Tang, Y. and Li, B. (2023). A nonparametric test for elliptical distribution based on kernel embedding of probabilities. *arXiv:2306.10594*.
- Tian, X., Lu, Y., and Li, W. (2015). A robust test for sphericity of high-dimensional covariance matrices. *Journal of Multivariate Analysis*, 141:217–227.
- Tyler, D. E. (1987). A distribution-free M-estimator of multivariate scatter. *Annals of Statistics*, pages 234–251.
- van De Vijver, M. J., He, Y. D., van’t Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J., et al. (2002). A gene-expression signature as a predictor of survival in breast cancer. *New England Journal of Medicine*, 347(25):1999–2009.
- van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge.
- Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge.
- Wang, Q. and Yao, J. (2013). On the sphericity test with large-dimensional observations. *Electronic Journal of Statistics*, 7:2164–2192.
- Wang, S. and Lopes, M. E. (2023). A bootstrap method for spectral statistics in high-dimensional elliptical models. *Electronic Journal of Statistics*, 17(2):1848 – 1892.

- Yang, X., Zheng, X., and Chen, J. (2021). Testing high-dimensional covariance matrices under the elliptical distribution and beyond. *Journal of Econometrics*, 221(2):409–423.
- Yao, J. and Lopes, M. E. (2023). Rates of bootstrap approximation for eigenvalues in high-dimensional PCA. *Statistica Sinica*, 33:1461–1481.
- Yao, J., Zheng, S., and Bai, Z. (2015). *Sample Covariance Matrices and High-Dimensional Data Analysis*. Cambridge.
- Yin, Y. Q., Bai, Z., and Krishnaiah, P. R. (1988). On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probability Theory and Related Fields*, 78(4):509–521.
- Zellner, A. (1976). Bayesian and non-bayesian analysis of the regression model with multivariate student-t error terms. *Journal of the American Statistical Association*, 71(354):400–405.
- Zhang, Y., Hu, J., and Li, W. (2022). CLT for linear spectral statistics of high-dimensional sample covariance matrices in elliptical distributions. *Journal of Multivariate Analysis*.
- Zhao, Q. and Candès, E. J. (2022). An adaptively resized parametric bootstrap for inference in high-dimensional generalized linear models. *arXiv:2208.08944*.
- Zheng, S. (2012). Central limit theorems for linear spectral statistics of large dimensional F-matrices. In *Annales de l’IHP Probabilités et statistiques*, volume 48, pages 444–476.