

ON THE TESTING AND ESTIMATION OF HIGH-DIMENSIONAL  
COVARIANCE MATRICES

---

A Dissertation  
Presented to  
the Graduate School of  
Clemson University

---

In Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy  
Mathematical Sciences

---

by  
Thomas J. Fisher  
December 2009

---

Accepted by:  
Dr. Xiaoqian Sun, Committee Chair  
Dr. Colin M. Gallagher  
Dr. K. B. Kulasekera  
Dr. Chanseok Park

# Abstract

Many applications of modern science involve a large number of parameters. In many cases, the number of parameters,  $p$ , exceeds the number of observations,  $N$ . Classical multivariate statistics are based on the assumption that the number of parameters is fixed and the number of observations is large. Many of the classical techniques perform poorly, or are degenerate, in high-dimensional situations.

In this work, we discuss and develop statistical methods for inference of data in which the number of parameters exceeds the number of observations. Specifically we look at the problems of hypothesis testing regarding and the estimation of the covariance matrix.

A new test statistic is developed for testing the hypothesis that the covariance matrix is proportional to the identity. Simulations show this newly defined test is asymptotically comparable to those in the literature. Furthermore, it appears to perform better than those in the literature under certain alternative hypotheses.

A new set of Stein-type shrinkage estimators are introduced for estimating the covariance matrix in large-dimensions. Simulations show that under the assumption of normality of the data, the new estimators are comparable to those in the literature. Simulations also indicate the new estimators perform better than those in the literature in cases of extreme high-dimensions. A data analysis of DNA microarray data also appears to confirm our results of improved performance in the case of extreme high-dimensionality.

# Dedication

In loving memory of my grandparents, Charles Stanley and Anna Elizabeth Moser.

# Acknowledgments

No work of this magnitude can be completed entirely on your own. Many people have contributed to the work or helped me along the way in one form or another. I would like to thank several individuals for their contributions.

I owe an immeasurable amount of gratitude to Dr. Xiaoqian Sun for his leadership, friendship and the initial ideas for this work. Many of the primary results originate with his ideas. More than an advisor, he has provided guidance and been a good role model on how to conduct scientific research. The lessons learned under his guidance are much appreciated.

I'd like to thank Dr. Colin Gallagher for his willingness to work with me on the Martingale convergence proof. He has been a good friend my entire time in graduate school and his help is appreciated.

Dr. KB Kulasekera for working with me for my Master project, and continuing to be a positive influence during my PhD work.

Dr. Chanseok Park for his review and editorial changes improving the final version of this work.

Jenny Foster for her continuing love and support. Her support during the multiple emotional trials when I contemplated quitting cannot be measured. Without her, I would not be where I am today. I cannot thank her enough and only hope I can provide an equal amount of love and support in the future.

Dr. Charlie Toll of UMBC for his willingness to take a chance on me and recommend me for the graduate program at Clemson.

Jared Bunting for his friendship and helpfulness to proofread research papers I know didn't interest him. I would also like to thank him for his help with computational issues while I was working with the Center for Advanced Engineering Fibers and Films.

To my family and friends whose love and support over the years cannot be summed up here. Specifically I'd like to thank my father Charles Ronald Fisher, my sister Barbara and my niece Jorji.

Lastly, WWWQ-HD2 99X out of Atlanta, whose streaming radio was listened during the majority of the writing of this dissertation.

# Table of Contents

<b>Title Page</b> . . . . .	<b>i</b>
<b>Abstract</b> . . . . .	<b>ii</b>
<b>Dedication</b> . . . . .	<b>iii</b>
<b>Acknowledgments</b> . . . . .	<b>iv</b>
<b>List of Tables</b> . . . . .	<b>vii</b>
<b>List of Figures</b> . . . . .	<b>viii</b>
<b>1 Introduction</b> . . . . .	<b>1</b>
1.1 Estimation . . . . .	3
1.2 Hypothesis Testing . . . . .	3
1.3 Need for new Techniques . . . . .	4
1.4 Organization of this Work . . . . .	7
<b>2 Hypothesis Testing</b> . . . . .	<b>8</b>
2.1 Testing for an independence structure . . . . .	9
2.2 Testing for Sphericity . . . . .	12
2.3 Testing for Identity Covariance Matrix . . . . .	24
2.4 Concluding Remarks . . . . .	27
<b>3 Estimation</b> . . . . .	<b>29</b>
3.1 Estimation Techniques . . . . .	30
3.2 Stein-Type Shrinkage Estimators . . . . .	32
3.3 Concluding Remarks . . . . .	56
<b>Appendices</b> . . . . .	<b>57</b>
A Proof of $(n, p)$ -Consistent Test for Sphericity . . . . .	58
B Technical Details for Shrinkage Estimators . . . . .	108
C Selected Source Code . . . . .	111
<b>Bibliography</b> . . . . .	<b>114</b>

# List of Tables

2.1	ASL for $T$ in (2.35) . . . . .	17
2.2	ASL for $T_s$ (2.26) . . . . .	17
2.3	Simulated Power for $T$ in (2.35) under $\Sigma = \Lambda \sim Unif(0.5, 1.5)$ . . . . .	20
2.4	Simulated Power for $T_s$ in (2.26) under $\Sigma = \Lambda \sim Unif(0.5, 1.5)$ . . . . .	20
2.5	Simulated Power for $T$ in (2.35) under $\Sigma = \Lambda \sim Unif(0.5, 10.5)$ . . . . .	20
2.6	Simulated Power for $T_s$ in (2.26) under $\Sigma = \Lambda \sim Unif(0.5, 10.5)$ . . . . .	20
2.7	Simulated Power for $T$ in (2.35) under <i>near</i> Spherical Covariance Matrix . . . . .	21
2.8	Simulated Power for $T_s$ in (2.26) under <i>near</i> Spherical Covariance Matrix . . . . .	21
3.1	$\lambda$ estimation for $n = 40, p = 20, T = \mu I$ . . . . .	46
3.2	$\lambda$ estimation for $n = 30, p = 30, T = \mu I$ . . . . .	46
3.3	$\lambda$ estimation for $n = 5, p = 100, T = \mu I$ . . . . .	46
3.4	Shrinkage estimation for $n = 40, p = 20, T = \mu I$ . . . . .	47
3.5	Shrinkage estimation for $n = 30, p = 30, T = \mu I$ . . . . .	47
3.6	Shrinkage estimation for $n = 5, p = 100, T = \mu I$ . . . . .	48
3.7	Shrinkage estimation for $n = 40, p = 20, T = I$ . . . . .	48
3.8	Shrinkage estimation for $n = 30, p = 30, T = I$ . . . . .	48
3.9	Shrinkage estimation for $n = 5, p = 100, T = I$ . . . . .	49
3.10	$\lambda$ estimation for $n = 40, p = 20, T = D$ . . . . .	49
3.11	$\lambda$ estimation for $n = 30, p = 30, T = D$ . . . . .	49
3.12	$\lambda$ estimation for $n = 5, p = 100, T = D$ . . . . .	50
3.13	Shrinkage estimation for $n = 40, p = 20, T = D$ . . . . .	50
3.14	Shrinkage estimation for $n = 30, p = 30, T = D$ . . . . .	50
3.15	Shrinkage estimation for $n = 5, p = 100, T = D$ . . . . .	50
3.16	Condition Numbers (and $\hat{\lambda}$ ) for estimators and common targets on E.coli data . . . . .	55

# List of Figures

2.1	Normal QQ-Plot for $T$ in (2.35) under $H_0$ . . . . .	18
2.2	Normal QQ-Plot for $T$ in (2.35) under $H_A$ . . . . .	19
3.1	Timing Study as $n$ increases, $p$ fixed . . . . .	52
3.2	Timing Study as $p$ increases, $n$ fixed . . . . .	53
3.3	Timing Study as $n = p$ increases . . . . .	54



# Chapter 1

## Introduction

Multivariate statistics concerns the analysis of data consisting of more than one measurement on a number of individuals or objects. We consider the number of measurements to be the dimension of our analysis, typically denoted with  $p$ . The number of observations,  $N = n + 1$ , are drawn randomly from a  $p$ -dimensional population,  $\Theta$ , and are known as the random sample of size  $N$ . The measurements made on a single observation are typically assembled into a column vector, i.e.

$$\mathbf{x}_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}$$

where  $i$  represents the  $i^{\text{th}}$ ,  $i = 1, \dots, N$ , observation from the random sample. The set of measurements on all observations in a sample set side-by-side make up a matrix of observations,  $\mathbf{X}$  such that,

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_N \end{pmatrix}$$

and  $\mathbf{X}$  is a  $p \times N$ -matrix. We assume each vector to be from a multivariate population. When an observation is drawn randomly, we consider the vector to have a set of probability laws describing the population, known as a distribution. The  $p$ -dimensional population is assumed to have a  $p \times 1$ -mean

vector  $\mu$  and a  $p \times p$ -covariance matrix  $\Sigma$  such that

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{pmatrix}$$

where  $\mu_i = E[x_i]$  and  $\sigma_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)]$  for  $i, j < p$ . The expectation is defined in the typical fashion,

$$E[X] = \int x f(x) dx$$

where  $f$  is the distribution function of the random variable  $X$ .  $\Sigma$  is positive definite, typically denoted  $\Sigma > 0$  and can be expressed using the matrix notation,

$$\Sigma = E[(\mathbf{x}_i - E[\mathbf{x}_i])(\mathbf{x}_i - E[\mathbf{x}_i])']$$

where  $\mathbf{x}_i'$  denotes the transpose of the  $p \times 1$  vector  $\mathbf{x}_i$ .

Many of the *classical* statistical methods that have been developed and assessed can be put in the context of the multivariate Normal, or Gaussian, distribution, denoted with  $\mathbf{x}_i \sim N_p(\mu, \Sigma)$ . The probability density function for  $p \times 1$ -dimensional random vector  $\mathbf{x}$  from the multivariate normal distribution is defined as,

$$f(\mathbf{x}) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu)\right),$$

where  $|\Sigma|$  denotes the determinant operation on the matrix  $\Sigma$  and  $|\Sigma| \neq 0$  since  $\Sigma$  is positive definite.  $\Sigma^{-1}$  is the matrix inverse of  $\Sigma$  and is known as the precision matrix.

The *normality* assumption can generally be relaxed when applying many of the methods discussed here, a property known as *robustness*. Regardless of the assumed distribution, in the classical case there is the assumption that the number of observations is greater than the dimensionality of our data, i.e.  $N > p$ .

## 1.1 Estimation

We are generally interested in the two parameters, the mean vector,  $\mu$ , and the covariance matrix,  $\Sigma$ . However each is unknown and must be estimated. The typical estimates are the sample mean vector,  $\bar{\mathbf{x}}$ , and sample covariance matrix,  $\mathbf{S}$ , both from a sample of size  $N = n + 1$ , where

$$\bar{\mathbf{x}} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{pmatrix}$$

with

$$\bar{x}_i = \frac{1}{N} \sum_{j=1}^N x_{ij}$$

is the sample mean of the  $i^{\text{th}}$  covariate. The sample covariance matrix is typically defined as

$$\mathbf{S} = \frac{1}{N-1} \sum_{j=1}^N (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' \quad (1.1)$$

and  $\mathbf{S}$  can be written in matrix form

$$\mathbf{S} = \frac{1}{N-1} (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})' \quad (1.2)$$

where  $\bar{\mathbf{X}}$  is a  $p \times N$  matrix with each column comprised of  $\bar{\mathbf{x}}$ .

It is well-known that both  $\bar{\mathbf{x}}$  and  $\mathbf{S}$  are based on the maximum likelihood estimators, are unbiased and consistent, when  $n \rightarrow \infty$  with  $p$  fixed, for  $\mu$  and  $\Sigma$  respectively. They typically are considered the *best* estimators available in the classical statistical case of  $N > p$ .

## 1.2 Hypothesis Testing

Decision making about  $\mu$  and  $\Sigma$  is generally achieved through likelihood ratio criterion. For hypothesis test regarding  $\mu$  this test is based on the  $T^2$ -statistic that is

$$T^2 = N(\bar{\mathbf{x}} - \mu)' \mathbf{S}^{-1} (\bar{\mathbf{x}} - \mu). \quad (1.3)$$

Tests regarding the covariance matrix are also based on the likelihood ratio criterion. These tests are based on a function of the eigenvalues of the sample covariance matrix. Specific Likelihood Ratio Tests (LRT) for varying hypothesis are described in Chapter 2. In all the hypothesis testing thus described, we are assuming the sample size,  $N$ , exceeds that of the dimensionality,  $p$ . The consistency of these tests and other properties are shown when  $n \rightarrow \infty$  with  $p$  fixed. Details are provided in many multivariate analysis, asymptotic statistical and mathematical statistical texts, see Anderson [3], Johnson and Wichern [63], van der Vaart [107], Casella and Berger [14], Rao [81], Dudewicz and Mishra [27], Lehmann [73] and others.

### 1.3 Need for new Techniques

Many applications in modern science and economics, (e.g. the analysis of DNA Microarrays, computer science network programs, portfolio analysis in economics) involve observations where the number of variables,  $p$ , is large. In many cases, the number of variables is of the same, or greater, magnitude as that of the number of observations,  $N$ . The theoretical framework for the classical approach is restricted to the case where the number of observations grow while the number of variables stays fixed. Specifically when working with the covariance matrix, new techniques are needed as the typical methods run into several severe problems in the modern applications. Modern techniques explore what is known as  $(n, p)$ -asymptotics, “general asymptotics,” “concentration asymptotics,” see Ledoit & Wolf [69], or “increasing dimension asymptotics,” see Serdobolskii [93]. The  $(n, p)$ -asymptotics are a generalization of the classical techniques, however we consider the case where both  $n \rightarrow \infty$  and  $p \rightarrow \infty$ . Furthermore, we typically relax restrictions on the dimensionality, as it can be larger than the sample size.

#### 1.3.1 Estimation in High-Dimensions

The sample covariance matrix,  $\mathbf{S}$ , defined in (1.1) becomes ill-conditioned or near singular as the number of variables,  $p$ , approaches the number of observations. It is well known that the eigenvalues of  $\mathbf{S}$  disperse from the true eigenvalues of  $\mathbf{\Sigma}$ . The smallest eigenvalues will go to zero and the largest will go to  $\infty$  as the dimensionality increases. When  $p > n$ ,  $\mathbf{S}$  is singular and the smallest eigenvalues are zero. In many cases an accurate estimate for  $\mathbf{\Sigma}$ , or  $\mathbf{\Sigma}^{-1}$ , the precision matrix, is required. As  $p$  approaches  $N$ , or even passes it, the typical estimate for  $\mathbf{\Sigma}$  becomes degenerate and

shouldn't be used.

Typically in high-dimensions, estimation of  $\Sigma$  is achieved by putting some predetermined structure, such as banding or diagonality, on the estimate for the covariance matrix. This structure is chosen to be well-conditioned, non-singular and if previous knowledge about the data is available, should be representative of the true covariance matrix.

Another popular method for estimation of the covariance matrix is that of Stein-type shrinkage estimation. A convex combination of the sample covariance matrix,  $\mathbf{S}$ , and some well-conditioned target matrix is used to estimate the covariance matrix. The idea is to reduce the weight on  $\mathbf{S}$  and put more weight on the target matrix when the dimensionality increases. This convex combination *shrinks* the eigenvalues of the sample covariance matrix to that of the target matrix. The target matrix is chosen to be well-conditioned and non-singular. Typically matrices such as the identity are used, but if previous knowledge of the dataset is known (e.g. stock market returns), a different target can be designed for that particular dataset.

Since the calculation of  $\mathbf{S}$  is well-known, and the target is predetermined, the computational issues associated with Stein-type shrinkage estimation are in determining an appropriate *weight*, typically called the shrinkage intensity. Recent work by Ledoit & Wolf [70] [72] shows an optimal intensity will always exist under the quadratic risk, however this optimal weight must also be estimated. We recap the estimators from the literature and introduce a new set of shrinkage estimators for three common target matrices. The newly suggested estimators are found under fairly general assumptions and are unbiased like many of those from the literature. A simulation study and data analysis indicate the newly suggested estimators are comparable to those in the literature, and appear to perform better in cases of extreme high-dimensions.

### 1.3.2 Hypothesis Testing in High-Dimensions

As aforementioned, the likelihood ratio criterion is typically used for hypothesis test for the covariance matrix. In the high-dimensional case, the likelihood ratio criterion fails or is computationally unstable. The likelihood ratio criterion is based on the eigenvalues of the sample covariance matrix. As discussed above, when  $p > n$ , only the first  $n$  eigenvalues will be non-zero. Also, the smallest eigenvalues will tend to zero pretty quickly as the dimensionality grows. The LRT uses all  $p$  eigenvalues and typically requires calculating the geometric mean of the eigenvalues. The geometric mean will always result in a value of zero when  $p > n$  and will be computationally close to zero

when  $p$  is large with respect to  $n$ . Thus, the Likelihood Ratio criterion is unstable and should not be used in cases of high-dimensions.

Recent work has looked at many of the common hypotheses regarding the covariance matrix when the dimensionality is large, including: the covariance is the identity, proportional to the identity (typically called sphericity), the covariance matrix is a diagonal matrix and multiple covariance matrices are equal.

Our work begins by recapping the new methods for testing if the covariance matrix is a diagonal matrix. The primary method is to look at the off-diagonal elements of the covariance or correlation matrix. Under the null hypothesis, the true matrix is diagonal, and we'd expect the off-diagonal elements of the sample covariance to be zero. A test statistic is constructed based on these off-diagonal elements. We generalize the results in the literature to the case of a block diagonal matrix.

We then discuss testing the hypothesis for sphericity of the covariance matrix. Several tests have been introduced in recent years utilizing the first and second arithmetic means of the eigenvalues of  $\mathbf{S}$ . We note that unlike the geometric mean, arithmetic means will not be adversely affected by zero eigenvalues. We develop a new statistic for testing sphericity based on the ratio of the fourth and second arithmetic means of the eigenvalues of the sample covariance matrix. The asymptotic distribution of the statistic is found under both the null and alternative hypotheses. The test statistic is shown to be consistent as both the sample size and dimensionality grow together. Simulations indicate this new test is asymptotically comparable to those in the literature and that under a certain class of matrices under the alternative distribution we call *near* spherical matrices, the newly defined test appears to be more powerful than those in the literature. Lastly a data analysis is performed for comparisons with the literature.

Lastly, we explore the results in the literature, as well as discuss how to modify the newly defined statistic for sphericity, to test the hypothesis that the covariance matrix is the identity. We note this is a special case of sphericity with proportion one. The idea of these procedures is to utilize the fact that all eigenvalues will be one under the null hypothesis. Thus, on average, we'd expect the eigenvalues of the sample covariance to be about one as well. We can look at squared or quartic difference between the eigenvalues and the theoretical value of one to develop testing procedure to see if the covariance matrix is the identity.

## 1.4 Organization of this Work

Chapter 2 provides a detailed discussion of much of the previous work in the literature regarding hypothesis testing of the covariance matrix when the dimensionality is large. We introduce several new testing procedures for multiple hypotheses and provide a simulation study indicating an improvement in certain situations. Chapter 3 provides a review of the estimation procedures of the covariance matrix, and introduces several new shrinkage estimators for the covariance matrix. We note we drop the boldface notation to distinguish vectors and matrices from univariate constants and random variables.

## Chapter 2

# Hypothesis Testing

We discuss advances in hypothesis testing of covariance matrices when the number of observations is less than the dimensionality of the matrix. This is a common problem in modern genetic research, medicine and economics. In the classical case, i.e.  $n > p$  and  $n$ -large, test based on the likelihood ratio are used. The likelihood ratio test becomes degenerate when  $p > n$ . Details will be provided in the individual sections below.

Much of the current work rests on the large body of literature regarding asymptotics for eigenvalues of random matrices, specifically the sample covariance matrix, such as Arharov [4], Bai [5], Narayanaswamy and Raghavarao [76], Girko [47] [45] [46], Serdobolskii [92] [93] [91], Silverstein [95], Yin and Krishnaiah [112] and others. We build on the substantial list of work completed on statistical testing in high-dimensions, such as Bai, Krishnaiah, and Zhao [8], Saranadasa [84], Kim and Press [65], Girko [48] and most recently the work completed by Ledoit and Wolf [69], Srivastava [99] [100] [97] and Schott [87], [88], [89], [90].

This chapter contains four sections discussing three common hypotheses about the covariance matrix and concluding remarks. We begin by discussing the work completed by Srivastava [99] and Schott [88] for testing independence of variables in the covariance matrix, i.e. the covariance matrix is a diagonal matrix. Using the methodology of Srivastava [99] we discuss how to create a test for block-diagonality of the covariance matrix. In the second section we explore the existing tests for sphericity of the covariance matrix and develop a new test based on the Cauchy-Schwarz inequality. Further exploration into the performance of our newly defined test is achieved through a simulation study and the application of our new test on some microarray data. In the third section



we discuss the work that has been done on testing that the covariance matrix is the identity. We conclude with some observations, remarks and discussion of potential future work.

## 2.1 Testing for an independence structure

Let  $x_1, x_2, \dots, x_N$  be iid  $N_p(\mu, \Sigma)$ , sample size  $N = n + 1$ . We consider the problem of testing the hypothesis of complete independence, i.e.  $H_0 : \Sigma = \text{diag}(\sigma_{11}^2, \sigma_{22}^2, \dots, \sigma_{pp}^2)$ . We begin by discussing the work of Schott [88] and Srivastava [99].

Schott [87] notes the above hypothesis is equivalent to  $\sigma_{ij} = 0$  for all  $1 \leq i < j \leq p$ . A simple and intuitive statistic for testing would be based on the  $s_{ij}$ 's from the sample covariance matrix  $S$ . Under the assumptions that the data comes from a multivariate normal distribution, Schott [87] looks at the sum of squares of the  $s_{ij}$ s and centralizes to mean zero. He derives

$$t_{np} = \sum_{i=2}^p \sum_{j=1}^{i-1} (s_{ij}^2 - \frac{s_{ii}s_{jj}}{n}) \quad (2.1)$$

and finds

$$\hat{\sigma}_{np}^2 = \frac{2(n-1)}{n^2(n+2)} \sum_{i=2}^p \sum_{j=1}^{i-1} s_{ii}^2 s_{jj}^2 \quad (2.2)$$

is a consistent estimator for the variance of  $t_{np}$  from (2.1). He then shows that  $t_{np}^* = t_{np}/\hat{\sigma}_{np}^2 \sim N(0, 1)$  as  $(n, p) \rightarrow \infty$ . In [88] he provides a similar result using the sample correlations.

Srivastava [99] develops a similar test using the sample correlation matrix. Noting that an equivalent null hypothesis is  $\rho_{ij} = 0$  when  $\rho_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$ . We let  $r_{ij}$  be the sample correlation, or

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}}, i \neq j \quad (2.3)$$

and he constructs a test based on the sum of squares of  $r_{ij}$ s for all  $i \neq j$ . Srivastava [99] [100] shows as  $(n, p) \rightarrow \infty$

$$T_s^* = \frac{n \sum_{i < j}^p r_{ij}^2 - q}{\sqrt{2q}} \sim N(0, 1), \quad (2.4)$$

where

$$q = \frac{1}{2}p(p-1).$$

We build on the idea of independence to the more general case of a block diagonal structure,

specifically we are interested in testing  $H_0 : \Sigma = \Sigma^*$  against  $H_A : \Sigma \neq \Sigma^*$  where

$$\Sigma^* = \begin{pmatrix} \Sigma_{11} & 0 & \dots & 0 \\ 0 & \Sigma_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Sigma_{kk} \end{pmatrix}, \quad (2.5)$$

with each  $\Sigma_{ii}$  being a  $p_i \times p_i$  positive definite covariance matrix, for  $i = 1, \dots, k$ . In Schott [87], he provides a generalization of his above result for testing against a block diagonal structure.

We provide a generalization of the result provided by Srivastava [99]. This hypothesis equates to testing  $H_0 : \sigma_{ij} = 0$  for all  $(i, j)$  off-block diagonals entries, against the alternative  $\sigma_{ij} \neq 0$  for at least one-pair  $(i, j)$ . Without loss of generality we can consider the hypothesis based on the correlation,  $H_0 : \rho_{ij} = 0$  versus  $H_1 : \rho_{ij} \neq 0$  for at least one pair  $(i, j)$  on the off diagonal block entries where  $\rho_{ij} = \sigma_{ij}/(\sigma_{ii}\sigma_{jj})^{1/2}$  for  $i \neq j$ . Let  $r_{ij}$  be the sample correlation coefficient defined above in (2.3). We define

$$q = p_1p_2 + p_1p_3 + \dots + p_1p_k + p_2p_3 + \dots + p_2p_k + \dots + p_{k-1}p_k \quad (2.6)$$

and

$$r = (r_{1,p_1+1}, \dots, r_{1,p}, r_{2,p_1+1}, \dots, r_{p_1+\dots+p_{k-1}, p_1+\dots+p_{k-1}+1}, \dots, r_{p_1+\dots+p_{k-1}, p})'. \quad (2.7)$$

That is,  $r$  is a vector of the sample correlations corresponding to the off-block diagonal entries with length  $q$ . Let

$$\rho = (\rho_{1,p_1+1}, \rho_{1,p_1+2}, \dots, \rho_{p_1+\dots+p_{k-1}, p})' \quad (2.8)$$

be the corresponding *true* correlations for the off-block diagonal entries. Under  $H_0$ , each  $\rho_{ij} = 0$  and hence  $\rho$  is a vector comprised of  $q$  zero elements. By Hsu [58],

$$\sqrt{nr} \sim N_q(0, \Omega)$$

where the covariance matrix,  $\Omega$ , has diagonal elements  $(1 - \rho_{ij}^2)^2$  for each  $i, j$  in  $\rho$ , which equates to

1 under  $H_0$ , and off-diagonal elements given by

$$\begin{aligned} \text{cov}(r_{ij}, r_{kl}) &= \rho_{ik}\rho_{jl} + \rho_{il}\rho_{jk} - \rho_{ij}(\rho_{ik}\rho_{il} + \rho_{jk}\rho_{jl}) - \rho_{jk}(\rho_{ik}\rho_{kl} + \rho_{il}\rho_{jl}) \\ &\quad + \frac{1}{2}\rho_{ij}\rho_{kl}(\rho_{ik}^2 + \rho_{il}^2 + \rho_{jk}^2 + \rho_{jl}^2). \end{aligned} \quad (2.9)$$

(2.9) can be simplified to

$$\text{cov}(r_{ij}, r_{kl}) = \rho_{ik}\rho_{jl} + \rho_{il}\rho_{jk} \quad (2.10)$$

under the null hypothesis.

**Lemma 2.1.1.**

$$nr'\Omega^{-1}r \xrightarrow{D} \chi_q^2 \text{ as } n \rightarrow \infty$$

*Proof.* A common result in Linear Models, see Graybill [50] □

**Theorem 2.1.1.** As  $n \rightarrow \infty$ ,

$$\frac{nr'\Omega^{-1}r - q}{\sqrt{2q}} \xrightarrow{D} N(0, 1).$$

*Proof.* An application of the Central Limit Theorem on the asymptotically independent  $\chi^2$  random variables provides the result. □

Application of Lemma 2.1.1 and Theorem 2.1.1 allows us to test  $H_0 : \Sigma = \Sigma^*$  vs  $H_A : \Sigma \neq \Sigma^*$  with the statistic

$$Z = \frac{nr'\hat{\Omega}^{-1}r - q}{\sqrt{2q}} \quad (2.11)$$

where  $\hat{\Omega}$  is the consistent estimator of  $\Omega$  comprised of  $r_{ij}$ s.

**Theorem 2.1.2.** As  $(n, p) \rightarrow \infty$ , the test statistic,  $Z \xrightarrow{D} N(0, 1)$

*Proof.* We provide the general argument for the proof. Each  $r_{ij} \rightarrow \rho_{ij}$  as  $n \rightarrow \infty$ , hence  $\hat{\Omega} \rightarrow \Omega$  and  $\hat{\Omega}^{-1} \rightarrow \Omega^{-1}$  by Slutsky's Theorem and continuous mapping theorem. Whence  $nr'\hat{\Omega}^{-1}r \xrightarrow{D} \chi_q^2$ . This argument shows  $n$ -consistency and is not based on the behavior of  $p$ .

An application of the central limit theorem with respect to  $p$ , and hence  $q$ , on  $nr'\hat{\Omega}^{-1}r$  gives us the result  $Z = \frac{nr'\hat{\Omega}^{-1}r - q}{\sqrt{2q}} \xrightarrow{D} N(0, 1)$  as  $(n, p) \rightarrow \infty$ .

Further details for the convergence can be found in Srivastava [100]. □

## 2.2 Testing for Sphericity

We consider the problem of testing for sphericity of the covariance matrix, or that the covariance matrix is proportional to the identity matrix. Explicitly written as  $H_0 : \Sigma = \sigma^2 I$  vs  $H_A : \Sigma \neq \sigma^2 I$ , where  $\sigma^2$  is an unknown scalar proportion. As previously discussed, when  $n > p$  the appropriate test is the likelihood ratio test (LRT) defined below in (2.12). It is a test based on the eigenvalues of the sufficient statistic  $S$ . The LRT is

$$W = \left( \frac{\prod_{i=1}^p l_i^{1/p}}{(1/p) \sum_{i=1}^p l_i} \right)^p \quad (2.12)$$

where  $l_i, i = 1, \dots, p$ , are the eigenvalues of  $S$ . From Anderson [3]

$$P(-n\rho \log W \leq z) = P(\chi_f^2 \leq z) + \omega_2 (P(\chi_{f+4}^2 \leq z) - P(\chi_f^2 \leq z)) + O(n^{-3}) \quad (2.13)$$

where

$$f = \frac{1}{2}p(p+1) - 1, \quad (2.14)$$

$$\rho = 1 - \frac{2p^2 + p + 2}{6pn}, \quad (2.15)$$

$$\omega_2 = \frac{(p+2)(p-1)(p-2)(2p^3 + 6p^2 + 3p + 2)}{288p^2 n^2 \rho^2}. \quad (2.16)$$

The LRT has been shown to have a monotone power function by Carter and Srivastava [13]. The LRT depends on the geometric mean of the sample eigenvalues. When  $n < p$ , the likelihood ratio test is degenerate since only the first  $n$  eigenvalues of the sample covariance matrix will be non-zero, resulting in a geometric mean of zero. Furthermore, as  $p \approx n$ ,  $S$  becomes ill-conditioned, and the eigenvalues disperse from the true eigenvalues. Bai and Yin [7] show the smallest non-zero eigenvalues will approach a limit close to zero creating an ill-conditioned, or degenerate, test. New methodology is necessary when the number of variables,  $p$ , is of the same order, or larger, as the number of observations,  $n$ .

Ledoit and Wolf [69] show the locally best invariant test based on John's U statistic [62], see (2.19), to be  $(n, p)$ -consistent when  $(p/n) \rightarrow c < \infty$  and  $c$  is a constant known as the concentra-

tion. However the distribution of the test statistic under the alternative hypothesis is not available. Srivastava [99] proposes a test based on consistent estimators of the trace of powers of  $\Sigma$ . His test, like that of John, is based on the first and second arithmetic means of the sample eigenvalues but only requires the more general condition  $n = O(p^\delta)$ ,  $0 < \delta \leq 1$ . Furthermore, in Srivastava [100] he proposes a modified version of the LRT in which only the first  $n$  eigenvalues are used. This test is applicable under the assumptions  $n/p \rightarrow 0$  and  $n$  fixed. Motivated by the result in [99] we propose a test based on consistent estimators of the second and fourth arithmetic means of the sample eigenvalues.

We begin by highlighting the technical results of the work done in the literature. Under the assumptions that  $(p/n) \rightarrow c < \infty$ , normality of the data, and finite fourth arithmetic means of the eigenvalues, Ledoit and Wolf [69] explore John's U-statistic [62]

$$U = \frac{1}{p} \text{tr} \left[ \left( \frac{S}{(1/p)\text{tr}(S)} - I \right)^2 \right] = \frac{(1/p)\text{tr}(S^2)}{[(1/p)\text{tr}(S)]^2} - 1. \quad (2.17)$$

They show that as  $(n, p) \rightarrow \infty$

$$nU - p \xrightarrow{D} N(1, 4) \quad (2.18)$$

and hence

$$U_J = \frac{nU - p - 1}{2} \xrightarrow{D} N(0, 1) \text{ as } (n, p) \rightarrow \infty. \quad (2.19)$$

Srivastava [100] proposes an adapted version of the likelihood ratio test when  $n > p$  to the case  $n < p$  simply by interchanging  $n$  and  $p$ . We let

$$\begin{aligned} c_1 &= \frac{(n+1)(n-1)(n+2)(2n^3 + 6n^2 + 3n + 2)}{288n^2}, \\ m_1 &= p - \frac{2n^2 + n + 2}{6n}, \\ g_1 &= \frac{1}{2}n(n+1) - 1, \\ Q_1 &= -m_1 \log L_1, \end{aligned} \quad (2.20)$$

where

$$L_1 = \frac{\prod_{i=1}^n l_i}{\left( \frac{1}{n} \sum_{i=1}^n l_i \right)^n} \quad (2.21)$$

and  $l_i$  are the eigenvalues of the sample covariance matrix. We note  $L_1$  is based on the first  $n$  non-zero eigenvalues of  $S$ . Srivastava [100] provides the following result

$$P(Q_1 \geq z) = P(\chi_{g_1}^2 \geq z) + c_1 m_1^{-2} [P(\chi_{g_1+4}^2 \geq z) - P(\chi_{g_1}^2 \geq z)] + O(m_1^3). \quad (2.22)$$

We next describe the test defined in Srivastava [99] and use it to introduce our newly defined test. Ledoit and Wolf [69] discuss concerns over John's  $U$  statistic and describe how no unbiased tests are known for high-dimensional of the form of  $M(r)/M(t)$ , where  $M(r)$  is the  $r^{\text{th}}$  arithmetic mean, for  $r, t > 0$ . Using this as motivation, we let  $x_1, \dots, x_N$  be iid  $N_p(\mu, \Sigma)$  and  $N = n + 1$ . The covariance matrix,  $\Sigma$ , is assumed to be a positive definite covariance matrix and  $\mu$  is the mean vector. Let  $a_i = (\text{tr}\Sigma^i/p)$ , where  $\Sigma^i$  is a short notation for the matrix multiplication of  $i$   $\Sigma$ s, i.e.  $\Sigma^3 = \Sigma \times \Sigma \times \Sigma$ . Assume the following,

$$(A.a) : \text{As } p \rightarrow \infty, a_i \rightarrow a_i^0, 0 < a_i^0 < \infty, i = 1, \dots, 8.$$

$$(B.a) : n = O(p^\delta), 0 \leq \delta \leq 1,$$

where  $O$  denotes Big-Oh notation.

Like that of the LRT, testing remains invariant under the transformation  $x \rightarrow Gx$ , where  $G$  is an orthogonal matrix. The test is also invariant under the scalar transformation  $x \rightarrow cx$ ; thus we may assume without loss of generality  $\Sigma = \text{diag}(\lambda_1, \dots, \lambda_p)$ . From the Cauchy-Schwarz inequality, it follows that

$$\left( \sum_{i=1}^p \lambda_i^r \times 1^r \right)^2 \leq p \left( \sum_{i=1}^p \lambda_i^{2r} \right)$$

with equality holding if and only if  $\lambda_1 = \dots = \lambda_p = \lambda$  for constant  $\lambda$ . Thus the ratio

$$\psi_r = \frac{\left( \sum_{i=1}^p \lambda_i^{2r} / p \right)}{\left( \sum_{i=1}^p \lambda_i^r / p \right)^2} \geq 1 \quad (2.23)$$

with equality holding if and only if  $\lambda_i = \lambda$ , some constant  $\lambda$ , for all  $i = 1, \dots, p$ . Thus, we may consider testing  $H_0 : \psi_r = 1$  vs  $H_A : \psi_r > 1$ . We note this test is based on the ratio of arithmetic means of the sample eigenvalues.

Srivastava [99] considers the case where  $r = 1$  and finds

$$\hat{\psi}_1 = \frac{n^2}{(n-1)(n+2)} \frac{1}{p} \left[ \text{tr} S^2 - \frac{1}{n} (\text{tr} S)^2 \right] / (\text{tr} S/p)^2 = \frac{\hat{a}_2}{\hat{a}_1^2} \quad (2.24)$$

and provides the distribution under the null and alternative hypothesis of the test statistic under the two assumptions (A.a) and (B.a). We summarize his results from [99].

**Theorem 2.2.1.** *Under assumptions (A.a) and (B.a), as  $(n, p) \rightarrow \infty$  asymptotically*

$$\left( \frac{n}{2} \right) \left( \frac{\hat{a}_2}{\hat{a}_1^2} - \psi_1 \right) \xrightarrow{D} N(0, \xi_1^2)$$

where

$$\xi_1^2 = \frac{2n(a_4 a_1^2 - 2a_1 a_2 a_3 + a_2^2)}{p a_1^6} + \frac{a_2^2}{a_1^4}. \quad (2.25)$$

**Corollary 2.2.1.** *Under the null hypothesis that  $\psi_1 = 1$ ,  $\xi_1^2 = 1$  and under assumptions (A.a) and (B.a), as  $(n, p) \rightarrow \infty$*

$$T_s = \frac{n}{2} \left( \frac{\hat{a}_2}{\hat{a}_1^2} - 1 \right) \xrightarrow{D} N(0, 1) \quad (2.26)$$

Results for the simulated attained significance level, or size, and power are provided in [100] and in our analysis below.

We construct a consistent test in the case of  $r = 2$  to compare with the test results provided in the literature. Consider the adapted assumptions similar to that of Srivastava [99] and Ledoit and Wolf [69],

(A) : As  $p \rightarrow \infty$ ,  $a_i \rightarrow a_i^0$ ,  $0 < a_i^0 < \infty$ ,  $i = 1, \dots, 16$ ,

(B) : As  $n, p \rightarrow \infty$ ,  $\frac{p}{n} \rightarrow c$  where  $0 < c < \infty$ .

Consider the following constants

$$b = -\frac{4}{n}, \quad (2.27)$$

$$c^* = -\frac{2n^2 + 3n - 6}{n(n^2 + n + 2)}, \quad (2.28)$$

$$d = \frac{2(5n+6)}{n(n^2+n+2)}, \quad (2.29)$$

$$e = -\frac{5n+6}{n^2(n^2+n+2)}, \quad (2.30)$$

and

$$\tau = \frac{n^5(n^2+n+2)}{(n+1)(n+2)(n+4)(n+6)(n-1)(n-2)(n-3)}. \quad (2.31)$$

**Theorem 2.2.2.** *An unbiased and  $(n, p)$ -consistent estimator of  $a_4 = \sum_{i=1}^p \lambda_i^4/p$  is given by*

$$\hat{a}_4 = \frac{\tau}{p} \left[ \text{tr}S^4 + b \cdot \text{tr}S^3 \text{tr}S + c^* \cdot (\text{tr}S^2)^2 + d \cdot \text{tr}S^2 (\text{tr}S)^2 + e \cdot (\text{tr}S)^4 \right]. \quad (2.32)$$

*Proof.* The result is provided in Theorems A.2 and A.3 in the Appendix.  $\square$

Thus an  $(n, p)$ -consistent estimator for  $\psi_2$  is given by

$$\hat{\psi}_2 = \frac{\hat{a}_4}{\hat{a}_2^2}$$

where

$$\hat{a}_2 = \frac{n^2}{(n-1)(n+2)p} \left[ \text{tr}S^2 - \frac{1}{n} (\text{tr}S)^2 \right] \quad (2.33)$$

is provided in Srivastava [99]. The derivation and justification for our estimator is provided in the Appendix.

**Theorem 2.2.3.** *Under assumptions (A) and (B), as  $(n, p) \rightarrow \infty$*

$$\left( \frac{n}{\sqrt{8(8+12c+c^2)}} \right) \left( \frac{\hat{a}_4}{\hat{a}_2^2} - \psi_2 \right) \xrightarrow{D} N(0, \xi_2^2)$$

where

$$\xi_2^2 = \frac{1}{(8+12c+c^2)a_2^6} \left( \frac{4}{c}a_4^3 - \frac{8}{c}a_4a_2a_6 - 4a_4a_2a_3^2 + \frac{4}{c}a_2^2a_8 + 4a_6a_3^3 \right. \\ \left. + 8a_2^2a_5a_3 + 4ca_4a_2^4 + 8ca_3^2a_2^3 + c^2a_2^6 \right) \quad (2.34)$$

*Proof.* The result is provided in Theorem A.13 in the Appendix.  $\square$



$p = cn$	$c = 1$	$c = 2$	$c = 4$	$c = 5$
$n = 25$	0.036	0.040	0.050	0.056
$n = 50$	0.050	0.061	0.050	0.058
$n = 100$	0.060	0.052	0.048	0.054
$n = 150$	0.049	0.048	0.049	0.047
$n = 200$	0.047	0.055	0.057	0.051

Table 2.1: ASL for  $T$  in (2.35)

$p = cn$	$c = 1$	$c = 2$	$c = 4$	$c = 5$
$n = 25$	0.050	0.067	0.057	0.057
$n = 50$	0.055	0.049	0.051	0.053
$n = 100$	0.057	0.053	0.056	0.060
$n = 150$	0.054	0.046	0.050	0.040
$n = 200$	0.041	0.043	0.052	0.042

Table 2.2: ASL for  $T_s$  (2.26)

**Corollary 2.2.2.** *Under the null hypothesis,  $\psi_2 = 1$ , and under the assumptions (A) and (B), as  $(n, p) \rightarrow \infty$*

$$T = \left( \frac{n}{\sqrt{8(8 + 12c + c^2)}} \right) \left( \frac{\hat{a}_4}{\hat{a}_2^2} - 1 \right) \xrightarrow{D} N(0, 1) \quad (2.35)$$

*Proof.* The result is provided in Corollary A.2 in the appendix. □

**Theorem 2.2.4.** *Under assumptions (A) and (B), as  $(n, p) \rightarrow \infty$  the test  $T$  is  $(n, p)$ -consistent.*

*Proof.* See Theorem A.14 in the appendix. □

## 2.2.1 Simulation Study

We provide a simulation study to show the effectiveness of our test statistic. We first provide a study to test the normality of our test statistic. We look at the Attained Significance Level (ASL), or simulated size, as well as the QQ-Plot for Normality. We first draw an independent sample of size  $N = n + 1$  from a valid null distribution ( $\lambda = 1$ ). We replicate this 1000 times. Letting

$$T = \left( \frac{n}{\sqrt{8(8 + 12c + c^2)}} \right) \left( \frac{\hat{a}_4}{\hat{a}_2^2} - 1 \right)$$

we calculate

$$ASL(T) = \frac{(\#T > z_\alpha)}{1000}$$

denoting the ASL of  $T$  where  $z_\alpha$  is the upper  $100\alpha\%$  critical point of the standard normal distribution. We test with  $\alpha = 0.05$ . Table 2.1 and 2.2 provide the results for an assortment of  $c = \frac{p}{n}$  values for both our test statistic, and that of Srivastava [99] since it is most similar to our test in construction. Ledoit and Wolf [69] and Srivastava [100] provide analogous results for the  $U_J$  and  $Q_1$  test statistics in (2.19) and (2.20), respectively.

We see the resulting ASL values are comparable between the two tests, and appear to be

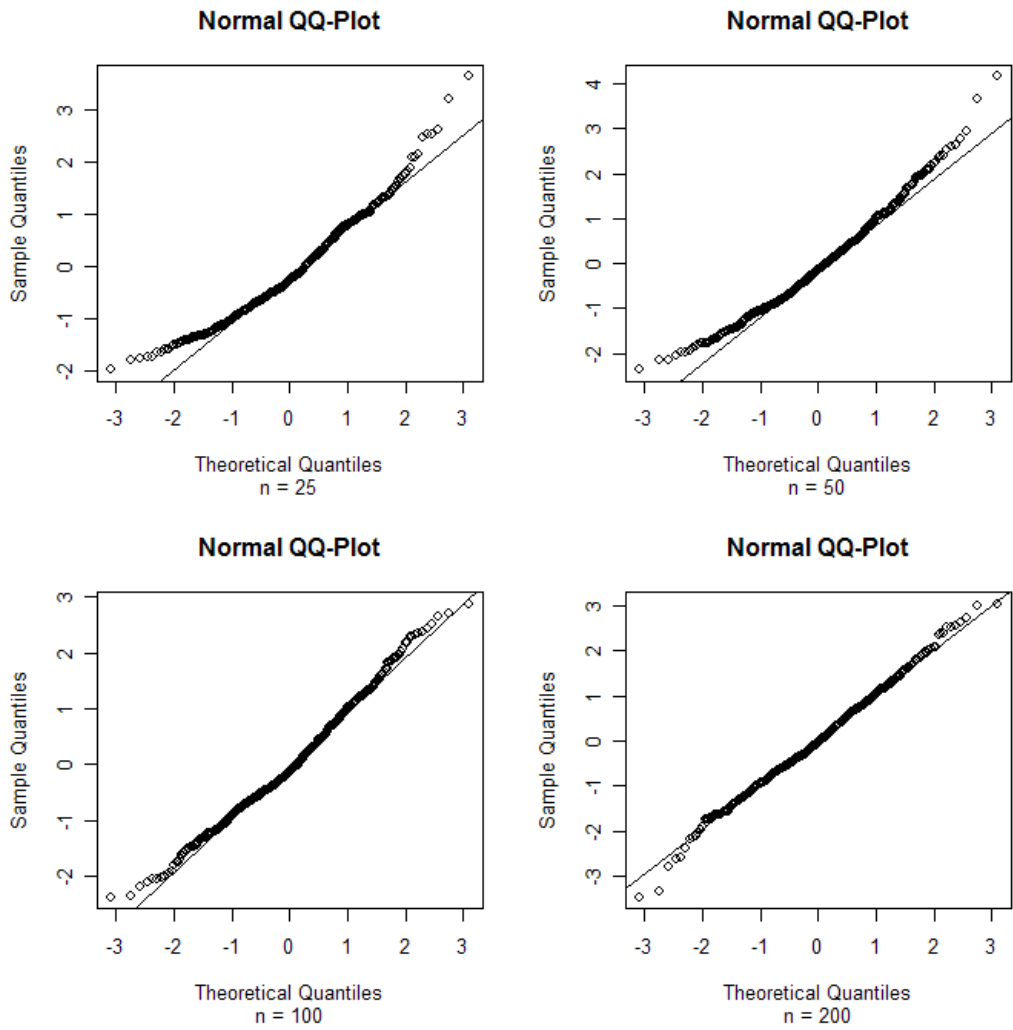


Figure 2.1: Normal QQ-Plot for  $T$  in (2.35) under  $H_0$

approximately normal, particularly when the sample size increases. Next we further confirm the normality result of our test statistic through a series of QQ-Plots. We look at the case of  $c = 2$  so  $p = 2n$ . We draw a random sample of size  $N = n + 1$  under the null distribution with  $\lambda = 1$ , replicate 500 times and plot the Normal QQ plot for the 500 observed values of  $T$ . We repeat the task for  $n = 25, 50, 100, 200$  with respective values of  $p$ . The results are in Figure 2.1.

We see from Figure 2.1 that the test statistic defined in (2.35) appears to be normal as  $(n, p) \rightarrow \infty$  under assumptions (A) and (B). We repeat the same simulation under the alternative distribution with  $\Sigma = \Lambda$  with  $\Lambda$  as a diagonal matrix with eigenvalues that are  $Unif(0.5, 1.5)$ . Figure 2.2 provides the results. We see from those QQ-Plots that under assumptions (A) and (B) and an

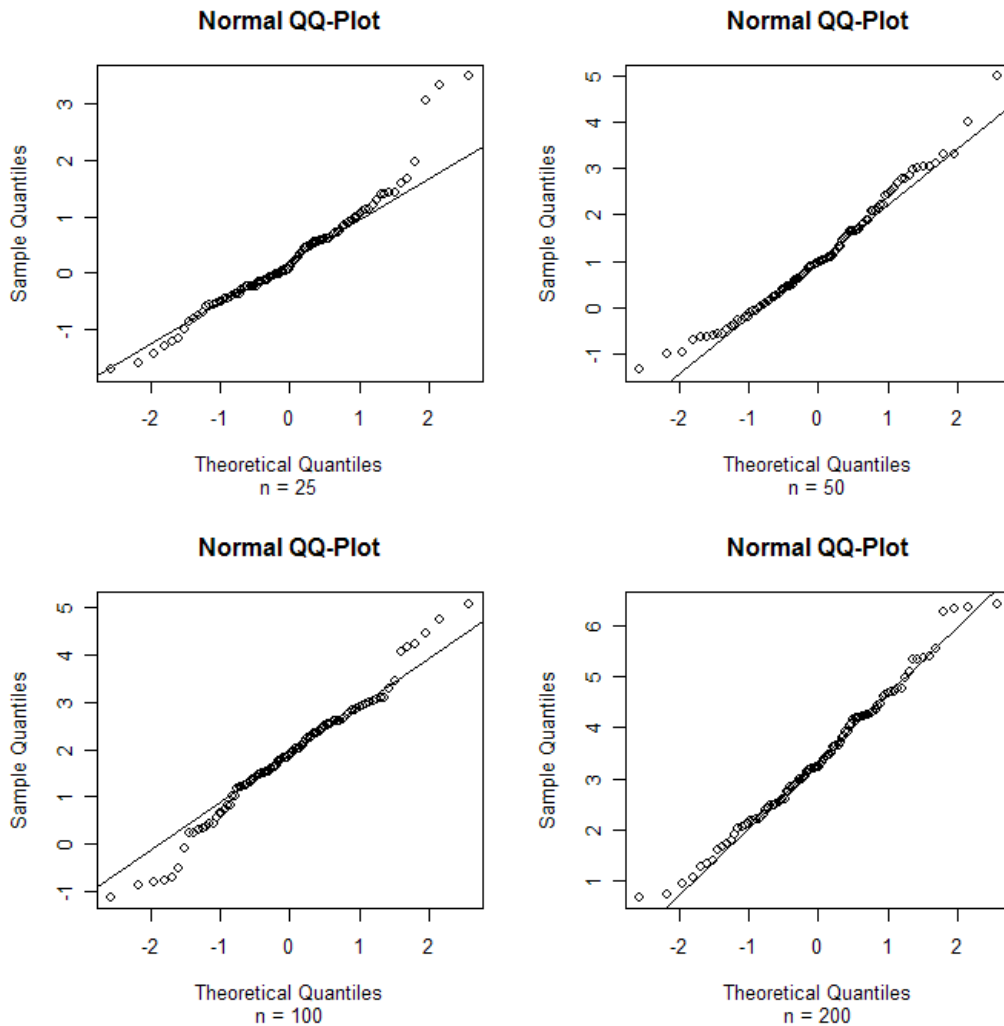


Figure 2.2: Normal QQ-Plot for  $T$  in (2.35) under  $H_A$

alternative hypothesis, the normality result appears to be confirmed as  $(n, p) \rightarrow \infty$ .

We next conduct a series of power simulations. From Theorem 2.2.4, the power of our test should converge to 1 as  $(n, p) \rightarrow \infty$  under assumptions (A) and (B). To make comparisons with Srivastava [100] we perform a similar test. We first carry out a simulation to obtain the critical point of our test statistic and that of Srivastava. We sample  $m = 1000$  observations under  $H_0$  of our test statistic and find  $T_\alpha$  such that

$$P(T > T_\alpha) = \alpha.$$

$p = cn$	$c = 1$	$c = 2$	$c = 4$	$c = 5$
$n = 25$	0.149	0.152	0.108	0.105
$n = 50$	0.308	0.220	0.139	0.140
$n = 100$	0.778	0.631	0.371	0.259
$n = 150$	0.955	0.848	0.504	0.434
$n = 200$	0.997	0.946	0.724	0.594
$n = 300$	1.000	0.999	0.956	0.937

Table 2.3: Simulated Power for  $T$  in (2.35) under  $\Sigma = \Lambda \sim Unif(0.5, 1.5)$

$p = cn$	$c = 1$	$c = 2$	$c = 4$	$c = 5$
$n = 25$	0.317	0.274	0.274	0.317
$n = 50$	0.677	0.744	0.736	0.675
$n = 100$	0.987	0.996	0.990	0.995
$n = 150$	1.000	1.000	1.000	1.000
$n = 200$	1.000	1.000	1.000	1.000
$n = 300$	1.000	1.000	1.000	1.000

Table 2.4: Simulated Power for  $T_s$  in (2.26) under  $\Sigma = \Lambda \sim Unif(0.5, 1.5)$

$p = cn$	$c = 1$	$c = 2$	$c = 4$	$c = 5$
$n = 25$	0.410	0.368	0.214	0.225
$n = 50$	0.849	0.676	0.402	0.358
$n = 100$	1.000	0.993	0.868	0.760
$n = 150$	1.000	1.000	0.986	0.958
$n = 200$	1.000	1.000	1.000	0.999

Table 2.5: Simulated Power for  $T$  in (2.35) under  $\Sigma = \Lambda \sim Unif(0.5, 10.5)$

$p = cn$	$c = 1$	$c = 2$	$c = 4$	$c = 5$
$n = 25$	0.974	0.970	0.962	0.975
$n = 50$	1.000	1.000	1.000	1.000
$n = 100$	1.000	1.000	1.000	1.000
$n = 150$	1.000	1.000	1.000	1.000
$n = 200$	1.000	1.000	1.000	1.000

Table 2.6: Simulated Power for  $T_s$  in (2.26) under  $\Sigma = \Lambda \sim Unif(0.5, 10.5)$

$T_\alpha$  is the estimated critical point at significance level  $\alpha$ . We then simulate again with sample size  $N = n + 1$  drawn from  $N_p(0, \Lambda)$  where  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$ . For comparison to [100] the values of  $\lambda_i$  are obtained by taking  $p$  iid random observations from the uniform distribution over the domain  $(0.5, 1.5)$ . The sample is replicated  $m = 1000$  times and the percentage of times the statistic  $T$  exceeds  $T_\alpha$  is recorded as the simulated power for the statistic. Tables 2.3 and 2.4 provide results for both our test statistic and that defined in Srivastava [99].

We see from tables 2.3 and 2.4 that both test statistics appear to be consistent as  $(n, p) \rightarrow \infty$  under assumptions (A) and (B). We then can conclude the two tests appear to be asymptotically comparable as  $(n, p) \rightarrow \infty$ . It does appear the power of the test defined in Srivastava converges quicker to 1. Discussion of this and other performance behavior is left for the Remarks section below. We also look at a similar power simulation with the eigenvalues  $\lambda_i$  as Uniform random variables over the domain  $(0.5, 10.5)$ . The results are provided in Tables 2.5 and 2.6. These tables show that when the diagonal entries of the covariance matrix have a greater range, in both test, the power converges to 1 quicker.

We next note that both (2.26) and (2.35) are approximately  $Z$ -test for large  $n$  and  $p$ . We further note that if

$$\frac{a_4}{a_2^2} > 1 + \sqrt{2(8 + 12c + c^2)} \left( \frac{a_2}{a_1^2} - 1 \right), \quad (2.36)$$

$p = cn$	$c = 1, \phi = 3$	$c = 2, \phi = 4$
$n = 25$	0.505	0.580
$n = 50$	0.647	0.750
$n = 100$	0.794	0.901
$n = 150$	0.858	0.938
$n = 200$	0.903	0.969

Table 2.7: Simulated Power for  $T$  in (2.35) under *near* Spherical Covariance Matrix

$p = cn$	$c = 1, \phi = 3$	$c = 2, \phi = 4$
$n = 25$	0.427	0.463
$n = 50$	0.489	0.599
$n = 100$	0.529	0.641
$n = 150$	0.565	0.680
$n = 200$	0.624	0.710

Table 2.8: Simulated Power for  $T_s$  in (2.26) under *near* Spherical Covariance Matrix

the test defined in (2.35) should be more powerful than that described in (2.26). One such case is that of a *near* spherical covariance matrix, i.e. covariance matrices of the form

$$\Sigma = \sigma^2 \begin{pmatrix} \phi & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}, \quad (2.37)$$

where  $\sigma^2$  is the unknown *proportion* and  $\phi$  is a non-one constant. In general, we can assume the covariance to be the identity with the exception of one element,  $\phi \neq 1$ . We perform a power simulation on *near* spherical matrices as defined in (2.37). The results in Tables 2.7 and 2.8 suggest our test statistic is more powerful than that described in Srivastava [99] for these *near* spherical matrices.

## 2.2.2 Data Examples

In this section, for further comparison we test the hypothesis of sphericity against the following *classic* data sets.

### 2.2.2.1 Colon Datasets

In this dataset, expression levels of 6500 human genes are measured using Affymetrix microarray technology on 40 tumors and 22 normal colon tissues. A selection of 2000 genes with the highest minimal intensity across the samples has been made by Alon, Barkai, Notterman, Gishdagger, Mack and Levine [2]. Our dimensionality,  $p = 2000$  and the degrees of freedom available to estimate the covariance matrix is only 60. The data is further described and is available at “<http://microarray.princeton.edu/oncology/affydata/index.html>”. A base-10 logarithmic

transformation is applied.

### 2.2.2.2 Leukemia Datasets

This dataset contains gene expression levels of 72 patients either suffering from acute lymphoblastic leukemia (ALL) or acute myeloid leukemia (AML). There are 47 and 25 patients for each respective case and they are obtained on Affymetric oligonucleotide microarrays. The data is attributed to Golub, Slonim, Tamayo, Huard, Gaasenbeek, Mesirov, Coller, Loh, Downing, Caligiuri, Bloomfield, and Lander [49]. We follow the preprocessing protocol attributed to Dudoit, Fridlyand, and Speed [28] and Dettling and Bühlmann [24] by a thresholding, filtering, a logarithmic transformation but do not follow standardization as to compare to the results in Srivastava [100]. The data is finally comprised of  $p = 3571$  genes and the degrees of freedom available is only 70. The data is available and described further at “<http://www.broad.mit.edu/science/data>”.

### 2.2.2.3 Test Performance

Both sets of the preprocessed data are available at the website of Prof. Tatsuya Kubokawa, see “<http://www.e.u-tokyo.ac.jp/~tatsuya/index.html>”. In each of the datasets, we treat them as two samples. In the colon data, we assume

$$X_{tumor} \sim N(\mu_{tumor}, \Sigma), \quad (2.38)$$

$$X_{tissue} \sim N(\mu_{tissue}, \Sigma). \quad (2.39)$$

We treat each sample individually, centralizing each  $X_{tumor}$  and  $X_{tissue}$  by their respective sample means. We then combine back into one sample of size  $N = 62$  and computing the sample covariance matrix  $S$  with  $n = 60$  degrees of freedom. We note this is equivalent to treating each of the samples individually and finding the pooled covariance matrix as an estimate for  $\Sigma$ .

For the colon data, we get test values of  $T = 185.8071$ ,  $T_s = 2771.6538$ ,  $Q_1 = 82086.3214$  and  $U_J = 2816.2916$  for (2.35), (2.26), (2.20) and (2.19) respectively. The leukemia data is treated the same way (as two samples, with  $n = 70$  degrees of freedom) and we get the value observed test values  $T = 242.4386$ ,  $T_s = 2294.9184$ ,  $Q_1 = 86120.8290$  and  $U_J = 2326.7520$  for (2.35), (2.26), (2.20) and (2.19) respectively. We note in each case we get a p-value=0 indicating any assumption of sphericity in the case of these datasets to be false.

## 2.2.3 Remarks and further Development

We have proposed a new test for sphericity of the covariance matrix. Like that of Srivastava [99], our test is based on the Cauchy-Schwartz inequality. Unlike John's U-statistic and Srivastava's  $T_s$  test, we look at the second and fourth arithmetic means of the sample eigenvalues, as compared to the first and second. The newly defined test statistic  $T$  (2.35) appears to perform better in some *near* spherical cases and is asymptotically comparable.

### 2.2.3.1 Changes in Assumptions

Of all the test discussed, the test discussed in [99] has the weakest assumptions in terms of the growth of  $p$  and  $n$ . Srivastava simply requires  $n = O(p^\delta)$  for some  $0 < \delta \leq 1$ . Our newly defined test, and that of Ledoit and Wolf [69], require  $p/n \rightarrow c$ , for some finite concentration such that  $0 < c < \infty$ . Although this is a stricter assumption, we note that in practice the concentration  $c$  is easily estimated by  $p/n$ . For example, in both of the analyzed data sets,  $c$  is approximated by  $2000/60 = 33.333$  and  $3571/70 = 51.01428$  respectively for the colon and leukemia datasets. The concentration is required for the asymptotic result but should not hinder any application of these test statistics.

The U-test of John only requires convergence of the fourth arithmetic mean of the eigenvalues of  $\Sigma$ . The test described in Srivastava [99] requires convergence of the eighth arithmetic mean. We require convergence up to the sixteenth arithmetic mean. We found no violations of this assumption during preliminary simulation tests, however we note convergence of higher order means may be infeasible in certain problems.

### 2.2.3.2 Limitations

We discovered several limitations in the test defined in (2.35) and, in general, tests of the form (2.23). When  $r$  is limited to the positive integers, as  $r$  increases, we require higher and higher arithmetic means of eigenvalues to converge. At  $r = 1$ , Srivastava [99] requires up to the eighth arithmetic mean to converge. At  $r = 2$ , we require up to the sixteenth. Continuing with this methodology, a test based on  $r = 3$  would require up to the thirty-second arithmetic mean to converge. It can easily be seen that as  $r$  increases, the assumption of the convergence in arithmetic mean may become infeasible.

We further note the increase in the variance of our test statistic, compared to that of Srivastava [99]. As we look at higher arithmetic means, the variance of our test statistic increases. This can be seen as (2.25) < (2.34) in general. The power of the test defined in (2.35) appears to converge slower than that described in [99]. Although the two tests are asymptotically equivalent, and the newly defined test in (2.35) appears to outperform that of Srivastava in *near* spherical cases of  $\Sigma$ , the larger variance of  $T$  may be a problem in certain cases.

### 2.2.3.3 Future Work and Recommendations

We have defined a new test of the form (2.23) with  $r = 2$ . This builds upon the work of Srivastava who defined a test based on  $r = 1$ . Future work may look at  $r = 3, 4, \dots$ . We conjecture that these tests will be more powerful than those already defined in certain cases of  $\Sigma$ , however we also note these tests will make more restrictive assumptions and the variance of the test statistic will grow to the point where it may be infeasible to use the statistic. In the case of  $r$  being a fraction, e.g.  $r = 0.5$ , we suspect the test may be more powerful in some cases of  $\Sigma$  and in general will not be hindered by assumptions and a large variance. However, the distribution of terms like  $a_{\frac{1}{2}}$  is difficult to determine and we leave this question open.

Although each of the tests described is asymptotically equivalent, each test seems to perform better under certain circumstances. We recommend our newly defined test,  $T$  (2.35), when a *near* spherical covariance matrix is suspected. In general, the test of [99] or that based on John's  $U$  statistics as defined in Ledoit and Wolf [69], do not appear to be hindered by larger variance terms and are comparable in performance.

## 2.3 Testing for Identity Covariance Matrix

We now explore the results in the literature, namely Ledoit and Wolf [69] and Srivastava [99] [100], in testing whether the covariance matrix equals a  $p \times p$  identity matrix, i.e.  $H_0 : \Sigma = I$  versus  $H_A : \Sigma \neq I$ . We note this is a special case of the sphericity test when  $\sigma^2 = 1$ . Under the classical  $n$ -asymptotics we typically will look at the test statistic based on the likelihood ratio criteria that can be found in Anderson [3], et. al. Let,

$$\lambda_1^* = e^{\frac{1}{2}pn} \left( |S| e^{-\text{tr}(S)} \right)^{\frac{1}{2}n} \quad (2.40)$$



and the distribution, as  $n \rightarrow \infty$ , can be found by

$$P(-2\rho \log \lambda_1^* \leq x) = P(\chi_f^2 \leq z) + \frac{\gamma_2}{\rho^2 N^2} (P(\chi_{f+4}^2 \leq z) - P(\chi_f^2 \leq z)) + O(N^{-3}) \quad (2.41)$$

where

$$\begin{aligned} \rho &= 1 - \frac{2p^2 + 3p - 1}{6N(p+1)}, \\ \gamma_2 &= \frac{p(2p^4 + 6p^3 + p^2 - 12p - 13)}{288(p+1)}, \\ f &= \frac{1}{2}p(p+1). \end{aligned}$$

Nagao proposes an invariant test based on the eigenvalues of  $S$ . Consider,

$$V = \frac{1}{p} \text{tr} [(S - I)^2] \quad (2.42)$$

which Nagao [75] derived from John's  $U$  statistic for the case of  $\Sigma = I$ . He shows under the null hypothesis, as  $n \rightarrow \infty$ ,  $V$  has a limiting  $\chi^2$ -distribution with  $\frac{1}{2}p(p+1)$  degrees of freedom.

Ledoit and Wolf [69] propose a test statistic based on  $V$  that has desirable  $(n, p)$ -asymptotic properties. Let

$$W = \frac{1}{p} \text{tr} [(S - I)^2] - \frac{p}{n} \left[ \frac{1}{p} \text{tr}(S) \right]^2 + \frac{p}{n}. \quad (2.43)$$

Under the same assumptions as for the convergence of the  $U$  statistic, namely  $p/n \rightarrow c < \infty$ , normality of the data, and convergence of the third and fourth arithmetic means of the eigenvalues of  $\Sigma$ , they show under the null hypothesis of  $\Sigma = I$ ,

$$\frac{nW - p - 1}{2} \xrightarrow{D} N(0, 1) \text{ as } (n, p) \rightarrow \infty. \quad (2.44)$$

They show the test based on  $W$  to be equivalent to Nagao's test based on  $V$  in the classical  $n$ -consistency sense. We note that the distribution under the alternative hypothesis is unavailable. Birke and Dette [11] look at the test statistics studied by Ledoit and Wolf [69] and study the behavior when  $c \rightarrow 0$  or  $c \rightarrow \infty$ .

Srivastava [100] proposes a test based on the modified likelihood ratio test except under the

case when  $n < p$ , consider

$$L_2 = \left(\frac{e}{p}\right)^{\frac{1}{2}pn} \left(\prod_{i=1}^n \tilde{l}_i\right)^{\frac{1}{2}p} e^{\frac{1}{2} \sum_{i=1}^n \tilde{l}_i} \quad (2.45)$$

where  $\tilde{l}_i$  are the eigenvalues of  $nS$ . Let

$$\begin{aligned} g_2 &= \frac{1}{2}n(n+1), \\ m_2 &= p - \frac{2n^2 + 3n + 1}{6(n+1)}, \\ c_2 &= \frac{n}{288(n+1)}(2n^4 + 6n^3 + n^2 - 12n - 13), \\ Q_2 &= -\left(\frac{2m_2}{p}\right) \log L_2. \end{aligned}$$

Srivastava shows under the null hypothesis  $\Sigma = I$ , when  $n/p \rightarrow 0$ , the asymptotic distribution of  $Q_2$  is given by

$$P(Q_2 \geq z) = P(\chi_{g_2}^2 \geq z) + c_2 m_2^{-2} (P(\chi_{g_2+4}^2 \geq z) - P(\chi_{g_2}^2 \geq z)) + O(m_2^{-3}). \quad (2.46)$$

In Srivastava [99], he proposes a test based on the sum of the eigenvalues, if  $\lambda_i = 1$  for all  $i$ , then

$$\frac{1}{p} \sum_{i=1}^p (\lambda_i - 1)^2 = 0$$

thus

$$\begin{aligned} \frac{1}{p} \sum_{i=1}^p (\lambda_i - 1)^2 &= \frac{1}{p} \left[ \sum_{i=1}^p \lambda_i^2 - 2 \sum_{i=1}^p \lambda_i + p \right] \\ &= \frac{1}{p} [\text{tr} \Sigma^2 - 2\text{tr} \Sigma + p] \\ &= a_2 - 2a_1 + 1 = 0 \end{aligned}$$

where  $a_2$  and  $a_1$  are defined as the second and first arithmetical means of the eigenvalues of  $\Sigma$ . As previously discussed in Section 2.2, Srivastava finds  $(n, p)$ -consistent estimators for  $a_2$  and  $a_1$ , see (2.33) and Srivastava [99]. Under the assumptions (A.a) and (B.a) from Section 2.2, he derives the distribution of the test statistic

$$T_s^{**} = \hat{a}_2 - 2\hat{a}_1 + 1, \quad (2.47)$$

as

$$\binom{n}{2} (T_s^{**} - (a_2 - 2a_1) - 1) \xrightarrow{D} N(0, \xi_*^2),$$

where

$$\xi_*^2 = \frac{2n}{p}(a_2 - 2a_3 + a_4) + a_2^2$$

as  $(n, p) \rightarrow \infty$ . Under the null hypothesis, as  $(n, p) \rightarrow \infty$

$$\binom{n}{2} T_s^{**} \xrightarrow{D} N(0, 1).$$

The test based on  $T_s^{**}$  is a one-sided Z-test for large  $(n, p)$ .

We note that using methodology similar to that of Section 2.2, we can propose a new test based on,

$$\begin{aligned} \frac{1}{p} \sum_{i=1}^p (\lambda_i^2 - 1)^2 &= \frac{1}{p} \left[ \sum_{i=1}^p \lambda_i^4 - 2 \sum_{i=1}^p \lambda_i^2 + p \right] \\ &= \frac{1}{p} [\text{tr} \Sigma^4 - 2 \text{tr} \Sigma^2 + p] \\ &= a_4 - 2a_2 + 1 = 0. \end{aligned}$$

Consistent estimators of  $a_4$  and  $a_2$  are derived in the section on Sphericity. However it was noted that the variance term for  $\hat{a}_4$  is quite large, and can hinder the performance of our newly defined test statistic. However, that test seemed appropriate in the case of *near* spherical covariance matrices. The same methodology can be implemented here to define a new test statistic, however we question whether there will be any statistical advantage in this newly defined test. Additional work is necessary to determine if test-statistics involving the fourth, or other arithmetic, means of the sample eigenvalues should be pursued.

## 2.4 Concluding Remarks

In this work, we have studied three common hypothesis tests for the covariance matrix in the framework of general asymptotics. We began by highlighting the previous work in the literature describing test for complete independence, or that the covariance matrix has a diagonal structure. We proposed a generalization of the test Srivastava [99] proposed based on the sample correlation

matrix. We highlighted the work completed on testing for sphericity of the covariance matrix, and proposed a new test based on the fourth and second arithmetic means of the eigenvalues of  $\Sigma$ . Monte Carlo simulations confirm our asymptotic results and we apply the newly defined test against large dimensional microarray data. Our newly defined test is comparable to that in the literature and appears to perform better when the covariance matrix has a *near* spherical structure. Lastly we study the existing test for an Identity covariance matrix and discuss a proposal for a new test based on the fourth and second arithmetic mean. The distribution under the alternative and null hypothesis can be explicitly calculated using the methodology described in the Appendix. Due to the high variance of our unbiased and consistent estimator of  $a_4$ , at this time, we cannot justify the construction of such a test. Future work may explore the usefulness of a test based on the fourth and second arithmetic moments. Lastly we note a key contribution of this work and that of Srivastava [99]: Unlike the previous work, the asymptotic alternative distribution is available for the sphericity test and identity test defined.

Directions for future work include: exploration into test statistics involving higher order arithmetic means for sphericity and identity, searching for the most powerful tests (within specific alternative frameworks, e.g. *near* spherical), relaxing the normality assumption, exploration of the rate of convergence.

## Chapter 3

# Estimation

We discuss the advances in the estimation of the covariance matrix  $\Sigma$ . We begin by describing the *classical* approach and its limitations. Typically,  $\Sigma$  is estimated with its empirical counterpart, the sample covariance matrix (1.2). Given a sample of size  $N = n + 1$  and dimension  $p$ , we find

$$S = \frac{1}{n}(X - \bar{X})(X - \bar{X})'$$

where we have dropped the bold notation indicating matrices.  $S$  is a  $p \times p$  semi-positive definite matrix. If  $n > p$ ,  $S$  is positive definite with probability 1, and hence is nonsingular.  $S$  is based on the maximum likelihood estimator and is unbiased for  $\Sigma$ . In this classical case,  $S$  is also consistent for  $\Sigma$  as  $n \rightarrow \infty$  with  $p$  fixed. The eigenvalues of  $S$  are good estimates of the eigenvalues of  $\Sigma$ . Since  $S$  is nonsingular,  $S^{-1}$  can be used to estimate the precision matrix  $\Sigma^{-1}$ . In the classical case,  $S$  is typically considered the *best* estimator for  $\Sigma$ .

Although  $S$  has several desirable properties, in practice the sample covariance matrix has some limitations. In many applications an estimator for the precision matrix is needed. When  $p > n$ ,  $S$  is singular and cannot be inverted. As  $n \approx p$ ,  $S$  becomes ill-conditioned and near-singular. Furthermore as  $p \rightarrow n$  or  $p > n$ , the eigenvalues of  $S$  disperse from the true eigenvalues of  $\Sigma$ , see Ledoit and Wolf [72], Schäfer and Strimmer [85], Bai, Silverstein and Yin [6], Bai and Yin [7], Bickel and Levina [10] and others for details of sample eigenvalue dispersion. In many applications of modern statistical methods, such as economics, portfolio selection (Ledoit and Wolf [70] [71], Frost and Savarino [41], Jorion [64]), clustering of genes from microarray data (Eisen, et al. [36]), gene

association networks (Toh and Horimoto [105], Dobra et al. [26]) and other applications where the dimensionality is large, an accurate estimate of the covariance matrix or precision matrix is needed. Besides an increase in efficiency and accuracy, it would be desirable that estimates of the covariance matrix (and precision matrix) should exhibit two characteristics not always found in  $S$ : they should be (1) positive definite and (2) well-conditioned.

### 3.1 Estimation Techniques

In the contents of estimating the mean vector of a multivariate normal distribution, Stein [101] demonstrated that we can improve upon the maximum likelihood estimator when the dimensionality,  $p$ , is non-negligible. Stein [102] and James and Stein [61] explore the idea of improving upon the MLE further. More insight into the so called “Stein-phenomenon” or “Stein-affect” on the performance of the MLE is provided in Efron [32].

We will highlight some of the historical approaches used to improve estimation of the covariance matrix. Stein [103] proposes an estimator that keeps the eigenvectors of the sample covariance matrix but replaces its eigenvalues. Isotonic regression is applied before recombining the *corrected* eigenvalues with the eigenvectors to ensure they are positive. Dey and Srinivasan [25] derive an estimator that is minimax under certain loss functions. It scales the sample eigenvalues by a constant. Haff [51] [54] introduced a new type of estimator for the precision matrix based on two different loss functions based on an identity for Wishart distribution [52] [55]. Efron and Morris [34] and Yang and Berger [111] use a Bayesian approach to estimate the precision matrix and covariance matrix, respectively. The list of literature that much of the current work is indebted cannot be justly summarized here. The basic technique was to reduce  $L(\hat{\Sigma}, \Sigma)$  (or  $L(\hat{\Sigma}^{-1}, \Sigma^{-1})$ ), for some loss-function  $L$ , compared to  $L(S, \Sigma)$  (or  $L(S^{-1}, \Sigma^{-1})$ ), and find an appropriate estimator meeting the reduction criterion. Various loss functions and methodology have been explored and the work continues. We now explore some of the more recent advances.

We briefly describe some of the work that does not fall into the “shrinkage” realm. Most of these techniques assume some structure in the true covariance matrix and exploit it for a better estimator.

Eaton and Olkin [30], Pourahmadi [77] [78] [79], Pourahmadi, Daniels and Park [80], Huang, Liu, Pourahmadi and Liu [60], Chen and Dunson [18], Roverato [82] and Smith and Kohn [96] use a

Cholesky decomposition approach where they decompose the covariance matrix. The decomposed model is then estimated using one of several methods: maximum likelihood, a penalized likelihood approach, simultaneous estimation, minimizing of a risk function, a Bayesian approach, or iterative techniques such as expectation-maximization.

Fan, Fan and Lv [37] assume the covariance to follow the form of a factor model and find an appropriate estimator using multivariate factor analysis techniques. Chaudhuri, Drton and Richardson [16] provide an algorithm for computing the maximum likelihood estimate for the covariance matrix under the constraint that certain covariance terms are zero. Cao and Bouman [12] estimate  $\Sigma$  based on the modeling assumption that the eigenvalues of  $\Sigma$  are represented by a sparse matrix transform. Bickel and Levina [10] assume banding, or tapering, of the covariance matrix. They constrain the estimator for  $\Sigma$  and  $\Sigma^{-1}$  by banding it. The level of banding, or tapering, is chosen by minimizing the risk with respect to the  $L_2$  matrix norm. Werner, Jansson and Stoica [109] estimate covariance matrices when they are assumed to have a Kronecker product structure, which is common in wireless communication channels and signal processing. Furrer [42] estimates the covariance matrix of a correlated multivariate spatial process. Furrer and Bengtsson [43] provide estimators for the covariance matrix based on Monte-Carlo Kalman filter variants.

Several authors explore estimation of the covariance matrix with applications to longitudinal data. Wu and Pourahmadi [110] propose a nonparametric estimator for the covariance matrix. Huang, Liu and Liu [59] propose a estimator for the covariance matrix based on a smoothing-based regulation and the use of the modified Cholesky-decomposition. Pourahmadi [77] explores joint mean-variance models. Fan and Wu [38] propose a semiparametric technique where the variance function is estimated nonparametrically and the correlation function is parametrically estimated. Daniels [20] and Daniels and Pourahmadi [23] use a Bayesian approach by exploring various priors for dependent data. Sun, Zhang and Tong [104] propose a random effect varying-coefficient model and propose an estimation procedure for the model.

Kubokawa and Srivastava [68] estimate the precision matrix under different loss functions utilizing the Moore-Penrose inverse of the sample covariance matrix by establishing a Stein-Haff identity for the singular Wishart distribution. Tsukuma and Konno [106] estimate the precision matrix by adapting several estimates for the covariance matrix under the squared loss function.

Champion [15] derives empirical Bayes estimators of normal variances and covariances using an inverse Wishart prior and the Kullback-Leibler distance for loss. Haff [56] provides a general rep-

resentation for the formal Bayes estimator of the mean vector and covariance matrix by minimizing the formal Bayes risk. Daniels [21] proposes several estimators of the covariance matrix based on several common Bayesian regression models.

Hemmerle and Hartley [57] show how to compute the objective function and its derivatives for the maximum likelihood estimation of variance components. This work was later extended by Corbeil and Searle [19] for restricted maximum likelihood estimation. Fellner [39] discusses the use of sparse matrices in expectation-maximization algorithms. Fraley and Burns [40] show how to compute the likelihood functions and derivatives via sparse matrix methods, generalizing the overall results.

This highlights some of the recent advances in estimation of the covariance matrix and precision matrix. However, we have ignored an entire class of estimators for the covariance matrix, and the precision matrix, typically called “Stein-Type” shrinkage, or just Shrinkage estimators. These are discussed in the next section.

## 3.2 Stein-Type Shrinkage Estimators

A common approach to improve the estimator of the covariance is that of “shrinking” or “biased estimation”. This was explored extensively by Efron [31] and Efron and Morris [33] [35]. Kubokawa [67] provides a review on the subject. We begin by recalling the bias-variance decomposition of the mean squared error for the sample covariance,

$$\text{MSE}(S) = \text{Bias}(S)^2 + \text{Var}(S). \tag{3.1}$$

For the sample covariance matrix,  $S$ , the bias is zero, hence the only way to improve on the accuracy of  $S$  is by reducing its variance. The basic idea of a shrinkage estimator is a trade-off between the error due to bias and the error due to variance. The error due to  $S$  is all variance. By introducing a biased estimator, we can actually reduce the variance and improve the mean squared error of our estimator. This is the central idea to the James and Stein [61] approach. We consider a convex combination of the empirical sample covariance matrix with that of a target matrix,

$$\lambda T + (1 - \lambda)S, \tag{3.2}$$



where  $\lambda \in [0, 1]$  is known as the shrinkage *intensity* and  $T$  is a shrinkage *target* matrix.  $T$  is chosen to have several properties we want in our estimator.  $T$  should be structured, positive definite, well-conditioned and represent the true covariance matrix for our application. We note that  $S$  is unbiased but with high variance, and  $T$  will be biased, but with its well-defined structure will have low (or possibly no) variance due to the number of free parameters. The intensity  $\lambda$  is then found to produce an estimator to improve on the estimator  $S$ . If  $n$  is large compared to  $p$ ,  $S$  will have a smaller variance and should be a good estimator, hence  $\lambda \rightarrow 0$ . Likewise, if  $p$  is large,  $S$  will have a greater variance, and more weight should be applied to the target, i.e.  $\lambda \rightarrow 1$ . Historically,  $\lambda$  is chosen via a bootstrap method, cross-validation, or Markov Chain Monte Carlo (MCMC) methods.

The shrinkage approach can be applied to any estimator for a parameter. Ghosh, Mergel and Datta [44] study the Stein-phenomenon when looking at the mean. Warton [108] shrinks the sample correlation matrix,  $R$ , to the identity matrix by maximizing cross-validation. Daniels and Kass [22] use a Bayesian approach and shrink the correlations and perform a Givens rotation to produce a shrinkage estimator for the covariance matrix. Barnard, McCulloch and Meng [9] model a covariance matrix in terms of its corresponding standard deviations and correlation matrix. Several choices for priors are presented and a straightforward computational strategy is presented for obtaining the posterior of the covariance matrix. Srivastava and Bilodeau [98] propose a Stein-estimator for the least squares regressors in the general linear regression model. Kubokawa [66] performs a double shrinkage algorithm on the regression estimators in generalized least squares. Ahmed [1] looks at a Stein-type shrinkage estimator for the eigenvalues of the covariance matrix.

In the remainder of this chapter, we extensively work with shrinkage estimators of the covariance matrix, with respect to minimizing the expected quadratic loss, or risk function. Ledoit and Wolf [70] provide a theorem that expresses the optimal shrinkage intensity in terms of the variance and covariance of the sample matrix and the target matrix. Under the expected quadratic loss with respect to the Frobenius norm, this optimal intensity will always exist. Haff [53] explored shrinkage estimates of this type and under this loss function, but did not provide any optimal results, his coefficients do not depend on the observations of  $X$ . The optimal intensity must be estimated from the data. Ledoit and Wolf [70] and [71] provide shrinkage estimators for application of portfolio selection in economics. In Ledoit and Wolf [70], the target is a single index model of historical market returns. In [71] they look at a constant-correlation model. Each of these targets are described in the respective papers and highlighted in Schäfer and Strimmer [85]. Both tend to be positively

correlated which matches the historical market data. Sancetta [83] applies a shrinkage estimate to a time-series model. The result is similar to that of Ledoit and Wolf [72] except the target matrix has additional structure for dependent data. Typically the target matrix should be well-structured and have few free parameters. Details for three popular targets is provided below.

### 3.2.1 Shrinkage to diagonal, common variance

We first explore the work in the literature when the target matrix,  $T$ , is of the form  $T = vI$ . That is,  $T$  is a diagonal matrix with a common variance  $v$ . The target will contribute very little variance to the estimator since there is only one-free parameter. The shrinkage estimator is written explicitly as

$$S^* = \lambda vI + (1 - \lambda)S. \quad (3.3)$$

Ledoit and Wolf [72] find the optimal common value,  $v$ , and shrinkage intensity,  $\lambda$ , by exploring the objective function in a finite sample and then minimizing the quadratic risk. Following Leung and Muirhead [74] they consider the Frobenius norm:  $\|A\| = \sqrt{\text{tr}(AA')/p}$ . The  $1/p$  is included so the norm of the identity matrix will be 1. Let  $X$  denote a  $p \times n$  matrix of  $n$  independent and identically distributed observations on a system of  $p$  random variables having mean zero and covariance  $\Sigma$  with finite fourth moments. They find a Stein-type shrinkage estimator of the form  $\Sigma^* = \lambda vI + (1 - \lambda)S$ , where  $I$  is the identity matrix and  $S = XX'/n$  is the sample covariance matrix, that minimizes the expected quadratic loss  $E[\|\Sigma^* - \Sigma\|^2]$ . We begin by highlighting their work in detail.

#### 3.2.1.1 Analysis in Finite Sample

The squared Frobenius norm  $\|\cdot\|^2$  is a quadratic form whose associated inner product is  $\langle A_1, A_2 \rangle = \text{tr}(A_1 A_2')/p$ . We define the four scalars:  $\mu = \langle \Sigma, I \rangle$ ,  $\alpha^2 = \|\Sigma - \mu I\|^2$ ,  $\beta^2 = E[\|S - \Sigma\|^2]$ , and  $\delta^2 = E[\|S - \mu I\|^2]$ . We note the fourth moments of  $X$  are needed so  $\beta^2$  and  $\delta^2$  are finite. As shown in the Lemma 2.1 in Ledoit and Wolf [72], we have the following decomposition,  $\delta^2 = \alpha^2 + \beta^2$ . The Lemma allows us to solve for the optimal values for  $v$  and  $\lambda$  explicitly. A calculus based minimization of the objective function,  $E[\|\Sigma^* - \Sigma\|^2]$ , with respect to  $\lambda$  and  $v$  provide the optimal values,  $v = \mu = \text{tr}\Sigma/p$  and  $\lambda = \beta^2/(\alpha^2 + \beta^2) = \beta^2/\delta^2$ . It is then noted that unfortunately  $\Sigma^* = \frac{\beta^2}{\delta^2}\mu I + \frac{\alpha^2}{\delta^2}S$  is not a *bona fide* estimator since it depends on knowledge of the covariance matrix  $\Sigma$ . They then develop an estimator with the same properties as  $\Sigma^*$  asymptotically as the

number of observations and the number of variables go to infinity together. We provide an alternative estimator for  $\Sigma^*$  with similar assumptions that is comparable under general-asymptotics.

### 3.2.1.2 Ledoit and Wolf Estimator

Ledoit and Wolf [72] find a consistent estimator, with respect to quartic mean, for  $\Sigma^*$  under the following assumptions. Let  $n = 1, 2, \dots$  index a sequence of statistical observations. For every  $n$ ,  $X_n$  is a  $p_n \times n$  matrix of  $n$  iid observations on a system of  $p_n$  random variables with mean zero and covariance matrix  $\Sigma_n$ . The subscript  $n$  is supplied to indicate we are exploring asymptotically. The number of variables  $p_n$  can change and even go to infinity with the number of observation  $n$ , as long as we satisfy the below assumptions. We first note we can decompose the covariance matrix as follows:  $\Sigma_n = \Gamma_n \Lambda_n \Gamma_n'$  where  $\Lambda_n$  is a diagonal matrix of eigenvalues, and  $\Gamma_n$  an orthogonal matrix of eigenvectors.  $Y_n = \Gamma_n' X_n$  is a  $p_n \times n$  matrix of  $n$  iid observations on a system of  $p_n$  uncorrelated and mean zero random variables that span the same space as the original system. Let  $(y_{1i}^n, y_{2i}^n, \dots, y_{p_n i}^n)'$  denote the  $i^{\text{th}}$  column of the matrix  $Y_n$ . They make the following additional assumptions.

Assumption 1: There exists a constant  $K_1$  independent of  $n$  such that  $\frac{p_n}{n} \leq K_1$

Assumption 2: There exists a constant  $K_2$  independent of  $n$  such that  $\frac{1}{p_n} \sum_{i=1}^{p_n} E[(y_{i1}^n)^8] \leq K_2$

Assumption 3:  $\lim_{n \rightarrow \infty} \frac{p_n}{n} \times \frac{\sum_{(i,j,k,l) \in Q_n} (\text{Cov}[y_{i1}^n y_{j1}^n, y_{k1}^n y_{l1}^n])^2}{|Q_n|} = 0$

where  $|Q_n|$  is the cardinality of  $Q_n$ , the set of all quadruples that are made of four distinct integers between 1 and  $p_n$ .

We note that Assumption 1 includes standard asymptotics where  $p_n$  will stay fixed. Furthermore it does not require the ratio  $p_n/n$  to converge, merely to be bounded. Assumption 2 states that the eighth moment is bounded, on average. Assumption 3 states the products of uncorrelated random variables are uncorrelated, on average and in the limit.

Under the three assumption, Ledoit and Wolf [72] provide several asymptotic results and suggest estimators for the asymptotic components of  $\alpha_n^2$ ,  $\beta_n^2$ ,  $\delta_n^2$  and  $\mu_n$ , where the subscript  $n$  indicates that the results will hold asymptotically.  $m_n = \langle S_n, I_n \rangle_n = \text{tr}(S_n)/p$ , where  $S_n = X_n X_n' / n$  and  $I$  is the  $p_n \times p_n$  identity matrix. They show, as  $n \rightarrow \infty$ ,  $m_n \xrightarrow{q.m.} \mu_n$ . Likewise,  $d_n^2 = \|S_n - m_n I_n\|^2$

will converge, in quartic mean, to  $\delta_n^2$  as  $n \rightarrow \infty$ . Similarly, if we let  $x_k^n$  denote the  $p_n \times 1$   $k^{\text{th}}$  column of the observation matrix  $X_n$ , for  $k = 1, \dots, n$ . We define  $\tilde{b}_n^2 = \frac{1}{n^2} \sum_{k=1}^n \|x_k^n (x_k^n)'\|_F^2$  and  $b_n^2 = \min(\tilde{b}_n^2, d_n^2)$ , then  $\tilde{b}_n^2 \xrightarrow{q.m.} \beta_n^2$  and  $b_n^2 \xrightarrow{q.m.} \beta_n^2$ . The constrained estimator  $b_n^2$  is included because  $\beta_n^2 \leq \delta_n^2$  by Lemma 3.2.2. In general the constraint is not necessary, but it is included to assure the estimator of  $\alpha_n^2$  is nonnegative.  $a_n^2 = d_n^2 - b_n^2$  is a consistent estimator of  $\alpha_n^2$  under quartic mean convergence. This leads to the following key result from Ledoit and Wolf [72].

**Theorem 3.2.1.**

$$S_n^* = \frac{b_n^2}{d_n^2} m_n I_n + \frac{a_n^2}{d_n^2} S_n \quad (3.4)$$

is a consistent estimator of  $\Sigma_n^*$  under quartic mean convergence. As a consequence,  $S_n^*$  has the same asymptotic expected loss (or risk) as  $\Sigma_n^*$ .

### 3.2.1.3 Chen, Wiesel, Hero Approach

Chen, Wiesel and Hero [17] improve on the estimator of Ledoit and Wolf [72] by utilizing the Rao-Blackwell theorem. Under the assumption of normality of the data, they condition on the sufficient statistic  $S$  and find the following estimate for the shrinkage intensity,

$$\hat{\lambda} = \min \left( \frac{(n-2)/n \cdot \text{tr}(S^2) + \text{tr}(S)^2}{(n+2) [\text{tr}(S^2) - \text{tr}(S)^2/p]}, 1 \right). \quad (3.5)$$

By the Rao-Blackwell theorem, see Casella and Berger [14], the shrinkage estimator using this estimator for the optimal intensity should dominate that of Ledoit and Wolf [72]. They use this approach with a motivation for time-series data and their simulations follow with examples showing the estimator's effectiveness. We use a similar approach in that we assume normality of the data and explicitly find the optimal shrinkage intensity in terms of the trace of  $\Sigma$  and  $\Sigma^2$ .

### 3.2.1.4 Schäfer-Strimmer Estimator

Schäfer and Strimmer [85] provide a detailed discussion of Shrinkage estimators of the covariance matrix and finding the optimal intensity. They explore small sample inference. Rather than finding a consistent estimator for  $\mu$  and  $\lambda$ , they estimate  $\lambda$  by finding unbiased sample counterparts. Like that described in Ledoit and Wolf [72] they find an estimator for  $\mu$  by  $v = \text{avg}_i s_{ii}$  where  $s_{ii}$  are the diagonal elements of the sample covariance matrix  $S$ . They suggest the following unbiased

estimator for the shrinkage intensity,

$$\hat{\lambda} = \frac{\sum_{i \neq j} \widehat{\text{Var}}(s_{ij}) + \sum_i \widehat{\text{Var}}(s_{ii})}{\sum_{i \neq j} s_{ij}^2 + \sum_i (s_{ii} - v)^2}. \quad (3.6)$$

It is noted that in order to compute the optimal shrinkage intensity it is necessary to estimate the variances of the individual entries of  $S$ . They provide technical details of this computation in the Appendix of their work, see [85].

### 3.2.1.5 A New Estimator under optimal conditions

Similar to Ledoit and Wolf [72], we utilize a general asymptotic estimator but with a different approach for finite application. In Theorem 3.2.2 we determine the optimal value for  $v$  and  $\lambda$  are  $\mu$  and  $\frac{\beta^2}{\delta^2}$ , respectively. Rather than find consistent estimators for  $\delta^2$ ,  $\beta^2$ ,  $\alpha^2$  and  $\mu$  in a purely asymptotic setting, we explicitly calculate these scalars in terms of the trace of  $\Sigma$  and  $\Sigma^2$  and suggest unbiased estimators for those terms under the following assumptions. We note this approach is similar to that utilized in Chen, Wiesel, and Hero [17].

We assume the data to come from a multivariate normal distribution with some mean vector and covariance matrix  $\Sigma > 0$ . We define  $a_i = \text{tr}\Sigma^i/p$  to be the  $i^{\text{th}}$  arithmetic mean of the eigenvalues of  $\Sigma$ .

We will utilize the following well known result with proof in Appendix B.

**Lemma 3.2.1.**

$$E[\text{tr}(S^2)] = \frac{n+1}{n} \text{tr}\Sigma^2 + \frac{1}{n} (\text{tr}\Sigma)^2 \quad (3.7)$$

We note,  $\mu = \langle \Sigma, I \rangle = \text{tr}(\Sigma)/p = a_1$ . Considering (3.7) and the decomposition of  $\delta^2$  as follows when the data is Normal:

$$\begin{aligned} \delta^2 = E[\|S - \mu I\|^2] &= E[\|S\|^2] - 2\mu E[\langle S, I \rangle] + \mu^2 \|I\|^2 \\ &= E[\text{tr}(S^2)/p] - 2\mu E[\text{tr}(S)/p] + \mu^2 \\ &= \frac{n+1}{n} \frac{1}{p} \text{tr}\Sigma^2 + \frac{1}{n} \frac{1}{p} (\text{tr}\Sigma)^2 - 2\frac{\mu}{p} \text{tr}\Sigma + a_1^2 \\ &= \frac{n+1}{n} a_2 + \frac{p-n}{n} a_1^2. \end{aligned}$$

Likewise, we expand the term  $\alpha^2$  as follows

$$\begin{aligned}\alpha^2 = \|\Sigma - \mu I\|^2 &= \|\Sigma\|^2 - 2\mu\langle\Sigma, I\rangle + \mu^2\|I\|^2 \\ &= \frac{1}{p}\text{tr}\Sigma^2 - 2\mu\frac{1}{p}\text{tr}\Sigma + \mu^2 \\ &= a_2 - a_1^2.\end{aligned}$$

In a finite setting, the problem has been reduced to finding good estimators for  $a_1$  and  $a_2$ .

We utilize the results from Srivastava [99]. He begins with the following assumptions,

$$(A) : \text{As } p \rightarrow \infty, a_i \rightarrow a_i^0, 0 < a_i^0 < \infty, i = 1, \dots, 4,$$

$$(B) : n = O(p^\delta), 0 \leq \delta \leq 1.$$

Srivastava finds unbiased and consistent estimators, with respect to convergence in probability, for  $a_1$  and  $a_2$ . We recall,

$$\hat{a}_1 = \text{tr}S/p$$

and

$$\hat{a}_2 = \frac{n^2}{(n-1)(n+2)} \frac{1}{p} \left[ \text{tr}S^2 - \frac{1}{n}(\text{tr}S)^2 \right]$$

are unbiased and consistent estimators for  $a_1$  and  $a_2$  respectively as  $(n, p) \rightarrow \infty$ . Thus in a finite situation,  $\hat{\mu} = \hat{a}_1$ ,  $\hat{\alpha}^2 = \hat{a}_2 - \hat{a}_1^2$ ,  $\hat{\delta}^2 = \frac{n+1}{n}\hat{a}_2 + \frac{p-n}{n}\hat{a}_1^2$  and  $\hat{\beta}^2 = \hat{\delta}^2 - \hat{\alpha}^2$  will be unbiased estimators for  $\mu$ ,  $\alpha^2$ ,  $\delta^2$  and  $\beta^2$ , respectively. Furthermore, due to the weak assumption (B), both  $\hat{a}_2$  and  $\hat{a}_1$  should be good estimators for large  $p$  situations, i.e. when  $p \gg n$ , consistency will hold. Simulations indicate an improvement with these estimators, over that suggested in the literature, when  $p$  is large and  $n$  is relatively small. In situations where both  $n$  and  $p$  are large (over 25), our newly suggested estimators are comparable to that in the literature.

### 3.2.2 Shrinkage to diagonal, unit variance

We now explore a similar target matrix to that described above. The target matrix is a diagonal with unit variance, or more commonly called the Identity. This target contributed no variance to the shrinkage estimator since the target is a constant matrix. An analogous optimal value of  $\lambda$  can easily be found using the methodology of Ledoit and Wolf [72]. Let  $\delta^2 = \|S - I\|^2$ ,

$\alpha^2 = \|\Sigma - I\|^2$ , and  $\beta^2 = E[\|S - \Sigma\|^2]$ . We utilize the following elementary decomposition of  $\delta^2$ .

**Lemma 3.2.2.**  $\delta^2 = \alpha^2 + \beta^2$

*Proof.* We first recall that  $E[S] = \Sigma$ .

$$\begin{aligned}\delta^2 = E[\|S - I\|^2] &= E[\|S - \Sigma\|^2] + E[\|\Sigma - I\|^2] + 2E[\langle S - \Sigma, \Sigma - I \rangle] \\ &= E[\|S - \Sigma\|^2] + \|\Sigma - I\|^2 + 2\langle E[S - \Sigma], \Sigma - I \rangle \\ &= E[\|S - \Sigma\|^2] + \|\Sigma - I\|^2 = \alpha^2 + \beta^2\end{aligned}$$

□

The decomposition of  $\delta^2$  makes it easy to calculate the optimal value of  $\lambda$  in terms of  $\alpha^2$ ,  $\beta^2$  and  $\delta^2$ .

**Theorem 3.2.2.** *Consider the optimization problem*

$$\begin{aligned}\min_{\lambda} E[\|\Sigma^* - \Sigma\|^2] \\ \text{s.t. } \Sigma^* = \lambda I + (1 - \lambda)S,\end{aligned}$$

where the coefficient  $\lambda$  is nonrandom. The solution verifies

$$\Sigma^* = \frac{\beta^2}{\delta^2} I + \frac{\alpha^2}{\delta^2} S, \quad (3.8)$$

with

$$E[\|\Sigma^* - \Sigma\|^2] = \frac{\alpha^2 \beta^2}{\delta^2}. \quad (3.9)$$

*Proof.* This proof follows that of Theorem 2.1 in Ledoit and Wolf [72]. We use the fact that  $E[S] = \Sigma$  as in Lemma 3.2.2, we can rewrite the objective function as

$$E[\|\Sigma^* - \Sigma\|^2] = \lambda^2 \|\Sigma - I\|^2 + (1 - \lambda)^2 E[\|S - \Sigma\|^2]. \quad (3.10)$$

Differentiating with respect to  $\lambda$  and the first-order condition is:  $2\lambda\alpha^2 - 2(1 - \lambda)\beta^2 = 0$ . The solution is found to be:  $\lambda = \beta^2/(\alpha^2 + \beta^2) = \beta^2/\delta^2$ . We note that  $1 - \lambda = \alpha^2/\delta^2$  and at the optimum, the

objective takes on the value:  $(\beta^2/\delta^2)\alpha^2 + (\alpha^2/\delta^2)\beta^2 = \alpha^2\beta^2/\delta^2$ . A second derivative will show a minimization.  $\square$

### 3.2.2.1 Ledoit-Wolf Type Estimator

Like in Ledoit and Wolf [72], unfortunately  $\lambda = \beta^2/\delta^2$  is not a true estimator since it depends on  $\Sigma$ . Consistent estimators, analogous to that described in Ledoit and Wolf can easily be found,  $d_n^2 = \|S_n - I_n\|^2 \xrightarrow{q.m.} \delta_n^2$  as  $n \rightarrow \infty$ . As before, if we let  $x_{.k}^n$  denote the  $p_n \times 1$   $k^{\text{th}}$  column of the observation matrix  $X_n$ , for  $k = 1, \dots, n$ . We define  $\tilde{b}_n^2 = \frac{1}{n^2} \sum_{k=1}^n \|x_{.k}^n (x_{.k}^n)'\|_F^2$  and  $b_n^2 = \min(\tilde{b}_n^2, d_n^2)$ , then  $\tilde{b}_n^2 \xrightarrow{q.m.} \beta_n^2$  and  $b_n^2 \xrightarrow{q.m.} \beta_n^2$ . The constrained estimator  $b_n^2$  is included because  $\beta_n^2 \leq \delta_n^2$  by Lemma 3.2.2.  $a_n^2 = d_n^2 - b_n^2$  will be a consistent, with respect to quadratic mean, estimator of  $\alpha^2$ . We note that Ledoit and Wolf [72] do not provide these estimators, however they are analogous to those provided for the case of the target matrix  $T = \mu I$ . We will refer to these estimators as Ledoit-Wolf type estimators in our simulations below. We also note that a Rao-Blackwell type approach utilized by Chen, Wiesel and Hero [17] can be applied here as well. Due to the similarity to our recommended estimator below, and the simulation results (see section 3.2.4) indicating an improved performance over that in the literature, we exclude a derivation and discussion on this type of estimator for the optimal shrinkage intensity.

### 3.2.2.2 Schäfer-Strimmer Estimator

Schäfer and Strimmer [85] estimate  $\lambda$  by finding unbiased sample counterparts. They suggest the following unbiased estimator for the shrinkage intensity,

$$\hat{\lambda} = \frac{\sum_{i \neq j} \widehat{\text{Var}}(s_{ij}) + \sum_i \widehat{\text{Var}}(s_{ii})}{\sum_{i \neq j} s_{ij}^2 + \sum_i (s_{ii} - 1)^2}. \quad (3.11)$$

We note this is very similar to (3.11) with the exception of a 1 in replace of the sample  $v$  in the denominator. They provide technical details of this computation in the Appendix of Schäfer and Strimmer [85].



### 3.2.2.3 A New Estimator

Like that in the previous section, under the assumption of normality of our data, we can explicitly calculate  $\alpha^2$ ,  $\beta^2$  and  $\delta^2$  in terms of the trace of  $\Sigma$  and  $\Sigma^2$ ,

$$\begin{aligned}\alpha^2 = \|\Sigma - I\|^2 &= \|\Sigma\|^2 - 2\langle \Sigma, I \rangle + \|I\|^2 \\ &= \frac{1}{p}\text{tr}\Sigma^2 - 2\frac{1}{p}\text{tr}\Sigma + 1 \\ &= a_2 - 2a_1 + 1\end{aligned}$$

and

$$\begin{aligned}\delta^2 = E[\|S - I\|^2] &= E[\|S\|^2] - 2E[\langle S, I \rangle] + \|I\|^2 \\ &= E[\text{tr}(S^2)/p] - 2E[\text{tr}(S)/p] + 1 \\ &= \frac{n+1}{n}\frac{1}{p}\text{tr}\Sigma^2 + \frac{1}{n}\frac{1}{p}(\text{tr}\Sigma)^2 - 2\frac{1}{p}\text{tr}\Sigma + 1 \\ &= \frac{n+1}{n}a_2 + \frac{p}{n}a_1^2 - 2a_1 + 1.\end{aligned}$$

We can replace  $a_1$  and  $a_2$  by the unbiased and consistent estimators  $\hat{a}_1$  and  $\hat{a}_2$ . An estimator for  $\lambda$  then can be found easily by  $\hat{\lambda} = \hat{\alpha}^2/\hat{\delta}^2$  where  $\hat{\alpha}^2 = \hat{a}_2 - 2\hat{a}_1 + 1$  and  $\hat{\delta}^2 = \frac{n+1}{n}\hat{a}_2 + \frac{p}{n}\hat{a}_1^2 - 2\hat{a}_1 + 1$ . As before, due to Assumption (B), we can expect our estimator to perform quite well in high-dimensional, low sample size situations.

### 3.2.3 Shrinkage to diagonal, unequal variance

We now explore the optimal shrinkage intensity for the shrinkage estimator with diagonal, unequal variance target. This model represents a compromise between the low-dimensional targets described above, and the correlation models described in Ledoit and Wolf [70], [71]. This target will meet the criteria stated above, structured, always positive definite and well-conditioned in practice. The target contributes more variance to the estimator since it has  $p$  free parameters, but less bias since it is composed of  $p$  unbiased components. Explicitly written as,  $S^* = \lambda D + (1 - \lambda)S$  where  $D$  is a diagonal matrix consisting of the diagonal entries of  $S$ . Using the theorem of optimization from Ledoit and Wolf [70], Schäfer and Strimmer [85] propose an unbiased estimator for the shrinkage intensity. Using methodology similar to that of Ledoit and Wolf [72] and that in the previous

sections, we explore the optimal in terms of the quadratic loss with respect to the Frobenius norm.

### 3.2.3.1 Schäfer and Strimmer Estimator

Schäfer and Strimmer [85] study this shrinkage estimator type in detail. They suggest a simple unbiased estimator for the shrinkage intensity,

$$\hat{\lambda} = \frac{\sum_{i \neq j} \widehat{\text{Var}}(s_{ij})}{\sum_{i \neq j} s_{ij}^2}. \quad (3.12)$$

The computational details of calculating  $\widehat{\text{Var}}(s_{ij})$  are provided in the appendix in Schäfer and Strimmer [85]. They developed an algorithm that allows quick computation of their shrinkage estimator which is provided in the `corpcor` package in the GNU R-Project. Their algorithm is as efficient as that of computing the sample covariance matrix  $S$ .

### 3.2.3.2 A New Estimator

We approach the problem similar to how Ledoit and Wolf [72] did with the target,  $T = \mu I$ . We will calculate the optimal shrinkage intensity with respect the following scalars.  $\delta^2 = E[\|S - D\|^2]$ ,  $\alpha^2 = E[\|S - \Sigma\|^2]$ ,  $\beta^2 = E[\|\Sigma - D\|^2]$ , and  $\gamma^2 = E[\langle S - \Sigma, \Sigma - D \rangle]$ . We begin by considering the following simple lemma that decomposes  $\delta^2$ .

**Lemma 3.2.3.**  $\delta^2 = \alpha^2 + \beta^2 + 2\gamma^2$

*Proof.*

$$\begin{aligned} \delta^2 = E[\|S - D\|^2] &= E[\|S - \Sigma\|^2] + E[\|\Sigma - D\|^2] + 2E[\langle S - \Sigma, \Sigma - D \rangle] \\ &= \alpha^2 + \beta^2 + 2\gamma^2 \end{aligned}$$

□

This decomposition will allow us to find the optimal shrinkage intensity. Consider the following theorem.

**Theorem 3.2.3.** *The solution to the minimization*

$$\begin{aligned} \min_{\lambda} E[\|\Sigma^* - \Sigma\|^2] \\ \text{s.t. } \Sigma^* = \lambda D + (1 - \lambda)S, \end{aligned}$$

where the coefficient  $\lambda$  is nonrandom is

$$\Sigma^* = \frac{\alpha^2 + \gamma^2}{\delta^2} D + \frac{\beta^2 + \gamma^2}{\delta^2} S. \quad (3.13)$$

*Proof.* We use Lemma 3.2.3 and can rewrite the objective function as

$$\begin{aligned} E[\|\Sigma^* - \Sigma\|^2] &= E[\|\lambda D - \lambda \Sigma + (1 - \lambda)S - (1 - \lambda)\Sigma\|^2] \\ &= \lambda^2 E[\|D - \Sigma\|^2] + (1 - \lambda)^2 E[\|S - \Sigma\|^2] + 2\lambda(1 - \lambda)E[\langle S - \Sigma, D - \Sigma \rangle] \\ &= \lambda^2 \beta^2 + (1 - \lambda)^2 \alpha^2 - 2\lambda(1 - \lambda)\gamma^2 \end{aligned}$$

Differentiate with respect to  $\lambda$  and find the first-order condition.

$$\begin{aligned} \frac{\partial E[\|\Sigma^* - \Sigma\|^2]}{\partial \lambda} &= 2\lambda\beta^2 - 2(1 - \lambda)\alpha^2 - 2\gamma^2 + 4\lambda\gamma^2 = 0 \\ &\quad -2\alpha^2 - 2\gamma^2 + 2\lambda\alpha^2 + 2\lambda^2\beta^2 + 4\lambda\gamma^2 = 0 \\ &\quad \lambda(\alpha^2 + \beta^2 + 2\gamma^2) = \alpha^2 + \gamma^2 \\ \Rightarrow \lambda &= \frac{\alpha^2 + \gamma^2}{\delta^2} \end{aligned}$$

Note that  $(1 - \lambda) = \frac{\beta^2 + \gamma^2}{\delta^2}$ . A second derivative shows the shrinkage intensity that minimizes the objective function.  $\square$

Let  $\Psi = \text{diag}(\Sigma)$ , that is,

$$\Psi = \begin{pmatrix} \sigma_{11} & 0 & \dots & 0 \\ 0 & \sigma_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{pp} \end{pmatrix} \quad (3.14)$$

and we define  $a_i^* = \text{tr}\Psi^i/p$ .

As before, under the assumption that our data is normally distributed, we can explicitly find the scalar terms  $\delta^2$ ,  $\alpha^2$ ,  $\beta^2$  and  $\gamma^2$  in terms of the trace of  $\Sigma$  and now  $\Psi$ . We utilize the Frobenius norm with the  $1/p$  term included. Unlike that in Ledoit and Wolf [72], who include  $1/p$  to normalize the norm of the identity matrix, we include it for theoretical benefits. We begin with additional Lemmas, proofs are provided in Appendix B

**Lemma 3.2.4.**

$$E[\text{tr}(D^2)] = \frac{n+2}{n} \text{tr}\Psi^2 \quad (3.15)$$

and

**Lemma 3.2.5.**

$$E[\text{tr}(SD)] = E[\text{tr}(DS)] = E[\text{tr}(D^2)] \quad (3.16)$$

Utilizing (3.7), (3.15) and (3.16) we get the following

$$\begin{aligned} \delta^2 = E[\|S - D\|^2] &= \frac{1}{p} E[\text{tr}(S^2) - 2\text{tr}(SD) + \text{tr}(D^2)] \\ &= \frac{n+1}{np} \text{tr}\Sigma^2 + \frac{1}{np} (\text{tr}\Sigma)^2 - \frac{n+2}{np} \text{tr}\Psi^2 \\ &= \frac{n+1}{n} a_2 + \frac{p}{n} a_1^2 - \frac{n+2}{n} a_2^*. \end{aligned}$$

Likewise

$$\begin{aligned} \gamma^2 = E[\langle S - \Sigma, \Sigma - D \rangle] &= \frac{1}{p} E[\text{tr}(S\Sigma) - \text{tr}(SD) - \text{tr}\Sigma^2 + \text{tr}(\Sigma D)] \\ &= \frac{1}{p} \left( \text{tr}\Sigma^2 - \frac{n+2}{n} \text{tr}\Psi^2 - \text{tr}\Sigma^2 + \text{tr}\Psi^2 \right) \\ &= -\frac{2}{np} \text{tr}\Psi^2 = -\frac{2}{n} a_2^*, \end{aligned}$$

with the third line justified since  $E[\text{tr}(S\Sigma)] = \text{tr}(E[S]\Sigma) = \text{tr}(\Sigma^2)$  and likewise  $E[\text{tr}(D\Sigma)] = \text{tr}(\Psi^2)$ .

Lastly,

$$\begin{aligned}
\alpha^2 = E[\|S - \Sigma\|^2] &= \frac{1}{p} E[\text{tr}(S^2) - 2\text{tr}(\Sigma S) + \text{tr}\Sigma^2] \\
&= \frac{1}{p} (E[\text{tr}(S^2)] - \text{tr}\Sigma) \\
&= \frac{n+1}{np} \text{tr}\Sigma^2 + \frac{1}{np} (\text{tr}\Sigma)^2 - \frac{1}{p} \text{tr}\Sigma^2 \\
&= \frac{1}{n} (a_2 + pa_1^2).
\end{aligned}$$

We find an estimator for  $\lambda$  in a finite sample using estimators defined for general asymptotics. When considering data from a multivariate normal distribution and recalling  $a_i = \text{tr}\Sigma^i/p$  and  $a_i^* = \text{tr}\Psi^i/p$  where  $\Psi$  is the diagonal elements of  $\Sigma$ . Consider the assumptions,

$$(A) : \text{As } p \rightarrow \infty, a_i \rightarrow a_i^0, 0 < a_i^0 < \infty, i = 1, \dots, 4,$$

$$(A.2) : \text{As } p \rightarrow \infty, a_i^* \rightarrow a_i^{*0}, 0 < a_i^{*0} < \infty, i = 1, \dots, 4,$$

$$(B) : n = O(p^\delta), 0 \leq \delta \leq 1.$$

We note these assumptions are equivalent to that of our assumption in the previous sections and that of Srivastava [99]. Assumption (A) will imply assumption (A.2). Here we write it explicitly to remind the reader. Under our assumptions,  $\hat{a}_2^* = \frac{n}{n+2} \text{tr}(D^2)/p$  will be an unbiased and consistent estimator for  $a_2^*$  under general asymptotics, see Theorem B.1 in Appendix B. To estimate  $\alpha^2$ ,  $\beta^2$ ,  $\gamma^2$ , and  $\delta^2$  we simply replace  $a_1$ ,  $a_2$  and  $a_2^*$  with their respective estimators,  $\hat{a}_1$ ,  $\hat{a}_2$ , and  $\hat{a}_2^*$  respectively. Due to the assumption (B), we expect our estimators to be fairly accurate for large  $p$ , small  $n$  situations. We estimate  $\lambda$  with  $\hat{\lambda} = \frac{\hat{\alpha}^2 + \hat{\gamma}^2}{\hat{\delta}^2}$  where  $\hat{\alpha}^2 = \frac{1}{n} (\hat{a}_2 + p\hat{a}_1^2)$ ,  $\hat{\gamma}^2 = -\frac{2}{n} \hat{a}_2^*$  and  $\hat{\delta}^2 = \frac{n+1}{n} \hat{a}_2 + \frac{p}{n} \hat{a}_1^2 - \frac{n+2}{n} \hat{a}_2^*$ .

### 3.2.4 Simulation Study

We conduct a short simulation study to show the effectiveness of our suggested estimators. We sample  $n+1$  observations from a  $p$ -dimensional multivariate normal distribution with zero mean vector and covariance matrix  $\Sigma$  that is positive definite. The eigenvalues of  $\Sigma$  are drawn from a Uniform distribution over (0.5, 10.5). We then find a positive definite matrix with those eigenvalues utilizing the method provided by Dr. Ravi Varadhan, see the selected source code in the Appendix C. The  $n+1$  samples of  $p$  dimension are then used to estimate the various shrinkage estimators.

	Optimal $\lambda$	$\lambda_{new}$	$\lambda_{LW}$	$\lambda_{RBLW}$	$\lambda_{Schaf}$
Simulated Mean	0.6503192	0.6595865	0.6265542	0.6424440	0.6407515
Standard Error	0	0.0000602	0.0000616	0.0000588	0.0000608

Table 3.1:  $\lambda$  estimation for  $n = 40, p = 20, T = \mu I$

	Optimal $\lambda$	$\lambda_{new}$	$\lambda_{LW}$	$\lambda_{RBLW}$	$\lambda_{Schaf}$
Simulated Mean	0.8635683	0.8685150	0.8121008	0.8392271	0.8388833
Standard Error	0	0.0000626	0.0000630	0.0000604	0.0000624

Table 3.2:  $\lambda$  estimation for  $n = 30, p = 30, T = \mu I$

$m = 1000$  simulations are run for each  $\Sigma$ .

Our first results explore the shrinkage target of the form (3.3) using the estimated values of Ledoit and Wolf [72], the Rao-Blackwell approach by Chen, Wiesel and Hero [17], Schäfer and Strimmer [85] and our suggested estimators using  $\hat{a}_1$  and  $\hat{a}_2$ . We provide two main results. First, we sample from the described multivariate normal distribution and we calculate,  $m = 1000$ , *observed* values of  $\lambda_{new}$ ,  $\lambda_{LW}$ ,  $\lambda_{RBLW}$ , and  $\lambda_{Schaf}$  as the estimated optimal shrinkage intensity for our newly suggested estimator, that of Ledoit and Wolf [72], the Rao-Blackwell approach by Chen, Wiesel and Hero [17] and Schäfer and Strimmer [85], respectively. We then compare these estimates to the *true* optimal value of  $\lambda = \beta^2/\delta^2$  from our analysis under the normality condition. We provide three such cases, Table 3.1 provides results for  $n = 40, p = 20$  for comparison with the results provided in Ledoit and Wolf [72], Table 3.2 for  $n = 30, p = 30$  and Table 3.3 for  $n = 5, p = 100$  as an extreme case of high dimension compared to the number of observations. We note that the *Simulated Mean* value for the *Optimal*  $\lambda$  is constant. We see in Tables 3.1 and 3.2 that our suggested estimate is comparable to that of Ledoit and Wolf [72], Chen, Wiesel and Hero [17] and Schäfer and Strimmer [85] and each does a fairly accurate job, on average, in estimating the optimal shrinkage intensity with comparable standard error. We see in Table 3.3 that our suggested estimator shows an improvement compared to the other estimators, with Chen, Wiesel and Hero [17] and Schäfer and Strimmer's [85] performing better than that of Ledoit and Wolf [72]. We also note, that as expected, the optimal intensity increases as the dimension increases. Our second simulation study looks at how well the shrinkage

	Optimal $\lambda$	$\lambda_{new}$	$\lambda_{LW}$	$\lambda_{RBLW}$	$\lambda_{Schaf}$
Simulated Mean	0.9868715	0.9887804	0.6634387	0.7909521	0.7950775
Standard Error	0	0.0000218	0.0000532	0.0000176	0.0000194

Table 3.3:  $\lambda$  estimation for  $n = 5, p = 100, T = \mu I$

Estimator	$S$	$S_{new}^*$	$S_{LW}^*$	$S_{RBLW}^*$	$S_{Schaf}^*$
Risk	13.57979441	5.60630692	5.62695307	5.60430548	5.60843356
SE on Risk	0.07783102	0.01608281	0.01762772	0.01662065	0.01682610
PRIAL	0.00000000	58.71581891	58.56378310	58.73055725	58.70015857
Cond. Num.	61.74148807	3.09899164	3.37202161	3.23545453	3.25240091

Table 3.4: Shrinkage estimation for  $n = 40, p = 20, T = \mu I$

Estimator	$S$	$S_{new}^*$	$S_{LW}^*$	$S_{RBLW}^*$	$S_{Schaf}^*$
Risk	23.56304	6.58705963	6.64572833	6.58960045	6.59754589
SE on Risk	0.1081352	0.01211236	0.01492722	0.01299565	0.01308563
PRIAL	0	72.04494536	71.79595918	72.03416227	72.00044236
Cond. Num.	3679878	2.63637030	3.06365648	2.85155433	2.85244013

Table 3.5: Shrinkage estimation for  $n = 30, p = 30, T = \mu I$

estimates perform compared to the empirical sample covariance matrix  $S$ . The main comparisons match that of the simulation study in Ledoit and Wolf [72]. We look at the simulated risk, with respect to the true covariance matrix  $\Sigma$ , of  $S$ ,  $S_{new}^*$ ,  $S_{LW}^*$ ,  $S_{RBLW}^*$  and  $S_{Schaf}^*$ , for the sample covariance matrix, the shrinkage estimate using our suggested intensity estimate, that of Ledoit and Wolf [72], Chen, Wiesel and Hero [17] and Schäfer and Strimmer [85], respectively. We also report the percentage relative improvement in average loss (PRIAL) of the various estimated shrinkage estimates. This is defined as:

$$PRIAL(S^*) = \frac{E[\|S - \Sigma\|^2] - E[\|S^* - \Sigma\|^2]}{E[\|S - \Sigma\|^2]} \times 100.$$

If the simulated risk is lower, and the simulated PRIAL is positive, this indicates the shrinkage estimate has performed better, on average, than the sample covariance matrix. We also report the average condition number of each estimate of  $\Sigma$ . This condition number is calculated as the ratio of the maximum and minimum eigenvalues of the estimator as done in Schäfer and Strimmer [85]. Like our previous study, we look at three cases. Table 3.4 for  $n = 40, p = 20$  for comparison with the results provided in Ledoit and Wolf [72], Table 3.5 for  $n = 30, p = 30$  and Table 3.6 for  $n = 5, p = 100$  as an extreme case of high dimension compared to the number of observations. We see from Tables 3.4 and 3.5 that our recommended estimator is comparable to that in the literature, with a slight improvement over that in Ledoit and Wolf [72] in terms of simulated risk, simulated PRIAL and the average condition number. In Table 3.6 we see a substantial improvement over that in the literature. Our estimator using  $\hat{a}_1$  and  $\hat{a}_2$  has lower simulated-risk, with a smaller error, a 98% improvement

Estimator	$S$	$S_{new}^*$	$S_{LW}^*$	$S_{RBLW}^*$	$S_{Schaf}^*$
Risk	529.331706	8.7998516	68.5254660	29.4456277	28.7813665
SE on Risk	2.535637	0.0195423	0.7620129	0.1927102	0.2056495
PRIAL	0	98.3375544	87.0543432	94.4372069	94.5626974
Cond. Num.	$\infty$	1.7015159	15.9457052	8.6021882	8.4549154

Table 3.6: Shrinkage estimation for  $n = 5, p = 100, T = \mu I$

Estimator	$S$	$S_{new}^*$	$S_{LW}^*$	$S_{Schaf}^*$
Risk	13.57979441	6.92282199	7.13536939	7.04202294
SE on Risk	0.07783102	0.03411014	0.03649461	0.03532218
PRIAL	0	49.02115763	47.45598368	48.14337589
Cond. Num.	61.74148807	6.54504929	6.96354951	6.783513571

Table 3.7: Shrinkage estimation for  $n = 40, p = 20, T = I$

over  $S$  and is very well-conditioned. The simulation results suggest our newly suggested estimator, under the assumption of normality, performs better in extreme cases of high dimensions, that is,  $p \gg n$ . We also note, that as expected, the empirical sample covariance matrix performs worse as the dimensionality increases.

We now conduct a short study for the effectiveness of our newly defined estimator for the shrinkage target  $T = I$ . We omit the results for estimating  $\lambda$  as they are analogous to that above. We look at the simulated risk, with respect to the true covariance matrix  $\Sigma$ , of  $S$ ,  $S_{new}^*$ ,  $S_{LW}^*$  and  $S_{Schaf}^*$ , for the sample covariance matrix, the shrinkage estimate using our suggested intensity estimate, that of the Ledoit and Wolf [72] type (i.e. consistent in quadratic mean), and Schäfer and Strimmer [85], respectively. We also report the percentage relative improvement in average loss (PRIAL) of the various estimated shrinkage estimates and the average condition number of each estimate of  $\Sigma$ . Like our previous study, we look at three cases. Table 3.7 for  $n = 40, p = 20$ , Table 3.8 for  $n = 30, p = 30$  and Table 3.9 for  $n = 5, p = 100$  as an extreme case of high dimension compared to the number of observations. We see analogous results to our previous study. Tables 3.7 and 3.8 show that our newly suggested shrinkage estimator is comparable to that in the literature.

Estimator	$S$	$S_{new}^*$	$S_{LW}^*$	$S_{Schaf}^*$
Risk	23.56304	8.02733238	8.53856215	8.28631740
SE on Risk	0.1081352	0.03046508	0.03568145	0.03286028
PRIAL	0	65.93252105	63.76289499	64.83340537
Cond. Num.	3679878	5.29158187	5.91095204	5.60755595

Table 3.8: Shrinkage estimation for  $n = 30, p = 30, T = I$



Estimator	$S$	$S_{new}^*$	$S_{LW}^*$	$S_{Schaf}^*$
Risk	529.331706	9.23354172	75.7840383	34.1198695
SE on Risk	2.535637	0.03560733	0.8071637	0.2346702
PRIAL	0	98.25562278	85.6830722	93.5541610
Cond. Num.	$\infty$	2.36223698	17.3008312	9.5728825

Table 3.9: Shrinkage estimation for  $n = 5, p = 100, T = I$

	Optimal $\lambda$	$\lambda_{new}$	$\lambda_{Schaf}$
Simulated Mean	0.6780003	0.6865272	0.6867752
Standard Error	0	0.0000715	0.0000688

Table 3.10:  $\lambda$  estimation for  $n = 40, p = 20, T = D$

However, Table 3.9 shows an improvement in risk, error, PRIAL and condition number over that in the literature.

We now conduct a simulation study for the shrinkage target,  $T = D$ , the diagonal elements of  $S$ . Like the previous studies, we look at estimating the shrinkage intensity as well as the overall performance of our newly suggested shrinkage estimator. Table 3.10 shows the result for  $n = 40, p = 20$ , Table 3.11 provides the case for  $n = 30, p = 30$  and Table 3.12 is provided to show an extreme case of high dimensionality compared to the sample size. We only look at our newly suggested estimator for  $\lambda$  and compare it to that in Schäfer and Strimmer [85].

We see similar results to our previous simulations. Tables 3.10 and 3.11 show our recommended estimator for  $\lambda$  is comparable to that of Schäfer and Strimmer [85] and are fairly accurate in estimating the optimal shrinkage intensity. Table 3.12 indicates that our recommended estimator performs better than that provided in the literature. We next conduct a series of simulations and comparing the shrinkage estimators,  $S^*$  to the sample covariance matrix. As before, we look at the simulated risk, PRIAL and average condition number for the case of  $n = 40, p = 20$ , Table 3.13,  $n = 30, p = 30$ , Table 3.11 and  $n = 5, p = 100$ , Table 3.15, as an extreme case of high dimensions compared to sample size. We see analogous results to our previous simulation studies. In Tables 3.13 and 3.14 we see that our recommended shrinkage estimator is comparable to that in Schäfer and Strimmer [85] and both show an improvement over the sample covariance matrix. Table 3.15

	Optimal $\lambda$	$\lambda_{new}$	$\lambda_{Schaf}$
Simulated Mean	0.7040765	0.7117110	0.7038033
Standard Error	0	0.0000552	0.0000528

Table 3.11:  $\lambda$  estimation for  $n = 30, p = 30, T = D$

	Optimal $\lambda$	$\lambda_{new}$	$\lambda_{Schaf}$
Simulated Mean	0.9878116	0.9876334	0.7945817
Standard Error	0	0.0000219	0.0000146

Table 3.12:  $\lambda$  estimation for  $n = 5, p = 100, T = D$

Estimator	$S$	$S_{new}^*$	$S_{Schaf}^*$
Risk	13.57979441	6.33127583	6.30745324
SE on Risk	0.07783102	0.02116759	0.02099693
PRIAL	0	53.37723360	53.55266031
Cond. Num.	61.74148807	4.87310095	4.92759049

Table 3.13: Shrinkage estimation for  $n = 40, p = 20, T = D$

Estimator	$S$	$S_{new}^*$	$S_{Schaf}^*$
Risk	23.56304	7.65010548	7.63223088
SE on Risk	0.1081352	0.01829089	0.01799747
PRIAL	0	67.53344758	67.60930623
Cond. Num.	3679878	5.10986390	5.17330154

Table 3.14: Shrinkage estimation for  $n = 30, p = 30, T = D$

Estimator	$S$	$S_{new}^*$	$S_{Schaf}^*$
Risk	529.331706	19.03195992	38.078744
SE on Risk	2.535637	0.07905792	0.195695
PRIAL	0	96.40453053	92.806261
Cond. Num.	$\infty$	46.13478905	115.857786

Table 3.15: Shrinkage estimation for  $n = 5, p = 100, T = D$

shows an improvement over shrinkage estimator provided by Schäfer and Strimmer [85]. We note that none of the results in this particular case appear to be very well-conditioned. This is due to the target matrix construction. Each  $s_{ii}$  in  $D$  is estimated with only  $n = 5$  degrees of freedom. However, our recommended estimator still shows an improvement over that in the literature or the sample covariance matrix.

### 3.2.5 Timing Study

It is noted in Schäfer and Strimmer [85] that their algorithm for estimating the optimal shrinkage estimator, in the diagonal target matrix case, for the covariance matrix is *efficient*. They claim their algorithm is on the same order as estimating the empirical sample covariance matrix. The estimator  $\hat{a}_2$  requires calculation of the square of the sample covariance matrix. We note that in theory this will require  $O(p^2)$  operations and should be on the same order as calculating  $S$ . We conduct a brief timing study to show that our recommended estimators is of the same order of computational efficiency as that described in Schäfer and Strimmer [85]. We note that the purpose of this timing study is not to perform a comprehensive study on the efficiency of our estimator, or that of Schäfer and Strimmer [85], but rather to compare the two.

The study is structured as such: For a given  $n$  and  $p$ , we find a random covariance matrix,  $\Sigma$ , using Dr. Varadhan’s method in Appendix C. We set the seed for the random number generator (RNG) and generate  $m = 10,000$  samples, of size  $N = n + 1$ , from a multivariate normal distribution with mean zero and covariance  $\Sigma$ . The shrinkage estimator using our suggested estimators for the intensity is computed for each of the  $m = 10,000$  iterations. The run-time, in seconds, is calculated and recorded. The process is then repeated by resetting the seed for the RNG, hence each generated normal sample will be the same, and the shrinkage estimator defined by Schäfer and Strimmer [85] is computed and the run-time taken to perform  $m = 10,000$  iterations is calculated and recorded.

A *driver* and *wrapper* program control the sample size and dimensionality and record the results. Coding samples are provided in the Appendix C. The timing study was run on a Dell Vostro 1510, Intel(R) Core(TM)2 Duo T9300 2.5GHz with 2GB of RAM, running Ubuntu release 9.04 (jaunty jackalope) Linux kernel version 2.6.28-2 on GNU R-Project 2.8.1.

We perform three brief timing studies. In the first, the dimensionality is fixed at  $p = 20$ , and the sample size  $n$  increases. Figure 3.1 provides the results as  $n$  increases with  $p$  fixed. We see that both our nearly defined shrinkage estimator and that of Schäfer and Strimmer [85] appear

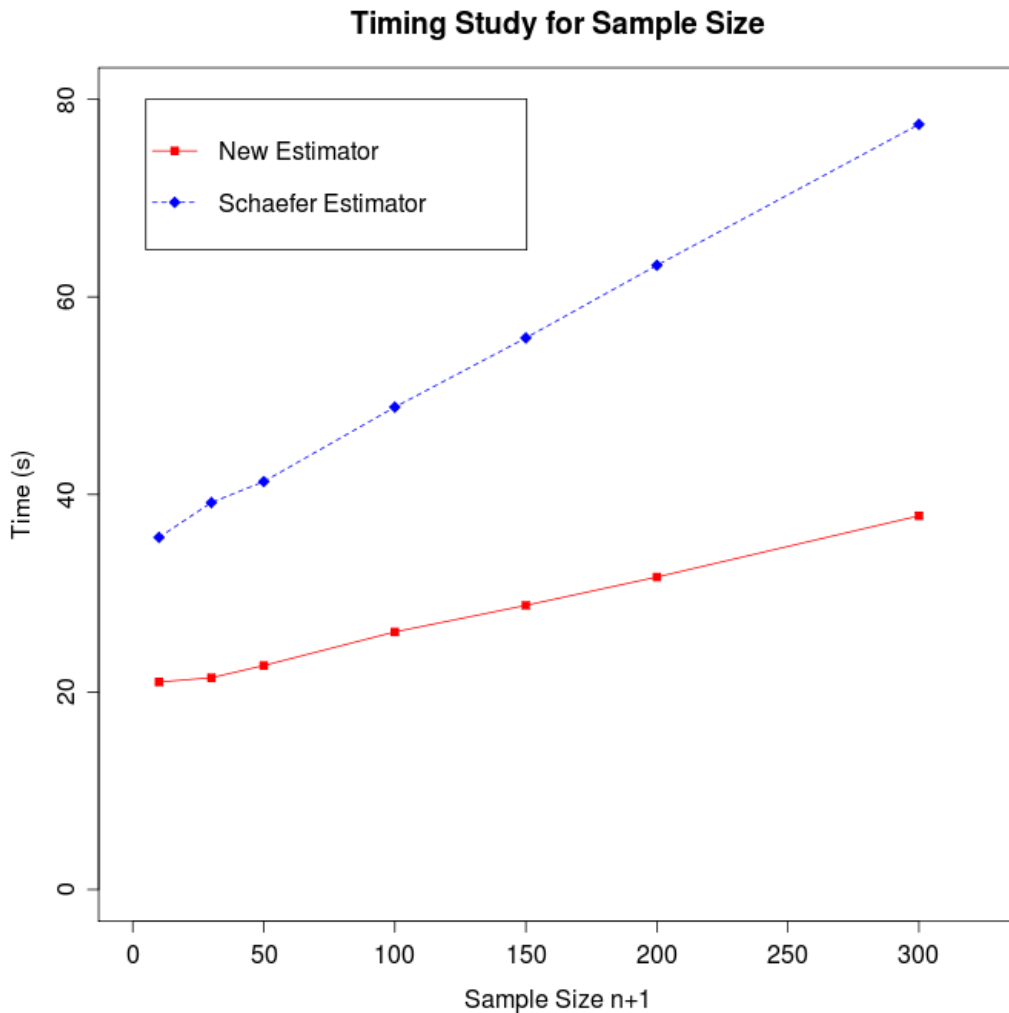


Figure 3.1: Timing Study as  $n$  increases,  $p$  fixed

to have linear growth with respect to  $n$  increasing. Furthermore, it appears that the linear rate is higher in the estimator in Schäfer and Strimmer [85]. Although the units of time are dependent (and hence irrelevant to our study) on the computer performing the study, it does appear that the new shrinkage estimator is approximately twice as fast in this particular case. We next explore the behavior of the estimators as the dimensionality increases with the sample size fixed. We let  $n = 30$  and let  $p$  increase. This situation is an example of extreme high-dimensionality. Figure 3.2 provides the result. We see that both of the shrinkage estimators appear to have a quadratic growth. Not only are they on the same order, they appear to be very similar. Lastly we explore when both the sample size and dimensionality increase. We let  $n = p$  increase together to perform our study. We

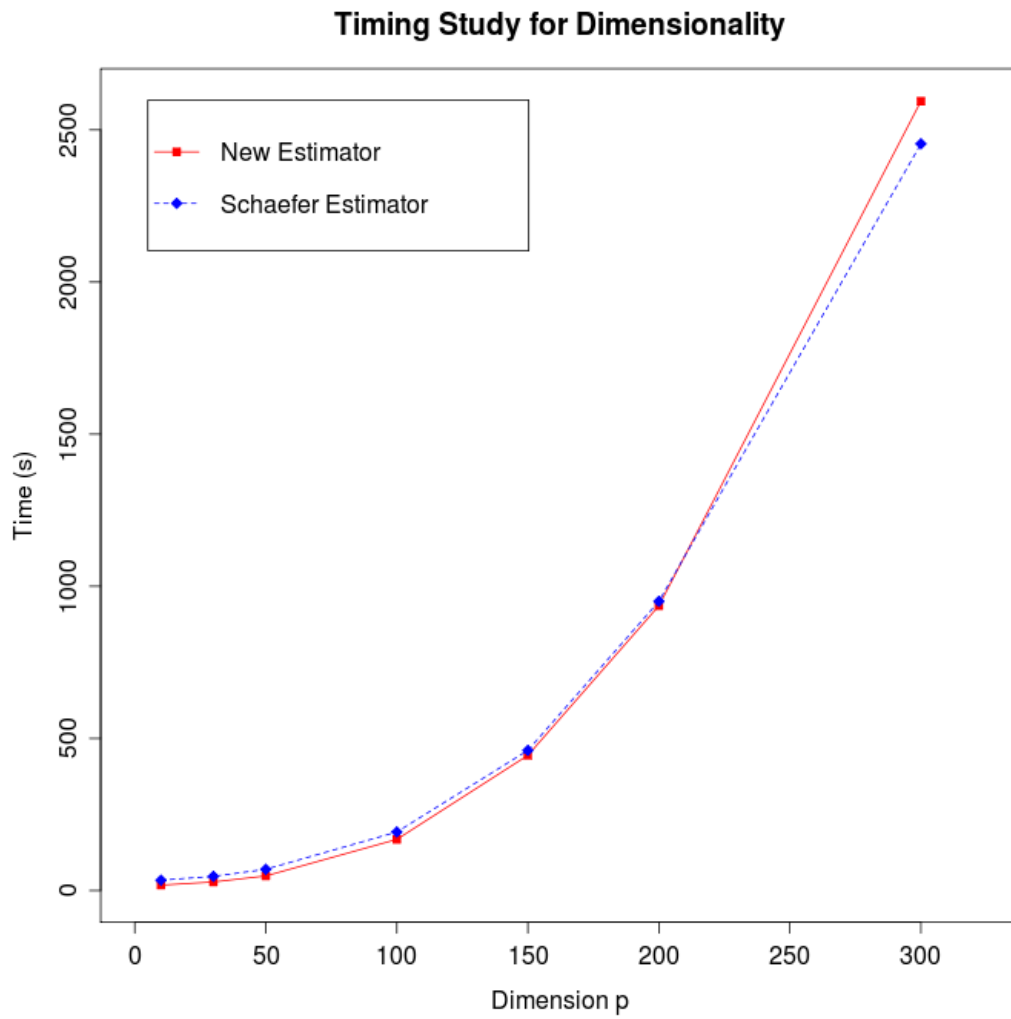


Figure 3.2: Timing Study as  $p$  increases,  $n$  fixed

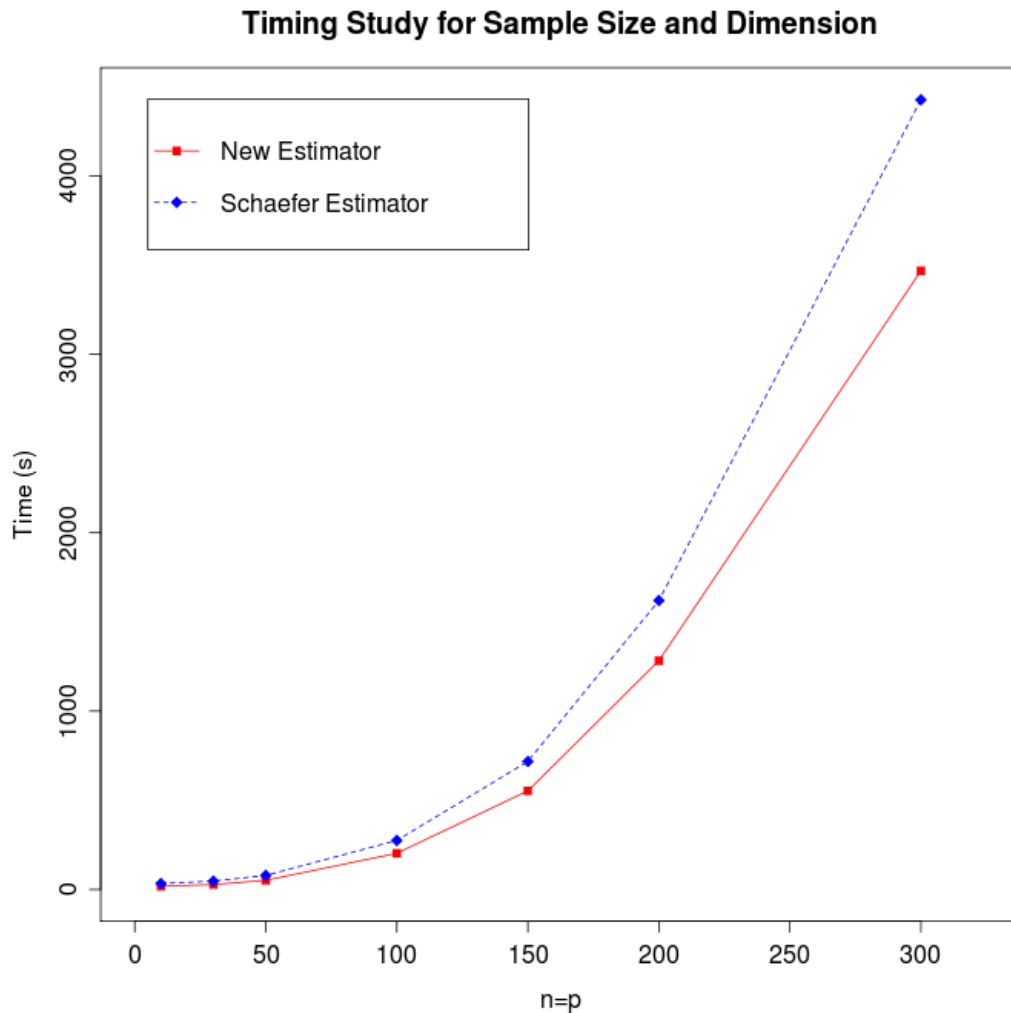


Figure 3.3: Timing Study as  $n = p$  increases

see in Figure 3.3 that as both  $n$  and  $p$  increase together, both estimators appear to have quadratic growth (as expected since  $n$  increasing is linear and  $p$  increasing is quadratic). The simulated timing study seems to indicate that the newly recommended estimator may be slightly more efficient than that in Schäfer and Strimmer [85] but on the same quadratic order. Overall, we can conclude there is no loss in computational efficiency when using our newly suggested shrinkage estimator.

### 3.2.6 Data Example

Here we apply our recommended estimator to a set of real data. The Institute of Applied Microbiology, University of Agricultural Sciences of Vienna, collected microarray data measuring the stress response of the microorganism *Escherichia coli* (more commonly known as *E.coli*) during the expression of a recombinant protein. The data monitors all 4,289 protein coding genes at 8, 15, 22, 45, 68, 90, 150 and 180 minutes after induction of the recombinant protein. In a comparison with pooled samples before induction 102 genes were identified by Schmidt-Heck et al. [86] as differentially expressed in one or more samples after induction. We note we have  $p = 102$  variables with only  $N = n + 1 = 8$  observations. This is a case of extreme high-dimension. Our simulations indicate our newly recommended set of estimators outperforms those in the literature in these cases. We provide shrinkage estimates using the three targets discussed in this chapter,  $\mu I$ ,  $I$ , and  $\text{Diag}(S)$ . For each target we explore all discussed estimates for the shrinkage intensity. For the target  $\mu I$ , this includes the estimator from Ledoit and Wolf [72], the Rao-Blackwell theorem based estimator supplied by Chen, Wiesel and Hero [17], the unbiased estimator provided by Schäfer and Strimmer [85] and our newly introduced estimator. For the target  $I$ , we supply a Ledoit and Wolf [72] type, and the Schäfer and Strimmer [85] suggested estimator, along with our own. Lastly, for the target consisting of the diagonal elements of  $S$ , we look at our newly suggested estimator and compare it to that of Schäfer and Strimmer [85]. Since the true covariance matrix is unavailable, the only way to compare the

Target	$T = \mu I$	$T = I$	$T = \text{diag}(S)$
New Estimators	156.73 (.33)	155.95 (.33)	468.37 (.33)
LW-Type	384.89 (.17)	382.97 (.17)	NA
RBLW-Type	212.23 (.27)	NA	NA
Schäfer-Strimmer	288.79 (.21)	287.35 (.21)	715.25 (.24)

Table 3.16: Condition Numbers (and  $\hat{\lambda}$ ) for estimators and common targets on *E.coli* data

estimates is the condition number of the estimator. We also provide the estimated optimal shrinkage intensity. The sample covariance matrix,  $S$ , has rank 7 and is ill-conditioned. We see in Table 3.16 that for each target matrix, our recommended shrinkage estimator is the best conditioned of those explored. In all cases, the stein-type shrinkages estimators are of full rank (102).

### 3.3 Concluding Remarks

We have explored many of the methods to estimate the covariance matrix in high dimensions. We have introduced three new estimates for optimal Stein-type shrinkage estimators. Simulations indicate our newly suggested estimators for the shrinkage intensity  $\lambda$  are comparable to that in the literature. Simulations also indicate that in the case of very high-dimensions, compared to the sample size, our suggested estimators dominate those in the literature. A brief timing study shows that our estimator is comparable to that in the literature in terms of computational efficiency. Due to the improvement in extreme high-dimensional cases, and the comparable performance in terms of computational efficiency, we recommend our newly suggested estimators based on  $\hat{a}_1$ ,  $\hat{a}_2$  and  $\hat{a}_2^*$  be used for Stein-type shrinkage estimation. A brief data analysis validates the newly suggested estimates as they are the best conditioned.



# Appendices

## Appendix A Proof of $(n, p)$ -Consistent Test for Sphericity

### A.1 Expression of estimator for $a_4$

We obtain expressions for  $\text{tr}S$ ,  $\text{tr}S^2$ ,  $(\text{tr}S)^2$ ,  $\text{tr}S^2(\text{tr}S)^2$ ,  $\text{tr}S^3\text{tr}S$ ,  $\text{tr}S^4$  and  $(\text{tr}S)^4$  in terms of chi-squared random variables. We make use of the following well known theorem from Serdobolskii et al. [91]

**Theorem A.1.** *Consider the sample covariance matrix and recalling  $N = n + 1$ ,*

$$S = \frac{1}{n} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})'.$$

*There exist an orthogonal transformation of vectors*

$$y_k = \sum_{i=1}^N \Omega_{ki} x_i$$

*such that the vectors  $y_N = \sqrt{N}\bar{x}$  and  $y_k \sim N(0, \Sigma)$ ,  $k = 1, \dots, n$ , are independent, and the sample covariance matrix is equal to*

$$S = \frac{1}{n} \sum_{i=1}^n y_i y_i'.$$

Let  $V = nS = YY' \sim W_p(\Sigma, n)$ , where  $Y = (y_1, y_2, \dots, y_n)$  and each  $y_i \sim N_p(0, \Sigma)$  and independent. By orthogonal decomposition,  $\Sigma = \Gamma' \Lambda \Gamma$ , where  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$  with  $\lambda_i$  being the  $i^{\text{th}}$  eigenvalue of  $\Sigma$  and  $\Gamma$  is an orthogonal matrix. Let  $U = (u_1, u_2, \dots, u_n)$ , where  $u_i$  are iid  $N_p(0, I)$  and we can write  $Y = \Sigma^{1/2}U$  where  $\Sigma^{1/2}\Sigma^{1/2} = \Sigma$ . Define  $W' = (w_1, w_2, \dots, w_p) = U'\Gamma'$  and each  $w_i$  are iid.  $N_n(0, I)$ . Thus, define  $v_{ii} = w_i' w_i$  are iid chi-squared random variables with  $n$  degree of freedom.

From Srivastava [99],

$$\begin{aligned} \text{tr}S &= \frac{1}{n} \text{tr}U'\Sigma U \\ &= \frac{1}{n} \text{tr}U'\Gamma'\Lambda\Gamma U \\ &= \frac{1}{n} \text{tr}W'\Lambda W \\ &= \frac{1}{n} \sum_{i=1}^p \lambda_i w_i' w_i. \end{aligned} \tag{A.1}$$

Likewise

$$n^2(\text{tr}S)^2 = \sum_{i=1}^p \lambda_i^2 v_{ii}^2 + 2 \sum_{i<j}^p \lambda_i \lambda_j v_{ii} v_{jj} \quad (\text{A.2})$$

and

$$n^2 \text{tr}S^2 = \sum_{i=1}^p \lambda_i^2 v_{ii}^2 + 2 \sum_{i<j}^p \lambda_i \lambda_j v_{ij}^2. \quad (\text{A.3})$$

Using the same approach and the commutative property of the trace operation (i.e.  $\text{tr}(ABC)$ )

= tr(CAB)), we derive,

$$\begin{aligned}
n^4 \text{tr} S^4 &= \text{tr}(W' \Lambda W)(W' \Lambda W)(W' \Lambda W)(W' \Lambda W) \\
&= \text{tr} \left[ \left( \sum_{i=1}^p \lambda_i w_i w_i' \right)^4 \right] \\
&= \text{tr} \left[ \sum_{i=1}^p \lambda_i^4 w_i w_i' w_i w_i' w_i w_i' w_i w_i' + 4 \sum_{i \neq j}^p \lambda_i^3 \lambda_j w_i w_i' w_i w_i' w_i w_i' w_j w_j' \right. \\
&\quad + \sum_{i < j}^p \lambda_i^2 \lambda_j^2 (4 w_i w_i' w_i w_i' w_j w_j' w_j w_j' + 2 w_i w_i' w_j w_j' w_i w_i' w_j w_j') \\
&\quad + \sum_{i \neq j < k}^p \lambda_i^2 \lambda_j \lambda_k (8 w_i w_i' w_i w_i' w_j w_j' w_k w_k' + 4 w_i w_i' w_j w_j' w_i w_i' w_k w_k') \\
&\quad + \left. \sum_{i < j < k < l}^p \lambda_i \lambda_j \lambda_k \lambda_l (8 w_i w_i' w_j w_j' w_k w_k' w_l w_l' + 8 w_i w_i' w_j w_j' w_l w_l' w_k w_k' \right. \\
&\quad \quad \quad \left. + 8 w_i w_i' w_k w_k' w_j w_j' w_l w_l') \right] \\
&= \sum_{i=1}^p \lambda_i^4 (w_i' w_i)^4 + 4 \sum_{i \neq j}^p \lambda_i^3 \lambda_j (w_i' w_i)^2 (w_i' w_j)^2 \\
&\quad + \sum_{i < j}^p \lambda_i^2 \lambda_j^2 (4 (w_i' w_i) (w_j' w_j) (w_i' w_j)^2 + 2 w (w_i' w_j)^4) \\
&\quad + \sum_{i \neq j < k}^p \lambda_i^2 \lambda_j \lambda_k (8 (w_i' w_i) (w_j' w_j) (w_j' w_k) (w_i' w_k) + 4 (w_i' w_j)^2 (w_i' w_k)^2) \\
&\quad + \sum_{i < j < k < l}^p \lambda_i \lambda_j \lambda_k \lambda_l (8 (w_i' w_j) (w_j' w_k) (w_k' w_l) (w_i' w_l) + 8 (w_i' w_j) (w_j' w_l) (w_l' w_k) (w_i' w_k) \\
&\quad \quad \quad + 8 (w_i' w_k) (w_k' w_j) (w_j' w_l) (w_i' w_l)) \\
&= \sum_{i=1}^p \lambda_i^4 v_{ii}^4 + 4 \sum_{i \neq j}^p \lambda_i^3 \lambda_j v_{ii}^2 v_{ij}^2 + \sum_{i < j}^p \lambda_i^2 \lambda_j^2 (4 v_{ii} v_{jj} v_{ij}^2 + 2 v_{ij}^4) \\
&\quad + \sum_{i \neq j < k}^p \lambda_i^2 \lambda_j \lambda_k (8 v_{ii} v_{ij} v_{jk} v_{ik} + 4 v_{ij}^2 v_{ik}^2) \tag{A.4} \\
&\quad + \sum_{i < j < k < l}^p \lambda_i \lambda_j \lambda_k \lambda_l (8 v_{ij} v_{jk} v_{kl} v_{il} + 8 v_{ij} v_{jl} v_{kl} v_{ik} + 8 v_{ik} v_{jk} v_{jl} v_{il}).
\end{aligned}$$

Likewise we find,

$$\begin{aligned}
n^4 \text{tr} S^3 \text{tr} S &= \sum_{i=1}^p \lambda_i^4 v_{ii}^4 + \sum_{i \neq j}^p \lambda_i^3 \lambda_j (3v_{ii}^2 v_{ij}^2 + v_{ii}^3 v_{jj}) + 6 \sum_{i < j}^p \lambda_i^2 \lambda_j^2 v_{ii} v_{jj} v_{ij}^2 \\
&+ \sum_{i \neq j < k}^p \lambda_i^2 \lambda_j \lambda_k (3v_{ii} v_{ij}^2 v_{kk} + 3v_{ii} v_{ik}^2 v_{jj} + 6v_{ii} v_{ij} v_{ik} v_{jk}) \\
&+ 6 \sum_{i < j < k < l}^p \lambda_i \lambda_j \lambda_k \lambda_l (v_{ii} v_{jk} v_{jl} v_{kl} + v_{jj} v_{ik} v_{il} v_{kl} + v_{kk} v_{ij} v_{il} v_{jl} + v_{ll} v_{ij} v_{ik} v_{jk}),
\end{aligned} \tag{A.5}$$

$$\begin{aligned}
n^4 (\text{tr} S^2)^2 &= \sum_{i=1}^p \lambda_i^4 v_{ii}^4 + 4 \sum_{i \neq j}^p \lambda_i^3 \lambda_j v_{ii}^2 v_{ij}^2 + \sum_{i < j}^p \lambda_i^2 \lambda_j^2 (4v_{ij}^4 + 2v_{ii}^2 v_{jj}^2) \\
&+ \sum_{i \neq j < k}^p \lambda_i^2 \lambda_j \lambda_k (4v_{ii}^2 v_{jk}^2 + 8v_{ij}^2 v_{ik}^2) + 8 \sum_{i < j < k < l}^p \lambda_i \lambda_j \lambda_k \lambda_l (v_{ij}^2 v_{kl}^2 + v_{ik}^2 v_{jl}^2 + v_{il}^2 v_{jk}^2),
\end{aligned} \tag{A.6}$$

$$\begin{aligned}
n^4 \text{tr} S^2 (\text{tr} S)^2 &= \sum_{i=1}^p \lambda_i^4 v_{ii}^4 + \sum_{i \neq j}^p \lambda_i^3 \lambda_j (2v_{ii}^2 v_{ij}^2 + 2v_{ii}^3 v_{jj}) + \sum_{i < j}^p \lambda_i^2 \lambda_j^2 (4v_{ii} v_{jj} v_{ij}^2 + 2v_{ii}^2 v_{jj}^2) \\
&+ \sum_{i \neq j < k}^p \lambda_i^2 \lambda_j \lambda_k (4v_{ii} v_{ij}^2 v_{kk} + 4v_{ii} v_{ik}^2 v_{jj} + 2v_{ii}^2 v_{jj} v_{kk} + 2v_{ii}^2 v_{jk}^2) \\
&+ 4 \sum_{i < j < k < l}^p \lambda_i \lambda_j \lambda_k \lambda_l \left( (v_{ij}^2 v_{kk} + v_{ik}^2 v_{jj} + v_{jk}^2 v_{ii}) v_{ll} \right. \\
&\quad \left. + (v_{il}^2 v_{jj} + v_{jl}^2 v_{ii}) v_{kk} + v_{kl}^2 v_{ii} v_{jj} \right),
\end{aligned} \tag{A.7}$$

and

$$\begin{aligned}
n^4 (\text{tr} S)^4 &= \sum_{i=1}^p \lambda_i^4 v_{ii}^4 + 4 \sum_{i \neq j}^p \lambda_i^3 \lambda_j v_{ii}^3 v_{jj} + 6 \sum_{i < j}^p \lambda_i^2 \lambda_j^2 v_{ii}^2 v_{jj}^2 \\
&+ 12 \sum_{i \neq j < k}^p \lambda_i^2 \lambda_j \lambda_k v_{ii}^2 v_{jj} v_{kk} + 24 \sum_{i < j < k < l}^p \lambda_i \lambda_j \lambda_k \lambda_l v_{ii} v_{jj} v_{kk} v_{ll}.
\end{aligned} \tag{A.8}$$

Consider the constants  $b, c^*, d, e$  defined above in (2.27), (2.28), (2.29), (2.30),

$$\frac{\text{tr} S^4 + b \cdot \text{tr} S^3 \text{tr} S + c^* \cdot (\text{tr} S^2)^2 + d \cdot \text{tr} S^2 (\text{tr} S)^2 + e \cdot (\text{tr} S)^4}{p} = \eta_1 + \eta_2 + \eta_3 + \eta_4 + \eta_5, \tag{A.9}$$

where

$$\eta_1 = \frac{n^4 - 5n^3 + 5n^2 + 5n - 6}{n^2(n^2 + n + 2)} \frac{1}{n^4 p} \sum_{i=1}^p \lambda_i^4 v_{ii}^4, \quad (\text{A.10})$$

$$\begin{aligned} \eta_2 &= \frac{1}{n^4 p} \sum_{i \neq j}^p \lambda_i^3 \lambda_j (4v_{ii}^2 v_{ij}^2 + b(3v_{ii}^2 v_{ij}^2 + v_{ii}^3 v_{jj}) + 4cv_{ii}^2 v_{ij}^2 + d(2v_{ii}^2 v_{ij}^2 + 2v_{ii}^3 v_{jj}) + 4ev_{ii}^3 v_{jj}) \\ &= \frac{4}{n^4 p} \sum_{i \neq j}^p \lambda_i^3 \lambda_j \left( \frac{v_{ii}^2 v_{ij}^2 (n^4 - 4n^3 + n^2 + 6n) + v_{ii}^3 v_{jj} (-n^3 + 4n^2 - n - 6)}{n^2(n^2 + n + 2)} \right), \end{aligned} \quad (\text{A.11})$$

$$\begin{aligned} \eta_3 &= \frac{2}{n^4 p} \sum_{i < j}^p \lambda_i^2 \lambda_j^2 \left( (2v_{ii}^2 v_{jj} v_{ij}^2 + v_{ij}^4) + 3bv_{ii} v_{ij}^2 v_{jj} + c(2v_{ij}^4 + v_{ii}^2 v_{jj}^2) \right. \\ &\quad \left. + d(2v_{ii} v_{ij}^2 v_{jj} + v_{ii}^2 v_{jj}^2) + 3ev_{ii}^2 v_{jj}^2 \right) \\ &= \frac{2}{n^4 p} \sum_{i < j}^p \lambda_i^2 \lambda_j^2 \left( \frac{v_{ii} v_{ij}^2 v_{jj} (2n^4 - 10n^3 + 12n^2) + v_{ij}^4 (n^4 - 3n^3 - 4n^2 + 12n)}{n^2(n^2 + n + 2)} \right. \\ &\quad \left. + \frac{v_{ii}^2 v_{jj}^2 (-2n^3 + 7n^2 + 3n - 18)}{n^2(n^2 + n + 2)} \right), \end{aligned} \quad (\text{A.12})$$

$$\begin{aligned} \eta_4 &= \frac{1}{n^4 p} \sum_{i \neq j < k}^p \lambda_i^2 \lambda_j \lambda_k \left( 8v_{ii} v_{ij} v_{jk} v_{ik} + 4v_{ij}^2 v_{ik}^2 + b(3v_{ii} v_{ij}^2 v_{kk} + 3v_{ii} v_{ik}^2 v_{jj} + 6v_{ii} v_{ij} v_{ik} v_{jk}) \right. \\ &\quad \left. + c(4v_{ii}^2 v_{jk}^2 + 8v_{ij}^2 v_{ik}^2) + d(4v_{ii} v_{ij}^2 v_{kk} + 4v_{ii} v_{ik}^2 v_{jj} + 2v_{ii}^2 v_{jj} v_{kk} + 2v_{ii}^2 v_{jk}^2) + e(12v_{ii}^2 v_{jj} v_{kk}) \right) \\ &= \frac{4}{n^4 p} \sum_{i \neq j < k}^p \lambda_i^2 \lambda_j \lambda_k \left( \frac{v_{ii}^2 v_{jk}^2 (-2n^3 + 2n^2 + 12n) + v_{ij}^2 v_{ik}^2 (n^4 - 3n^3 - 4n^2 + 12n)}{n^2(n^2 + n + 2)} \right. \\ &\quad \left. + \frac{(v_{ii} v_{ij}^2 v_{kk} + v_{ii} v_{ik}^2 v_{jj}) (-3n^3 + 7n^2 + 6n) + v_{ii}^2 v_{jj} v_{kk} (5n^2 - 9n - 18)}{n^2(n^2 + n + 2)} \right. \\ &\quad \left. + \frac{v_{ii} v_{ij} v_{ik} v_{jk} (2n^4 - 4n^3 - 2n^2 - 12n)}{n^2(n^2 + n + 2)} \right), \end{aligned} \quad (\text{A.13})$$

where the index is read as  $i \neq j$ ,  $i \neq k$  and  $j < k$ , and

$$\begin{aligned}
\eta_5 &= \frac{1}{n^4 p} \sum_{i < j < k < l}^p \lambda_i \lambda_j \lambda_k \lambda_l \left( 8(v_{ij} v_{jk} v_{kl} v_{il} + v_{ij} v_{jl} v_{kl} v_{ik} + v_{ik} v_{jk} v_{jl} v_{il}) \right. \\
&\quad + 6b(v_{ii} v_{jk} v_{jl} v_{kl} + v_{jj} v_{ik} v_{il} v_{kl} + v_{kk} v_{ij} v_{il} v_{jl} + v_{ll} v_{ij} v_{ik} v_{jk}) + 8c(v_{ij}^2 v_{kl}^2 + v_{ik}^2 v_{jl}^2 + v_{il}^2 v_{jk}^2) \\
&\quad \left. + 4d(v_{ij}^2 v_{kk} v_{ll} + v_{ik}^2 v_{jj} v_{ll} + v_{il}^2 v_{jj} v_{kk} + v_{jk}^2 v_{ii} v_{ll} + v_{jl}^2 v_{ii} v_{kk} + v_{kl}^2 v_{ii} v_{jj}) + 24e v_{ii} v_{jj} v_{kk} v_{ll} \right) \\
&= \frac{8}{n^4 p} \sum_{i < j < k < l}^p \lambda_i \lambda_j \lambda_k \lambda_l \left( \frac{n^2(n^2 + n + 2)(v_{ij} v_{jk} v_{kl} v_{il} + v_{ij} v_{jl} v_{kl} v_{ik} + v_{ik} v_{jk} v_{jl} v_{il})}{n^2(n^2 + n + 2)} \right. \\
&\quad - \frac{3n(n^2 + n + 2)(v_{ii} v_{jk} v_{jl} v_{kl} + v_{jj} v_{ik} v_{il} v_{kl} + v_{kk} v_{ij} v_{il} v_{jl} + v_{ll} v_{ij} v_{ik} v_{jk})}{n^2(n^2 + n + 2)} \\
&\quad - \frac{n(2n^2 + 3n - 6)(v_{ij}^2 v_{kl}^2 + v_{ik}^2 v_{jl}^2 + v_{il}^2 v_{jk}^2)}{n^2(n^2 + n + 2)} \\
&\quad \left. + \frac{n(5n + 6)(v_{ij}^2 v_{kk} v_{ll} + v_{ik}^2 v_{jj} v_{ll} + v_{il}^2 v_{jj} v_{kk} + v_{jk}^2 v_{ii} v_{ll} + v_{jl}^2 v_{ii} v_{kk} + v_{kl}^2 v_{ii} v_{jj})}{n^2(n^2 + n + 2)} \right. \\
&\quad \left. - \frac{3(5n + 6)v_{ii} v_{jj} v_{kk} v_{ll}}{n^2(n^2 + n + 2)} \right). \tag{A.14}
\end{aligned}$$

## A.2 Expected Value of Estimator

### A.2.1 Expected Value Preliminaries

**Lemma A.1.** For  $v_{ii} = (w'_i w_i)$  and  $v_{ij} = (w'_i w_j)$  for any  $i \neq j$ ,

$$\begin{aligned}
E[v_{ii}^4] &= n(n+2)(n+4)(n+6), & E[v_{ii}^3] &= n(n+2)(n+4) \\
E[v_{ii}^2] &= n(n+2), & E[v_{ii}] &= n, \\
E[v_{ii} v_{ij}^2] &= n(n+2), & E[v_{ii}^2 v_{ij}^2] &= n(n+2)(n+4), \\
E[v_{ij}^4] &= 3n(n+2), & E[v_{ij}^2] &= n, \\
E[v_{ii} v_{ij}^2 v_{jj}] &= n(n+2)^2, & E[v_{ij}^2 v_{jk}^2] &= n(n+2), \\
E[v_{ij} v_{ik} v_{jk}] &= n, & E[v_{ij} v_{il} v_{jk} v_{kl}] &= n, \\
E[v_{ii} v_{ij} v_{ik} v_{jk}] &= n(n+2).
\end{aligned}$$

*Proof.* The first 10 results are the moments of a  $\chi^2$  r.v. or can be found in Srivastava [99]. Using

$v_{ij} = (w'_i w_j) = (w'_j w_i)$  and  $E[w_j w'_j] = I$  we obtain the following:

$$\begin{aligned} E[v_{ij} v_{ik} v_{jk}] &= E\left[(w'_i w_j)(w'_j w_k)(w'_k w_i)\right] \\ &= E\left[w'_i (w_j w'_j) (w_k w'_k) w_i\right] \\ &= E[w'_i w_i] = E[v_{ii}] = n \end{aligned}$$

$$\begin{aligned} E[v_{ij} v_{il} v_{jk} v_{kl}] &= E\left[(w'_i w_j)(w'_j w_k)(w'_k w_l)(w'_l w_i)\right] \\ &= E[w'_i w_i] = E[v_{ii}] = n \end{aligned}$$

$$\begin{aligned} E[v_{ii} v_{ij} v_{ik} v_{jk}] &= E\left[(w'_i w_i)(w'_i w_j)(w'_j w_k)(w'_k w_i)\right] \\ &= E\left[(w'_i w_i)(w'_i w_i)\right] = E[v_{ii}^2] = n(n+2) \end{aligned}$$

□

### A.2.2 Expected Value of $\hat{a}_4$

**Lemma A.2.** For  $\eta_2, \eta_3, \eta_4, \eta_5$  above,

$$E[\eta_2] = E[\eta_3] = E[\eta_4] = E[\eta_5] = 0$$

*Proof.* Using the linearity of *expected value* and Lemma A.1 we find,

$$\begin{aligned} E[\eta_2] &= \frac{4}{n^4 p} \sum_{i \neq j}^p \lambda_i^3 \lambda_j \left( \frac{n(n+2)(n+4)(n^4 - 4n^3 + n^2 + 6n)}{n^2(n^2 + n + 2)} \right. \\ &\quad \left. + \frac{n^2(n+2)(n+4)(-n^3 + 4n^2 - n - 6)}{n^2(n^2 + n + 2)} \right) \\ &= \frac{4}{n^4 p} \sum_{i \neq j}^p \lambda_i^3 \lambda_j \frac{n(n+2)(n+4)}{n^2(n^2 + n + 2)} (n^4 - 4n^3 + n^2 + 6n - n^4 + 4n^3 - n^2 - 6n) = 0, \end{aligned}$$



$$\begin{aligned}
E[\eta_3] &= \frac{2}{n^4 p} \sum_{i < j}^p \lambda_i^2 \lambda_j^2 \left( \frac{n(n+2)^2(2n^4 - 10n^3 + 12n^2) + 3n(n+2)(n^4 - 3n^3 - 4n^2 + 12n)}{n^2(n^2 + n + 2)} \right. \\
&\quad \left. + \frac{n^2(n+2)^2(-2n^3 + 7n^2 + 3n - 18)}{n^2(n^2 + n + 2)} \right) \\
&= \frac{2}{n^4 p} \sum_{i < j}^p \lambda_i^2 \lambda_j^2 \frac{n(n+2)}{n^2(n^2 + n + 2)} \left( (n+2)(2n^4 - 10n^3 + 12n^2) \right. \\
&\quad \left. + 3(n^4 - 3n^3 - 4n^2 + 12n) + n(n+2)(-2n^3 + 7n^2 + 3n - 18) \right) \\
&= \frac{2}{n^4 p} \sum_{i < j}^p \lambda_i^2 \lambda_j^2 \frac{n(n+2)}{n^2(n^2 + n + 2)} \left( 2n^5 - 6n^4 - 8n^3 + 24n^2 \right. \\
&\quad \left. + 3n^4 - 9n^3 - 12n^2 + 36n - 2n^5 + 3n^4 + 17n^3 - 12n^2 - 36n \right) = 0,
\end{aligned}$$

$$\begin{aligned}
E[\eta_4] &= \frac{4}{n^4 p} \sum_{i \neq j < k}^p \lambda_i^2 \lambda_j \lambda_k \left( \frac{n^2(n+2)(-2n^3 + 2n^2 + 12n) + n(n+2)(n^4 - 3n^3 - 4n^2 + 12n)}{n^2(n^2 + n + 2)} \right. \\
&\quad \left. + \frac{(2n^2(n+2))(-3n^3 + 7n^2 + 6n) + n^3(n+2)(5n^2 - 9n - 18)}{n^2(n^2 + n + 2)} \right. \\
&\quad \left. + \frac{n(n+2)(2n^4 - 4n^3 - 2n^2 - 12n)}{n^2(n^2 + n + 2)} \right) \\
&= \frac{4}{n^4 p} \sum_{i \neq j < k}^p \lambda_i^2 \lambda_j \lambda_k \frac{n(n+2)}{n^2(n^2 + n + 2)} \left( n(-2n^3 + 2n^2 + 12n) + (n^4 - 3n^3 - 4n^2 + 12n) \right. \\
&\quad \left. + 2n(-3n^3 + 7n^2 + 6n) + n^2(5n^2 - 9n - 18) \right. \\
&\quad \left. + (2n^4 - 4n^3 - 2n^2 - 12n) \right) \\
&= \frac{4}{n^4 p} \sum_{i \neq j < k}^p \lambda_i^2 \lambda_j \lambda_k \frac{n(n+2)}{n^2(n^2 + n + 2)} \left( -2n^4 + 2n^3 + 12n^2 - n^4 + 3n^3 + 4n^2 - 12n \right. \\
&\quad \left. - 6n^4 + 14n^3 + 12n^2 + 5n^4 - 9n^3 - 18n^2 \right. \\
&\quad \left. + 2n^4 - 4n^3 - 2n^2 - 12n \right) = 0,
\end{aligned}$$

$$\begin{aligned}
E[\eta_5] &= \frac{8}{n^4 p} \sum_{i < j < k < l}^p \lambda_i \lambda_j \lambda_k \lambda_l \left( \frac{n^2(n^2 + n + 2)(3n)}{n^2(n^2 + n + 2)} - \frac{3n(n^2 + n + 2)(4n^2)}{n^2(n^2 + n + 2)} \right. \\
&\quad \left. - \frac{n(2n^2 + 3n - 6)(3n^2)}{n^2(n^2 + n + 2)} + \frac{n(5n + 6)(6n^3)}{n^2(n^2 + n + 2)} - \frac{3(5n + 6)n^4}{n^2(n^2 + n + 2)} \right) \\
&= \frac{8}{n^4 p} \sum_{i < j < k < l}^p \lambda_i \lambda_j \lambda_k \lambda_l \frac{n^3}{n^2(n^2 + n + 2)} \left( 3(n^2 + n + 2) - 12(n^2 + n + 2) \right. \\
&\quad \left. - 3(2n^2 + 3n - 6) + 6n(5n + 6) - 3n(5n + 6) \right) \\
&= \frac{8}{n^4 p} \sum_{i < j < k < l}^p \lambda_i \lambda_j \lambda_k \lambda_l \frac{n^2}{n^2(n^2 + n + 2)} \left( 3n^2 + 3n + 6 - 12n^2 - 12n - 24 \right. \\
&\quad \left. - 6n^2 - 9n + 18 + 30n^2 + 36n - 15n^2 - 18n \right) = 0.
\end{aligned}$$

□

**Lemma A.3.** For  $\eta_1$  defined above

$$E[\eta_1] = \frac{n(n+2)(n+4)(n+6)(n-1)(n-2)(n-3)(n+1)}{pn^6(n^2+n+2)} \sum_{i=1}^p \lambda_i^4.$$

**Theorem A.2.** With  $b, c^*, d, e$  defined above in (2.27), (2.28), (2.29), (2.30)

$$\frac{\tau}{p} \left[ \text{tr}S^4 + b \cdot \text{tr}S^3 \text{tr}S + c \cdot (\text{tr}S^2)^2 + d \cdot \text{tr}S^2 (\text{tr}S)^2 + e \cdot (\text{tr}S)^4 \right] \quad (\text{A.15})$$

is an unbiased estimator for  $a_4 = (\text{tr}\Sigma^4/p)$  where  $\tau$  is define above (2.31).

*Proof.* This follows from Lemma A.2 and A.3

$$\begin{aligned}
E &\left[ \frac{\tau}{p} (\text{tr}S^4 + b \text{tr}S^3 \text{tr}S + c (\text{tr}S^2)^2 + d \text{tr}S^2 (\text{tr}S)^2 + e (\text{tr}S)^4) \right] \\
&= \tau \left( \frac{n(n+2)(n+4)(n+6)(n-1)(n-2)(n-3)(n+1)}{pn^6(n^2+n+2)} \sum_{i=1}^p \lambda_i^4 \right) \\
&= \frac{1}{p} \sum_{i=1}^p \lambda_i^4 = a_4.
\end{aligned}$$

□

### A.3 Variance of Estimator

We shall now find the variance of the estimator in (A.15). We first establish several Lemmas that will allow us to derive the variance of the estimator in (A.15) based on the linear combination of  $\chi^2$  random variables.

#### A.3.1 Variance Preliminaries

**Lemma A.4.** *For a random variable from the Standard Normal distribution, i.e.  $Z \sim N(0, 1)$ , all odd central moments are zero. The first six even moments are as such,*

$$\begin{aligned} E[Z^2] &= 1, & E[Z^4] &= 3 \\ E[Z^6] &= 15, & E[Z^8] &= 105 \\ E[Z^{10}] &= 945, & E[Z^{12}] &= 10395 \end{aligned}$$

**Lemma A.5.** *Let  $Q$  be an orthogonal matrix such that  $A_j = Q'DQ$  with  $D = \text{diag}(w'_j w_j, 0, \dots, 0)$  and  $D = QA_j Q'$ . Given  $A_j$ , we can find  $x_i = Qw_i \sim N_n(0, I)$  and it follows that  $x_i$  is independently distributed of  $A_j$ . Now  $x_i = (x_{i1}, \dots, x_{in}) = Qw_i \sim N_n(0, I)$  and thus  $x_{i1}$  is independent of  $w'_j w_j$  and also  $x_{ik}$  for  $k = 2, \dots, n$ , hence*

$$E[v_{ij}^2] = E[w'_i w_j w'_j w_i] = E[w'_i A_j w_i] = E[x_{i1}^2 w'_j w_j]$$

and

$$E[v_{ij}^4] = E[(w'_i w_j w'_j w_i)^2] = E[(w'_i A_j w_i)^2] = E[x_{i1}^4 (w'_j w_j)^2]$$

*Proof.*  $A_j$  is a function of the random variable  $w_j$ ,  $x_i$  is a function of the random variable  $w_i$ .  $w_i$  and  $w_j$  are independent by definition, hence  $x_i$  and  $A_j$  are independent. Furthermore an orthogonal transformation does not alter the distribution of a Normal random variable. Matrix algebra provides the remaining of the derivation.  $\square$

Throughout this chapter we will make extensive use of the alternative form of the variance of a random variable, i.e.

$$\text{Var}[X] = E[X^2] - E[X]^2$$

### A.3.2 Variance of $\eta_1$

We will begin by finding the variance of  $\eta_1$ . First consider the following elementary lemma

**Lemma A.6.** For  $v_{ii} = (w'_i w_i)$

$$\text{Var}[v_{ii}^4] = 32n(n+2)(n+4)(n+6)(n+7)(n^2+14n+60)$$

*Proof.*

$$\begin{aligned} \text{Var}[v_{ii}^4] &= E[v_{ii}^8] - E[v_{ii}^4]^2 \\ &= n(n+2)(n+4)(n+6)(n+8)(n+10)(n+12)(n+14) - n^2(n+2)^2(n+4)^2(n+6)^2 \\ &= 32n(n+2)(n+4)(n+6)(n+7)(n^2+14n+60) \end{aligned}$$

□

Now consider  $\eta_1$  defined in (A.10)

$$\eta_1 = \frac{n^4 - 5n^3 + 5n^2 + 5n - 6}{n^2(n^2 + n + 2)} \frac{1}{n^4 p} \sum_{i=1}^p \lambda_i^4 v_{ii}^4$$

**Lemma A.7.**

$$\text{Var}[\eta_1] = \frac{32(n+2)(n+4)(n+6)(n+7)(n^2+14n+60)(n^4-5n^3+5n^2+5n-6)^2}{n^{11}(n^2+n+2)^2 p} a_8$$

*Proof.*

$$\begin{aligned} \text{Var}[\eta_1] &= \text{Var} \left[ \frac{n^4 - 5n^3 + 5n^2 + 5n - 6}{n^2(n^2 + n + 2)} \frac{1}{n^4 p} \sum_{i=1}^p \lambda_i^4 v_{ii}^4 \right] \\ &= \left( \frac{n^4 - 5n^3 + 5n^2 + 5n - 6}{n^6(n^2 + n + 2)p^2} \right)^2 \sum_{i=1}^p \lambda_i^8 \text{Var}[v_{ii}^4] \\ &= \frac{32(n+2)(n+4)(n+6)(n+7)(n^2+14n+60)(n^4-5n^3+5n^2+5n-6)^2}{n^{11}(n^2+n+2)^2 p^2} \sum_{i=1}^p \lambda_i^8 \\ &= \frac{32(n+2)(n+4)(n+6)(n+7)(n^2+14n+60)(n^4-5n^3+5n^2+5n-6)^2}{n^{11}(n^2+n+2)^2 p} a_8 \end{aligned}$$

□

### A.3.3 Variance of $\eta_2$

Here we provide details on the derivation of the variance of the  $\eta_2$  component of our estimator. Consider  $\eta_2$  defined in (A.11),

$$\eta_2 = \frac{4}{n^4 p} \sum_{i \neq j}^p \lambda_i^3 \lambda_j \left( \frac{v_{ii}^2 v_{ij}^2 (n^4 - 4n^3 + n^2 + 6n) + v_{ii}^3 v_{jj} (-n^3 + 4n^2 - n - 6)}{n^2 (n^2 + n + 2)} \right),$$

From Lemma A.2, we know  $E[\eta_2] = 0$  and hence  $Var[\eta_2] = E[\eta_2^2]$ . We rewrite  $\eta_2$  in the form

$$\eta_2 = \frac{C_{\eta_2}(n)}{p} \sum_{i \neq j}^p \lambda_i^3 \lambda_j V_{i,j}(\eta_2)$$

where

$$C_{\eta_2}(n) = \frac{4}{n^6 (n^2 + n + 2)}$$

and

$$\begin{aligned} V_{i,j}(\eta_2) &= v_{ii}^2 v_{ij}^2 (n^4 - 4n^3 + n^2 + 6n) + v_{ii}^3 v_{jj} (-n^3 + 4n^2 - n - 6) \\ &= v_{ii}^2 v_{ij}^2 n_1 + v_{ii}^3 v_{jj} n_2 \end{aligned} \quad (\text{A.16})$$

where  $n_1 = n^4 - 4n^3 + n^2 + 6n$  and  $n_2 = -n^3 + 4n^2 - n - 6$ . For simplification of notation, we call  $V_{i,j}(\eta_2) = V_{i,j}$ . Then  $\eta_2^2$  can be expressed as

$$\begin{aligned} \frac{p^2}{C_{\eta_2}(n)^2} \eta_2^2 &= \sum_{i \neq j}^p \lambda_i^6 \lambda_j^2 V_{ij}^2 + 2 \sum_{i \neq j < k}^p \lambda_i^6 \lambda_j \lambda_k V_{ij} V_{ik} + 2 \sum_{i < j}^p \lambda_i^4 \lambda_j^4 V_{ij} V_{ji} + 2 \sum_{i \neq j \neq k}^p \lambda_i^4 \lambda_j^3 \lambda_k V_{ik} V_{ji} \\ &\quad + \sum_{i < j \neq k}^p \lambda_i^3 \lambda_j^3 \lambda_k^2 V_{ik} V_{jk} + 2 \sum_{i < j \neq k < l}^p \lambda_i^3 \lambda_j^3 \lambda_k \lambda_l (V_{ik} V_{jl} + V_{il} V_{jk}) \end{aligned}$$

To compute the variance of  $\eta_2$  we will find the expectation of all the terms above. Consider the preliminary lemma

**Lemma A.8.** For  $v_{ii} = (w'_i w_i)$  and  $v_{ij} = (w'_i w_j)$ ,

$$\begin{aligned} E[v_{ii}^4 v_{ij}^4] &= 3n(n+2)(n+4)(n+6)(n+8)(n+10) \\ E[v_{ii}^5 v_{ij}^2 v_{jj}] &= n(n+2)^2(n+4)(n+6)(n+8)(n+10) \\ E[v_{ii}^6 v_{ij}^2] &= n^2(n+2)^2(n+4)(n+6)(n+8)(n+10) \end{aligned}$$

*Proof.* Some applications of Lemma A.5 along with the following derivations provide the result

$$\begin{aligned} E[v_{ii}^4 v_{ij}^4] &= E[(w'_i w_i)^4 (w'_i A_j w_i)^2] \text{ with } A_j \text{ defined in Lemma A.5} \\ &= E \left[ \left( \sum_{k=1}^n x_{ik}^2 \right)^4 \cdot x_{i1}^4 \cdot (w'_j w_j)^2 \right] \\ &= E \left[ \left( \sum_{k=1}^n x_{ik}^2 \right)^4 \cdot x_{i1}^4 \right] E[(w'_j w_j)^2] \end{aligned}$$

Exploring the left side only,

$$\begin{aligned} &E \left[ \left( \sum_{k=1}^n x_{ik}^2 \right)^4 \cdot x_{i1}^4 \right] \\ &= E \left[ x_{i1}^{12} + 4x_{i1}^{10} \left( \sum_{k=2}^n x_{ik}^2 \right) + 6x_{i1}^8 \left( \sum_{k=2}^n x_{ik}^2 \right)^2 + 4x_{i1}^6 \left( \sum_{k=2}^n x_{ik}^2 \right)^3 + x_{i1}^4 \left( \sum_{k=2}^n x_{ik}^2 \right)^4 \right] \\ &= 10395 + \left( 4 \cdot 945 + 6 \cdot 105(n+1) + 4 \cdot 15(n+1)(n+3) + 3(n+1)(n+3)(n+5) \right) (n-1) \\ &= 3(n+4)(n+6)(n+8)(n+10) \end{aligned}$$

resulting in

$$E[v_{ii}^4 v_{ij}^4] = 3n(n+2)(n+4)(n+6)(n+8)(n+10)$$

$$\begin{aligned} E[v_{ii}^5 v_{ij}^2 v_{jj}] &= E[(w'_i w_i)^5 (w'_i A_j w_i) (v'_j w_j)] \\ &= E \left[ \left( \sum_{k=1}^n x_{ik}^2 \right)^5 \cdot x_{i1}^2 \right] E[v_{jj}^2] \end{aligned}$$

Exploring the left side of the equation

$$\begin{aligned}
& E \left[ \left( \sum_{k=1}^n x_{ik}^2 \right)^5 \cdot x_{i1}^2 \right] \\
= & E \left[ x_{i1}^{12} + 5x_{i1}^{10} \left( \sum_{k=2}^n x_{ik}^2 \right) + 10x_{i1}^8 \left( \sum_{k=2}^n x_{ik}^2 \right)^2 + 10x_{i1}^6 \left( \sum_{k=2}^n x_{ik}^2 \right)^3 \right. \\
& \left. + 5x_{i1}^4 \left( \sum_{k=2}^n x_{ik}^2 \right)^4 + x_{i1}^2 \left( \sum_{k=2}^n x_{ik}^2 \right)^5 \right] \\
= & 10395 + 5 \cdot 945(n-1) + 10 \cdot 105(n-1)(n+1) + 10 \cdot 15(n-1)(n+1)(n+3) \\
& + 5 \cdot 3(n-1)(n+1)(n+3)(n+5) + (n-1)(n+1)(n+3)(n+5)(n+7) \\
= & (n+2)(n+4)(n+6)(n+8)(n+10)
\end{aligned}$$

including the  $E[v_{jj}^2] = n(n+2)$  portion and we get

$$E[v_{ii}^5 v_{ij}^2 v_{jj}] = n(n+2)^2(n+4)(n+6)(n+8)(n+10)$$

The result

$$E[v_{ii}^6 v_{jj}^2] = n^2(n+2)^2(n+4)(n+6)(n+8)(n+10)$$

follows from independent  $\chi^2$  random variables □

Thus

$$\begin{aligned}
E[V_{ij}^2] &= E \left[ (v_{ii}^2 v_{ij}^2 (n^4 - 4n^3 + n^2 + 6n) + v_{ii}^3 v_{jj} (-n^3 + 4n^2 - n - 6))^2 \right] \\
&= E \left[ n_1^2 v_{ii}^4 v_{ij}^4 + 2n_1 n_2 v_{ii}^5 v_{ij}^2 v_{jj} + n_2^2 v_{ii}^6 v_{jj}^2 \right] \\
&= n(n+2)(n+4)(n+6)(n+8)(n+10) (3n_1^2 + 2n_1 n_2 (n+2) + n_2^2 n (n+2)) \\
&= 2n^2(n-1)(n-2)^2(n-3)^2(n+1)^2(n+2)(n+4)(n+6)(n+8)(n+10)
\end{aligned}$$

Now consider the additional lemmas

**Lemma A.9.** For  $v_{ii} = (w'_i w_i)$  and  $v_{ij} = (w'_i w_j)$ ,

$$\begin{aligned} E[v_{ii}^4 v_{ij}^2 v_{ik}^2] &= n(n+2)(n+4)(n+6)(n+8)(n+10) \\ E[v_{ii}^5 v_{ij}^2 v_{kk}] &= n^2(n+2)(n+4)(n+6)(n+8)(n+10) \\ E[v_{ii}^6 v_{jj} v_{kk}] &= n^3(n+2)(n+4)(n+6)(n+8)(n+10) \end{aligned}$$

*Proof.* The proof of these three expectation follow from  $E[w_j w'_j] = I$ ,

$$\begin{aligned} E[v_{ii}^4 v_{ij}^2 v_{ik}^2] &= E[(w'_i w_i)^4 (w'_i w_j w'_j w_i) (w'_i w_k w'_k w_i)] \\ &= E[(w'_i w_i)^4 (w'_i w_i) (w'_i w_i)] \\ &= E[(w'_i w_i)^6] \\ &= n(n+2)(n+4)(n+6)(n+8)(n+10) \end{aligned}$$

likewise

$$\begin{aligned} E[v_{ii}^5 v_{ij}^2 v_{kk}] &= E[(w'_i w_i)^5 (w'_i w_j w'_j w_i)] E[v_{kk}] \\ &= E[(w'_i w_i)^5 (w'_i w_i)] E[v_{kk}] \\ &= n^2(n+2)(n+4)(n+6)(n+8)(n+10) \end{aligned}$$

and by independence

$$\begin{aligned} E[v_{ii}^6 v_{jj} v_{kk}] &= E[v_{ii}^6] E[v_{jj}] E[v_{kk}] \\ &= n^3(n+2)(n+4)(n+6)(n+8)(n+10) \end{aligned}$$

□



Thus for  $j \neq k$

$$\begin{aligned}
E[V_{ij}V_{ik}] &= E[(v_{ii}^2v_{ij}^2n_1 + v_{ii}^3v_{jj}n_2)(v_{ii}^2v_{ik}^2n_1 + v_{ii}^3v_{kk}n_2)] \\
&= E[v_{ii}^4v_{ij}^2v_{ik}^2n_1^2 + v_{ii}^5v_{ij}^2v_{kk}n_1n_2 + v_{ii}^5v_{ik}^2v_{jj}n_1n_2 + v_{ii}^6v_{jj}v_{kk}n_2^2] \\
&= n(n+2)(n+4)(n+6)(n+8)(n+10) (n_1^2 + 2n_1n_2n + n_2^2n^2) \\
&= 0
\end{aligned}$$

**Lemma A.10.** For  $v_{ii} = (w'_i w_i)$  and  $v_{ij} = (w'_i w_j)$ ,

$$\begin{aligned}
E[v_{ii}^2v_{ij}^4v_{jj}^2] &= 3n(n+2)(n+4)^2(n+6)^2 \\
E[v_{ii}^3v_{ij}^2v_{jj}^3] &= n(n+2)^2(n+4)^2(n+6)^2 \\
E[v_{ii}^4v_{jj}^4] &= n^2(n+2)^2(n+4)^2(n+6)^2
\end{aligned}$$

*Proof.*

$$\begin{aligned}
E[v_{ii}^2v_{ij}^4v_{jj}^2] &= E[(w'_i w_i)^2 (w'_i w_j w'_j w_i)^2 (w'_j w_j)^2] \\
&= E[(w'_i w_i)^2 (w'_i A_j w_i)^2 (w'_j w_j)^2]
\end{aligned}$$

and using Lemma A.5

$$\begin{aligned}
&= E \left[ \left( \sum_{k=1}^n x_{ik}^2 \right)^2 \cdot x_{i1}^4 \cdot (w'_j w_j)^4 \right] \\
&= E \left[ \left( \sum_{k=1}^n x_{ik}^2 \right)^2 \cdot x_{i1}^4 \right] E [(w'_j w_j)^4] \\
&= E \left[ x_{i1}^8 + 2x_{i1}^6 \left( \sum_{k=2}^n x_{ik}^2 \right) + x_{i1}^4 \left( \sum_{k=1}^n x_{ik}^2 \right)^2 \right] E [(w'_j w_j)^4] \\
&= (105 + 2 \cdot 15(n-1) + 3(n-1)(n+1)) n(n+2)(n+4)(n+6) \\
&= 3n(n+2)(n+4)^2(n+6)^2
\end{aligned}$$

$$E[v_{ii}^3v_{ij}^2v_{jj}^3] = E[(w'_i w_i)^3 (w'_i A_j w_i) (w'_j w_j)^3]$$

and by Lemma A.5

$$\begin{aligned}
&= E \left[ \left( \sum_{k=1}^n x_{ik}^2 \right)^3 \cdot x_{i1}^2 \cdot (w'_j w_j)^4 \right] \\
&= E \left[ \left( \sum_{k=1}^n x_{ik}^2 \right)^3 \cdot x_{i1}^2 \right] E [(w'_j w_j)^4] \\
&= E \left[ x_{i1}^8 + 3x_{i1}^6 \left( \sum_{k=2}^n x_{ik}^2 \right) + 3x_{i1}^4 \left( \sum_{k=1}^n x_{ik}^2 \right)^2 + x_{i1}^2 \left( \sum_{k=2}^n x_{ik}^2 \right)^3 \right] E [(w'_j w_j)^4] \\
&= \left( 105 + 3 \cdot 15(n-1) + 3 \cdot 3(n-1)(n+1) + (n-1)(n+1)(n+3) \right) n(n+2)(n+4)(n+6) \\
&= n(n+2)^2(n+4)^2(n+6)^2
\end{aligned}$$

and  $E[v_{ii}^4 v_{jj}^4] = n^2(n+2)^2(n+4)^2(n+6)^2$  by independent  $\chi^2$  random variables.  $\square$

and thus,

$$\begin{aligned}
E[V_{ij} V_{ji}] &= E[(v_{ii}^2 v_{ij}^2 n_1 + v_{ii}^3 v_{jj} n_2)(v_{jj}^2 v_{ij}^2 n_1 + v_{jj}^3 v_{ii} n_2)] \\
&= E[n_1^2 v_{ii}^2 v_{ij}^4 v_{jj}^2 + 2n_1 n_2 v_{ii}^3 v_{ij}^2 v_{jj}^3 + n_2^2 v_{ii}^4 v_{jj}^4] \\
&= 2n^2(n+1)^2(n+2)(n+4)^2(n+6)^2(n-1)(n-2)^2(n-3)^2
\end{aligned}$$

**Lemma A.11.** For  $v_{ii} = (w'_i w_i)$  and  $v_{ij} = (w'_i w_j)$ ,

$$\begin{aligned}
E[v_{ii}^2 v_{ij}^2 v_{jj}^2 v_{ik}^2] &= n(n+2)^2(n+4)^2(n+6) \\
E[v_{ii}^3 v_{ik}^2 v_{jj}^3] &= n^2(n+2)^2(n+4)^2(n+6) \\
E[v_{ii}^3 v_{ij}^2 v_{jj}^2 v_{kk}] &= n^2(n+2)^2(n+4)^2(n+6) \\
E[v_{ii}^4 v_{ij}^3 v_{kk}] &= n^3(n+2)^2(n+4)^2(n+6)
\end{aligned}$$

*Proof.*

$$E[v_{ii}^2 v_{ij}^2 v_{jj}^2 v_{ik}^2] = E[(w'_i w_i)^2 (w'_i A_j w_i) (w'_j w_j)^2 (w'_i w_k)^2]$$

By Lemma A.5 we can rewrite and isolate the  $w_j$  random variables

$$= E \left[ \left( \sum_{m=1}^n x_{im}^2 \right)^2 \left( \sum_{m=1}^n x_{im} x_{km} \right)^2 \cdot x_{i1}^2 \cdot (w'_j w_j)^3 \right]$$

and we'll look at the left side of the derivation

$$\begin{aligned}
& E \left[ \left( \sum_{m=1}^n x_{im}^2 \right)^2 \left( \sum_{m=1}^n x_{im} x_{km} \right)^2 \cdot x_{i1}^2 \right] \\
= & E \left[ x_{i1}^8 x_{k1}^2 + x_{i1}^7 (L) + 2x_{i1}^6 x_{k1}^2 \left( \sum_{m=2}^n x_{im}^2 \right) + x_{i1}^6 \left( \sum_{m=1}^n x_{im} x_{km} \right)^2 + x_{i1}^5 (L) \right. \\
& + x_{i1}^4 x_{k1}^2 \left( \sum_{m=2}^n x_{im}^2 \right)^2 + 2x_{i1}^4 \left( \sum_{m=2}^n x_{im}^2 \right) \left( \sum_{m=1}^n x_{im} x_{km} \right)^2 \\
& \left. + x_{i1}^3 (L) + x_{i1}^2 \left( \sum_{m=2}^n x_{im}^2 \right)^2 \left( \sum_{m=1}^n x_{im} x_{km} \right)^2 \right]
\end{aligned}$$

noting that  $(L)$  are random variables that do not contribute to the expected value due to the odd-moment of a normal rv and using previous results

$$\begin{aligned}
E \left[ \left( \sum_{m=1}^n x_{im} x_{km} \right)^2 \right] &= E[(w'_i w_k)^2] \\
&= E[(w'_i w_k w'_k w_i)] \\
&= E[w'_i w_i] \\
&= n - 1
\end{aligned}$$

when  $w_i$  is  $N_{n-1}(0, I)$ , hence

$$\begin{aligned}
& E \left[ \left( \sum_{m=1}^n x_{im}^2 \right)^2 \left( \sum_{m=1}^n x_{im} x_{km} \right)^2 \cdot x_{i1}^2 \right] \\
= & 105 + 3 \cdot 15(n-1) + 3(n-1)(n+1) + 2 \cdot 3(n-1)(n+1) + (n-1)(n+1)(n+3) \\
= & (n+2)(n+4)(n+6)
\end{aligned}$$

and

$$E[v_{ii}^2 v_{ij}^2 v_{jj}^2 v_{ik}^2] = n(n+2)^2(n+4)^2(n+6)$$

$$\begin{aligned}
E[v_{ii}^3 v_{ik}^2 v_{jj}^3] &= E[v_{ii}^3 v_{ik}^2] E[v_{jj}^3] \\
&= E[(w'_i w_i)^3 (w'_i w_k w'_k w_i)] E[(w'_j w_j)^3] \\
&= E[(w'_i w_i)^4] E[(w'_j w_j)^3] \\
&= n^2 (n+2)^2 (n+4)^2 (n+6)
\end{aligned}$$

$$E[v_{ii}^3 v_{ij}^2 v_{jj}^2 v_{kk}] = E[(w'_i w_i)^3 (w'_i A_j w_i) (w'_j w_j)^2] E[(w'_k w_k)]$$

and by lemma A.5, looking at the left portion of the expectation

$$E[(w'_i w_i)^3 (w'_i A_j w_i) (w'_j w_j)^2] = E[(w'_i w_i)^3 x_{i1}^2] E[(w'_j w_j)^3]$$

with

$$\begin{aligned}
E[(w'_i w_i)^3 x_{i1}^2] &= E \left[ x_{i1}^8 + 3x_{i1}^6 \left( \sum_{m=1}^n x_{im}^2 \right) + 3x_{i1}^4 \left( \sum_{m=1}^n x_{im}^2 \right)^2 + x_{i1}^2 \left( \sum_{m=1}^n x_{im}^2 \right)^3 \right] \\
&= 105 + 3 \cdot 15(n-1) + 3 \cdot 3(n-1)(n+1) + (n-1)(n+1)(n+3) \\
&= (n+2)(n+4)(n+6)
\end{aligned}$$

and

$$E[v_{ii}^3 v_{ij}^2 v_{jj}^2 v_{kk}] = n^2 (n+2)^2 (n+4)(n+6)$$

and

$$E[v_{ii}^4 v_{jj}^3 v_{kk}] = n^3 (n+2)^2 (n+4)^2 (n+6)$$

follows from independent  $\chi^2$  random variables. □

Thus

$$\begin{aligned}
E[V_{ik} V_{ji}] &= E[n_1^2 v_{ii}^2 v_{ik}^2 v_{ij}^2 v_{jj} + n_1 n_2 v_{ii}^3 v_{ik}^2 v_{jj}^4 + n_1 n_2 v_{ii}^3 v_{ij}^2 v_{ij}^2 v_{kk} + n_2^2 v_{ii}^4 v_{jj}^3 v_{kk}] \\
&= 0
\end{aligned}$$

**Lemma A.12.** For  $v_{ii} = (w'_i w_i)$  and  $v_{ij} = (w'_i w_j)$ ,

$$\begin{aligned} E[v_{ii}^2 v_{ik}^2 v_{jj}^2 v_{jk}^2] &= n(n+2)^3(n+4)^2 \\ E[v_{ii}^2 v_{ik}^2 v_{jj}^3 v_{kk}] &= n^2(n+2)^3(n+4)^2 \\ E[v_{ii}^3 v_{jj}^3 v_{kk}^2] &= n^3(n+2)^3(n+4)^2 \end{aligned}$$

*Proof.*

$$E[v_{ii}^2 v_{ik}^2 v_{jj}^2 v_{jk}^2] = E[(w'_i w_i)^2 (w'_i A_k w_i) (w'_j A_k w_j) (w'_j w_j)^2]$$

and apply Lemma A.5 with respect to  $A_k$  resulting in

$$E[v_{ii}^2 v_{ik}^2 v_{jj}^2 v_{jk}^2] = E[(w'_i w_i)^2 x_{i1}^2] E[(w'_j w_j)^2 x_{j1}^2] E[(w'_k w_k)^2]$$

by independence. We explore the terms of  $w_i$

$$\begin{aligned} E[(w'_i w_i)^2 x_{i1}^2] &= E \left[ \left( \sum_{m=1}^n x_{im}^2 \right)^2 \cdot x_{i1}^2 \right] \\ &= E \left[ x_{i1}^6 + 2x_{i1}^4 \left( \sum_{m=2}^n x_{im}^2 \right) + x_{i1}^2 \left( \sum_{m=2}^n x_{im}^2 \right)^2 \right] \\ &= 15 + 2 \cdot 3(n-1) + 1(n-1)(n+1) \\ &= (n+2)(n+4) \end{aligned}$$

and similar for the  $w_j$  term, hence

$$\begin{aligned} E[v_{ii}^2 v_{ik}^2 v_{jj}^2 v_{jk}^2] &= E[(w'_i w_i)^2 x_{i1}^2] E[(w'_j w_j)^2 x_{j1}^2] E[(w'_k w_k)^2] \\ &= (n+2)(n+4)(n+2)(n+4)n(n+2) \\ &= n(n+2)^3(n+4)^2 \end{aligned}$$

$$E[v_{ii}^2 v_{ik}^2 v_{jj}^3 v_{kk}] = E[(w'_i w_i)^2 (w'_i A_k w_i) (w'_k w_k) (w'_j w_j)^3]$$

and we apply Lemma A.5 and we can rewrite the expectation as

$$E[v_{ii}^2 v_{ik}^2 v_{jj}^3 v_{kk}] = E[(w'_i w_i)^2 x_{i1}^2] E[(w'_k w_k)^2] E[(w'_j w_j)^3]$$

and using the argument above for the  $w_i$  component we get

$$\begin{aligned} E[v_{ii}^2 v_{ik}^2 v_{jj}^3 v_{kk}] &= E[(w'_i w_i)^2 x_{i1}^2] E[(w'_k w_k)^2] E[(w'_j w_j)^3] \\ &= (n+2)(n+4)n(n+2)n(n+2)(n+4) \\ &= n^2(n+2)^3(n+4)^2 \end{aligned}$$

and lastly

$$E[v_{ii}^3 v_{jj}^3 v_{kk}^2] = n^3(n+2)^3(n+4)^2$$

follows from independence. □

$$\begin{aligned} E[V_{ik} V_{jk}] &= EE[n_1^2 v_{ii}^2 v_{ik}^2 v_{jj}^2 v_{jk}^2 + n_1 n_2 v_{ii}^2 v_{ik}^2 v_{jj}^3 v_{kk} + n_1 n_2 v_{ii}^3 v_{jj}^2 v_{jk}^2 v_{kk} + n_2^2 v_{ii}^3 v_{jj}^3 v_{kk}^2] \\ &= n_1^2 n(n+2)^3(n+4)^2 + 2n_1 n_2 n^2(n+2)^3(n+4)^2 + n_2^2 n^3(n+2)^3(n+4)^2 \\ &= 0 \end{aligned}$$

Due to independence, we see the following

$$\begin{aligned} E[V_{ik} V_{jl}] &= E[V_{ik}] E[V_{jl}] \\ &= 0 \end{aligned}$$

and likewise for  $V_{il} V_{jk}$ , which allows us to compute

Thus

$$\begin{aligned} E\left[\frac{p^2}{C_{\eta_2}(n)^2} \eta_2^2\right] &= 2n^2(n-1)(n-2)^2(n-3)^2(n+1)^2(n+2)(n+4)(n+6) \\ &\quad \times \left( (n+8)(n+10) \sum_{i \neq j}^p \lambda_i^6 \lambda_j^2 + 2(n+4)(n+6) \sum_{i < j}^p \lambda_i^4 \lambda_j^4 \right) \end{aligned}$$

with the remaining terms having an expectation of zero. We note

$$\begin{aligned}\sum_{i \neq j}^p \lambda_i^6 \lambda_j^2 &= \left( \sum_{i=1}^p \lambda_i^6 \right) \left( \sum_{j=1}^p \lambda_j^2 \right) - \left( \sum_{i=1}^p \lambda_i^8 \right) \\ &= p^2 a_6 a_2 - p a_8 = p(p a_6 a_2 - a_8)\end{aligned}$$

and

$$\begin{aligned}2 \sum_{i < j}^p \lambda_i^4 \lambda_j^4 &= \left( \sum_{i=1}^p \lambda_i^4 \right) \left( \sum_{j=1}^p \lambda_j^4 \right) - \left( \sum_{i=1}^p \lambda_i^8 \right) \\ &= p^2 a_4^2 - p a_8 = p(p a_4^2 - a_8)\end{aligned}\tag{A.17}$$

and then

$$\begin{aligned}Var(\eta_2) = E[\eta_2^2] &= C_{\eta_2}(n)^2 n^2 (n-1)(n-2)^2 (n-3)^2 (n+1)^2 (n+2)(n+4)(n+6) \\ &\quad \times ((n+8)(n+10)p(p a_6 a_2 - a_8) + (n+4)(n+6)p(p a_4^2 - a_8)) \\ &= \frac{32(n-1)(n-2)^2 (n-3)^2 (n+1)^2 (n+2)(n+4)(n+6)}{n^{10}(n^2+n+2)^2} \\ &\quad \times \left( (n^2+18n+80)a_6 a_2 + (n^2+10n+24)a_4^2 - \frac{2(n^2+14n+52)}{p} a_8 \right)\end{aligned}$$

### A.3.4 Variance of $\eta_3, \eta_4$ and $\eta_5$

Following the same derivation in the calculation for the variance of the  $\eta_2$  terms in section A.3.3 we can find the variance of  $\eta_3, \eta_4$  and  $\eta_5$ . We leave out some of the tedious details of calculating the expectations of the random components.

Consider (A.12) and

$$\begin{aligned}\eta_3 &= \frac{2}{n^4 p} \sum_{i < j}^p \lambda_i^2 \lambda_j^2 \left( \frac{v_{ii} v_{ij}^2 v_{jj} (2n^4 - 10n^3 + 12n^2) + v_{ij}^4 (n^4 - 3n^3 - 4n^2 + 12n)}{n^2 (n^2 + n + 2)} \right. \\ &\quad \left. + \frac{v_{ii}^2 v_{jj}^2 (-2n^3 + 7n^2 + 3n - 18)}{n^2 (n^2 + n + 2)} \right) \\ &= \frac{C_{\eta_3}(n)}{p} \sum_{i < j}^p \lambda_i^2 \lambda_j^2 V_{i,j}(\eta_3)\end{aligned}$$

where

$$C_{\eta_3}(n) = \frac{2}{n^6(n^2 + n + 2)}$$

and  $V_{i,j}(\eta_3)$  represents the random component.

We find

$$E[V_{i,j}(\eta_3)V_{i,k}(\eta_3)] = E[V_{i,j}(\eta_3)V_{j,k}(\eta_3)] = 0$$

and

$$E[V_{i,j}^2] = 8n^2(n-1)(n-2)^2(n-3)^2(n+1)(n+2)(n+4)(n+6)(n^3 + 15n^2 + 69n + 54)$$

Recalling (A.17) and using the property that  $\eta_3$  has mean zero and the fact the cross-products do not contribute to the second moment, we can find the variance by

$$\begin{aligned} \text{Var}[\eta_3] &= \frac{C_{\eta_3}(n)^2}{p^2} 8n^2(n-1)(n-2)^2(n-3)^2(n+1)(n+2)(n+4)(n+6) \\ &\quad \times (n^3 + 15n^2 + 69n + 54) \sum_{i < j}^p \lambda_i^4 \lambda_j^4 \\ &= \frac{16(n-1)(n-2)^2(n-3)^2(n+1)(n+2)(n+4)(n+6)}{n^{10}(n^2 + n + 2)^2} \\ &\quad \times (n^3 + 15n^2 + 69n + 54) \left( a_4^2 - \frac{a_8}{p} \right) \end{aligned}$$

Next we consider the term (A.13)

$$\begin{aligned} \eta_4 &= \frac{4}{n^4 p} \sum_{i \neq j < k}^p \lambda_i^2 \lambda_j \lambda_k \left( \frac{v_{ii}^2 v_{jk}^2 (-2n^3 + 2n^2 + 12n) + v_{ij}^2 v_{jk}^2 (-4n^3 - 6n^2 + 12n)}{n^2(n^2 + n + 2)} \right. \\ &\quad + \frac{(v_{ii} v_{ij}^2 v_{kk} + v_{ii} v_{ik}^2 v_{jj}) (-3n^3 + 7n^2 + 6n) + v_{ii}^2 v_{jj} v_{kk} (5n^2 - 9n - 18)}{n^2(n^2 + n + 2)} \\ &\quad \left. + \frac{v_{ii} v_{ij} v_{ik} v_{jk} (2n^4 - 4n^3 - 2n^2 - 12n) + v_{ij}^2 v_{ik}^2 (n^4 + n^3 + 2n^2)}{n^2(n^2 + n + 2)} \right) \\ &= \frac{C_{\eta_4}(n)}{p} \sum_{i \neq j < k}^p \lambda_i^2 \lambda_j \lambda_k V_{i,j,k}(\eta_4) \end{aligned}$$

where we define

$$C_{\eta_4}(n) = \frac{4}{n^6(n^2 + n + 2)}$$



and  $V_{i,j,k}(\eta_4)$  is the random component. We find

$$E[V_{i,j,k}(\eta_4)^2] = 4n^2(n+2)(n^{10}+12n^9-138n^7-81n^6+3102n^5+200n^4-5316n^3-912n^2-144n+3456)$$

and

$$\begin{aligned} E[V_{i,j,k}(\eta_4)V_{j,i,k}(\eta_4)] &= E[V_{i,j,k}(\eta_4)V_{k,i,j}(\eta_4)] \\ &= 4n^2(n+2)(n+4)\left(n^9+5n^8-44n^7-166n^6\right. \\ &\quad \left.+493n^5-79n^4-1554n^3+1380n^2+792n-864\right) \end{aligned}$$

That is, the expectation of all possible combinations of  $i, j, k$  with the restrictions of the subscript read as  $i \neq j$ ,  $i \neq k$  and  $j < k$ . Cross-product terms involving four or more subscript values have expectation zero, i.e.

$$E[V_{i,j,k}(\eta_4)V_{i,j,l}(\eta_4)] = E[V_{i,j,k}(\eta_4)V_{i,k,l}(\eta_4)] = 0$$

and so on. This allows us to find the cross-products in the second moment of  $\eta_4$  and the expectation.

We note,

$$\begin{aligned} 2 \sum_{i \neq j < k}^p \lambda_i^4 \lambda_j^2 \lambda_k^2 &= \left( \sum_{i=1}^p \lambda_i^4 \right) \left( \sum_{j=1}^p \lambda_j^2 \right)^2 - \left( \sum_{i=1}^p \lambda_i^4 \right)^2 - 2 \left( \sum_{i=1}^p \lambda_i^6 \right) \left( \sum_{j=1}^p \lambda_j^2 \right) + 2 \left( \sum_{i=1}^p \lambda_i^8 \right) \\ &= p^3 a_4 a_2^2 - p^2 a_4^2 - 2p^2 a_6 a_2 + 2pa_8 = p(p^2 a_4 a_2^2 - pa_4^2 - 2pa_6 a_2 + 2a_8) \end{aligned}$$

and

$$\begin{aligned} 2 \sum_{i \neq j < k}^p \lambda_i^2 \lambda_j^3 \lambda_k^3 &= \left( \sum_{i=1}^p \lambda_i^2 \right)^2 \left( \sum_{j=1}^p \lambda_j^3 \right) - \left( \sum_{i=1}^p \lambda_i^6 \right) \left( \sum_{i=1}^p \lambda_i^2 \right) \\ &\quad - 2 \left( \sum_{i=1}^p \lambda_i^5 \right) \left( \sum_{j=1}^p \lambda_j^3 \right) + 2 \left( \sum_{i=1}^p \lambda_i^8 \right) \\ &= p^3 a_3^2 a_2 - p^2 a_6 a_2 - 2p^2 a_5 a_3 + 2pa_8 = p(p^2 a_3^2 a_2 - pa_6 a_2 - 2pa_5 a_3 + 2a_8) \end{aligned}$$

We find the variance of  $\eta_4$  by calculating its second moment and using the expectations above

$$\begin{aligned}
Var[\eta_4] &= \frac{C_{\eta_4}(n)^2}{p^2} \left[ 4n^2(n+2)p(p^2a_4a_2^2 - pa_4^2 - 2pa_6a_8 + 2a_8)/2 \right. \\
&\quad \times (n^{10} + 12n^9 - 138n^7 - 81n^6 + 3102n^5 + 200n^4 - 5316n^3 - 912n^2 - 144n + 3456) \\
&\quad + 4(n^9 + 5n^8 - 44n^7 - 166n^6 + 493n^5 - 79n^4 - 1554n^3 + 1380n^2 + 792n - 864) \\
&\quad \left. \times n^2(n+2)(n+4)p(p^2a_3^2a_2 - pa_6a_2 - 2pa_5a_3 + 2a_8) \right] \\
&= \frac{64(n+2)}{n^{10}(n^2+n+2)^2} \left[ \frac{pa_4a_2^2 - a_4^2 - 2a_6a_2 + \frac{2}{p}a_8}{2} \right. \\
&\quad \times (n^{10} + 12n^9 - 138n^7 - 81n^6 + 3102n^5 + 200n^4 - 5316n^3 - 912n^2 - 144n + 3456) \\
&\quad + (n+4)(pa_3^2a_2 - a_6a_2 - 2a_5a_3 + \frac{2}{p}a_8) \\
&\quad \left. \times (n^9 + 5n^8 - 44n^7 - 166n^6 + 493n^5 - 79n^4 - 1554n^3 + 1380n^2 + 792n - 864) \right]
\end{aligned}$$

Likewise, it determining the variance of (A.14)

$$\begin{aligned}
\eta_5 &= \frac{8}{n^4p} \sum_{i<j<k<l}^p \lambda_i \lambda_j \lambda_k \lambda_l \left( \frac{n^2(n^2+n+2)(v_{ij}v_{jk}v_{kl}v_{il} + v_{ij}v_{jl}v_{kl}v_{ik} + v_{ik}v_{jk}v_{jl}v_{il})}{n^2(n^2+n+2)} \right. \\
&\quad - \frac{3n(n^2+n+2)(v_{ii}v_{jk}v_{jl}v_{kl} + v_{jj}v_{ik}v_{il}v_{kl} + v_{kk}v_{ij}v_{il}v_{jl} + v_{ll}v_{ij}v_{ik}v_{jk})}{n^2(n^2+n+2)} \\
&\quad - \frac{n(2n^2+3n-6)(v_{ij}^2v_{kl}^2 + v_{ik}^2v_{jl}^2 + v_{il}^2v_{jk}^2)}{n^2(n^2+n+2)} \\
&\quad + \frac{n(5n+6)(v_{ij}^2v_{kk}v_{ll} + v_{ik}^2v_{jj}v_{ll} + v_{il}^2v_{jj}v_{kk} + v_{jk}^2v_{ii}v_{ll} + v_{jl}^2v_{ii}v_{kk} + v_{kl}^2v_{ii}v_{jj})}{n^2(n^2+n+2)} \\
&\quad \left. - \frac{3(5n+6)v_{ii}v_{jj}v_{kk}v_{ll}}{n^2(n^2+n+2)} \right). \\
&= \frac{C_{\eta_5}(n)}{p} \sum_{i<j<k<l}^p \lambda_i \lambda_j \lambda_k \lambda_l V_{i,j,k,l}(\eta_5)
\end{aligned}$$

with

$$C_{\eta_5}(n) = \frac{8}{n^6(n^2+n+2)}$$

and  $V_{i,j,k,l}(\eta_5)$  represents the random component. The cross products terms have expectation zero, that is

$$E[V_{i,j,k,l}(\eta_5)V_{i,j,k,m}(\eta_5)] = 0$$

for any combination of  $i, j, k, l, m$  with 5 or more indices. The only term contributing to the variance

is

$$E[V_{i,j,k,l}(\eta_5)^2] = 3n^3(n-1)(n+2)(n^2+n+2)(n^5+6n^4+9n^3-56n^2+132n+144)$$

We note

$$\begin{aligned} 24 \sum_{i < j < k < l}^p \lambda_i^2 \lambda_j^2 \lambda_k^2 \lambda_l^2 &= p^4 a_2^4 - 6p^3 a_4 a_2^2 + 8p^2 a_6 a_2 + 3p^2 a_4^2 - 6pa_8 \\ &= p(p^3 a_2^4 - 6p^2 a_4 a_2^2 + 8pa_6 a_2 + 3pa_4^2 - 6a_8) \end{aligned}$$

and we find the variance of  $\eta_5$  by calculating its second moment

$$\begin{aligned} \text{Var}[\eta_5] &= \frac{8(n-1)(n+2)}{n^9(n^2+n+2)}(n^5+6n^4+9n^3-56n^2+132n+144) \\ &\quad \times (p^2 a_2^4 - 6pa_4 a_2^2 + 8a_6 a_2 + 3a_4^2 - \frac{6}{p} a_8) \end{aligned}$$

### A.3.5 Covariance terms

To determine the covariance terms of  $\eta_1, \eta_2, \eta_3, \eta_4,$  and  $\eta_5$  we utilize the fact that  $E[\eta_i] = 0$  for  $i = 2, 3, 4, 5$ . Therefore

$$\text{Cov}(\eta_1, \eta_2) = E[\eta_1 \eta_2]$$

and due to independence of many of the random variables, we only have to explore the variables of the form  $v_{ii}^4 V_{i,j}(\eta_2)$  and  $v_{jj}^4 V_{i,j}(\eta_2)$  (i.e.  $v_{ii}$  and  $V_{j,k}(\eta_2)$  are independent). We recall  $V_{i,j}(\eta_2)$  from (A.16) and see

$$\begin{aligned} v_{ii}^4 V_{i,j}(\eta_2) &= v_{ii}^4 (v_{ii}^2 v_{ij}^2 n_1 + v_{ii}^3 v_{jj} n_2) \\ &= v_{ii}^6 v_{ij}^2 n_1 + v_{ii}^7 v_{jj} n_2 \end{aligned}$$

where the  $v_{ii}^4$  component from  $\eta_1$  essentially adds 4 moments to the random variable. Taking expectations we see

$$\begin{aligned} E[v_{ii}^4 V_{i,j}(\eta_2)] &= n(n+2)(n+4)(n+6)(n+8)(n+10)(n+12)(n_1 + nn_2) \\ &= 0 \text{ with } n_1, n_2 \text{ defined in (A.16)}. \end{aligned}$$

A similar result holds for  $v_{jj}^4 V_{i,j}(\eta_2)$  except fourth moments of the  $v_{jj}$  are incorporated. This concept can easily be seen in Lemma A.1 in section A.2.1, specifically with the expected values of  $v_{ii}v_{ij}^2$  and  $v_{ii}^2v_{ij}^2$ . The additional  $v_{ii}$  will add an additional moment resulting in the  $(n+4)$  in the expected value. In the case of  $\eta_1$  and  $\eta_2$ , we add a fourth moment of  $v_{ii}$  and  $v_{jj}$  in the respective calculations to both parts of  $V_{i,j}(\eta_2)$ . Since both covariance terms are zero, and the other terms are zero by independence, we determine  $Cov(\eta_1, \eta_2) = 0$ . Analogous results hold for the covariance terms of  $\eta_1$  with  $\eta_3, \eta_4$  and  $\eta_5$  respectfully.

When exploring  $Cov(\eta_2, \eta_3)$  we find the terms  $V_{i,j}(\eta_2)$  and  $V_{i,j}(\eta_3)$  are correlated, along with  $V_{j,i}(\eta_2)$  and  $V_{i,j}(\eta_3)$  with the remaining terms uncorrelated. Derivation similar to that of section A.3.3 leads to the result

$$\begin{aligned} Cov(\eta_2, \eta_3) &= \frac{32n^2(n-1)(n-2)^2(n-3)^2(n+1)^2(n+2)(n+4)(n+6)^2(n+8)}{n^{12}(n^2+n+2)^2p^2} \sum_{i \neq j}^p \lambda_i^5 \lambda_j^3 \\ &= \frac{32(n-1)(n-2)^2(n-3)^2(n+1)^2(n+2)(n+4)(n+6)^2(n+8)}{n^{10}(n^2+n+2)^2} (a_5 a_3 - \frac{a_8}{p}). \end{aligned}$$

Similar expansions to find  $Cov(\eta_2, \eta_4)$  and  $Cov(\eta_2, \eta_5)$  find no terms to be correlated.

Using analogous methods, through expansion and calculation of the expectations, we see  $Cov(\eta_3, \eta_4) = 0$ ,  $Cov(\eta_3, \eta_5) = 0$  and  $Cov(\eta_4, \eta_5) = 0$ .

### A.3.6 Interaction between $\hat{a}_4$ and $\hat{a}_2$

We find the covariance, or interaction, of the terms of  $\hat{a}_4$  and  $\hat{a}_2$ . We note that the covariance between  $q_1$  of  $\hat{a}_2$  and the terms  $\eta_2, \eta_3, \eta_4$ , and  $\eta_5$  is analogous to that of  $\eta_1$  with the respective terms, resulting in

$$Cov(\eta_2, q_1) = Cov(\eta_3, q_1) = Cov(\eta_4, q_1) = Cov(\eta_5, q_1) = 0$$

Careful expansion and taking expectations, similar to that provided above, of the random variables finds there are no correlated terms between  $q_2$  and  $\eta_1, \eta_4$  or  $\eta_5$  resulting in

$$Cov(\eta_1, q_2) = Cov(\eta_4, q_2) = Cov(\eta_5, q_2) = 0$$

Expansion of  $E[\eta_2 q_2]$  and  $[\eta_3 q_2]$  and finding expectations provides the following results.

$$\begin{aligned} Cov(\eta_2, q_2) &= \frac{16n(n-1)(n-2)(n-3)(n+1)(n+2)(n+4)(n+6)}{n^8(n^2+n+2)p^2} \sum_{i \neq j}^p \lambda_i^4 \lambda_j^2 \\ &= \frac{16(n-1)(n-2)(n-3)(n+1)(n+2)(n+4)(n+6)}{n^7(n^2+n+2)} (a_4 a_2 - \frac{a_6}{p}) \end{aligned}$$

$$\begin{aligned} Cov(\eta_3, q_2) &= \frac{16n(n-1)(n-2)(n-3)(n+1)(n+2)(n+4)(n+6)}{n^8(n^2+n+2)p^2} \sum_{i < j}^p \lambda_i^3 \lambda_j^3 \\ &= \frac{8(n-1)(n-2)(n-3)(n+1)(n+2)(n+4)(n+6)}{n^7(n^2+n+2)} (a_3^2 - \frac{a_6}{p}) \end{aligned}$$

We leave the  $Cov(\eta_1, q_1)$  term for a later derivation.

#### A.4 Asymptotic Variances and Consistency

We simplify our variance and covariance terms by finding their asymptotic values under assumptions (A) and (B) and as  $(n, p) \rightarrow \infty$ .

$$\begin{aligned} V(\eta_1) &\simeq \frac{32}{np} a_8, \\ V(\eta_2) &\simeq \frac{32}{n^2} (a_6 a_2 + a_4^2 - \frac{2}{p} a_8) \simeq \frac{32}{n^2} (a_6 a_2 + a_4^2), \\ V(\eta_3) &\simeq \frac{16}{n^2} (a_4^2 - \frac{a_8}{p}) \simeq \frac{16}{n^2} a_4^2, \\ V(\eta_4) &\simeq \frac{64}{n^3} \left( \frac{p a_4 a_2^2 - (a_4^2 + 2a_6 a_2) + \frac{2}{p} a_8}{2} + p a_3^2 a_2 - (a_6 a_2 + 2a_5 a_3) + \frac{2}{p} a_8 \right) \\ &\simeq \frac{64}{n^2} c (a_4 a_2^2 / 2 + a_3^2 a_2), \\ V(\eta_5) &\simeq \frac{8}{n^4} (p^2 a_2^4 - 6p a_4 a_2^2 + (9a_6 a_2 + 3a_4^2) - \frac{6}{p} a_8) \simeq \frac{8}{n^2} c^2 a_2^4, \end{aligned}$$

and

$$\begin{aligned} V(q_1) &\simeq \frac{8}{np} a_4, \\ V(q_2) &\simeq \frac{4}{n^2} (a_2^2 - \frac{a_4}{p}) \simeq \frac{4}{n^2} a_2^2, \end{aligned}$$

are provided in Srivastava [99]. Likewise,

$$\begin{aligned}
Cov(q_1, \eta_1) &\simeq \frac{16}{np}a_6, \\
Cov(\eta_2, \eta_3) &\simeq \frac{32}{n^2}(a_5a_3 - \frac{a_8}{p}) \simeq \frac{32}{n^2}a_5a_3, \\
Cov(\eta_2, q_2) &\simeq \frac{16}{n^2}(a_4a_2 - \frac{a_6}{p}) \simeq \frac{16}{n^2}a_4a_2, \\
Cov(\eta_3, q_2) &\simeq \frac{8}{n^2}(a_3^2 - \frac{a_6}{p}) \simeq \frac{8}{n^2}a_3^2,
\end{aligned}$$

and we note that for  $\tau$  in (2.31),  $\tau^2 \simeq 1$  as  $n \rightarrow \infty$ .

Recall Chebyshev inequality

**Lemma A.13.** *Chebyshev's Inequality*

For positive  $\epsilon$  and random variable  $X$ ,

$$P[|X - E[X]| > \epsilon] \leq \frac{1}{\epsilon^2}Var[X]$$

which leads to the following result

**Theorem A.3.** *The estimator  $\hat{a}_4$  defined in (2.32) is an  $(n, p)$  consistent estimator for  $a_4$ .*

*Proof.* By Theorem A.2,  $\hat{a}_4$  is unbiased for  $a_4$ . By Chebyshev's inequality, Lemma A.13, for any  $\epsilon > 0$

$$\begin{aligned}
P[|\hat{a}_4 - a_4| > \epsilon] &\leq \frac{1}{\epsilon^2}Var[\hat{a}_4] \\
&\simeq \frac{1}{\epsilon^2} \left( \frac{32}{np}a_8 + \frac{32}{n^2}(a_6a_2 + a_4^2) + \frac{16}{n^2}(a_4^2) \right. \\
&\quad \left. + \frac{64}{n^2}(ca_4a_2^2/2 + ca_3^2a_2) + \frac{8}{n^2}(c^2a_2^4) + 2\frac{32}{n^2}(a_5a_3) \right) \\
&\rightarrow 0 \text{ as } n, p \rightarrow \infty
\end{aligned}$$

hence  $\hat{a}_4 \rightarrow a_4$  in probability and  $\hat{a}_4$  is an  $(n, p)$  consistent estimator for  $a_4$  under the two assumptions, (A) and (B).  $\square$

## A.5 Asymptotic Distribution Preliminaries

To find the asymptotic distribution of our statistic, we utilize the theory of martingale-differences.

**Lemma A.14.** *Let  $X_{n,p}$  be a sequence of random variables with  $\mathcal{F}_{n,p}$  the  $\sigma$ -field generated by the random variables  $(w_1, \dots, w_p)$ , then  $\mathcal{F}_{n,0} \subset \mathcal{F}_{n,1} \subset \dots \subset \mathcal{F}_{n,p}$ . If  $E[X_{n,p}|\mathcal{F}_{n,p-1}] = 0$  a.s. then  $(X_{n,p}, \mathcal{F}_{n,p})$  is known as a martingale-difference array. If*

$$(1) \sum_{j=0}^p E \left[ (X_{n,j})^2 | \mathcal{F}_{n,j-1} \right] \xrightarrow{P} \sigma^2 \text{ as } (n,p) \rightarrow \infty$$

$$(2) \sum_{j=0}^p E \left[ X_{n,j}^2 I(X_{n,j} > \epsilon) | \mathcal{F}_{n,j-1} \right] \xrightarrow{P} 0$$

then  $Y_{n,p} = \sum_{j=0}^p X_{n,p} \xrightarrow{D} N(0, \sigma^2)$ .

The second condition is known as the Lindeberg condition. The result can be found in numerous texts, see Durrett [29] or Shiryaev [94]. The second condition can be satisfied with the stronger Lyapounov type condition

$$\sum_{j=0}^p E[X_{n,j}^4] \rightarrow 0.$$

For the  $q_1$  and  $\eta_1$  terms, we use a more traditional central limit theorem but note they two will satisfy the martingale different requirements since they are iid. Consider

$$u_{2i} = \frac{\lambda_i^2 [v_{ii}^2 - n(n+2)]}{\sqrt{n(n+2)(n+3)}}, u_{4i} = \frac{\lambda_i^4 [v_{ii}^4 - n(n+2)(n+4)(n+6)]}{\sqrt{n(n+2)(n+4)(n+6)(n^3 + 21n^2 + 158n + 420)}} \quad (\text{A.18})$$

and

$$\epsilon_n = \frac{\sqrt{(n+4)(n+6)(n+5)}}{\sqrt{(n+3)(n+7)(n^2 + 14n + 60)}} \quad (\text{A.19})$$

then

$$E[u_{2i}] = 0, E[u_{4i}] = 0 \quad (\text{A.20})$$

$$\text{Var}[u_{2i}] = 8\lambda_i^4, \text{Var}[u_{4i}] = 32\lambda_i^8, \quad (\text{A.21})$$

$$\text{Cov}[u_{2i}, u_{4i}] = 16\lambda_i^6 \epsilon_n \quad (\text{A.22})$$

thus  $u_i = \begin{pmatrix} u_{2i} \\ u_{4i} \end{pmatrix}$  are independently distributed random vectors for  $i = 1, \dots, p$ , with the mean zero vector and the  $2 \times 2$  covariance matrix  $M_{in}$  given by

$$M_{in} = \begin{pmatrix} 8\lambda_i^4 & 16\epsilon_n\lambda_i^6 \\ 16\epsilon_n\lambda_i^6 & 32\lambda_i^8 \end{pmatrix} \quad (\text{A.23})$$

We note that

$$M_n = \frac{1}{p} \sum_{i=1}^p M_{in} = \begin{pmatrix} 8a_4 & 16\epsilon_n a_6 \\ 16\epsilon_n a_6 & 32a_8 \end{pmatrix} \neq 0 \quad (\text{A.24})$$

and for any fixed  $n$ ,  $M_n \rightarrow M_n^0 \neq 0$  as  $p \rightarrow \infty$  since each  $a_i$  is assumed to converge by assumption (A) and  $\epsilon_n$  does not depend on  $p$ .  $M_n^0$  is defined to be

$$M_n^0 = \begin{pmatrix} 8a_4^0 & 16\epsilon_n a_6^0 \\ 16\epsilon_n a_6^0 & 32a_8^0 \end{pmatrix} \quad (\text{A.25})$$

and  $M_n^0 \rightarrow M^0$  as  $n \rightarrow \infty$  since  $\epsilon_n \rightarrow 1$  with  $M^0$  defined as

$$M^0 = \begin{pmatrix} 8a_4^0 & 16a_6^0 \\ 16a_6^0 & 32a_8^0 \end{pmatrix} \quad (\text{A.26})$$

similarly as  $n \rightarrow \infty$ ,  $M_{in} \rightarrow M_i$  such that

$$M_i = \begin{pmatrix} 8\lambda_i^4 & 16\lambda_i^6 \\ 16\lambda_i^6 & 32\lambda_i^8 \end{pmatrix} \quad (\text{A.27})$$

We will make use the Lyapunov-type Central Limit Theorem from Rao [81],

**Theorem A.4.** *Let  $X_1, X_2, \dots$  be a sequence of independent  $k$  dimensional random variables such that  $E[X_i] = 0$  and  $\Sigma_i$  is the  $k \times k$  covariance matrix of  $X_i$  and suppose that*

$$\frac{1}{p} \sum_{i=1}^p \Sigma_i \rightarrow \Sigma^0 \neq 0$$



and for every  $\epsilon > 0$

$$\frac{1}{p} \sum_{i=1}^p \int_{\|X\| > \epsilon \sqrt{p}} \|X\|^2 dF_i \rightarrow 0$$

where  $F_i$  is the distribution function of  $X_i$  and  $\|X\|$  is the standard Euclidean norm of the vector  $X$ . Then the random variable  $\frac{X_1 + X_2 + \dots + X_p}{\sqrt{p}}$  converges to the  $k$ -variate normal distribution with mean zero and dispersion matrix  $\Sigma$ .

We also make use the the following inequality for random variables, found in Rao [81]

**Lemma A.15.** For random variables  $X$  and  $Y$

$$E[(X + Y)^r] \leq C_r (E[X^r] + E[Y^r])$$

where  $C_r$  satisfies

$$C_r = \begin{cases} 1 & \text{if } r \leq 1 \\ 2^{r-1} & \text{if } r > 1 \end{cases}$$

By (A.24), our covariance matrix converges to a non-zero matrix. Let  $F_i$  represent the distribution function of  $u_i$ .

$$\begin{aligned} \frac{1}{p} \sum_{i=1}^p \int_{\sqrt{(u'_i u_i)} > \sqrt{p}\epsilon} (u'_i u_i) dF_i &= \frac{1}{p} \sum_{i=1}^p \int_{(u'_i u_i) > p\epsilon^2} (u'_i u_i) dF_i \\ &\leq \frac{1}{p} \sum_{i=1}^p \frac{1}{p\epsilon^2} \int (u'_i u_i)^2 dF_i \\ &= \frac{1}{p^2 \epsilon^2} \sum_{i=1}^p E[(u_{2i}^2 + u_{4i}^2)^2] \\ &\leq \frac{2}{p^2 \epsilon^2} \sum_{i=1}^p E[u_{2i}^4 + u_{4i}^4] \end{aligned}$$

using Lemma A.15. It is easy to see that when assumption (A) is satisfied

$$\begin{aligned} \frac{2}{p^2} \sum_{i=1}^p E[u_{2i}^4] &= \frac{2}{p^2} \sum_{i=1}^p \lambda_i \frac{E[(v_i i^2 - n(n+2))^4]}{n^2(n+2)^2(n+3)^2} \\ &= f_2(n) \frac{a_8}{p} = O\left(\frac{1}{p}\right) \rightarrow 0 \text{ as } p \rightarrow \infty \end{aligned}$$

where  $f_2(n)$  is a function of  $n$ . That is, the 4th moment of the centralized  $\chi^2$  random variable only

depends on  $n$ . Likewise,

$$\begin{aligned} \frac{2}{p^2} \sum_{i=1}^p E[u_{4i}^4] &= \frac{2}{p^2} \sum_{i=1}^p \lambda_i^{16} \frac{E[(v_{ii}^4 - n(n+2)(n+4)(n+6))^2]}{n^2(n+2)^2(n+4)^2(n+6)^2(n^3+21n^2+158n+420)^2} \\ &= f_4(n) \frac{\alpha_{16}}{p} = O\left(\frac{1}{p}\right) \rightarrow 0 \text{ as } p \rightarrow \infty \end{aligned}$$

whence,

$$\frac{1}{p} \sum_{i=1}^p \int_{\sqrt{(u'_i u_i)} > \sqrt{p}\epsilon} (u'_i u_i) dF_i \leq \frac{2}{p^2 \epsilon^2} \sum_{i=1}^p E[u_{2i}^4 + u_{4i}^4] \rightarrow 0. \quad (\text{A.28})$$

By (A.24) and (A.28), we can apply Theorem A.4 and

$$\frac{1}{\sqrt{p}} \sum_{i=1}^p u_i \sim N_2(0, M_n^0) \text{ as } p \rightarrow \infty \quad (\text{A.29})$$

and we note

$$\frac{1}{\sqrt{p}} \sum_{i=1}^p u_i = \frac{1}{\sqrt{np}} \sum_{i=1}^p \left( \frac{\frac{\lambda_i^2 (v_{ii}^2 - n(n+2))}{\sqrt{(n+2)(n+3)}}}{\frac{\lambda_i^4 (v_{ii}^4 - n(n+2)(n+4)(n+6))}{\sqrt{(n+2)(n+4)(n+6)(n^3+21n^2+158n+420)}}} \right). \quad (\text{A.30})$$

As  $p \rightarrow \infty$  and  $n \rightarrow \infty$  it follows

$$\frac{1}{\sqrt{p}} \sum_{i=1}^p u_i \rightarrow N_2(0, M^0). \quad (\text{A.31})$$

Likewise by the standard multivariate Central Limit Theorem, we note that as  $n \rightarrow \infty$  with  $p$  fixed,

$$u_i \rightarrow N_2(0, M_i), i = 1, \dots, p$$

for  $M_i$  defined in (A.27). Letting

$$M = \frac{1}{p}(M_1 + \dots + M_p)$$

which will converge to  $M^0$  as  $p \rightarrow \infty$  by assumption (A). The  $u_i$  are asymptotically independently distributed, it follows from the above that as  $n \rightarrow \infty$  and  $p \rightarrow \infty$

$$\frac{1}{\sqrt{p}} \sum_{i=1}^p u_i \rightarrow N_2(0, M^0).$$

We note this allows us to interchange the convergence order and provides us with a general  $(n, p)$  convergence result.

## A.6 Asymptotic Normality of Test Statistic

Let

$$M = \frac{1}{p}(M_1 + M_2 + \dots + M_p) = \begin{pmatrix} 8a_4 & 16a_6 \\ 16a_6 & 32a_8 \end{pmatrix} \quad (\text{A.32})$$

and  $M \rightarrow M^0$  as  $p \rightarrow \infty$  due to assumption (A). Without loss of generality we replace  $M^0$  with  $M$  and note that

$$\begin{aligned} q_1 &= \frac{n-1}{n^3 p} \sum_{i=1}^p \lambda_i^2 v_{ii}^2 \\ &\simeq \frac{1}{n^2 p} \sum_{i=1}^p \lambda_i^2 v_{ii}^2 \end{aligned}$$

and

$$\begin{aligned} \eta_1 &= \frac{n^4 - 5n^3 + 5n^2 + 5n - 6}{n^2(n^2 + n + 2)} \frac{1}{n^4 p} \sum_{i=1}^p \lambda_i^4 v_{ii}^4 \\ &\simeq \frac{1}{n^4 p} \sum_{i=1}^p \lambda_i^4 v_{ii}^4. \end{aligned}$$

This leads to the following theorems

**Theorem A.5.** *As  $n$  and  $p \rightarrow \infty$*

$$\sqrt{np} \begin{pmatrix} q_1 \\ \eta_1 \end{pmatrix} \rightarrow N_2 \left( \begin{pmatrix} a_2 \\ a_4 \end{pmatrix}, M \right)$$

*Proof.* Some manipulations of the  $u_i$  in (A.30) provide the result.

$$\begin{aligned} q_1 &\simeq \frac{1}{n^2 p} \sum_{i=1}^p \lambda_i^2 v_{ii}^2 \\ &= \frac{1}{np} \sum_{i=1}^p \frac{\lambda_i^2 v_{ii}^2}{n} \end{aligned}$$

which appears as a biased version of the  $u_i$  component for  $\lambda_i^2$  multiplied by  $\frac{1}{\sqrt{np}}$  for large  $n$ . Applying

Theorem A.4 and multiplying by an additional  $\sqrt{np}$  provides the result.

$$q_1 \simeq \sqrt{np} \frac{1}{np} \sum_{i=1}^p \frac{\lambda_i^2 v_{ii}^2}{n} = \frac{1}{\sqrt{np}} \sum_{i=1}^p \frac{\lambda_i^2 v_{ii}^2}{n} \sim N(a_2, 8a_4)$$

Likewise for the  $\eta_1$  component of  $u_i$

$$\begin{aligned} \eta_1 &\simeq \frac{1}{n^4 p} \sum_{i=1}^p \lambda_i^4 v_{ii}^4 \\ &= \frac{1}{np} \sum_{i=1}^p \frac{\lambda_i^4 v_{ii}^4}{n^3} \end{aligned}$$

which appears as a biased version of the  $\lambda_i^4$  component of  $u_i$  for large  $n$ . As before,

$$\eta_1 \simeq \sqrt{np} \frac{1}{np} \sum_{i=1}^p \frac{\lambda_i^4 v_{ii}^4}{n^3} = \frac{1}{\sqrt{np}} \sum_{i=1}^p \frac{\lambda_i^4 v_{ii}^4}{n^3} \sim N(a_4, 32a_8)$$

Combining into a vector completes the proof.  $\square$

The normality of the other components require use of the Martingale Difference approach. We provide all the details of the convergence for  $\eta_2$  to normality under general asymptotics. Some of the tedious algebraic details for the other terms are not provided and are straightforward expectations, or conditional expectations. Recall that each term,  $q_2$ ,  $\eta_2$ ,  $\eta_3$ ,  $\eta_4$ , and  $\eta_5$ , was constructed to have mean zero. With assumption (B), the argument for the asymptotic normality of each term follows the same methodology. We further note that the variance of each term is written with a  $\frac{p}{n}$  term rather than the concentration. We keep the term written this way as it will disappear following an application of the delta method in later work. First we recall a result from Srivastava [99],

**Theorem A.6.** *Under assumption (A) and (B), for  $q_2$  defined above*

$$\sqrt{np} q_2 \xrightarrow{D} N\left(0, 4 \frac{p}{n} a_2^2\right) \quad (\text{A.33})$$

**Theorem A.7.** *Under assumptions (A) and (B), for  $\eta_2$  defined in (A.11)*

$$\sqrt{np} \eta_2 \xrightarrow{D} N\left(0, 32 \frac{p}{n} (a_6 a_2 + a_4^2)\right) \quad (\text{A.34})$$

*Proof.* Define the  $\sigma$ -field generated by the random variables  $\mathcal{F}_{n,j} = \sigma\{w_1, w_2, \dots, w_j\}$ . We then

rewrite  $\eta_2$  as

$$\begin{aligned} n\eta_2 &= \frac{K_2}{p} \sum_{i \neq j}^p \lambda_i^3 \lambda_j V_{ij} = \frac{K_2}{p} \sum_{i < j}^p \lambda_i^3 \lambda_j V_{ij} + \lambda_i \lambda_j^3 V_{ji} \\ &= \frac{K_2}{p} \sum_{j=2}^p \sum_{i=1}^{j-1} \lambda_i^3 \lambda_j V_{ij} + \lambda_i \lambda_j^3 V_{ji} \end{aligned}$$

using the notation from Section A.3.3 and

$$K_2 = \frac{4}{n^5(n^2 + n + 2)}.$$

Define

$$X_{n,j} = \sum_{i=1}^{j-1} \frac{K_2}{p} (\lambda_i^3 \lambda_j V_{ij} + \lambda_i \lambda_j^3 V_{ji}) = \sum_{i=1}^{j-1} Y_{ij}$$

and note the following conditional expectations

$$\begin{aligned} E[v_{ii}^2 v_{ij}^2 | \mathcal{F}_{n,j-1}] &= v_{ii}^3, & E[v_{ii}^3 v_{jj} | \mathcal{F}_{n,j-1}] &= n v_{ii}^3, \\ E[v_{jj}^2 v_{ij}^2 | \mathcal{F}_{n,j-1}] &= (n+2)(n+4)v_{ii}, & E[v_{jj}^3 v_{ii} | \mathcal{F}_{n,j-1}] &= n(n+2)(n+4)v_{ii}, \end{aligned}$$

hence  $E[V_{ij} | \mathcal{F}_{n,j-1}] = 0$  and  $E[V_{ji} | \mathcal{F}_{n,j-1}] = 0$ , so  $E[X_{n,j} | \mathcal{F}_{n,j-1}] = 0$ . Following the methodology from Section A.3.3 we can calculate and show that condition (1) from Lemma A.14 holds. Begin by noting

$$X_{n,j}^2 = \sum_{i=1}^{j-1} Y_{ij}^2 + 2 \sum_{i < k}^{j-1} Y_{ij} Y_{kj}. \quad (\text{A.35})$$

We use the well known result about expectations

$$E[E[X_{n,j}^2 | \mathcal{F}_{n,j-1}]] = E[X_{n,j}^2]$$

and

$$\begin{aligned} E[X_{n,j}^2] &= E \left[ \left( \sum_{i=1}^{j-1} Y_{ij} \right)^2 \right] \\ &= \sum_{i=1}^{j-1} E[Y_{ij}^2] + 2 \sum_{i < k}^{j-1} E[Y_{ij} Y_{kj}]. \end{aligned}$$

By the methodology from Section A.3.3,

$$E[Y_{ij}Y_{kj}] = 0$$

and for large  $n$

$$\begin{aligned} \sum_{i=1}^{j-1} E[Y_{ij}^2] &= \sum_{i=1}^{j-1} \frac{K_2^2}{p^2} (\lambda_i^6 \lambda_j^2 E[V_{ij}^2] + 2\lambda_i^4 \lambda_j^4 E[V_{ij}V_{ji}] + \lambda_i^2 \lambda_j^6 E[V_{ji}^2]) \\ &= \sum_{i=1}^{j-1} \frac{16}{O(n^{14})} \frac{1}{p^2} (\lambda_i^6 \lambda_j^2 O(n^{14}) + 2\lambda_i^4 \lambda_j^4 O(n^{14}) + \lambda_i^2 \lambda_j^6 O(n^{14})) \\ &= \frac{16}{p} (\lambda_j^2 a_6 + 2\lambda_j^4 a_4 + \lambda_j^6 a_2). \end{aligned}$$

Hence

$$\begin{aligned} E \left[ \sum_{j=2}^p E[X_{n,j}^2 | \mathcal{F}_{n,j-1}] \right] &= \sum_{j=2}^p E[X_{n,j}^2] \\ &= \sum_{j=2}^p \frac{16}{p} (\lambda_j^2 a_6 + 2\lambda_j^4 a_4 + \lambda_j^6 a_2) \\ &= 32(a_6 a_2 + a_4^2). \end{aligned}$$

If we can show that

$$\text{Var} \left[ \sum_{j=2}^p E[X_{n,j}^2 | \mathcal{F}_{n,j-1}] \right] \rightarrow 0 \text{ as } (n, p) \rightarrow \infty,$$

then by the law of large numbers, condition (1) of Lemma A.14 will be satisfied.

Using (A.35) we can find the conditional expectation given the  $\sigma$ -field  $\mathcal{F}_{n,j-1}$ . It is fairly straightforward to show

$$E[Y_{ij}Y_{kj} | \mathcal{F}_{n,j-1}] = 0.$$

Hence

$$E[X_{n,j}^2 | \mathcal{F}_{n,j-1}] = \sum_{i=1}^{j-1} E[Y_{ij}^2 | \mathcal{F}_{n,j-1}]$$

and for large  $n$

$$E[Y_{ij}^2 | \mathcal{F}_{n,j-1}] = \frac{K_2^2}{p^2} (\lambda_i^6 \lambda_j^2 O(n^8) v_{ii}^6 + 2\lambda_i^4 \lambda_j^4 O(n^{10}) v_{ii}^4 + \lambda_i^2 \lambda_j^6 O(n^{12}) v_{ii}^2). \quad (\text{A.36})$$

Utilizing the triangular sequencing allows us to compute

$$\text{Var} \left[ \sum_{j=2}^p E[X_{n,j}^2 | \mathcal{F}_{n,j-1}] \right] = \sum_{i=1}^{p-1} \text{Var} \left[ \sum_{j=i+1}^p E[Y_{ij}^2 | \mathcal{F}_{n,j-1}] \right]$$

and from (A.36)

$$\begin{aligned} \sum_{i=1}^{p-1} \text{Var} \left[ \sum_{j=i+1}^p E[Y_{ij}^2 | \mathcal{F}_{n,j-1}] \right] &\leq \frac{K_2^4}{p^2} \sum_{i=1}^p \left( \lambda_i^{12} a_2^2 O(n^{16}) \text{Var}[v_{ii}^6] + \lambda_i^8 a_4^2 O(n^{20}) \text{Var}[v_{ii}^4] \right. \\ &\quad + \lambda_i^4 a_6^2 O(n^{24}) \text{Var}[v_{ii}^2] + \lambda_i^{10} a_2 a_4 O(n^{18}) \text{Cov}(v_{ii}^6, v_{ii}^4) \\ &\quad \left. + \lambda_i^8 a_2 a_6 O(n^{20}) \text{Cov}(v_{ii}^6, v_{ii}^2) + \lambda_i^6 a_6 a_4 O(n^{22}) \text{Cov}(v_{ii}^4, v_{ii}^2) \right) \end{aligned}$$

Its straightforward to calculate  $\text{Var}[v_{ii}^6] = O(n^{11})$  and the other variance and covariance terms. When including the  $K_2^4 = O(n^{-28})$  it is easy to see that  $\text{Var} [E[Y_{ij}^2 | \mathcal{F}_{n,j-1}]] = O(n^{-1} p^{-2})$ . From here its clear that  $\text{Var} \left[ \sum_{j=2}^p E[X_{n,j}^2 | \mathcal{F}_{n,j-1}] \right] = O((np)^{-1}) \rightarrow 0$  as  $(n, p) \rightarrow \infty$ .

To show the Lyapounov type condition consider

$$X_{n,j}^4 = \sum_{i=1}^{j-1} Y_{ij}^4 + 4 \sum_{i \neq k}^{j-1} Y_{ij}^3 Y_{kj} + 6 \sum_{i < k}^{j-1} Y_{ij}^2 Y_{kj}^2 + 12 \sum_{i \neq k < l}^{j-1} Y_{ij}^2 Y_{kj} Y_{lj} + 24 \sum_{i < k < l < m}^{j-1} Y_{ij} Y_{kj} Y_{lj} Y_{mj}$$

and only the  $Y_{ij}^4$  and  $Y_{ij}^2 Y_{kj}^2$  have non-zero expectation. For large  $n$

$$\sum_{i=1}^{j-1} E[Y_{ij}^4] = \frac{256}{p^3} (\lambda_j^4 a_{12} + 4\lambda_j^6 a_{10} + 6\lambda_j^8 a_8 + 4\lambda_j^{10} a_6 + \lambda_j^{12} a_4)$$

and

$$\sum_{i < k}^{j-1} E[Y_{ij}^2 Y_{kj}^2] = \frac{256}{p^2} (\lambda_j^4 a_6^2 + 4\lambda_j^6 a_4 a_6 + 4\lambda_j^8 a_4^2 + 2\lambda_j^8 a_2 a_6 + 4\lambda_j^{10} a_2 a_4 + \lambda_j^{12} a_2^2).$$

Hence

$$\sum_{j=2}^p E[X_{n,j}^4] = \frac{256}{p^2} O(1) + \frac{256}{p} O(1) \rightarrow 0 \text{ as } (n, p) \rightarrow \infty$$

and we have satisfied the second condition of Lemma A.14. This completes the proof for the asymptotic normality of  $n\eta_2$ . We utilize assumption (B) to rewrite our result with the same convergent rate of  $\sqrt{np}$ .  $\square$

**Theorem A.8.** Under assumptions (A) and (B), for  $\eta_3$  defined in (A.12)

$$\sqrt{np}\eta_3 \xrightarrow{D} N\left(0, 16\frac{p}{n}a_4^2\right) \quad (\text{A.37})$$

*Proof.* This proof follows that of  $\eta_2$ , we begin by writing

$$n\eta_3 = \frac{K_3}{p} \sum_{j=2}^p \sum_{i=1}^{j-1} \lambda_i^2 \lambda_j^2 W_{ij} = \sum_{j=2}^p X_{n,j}$$

with

$$K_3 = nC_{\eta_3}(n) = \frac{2}{n^5(n^2 + n + 2)}$$

and  $W_{ij} = V_{i,j}(\eta_3)$  being the random components of  $\eta_3$ . With respect to the  $\sigma$ -field generated as  $\mathcal{F}_{n,j} = \sigma\{w_1, \dots, w_j\}$  we find the conditional expectations

$$\begin{aligned} E[v_{ii}v_{ij}^2v_{jj}|\mathcal{F}_{n,j-1}] &= v_{ii}^2(n+2) \\ E[v_{ij}^4|\mathcal{F}_{n,j-1}] &= 3v_{ii}^2 \\ E[v_{ii}^2v_{jj}^2|\mathcal{F}_{n,j-1}] &= v_{ii}^2n(n+2) \end{aligned}$$

and thus  $E[W_{ij}|\mathcal{F}_{n,j-1}] = 0$  and  $\eta_3$  satisfies the conditions of martingale difference.

We note for large  $n$

$$\begin{aligned} E\left[\left(\sum_{i=1}^{j-1} \frac{K_3}{p} \lambda_i^2 \lambda_j^2 W_{ij}\right)^2\right] &= \frac{K_3^2}{p^2} \sum_{i=1}^{j-1} \lambda_i^4 \lambda_j^4 E[W_{ij}^4] \\ &= \frac{16}{p} \lambda_j^4 a_4 \end{aligned}$$

and from this we see

$$E\left[\sum_{j=2}^p X_{n,j}^2\right] = 16a_4^2$$

Like in  $\eta_2$ , if we can show that

$$\text{Var}\left[\sum_{j=2}^p E[X_{n,j}^2|\mathcal{F}_{n,j-1}]\right] \rightarrow 0 \text{ as } (n, p) \rightarrow \infty$$



then by the law of large numbers condition 1 of Lemma A.14 will be satisfied. It is straightforward to show

$$E[W_{ij}W_{kj}|\mathcal{F}_{n,j-1}] = 0.$$

and hence

$$E[X_{n,j}^2|\mathcal{F}_{n,j-1}] = \sum_{i=1}^p \lambda_i^4 \lambda_j^4 E[W_{ij}^2|\mathcal{F}_{n,j-1}].$$

Again we use the triangular sequencing of the summation and see

$$\text{Var} \left[ \sum_{j=2}^p E[X_{n,j}^2|F_{n,j-1}] \right] = \sum_{i=1}^{p-1} \text{Var} \left[ \sum_{j=i+1}^p E[W_{ij}^2|\mathcal{F}_{n,j-1}] \right].$$

As with the  $\eta_2$  component, when including the  $K_3^4$  term, its a straightforward calculation to determine that  $\text{Var} \left[ \sum_{j=i+1}^p E[W_{ij}^2|\mathcal{F}_{n,j-1}] \right] \leq O(n^{-1}p^{-2})$  and hence  $\text{Var} \left[ \sum_{j=2}^p E[X_{n,j}^2|F_{n,j-1}] \right] \leq O((np)^{-1})$ . Hence we know the first condition of Lemma A.14 is satisfied. To show condition two, consider the Lyapounov condition.

We begin by noting that for large  $n$

$$\begin{aligned} E \left[ \left( \sum_{i=1}^{j-1} \frac{K_3}{p} \lambda_i^2 \lambda_j^2 W_{ij} \right)^4 \right] &= \frac{K_3^4}{p^4} \left( \sum_{i=1}^{j-1} \lambda_i^8 \lambda_j^8 E[W_{ij}^4] + 6 \sum_{i < k}^{j-1} \lambda_i^4 \lambda_k^4 \lambda_j^8 E[W_{ij}^2 W_{kj}^2] \right) \\ &= \frac{256}{p^3} \lambda_j^8 a_8 + \frac{256}{p^2} \lambda_j^8 a_4^2 \end{aligned}$$

and thus

$$E \sum_{j=2}^p [X_{n,j}^4] = 256a_8^2 O\left(\frac{1}{p^2}\right) + 256a_8a_4^2 O\left(\frac{1}{p}\right) \rightarrow 0.$$

All the conditions of Lemma A.14 are satisfied and this completes the proof. Using assumption (B) we can write the result with a  $\sqrt{np}$  convergent rate.  $\square$

**Theorem A.9.** *Under assumptions (A) and (B), for  $\eta_4$  defined in (A.13)*

$$\sqrt{np}\eta_4 \xrightarrow{D} N \left( 0, 64c \frac{p}{n} \left( \frac{a_4 a_2^2}{2} + a_3^2 a_2 \right) \right) \quad (\text{A.38})$$

*Proof.* We begin the proof by first rewriting  $\eta_4$  as

$$n\eta_4 = \frac{K_4}{p} \sum_{k=3}^p \sum_{j=2}^{k-1} \sum_{i=1}^{j-1} \lambda_i^2 \lambda_j \lambda_k V_{ijk} + \lambda_i \lambda_j^2 \lambda_k V_{jik} + \lambda_i \lambda_j \lambda_k^2 V_{kij} = \sum_{k=3}^p X_{n,k}$$

with

$$K_4 = \frac{4}{n^5(n^2 + n + 2)}$$

and  $V_{ijk}$ ,  $V_{jik}$  and  $V_{kij}$  being the random components. Define the  $\sigma$ -field generated by the first  $k$  random variables,  $\mathcal{F}_{n,k} = \sigma\{w_i, \dots, w_k\}$  and straightforward calculations show

$$\begin{aligned} E[v_{ii}^2 v_{jk}^2 | \mathcal{F}_{n,k-1}] &= v_{ii}^2 v_{jj} \\ E[v_{ij}^2 v_{ik}^2 | \mathcal{F}_{n,k-1}] &= v_{ii} v_{ij}^2 \\ E[v_{ii} v_{ij}^2 v_{kk} | \mathcal{F}_{n,k-1}] &= v_{ii} v_{ij}^2 n \\ E[v_{ii} v_{ik}^2 v_{jj} | \mathcal{F}_{n,k-1}] &= v_{ii}^2 v_{jj} \\ E[v_{ii}^2 v_{jj} v_{kk} | \mathcal{F}_{n,k-1}] &= v_{ii}^2 v_{jj} n \\ E[v_{ii} v_{ij} v_{ik} v_{jk} | \mathcal{F}_{n,k-1}] &= v_{ii} v_{ij}^2 \end{aligned}$$

and hence  $E[V_{ijk} | \mathcal{F}_{n,k-1}] = 0$ . Likewise

$$\begin{aligned} E[v_{jj}^2 v_{ik}^2 | \mathcal{F}_{n,k-1}] &= v_{jj}^2 v_{ii} \\ E[v_{ij}^2 v_{jk}^2 | \mathcal{F}_{n,k-1}] &= v_{jj} v_{ij}^2 \\ E[v_{jj} v_{ij}^2 v_{kk} | \mathcal{F}_{n,k-1}] &= v_{jj} v_{ij}^2 n \\ E[v_{jj} v_{jk}^2 v_{ii} | \mathcal{F}_{n,k-1}] &= v_{jj}^2 v_{ii} \\ E[v_{jj}^2 v_{ii} v_{kk} | \mathcal{F}_{n,k-1}] &= v_{jj}^2 v_{ii} n \\ E[v_{jj} v_{ij} v_{jk} v_{ik} | \mathcal{F}_{n,k-1}] &= v_{jj} v_{ij}^2 \end{aligned}$$

and hence  $E[V_{jik} | \mathcal{F}_{n,k-1}] = 0$ . The conditional expectation of the  $V_{kij}$  component can be tricky. We

note the first few conditional expectations

$$\begin{aligned}
E[v_{kk}^2 v_{ij}^2 | \mathcal{F}_{n,k-1}] &= v_{ij}^2 n(n+2) \\
E[v_{kk} v_{jk}^2 v_{ii} | \mathcal{F}_{n,k-1}] &= v_{ii} v_{jj} (n+2) \\
E[v_{kk} v_{ik}^2 v_{jj} | \mathcal{F}_{n,k-1}] &= v_{ii} v_{jj} (n+2) \\
E[v_{kk}^2 v_{ii} v_{jj} | \mathcal{F}_{n,k-1}] &= v_{ii} v_{jj} n(n+2) \\
E[v_{kk} v_{ik} v_{jk} v_{ij} | \mathcal{F}_{n,k-1}] &= v_{ij}^2 (n+2)
\end{aligned}$$

as they are straightforward with an application of Lemma A.5. To find the conditional expectation of the  $v_{ik}^2 v_{jk}^2$  term we perform the following calculation

$$E[v_{ik}^2 v_{jk}^2 | \mathcal{F}_{n,k-1}] = E \left[ \left( \sum_{m=1}^n x_{im} x_{km} \right)^2 \left( \sum_{m=1}^n x_{jm} x_{km} \right)^2 \middle| \mathcal{F}_{n,k-1} \right].$$

This comes from the definition of the  $w_i$ s. By expanding out the summations and calculating the individual conditional expectations we see many of the terms are zero. We are left with

$$E[v_{ik}^2 v_{jk}^2 | \mathcal{F}_{n,k-1}] = 2v_{ij}^2 + v_{ii} v_{jj}.$$

From here, we see that  $E[V_{kij} | \mathcal{F}_{n,k-1}] = 0$  and hence  $\eta_4$  satisfies the conditions to be a martingale difference. As before, we will consider the expectation of the second moment. We note we need to use assumption (B),  $\frac{p}{n} \simeq c$  for large  $(n, p)$ . Consider for large  $n$ ,

$$\begin{aligned}
E[E[X_{n,k}^2 | \mathcal{F}_{n,k-1}]] &= E[X_{n,k}^2] \\
&= \frac{16}{O(n^{14})} \frac{1}{p^2} \sum_{j=2}^{k-1} \sum_{i=1}^{j-1} O(4n^{13}) \left( \lambda_i^4 \lambda_j^2 \lambda_k^2 + 2\lambda_i^3 \lambda_j^3 \lambda_k^2 + 2\lambda_i^3 \lambda_j^2 \lambda_k^3 + \lambda_i^2 \lambda_j^4 \lambda_k^2 \right. \\
&\quad \left. + 2\lambda_i^2 \lambda_j^3 \lambda_k^3 + \lambda_i^2 \lambda_j^2 \lambda_k^4 \right) \\
&= \frac{64}{p} c \left( \frac{a_4 a_2 \lambda_k^2}{2} + a_3^2 \lambda_k^2 + a_2 a_3 \lambda_k^3 + \frac{a_2^2 \lambda_k^4}{2} \right)
\end{aligned}$$

which leads to the following result

$$E \left[ \sum_{k=3}^p E[X_{n,k}^2 | \mathcal{F}_{n,k-1}] \right] = \sum_{k=3}^p E[X_{n,k}^2] = 64c \left( \frac{a_4 a_2^2}{2} + a_3^2 a_2 \right).$$

As with the previous proofs, if we can show the variance goes to zero, we have satisfied condition 1 of Lemma A.14. Based on how we rewrote  $\eta_4$  above, elementary calculations show that

$$E[X_{n,k}^2 | \mathcal{F}_{n,k-1}] = \sum_{j=2}^{k-1} \sum_{i=1}^{j-1} E[Y_{ijk}^2 | \mathcal{F}_{n,k-1}]$$

where the  $Y_{ijk}^2$  represents the eigenvalues and random components. Thus

$$\text{Var} \left[ \sum_{k=3}^p E[X_{n,k-1} | \mathcal{F}_{n,k-1}] \right] = \sum_{j=2}^p \sum_{i=1}^{j-1} \text{Var} \left[ \sum_{k=3}^p E[Y_{ijk}^2 | \mathcal{F}_{n,k-1}] \right]$$

As before, we bound the Variance term by the second moment. Using the methodology similar to calculating the variance of  $\eta_4$ , recalling  $K_4$ , calculations show that under assumptions (A)

$$\text{Var} \left[ \sum_{k=3}^p E[Y_{ijk}^2 | \mathcal{F}_{n,k-1}] \right] \leq O \left( \frac{1}{p^2 n^2} \right).$$

Now with assumption (B) and handling the double summation, we see

$$\text{Var} \left[ \sum_{k=3}^p E[X_{n,k-1} | \mathcal{F}_{n,k-1}] \right] \leq O \left( \frac{1}{np} \right)$$

and hence by the Law of Large numbers condition 1 is satisfied.

Using similar methods to above, and like that of the previous proofs. Utilizing, assumption (B),

$$\sum_{k=3}^p E[X_{n,k}^4] = O \left( \frac{1}{p} \right) \rightarrow 0.$$

Hence condition 2 is also satisfied. Thus by the Martingale-Difference approach,  $n\eta_4$  is approximately a normal random variable. Utilizing assumption (B) allows us to write it with the same convergence rate as  $q_1$  and  $\eta_1$ . □

**Theorem A.10.** *Under assumptions (A) and (B), for  $\eta_5$  defined in (A.14)*

$$\sqrt{np}\eta_5 \xrightarrow{D} N \left( 0, 8 \frac{p}{n} c^2 a_2^4 \right) \tag{A.39}$$

*Proof.* Rewrite  $\eta_5$  as

$$n\eta_5 = \frac{K_5}{p} \sum_{l=4}^p \sum_{k=3}^{l-1} \sum_{j=2}^{k-1} \sum_{i=1}^{j-1} \lambda_i \lambda_j \lambda_k \lambda_l V_{ijkl} = \sum_{l=4}^p X_{n,l}$$

with

$$K_5 = \frac{8}{n^5(n^2 + n + 2)}.$$

We calculate the following conditional expectations when conditioning on the  $\sigma$ -field defined as

$$\mathcal{F}_{n,l} = \sigma\{w_1, \dots, w_l\},$$

$$\begin{aligned} E[v_{ij}v_{jk}v_{kl}v_{il}|\mathcal{F}_{n,l-1}] &= v_{ij}v_{jk}v_{ik} \\ E[v_{ij}v_{jl}v_{kl}v_{ik}|\mathcal{F}_{n,l-1}] &= v_{ij}v_{jk}v_{ik} \\ E[v_{ik}v_{jk}v_{jl}v_{il}|\mathcal{F}_{n,l-1}] &= v_{ij}v_{jk}v_{ik} \\ E[v_{ii}v_{jk}v_{jl}v_{kl}|\mathcal{F}_{n,l-1}] &= v_{ii}v_{jk}^2 \\ E[v_{jj}v_{ik}v_{il}v_{kl}|\mathcal{F}_{n,l-1}] &= v_{jj}v_{ik}^2 \\ E[v_{kk}v_{ij}v_{il}v_{jl}|\mathcal{F}_{n,l-1}] &= v_{kk}v_{ij}^2 \\ E[v_{ll}v_{ij}v_{ik}v_{jk}|\mathcal{F}_{n,l-1}] &= n \cdot v_{ij}v_{ik}v_{jk} \\ E[v_{ij}^2v_{kl}^2|\mathcal{F}_{n,l-1}] &= v_{ij}^2v_{kk} \\ E[v_{ik}^2v_{jl}^2|\mathcal{F}_{n,l-1}] &= v_{ik}^2v_{jj} \\ E[v_{il}^2v_{jk}^2|\mathcal{F}_{n,l-1}] &= v_{jk}^2v_{ii} \\ E[v_{ij}^2v_{kk}v_{ll}|\mathcal{F}_{n,l-1}] &= n \cdot v_{ij}^2v_{kk} \\ E[v_{ik}^2v_{jj}v_{ll}|\mathcal{F}_{n,l-1}] &= n \cdot v_{ik}^2v_{jj} \\ E[v_{il}^2v_{jj}v_{kk}|\mathcal{F}_{n,l-1}] &= v_{ii}v_{jj}v_{kk} \\ E[v_{jk}^2v_{ii}v_{ll}|\mathcal{F}_{n,l-1}] &= n \cdot v_{jk}^2v_{ii} \\ E[v_{jl}^2v_{ii}v_{kk}|\mathcal{F}_{n,l-1}] &= v_{ii}v_{jj}v_{kk} \\ E[v_{kl}^2v_{ii}v_{jj}|\mathcal{F}_{n,l-1}] &= v_{ii}v_{jj}v_{kk} \\ E[v_{ii}v_{jj}v_{kk}v_{ll}|\mathcal{F}_{n,l-1}] &= n \cdot v_{ii}v_{jj}v_{kk}. \end{aligned}$$

Thus we calculate  $E[X_{n,l}|\mathcal{F}_{n,l-1}] = 0$  and we can apply the martingale difference central limit

theorem. Straightforward calculations show

$$E \left[ \sum_{l=4}^p E[X_{n,l}^2 | \mathcal{F}_{n,l-1}] \right] = \sum_{l=4}^p E[X_{n,l}^2] = 8c^2 a_2^4$$

Its fairly straightforward to show

$$\text{Var} \left[ \sum_{l=4}^p E[X_{n,l}^2 | \mathcal{F}_{n,l-1}] \right] = \frac{K_5^4}{p^4} \sum_{k=3}^{p-1} \sum_{j=2}^{k-1} \sum_{i=1}^{j-1} \text{Var} \left[ \sum_{l=4}^p \lambda_i^2 \lambda_j^2 \lambda_k^2 \lambda_l^4 E[V_{ijkl}^2 | \mathcal{F}_{n,l-1}] \right].$$

Recalling assumption (B) to handle all the summations, we find

$$\text{Var} \left[ \sum_{l=4}^p E[X_{n,l}^2 | \mathcal{F}_{n,l-1}] \right] \leq O \left( \frac{1}{np} \right) \rightarrow 0,$$

Using similiar methods we find

$$\sum_{l=4}^p E[X_{n,l}^4] = O \left( \frac{1}{p} \right) \rightarrow 0$$

and hence both conditions of Lemma A.14 are satisfied. We utilize assumption (B) to express the result with a  $\sqrt{np}$  convergence rate completes the result.  $\square$

The marginal distribution of each term and the martingale-difference approach provides the following important result,

**Theorem A.11.** *Under assumptions (A) and (B), as  $(n, p) \rightarrow \infty$*

$$\sqrt{np} \begin{pmatrix} q_1 \\ \eta_1 \\ q_2 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \end{pmatrix} \xrightarrow{D} N \left( \begin{pmatrix} a_2 \\ a_4 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{q_1}^2 & \sigma_{q_1 \eta_1} & 0 & 0 & 0 & 0 & 0 \\ \sigma_{q_1 \eta_1} & \sigma_{\eta_1}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{q_2}^2 & \sigma_{q_2 \eta_2} & \sigma_{q_2 \eta_3} & 0 & 0 \\ 0 & 0 & \sigma_{q_2 \eta_2} & \sigma_{\eta_2}^2 & \sigma_{\eta_2 \eta_3} & 0 & 0 \\ 0 & 0 & \sigma_{q_2 \eta_3} & \sigma_{\eta_2 \eta_3} & \sigma_{\eta_3}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{\eta_4}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{\eta_5}^2 \end{pmatrix} \right),$$

where  $\sigma_{q_1}^2$ ,  $\sigma_{\eta_1}^2$  and  $\sigma_{q_1 \eta_1}$  are found in the  $M$  matrix defined in (A.32) and  $\sigma_{q_2}^2$ ,  $\sigma_{\eta_2}^2$ ,  $\sigma_{\eta_3}^2$ ,  $\sigma_{\eta_4}^2$ , and  $\sigma_{\eta_5}^2$  are the asymptotic variances in (A.33), (A.34), (A.37), (A.38) and (A.39) respectively and  $\sigma_{q_2 \eta_2}$ ,  $\sigma_{q_2 \eta_3}$ , and  $\sigma_{\eta_2 \eta_3}$  are the asymptotic covariance terms of  $q_2$  and  $\eta_2$  and  $\eta_3$  defined in section A.3.6

with the  $\sqrt{np}$  convergence rate.

*Proof.* Consider a set of arbitrary non-zero constants  $k_i$ s such that

$$\sqrt{np}K = \sqrt{np}(k_1(q_1 - a_2) + k_2(\eta_1 - a_4) + k_3q_2 + k_4\eta_2 + k_5\eta_3 + k_6\eta_4 + k_7\eta_5)$$

and without loss of generality,  $k_1 + \dots + k_7 = 1$ . With respect to the increasing set of  $\sigma$ -fields,  $F_{n,l} = \sigma\{w_1, \dots, w_l\}$  we note that  $K$  will satisfy the conditions of Lemma A.14 since each term also satisfies the requirements. Calculate the variance of  $K$  (condition (1)) by computing the variance and covariance of the individual terms. To satisfy condition (2) of Lemma A.14 we will use the well known inequality for any random variables  $Y_1, \dots, Y_n$

$$E \left| \sum_{i=1}^n Y_i \right|^p \leq n^{p-1} \sum_{i=1}^n E[|Y_i|^p]$$

and the Lyapounov condition. That is,

$$E[K^4] \leq 7^3 (k_1^4 E[(q_1 - a_2)^4] + \dots + k_7^4 E[\eta_5^4])$$

and if each component goes to zero, then the fourth moment of  $K$  will also go to zero. This results in asymptotic joint-normality.  $\square$

A simple linear transformation provides the result

**Corollary A.1.**

$$\sqrt{np} \begin{pmatrix} \hat{a}_2 \\ \hat{a}_4 \end{pmatrix} \xrightarrow{D} N_2 \left( \begin{pmatrix} a_2 \\ a_4 \end{pmatrix}, \begin{pmatrix} \sigma_2^2 & \sigma_{24} \\ \sigma_{24} & \sigma_4^2 \end{pmatrix} \right)$$

where  $\sigma_2^2$ ,  $\sigma_{24}$  and  $\sigma_4^2$  are the asymptotic variance and covariance of  $\hat{a}_2$  and  $\hat{a}_4$  defined as

$$\sigma_2^2 = 8a_4 + 4\frac{p}{n}a_2^2$$

$$\sigma_{24} = 16a_6 + \frac{16p}{n}a_4a_2 + \frac{8p}{n}a_3^2$$

and

$$\begin{aligned}
\sigma_4^2 &= 32a_8 + \frac{32}{n}(pa_6a_2 + pa_4^2) + \frac{16}{n}pa_4^2 \\
&\quad + \frac{64pa_5a_3}{n} + \frac{32pca_4a_2^2}{n} + \frac{64pca_3^2a_2}{n} + \frac{8pc^2a_2^4}{n} \\
&= 32a_8 + \frac{48p}{n}a_4^2 + \frac{32p}{n}a_6a_2 + \frac{64p}{n}a_5a_3 + \frac{32p}{n}a_4a_2^2c + \frac{64p}{n}a_3^2a_2c + \frac{8p}{n}a_2^4c^2
\end{aligned}$$

Here we include a  $\frac{p}{n}$  term in many of the terms, but note they will go away following an application of the delta method and some algebra.

We note our test statistic is

$$T = \hat{\psi}_2 - 1 = \frac{\hat{a}_4}{\hat{a}_2^2} - 1$$

and recalling that  $\hat{a}_4$  and  $\hat{a}_2$  are  $(n, p)$ -consistent estimators for  $a_4$  and  $a_2$  respectively, an application of the delta method leads to the result.

**Theorem A.12.** For large  $(n, p)$ ,  $T$  is approximately distributed as

$$T \sim N(\psi_2 - 1, \xi^2)$$

where

$$\xi^2 = \frac{32a_8}{npa_2^4} + \frac{32a_6}{n^2a_2^3} + \frac{32a_4^3}{npa_2^6} + \frac{32ca_4}{n^2a_2^2} + \frac{64ca_3^2}{n^2a_2^3} - \frac{64a_4a_6}{npa_2^5} - \frac{32a_4a_3^2}{n^2a_2^5} + \frac{64a_5a_3}{n^2a_2^4} + \frac{8c^2}{n^2}$$

*Proof.*

$$\frac{\partial T}{\partial \hat{a}_2} = -\frac{2\hat{a}_4}{\hat{a}_2^3} \quad \text{and} \quad \frac{\partial T}{\partial \hat{a}_4} = \frac{1}{\hat{a}_2^2}$$

Thus asymptotically,  $T \sim N(\psi_2 - 1, \xi^2)$  where

$$\begin{aligned}
\xi^2 &= \frac{1}{np} \left( -\frac{2\hat{a}_4}{\hat{a}_2^3}, \frac{1}{\hat{a}_2^2} \right) \begin{pmatrix} \sigma_2^2 & \sigma_{24} \\ \sigma_{24} & \sigma_4^2 \end{pmatrix} \begin{pmatrix} -\frac{2\hat{a}_4}{\hat{a}_2^3} \\ \frac{1}{\hat{a}_2^2} \end{pmatrix} \\
&\simeq \frac{32a_8}{npa_2^4} + \frac{32a_6}{n^2a_2^3} + \frac{32a_4^3}{npa_2^6} + \frac{32ca_4}{n^2a_2^2} + \frac{64ca_3^2}{n^2a_2^3} - \frac{64a_4a_6}{npa_2^5} - \frac{32a_4a_3^2}{n^2a_2^5} + \frac{64a_5a_3}{n^2a_2^4} + \frac{8c^2}{n^2}
\end{aligned}$$



and we note that under  $H_0$ , where  $\lambda_i = \lambda$  for all  $i = 1, \dots, p$  we have

$$\xi^2 = \frac{8(8 + 12c + c^2)}{n^2}$$

□

This leads to the following theorem,

**Theorem A.13.** *Proof of Theorem 2.2.3*

*Under assumptions (A) and (B),*

$$\frac{n}{\sqrt{8(8 + 12c + c^2)}} (T - \gamma_2 + 1) \xrightarrow{D} N(0, \xi_2^2) \quad (\text{A.40})$$

*with*

$$\begin{aligned} \xi_2^2 = & \frac{1}{(8 + 12c + c^2)a_2^6} \left( \frac{4}{c}a_4^3 - \frac{8}{c}a_4a_2a_6 - 4a_4a_2a_3^2 + \frac{4}{c}a_2^2a_8 \right. \\ & \left. + 4a_6a_2^3 + 8a_2^2a_5a_3 + 4ca_4a_2^4 + 8ca_3^2a_2^3 + c^2a_2^6 \right). \end{aligned} \quad (\text{A.41})$$

*Proof.* The results follows from Theorem A.12

□

**Corollary A.2.** *Proof of Corollary 2.2.2*

*Under  $H_0 : \gamma_2 = 1$  and assumptions (A) and (B),*

$$\frac{n}{\sqrt{8(8 + 12c + c^2)}} \left( \frac{\hat{a}_4}{\hat{a}_2^2} - 1 \right) \xrightarrow{D} N(0, 1) \quad (\text{A.42})$$

*Proof.* This follows from the above theorems.

□

### A.6.1 Asymptotic Properties

Here we consider the  $(n, p)$ -asymptotic properties of our test statistic. In the previous section we showed the test statistic has a Normal distribution as  $(n, p) \rightarrow \infty$ . We now consider the power

function of our test statistic.

$$\begin{aligned}
P(\text{Reject } H_0 | H_1 \text{ true}) &= P\left(\frac{n}{\sqrt{8(8+12c+c^2)}}\left(\hat{a}_2^4 - 1\right) > z_\alpha \mid (A.40)\right) \\
&= P\left(\frac{\frac{n}{\sqrt{8(8+12c+c^2)}}\left(\hat{a}_2^4 - \frac{a_4}{a_2^2}\right)}{\xi_2} > \frac{z_\alpha - \frac{n}{\sqrt{8(8+12c+c^2)}}\left(\frac{a_4}{a_2^2} - 1\right)}{\xi_2}\right) \\
&= \Phi\left(\frac{\frac{n}{\sqrt{8(8+12c+c^2)}}\left(\frac{a_4}{a_2^2} - 1\right) - z_\alpha}{\xi_2}\right)
\end{aligned}$$

To determine the behavior of the power function we first consider the behavior of  $\xi_2$ .

**Lemma A.16.** *Under assumptions (A) and (B), as  $(n, p) \rightarrow \infty$ ,  $\xi_2^2 \rightarrow O(1)$ .*

*Proof.* The proof can be seen by (A.41). □

For large  $(n, p)$ , we can rewrite the power function as follows

$$\Phi\left(\frac{n\left(\frac{a_4}{a_2^2} - 1\right)}{\xi_2 \sqrt{8(8+12c+c^2)}} - \frac{z_\alpha}{\xi_2}\right) \approx \Phi\left(\Xi_2 - \frac{z_\alpha}{\xi_2}\right)$$

**Lemma A.17.** *Under assumptions (A) and (B), as  $(n, p) \rightarrow \infty$ ,  $\Xi_2 \rightarrow O(n) \rightarrow \infty$  and  $\frac{z_\alpha}{\xi_2} \rightarrow O(1)$ .*

*Proof.* The results follow from Lemma A.16 and algebra. □

Hence we can determine the behavior of the power function under general asymptotics.

**Theorem A.14.** *Proof of Theorem 2.2.4*

*Under assumptions (A) and (B), as  $(n, p) \rightarrow \infty$ , the power function,*

$$P(\text{Reject } H_0 | H_1 \text{ true}) \simeq \Phi\left(\Xi_2 - \frac{z_\alpha}{\xi_2}\right) \text{ as } (n, p) \rightarrow \infty$$

*will converge to 1 and hence our test is  $n, p$  consistent.*

*Proof.* The proof follows Lemmas A.16, A.17 and the well known properties of the Standard Normal CDF ( $\Phi$ ). □

### A.6.2 Comparison to Srivastava's Test Statistic for Sphericity

We begin by noting the power of the test defined in Srivastava [99] for large  $(n, p)$ ,

$$\Phi \left( \frac{\frac{n}{2} \left( \frac{a_2}{a_1^2} - 1 \right) - z_\alpha}{\xi_1} \right) \quad (\text{A.43})$$

where

$$\xi_1^2 = \frac{2n (a_4 a_1^2 - 2a_1 a_2 a_3 + a_2^3)}{p a_1^6} + \frac{a_2^2}{a_1^4}$$

We note the asymptotic behavior of the variance under assumptions (A) and (B). As  $(n, p) \rightarrow \infty$ ,  $\xi_1^2 \rightarrow O(1)$ . Hence we can determine the power of (A.43). As  $(n, p) \rightarrow \infty$ , power of the test defined in Srivastava [99] (A.43)  $\rightarrow 1$ .

Thus we can consider the two test to be asymptotically equivalent.

## Appendix B Technical Details for Shrinkage Estimators

In this section, we build on some of the work from the previous Appendix with the intent to apply it to the shrinkage estimation problem from Chapter 3. We start by providing proof to the well known result

**Lemma B.1.**

$$E[\text{tr}(S^2)] = \frac{n+1}{n} \text{tr}\Sigma^2 + \frac{1}{n} (\text{tr}\Sigma)^2 \quad (\text{B.1})$$

*Proof.* Recall equation (A.3)

$$\text{tr}S^2 = \frac{1}{n^2} \left[ \sum_{i=1}^p \lambda_i^2 v_{ii}^2 + 2 \sum_{i<j}^p \lambda_i \lambda_j v_{ij}^2 \right]$$

Taking expectations and using Lemma A.1 we see

$$\begin{aligned} E[\text{tr}S^2] &= \frac{1}{n^2} \left[ \sum_{i=1}^p \lambda_i^2 E[v_{ii}^2] + 2 \sum_{i<j}^p \lambda_i \lambda_j E[v_{ij}^2] \right] \\ &= \frac{n(n+2)}{n^2} \sum_{i=1}^p \lambda_i^2 + 2 \frac{n}{n^2} \sum_{i<j}^p \lambda_i \lambda_j \\ &= \frac{n+2}{n} \text{tr}\Sigma^2 + \frac{1}{n} ((\text{tr}\Sigma)^2 - \text{tr}\Sigma^2) \\ &= \frac{n+1}{n} \text{tr}\Sigma^2 + \frac{1}{n} (\text{tr}\Sigma)^2 \end{aligned}$$

□

**Lemma B.2.**

$$E[\text{tr}(D^2)] = \frac{n+2}{n} \text{tr}\Psi^2 \quad (\text{B.2})$$

*Proof.* It is well known that for each  $s_{ii}$  from a sample of size  $N = n + 1$ ,

$$n \frac{s_{ii}}{\sigma_{ii}} \sim \chi_n^2$$

hence

$$\text{tr}(D) = \sum_{i=1}^p s_{ii} = \frac{1}{n} \sum_{i=1}^p \sigma_{ii} v_{ii}$$

where  $v_{ii}$  is a  $\chi^2$  random variable with  $n$  degrees of freedom. Because  $D$  is a diagonal matrix,

$$\text{tr}(D^2) = \sum_{i=1}^p s_{ii}^2 = \frac{1}{n^2} \sum_{i=1}^p \sigma_{ii}^2 v_{ii}^2.$$

Taking expectations provides the result,

$$\begin{aligned} E[\text{tr}(D^2)] &= \frac{1}{n} \sum_{i=1}^p \sigma_{ii}^2 E[v_{ii}^2] \\ &= \frac{n(n+2)}{n^2} \sigma_{ii}^2 \\ &= \frac{n+2}{n} \text{tr}\Psi^2 \end{aligned}$$

□

**Lemma B.3.**

$$E[\text{tr}(SD)] = E[\text{tr}(DS)] = E[\text{tr}(D^2)] \quad (\text{B.3})$$

*Proof.* This proof follows from some linear algebra.  $\text{tr}(DS) = \text{tr}(SD)$  by Properties of the trace operator. Furthermore,  $D$  is a diagonal matrix of the diagonal elements of  $S$ .  $SD$  will have diagonal elements  $(s_{11}^2, \dots, s_{pp}^2)$  which is the same as  $D^2$ . □

We define

$$\hat{a}_2^* = \frac{n}{n+2} \text{tr}(D^2)/p \quad (\text{B.4})$$

$\hat{a}_2^*$  is clearly unbiased for  $a_2^* = \text{tr}\Psi^2/p$ , see Lemma B.2.

**Lemma B.4.**

$$\text{Var}(\hat{a}_2^*) \simeq \frac{8}{np} a_4^*$$

*Proof.* We first note that  $\frac{n}{n+2} \simeq 1$  for large  $n$ .

$$\begin{aligned}
\text{Var}(\hat{a}_2^*) &\simeq \frac{1}{p} \text{Var}(\text{tr}(D^2)) = \frac{1}{n^4 p^2} \sum_{i=1}^p \sigma_{ii}^4 \text{Var}(v_{ii}^2) \\
&= \frac{1}{n^4 p^2} \sum_{i=1}^p \sigma_{ii}^4 (E[v_{ii}^4] - E[v_{ii}^2]^2) \\
&= \frac{1}{n^4 p^2} \sum_{i=1}^p \sigma_{ii}^4 (n(n+2)(n+4)(n+6) - n^2(n+2)^2) \\
&= \frac{n(n+2)}{n^4 p^2} \sum_{i=1}^p \sigma_{ii}^4 (n^2 + 10n + 24 - n^2 - 2n) \\
&= \frac{8n(n+2)(n+3)}{n^4 p} a_4^* \simeq \frac{8}{np} a_4^*
\end{aligned}$$

□

**Theorem B.1.**  $\hat{a}_2^* \xrightarrow{p} a_2^*$  as  $(n, p) \rightarrow \infty$

*Proof.* An application of Chebyshev's inequality will provide the result. This follows from the fact that  $\hat{a}_2^*$  is unbiased for  $a_2^*$  and by Lemma B.4, the variance will go to zero as  $(n, p) \rightarrow \infty$ . □

## Appendix C Selected Source Code

### C.1 Ravi Varadhan Random Positive Definite Matrix

```
# Generating a random positive-definite matrix with user-specified eigenvalues
# If eigenvalues are not specified, they are generated from a uniform dist.
Posdef <- function (n, ev = runif(n, 0, 10)) {
  Z <- matrix(ncol=n, rnorm(n^2))
  decomp <- qr(Z)
  Q <- qr.Q(decomp)
  R <- qr.R(decomp)
  d <- diag(R)
  ph <- d / abs(d)
  O <- Q %*% diag(ph)
  Z <- t(O) %*% diag(ev) %*% O
  return(Z)
}
```

### C.2 Shrinkage to Diagonal Algorithm

```
# Given a (n+1)xp matrix of observation, we find the optimal shrinkage
# estimator for the covariance matrix of the form:
#          S* = lambda*D + (1-lambda)*S
# where lambda is the optimal shrinkage intensity and D is the matrix
# consisting of the diagonal elements of S
cov.Diag.Shrink <- function(X) {
  N <- dim(X)[1];
  p <- dim(X)[2];
  n <- N - 1;
  S <- cov(X);
  S.2 <- S%*%S;
  D <- diag(S);
  D.2 <- D*D;
```

```

a1.hat <- mean(diag(S));
a2.hat <- (n^2/(n-1)/(n+2))*(sum(diag(S.2)) - sum(diag(S))^2/n)/p;
psi2.hat <- n/(n+2)*mean(D.2);
alpha <- (a2.hat + p*a1.hat^2)/n;
delta <- (n+1)/n*a2.hat + p/n*a1.hat^2 - (n+2)/n*psi2.hat;
Gamma <- -2/n*psi2.hat;
Rho <- (alpha + Gamma)/delta;
Rho*diag(D) + (1-Rho)*S;
}

```

### C.3 Timing Study Code

```

# Here we calculate the respective shrinkage estimators.
# We note that the same from the MVNormal will be the
# same since the seeds for the RNG are reset.
runDiagShrink <- function(n, Sigma) {
  p <- dim(Sigma)[1];
  X <- rmvnorm(n+1, rep(0, p), Sigma);
  S.new <- cov.Diag.Shrink(X);
  0; #Doesn't matter what we return, we are interested in time.
}

runSchaefShrink <- function(n, Sigma) {
  p <- dim(Sigma)[1];
  X <- rmvnorm(n+1, rep(0, p), Sigma);
  S.schaef <- cov.shrink(X, lambda.var=0, verbose=FALSE);
  0;
}

wrapperRuns <- function(n, p, m, seed) {
  # We find a random Positive Definite Matrix Sigma
  ev <- runif(p, 0.5, 10.5);
  Sigma <- Posdef(p, ev);
  # Set the seed to control the samples, calculate the

```



```

# time necessary for our estimator.
if(!(seed==Inf)) set.seed(seed,kind=NULL);
N <- rep(n,m);
ptm <- proc.time();
res1 <- sapply(N, runDiagShrink, Sigma=Sigma);
time.new <- (proc.time()-ptm)[3];
# Reset the seed, calculate the time for Schaefer
# and Strimmers estimator.
if(!(seed==Inf)) set.seed(seed,kind=NULL);
ptm <- proc.time();
res1 <- sapply(N, runSchaefShrink, Sigma=Sigma);
time.sch <- (proc.time()-ptm)[3];
# Return the two run-times for m-runs
c(time.new, time.sch);
}

timingDriverN <- function(p=20, m=1000, seed=Inf) {
if(!(seed==Inf)) set.seed(seed,kind=NULL);
x.axis<-c(10, 30, 50, 100, 150, 200, 300);
y.axis<-sapply(x.axis, wrapperRuns, p=p, m=m, seed=seed);
x.limit <- max(x.axis)+25;
y.limit <- ceiling(max(y.axis)+2);
plot(x.axis, y.axis[1,], type='p', pch=22, bg="red", col="red",
      xlim=c(0,x.limit), ylim=c(0,y.limit), xlab="Sample Size n+1",
      ylab="Time (s)", main="Timing Study for Sample Size" );
lines(x.axis, y.axis[1,], lty="solid", col="red");
points(x.axis, y.axis[2,], pch=23, bg="blue", col="blue");
lines(x.axis, y.axis[2,], lty="dashed", col="blue");
legend(5, (y.limit), legend = c("New Estimator", "Schaefer Estimator"),
      col = c("red", "blue"), pch=c(22, 23),
      pt.cex=1, pt.bg=c("red","blue"), lty = c(1, 2))
}

```

# Bibliography

- [1] S. E. Ahmed. Large-sample estimation strategies for eigenvalues of a Wishart matrix. *Metrika*, 47(1):35–45, 1998.
- [2] U. Alon, N. Barkai, D. A. Notterman, K. Gish, S. Ybarra, D. Mack, and A. J. Levine. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proceedings of the National Academy of Sciences of the United States of America*, 96(12):6745–6750, June 1999.
- [3] T. W. Anderson. *An introduction to multivariate statistical analysis*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, third edition, 2003.
- [4] L. V. Arharov. Limit theorems for the characteristic roots of a sample covariance matrix. *Dokl. Akad. Nauk SSSR*, 199:994–997, 1971.
- [5] Z. D. Bai. Convergence rate of expected spectral distributions of large random matrices. II. Sample covariance matrices. *Ann. Probab.*, 21(2):649–672, 1993.
- [6] Z. D. Bai, Jack W. Silverstein, and Y. Q. Yin. A note on the largest eigenvalue of a large-dimensional sample covariance matrix. *J. Multivariate Anal.*, 26(2):166–168, 1988.
- [7] Z. D. Bai and Y. Q. Yin. Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.*, 21(3):1275–1294, 1993.
- [8] Zhi Dong Bai, Paruchuri R. Krishnaiah, and Lin Cheng Zhao. On rates of convergence of efficient detection criteria in signal processing with white noise. *IEEE Trans. Inform. Theory*, 35(2):380–388, 1989.
- [9] John Barnard, Robert McCulloch, and Xiao-Li Meng. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statist. Sinica*, 10(4):1281–1311, 2000.
- [10] Peter J. Bickel and Elizaveta Levina. Regularized estimation of large covariance matrices. *Ann. Statist.*, 36(1):199–227, 2008.
- [11] Melanie Birke and Holger Dette. A note on testing the covariance matrix for large dimension. *Statist. Probab. Lett.*, 74(3):281–289, 2005.
- [12] Guangzhi Cao and Charles Bouman. Covariance estimation for high dimensional data vectors using the sparse matrix transform. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems 21*, pages 225–232. 2009.
- [13] E. M. Carter and M. S. Srivastava. Monotonicity of the power functions of modified likelihood ratio criterion for the homogeneity of variances and of the sphericity test. *J. Multivariate Anal.*, 7(1):229–233, 1977.

- [14] George Casella and Roger L. Berger. *Statistical inference*. The Wadsworth & Brooks/Cole Statistics/Probability Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1990.
- [15] Colin J. Champion. Empirical Bayesian estimation of normal variances and covariances. *J. Multivariate Anal.*, 87(1):60–79, 2003.
- [16] Sanjay Chaudhuri, Mathias Drton, and Thomas S. Richardson. Estimation of a covariance matrix with zeros. *Biometrika*, 94(1):199–216, 2007.
- [17] Yilun Chen, Ami Wiesel, and Alfred O. Hero. Shrinkage estimation of high dimensional covariance matrices. *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, 0:2937–2940, 2009.
- [18] Zhen Chen and David B. Dunson. Random effects selection in linear mixed models. *Biometrics*, 59(4):762–769, 2003.
- [19] R. R. Corbeil and S. R. Searle. Restricted maximum likelihood (REML) estimation of variance components in the mixed model. *Technometrics*, 18(1):31–38, 1976.
- [20] Michael J. Daniels. A class of shrinkage priors for the dependence structure in longitudinal data. *J. Statist. Plann. Inference*, 127(1-2):119–130, 2005.
- [21] Michael J. Daniels. Bayesian modeling of several covariance matrices and some results on propriety of the posterior for linear regression with correlated and/or heterogeneous errors. *J. Multivariate Anal.*, 97(5):1185–1207, 2006.
- [22] Michael J. Daniels and Robert E. Kass. Shrinkage estimators for covariance matrices. *Biometrics*, 57(4):1173–1184, 2001.
- [23] Michael J. Daniels and Mohsen Pourahmadi. Bayesian analysis of covariance matrices and dynamic models for longitudinal data. *Biometrika*, 89(3):553–566, 2002.
- [24] M. Dettling and P. Bühlmann. Boosting for tumor classification with gene expression data. *Bioinformatics*, 19(9):1061–1069, June 2003.
- [25] D. K. Dey and C. Srinivasan. Estimation of a covariance matrix under Stein’s loss. *The Annals of Statistics*, 13(4):1581–1591, 1985.
- [26] Adrian Dobra, Chris Hans, Beatrix Jones, Joseph R. Nevins, Guang Yao, and Mike West. Sparse graphical models for exploring gene expression data. *J. Multivariate Anal.*, 90(1):196–212, 2004.
- [27] Edward J. Dudewicz and Satya N. Mishra. *Modern mathematical statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1988.
- [28] Sandrine Dudoit, Jane Fridlyand, and Terence P. Speed. Comparison of discrimination methods for the classification of tumors using gene expression data. *J. Amer. Statist. Assoc.*, 97(457):77–87, 2002.
- [29] Richard Durrett. *Probability: theory and examples*. Duxbury Press, Belmont, CA, second edition, 1996.
- [30] Morris L. Eaton and Ingram Olkin. Best equivariant estimators of a Cholesky decomposition. *Ann. Statist.*, 15(4):1639–1650, 1987.

- [31] Bradley Efron. Biased versus unbiased estimation. *Advances in Math.*, 16:259–277, 1975.
- [32] Bradley Efron. Maximum likelihood and decision theory. *Ann. Statist.*, 10(2):340–356, 1982.
- [33] Bradley Efron and Carl Morris. Data analysis using stein’s estimator and its generalizations. *Journal of the American Statistical Association*, 70(350):311–319, 1975.
- [34] Bradley Efron and Carl Morris. Multivariate empirical Bayes and estimation of covariance matrices. *Ann. Statist.*, 4(1):22–32, 1976.
- [35] Bradley Efron and Carl Morris. Stein’s paradox in statistics. *Scientific American*, 236(5):119–127, 1977.
- [36] Michael B. Eisen, Paul T. Spellman, Patrick O. Brown, and David Botstein. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America*, 95(25):14863–14868, December 1998.
- [37] Jianqing Fan, Yingying Fan, and Jinchi Lv. High dimensional covariance matrix estimation using a factor model. *J. Econometrics*, 147(1):186–197, 2008.
- [38] Jianqing Fan and Yichao Wu. Semiparametric estimation of covariance matrixes for longitudinal data. *J. Amer. Statist. Assoc.*, 103(484):1520–1533, 2008.
- [39] William H. Fellner. Sparse matrices, and the estimation of variance components by likelihood methods. *Comm. Statist. B—Simulation Comput.*, 16(2):439–463, 1987.
- [40] Chris Fraley and Patrick J. Burns. Large-scale estimation of variance and covariance components. *SIAM J. Sci. Comput.*, 16(1):192–209, 1995.
- [41] Peter A. Frost and James E. Savarino. An empirical bayes approach to portfolio selection. *J. Financial and Quantitive Anal.*, 21:293–305, 1986.
- [42] Reinhard Furrer. Covariance estimation under spatial dependence. *J. Multivariate Anal.*, 94(2):366–381, 2005.
- [43] Reinhard Furrer and Thomas Bengtsson. Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *J. Multivariate Anal.*, 98(2):227–255, 2007.
- [44] Malay Ghosh, Victor Mergel, and Gauri Sankar Datta. Estimation, prediction and the Stein phenomenon under divergence loss. *J. Multivariate Anal.*, 99(9):1941–1961, 2008.
- [45] V. L. Girko. Random matrices. In *Handbook of algebra, Vol. 1*, pages 27–78. North-Holland, Amsterdam, 1996.
- [46] V. L. Girko. Strong law for the eigenvalues and eigenvectors of empirical covariance matrices. *Random Oper. Stochastic Equations*, 4(2):179–204, 1996.
- [47] V. L. Girko. Strong law for the eigenvalues of empirical covariance matrices. In *Multidimensional statistical analysis and theory of random matrices (Bowling Green, OH, 1996)*, pages 61–92. VSP, Utrecht, 1996.
- [48] Vyacheslav L. Girko. *Statistical analysis of observations of increasing dimension*, volume 28 of *Theory and Decision Library. Series B: Mathematical and Statistical Methods*. Kluwer Academic Publishers, Dordrecht, 1995. Translated from the Russian.

- [49] T. Golub, D. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. Mesirov, H. Coller, M. Loh, J. Downing, M. Caligiuri, C. Bloomfield, and E. Lander. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science*, 286(5439):531–537, October 1999.
- [50] Franklin A. Graybill. *Theory and application of the linear model*. Duxbury Press, North Scituate, Mass., 1976.
- [51] L. R. Haff. Estimation of the inverse covariance matrix: random mixtures of the inverse Wishart matrix and the identity. *Ann. Statist.*, 7(6):1264–1276, 1979.
- [52] L. R. Haff. An identity for the Wishart distribution with applications. *J. Multivariate Anal.*, 9(4):531–544, 1979.
- [53] L. R. Haff. Empirical Bayes estimation of the multivariate normal covariance matrix. *Ann. Statist.*, 8(3):586–597, 1980.
- [54] L. R. Haff. Corrections to: “Estimation of the inverse covariance matrix: random mixtures of the inverse Wishart matrix and the identity” [Ann. Statist. **7** (1979), no. 6, 1264–1276; MR 81b:62011]. *Ann. Statist.*, 9(5):1132, 1981.
- [55] L. R. Haff. Erratum: “An identity for the Wishart distribution with applications” [J. Multivariate Anal. **9** (1979), no. 4, 531–544; MR 81c:62055]. *J. Multivariate Anal.*, 11(4):608–609, 1981.
- [56] L. R. Haff. The variational form of certain Bayes estimators. *Ann. Statist.*, 19(3):1163–1190, 1991.
- [57] W. J. Hemmerle and H. O. Hartley. Computing maximum likelihood estimates for the mixed A.O.V. model using the  $W$  transformation. *Technometrics*, 15:819–831, 1973.
- [58] P. L. Hsu. The limiting distribution of functions of sample means and application to testing hypotheses. In *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability, 1945, 1946*, pages 359–402, Berkeley and Los Angeles, 1949. University of California Press.
- [59] Jianhua Z. Huang, Linxu Liu, and Naiping Liu. Estimation of large covariance matrices of longitudinal data with basis function approximations. *J. Comput. Graph. Statist.*, 16(1):189–209, 2007.
- [60] Jianhua Z. Huang, Naiping Liu, Mohsen Pourahmadi, and Linxu Liu. Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika*, 93(1):85–98, 2006.
- [61] W. James and Charles Stein. Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I*, pages 361–379. Univ. California Press, Berkeley, Calif., 1961.
- [62] S. John. Some optimal multivariate tests. *Biometrika*, 58:123–127, 1971.
- [63] Richard A. Johnson and Dean W. Wichern. *Applied multivariate statistical analysis*. Pearson Prentice Hall, Upper Saddle River, NJ, sixth edition, 2007.
- [64] Philippe Jorion. Bayes-stein estimation for portfolio analysis. *J. Financial and Quantitative Anal.*, 21(3):279–292, 1986.
- [65] H. J. Kim and S. J. Press. Bayesian hypothesis testing of equality of normal covariance matrices. In *Multivariate analysis and its applications (Hong Kong, 1992)*, volume 24 of *IMS Lecture Notes Monogr. Ser.*, pages 323–330. Inst. Math. Statist., Hayward, CA, 1994.

- [66] Tatsuya Kubokawa. Double shrinkage estimation of common coefficients in two regression equations with heteroscedasticity. *J. Multivariate Anal.*, 67(2):169–189, 1998.
- [67] Tatsuya Kubokawa. The Stein phenomenon in simultaneous estimation: a review. In *Applied statistical science, III*, pages 143–173. Nova Sci. Publ., Commack, NY, 1998.
- [68] Tatsuya Kubokawa and Muni S. Srivastava. Estimation of the precision matrix of a singular Wishart distribution and its application in high-dimensional data. *J. Multivariate Anal.*, 99(9):1906–1928, 2008.
- [69] Olivier Ledoit and Michael Wolf. Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.*, 30(4):1081–1102, 2002.
- [70] Olivier Ledoit and Michael Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10(5):603–621, December 2003.
- [71] Olivier Ledoit and Michael Wolf. Honey, i shrunk the sample covariance matrix. *The J. of Portfolio Management*, 31(1):110–119, 2004.
- [72] Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivariate Anal.*, 88(2):365–411, 2004.
- [73] E. L. Lehmann. *Elements of large-sample theory*. Springer Texts in Statistics. Springer-Verlag, New York, 1999.
- [74] Pui Lam Leung and Robb J. Muirhead. Estimation of parameter matrices and eigenvalues in MANOVA and canonical correlation analysis. *Ann. Statist.*, 15(4):1651–1666, 1987.
- [75] Hisao Nagao. On some test criteria for covariance matrix. *Ann. Statist.*, 1:700–709, 1973.
- [76] C. R. Narayanaswamy and D. Raghavarao. Principal component analysis of large dispersion matrices. *J. Roy. Statist. Soc. Ser. C*, 40(2):309–316, 1991.
- [77] Mohsen Pourahmadi. Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika*, 86(3):677–690, 1999.
- [78] Mohsen Pourahmadi. Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika*, 87(2):425–435, 2000.
- [79] Mohsen Pourahmadi. Cholesky decompositions and estimation of a covariance matrix: orthogonality of variance-correlation parameters. *Biometrika*, 94(4):1006–1013, 2007.
- [80] Mohsen Pourahmadi, Michael J. Daniels, and Trevor Park. Simultaneous modelling of the Cholesky decomposition of several covariance matrices. *J. Multivariate Anal.*, 98(3):568–587, 2007.
- [81] C. Radhakrishna Rao. *Linear statistical inference and its applications*. John Wiley & Sons, New York-London-Sydney, second edition, 1973. Wiley Series in Probability and Mathematical Statistics.
- [82] Alberto Roverato. Cholesky decomposition of a hyper inverse Wishart matrix. *Biometrika*, 87(1):99–112, 2000.
- [83] Alessio Sancetta. Sample covariance shrinkage for high dimensional dependent data. *J. Multivariate Anal.*, 99(5):949–967, 2008.

- [84] Hewa Saranadasa. Asymptotic expansion of the misclassification probabilities of  $D$ - and  $A$ -criteria for discrimination from two high-dimensional populations using the theory of large-dimensional random matrices. *J. Multivariate Anal.*, 46(1):154–174, 1993.
- [85] Juliane Schäfer and Korbinian Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat. Appl. Genet. Mol. Biol.*, 4:Art. 32, 28 pp. (electronic), 2005.
- [86] Guthke R. Toepfer S. Reischer H. Duerrschmid K. Schmidt-Heck, W. and K. Bayer. Reverse engineering of the stress response during expression of a recombinant protein. In *Proceedings of the EUNITE symposium*, pages 407–412, June 2004. EUNITE 2004, June 10-12, 2004, Aachen, Germany, <http://www.eunite.org>.
- [87] James R. Schott. Testing for independence in high dimensions. Technical Report TR-2004-01, University of Central Florida, 2004. "[http://statistics.cos.ucf.edu/files/technical\\_reports/TR-2004-01.pdf](http://statistics.cos.ucf.edu/files/technical_reports/TR-2004-01.pdf)".
- [88] James R. Schott. Testing for complete independence in high dimensions. *Biometrika*, 92(4):951–956, 2005.
- [89] James R. Schott. A high-dimensional test for the equality of the smallest eigenvalues of a covariance matrix. *J. Multivariate Anal.*, 97(4):827–843, 2006.
- [90] James R. Schott. A test for the equality of covariance matrices when the dimension is large relative to the sample sizes. *Comput. Statist. Data Anal.*, 51(12):6535–6542, 2007.
- [91] V. Serdobolskii. *Multivariate statistical analysis*, volume 41 of *Theory and Decision Library. Series B: Mathematical and Statistical Methods*. Kluwer Academic Publishers, Dordrecht, 2000. A high-dimensional approach.
- [92] V. I. Serdobol'skiĭ. Spectral properties of sample covariance matrices. *Teor. Veroyatnost. i Primenen.*, 40(4):910–919, 1995.
- [93] V. I. Serdobol'skiĭ. Theory of essentially multivariate statistical analysis. *Uspekhi Mat. Nauk*, 54(2(326)):85–112, 1999.
- [94] A. N. Shiryaev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1996. Translated from the first (1980) Russian edition by R. P. Boas.
- [95] Jack W. Silverstein. Eigenvalues and eigenvectors of large-dimensional sample covariance matrices. In *Random matrices and their applications*, pages 153–159. Amer. Math. Soc., Providence, RI, 1986.
- [96] Michael Smith and Robert Kohn. Parsimonious covariance matrix estimation for longitudinal data. *J. Amer. Statist. Assoc.*, 97(460):1141–1153, 2002.
- [97] M. S. Srivastava. Multivariate theory for analyzing high dimensional data. *J. Japan Statist. Soc.*, 37(1):53–86, 2007.
- [98] M. S. Srivastava and M. Bilodeau. Stein estimation under elliptical distributions. *J. Multivariate Anal.*, 28(2):247–259, 1989.
- [99] Muni S. Srivastava. Some tests concerning the covariance matrix in high dimensional data. *J. Japan Statist. Soc.*, 35(2):251–272, 2005.

- [100] Muni S. Srivastava. Some tests criteria for the covariance matrix with fewer observations than the dimension. *Acta Comment. Univ. Tartu. Math.*, (10):77–93, 2006.
- [101] Charles Stein. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. I*, pages 197–206, Berkeley and Los Angeles, 1956. University of California Press.
- [102] Charles Stein. Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.*, 16:155–160, 1964.
- [103] Charles Stein. Estimation of a covariance matrix. In *Rietz Lecture, 39th Annual Meeting IMS, Atlanta, GA*, 1975.
- [104] Yan Sun, Wenyang Zhang, and Howell Tong. Estimation of the covariance matrix of random effects in longitudinal studies. *Ann. Statist.*, 35(6):2795–2814, 2007.
- [105] Hiroyuki Toh and Katsuhisa Horimoto. Inference of a genetic network by a combined approach of cluster analysis and graphical gaussian modeling. *Bioinformatics*, 18(2):287–297, 2002.
- [106] Hisayuki Tsukuma and Yoshihiko Konno. On improved estimation of normal precision matrix and discriminant coefficients. *J. Multivariate Anal.*, 97(7):1477–1500, 2006.
- [107] A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.
- [108] David I. Warton. Penalized normal likelihood and ridge regularization of correlation and covariance matrices. *J. Amer. Statist. Assoc.*, 103(481):340–349, 2008.
- [109] K. Werner, M. Jansson, and P. Stoica. On estimation of covariance matrices with kronecker product structure. *Signal Processing, IEEE Transactions on*, 56(2):478–491, Feb. 2008.
- [110] Wei Biao Wu and Mohsen Pourahmadi. Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika*, 90(4):831–844, 2003.
- [111] Ruo-yong Yang and James O. Berger. Estimation of a covariance matrix using the reference prior. *Ann. Statist.*, 22(3):1195–1211, 1994.
- [112] Y. Q. Yin and P. R. Krishnaiah. A limit theorem for the eigenvalues of product of two random matrices. *J. Multivariate Anal.*, 13(4):489–507, 1983.