Random Matrix Theory with Applications in Statistics and Finance

Nadia Abdel Samie Basyouni Kotb Saad

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the degree of
Doctorate of Philosophy in Mathematics [1]

Department of Mathematics and Statistics
Faculty of Science
University of Ottawa

[1] The Ph.D. program is a joint program with Carleton University, administered by the Ottawa-Carleton Institute of Mathematics and Statistics

Abstract

This thesis investigates a technique to estimate the risk of the mean-variance (MV) portfolio optimization problem. We call this technique the *Scaling* technique. It provides a better estimator of the risk of the MV optimal portfolio. We obtain this result for a general estimator of the covariance matrix of the returns which includes the correlated sampling case as well as the independent sampling case and the exponentially weighted moving average case. This gave rise to the paper, [CMcS].

Our result concerning the Scaling technique relies on the moments of the inverse of compound Wishart matrices. This is an open problem in the theory of random matrices. We actually tackle a much more general setup, where we consider any random matrix provided that its distribution has an appropriate invariance property (orthogonal or unitary) under an appropriate action (by conjugation, or by a left-right action). Our approach is based on Weingarten calculus. As an interesting byproduct of our study - and as a preliminary to the solution of our problem of computing the moments of the inverse of a compound Wishart random matrix, we obtain explicit moment formulas for the pseudo-inverse of Ginibre random matrices. These results are also given in the paper, [CMS].

Using the moments of the inverse of compound Wishart matrices, we obtain asymptotically unbiased estimators of the risk and the weights of the MV portfolio. Finally, we have some numerical results which are part of our future work.

Résumé

L'objet de cette thèse est d'étudier une technique pour estimer le risque ou la frontière de Markowitz dans le problème d'optimisation de portefeuille. Nous l'appelons "technique de scaling". Elle fournit un estimateur amélioré du risque du portefeuille VM-optimal. Nous obtenons ce résultat pour un estimateur général de la matrice de covariance des gains. L'estimateur de la covariance décrit le cas des échantillonnages coréllés ainsi que l'échantillonnage indépendent. Cette amélioration donne lieu à une nouvelle approche de l'estimation des risques des matrices de covariance financières qui impliquent des gains avec moyenne mobile à poids exponentiels. Ces résultats ont donné lieu au papier [CMcS]. Notre résultat sur la technique de "scaling" dépend des moments de l'inverse de matrices de Wishart composées. Nous nous attaquons à un cadre bien plus général, dans lequel nous considérons toute matrice aléatoire dont on suppose que sa distribution a une propriété d'invariance adéquate (orthogonale ou unitaire) sous une action appropriée (par conjugaison, ou bien sous une action à droite et à gauche). Notre approche se fonde sur le calcul de Weingarten. Une conséquence intéressante de notre étude - et une solution préliminaire à notre problème de calculer les moments de l'inverse d'une matrice de Wishart composée, est une formule explicite de moments pour les pseudo-inverses des matrices de Ginibre. Ces résultats ont donné lieu au papier [CMS]. En utilisant les moments de l'inverse d'une matrice de Wishart composée, nous obtenons un estimateur asymptotiquement non-biaisé du risque d'une portefeuille de Markowitz, et un estimateur non-biaisé de ses poids.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisors Dr. Benoît Collins and Dr. David McDonald for inspiring me all along the way throughout my PhD study and research, for their patience, motivation, enthusiasm and immense knowledge. Their provocative questions, thoughtful discussions and careful comments have greatly influenced me and are reflected throughout this work. This thesis could not be possible without their support and encouragement.

I owe my deepest gratitude to Dr. Sho Matsumoto, for whom a proper "thank you" would be larger than this thesis. I am grateful to him for his positivity throughout my studies.

It is an honor for me to thank Dr. Mayer Alvo, who has made available his support in a numerous of ways. I would like to thank him for all the guidance, encouragement, inspiration and help.

I have been most fortunate to have been able to discuss my work with Dr. Rafal Kulik, Dr. Mohamedou Ould Haye and Dr. Ion Nechita who always provide me with useful suggestions.

Special thanks go to my friends and colleagues for providing me with their experience, support and making my time more enjoyable at the university.

Finally, there has been no more constant a source of encouragement than my husband, my parents and my daughters. They have participated in this endeavor by supporting my efforts and making significant sacrifices.

Dedication

To my husband, my parents and my daughters

Aya, Habiba and Sondos.

For their unconditional love and never ending support.

Contents

List of Figures

Chapter 1

Introduction

The optimal portfolio selection problem is one of the most important topics in the fields of investment and financial research. Modern portfolio theory (MPT) dates from Markowitz's pioneering article [Mark] in 1952 and subsequent book [Mark2]. Markowitz constructed the mean-variance (MV) portfolio model, namely holding the variance constant while maximizing the expected return, or holding the expected return constant while minimizing the variance. These two principles led to the formulation of an efficient frontier from which the investor could choose his or her preferred portfolio (called optimal portfolio), depending on individual risk return preferences.

There are two criticisms of the use of the variance as a measure of risk. The first criticism is that since the variance measures the dispersion of an asset's return around its expected return, it treats both the returns above and below the expected return identically. However, the investors don't view return outcomes above the expected return in the same way as they view returns below the expected return. Markowitz recognized this limitation and suggested a measure of downside risk (the risk of realizing an outcome below the expected return) called the *semi-variance*. The semi-variance is similar to the variance expect that in the calculation no consideration is given to returns above the expected return. However, because of the computational problems with using the semi-variance, he used the variance in developing portfolio theory. On the other hand, as suggested by Davison [D], modern finance assumes that all information about the future of a stock is properly weighed and considered by the market place, making it improbable that the market would be wrong about the central tendency of the return distribution. Therefore, return distributions must display a certain degree of symmetry around their mean. If complete symmetry around a known mean applies, then minimizing variance will minimize semi-variance.

The second criticism is that considering just the mean return and variance of return of a portfolio is a simplification relative to including additional moments that might more completely describe the distribution of returns of the portfolio. Many researchers offered alternative portfolio theories that included more moments such as skewness or were accurate for more realistic descriptions of the distribution of return (see Fama [F], Elton and Gruber [EG1]).

Nevertheless, Elton and Gruber [EG3] show that the mean variance theory has remained the cornerstone of modern portfolio theory despite these objections due to two reasons. First, the mean variance theory itself places large data requirements on the investor, and there is no evidence that adding additional moments improves the desirability of the portfolio selected. Second, the implications of mean variance portfolio theory are well developed, widely known, and have great intuitive appeal. Professionals who have never run an optimizer have learned that correlations as well as means and variances are necessary to understand the impact of adding a security to a portfolio.

Many researchers have contributed to the development of MPT (see [EG2], [Lu], [RU]) and several portfolio models, extending the MV model, have been proposed such as the mean-absolute deviation model ([K], [KSY]). The MV model was intended to be practical and implementable. But due to the error in estimating the parameters (including the mean return and the covariance matrix of returns) of the MV model, the applicability of the MV model is limited. In this thesis, we consider a general estimator of the covariance matrix (including the correlated sampling case) of returns and using techniques from random matrix theory (RMT) we study the effect of the noise induced by estimating the covariance matrix on the risk and the compositions of the MV portfolio model.

1.1 Problem Definition

The problem studied in this thesis has two parts. The first part describes a mathematical finance problem as we will see in Section 1.1.1 . While Section 1.1.2 illustrates the second part of the problem which poses an interesting question in random matrix theory (RMT).

1.1.1 From the Perspective of Mathematical Finance

The concept of financial risk attempts to quantify the uncertainty of the outcome of an investment and hence the magnitude of possible losses. Portfolio optimization aims to give a recipe for the composition of portfolios such that the overall risk is minimized for a given reward, or, alternatively, reward is maximized for a given risk.

The classical portfolio optimization problem formulated first by Markowitz ([Mark], [Mark2]) relies on the variance to measure the risk and on expected return to measure the reward. Since the return on a portfolio is a linear combination of the returns of the assets forming the portfolio with weights given by the proportion of wealth invested in the assets, the portfolio variance can be expressed as a quadratic form of these weights with the volatilities and correlations as coefficients.

For any practical use of the theory, it is necessary to have reliable estimates of the volatilities and correlations, which, in most cases, are obtained from historical return data. Actually, volatility and correlation estimates extracted from historical data have also become standard tools for several other risk management practices widely utilized in the financial industry. However, if one estimates an $n \times n$ correlation (or covariance) matrix from n time series of length T each, with T bounded for evident practical reasons, one inevitably introduces estimation error, which for large n can become so overwhelming that the whole applicability of the theory becomes questionable. This difficulty has been well known by economists for a long time (see e.g. [EG2]).

In ([GBP], [LCBP1]), the problem has been approached from the point of view of random matrix theory (RMT). These studies have shown that empirical correlation matrices deduced from financial return series contain a high amount of noise. Apart from a few large eigenvalues and the corresponding eigenvectors, their structure can essentially be regarded as random. In [LCBP1], it is reported that about 94% of the spectrum of correlation matrices determined from return series on the S&P 500 stocks can be fitted by that of a random matrix. The authors conclude that "Markowitz's portfolio optimization scheme based on a purely historical determination of the correlation matrix is inadequate". Two subsequent studies ([LCBP2], [RGPS]) found that the risk level of optimized portfolios could be improved if prior to optimization one filtered out the lower part of the eigenvalue spectrum of the empirical correlation matrix, thereby removing the noise (at least partially).

For the empirical covariance matrices with independent sampling, Pafka et al. [PK] and El Karoui [El-K], were able to compute the asymptotic effect of the noise resulting from estimating the covariance matrix on the optimal portfolio's risk. Covariance matrices with correlated sampling play a fundamental role in many fields. In finance, the exponentially weighted moving average (EWMA) is an example of such covariance matrices. The EWMA technique is introduced by Bollerslev [Bol] and it describes the current market conditions more accurately by giving more weight to the recent observations than the past ones. We start our work with the following question: "What is the asymptotic effect on the risk of the optimal portfolio of the noise resulting from estimating a covariance matrix with correlated sampling?"

1.1.2 From the Perspective of Random Matrices

The mathematical financial problem in Section 1.1.1 poses an interesting open question in RMT concerning the moments of the inverse of compound Wishart matrices. This is one of the important contributions of this Ph.D. thesis.

The Wishart distribution is the multivariate extension of the gamma distribution, although most statisticians use the Wishart distribution in the special case of integer degrees of freedom, in which case it simplifies to a multivariate generalization of the chi-square distribution. The Wishart distribution is used to model random covariance matrices. One generalization of Wishart matrices is compound Wishart matrices which are studied, for example, in [Sp, HP]. Compound Wishart matrices appear in many contexts such as spiked random matrices.

The study of the eigenvalues of Wishart matrices is quite well developed but a systematic study of the joint moments of their entries is more recent. The theoretical study of the inverse of Wishart matrices is also very important, however, the study of their local moments is much more recent, and was actually still open in the case of the inverse of the compound Wishart matrix.

1.2 Thesis Contribution

In this thesis, we focus on the noise induced by estimating the covariance matrix of returns and its effect on measuring the risk of the optimal portfolio. To cover the case of correlated sampling covariance matrices such as the EWMA, we consider a general estimator of the covariance matrix which describes the correlated sampling case as well as the independent one. To concentrate only on the noise resulting from estimating the covariance matrix, we consider a simplified version of the MV portfolio model which depends only on the empirical covariance matrix not on the mean return. The optimal portfolio can easily be found by introducing a Lagrange multiplier which leads to a linear problem where the empirical covariance matrix has to be inverted.

We define the impact of this noise, resulting from using the empirical covariance matrix, on the measure of the optimal portfolio's risk as the ratio between the *Predicted* risk, the measure of the portfolio's risk depending on the empirical covariance matrix, and the *True* risk, the measure of the portfolio's risk depending on the "True" covariance matrix. In practice, we can only obtain the *Predicted* risk while the *True* risk is unknown. To study the asymptotic behavior of this ratio, we need to obtain formulas for the joint moments of the entries of the inverse of compound Wishart matrices (which describes the distribution of the inverse of empirical covariance matrices).

This is an open problem in random matrix theory and to solve it, we consider random matrices that have invariance properties under the action of unitary groups (either a left-right invariance, or a conjugacy invariance). Using the results of Collins in [C]; Collins and Matsumoto in [CM], we represent the moments of these unitary (or orthogonal for the real case) invariant matrices in terms of functions of eigenvalues. Our main tools are the unitary (and orthogonal) Weingarten functions which are studied in many works (see [We], [C], [CS], [MN]). In our work, we need to introduce a modification of the Weingarten function, namely, a 'double' Weingarten function with two dimension parameters instead of one. As an application to statistics, we obtain new formulas for the pseudo inverse of Gaussian matrices and for the inverse of compound Wishart matrices by using the result of Matsumoto in [M2] concerning the global moments of the Wishart matrices and their inverses. These results are incorporated in the paper [CMS].

We also are able to derive an interesting property of inverse compound Wishart matrices. Using this property as well as our formula concerning the moments of the inverse of compound Wishart matrices, we derive the asymptotic effect of the noise induced by estimating the covariance matrix on computing the risk of the optimal portfolio. This in turn enables us scale the *Predicted* risk by a bias factor and get an asymptotically unbiased estimator of the risk of the optimal portfolio not only for the case of independent observations but also in the case of correlated observations. We call this technique the *Scaling* technique. In the case of independent observations, our results coincide with the results of Pafka et al. [PK] and El Karoui [El-K]. As an application, we obtain a new approach to estimating the risk of financial covariance matrices involving stock returns by using the exponentially weighted moving average. These results appear in the paper [CMcS].

To study the usefulness of the *Scaling* technique, we simulate the *Predicted* risk of the optimal portfolio before and after applying the *Scaling* technique for different estimators of the covariance matrix. Since the empirical data sets contain several sources of error (caused by nonstationarity, market microstructure etc.) in addition to the noise due to the finite length of the time series, we based our simulations on data generated from some toy models. This procedure offers a major advantage in that the covariance matrix and consequently the True risk is known. The simulations show a remarkable improvement in estimating the risk of the optimal portfolio after using the Scaling technique in the case of independent observations as well as in the case of the EWMA covariances. Simulations also show that the Scaling technique provides better estimation of the risk of the optimal portfolio than the Filtering technique which depends on eliminating the noisy eigenvalues of the empirical covariance (correlation) matrix. We also use our result concerning the moments of the inverse of the compound Wishart matrices to study the optimal weights and illustrate their asymptotically unbiased estimator.

1.3 Thesis Organization

The thesis is organized as follows: Chapter 2 introduces modern portfolio theory (MPT) and illustrates the MV portfolio model. For a simplified version (depending only on the covariance matrix of returns as a parameter) of the MV model, we demonstrate the risk of the optimal portfolio in terms of the entries (or, the eigenvalues) of the inverse of the covariance matrix of returns.

Chapter 3 reviews some basic concepts of random matrix theory (RMT).

Chapter 4 presents our results concerning the moments of the unitary or orthogonal invariant random matrices and as an application to statistics, we formulate the moments of the pseudo inverse of the Gaussian matrices and the moments of the inverse of compound Wishart matrices.

In Chapter 5, we use the results of Chapter 4 to study the effect of the noise induced by estimating the covariance matrix of returns in the case of independent observations as well as the correlated sampling situation. We also simulate our results and compare them with the results of another technique called the *Filtering* technique. Finally, we study the optimal weights and get a result concerning their asymptotically unbiased estimator.

Chapter 6 presents some numerical observations and topics which are of great interest for our future work.

Finally, we present the R Language code used in our simulations.

Chapter 2

Modern Portfolio Theory (MPT) and Risk Estimation

This chapter provides the financial background for this thesis. Modern portfolio theory (MPT) is the mathematical formulation of the concept of diversification in investing. The main idea is to allocate investments between different assets by considering the trade-off between risk and expected return. The theory attempts to maximize portfolio expected return for a given amount of portfolio risk, or alternatively minimize risk for a certain level of expected portfolio return, by carefully choosing the weights of various assets. MPT was developed in the 1950s through the early 1970s and was considered an important advance in the mathematical modeling of risk. In the following, we are going to discuss the parameters and concepts of the theory. Then, we illustrate the Markowitz mean-variance (MV) model and compute the risk of the model.

2.1 Portfolio's Expected Return

A *risky* asset is one for which the return that will be realized in the future is uncertain. There are assets for which the return that will be realized in the future is known with certain today. Such assets are referred to as *risk-free* or *riskless* assets.

The actual return R_P on a portfolio P of n assets over some specific time period is calculated using the following formula:

$$R_P = \sum_{i=1}^{n} \omega_i R_i, \tag{2.1.1}$$

where ω_i $(i = 1, 2, \ldots, n)$ is the amount of capital invested in the asset i at the beginning of the period, and $\{R_i\}$ are the returns of the individual assets. We denote the expected return $E(R_i)$ by μ_{R_i}. The portfolio return R_P is sometimes called the ex post return. According to (2.1.1), R_P is equal to the sum over all individual assets' weights in the portfolio times their respective return. While the expected portfolio return $E(R_P) = \mu_P$ is the weighted average of the expected return of each asset in the portfolio. That is,

$$\mu_P = \omega_1 \mu_{R_1} + \omega_2 \mu_{R_2} + \cdots + \omega_n \mu_{R_n}.$$

Note that, $E(R_P)$ is sometimes called the expected portfolio return over some specific time period.

2.2 Portfolio's Risk

A portfolio's risk is the possibility that an investment portfolio may not achieve its objectives. There are a number of factors that contribute a portfolio's risk, and while you are able to minimize them, you will never be able to fully eliminate them.

Systematic risk is one factor which contributes to portfolio's risk. It includes the risk associated with general economic cycle, interest rates, war and political instability. It is clear that this risk factor is unpredictable.

Unsystematic risk is risk that one can control or minimize. It relates to the risk associated with owning the shares of a specific company in a portfolio.

In [Mark], Harry Markowitz quantified the concept of risk using the well-known statistical measures of variance and covariance. Markowitz defined the portfolio's risk as the square root of the variance of the portfolio's return. In the case of an asset's return, the variance is a measure of the dispersion of the possible rate of return outcomes around the expected return.

For the portfolio P with n assets, the variance and covariance of individual assets are characterized by an $n \times n$ positive semi-definite matrix

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \vdots & \vdots & \dots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn} \end{pmatrix},$$

where σ_{ii} is the variance of asset i and σ_{ij} is the covariance between the assets i and j. The variance of the portfolio return, denoted by σ_P^2, can be computed by

$$\sigma_P^2 = \mathbf{w}^t \Sigma \mathbf{w} = \sum_{i,j=1}^{n} \omega_i \sigma_{ij} \omega_j, \tag{2.2.1}$$

where $\mathbf{w}^t$ is the transpose of w, and $\mathbf{w}$ is an $n-$dimensional vector whose i^{th} entry is ω_i the amount of capital invested in the asset i.

Remark 2.2.1 *A positive covariance means that the returns on two assets tend to move or change in the same direction, while a negative covariance means that the returns tend to move in opposite directions. The covariance is important because the variance of a portfolio's return depends on it and the key to diversification is the covariance of the asset returns.*

The principle of Markowitz diversification states that as the correlation (covariance) between the returns for assets that are combined in a portfolio decreases, so does the variance of the return for the portfolio. This is due to the degree of correlation between the expected asset returns.

2.3 The Set of Efficient Portfolios and the Optimal Portfolio

In the portfolio theory, different portfolios have different levels of expected return and risk. Also, the higher the level of expected return, the larger the risk. In the investment management process, the investor attempts to construct an efficient portfolio.

Definition 2.3.1 *An efficient portfolio is one that provides the greatest expected return for a given level of risk, or alternatively, the lowest risk for a given level of expected return.*

To construct an efficient portfolio, the investor must be able to quantify risk and provide the necessary inputs. There are three key inputs that are needed: expected return, variance of asset returns, and correlation (or covariance) of asset returns. The construction of an efficient portfolio based on the expected return of the portfolio and the variance of the portfolio's return is referred to as "mean-variance" portfolio management.

2.4 Markowitz Mean-Variance (MV) Model

MPT models the vector of assets in a portfolio as a multivariate normal random variable $N(\mu, \Sigma)$ where,

$$\mu = \begin{pmatrix} \mu_{R_1} \\ \mu_{R_2} \\ \vdots \\ \mu_{R_n} \end{pmatrix}.$$

As discussed before in Section 2.2 and Section 2.1, MPT defines the risk as the standard deviation of return, and models a portfolio as a weighted combination of assets so that the return of a portfolio is the weighted combination of the assets' returns.

The Markowitz mean-variance (MV) model has been used as the standard framework for optimal portfolio selection problem. In this MV, a portfolio is said to be optimal (MV efficient) if there is no portfolio having the same risk with a greater expected return and there is no portfolio having the same expected return with a lower risk. Therefore, a way to formulate the MV model mathematically is the following quadratic program:

$$\begin{cases} \min \mathbf{w}^t \Sigma \mathbf{w} \\ \mathbf{w}^t \mu = \beta, \quad \mathbf{w}^t \mathbf{e} = 1 \end{cases} \tag{2.4.1}$$

where β denotes the required expected reward.

In practice, Σ and $E(R_P)$ are unknown and we deal with estimators of them. Throughout the thesis, we are going to denote the estimators of Σ and $E(R_P)$ by $\widehat{\Sigma}$ and $\widehat{\mu}$, respectively. It is clear that using estimators of the required parameters will produce "noise". Since, in our study, we focus on the noise induced by estimating the covariance matrix and its effect on measuring the risk, then we will consider the following simplified version of the portfolio optimization problem in which we deal with risky assets:

$$\begin{cases} \min \mathbf{w}^t \Sigma \mathbf{w} \\ \mathbf{w}^t \mathbf{e} = 1. \end{cases} \tag{2.4.2}$$

where $\mathbf{e}$ is an $n \times 1$ vector with 1 in each entry.

2.5 Weights of the Optimal Portfolio

As discussed before, the goal of the portfolio optimization is to find a combination of assets $\{\omega_i\}$ that minimizes the risk of the portfolio for a given level of expected return or, in other words, a combination of assets that maximizes the expected return of the portfolio for a given level of risk. To do that, we need to solve an optimization problem as represented in (2.4.1) or in (2.4.2).

In [El-K], El-Karoui provides a solution for the following quadratic program:

$$\begin{cases} \min_{\mathbf{w} \in \mathbb{R}^n} \mathbf{w}^t \Sigma \mathbf{w} \\ \mathbf{w}^t \mathbf{v_i} = u_i, 1 \leq i \leq k. \end{cases} \tag{2.5.1}$$

where Σ is an $n \times n$ positive definite matrix, $\mathbf{v_i} \in \mathbb{R}^n$ and $u_i \in \mathbb{R}$.
El-Karoui depend on the method of Lagrange multipliers to solve the optimization problem in (2.5.1) and he stated the solution according to the following theorem.

Theorem 2.5.1 *[El-K] Let V be the $n \times k$ matrix whose i^{th}column is $\mathbf{v}_i$, $\mathbf{u}$ be the $k-$dimensional vector whose i^{th}entry is u_i and M be the $k \times k$ matrix such that*

$$M = V^t \Sigma^{-1} V,$$

where Σ^{-1} is the inverse of the covariance matrix Σ. Assume that $\mathbf{v}_i$'s are such that M is invertible. The solution of the quadratic program with linear equality constraints (2.5.1) is achieved by

$$\mathbf{w} = \Sigma^{-1} V M^{-1} U, \tag{2.5.2}$$

and we have

$$\mathbf{w}^t \Sigma \mathbf{w} = U^t M^{-1} U. \tag{2.5.3}$$

Proof: Let α be a $k-$dimensional vector of Lagrange multipliers. The Lagrangian function is, in matrix notation,

$$L(\mathbf{w}, \alpha) = \mathbf{w}^t \Sigma \mathbf{w} - 2\alpha^t (V^t \mathbf{w} - \mathbf{u}).$$

Then,

$$\frac{\partial L}{\partial \mathbf{w}} = 2\Sigma \mathbf{w} - 2V\alpha = 0.$$

So,

$$\mathbf{w} = \Sigma^{-1} V \alpha. \tag{2.5.4}$$

Now we know that

$$U = V^t \mathbf{w}. \tag{2.5.5}$$

By substituting from (2.5) into (2.5.5), we get

$$U = V^t \Sigma^{-1} V \alpha = M\alpha.$$

Therefore,

$$\mathbf{w} = \Sigma^{-1} V M^{-1} U.$$

It follows that

$$\mathbf{w}^t \Sigma \mathbf{w} = U^t M^{-1} U.$$

∎

Using Theorem 2.5.1, it is easy to formulate the optimal weights for the model 2.4.2 as shown in the following lemma.

Lemma 2.5.2 *For the optimization problem in (2.4.2), the optimal weights of the portfolio are given by*

$$\omega_i = \frac{\sum_{j=1}^{n} \sigma_{ij}^{(-1)}}{\sum_{j,k=1}^{n} \sigma_{jk}^{(-1)}} \quad (i = 1, \ldots, n) \tag{2.5.6}$$

where $(\sigma_{ij}^{(-1)})_{i,j=1}^{n}$ *are the entries of the matrix* Σ^{-1}.

Remark 2.5.3 *It is clear from Lemma 2.5.2 that the financial covariance matrices are the key input parameters to Markowitz's classical portfolio selection problem in (2.4.2).*

2.6 Risk of the Optimal Portfolio

As discussed in Section 2.2, the risk of the portfolio is the standard deviation of return. From (2.2.1), the risk σ_P of the portfolio P can be written as:

$$\sigma_P = \sqrt{\sum_{i,j=1}^{n} \omega_i \sigma_{ij} \omega_j} \tag{2.6.1}$$

As a consequence of Theorem 2.5.1, the risk of the optimal portfolio (2.4.2) can be written in terms of the entries of the inverse of the covariance matrix as shown in the following corollary.

Corollary 2.6.1 *The risk of the classical portfolio optimization problem in (2.4.2) is given by:*

$$\sigma_P = \frac{1}{\sqrt{\sum_{i,j=1}^{n} \sigma_{ij}^{(-1)}}}.$$

Also, the risk of the optimal portfolio (2.4.2) can be expressed in terms of the eigenvalues and the eigenvectors of the covariance matrix as shown in the following lemma.

Lemma 2.6.2 *The risk of the optimal portfolio (2.4.2) is given by:*

$$\sigma_P = \frac{1}{\sqrt{\sum_{r=1}^{n} \lambda_r^{-1}(1 + 2\sum_{i<j} o_{ri}o_{rj})}}$$

where, $(\lambda_r)_{r=1}^{n}$ *and* $(o_{r1}, o_{r2}, \ldots, o_{rn})^t$ *are the* r^{th} *eigenvalue and the corresponding eigenvector of* Σ*, respectively.*

Proof: The proof is straightforward by using Corollary 2.6.1 and the spectral decomposition of the symmetric matrix Σ. ∎

Chapter 3

Background in Random Matrices

Random matrix theory first gained attention in the 1950's in nuclear physics [W]. It was introduced by Eugene Wigner to describe the general properties of the energy levels of highly excited states of heavy nuclei. Random matrix theory (RMT) has found uses in a wide variety of problems in mathematics, physics and statistics. In multivariate statistics, random matrices were introduced by John Wishart [Wi], for statistical analysis of large samples. In this chapter, we will cover some basics and fundamentals of RMT.

3.1 RMT in Multivariate Statistics

As Johnstone [J] remarked:
"*It is a striking feature of the classical theory of multivariate statistical analysis that most of the standard techniques such as principal components, canonical correlations, multivariate analysis of variance (MANOVA) and discriminant analysis are founded on the eigenanalysis of covariance matrices*".

The sample covariance matrix is the most important random matrix in multivariate statistical inference. Thus it is not surprising that the methods of random matrix theory have important applications to multivariate statistical analysis. Many tests in statistics are defined by the eigenvalues of the covariance matrix. More recently, RMT is widely used in mathematics (operator algebra, mathematical physics, quantum information, etc). With vast data collection, data sets now have as many variables as the number of observations. In this context, the techniques and results of RMT have much to offer to multivariate statistics.

3.2 RMT in Finance

Empirical correlation matrices are of great importance for risk management and asset allocation. Results from the theory of random matrices are potentially of great interest to understand the statistical structure of the empirical correlation matrices appearing in the study of multivariate financial time series. RMT has recently been applied to noise filtering in financial time series, in particular, in large dimensional systems such as stock markets, by several authors including Plerou et al. ([PGRAS], [PGRAGS]) and Laloux et al. ([LCBP1], [LCBP2]). Both groups have analyzed US stock markets and have found that the eigenvalues of the correlation matrix of returns are consistent with those calculated using random returns, with the exception of a few large eigenvalues. Of particular interest was the demonstration ([LCBP2], [PGRAGS]), that filtering techniques based on RMT, could be beneficial in portfolio optimization to improve the forecast of the portfolio's risk. The applications of RMT to financial markets is a topic to which a considerable number of papers have been devoted to (see e.g. [LCBP1], [LCBP2], [PGRAGS], [PK], [GBP], [El-K]).

3.3 The Space of Random Matrices

A random variable is a measurable function on a probability space. An interesting feature of free probability theory [NiSp] is that it allows the algebras of random variables to be non-commutative. An example of such algebras is the algebra of random matrices. In this section, we will recall some definitions from RMT.

Definition 3.3.1 *A* non-commutative probability space *consists of a unital algebra* $\mathfrak{A}$ *with unital* $\mathbf{1}_{\mathfrak{A}}$ *over the space of complex numbers* $\mathbb{C}$ *and a unital linear functional*

$$\Phi : \mathfrak{A} \to \mathbb{C}; \qquad \Phi(\mathbf{1}_{\mathfrak{A}}) = 1.$$

We will denote it by $(\mathfrak{A}, \Phi)$*.*

Remark 3.3.2 *For a non-commutative probability space* $(\mathfrak{A}, \Phi)$*, if for every* $u, v \in \mathfrak{A}$

$$\Phi(uv) = \Phi(vu),$$

Then, $(\mathfrak{A}, \Phi)$ *is called tracial.*

Definition 3.3.3 *Let* $(\mathfrak{A}, \Phi)$ *be a non-commutative probability space.* $(\mathfrak{A}, \Phi)$ *is called a* $*$-probability space *if* $\mathfrak{A}$ *is a* $*$*-algebra and* Φ *is positive, i.e.*

$$\text{for every } u \in \mathfrak{A}, \qquad \Phi(u^* u) \geq 0$$

.

For a $*$-probability space $(\mathfrak{A}, \Phi)$, let us remind the reader that for $u \in \mathfrak{A}$

- if $u = u^*$ then, u is self-adjoint,
- if $u^* u = uu^* = 1$ then, u is unitary, and
- if $u^* u = uu^*$ then, u is a normal random variable.

Example 3.3.4 *Let* $M_n(\mathbb{C})$ *be the algebra of* $n \times n$ *complex matrices with usual matrix multiplication, and let* tr *be the normalized trace, i.e. for* $A = (a_{ij})_{i,j=1}^n$

$$\operatorname{tr}(A) := \frac{1}{n} \sum_{i=1}^{n} a_{ii}.$$

Then, $(M_n(\mathbb{C}), \operatorname{tr})$ *is a* $*$*-probability space where the* $*$*-operation is the adjoint of the matrix.*

Random matrices are matrices whose entries are classical random variables.

Definition 3.3.5 *A $*$-probability space of $n \times n$ random matrices is given by*

$$(M_n(L^{\infty-}(\Omega, \mu)), \mathrm{tr} \otimes E)$$

where $L^{\infty-}(\Omega, \mu)$ is the algebra of those random variables of the classical probability space (Ω, μ) which have finite moments of all orders and E denotes the expectation with respect to μ, i.e. for an $n \times n$ matrix $A = (a_{ij})_{i,j=1}^n$

$$\mathrm{tr} \otimes E(A) := \frac{1}{n} \sum_{i=1}^{n} \int_{\Omega} a_{ii}(\omega) d\mu(\omega).$$

3.4 Limiting Spectral Distribution (LSD)

In the space of random matrices, the most important information is contained in the eigenvalues of the matrices and the most important analytical object is the distribution of the eigenvalues.

Definition 3.4.1 *For any square matrix A, the probability distribution μ_A which puts equal mass on each eigenvalue of A is called the* empirical spectral distribution *(ESD) of A, i.e. if $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of the $n \times n$ matrix A, then*

$$\mu_A := \frac{1}{n}(\delta_{\lambda_1} + \delta_{\lambda_2} + \cdots + \delta_{\lambda_n})$$

where δ is the Kronecker delta.

Remark 3.4.2 *If λ is an eigenvalue of an $n \times n$ matrix A of multiplicity m, then the ESD μ_A puts mass m/n at λ.*

Remark 3.4.3 *Since we are interested in the large dimension matrices, then we are going to denote an $n \times n$ matrix A by A_n.*

Definition 3.4.4 *Let $(A_n)_{n=1}^{\infty}$ be a sequence of square matrices with the corresponding ESD $(\mu_n)_{n=1}^{\infty}$. The* limiting spectral distribution *(LSD) of the sequence is defined as the weak limit of the sequence (μ_n), if it exists.*

Remark 3.4.5 *Note that if the entries of the matrices (A_n) are random, then (μ_n) are random measures and the LSD is understood to be in some probabilistic sense, such as almost surely or in probability.*

Definition 3.4.6 *Let A_n be an $n \times n$ random matrix. If $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the n eigenvalues of A_n, then the* empirical spectral distribution function *(ESDF) of A_n is given by*

$$F_n(x, y) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{Re(\lambda_i) \leq x, Im(\lambda_i) \leq y\}$$

where $\mathbf{1}\{\cdot\}$ is the indicator of the event $\{\cdot\}$, Re and Im stand for the operations of taking the real and the imaginary parts of a complex number, respectively. The expected spectral distribution function of A_n is defined as $E(F_n(\cdot))$.

In the following, we describe the two most powerful tools which have been used in establishing LSDs. One is the moment method and the other is the method of Stieltjes Transform.

3.4.1 The Moment Method

Suppose (Y_n) is a sequence of real valued random variables. Suppose that there exists some (non-random) sequence such that for every positive integer k,

$$E(Y_n^k) \to \alpha_k$$

where (α_k) satisfies Carleman's condition. This condition, proposed by Torsten Carleman in 1922 (see [Akh]), requires the (α_k) to be chosen such that

$$\sum_{k=1}^{\infty} \alpha_{2k}^{-1/2k} = \infty.$$

It is well-known that there exists a distribution function μ, such that for all k,

$$\alpha_k = \int x^k d\mu(x) \qquad \text{and} \qquad Y_n \to \mu \ \text{ in distribution.}$$

For a positive integer k, the k-th moment of the ESD μ_n of an $n \times n$ matrix A_n with real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ has the following nice form:

$$\begin{aligned} \alpha_k(\mu_n) &= \int_{\mathbb{R}} x^k d\mu_n(x) \\ &= \frac{1}{n}\sum_{i=1}^{n} \lambda_i^k \\ &= \operatorname{tr}(A_n^k). \end{aligned}$$

Remark 3.4.7 *Note that the moments of the measure μ_n are exactly the moments of the matrix A_n with respect to the trace.*

Now, suppose $(A_n)_{n=1}^{\infty}$ is a sequence of square random matrices such that

$$\alpha_k(\mu_n) \to \alpha_k. \tag{3.4.1}$$

Remark 3.4.8 *In (3.4.1), the convergence takes place either "in probability" or "almost surely" and (α_k) are non-random.*

Now, if (α_k) satisfies Carleman's condition, then we can say that the LSD of the sequence (A_n) is μ (in the corresponding in probability or almost sure sense). We assume that the LSD has all moments finite. The method is not practically manageable in a wide variety of cases. The combinatorial arguments involved in the counting become quite unwieldy and even practically impossible as k and n increase. In cases where this method has been successful, the combinatorial arguments are very intricate. The relation (3.4.1) can often be verified by showing that $E(\alpha_k(\mu_n)) \to \alpha_k$ and $Var(\alpha_k(\mu_n)) \to 0$.

3.4.2 Stieltjes Transform Method

Stieltjes transforms play an important role in deriving LSDs. The Stieltjes transform is defined as follows.

Definition 3.4.9 *Let μ be a probability measure on $\mathbb{R}$. The Stieltjes transform of μ is the function m_μ defined on the upper half plane $\mathbb{C}^+ = \{u + iv : \ u, v \in \mathbb{R}, v > 0\}$ by the formula*

$$m_\mu(z) = \int_{\mathbb{R}} \frac{1}{z-t} d\mu(t), \qquad z \in \mathbb{C}^+.$$

Remark 3.4.10 *If a sequence of Stieltjes transforms converges uniformly on compact sets of $\mathbb{C}^+$, then the corresponding distributional convergence holds.*

Remark 3.4.11 *Suppose that μ is compactly supported. Let $r := \sup\{|t|, \ t \in$ support$(\mu_n)\}$. We then have the power series expansion:*

$$m_\mu(z) = \sum_{k=0}^{\infty} \frac{\alpha_k}{z^{k+1}}, \quad |z| > r$$

where α_k is the k-th moment of μ. Note that the previous expansion of m_μ around the point at infinity implies that for $z \in \mathbb{C}^+$

$$\lim_{|z| \to \infty} z m_\mu(z) = 1. \tag{3.4.2}$$

Let A_n be an $n \times n$ matrix with real eigenvalues. The Stieltjes transform of the ESD (μ_n) of A_n is

$$\begin{aligned} m_{\mu_n}(z) &= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\lambda_i - z} \\ &= \mathrm{tr}(A_n - zI_n)^{-1} \end{aligned}$$

where I_n is the $n \times n$ identity matrix.

Let (A_n) be a sequence of random matrices with real eigenvalues and let the corresponding sequence of Stieltjes transforms be (m_{μ_n}). If $m_{\mu_n} \to m$, in some suitable manner, where m is a Stieltjes transform, then the LSD of the sequence (A_n) is the unique probability on the real line whose Stieltjes transform is the function m.

Remark 3.4.12 *The convergence of the sequence (m_{μ_n}) is often verified by first showing that it satisfies some (approximate) recursion equation. Solving the limiting form of this equation identifies the Stieltjes transform of the LSD.*

For every $\varepsilon > 0$ and $t \in \mathbb{R}$, let

$$g_\varepsilon(t) := -\frac{1}{\pi}\ Im(m_\mu(t + \imath\varepsilon)).$$

The *Stieltjes inversion formula* states that

$$d\mu(t) = \lim_{\varepsilon\to 0} g_\varepsilon(t)dt. \tag{3.4.3}$$

The latter limit is considered in the weak topology on the space of probability measures on $\mathbb{R}$, and thus for every bounded continuous complex valued function $f : \mathbb{R} \to \mathbb{C}$,

$$\int_{\mathbb{R}} f(t)d\mu(t) = \lim_{\varepsilon\to 0} \int_{\mathbb{R}} f(t)g_\varepsilon(t)dt.$$

This method has been successfully applied to the Wigner matrices and the sample covariance type matrices. See [Bai] for more details. In the following, we are going to describe the LSDs for some interesting random matrices: Wigner matrices and sample covariance matrices.

3.4.3 Wigner matrix and the Semi-Circular Law

A *Wigner matrix* was introduced by Wigner [W]. This matrix is of considerable interest to physicists.

Definition 3.4.13 *An $n \times n$ Wigner matrix W_n of scale parameter σ is an $n \times n$ hermitian matrix whose entries above the diagonal are i.i.d. complex random variables with zero mean and finite variance σ^2 and whose diagonal elements are i.i.d. real random variables.*

An interesting special Wigner matrix is self-adjoint Gaussian random matrix which is defined as follows.

Definition 3.4.14 *A self-adjoint Gaussian random matrix* $G_n = (g_{ij})_{i,j=1}^n$ *is a Wigner matrix of scale* $1/\sqrt{n}$ *whose entries* g_{ij} $(i, j = 1, \ldots, n)$ *form a complex Gaussian family, the collection of their real and imaginary parts is a Gaussian family, is determined by the covariance*

$$E(g_{ij}g_{kl}) = \frac{1}{n}\delta_{il}\delta_{jk}.$$

In 1955, Wigner [W] showed that the semi-circular law with scale parameter σ arises as the LSD spectral distribution of $\frac{1}{\sqrt{n}}W_n$ (see also [AGZ] for more details). It has the density function

$$\mu^\sigma(x) = \begin{cases} \frac{1}{2\pi\sigma^2}\sqrt{4\sigma^2 - x^2} & \text{if } |x| \leq 2\sigma \\ 0 & \text{otherwise.} \end{cases}$$

All its odd moments are zero. The even moments are given in terms of the Catalan numbers C_k as follows

$$\int x^{2k} d\mu^\sigma(x) = C_k\sigma^{2k}. \tag{3.4.4}$$

Hence, the convergence of self-adjoint Gaussian matrices as shown in the following theorem.

Theorem 3.4.15 ***(Wigner's semicircle law)*** *Let* $(G_n)_{n=1}^\infty$ *be a sequence of self-adjoint Gaussian* $n \times n$ *matrices. Then as* n *tends to infinity,* (G_n) *converges in distribution towards a semicircle element with scale parameter* 1.

In [NiSp], it is shown that using the relation (3.4.4), the Stieltjes transform m_{μ_n} of self-adjoint Gaussian matrices satisfies the following recursion relation:

$$m_{\mu_n}(z) = \frac{1}{z} + \frac{1}{z}m_{\mu_n}(z)^2.$$

Using this, the Stieltjes transform satisfies the quadratic equation

$$m_{\mu_n}(z)^2 - z m_{\mu_n}(z) + 1 = 0, \qquad z \in \mathbb{C}^+.$$

This equation has two solutions for each $z \in \mathbb{C}^+$. From 3.4.2, the correct solution is

$$m_{\mu_n}(z) = z + \sqrt{z^2 - 4}$$

which is indeed the Stieltjes transform of the semicircular law with scale parameter 1.

3.4.4 Sample Covariance type matrices and the Marçenko-Pastur Law

Sample covariance matrices play a fundamental role in multivariate statistics and they are defined as follows.

Definition 3.4.16 *Suppose $(x_{ij}, i = 1, \ldots, n,\ j = 1, \ldots, T)$ is a double array of i.i.d. complex random variables with mean zero and finite variance σ^2. Write $\mathbf{x}_j = (x_{1j}, ..., x_{nj})^t$ and let $X^t = [\mathbf{x}_1 \mathbf{x}_2 \ldots \mathbf{x}_T]$. The matrix $S_{n,T} = \frac{1}{T} X^t X$ is called a sample covariance matrix.*

Remark 3.4.17 *If (x_{ij}) are real normal random variables with mean zero and variance one, then $S_{n,T}$ is a white Wishart matrix.*

The LSD μ_n of S_n was first established by Marçenko and Pastur [MP].

Theorem 3.4.18 *(**Marçenko Pastur law**) Suppose that $n, T \to \infty$ such that the ratio $n/T \to r \in (0, \infty)$. Then $\mu_n \to \mu$ (in distribution) where*

$$\mu(x) = \begin{cases} (1 - \frac{1}{r})\mathbf{1}\{x = 0\} + \frac{\sqrt{(b-x)(x-a)}}{2\pi r x \sigma^2}\mathbf{1}\{x \in [a, b]\} & \text{if } r > 1 \\ \frac{\sqrt{(b-x)(x-a)}}{2\pi r x \sigma^2}\mathbf{1}\{x \in [a, b]\} & \text{if } 0 \leq r \leq 1 \end{cases}$$

with $a = \sigma^2(1 - \sqrt{r})^2$ and $b = \sigma^2(1 + \sqrt{r})^2$.

Depending on this theorem, Laloux et al. [LCBP2] established a technique called the " *Filtering technique*" to clean the empirical covariance matrix of the returns of assets. This technique will be discussed in detail in Chapter 5.

In Chapter 4, we discuss a generalization of Wishart matrices called compound Wishart matrices. We will formulate the moments of their inverse which will play a fundamental role in our work.

Chapter 4

Integration of Invariant Matrices and Application to Statistics

In this chapter, we consider random matrices that have invariance properties under the action of unitary groups (either a left-right invariance, or a conjugacy invariance), and we give formulas for moments in terms of functions of eigenvalues. Our main tool is the Weingarten calculus. As an application to statistics, we obtain new formulas for the pseudo inverse of Gaussian matrices and for the inverse of compound Wishart matrices.

4.1 Introduction

Wishart matrices have been introduced and studied for the first time for statistical purposes in [Wi], and they are still a fundamental random matrix model related to theoretical statistics. One generalization of Wishart matrices is compound Wishart matrices which are studied, for example, in [Sp, HP].

The study of the eigenvalues of Wishart matrices is quite well developed but a systematic study of the joint moments of their entries (that we will call *local moments*) is more recent. On the other hand, the theoretical study of the inverse of Wishart matrices is also very important, in particular for mathematical finance purposes, as shown in ([CW] and [CMW]). However, the study of their local moments is much more recent, and is actually still open in the case of the inverse of the compound Wishart matrix.

Our approach is based on the Weingarten calculus. This tool is used to compute the local moments of random matrices distributed according to Haar measures on compact groups such as the unitary or the orthogonal group. It was introduced in [We] and then improved many times, with a first complete description in [C, CS].

In our work, we need to introduce a modification of the Weingarten function, namely, a 'double' Weingarten function with two dimension parameters instead of one. As far as we know it is the first time that such a double-parameter Weingarten function is needed. Beyond proving to be efficient in systematically computing moments, we believe that it will turn out to have important theoretical properties. The aim of this chapter is to provide a unified approach to the problem of computing the local moments of the above random matrix models.

As an interesting byproduct of our study - and as a preliminary to the solution of our problem of computing the moments of the inverse of a compound Wishart random matrix, we obtain explicit moment formulas for the pseudo-inverse of Ginibre random matrices.

The chapter is organized as follows. In Section 4.3 and Section 4.4, we recall known results about Weingarten calculus and Wishart matrices, respectively. Section 4.5 is devoted to the computation of moments of general invariant random matrices, and in Section 4.6, we systematically solve the problem of computing moments of inverses of compound Wishart matrices. In the following section, we are going to give the notation used in the chapter.

4.2 Notation

4.2.1 The Complex Case

Let k be a positive integer. A partition of k is a weakly decreasing sequence $\lambda = (\lambda_1, \dots, \lambda_l)$ of positive integers with $\sum_{i=1}^{l} \lambda_i = k$, i.e. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_l > 0$. We write $\lambda \vdash k$ and we denote the length l of λ by $\ell(\lambda)$.

Let S_k be the symmetric group acting on $[k] = \{1, 2, \dots, k\}$. A permutation $\pi \in S_k$ is decomposed into cycles. If the lengths of cycles are $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_l$, then the sequence $\mu = (\mu_1, \mu_2, \dots, \mu_l)$ is a partition of k. We will refer to μ as the *cycle-type of* π. Denote by $\kappa(\pi)$ the length $\ell(\mu)$ of the cycle-type of π, or equivalently the number of cycles of π.

For two sequences $\mathbf{i} = (i_1, \dots, i_k)$ and $\mathbf{i}' = (i'_1, \dots, i'_k)$ of positive integers and for a permutation $\pi \in S_k$, we define

$$\delta_\pi(\mathbf{i}, \mathbf{i}') = \prod_{s=1}^{k} \delta_{i_{\pi(s)}, i'_s}. \tag{4.2.1}$$

Given a square matrix A and a permutation $\pi \in S_k$ of cycle-type $\mu = (\mu_1, \dots, \mu_l)$, we define

$$\mathrm{Tr}_\pi(A) = \prod_{j=1}^{l} \mathrm{Tr}(A^{\mu_j}). \tag{4.2.2}$$

Example 4.2.1 *Let*

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 5 & 4 & 3 & 1 & 8 & 7 & 6 \end{pmatrix} \in S_8.$$

Then, π is decomposed as $\pi = (1\ 2\ 5)(3\ 4)(6\ 8)(7)$ and the cycle-type of π is the partition $(3, 2, 2, 1)$. Hence,

$$\kappa(\pi) = 4$$

and

$$\mathrm{Tr}_\pi(A) = \mathrm{Tr}(A^3)\,\mathrm{Tr}(A^2)^2\,\mathrm{Tr}(A).$$

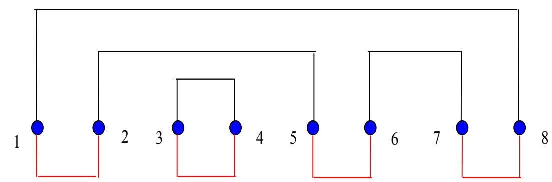


Figure 4.1: $\Gamma(\sigma)$

4.2.2 In the Real Case

Given $\sigma \in S_{2k}$, we attach an undirected graph $\Gamma(\sigma)$ with vertices $1, 2, \ldots, 2k$ and edge set consisting of

$$\big\{\{2i-1, 2i\} \mid i = 1, 2, \ldots, k\big\} \cup \big\{\{\sigma(2i-1), \sigma(2i)\} \mid i = 1, 2, \ldots, k\big\}.$$

Figure (4.1) describes the graph $\Gamma(\sigma)$ for $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 5 & 4 & 3 & 1 & 8 & 7 & 6 \end{pmatrix} \in S_8$.

Remark 4.2.2 *In the graph $\Gamma(\sigma)$, we distinguish every edge $\{2i-1, 2i\}$ from $\{\sigma(2j-1), \sigma(2j)\}$ even if these pairs coincide. Then each vertex of the graph lies on exactly two edges, and the number of vertices in each connected component is even.*

In the connected components of the graph, if the numbers of vertices are

$$2\mu_1 \geq 2\mu_2 \geq \cdots \geq 2\mu_l,$$

then the sequence $\mu = (\mu_1, \mu_2, \ldots, \mu_l)$ is a partition of k. We will refer to the μ as *the coset-type of* σ, see [Mac, VII.2] for more details.
Denote by $\kappa'(\sigma)$ the length $l(\mu)$ of the coset-type of σ, or equivalently the number of components of $\Gamma(\sigma)$.

Let M_{2k} be the set of all pair partitions of the set $[2k] = \{1, \ldots, 2k\}$. A pair partition $\sigma \in M_{2k}$ can be uniquely expressed in the form

$$\sigma = \big\{\{\sigma(1), \sigma(2)\}, \{\sigma(3), \sigma(4)\}, \ldots, \{\sigma(2k-1), \sigma(2k)\}\big\}$$

with $1 = \sigma(1) < \sigma(3) < \cdots < \sigma(2k-1)$ and $\sigma(2i-1) < \sigma(2i)$ $(1 \leq i \leq k)$.

Remark 4.2.3 *Let σ be a pair partition of the set $[2k]$. Then, σ can be regarded as a permutation*

$$\begin{pmatrix} 1 & 2 & \dots & 2k \\ \sigma(1) & \sigma(2) & \dots & \sigma(2k) \end{pmatrix} \in S_{2k}.$$

We thus embed M_{2k} into S_{2k}. In particular, the coset-type and the value of κ' for $\sigma \in M_{2k}$ are defined.

For a permutation $\sigma \in S_{2k}$ and a $2k$-tuple $\mathbf{i} = (i_1, i_2, \dots, i_{2k})$ of positive integers, we define

$$\delta'_\sigma(\mathbf{i}) = \prod_{s=1}^{k} \delta_{i_{\sigma(2s-1)}, i_{\sigma(2s)}}. \tag{4.2.3}$$

Remark 4.2.4 *In particular, if $\sigma \in M_{2k}$, then $\delta'_\sigma(\mathbf{i}) = \prod_{\{a,b\}\in\sigma} \delta_{i_a, i_b}$, where the product runs over all pairs in σ.*

For a square matrix A and $\sigma \in S_{2k}$ with coset-type $(\mu_1, \mu_2, \dots, \mu_l)$, we define

$$\mathrm{Tr}'_\sigma(A) = \prod_{j=1}^{l} \mathrm{Tr}(A^{\mu_j}). \tag{4.2.4}$$

Example 4.2.5 *As in Example 4.2.1, let*

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 5 & 4 & 3 & 1 & 8 & 7 & 6 \end{pmatrix} \in S_8.$$

Then, the coset-type of π is the partition $(3, 1)$, which gives

$$\kappa'(\pi) = 2$$

and

$$\mathrm{Tr}'_\pi(A) = \mathrm{Tr}(A^3)\,\mathrm{Tr}(A).$$

4.3 Weingarten Calculus

4.3.1 Unitary Weingarten Calculus

Here, we review some basic material on unitary integration and the unitary Weingarten function. A more complete exposition of these matters can be found in [C, CS, MN]. We use notation defined in Section 4.2.1.

Let $L(S_k)$ be the algebra of complex-valued functions on S_k with convolution

$$(f_1 * f_2)(\pi) = \sum_{\tau \in S_k} f_1(\tau) f_2(\tau^{-1}\pi) \qquad (f_1, f_2 \in L(S_k),\ \pi \in S_k).$$

The identity element in the algebra $L(S_k)$ is the Dirac function δ_e at the identity permutation $e = e_k \in S_k$ i.e., for $\pi \in S_k$

$$\delta_e(\pi) = \begin{cases} 1 & \text{if } \pi = e \\ 0 & \text{if } \pi \neq e \end{cases}$$

Let z be a complex number and consider the function $z^{\kappa(\cdot)}$ in $L(S_k)$ defined by

$$S_k \ni \pi \mapsto z^{\kappa(\pi)} \in \mathbb{C},$$

which belongs to the center $\mathcal{Z}(L(S_k))$ of $L(S_k)$. The *unitary Weingarten function*

$$S_k \ni \pi \mapsto \mathrm{Wg}^{\mathrm{U}}(\pi; z) \in \mathbb{C}$$

is, by definition, the pseudo-inverse element of $z^{\kappa(\cdot)}$ in $\mathcal{Z}(L(S_k))$ i.e., the unique element in $\mathcal{Z}(L(S_k))$ satisfying

$$z^{\kappa(\cdot)} * \mathrm{Wg}^{\mathrm{U}}(\cdot; z) * z^{\kappa(\cdot)} = z^{\kappa(\cdot)},$$

and

$$\mathrm{Wg}^{\mathrm{U}}(\cdot; z) * z^{\kappa(\cdot)} * \mathrm{Wg}^{\mathrm{U}}(\cdot; z) = \mathrm{Wg}^{\mathrm{U}}(\cdot; z).$$

The expansion of the unitary Weingarten function in terms of irreducible characters χ^λ of S_k is given by Collins et al. [CS] as follows

$$\mathrm{Wg}^{\mathrm{U}}(\pi; z) = \frac{1}{k!} \sum_{\substack{\lambda \vdash k \\ C_\lambda(z) \neq 0}} \frac{f^\lambda}{C_\lambda(z)} \chi^\lambda(\pi) \qquad (\pi \in S_k),$$

summed over all partitions λ of k satisfying $C_\lambda(z) \neq 0$. Here $f^\lambda = \chi^\lambda(e)$ and

$$C_\lambda(z) = \prod_{i=1}^{\ell(\lambda)} \prod_{j=1}^{\lambda_i} (z + j - i).$$

In particular, unless $z \in \{0, \pm 1, \pm 2, \ldots, \pm(k-1)\}$, the functions $z^{\kappa(\cdot)}$ and $\mathrm{Wg}^{\mathrm{U}}(\cdot; z)$ are inverses of each other and satisfy $z^{\kappa(\cdot)} * \mathrm{Wg}^{\mathrm{U}}(\cdot; z) = \delta_e$.

Proposition 4.3.1 ([C]) *Let $U = (u_{ij})_{1 \le i,j \le n}$ be an $n \times n$ Haar-distributed unitary matrix. For four sequences $\mathbf{i} = (i_1, i_2, \ldots, i_k)$, $\mathbf{j} = (j_1, j_2, \ldots, j_k)$, $\mathbf{i}' = (i'_1, i'_2, \ldots, i'_k)$, $\mathbf{j}' = (j'_1, j'_2, \ldots, j'_k)$ of positive integers in $[n]$, we have*

$$E[u_{i_1 j_1} \ldots u_{i_k j_k} \overline{u_{i'_1 j'_1} \cdots u_{i'_k j'_k}}] = \sum_{\sigma, \tau \in S_k} \delta_\sigma(\mathbf{i}, \mathbf{i}') \delta_\tau(\mathbf{j}, \mathbf{j}') \mathrm{Wg}^{\mathrm{U}}(\sigma^{-1}\tau; n). \tag{4.3.1}$$

We will need the following function later. Define the function $\mathrm{Wg}^{\mathrm{U}}(\cdot; z, w)$ on S_k with two complex parameters $z, w \in \mathbb{C}$ by the convolution

$$\mathrm{Wg}^{\mathrm{U}}(\cdot; z, w) = \mathrm{Wg}^{\mathrm{U}}(\cdot; z) * \mathrm{Wg}^{\mathrm{U}}(\cdot; w). \tag{4.3.2}$$

More precisely,

$$\mathrm{Wg}^{\mathrm{U}}(\cdot; z, w) = \frac{1}{k!} \sum_{\substack{\lambda \vdash k \\ C_\lambda(z) C_\lambda(w) \neq 0}} \frac{f^\lambda}{C_\lambda(z) C_\lambda(w)} \chi^\lambda.$$

4.3.2 Orthogonal Weingarten Calculus

We next review the theory of orthogonal integration and the orthogonal Weingarten function. See [CS, CM, M1, M2, M3] for more details. We use notation defined in Section 4.2.2.

Let H_k be the hyperoctahedral group of order $2^k k!$, which is the centralizer of t_k in S_{2k}, where $t_k \in S_{2k}$ is the product of the transpositions $(1\ 2), (3\ 4), \dots, (2k-1\ 2k)$. Let $L(S_{2k}, H_k)$ be the subspace of all H_k-bi-invariant functions in $L(S_{2k})$:

$$L(S_{2k}, H_k) = \{f \in L(S_{2k}) \mid f(\zeta\sigma) = f(\sigma\zeta) = f(\sigma) \quad (\sigma \in S_{2k},\ \zeta \in H_k)\}.$$

We introduce another product on $L(S_{2k}, H_k)$. For $f_1, f_2 \in L(S_{2k}, H_k)$, we define

$$(f_1 \sharp f_2)(\sigma) = \sum_{\tau \in M_{2k}} f_1(\sigma\tau) f_2(\tau^{-1}) \qquad (\sigma \in S_{2k}).$$

Remark 4.3.2 *Note that $f_1 \sharp f_2 = (2^k k!)^{-1} f_1 * f_2$. In fact, since M_{2k} gives the representative of cosets σH_k in S_{2k} and since f_1, f_2 are H_k-bi-invariant, we have*

$$\begin{aligned} (f_1 * f_2)(\sigma) &= \sum_{\tau \in M_{2k}} \sum_{\zeta \in H_k} f_1(\sigma(\tau\zeta)) f_2((\tau\zeta)^{-1}) \\ &= \sum_{\tau \in M_{2k}} \sum_{\zeta \in H_k} f_1(\sigma\tau) f_2(\tau^{-1}) = |H_k| (f_1 \sharp f_2)(\sigma). \end{aligned}$$

The new product $\sharp$ is almost the same as the convolution $*$ on $L(S_{2k}, H_k)$ up to the normalization factor $2^k k!$, but it will be convenient in the present context. We note that $L(S_{2k}, H_k)$ is a commutative algebra under the product $\sharp$ with the identity element

$$\mathbf{1}_{H_k}(\sigma) = \begin{cases} 1 & \text{if } \sigma \in H_k \\ 0 & \text{otherwise.} \end{cases}$$

Consider the function $z^{\kappa'(\cdot)}$ with a complex parameter z defined by

$$S_{2k} \ni \sigma \mapsto z^{\kappa'(\sigma)} \in \mathbb{C},$$

which belongs to $L(S_{2k}, H_k)$. The *orthogonal Weingarten function* $\mathrm{Wg}^{\mathrm{O}}(\sigma; z)$ $(\sigma \in S_{2k})$ is the unique element in $L(S_{2k}, H_k)$ satisfying

$$z^{\kappa'(\cdot)} \sharp \mathrm{Wg}^{\mathrm{O}}(\cdot; z) \sharp z^{\kappa'(\cdot)} = z^{\kappa'(\cdot)} \qquad \text{and} \qquad \mathrm{Wg}^{\mathrm{O}}(\cdot; z) \sharp z^{\kappa'(\cdot)} \sharp \mathrm{Wg}^{\mathrm{O}}(\cdot; z) = \mathrm{Wg}^{\mathrm{O}}(\cdot; z).$$

For each partition λ of k, the zonal spherical function ω^λ is defined by

$$\omega^\lambda = (2^k k!)^{-1} \chi^{2\lambda} * \mathbf{1}_{H_k},$$

where $2\lambda = (2\lambda_1, 2\lambda_2, \dots)$, and the family of ω^λ form a linear basis of $L(S_{2k}, H_k)$. The expansion of $\mathrm{Wg}^{\mathrm{O}}(\cdot; z)$ in terms of ω^λ is given by Collins and Matsumoto [CM] as follows

$$\mathrm{Wg}^{\mathrm{O}}(\sigma; z) = \frac{2^k k!}{(2k)!} \sum_{\substack{\lambda \vdash k \\ C'_\lambda(z) \neq 0}} \frac{f^{2\lambda}}{C'_\lambda(z)} \omega^\lambda(\sigma) \qquad (\sigma \in S_{2k}),$$

summed over all partitions λ of k satisfying $C'_\lambda(z) \neq 0$, where

$$C'_\lambda(z) = \prod_{i=1}^{\ell(\lambda)} \prod_{j=1}^{\lambda_i} (z + 2j - i - 1).$$

In particular, if $C'_\lambda(z) \neq 0$ for all partitions λ of k, functions $z^{\kappa'(\cdot)}$ and $\mathrm{Wg}^{\mathrm{O}}(\cdot; z)$ are the inverse of each other and satisfy $z^{\kappa'(\cdot)} \sharp \mathrm{Wg}^{\mathrm{O}}(\cdot; z) = \mathbf{1}_{H_k}$.

Definition 4.3.3 *Let* $\mathrm{O}(n)$ *be the real orthogonal group of degree* n*, equipped with its Haar probability measure.* $\mathrm{O}(n)$ *is called the group of* $n \times n$ Haar-distributed orthogonal matrices.

Proposition 4.3.4 ([CM]) *Let* $O = (o_{ij})_{1 \le i,j \le n}$ *be an* $n \times n$ *Haar-distributed orthogonal matrix. For two sequences* $\mathbf{i} = (i_1, \dots, i_{2k})$ *and* $\mathbf{j} = (j_1, \dots, j_{2k})$*, we have*

$$E[o_{i_1 j_1} o_{i_2 j_2} \cdots o_{i_{2k} j_{2k}}] = \sum_{\sigma, \tau \in M_{2k}} \delta'_\sigma(\mathbf{i}) \delta'_\tau(\mathbf{j}) \mathrm{Wg}^{\mathrm{O}}(\sigma^{-1}\tau; n). \tag{4.3.3}$$

Here $\sigma, \tau \in M_{2k}$ *are regarded as permutations in* S_{2k}*, and so is* $\sigma^{-1}\tau$*.*

We will also need the following function later. Define the function $\mathrm{Wg}^{\mathrm{O}}(\cdot; z, w)$ in $L(S_{2k}, H_k)$ with two complex parameters $z, w \in \mathbb{C}$ by

$$\mathrm{Wg}^{\mathrm{O}}(\cdot; z, w) = \mathrm{Wg}^{\mathrm{O}}(\cdot; z) \sharp \mathrm{Wg}^{\mathrm{O}}(\cdot; w). \tag{4.3.4}$$

More precisely,

$$\mathrm{Wg}^{\mathrm{O}}(\cdot; z, w) = \frac{2^k k!}{(2k)!} \sum_{\substack{\lambda \vdash k \\ C'_\lambda(z) C'_\lambda(w) \neq 0}} \frac{f^{2\lambda}}{C'_\lambda(z) C'_\lambda(w)} \omega^\lambda.$$

4.4 Wishart Matrices and their Inverse

4.4.1 Complex Wishart Matrices

Definition 4.4.1 *Let X be an $n \times p$ random matrix whose columns are i.i.d. complex vectors which follow n-dimensional complex normal distributions $\mathrm{N}_{\mathbb{C}}(\mathbf{0}, \Sigma)$, where Σ is an $n \times n$ positive definite Hermitian matrix. Then we call a random matrix $W = XX^*$ a (centered)* complex Wishart matrix.

We will need the computation of the local moments for the inverse W^{-1}.

Proposition 4.4.2 ([GLM]) *Let W be a complex Wishart matrix defined as above. Put $q = p - n$. If $\pi \in S_k$ and $q \geq k$, then*

$$E[\mathrm{Tr}_\pi(W^{-1})] = (-1)^k \sum_{\tau \in S_k} \mathrm{Wg}^{\mathrm{U}}(\pi\tau^{-1}; -q)\, \mathrm{Tr}_\tau(\Sigma^{-1}). \tag{4.4.1}$$

4.4.2 Real Wishart Matrices

Definition 4.4.3 *Let X be an $n \times p$ random matrix whose columns are i.i.d. vectors which follow n-dimensional real normal distributions $\mathrm{N}_{\mathbb{R}}(\mathbf{0}, \Sigma)$, where Σ is an $n \times n$ positive definite real symmetric matrix. Then we call a random matrix $W = XX^t$ a (centered)* real Wishart matrix.

Proposition 4.4.4 ([M2]) *Let W be a real Wishart matrix defined as above. Put $q = p - n - 1$.*

If $\pi \in M_{2k}$ and $q \geq 2k - 1$, then

$$E[\mathrm{Tr}'_\pi(W^{-1})] = (-1)^k \sum_{\tau \in M_{2k}} \mathrm{Wg}^{\mathrm{O}}(\pi\tau^{-1}; -q)\, \mathrm{Tr}'_\tau(\Sigma^{-1}). \tag{4.4.2}$$

4.5 Invariant Random Matrices

In this section we consider random matrices with invariance property and establish the link between local and global moments.

4.5.1 Conjugacy Invariance (Unitary Case)

Theorem 4.5.1 *Let $W = (w_{ij})$ be an $n \times n$ complex Hermitian random matrix with the invariant property such that UWU^* has the same distribution as W for any unitary matrix U. For two sequences $\mathbf{i} = (i_1, \ldots, i_k)$ and $\mathbf{j} = (j_1, \ldots, j_k)$, we have*

$$E[w_{i_1 j_1} w_{i_2 j_2} \ldots w_{i_k j_k}] = \sum_{\sigma, \tau \in S_k} \delta_\sigma(\mathbf{i}, \mathbf{j}) \mathrm{Wg}^{\mathrm{U}}(\sigma^{-1}\tau; n) E[\mathrm{Tr}_\tau(W)],$$

where $\delta_\sigma(\cdot)$ and $\mathrm{Tr}_\tau(\cdot)$ are defined in (4.2.1) and (4.2.2), respectively.

Before we prove this theorem we need the following lemma

Lemma 4.5.2 *Let W be as in Theorem 4.5.1. W has the same distribution as UDU^*, where U is a Haar distributed random unitary matrix, D is a diagonal matrix whose eigenvalues have the same distribution as those of W, and D, U are independent.*

Proof: Let U, D be matrices (U unitary, and D diagonal) such that $W = UDU^*$. It is possible to have U, D as measurable functions of W (if the singular values have no multiplicity this follows from the fact that U can be essentially chosen in a canonical way, and in the general case, it follows by an approximation argument). So, we may consider that U, D are also random variables and that the σ-algebra generated by U, D is the same as the σ-algebra generated by W.

Let V be a deterministic unitary matrix. The fact that VWV^* has the same distribution as W and our previous uniqueness considerations imply that VU has the same distribution as U. By uniqueness of the Haar measure, this implies that U has to be distributed according to the Haar measure.

To conclude the proof, we observe that instead of taking V to be a deterministic unitary matrix, we could have taken V random, independent from W, and distributed according to the Haar measure without changing the fact that VWV^* has the same distribution as W. This implies that U can be replaced by VU, and clearly, VU is Haar distributed, and independent from D, so the proof is complete. ∎

Proof: (**Proof of Theorem 4.5.1**)

From Lemma 4.5.2, each matrix entry w_{ij} has the same distribution as $\sum_{r=1}^{n} u_{ir} d_r \overline{u_{jr}}$, where $U = (u_{ij})$ and $D = \mathrm{diag}(d_1, \dots, d_n)$ are unitary and diagonal matrices respectively and U, D are independent. It follows that

$$\begin{aligned} & E[w_{i_1 j_1} w_{i_2 j_2} \cdots w_{i_k j_k}] \\ = & \sum_{\mathbf{r}=(r_1,\dots,r_k)} E[d_{r_1} d_{r_2} \cdots d_{r_k}] \cdot E[u_{i_1 r_1} u_{i_2 r_2} \cdots u_{i_k r_k} \overline{u_{j_1 r_1} u_{j_2 r_2} \cdots u_{j_k r_k}}]. \end{aligned}$$

The unitary Weingarten calculus (Proposition 4.3.1) gives

$$\begin{aligned} &= \sum_{\mathbf{r}=(r_1,\dots,r_k)} E[d_{r_1} d_{r_2} \cdots d_{r_k}] \sum_{\sigma,\tau \in S_k} \delta_\sigma(\mathbf{i},\mathbf{j}) \delta_\tau(\mathbf{r},\mathbf{r}) \mathrm{Wg}^{\mathrm{U}}(\sigma^{-1}\tau; n) \\ &= \sum_{\sigma,\tau \in S_k} \delta_\sigma(\mathbf{i},\mathbf{j}) \mathrm{Wg}^{\mathrm{U}}(\sigma^{-1}\tau; n) \sum_{\mathbf{r}=(r_1,\dots,r_k)} \delta_\tau(\mathbf{r},\mathbf{r}) E[d_{r_1} d_{r_2} \cdots d_{r_k}]. \end{aligned}$$

To conclude the proof, we have to show: For $\tau \in S_k$ and a diagonal matrix $D = \mathrm{diag}(d_1, \dots, d_n)$,

$$\sum_{\mathbf{r}=(r_1,\dots,r_k)} \delta_\tau(\mathbf{r},\mathbf{r}) d_{r_1} d_{r_2} \cdots d_{r_k} = \mathrm{Tr}_\tau(D). \tag{4.5.1}$$

We observe that $\delta_\tau(\mathbf{r},\mathbf{r})$ survives if and only if all r_i in each cycle of τ coincide. Hence, if τ has the cycle-type $\mu = (\mu_1, \dots, \mu_l)$, then

$$\sum_{\mathbf{r}=(r_1,\dots,r_k)} \delta_\tau(\mathbf{r},\mathbf{r}) d_{r_1} d_{r_2} \cdots d_{r_k} = \sum_{s_1,\dots,s_l} d_{s_1}^{\mu_1} \cdots d_{s_l}^{\mu_l} = \mathrm{Tr}(D^{\mu_1}) \cdots \mathrm{Tr}(D^{\mu_l}) = \mathrm{Tr}_\tau(D),$$

which proves (4.5.1). ∎

Example 4.5.3 *Let W be as in Theorem 4.5.1. For each $1 \le i \le n$ and $k \ge 1$,*

$$E[w_{ii}^k] = \frac{1}{n(n+1)\cdots(n+k-1)} \sum_{\mu \vdash k} \frac{k!}{z_\mu} E\left[\prod_{j=1}^{\ell(\mu)} \mathrm{Tr}(W^{\mu_j})\right] \tag{4.5.2}$$

summed over all partition μ of k. Here

$$z_\mu = \prod_{i \geq 1} i^{m_i(\mu)}\, m_i(\mu)!$$

with the multiplicities $m_i(\mu)$ of i in μ. In fact, Theorem 4.5.1 implies the identity $E[w_{ii}^k] = \sum_{\sigma \in S_k} \mathrm{Wg}^{\mathrm{U}}(\sigma; n) \cdot \sum_{\tau \in S_k} E[\mathrm{Tr}_\tau(W)]$, and the claim therefore is obtained by the following two known facts:

$$\sum_{\sigma \in S_k} \mathrm{Wg}^{\mathrm{U}}(\sigma; n) = \frac{1}{n(n+1)\cdots(n+k-1)};$$

the number of permutations in S_k of cycle-type μ is $k!/z_\mu$. When $k=1$ the equation (4.5.2) gives a trivial identity $E[w_{ii}] = \frac{1}{n} E[\mathrm{Tr}(W)]$. When $k = 2, 3$, it gives

$$\begin{aligned} E[w_{ii}^2] =& \frac{1}{n(n+1)} (E[\mathrm{Tr}(W^2)] + E[\mathrm{Tr}(W)^2]); \\ E[w_{ii}^3] =& \frac{1}{n(n+1)(n+2)} (2E[\mathrm{Tr}(W^3)] + 3E[\mathrm{Tr}(W^2)\,\mathrm{Tr}(W)] + E[\mathrm{Tr}(W)^3]). \end{aligned}$$

4.5.2 Conjugacy Invariance (Orthogonal Case)

Theorem 4.5.4 *Let $W = (w_{ij})$ be an $n \times n$ real symmetric random matrix with the invariant property such that UWU^t has the same distribution as W for any orthogonal matrix U. For any sequence $\mathbf{i} = (i_1, \dots, i_{2k})$, we have*

$$E[w_{i_1 i_2} w_{i_3 i_4} \dots w_{i_{2k-1} i_{2k}}] = \sum_{\sigma, \tau \in M_{2k}} \delta'_\sigma(\mathbf{i}) \mathrm{Wg}^{\mathrm{O}}(\sigma^{-1}\tau; n) E[\mathrm{Tr}'_\tau(W)],$$

where $\delta'_\sigma(\cdot)$ and $\mathrm{Tr}'_\tau(\cdot)$ are defined in (4.2.3) and (4.2.4), respectively.

Proof: As in Lemma 4.5.2, W has the same distribution as UDU^t, where $U = (u_{ij})$ and $D = \mathrm{diag}(d_1, \dots, d_n)$ are orthogonal and diagonal matrices respectively and U, D are independent. We have

$$\begin{aligned}
&E[w_{i_1 i_2} w_{i_3 i_4} \dots w_{i_{2k-1} i_{2k}}] \\
&= \sum_{\mathbf{r}=(r_1,\dots,r_k)} E[d_{r_1} d_{r_2} \cdots d_{r_k}] \cdot E[u_{i_1 r_1} u_{i_2 r_1} u_{i_3 r_2} u_{i_4 r_2} \cdots u_{i_{2k-1} r_k} u_{i_{2k} r_k}],
\end{aligned}$$

and the orthogonal Weingarten calculus (Proposition 4.3.4) gives

$$\begin{aligned}
&= \sum_{\mathbf{r}=(r_1,\dots,r_k)} E[d_{r_1} d_{r_2} \cdots d_{r_k}] \sum_{\sigma,\tau \in M_{2k}} \delta'_\sigma(\mathbf{i}) \delta'_\tau(\tilde{\mathbf{r}}) \mathrm{Wg}^{\mathrm{O}}(\sigma^{-1}\tau; n) \\
&= \sum_{\sigma,\tau \in M_{2k}} \delta'_\sigma(\mathbf{i}) \mathrm{Wg}^{\mathrm{O}}(\sigma^{-1}\tau; n) \sum_{\mathbf{r}=(r_1,\dots,r_k)} \delta'_\tau(\tilde{\mathbf{r}}) E[d_{r_1} d_{r_2} \cdots d_{r_k}],
\end{aligned}$$

where $\tilde{\mathbf{r}} = (r_1, r_1, r_2, r_2, \dots, r_k, r_k)$ for each $\mathbf{r} = (r_1, r_2, \dots, r_k)$.

Recall notation defined in Section 4.2.2. To conclude the proof, we have to show: For $\tau \in S_{2k}$ and a diagonal matrix $D = \mathrm{diag}(d_1, \dots, d_n)$,

$$\sum_{\mathbf{r}=(r_1,\dots,r_k)} \delta'_\tau(\tilde{\mathbf{r}}) d_{r_1} d_{r_2} \cdots d_{r_k} = \mathrm{Tr}'_\tau(D). \tag{4.5.3}$$

This equation follows from the following fact:

$\delta'_\tau(\tilde{\mathbf{r}})$ survives if and only if all the r_i in each component of the graph $\Gamma(\tau)$ coincide. ∎

Example 4.5.5 *Let W be as in Theorem 4.5.4. For each $1 \le i \le n$ and $k \ge 1$,*

$$E[w_{ii}^k] = \frac{1}{n(n+2)\cdots(n+2k-2)} \sum_{\mu \vdash k} \frac{2^k k!}{2^{\ell(\mu)} z_\mu} E\left[\prod_{i=1}^{\ell(\mu)} \mathrm{Tr}(W^{\mu_j})\right]. \tag{4.5.4}$$

In fact, Theorem 4.5.4 along with the following two facts gives the claim:

$$\sum_{\sigma \in M_{2k}} \mathrm{Wg}^{\mathrm{O}}(\sigma; n) = \frac{1}{n(n+2)\cdots(n+2k-2)};$$

and the number of pair partitions in M_{2k} *of coset-type* μ *is* $2^k k!/(2^{\ell(\mu)} z_\mu)$. *When* $k = 2, 3$, (4.5.4) *gives*

$$E[w_{ii}^2] = \frac{1}{n(n+2)}(2E[\mathrm{Tr}(W^2)] + E[\mathrm{Tr}(W)^2]);$$
$$E[w_{ii}^3] = \frac{1}{n(n+2)(n+4)}(8E[\mathrm{Tr}(W^3)] + 6E[\mathrm{Tr}(W^2)\,\mathrm{Tr}(W)] + E[\mathrm{Tr}(W)^3]).$$

4.5.3 Left-Right Invariance (Unitary Case)

Theorem 4.5.6 *Let* X *be a complex* $n \times p$ *random matrix which has the same distribution as* UXV *for any unitary matrices* U, V. *For four sequences* $\mathbf{i} = (i_1, \dots, i_k)$, $\mathbf{j} = (j_1, \dots, j_k)$, $\mathbf{i}' = (i'_1, \dots, i'_k)$, $\mathbf{j}' = (j'_1, \dots, j'_k)$,

$$\begin{aligned} &E[x_{i_1 j_1} \cdots x_{i_k j_k} \overline{x_{i'_1 j'_1} \cdots x_{i'_k j'_k}}] \\ &= \sum_{\sigma_1, \sigma_2, \pi \in S_k} \delta_{\sigma_1}(\mathbf{i}, \mathbf{i}') \delta_{\sigma_2}(\mathbf{j}, \mathbf{j}') \mathrm{Wg}^{\mathrm{U}}(\pi \sigma_1^{-1} \sigma_2; n, p) E[\mathrm{Tr}_\pi(XX^*)], \end{aligned}$$

where $\mathrm{Wg}^{\mathrm{U}}(\cdot; n, p)$ *is defined in* (4.3.2).

Proof: As in Lemma 4.5.2, we can see that X has the same distribution UDV^*, where U and V are Haar distributed $n \times n$ and $p \times p$ random unitary matrices, respectively, and D is an $n \times p$ diagonal matrix whose singular values have the same distribution as those of X. Moreover, D, U, V are independent.

Since each entry x_{ij} has the same distribution as $\sum_{r=1}^{\min(n,p)} u_{ir} d_r \overline{v_{jr}}$, it follows from the independence of U, D, and V that

$$\begin{aligned} &E[x_{i_1 j_1} \cdots x_{i_k j_k} \overline{x_{i'_1 j'_1} \cdots x_{i'_k j'_k}}] \\ = &\sum_{\mathbf{r} = (r_1, \dots, r_k)} \sum_{\mathbf{r}' = (r'_1, \dots, r'_k)} E[d_{r_1} \cdots d_{r_k} \overline{d_{r'_1} \cdots d_{r'_k}}] \\ &\qquad \times E[u_{i_1 r_1} \cdots u_{i_k r_k} \overline{u_{i'_1 r'_1} \cdots u_{i'_k r'_k}}] \times E[\overline{v_{j_1 r_1} \cdots v_{j_k r_k}} v_{j'_1 r'_1} \cdots v_{j'_k r'_k}]. \end{aligned}$$

Here r_s, r_s' run over $1, 2, \dots, \min(p, n)$. From the unitary Weingarten calculus (Proposition 4.3.1), we have

$$= \sum_{\sigma_1,\tau_1,\sigma_2,\tau_2 \in S_k} \delta_{\sigma_1}(\mathbf{i},\mathbf{i}')\delta_{\sigma_2}(\mathbf{j},\mathbf{j}')\mathrm{Wg}^{\mathrm{U}}(\sigma_1^{-1}\tau_1; n)\mathrm{Wg}^{\mathrm{U}}(\sigma_2^{-1}\tau_2; p)$$
$$\times \sum_{\mathbf{r}=(r_1,\dots,r_k)} \sum_{\mathbf{r}'=(r_1',\dots,r_k')} \delta_{\tau_1}(\mathbf{r},\mathbf{r}')\delta_{\tau_2}(\mathbf{r},\mathbf{r}')E[d_{r_1}\cdots d_{r_k}\overline{d_{r_1'}\cdots d_{r_k'}}]. \tag{4.5.5}$$

Since $\delta_{\tau_1}(\mathbf{r},\mathbf{r}')\delta_{\tau_2}(\mathbf{r},\mathbf{r}') = 1$ if and only if $r_s' = r_{\tau_2(s)}$ $(1 \le s \le k)$ and $\delta_{\tau_1^{-1}\tau_2}(\mathbf{r},\mathbf{r}) = 1$, we have

$$\sum_{\mathbf{r}=(r_1,\dots,r_k)} \sum_{\mathbf{r}'=(r_1',\dots,r_k')} \delta_{\tau_1}(\mathbf{r},\mathbf{r}')\delta_{\tau_2}(\mathbf{r},\mathbf{r}')d_{r_1}\cdots d_{r_k}\overline{d_{r_1'}\cdots d_{r_k'}}$$
$$= \sum_{\mathbf{r}=(r_1,\dots,r_k)} \delta_{\tau_1^{-1}\tau_2}(\mathbf{r},\mathbf{r})d_{r_1}\cdots d_{r_k}\overline{d_{r_1}\cdots d_{r_k}},$$

which equals $\mathrm{Tr}_{\tau_1^{-1}\tau_2}(DD^*)$ by (4.5.1). Substituting this fact into (4.5.5), we have

$$E[x_{i_1j_1}\cdots x_{i_kj_k}\overline{x_{i_1'j_1'}\cdots x_{i_k'j_k'}}] = \sum_{\sigma_1,\sigma_2 \in S_k} \delta_{\sigma_1}(\mathbf{i},\mathbf{i}')\delta_{\sigma_2}(\mathbf{j},\mathbf{j}')$$
$$\times \sum_{\tau_1,\tau_2 \in S_k} \mathrm{Wg}^{\mathrm{U}}(\sigma_1^{-1}\tau_1; n)\mathrm{Wg}^{\mathrm{U}}(\sigma_2^{-1}\tau_2; p)E[\mathrm{Tr}_{\tau_1^{-1}\tau_2}(XX^*)].$$

The proof of the theorem follows from the following observation.

$$\sum_{\tau_1,\tau_2 \in S_k} \mathrm{Wg}^{\mathrm{U}}(\sigma_1^{-1}\tau_1; n)\mathrm{Wg}^{\mathrm{U}}(\sigma_2^{-1}\tau_2; p)E(\mathrm{Tr}_{\tau_1^{-1}\tau_2}(XX^*))$$
$$= \sum_{\tau_2,\pi \in S_k} \mathrm{Wg}^{\mathrm{U}}(\sigma_1^{-1}\tau_2\pi; n)\mathrm{Wg}^{\mathrm{U}}(\sigma_2^{-1}\tau_2; p)E(\mathrm{Tr}_{\pi^{-1}}(XX^*)) \qquad (\because \tau_1 = \tau_2\pi)$$
$$= \sum_{\tau_2,\pi \in S_k} \mathrm{Wg}^{\mathrm{U}}(\pi\sigma_1^{-1}\tau_2; n)\mathrm{Wg}^{\mathrm{U}}(\tau_2^{-1}\sigma_2; p)E(\mathrm{Tr}_{\pi^{-1}}(XX^*)) \qquad (\because \mathrm{Wg}^{\mathrm{U}}(\sigma; z) = \mathrm{Wg}^{\mathrm{U}}(\sigma^{-1}; z))$$
$$= \sum_{\pi \in S_k} \mathrm{Wg}^{\mathrm{U}}(\pi\sigma_1^{-1}\sigma_2; n, p)E[\mathrm{Tr}_{\pi}(XX^*)].$$

At the last equality we have used the definition of $\mathrm{Wg}^{\mathrm{U}}(\cdot; n, p)$. ∎

Example 4.5.7 *If X satisfies the condition of Theorem 4.5.6, we have*

$$E[x_{ij}\overline{x_{i'j'}}] = \delta_{ii'}\delta_{jj'}\frac{1}{np}E[\mathrm{Tr}(XX^*)].$$

4.5.4 Left-Right Invariance (Orthogonal Case)

Theorem 4.5.8 *Let X be a real $n \times p$ random matrix which has the same distribution as UXV for any orthogonal matrices U, V. For two sequences $\mathbf{i} = (i_1, \dots, i_{2k})$ and $\mathbf{j} = (j_1, \dots, j_{2k})$,*

$$E[x_{i_1j_1} \cdots x_{i_{2k}j_{2k}}] = \sum_{\sigma_1,\sigma_2,\pi \in M_{2k}} \delta'_{\sigma_1}(\mathbf{i})\delta'_{\sigma_2}(\mathbf{j})\mathrm{Wg}^{\mathrm{O}}(\pi\sigma_1^{-1}\sigma_2; n, p)E[\mathrm{Tr}_\pi(XX^t)],$$

where $\mathrm{Wg}^{\mathrm{O}}(\cdot; n, p)$ is defined in (4.3.4).

Proof: In a similar way to the proof of Theorem 4.5.6, we have

$$\begin{aligned} &E[x_{i_1j_1} \cdots x_{i_{2k}j_{2k}}] \\ =& \sum_{\sigma_1,\sigma_2,\tau_1,\tau_2 \in M_{2k}} \delta'_{\sigma_1}(\mathbf{i})\delta'_{\sigma_2}(\mathbf{j})\mathrm{Wg}^{\mathrm{O}}(\sigma_1^{-1}\tau_1; n)\mathrm{Wg}^{\mathrm{O}}(\sigma_2^{-1}\tau_2; p) \\ &\times \sum_{\mathbf{r}=(r_1,\dots,r_{2k})} \delta'_{\tau_1}(\mathbf{r})\delta'_{\tau_2}(\mathbf{r})E[d_{r_1} \cdots d_{r_{2k}}]. \end{aligned}$$

We observe that $\delta'_{\tau_1}(\mathbf{r})\delta'_{\tau_2}(\mathbf{r}) = 1$ if and only if all r_i in each component of $\Gamma(\tau_1^{-1}\tau_2)$ coincide. Letting $(\mu_1, \dots, \mu_l)$ to be a coset-type of $\tau_1^{-1}\tau_2$ we have

$$\sum_{\mathbf{r}=(r_1,\dots,r_{2k})} \delta'_{\tau_1}(\mathbf{r})\delta'_{\tau_2}(\mathbf{r})d_{r_1} \cdots d_{r_{2k}} = \sum_{s_1,\dots,s_l} d_{s_1}^{2\mu_1} \cdots d_{s_l}^{2\mu_l} = \mathrm{Tr}'_{\tau_1^{-1}\tau_2}(DD^t) = \mathrm{Tr}'_{\tau_1^{-1}\tau_2}(XX^t).$$

We thus have proved

$$\begin{aligned} &E[x_{i_1j_1} \cdots x_{i_{2k}j_{2k}}] \\ =& \sum_{\sigma_1,\sigma_2,\tau_1,\tau_2 \in M_{2k}} \delta'_{\sigma_1}(\mathbf{i})\delta'_{\sigma_2}(\mathbf{j})\mathrm{Wg}^{\mathrm{O}}(\sigma_1^{-1}\tau_1; n)\mathrm{Wg}^{\mathrm{O}}(\sigma_2^{-1}\tau_2; p)E(\mathrm{Tr}'_{\tau_1^{-1}\tau_2}(XX^t)). \end{aligned}$$

The remaining step is shown in a similar way to the proof of Theorem 4.5.6. (Replace a sum $\sum_{\sigma \in M_{2k}}$ by $(2^k k!)^{-1} \sum_{\sigma \in S_{2k}}$.) ∎

Example 4.5.9 *If X satisfies the condition of Theorem 4.5.8, we have*

$$E[x_{i_1j_1}x_{i_2j_2}] = \delta_{i_1i_2}\delta_{j_1j_2}\frac{1}{np}E[\mathrm{Tr}(XX^t)].$$

4.6 Application to Statistics

4.6.1 Pseudo-Inverse of a Ginibre Matrix (Complex Case)

Definition 4.6.1 *An $n \times p$* complex Ginibre matrix *G is a random matrix whose columns are i.i.d. and distributed as n-dimensional normal distribution $\mathrm{N}_{\mathbb{C}}(0, \Sigma)$, where Σ is an $n \times n$ positive definite Hermitian matrix.*

If $G = UDV^*$ is a singular value decomposition of G, the matrix $G^- = VD^-U^*$ is the pseudo-inverse of G, where D^- is the $p \times n$ diagonal obtained by inverting pointwise the entries of D along the diagonal (and zero if the diagonal entry is zero).

Note that it is easy to check that the pseudo-inverse is well-defined in the sense that it does not depend on the decomposition $G = UDV^*$. Actually, in the same vein as in Section 4.3 where the pseudo-inverse is introduced in the context of Weingarten functions, the properties $GG^-G = G$, $G^-GG^- = G^-$ together with the fact that GG^-, G^-G are selfadjoint, suffice to define the inverse uniquely. If the matrix is invertible, the pseudo-inverse is the inverse (this notion of pseudo-inverse is sometimes known as the Moore-Penrose pseudo inverse).

Theorem 4.6.2 *Let $G^- = (g^{ij})$ be the pseudo-inverse matrix of an $n \times p$ complex Ginibre matrix associated with an $n \times n$ positive definite Hermitian matrix Σ. Put $q = p - n$ and suppose $n, q \geq k$. For four sequences $\mathbf{i} = (i_1, \ldots, i_k)$, $\mathbf{j} = (j_1, \ldots, j_k)$, $\mathbf{i}' = (i'_1, \ldots, i'_k)$, and $\mathbf{j}' = (j'_1, \ldots, j'_k)$, we have*

$$\begin{aligned} &E[g^{i_1j_1} \cdots g^{i_kj_k}\overline{g^{i'_1j'_1} \cdots g^{i'_kj'_k}}] \\ =&(-1)^k \sum_{\sigma,\rho \in S_k} \delta_\sigma(\mathbf{i}, \mathbf{i}')\mathrm{Wg}^{\mathrm{U}}(\sigma^{-1}\rho; p, -q)\overline{(\Sigma^{-1})_{j_{\rho(1)}j'_1} \cdots (\Sigma^{-1})_{j_{\rho(k)}j'_k}}, \end{aligned}$$

where $\mathrm{Wg}^{\mathrm{U}}(\cdot; p, -q)$ *is defined in* (4.3.2).

Proof: Let Z be an $n \times p$ matrix of i.i.d. $\mathrm{N}_{\mathbb{C}}(0,1)$ random variables. Then it is immediate to see that $\Sigma^{1/2} Z$ has the same distribution as G. Therefore each g^{ij} has the same distribution as $\sum_{r=1}^{n} z^{ir} (\Sigma^{-1/2})_{rj}$, where $Z^{-} = (z^{ij})_{1 \le i \le p, 1 \le j \le n}$ is the pseudo-inverse matrix of Z, and hence

$$E[g^{i_1 j_1} \cdots g^{i_k j_k} \overline{g^{i'_1 j'_1} \cdots g^{i'_k j'_k}}] = \sum_{\mathbf{r}=(r_1,\dots,r_k)} \sum_{\mathbf{r}'=(r'_1,\dots,r'_k)} \prod_{s=1}^{k} (\Sigma^{-1/2})_{r_s j_s} \overline{(\Sigma^{-1/2})_{r'_s j'_s}} \times E[z^{i_1 r_1} \cdots z^{i_k r_k} \overline{z^{i'_1 r'_1} \cdots z^{i'_k r'_k}}].$$

Since Z^{-} is a $p \times n$ matrix satisfying the condition on Theorem 4.5.6, we have

$$\begin{aligned} & E[z^{i_1 r_1} \cdots z^{i_k r_k} \overline{z^{i'_1 r'_1} \cdots z^{i'_k r'_k}}] \\ = & \sum_{\sigma, \rho, \pi \in S_k} \delta_{\sigma}(\mathbf{i}, \mathbf{i}') \delta_{\rho}(\mathbf{r}, \mathbf{r}') \mathrm{Wg}^{\mathrm{U}}(\sigma^{-1} \pi \rho; p, n) E[\mathrm{Tr}_{\pi}(Z^{-}(Z^{-})^{*})]. \end{aligned}$$

Moreover, from the condition of $q = p - n \ge k$, we can apply Proposition 4.4.2 with $W = ZZ^{*}$, and

$$E[\mathrm{Tr}_{\pi}(Z^{-}(Z^{-})^{*})] = E[\mathrm{Tr}_{\pi}(W^{-1})] = (-1)^k \sum_{\tau \in S_k} \mathrm{Wg}^{\mathrm{U}}(\pi \tau^{-1}; -q) \, \mathrm{Tr}_{\tau}(I_n),$$

where I_n is the $n \times n$ identity matrix. Note that $\mathrm{Tr}_{\tau}(I_n) = n^{\kappa(\tau)}$. Hence we have obtained

$$\begin{aligned} & E[g^{i_1 j_1} \cdots g^{i_k j_k} \overline{g^{i'_1 j'_1} \cdots g^{i'_k j'_k}}] \\ = & (-1)^k \sum_{\sigma, \rho, \pi, \tau \in S_k} \delta_{\sigma}(\mathbf{i}, \mathbf{i}') n^{\kappa(\tau)} \mathrm{Wg}^{\mathrm{U}}(\sigma^{-1} \pi \rho; p, n) \mathrm{Wg}^{\mathrm{U}}(\pi^{-1} \tau; -q) \\ & \times \sum_{\mathbf{r}=(r_1,\dots,r_k)} \sum_{\mathbf{r}'=(r'_1,\dots,r'_k)} \delta_{\rho}(\mathbf{r}, \mathbf{r}') \prod_{s=1}^{k} (\Sigma^{-1/2})_{r_s j_s} \overline{(\Sigma^{-1/2})_{r'_s j'_s}}. \end{aligned}$$

A direct calculation gives

$$\sum_{\pi, \tau \in S_k} n^{\kappa(\tau)} \mathrm{Wg}^{\mathrm{U}}(\sigma^{-1} \pi \rho; p, n) \mathrm{Wg}^{\mathrm{U}}(\pi^{-1} \tau; -q)$$

$$= \sum_{\pi,\tau \in S_k} \mathrm{Wg}^{\mathrm{U}}(\rho\sigma^{-1}\pi; p, n)\mathrm{Wg}^{\mathrm{U}}(\pi^{-1}\tau; -q) n^{\kappa(\tau^{-1})}$$
$$= [\mathrm{Wg}^{\mathrm{U}}(\cdot; p) * \mathrm{Wg}^{\mathrm{U}}(\cdot; n) * \mathrm{Wg}^{\mathrm{U}}(\cdot; -q) * n^{\kappa(\cdot)}](\rho\sigma^{-1}).$$

Since $n^{\kappa(\cdot)} * \mathrm{Wg}^{\mathrm{U}}(\cdot; n) = \delta_e$ when $n \geq k$, we have

$$\sum_{\pi,\tau \in S_k} n^{\kappa(\tau)} \mathrm{Wg}^{\mathrm{U}}(\sigma^{-1}\pi\rho; p, n)\mathrm{Wg}^{\mathrm{U}}(\pi^{-1}\tau; -q) = \mathrm{Wg}^{\mathrm{U}}(\sigma^{-1}\rho; p, -q).$$

On the other hand, it is easy to see that

$$\sum_{\mathbf{r}=(r_1,\dots,r_k)} \sum_{\mathbf{r}'=(r'_1,\dots,r'_k)} \delta_\rho(\mathbf{r}, \mathbf{r}') \prod_{s=1}^{k} (\Sigma^{-1/2})_{r_s j_s} \overline{(\Sigma^{-1/2})_{r'_s j'_s}}$$
$$= \sum_{r_1,\dots,r_k} \prod_{s=1}^{k} \overline{(\Sigma^{-1/2})_{j_s r_s} (\Sigma^{-1/2})_{r_{\rho(s)} j'_s}}$$
$$= \sum_{r_1,\dots,r_k} \prod_{s=1}^{k} \overline{(\Sigma^{-1/2})_{j_{\rho(s)} r_{\rho(s)}} (\Sigma^{-1/2})_{r_{\rho(s)} j'_s}}$$
$$= \prod_{s=1}^{k} \sum_{r} \overline{(\Sigma^{-1/2})_{j_{\rho(s)} r} (\Sigma^{-1/2})_{r j'_s}}$$
$$= \prod_{s=1}^{k} \overline{(\Sigma^{-1})_{j_{\rho(s)} j'_s}}.$$

We thus have completed the proof of the theorem. ∎

Example 4.6.3 *For G given as in Theorem 4.6.2,*

$$E[g^{ij}\overline{g^{i'j'}}] = \delta_{i,i'} \frac{1}{p(p-n)} \overline{(\Sigma^{-1})_{jj'}}.$$

4.6.2 Pseudo-Inverse of a Ginibre Matrix (Real Case)

Definition 4.6.4 *An $n \times p$* real Ginibre matrix *G is a random matrix whose columns are i.i.d. and distributed as n-dimensional normal distribution $\mathrm{N}_{\mathbb{R}}(0, \Sigma)$, where Σ is an $n \times n$ positive definite real symmetric matrix.*

Theorem 4.6.5 *Let $G^- = (g^{ij})$ be the pseudo-inverse matrix of an $n \times p$ real Ginibre matrix associated with an $n \times n$ positive definite real symmetric matrix Σ. Put $q = p - n - 1$ and suppose $n \geq k$ and $q \geq 2k-1$. For two sequences $\mathbf{i} = (i_1, \dots, i_{2k})$ and $\mathbf{j} = (j_1, \dots, j_{2k})$, we have*

$$E[g^{i_1 j_1} g^{i_2 j_2} \cdots g^{i_{2k} j_{2k}}] = (-1)^k \sum_{\sigma, \rho \in M_{2k}} \delta'_\sigma(\mathbf{i}) \mathrm{Wg}^{\mathrm{O}}(\sigma^{-1}\rho; p, -q) \prod_{\{a,b\} \in \rho} (\Sigma^{-1})_{j_a j_b},$$

where $\mathrm{Wg}^{\mathrm{O}}(\cdot; p, -q)$ is defined in (4.3.4).

Proof: The proof is similar to that of the complex case if we use Theorem 4.5.8, Proposition 4.4.4, and the following identity: for each $\sigma \in M_{2k}$,

$$\sum_{\mathbf{r}=(r_1,\dots,r_{2k})} \delta'_\sigma(\mathbf{r}) \prod_{s=1}^{2k} (\Sigma^{-1/2})_{r_s j_s} = \prod_{\{a,b\} \in \sigma} (\Sigma^{-1})_{j_a j_b}, \tag{4.6.1}$$

which is verified easily. ∎

Remark 4.6.6 *For $\sigma = \{\{1,2\},\{3,4\}\} \in M_4$,*

$$\prod_{\{a,b\} \in \sigma} (\Sigma^{-1})_{j_a j_b} = (\Sigma^{-1})_{j_1 j_2} (\Sigma^{-1})_{j_3 j_4}.$$

Example 4.6.7 *For G given as in Theorem 4.6.5,*

$$E[g^{i_1 j_1} g^{i_2 j_2}] = \delta_{i_1, i_2} \frac{1}{p(p-n-1)} (\Sigma^{-1})_{j_1 j_2}.$$

4.6.3 Inverse of Compound Wishart Matrix (Complex Case)

Definition 4.6.8 *Let Σ be an $n \times n$ positive definite Hermitian matrix and let B be a $p \times p$ complex matrix. Let Z be an $n \times p$ matrix of i.i.d. $\mathrm{N}_{\mathbb{C}}(0,1)$ random variables. Then we call a matrix*

$$W = \Sigma^{1/2} Z B Z^* \Sigma^{1/2}$$

a complex compound Wishart matrix *with shape parameter B and scale parameter Σ, where $\Sigma^{1/2}$ is the hermitian root of Σ.*

Remark 4.6.9 *If $\Sigma = I_n$, then the corresponding compound Wishart matrix is called white (or standard) compound Wishart. If B is a positive-definite matrix, then the corresponding compound Wishart matrix can be considered as a sample covariance matrix under correlated sampling as explained in [BJJNPZ].*

Theorem 4.6.10 *Let Σ be an $n \times n$ positive definite Hermitian matrix and B be a $p \times p$ complex matrix.*

Let $W^{-1} = (w^{ij})$ be the inverse matrix of an $n \times n$ complex compound Wishart matrix with shape parameter B and scale parameter Σ. Put $q = p - n$ and suppose $n, q \geq k$. For two sequences $\mathbf{i} = (i_1, \ldots, i_k)$ and $\mathbf{j} = (j_1, \ldots, j_k)$, we have

$$E[w^{i_1 j_1} \cdots w^{i_k j_k}] = (-1)^k \sum_{\sigma,\rho \in S_k} \mathrm{Tr}_\sigma(B^-) \mathrm{Wg}^{\mathrm{U}}(\sigma^{-1}\rho; p, -q) \overline{(\Sigma^{-1})_{i_{\rho(1)} j_1} \cdots (\Sigma^{-1})_{i_{\rho(k)} j_k}}.$$

Proof: The matrix W has the same distribution as GBG^*, where G is an $n \times p$ Ginibre matrix associated with Σ. If we write $B^- = (b^{ij})$ and $G^- = (g^{ij})$, then

$$E[w^{i_1 j_1} \cdots w^{i_k j_k}] = \sum_{\mathbf{r}=(r_1,\ldots,r_k)} \sum_{\mathbf{r}=(r'_1,\ldots,r'_k)} b^{r_1 r'_1} \cdots b^{r_k r'_k} E[\overline{g^{r_1 i_1} \cdots g^{r_k i_k}} g^{r'_1 j_1} \cdots g^{r'_k j_k}].$$

Moreover, it follows from Theorem 4.6.2 that

$$\begin{aligned} E[w^{i_1 j_1} \cdots w^{i_k j_k}] =& (-1)^k \sum_{\sigma,\rho \in S_k} \mathrm{Wg}^{\mathrm{U}}(\sigma^{-1}\rho; p, -q) \overline{(\Sigma^{-1})_{i_{\rho(1)} j_1} \cdots (\Sigma^{-1})_{i_{\rho(k)} j_k}} \\ & \times \sum_{\mathbf{r}=(r_1,\ldots,r_k)} \sum_{\mathbf{r}=(r'_1,\ldots,r'_k)} \delta_\sigma(\mathbf{r}, \mathbf{r}') b^{r_1 r'_1} \cdots b^{r_k r'_k}. \end{aligned}$$

We finally observe that

$$\sum_{\mathbf{r}=(r_1,\ldots,r_k)} \sum_{\mathbf{r}=(r'_1,\ldots,r'_k)} \delta_\sigma(\mathbf{r}, \mathbf{r}') b^{r_1 r'_1} \cdots b^{r_k r'_k} = \sum_{\mathbf{r}=(r_1,\ldots,r_k)} b^{r_1 r_{\sigma(1)}} \cdots b^{r_k r_{\sigma(k)}} = \mathrm{Tr}_\sigma(B^-).$$

∎

Remark 4.6.11 *If $\Sigma = I_n$ (in the white compound Wishart case), one can observe that a simplification occurs in the above formula.*

In turn, this simplification has the following probabilistic explanation: the joint distribution of the traces of $W, W^2, \ldots$ has the same law as the joint distribution of the traces of $\tilde{W}, \tilde{W}^2, \ldots$, where $\tilde{W}$ is a non-compound Wishart distribution of parameter $B^{1/2}$.

Therefore we can use existing results for the inverse of non-compound Wishart matrices in order to work out this case.

4.6.4 Inverse of Compound Wishart Matrix (Real Case)

Definition 4.6.12 *Let Σ be an $n \times n$ positive definite symmetric matrix and let B be a $p \times p$ real matrix. Let Z be an $n \times p$ matrix of i.i.d. $\mathrm{N}_{\mathbb{R}}(0,1)$ random variables. Then we call a matrix*

$$W = \Sigma^{1/2} Z B Z^* \Sigma^{1/2}$$

a real compound Wishart matrix *with shape parameter B and scale parameter Σ, where $\Sigma^{1/2}$ is the symmetric root of Σ.*

Theorem 4.6.13 *Let Σ be an $n \times n$ positive definite real symmetric matrix and B a $p \times p$ real matrix. Let $W^{-1} = (w^{ij})$ be the inverse matrix of an $n \times n$ real compound Wishart matrix with shape parameter B and scale parameter Σ. Put $q = p - n - 1$ and suppose $n \geq k$ and $q \geq 2k - 1$. For any sequence $\mathbf{i} = (i_1, \ldots, i_{2k})$, we have*

$$E[w^{i_1 i_2} \cdots w^{i_{2k-1} i_{2k}}] = (-1)^k \sum_{\sigma, \rho \in M_{2k}} \mathrm{Tr}'_\sigma(B^-) \mathrm{Wg}^{\mathrm{O}}(\sigma^{-1}\rho; p, -q) \prod_{\{u,v\} \in \rho} (\Sigma^{-1})_{i_u i_v}.$$

Proof: The proof is similar to the complex case. ∎

Chapter 5

Random Matrix Theory and Noisy Empirical Covariance Matrices : Risk Underestimation

5.1 Introduction

Random matrix theory (RMT) may be used to improve the estimation of the risk of the optimal portfolio. As shown in Chapter 2, covariance matrices are the key input parameters to Markowitz's optimization problem. Computation of the risk and the weights of the optimal portfolio depends essentially on the entries of the inverse of the covariance matrix.

In practical situations in mathematical finance, the covariance matrix of the returns is unknown and we always deal with an estimator of it. To estimate the covariance matrix for the returns of n different assets, we need to determine $n(n+1)/2$ entries from n time series of length T. Throughout the chapter, n will denote the number of the assets of the portfolio and T will denote the number of observations.

If T is not very large compared to n, which is the common situation in real life, one should expect that the determination of the covariances is noisy. Results from RMT reinforce the doubts about the accuracy of empirical covariance matrices. In ([LCBP1]), Laloux et al. showed that the covariance matrices determined from empirical financial time series appear to have such a high amount of noise such that except for a few large eigenvalues and corresponding eigenvectors, its structure can be regarded as random. This result conflicts with the fundamental role played by the covariance matrices in computing the risk of the optimal portfolio. Hence, Laloux et al. [LCBP1] concluded that "Markowitz's portfolio optimization scheme based on a purely historical determination of the covariance matrix is inadequate".

In the two subsequent studies ([LCBP2], [PGRAS]), based on historical data, the authors optimized the portfolio by using the empirical covariance matrix of the first half of the sample and after calculating the *predicted risk* (the standard deviation of the returns of the optimal portfolio in the first half of the sample). They used the second half of the sample to compute the *realized risk* (the standard deviation of the returns of the second part of the sample) and the authors found a significant difference between the predicted risk and the realized risk.

Improving the estimation of the risk of the optimal portfolio was an essential aim for many scientists (see [PGRAGS], [BiBouP], [RGPS] , [PK], [El-K]). In [LCBP2], it was found that the risk level of an optimized portfolio could be improved if prior to optimization, one got rid of the lower part of the eigenvalue spectrum of the empirical covariance matrix which coincides with the eigenvalue spectrum of the "noisy" random matrix. This method is called "*Filtering*" technique and it will be discussed in detail in Section 5.2.

For the maximum likelihood estimator (MLE) of the covariance matrix of the MV simplified model, Pafka et al. [PK] observed that the effect of the noise induced by estimating the covariance matrix of the returns on computing the risk of the optimal portfolio strongly depends on the ratio n/T. On the basis of numerical experiments and analytic results for some toy portfolio models, they showed that for large values of the ratio $\frac{n}{T}$ (e.g. $\frac{n}{T} = 0.6$) noise does have a strong effect on estimating the risk while for small values $\frac{n}{T} = 0.2$, the error in computing the risk reduces to acceptable levels. Pafka et al. defined the asymptotic effect of the noise on the estimation of the risk of the optimal portfolio as the ratio $\frac{1}{\sqrt{1-\frac{n}{T}}}$.

In our work, we deal with a more general estimator of the covariance matrix for which the MLE covariance matrix is a special case. Our aim is to measure this effect of the noise induced by estimating the covariance matrix not only for the independent observations but also for the correlated observations. We use the techniques of the random matrix theory (RMT) to quantify the asymptotic effect of the noise resulting from estimating the covariance matrix on predicting the risk of the optimal portfolio. In the case of independent sampling, our results agree with those of Pafka et al. [PK] and El Karoui [El-K].

The chapter is divided into eight parts. In Section 5.2, we give an overview of the "*Filtering*" technique. In Section 5.3, we introduce the "Scaling" technique to improve the estimation of the optimal portfolio's risk. The technique depends on our result concerning the asymptotic behavior of the effect of the noise induced by estimating the covariance matrix on computing the risk of the optimal portfolio. Some examples and simulations of the "*Scaling*" technique will be discussed in Section 5.4. As an application, Section 5.5 will illustrates the impact of the noise induced by estimating the covariance matrix for the exponentially weighted moving average (EWMA) covariance estimator which is often used in finance. Then, we are going to make a comparison between the "*Filtering*" and the "*Scaling*" techniques in Section 5.6. In Section 5.7, we discuss the estimation of the optimal weights. Finally, there will be a conclusion in Section 5.8.

5.2 Filtering Technique

The *Filtering* technique is used to improve the estimation of the risk of the optimal portfolio. The technique is discussed in many works as shown in ([LCBP2], [BiBouP], [RGPS], [RGPS]). The technique depends on converting the estimated covariance matrix into a corresponding correlation matrix and then on cleaning the correlation matrix by removing the noisy eigenvalues (those eigenvalues falling in the region of the eigenvalues of a random matrix). After cleaning the correlation matrix, it is converted back to the corresponding covariance matrix and then the risk of the optimal portfolio can be computed.

The idea starts with a paper of Laloux et al. [LCBP1]. Using results from the theory of random matrices, Laloux et al. found a remarkable agreement between the theoretical assumption that the correlation matrix is random and the density of eigenvalues of the empirical correlation matrix. In the case of $S\&P500$, Laloux et al. showed that 94% of the total number of eigenvalues of the empirical correlation matrix fall in the same region as the histogram of eigenvalues of the White Wishart matrices given by the Marcenko and Pastur law.

For T observations of n assets, let $C = (c_{ij})_{i,j=1}^{n}$ be an $n \times n$ empirical correlation matrix. For the time series of price changes x_{ti} (where i labels the asset and t labels the time),

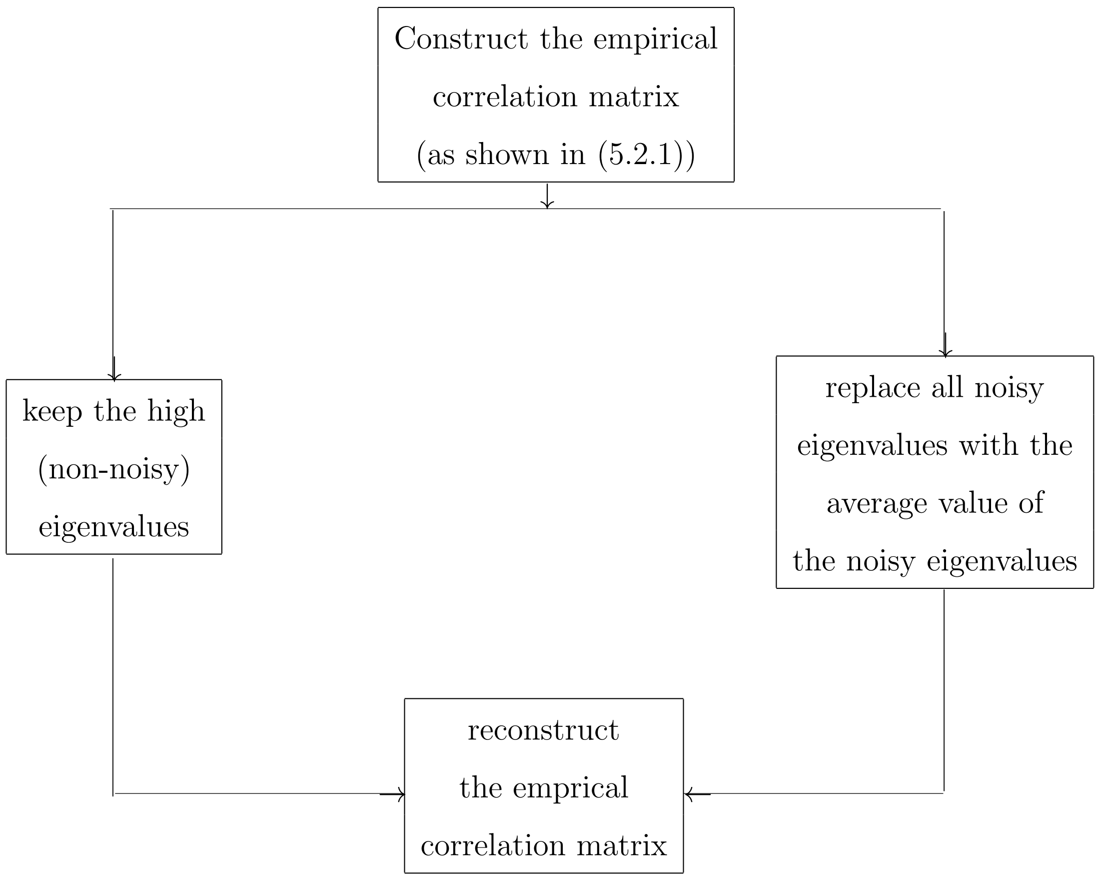
$$c_{ij} = \frac{1}{T} \sum_{t=1}^{T} x_{ti} x_{tj}. \tag{5.2.1}$$

Remark 5.2.1 *It is assumed that the average value of the x's has been subtracted off and also that the x's are scaled to have a constant unit volatility.*

The Filtering technique is used to reduce the noise in the empirical covariance matrices by eliminating the small eigenvalues of the empirical covariance matrix which match with those eigenvalues of the random matrix by using Marçenko-Pastur law. In ([LCBP2], [RGPS]), depending on historical data, the authors showed that the

Filtering technique reduces the error in computing the risk of the optimal portfolio.

For the $n \times n$ empirical correlation matrix C, the *Filtering* technique is described as shown in the following diagram:



5.3 Scaling Technique

In the Scaling technique, we deal directly with the covariance matrices (we don't need to find the correlation matrix as in the Filtering technique). We will consider a more general estimator of the covariance matrix which describes the correlated sampling case as well as the independent sampling one. Let $\widehat{\Sigma}$ be the estimator of the covariance matrix Σ such that

$$\widehat{\Sigma} = \frac{1}{\mathrm{Tr}(B)} Y^t B Y \tag{5.3.1}$$

where $Y = (y_{ij})$ is a $T \times n$ matrix whose rows are $n-$dimensional vectors of centered returns which are taken sequentially in time: $Y_1, Y_2, \ldots, Y_T$. We assume that these vectors are i.i.d. with distribution $N(0, \Sigma)$. So that y_{ij} is the return of the j^{th} asset at time i. Hence $Y \sim N(\mathbf{0}, I_T \otimes \Sigma)$ where $\otimes$ denotes the Kronecker product of matrices and B is a $T \times T$ known weighting matrix.

Remark 5.3.1 *Note that for $B = I_T$, the $T \times T$ identity matrix, $\widehat{\Sigma}$ is the maximum likelihood estimator (MLE) of the covariance matrix. If $B = (b_{ij})_{i,j=1}^n$ is a diagonal matrix such that $b_{ii} = \lambda^{i-1}$ for some $0 < \lambda < 1$ and for $i = 1, \ldots, n$ then, $\widehat{\Sigma}$ is the exponentially weighted moving average (EWMA) estimator of the covariance matrix which will be studied in detail in Section 5.5.*

Since Y has the same distribution as $X\Sigma^{\frac{1}{2}}$ where, X is a $T \times n$ matrix with $i.i.d$ standard normal entries and $\Sigma^{\frac{1}{2}}$ is the symmetric root of Σ. We write

$$Y \stackrel{\mathcal{L}}{=} X\Sigma^{\frac{1}{2}}. \tag{5.3.2}$$

From (5.3.1) and (5.3.2),

$$\widehat{\Sigma} \stackrel{\mathcal{L}}{=} \frac{1}{\mathrm{Tr}(B)} \Sigma^{\frac{1}{2}} X^t B X \Sigma^{\frac{1}{2}}. \tag{5.3.3}$$

From (5.3.3) and since the matrix X is left-right orthogonally invariant then, $\widehat{\Sigma}$ is a compound Wishart matrix with scale parameter Σ and shape parameter B.

Remark 5.3.2 *Since the matrix X in (5.3.3) has a left-right orthogonally invariant distribution then, the distribution of the estimator $\widehat{\Sigma}$ in (5.3.3) depends only on the eigenvalues of the matrix B. So, the shape parameter of the estimator $\widehat{\Sigma}$ can be defined as the matrix Λ_B where Λ_B is the diagonal matrix that is similar to the matrix B.*

Since we deal with an estimator of the covariance matrix instead of Σ itself then, for a portfolio with n assets and time series of financial observations of the returns of length T, we can define two kinds of risks; one using Σ and we will call it the *True* risk, where

$$\text{True risk} = \sqrt{\mathbf{w}^t \Sigma \mathbf{w}}, \tag{5.3.4}$$

with $\mathbf{w}$ denoting the vector of the optimal weights determined by using the entries of Σ^{-1} as shown in Lemma 2.5.2. The other kind of risk depends on $\widehat{\Sigma}$ and is called the *Predicted* risk, where

$$\text{Predicted risk} = \sqrt{\hat{\mathbf{w}}^t \widehat{\Sigma} \hat{\mathbf{w}}}, \tag{5.3.5}$$

with $\hat{\mathbf{w}}$ denoting the vector of the optimal weights determined by using the entries of $\widehat{\Sigma}^{-1}$.

Remark 5.3.3 *Note that, in practice, only the* Predicted *risk can be computed while the* True *risk is unknown.*

Let

$$Q = \frac{(True\ risk)^2}{(Predicted\ risk)^2} \tag{5.3.6}$$

Our goal is to have the ratio Q in (5.3.6) as close as possible to one. By Corollary 2.6.1, we can write

$$Q = \frac{\sum_{i,j=1}^{n} \hat{\sigma}_{ij}^{(-1)}}{\sum_{i,j=1}^{n} \sigma_{ij}^{(-1)}} \tag{5.3.7}$$

Clearly, this ratio is close to one as the sample size T tends to infinity while n remains fixed. By using Theorem 4.6.13, we can also consider cases where T and n tend to infinity and $T > n + 3$.

We aim to derive a deterministic bias factor which can be used to correct the above predicted risk. To do that, we need to prove an interesting property of the inverted compound Wishart matrices. In which, we show that for a compound

Wishart matrix W with a scale parameter Σ and a shape parameter B (we write $W \in \mathcal{W}(\Sigma, B)$), the ratio between the expected trace of W^{-1} and the expected sum of its entries equals to the ratio between the trace of Σ^{-1} and the sum of its entries:

Proposition 5.3.4 *For an $n \times n$ matrix $W \in \mathcal{W}(\Sigma, B)$,*

$$E(\mathrm{Tr}(W^{-1}))/E(\sum_{i,j=1}^{n} w_{ij}^{(-1)}) = \mathrm{Tr}(\Sigma^{-1})/\sum_{i,j=1}^{n} \sigma_{ij}^{(-1)}.$$

Before we prove this proposition we need to recall the following well-known fact:

Lemma 5.3.5 *Let M be an $n \times n$ orthogonally invariant matrix (for any $n \times n$ orthogonal matrix O, M and OMO^t have the same distribution). Then*
$(i) E(M) = \alpha I_n$, where α is some scalar and I_n is the $n \times n$ identity matrix.
$(ii) M^k$ is orthogonally invariant, for each $k \in \mathbb{Z}$.

Proof: Let Z be a $T \times n$ matrix of i.i.d. entries which are normally distributed with zero mean and unit variance i.e.,

$$Z = (z_{ij}) \quad (i = 1, \ldots, T;\ j = 1, \ldots, n) \quad \text{such that} \quad z_{ij} \sim N(0, 1). \tag{5.3.8}$$

Consider

$$A = Z^t B Z. \tag{5.3.9}$$

Then, A is orthogonally invariant. By Lemma 5.3.5 (ii) taking $k = -1$, A^{-1} is orthogonally invariant as well and

$$E(A^{-1}) = \alpha I_n, \tag{5.3.10}$$

for some scalar α. Another important remark is that,

$$\begin{aligned} E(\sum_{i,j=1}^{n} w_{ij}^{(-1)}) &= E(\mathrm{Tr}(\mathbf{e}^t W^{-1} \mathbf{e})) \\ &= \mathrm{Tr}(E(\mathbf{e}^t W^{-1} \mathbf{e})). \end{aligned} \tag{5.3.11}$$

Since $W \in \mathcal{W}(\Sigma, B)$ then, $W^{-1} \stackrel{\mathcal{L}}{=} \Sigma^{-\frac{1}{2}} A^{-1} \Sigma^{-\frac{1}{2}}$ and so,

$$E(\sum_{i,j=1}^{n} w_{ij}^{(-1)}) = E(\mathrm{Tr}(\mathbf{e}^t \Sigma^{-\frac{1}{2}} A^{-1} \Sigma^{-\frac{1}{2}} \mathbf{e}))$$

Since Tr is invariant under cyclic permutations then,

$$\begin{aligned} E(\sum_{i,j=1}^{n} w_{ij}^{(-1)}) &= E(\mathrm{Tr}(\Sigma^{-\frac{1}{2}} \mathbf{e}\mathbf{e}^t \Sigma^{-\frac{1}{2}} A^{-1})) \\ &= \mathrm{Tr}(\Sigma^{-\frac{1}{2}} \mathbf{e}\mathbf{e}^{\mathbf{t}} \Sigma^{-\frac{1}{2}} E(A^{-1})). \end{aligned}$$

So,

$$\begin{aligned} \mathrm{Tr}(\Sigma^{-1}) E(\sum_{i,j=1}^{n} w_{ij}^{(-1)}) &= \mathrm{Tr}(\Sigma^{-1})\,\mathrm{Tr}(\Sigma^{-\frac{1}{2}} \mathbf{e}\mathbf{e}^{\mathbf{t}} \Sigma^{-\frac{1}{2}} E(A^{-1})) \\ &= \mathrm{Tr}(\alpha \Sigma^{-1})\,\mathrm{Tr}(\mathbf{e}^{\mathbf{t}} \Sigma^{-1} \mathbf{e}) \quad \text{(from (5.3.10))} \\ &= \mathrm{Tr}(E(A^{-1})\Sigma^{-1}) \sum_{i,j=1}^{n} \sigma_{ij}^{(-1)} \\ &= \mathrm{Tr}(E(A^{-1}\Sigma^{-1})) \sum_{i,j=1}^{n} \sigma_{ij}^{(-1)} \\ &= E(\mathrm{Tr}(W^{-1})) \sum_{i,j=1}^{n} \sigma_{ij}^{(-1)}. \end{aligned}$$

∎

Remark 5.3.6 *Note that the $T \times T$ matrix B depends essentially on the dimension T. So, from now on we will denote B by B_T.*

In the following theorem, we study the asymptotic behavior of the ratio Q which will play a great role in improving the prediction of the risk of the optimal portfolio.

Theorem 5.3.7 *Let B_T be a $T \times T$ real matrix such that*

$$\lim_{T\to\infty} \frac{1}{T} (\mathrm{tr}(B_T))^2\,\mathrm{tr}(B_T^{-2}) = 0 \tag{5.3.12}$$

where tr *denotes the normalized trace of the matrix i.e., for an $n \times n$ matrix S,* $\mathrm{tr}(S) = \frac{1}{n}\,\mathrm{Tr}(S)$*. Let $\widehat{\Sigma}$ be as defined in (5.3.1). If $T > n+3$, then as n and T tend to infinity such that $n/T \to r < 1$ we have*

$$Q - \mathrm{Tr}(B_T)E(\mathrm{tr}((X^t B_T X)^{-1})) \xrightarrow{\text{P}} 0 \tag{5.3.13}$$

where X is a $T \times n$ matrix of i.i.d. standard normal entries.

Remark 5.3.8 *The condition $T > n+3$ is needed to compute the second moment of the inverse of a compound Wishart matrix (by Theorem 4.6.13) and get a formula for the variance of the difference $Q - \mathrm{Tr}(B_T)E(\mathrm{tr}((X^t B_T X)^{-1}))$.*

To prove Theorem 5.3.7, we need first to consider the following result concerning the variance of the ratio Q.

Proposition 5.3.9 *Let B_T be a $T \times T$ real matrix and let $\widehat{\Sigma}$ be as defined in (5.3.1). If $q = T - n - 1$, then for $q > 2$,*

$$Var(Q) = \frac{(\mathrm{Tr}(B_T))^2}{T^2(T+2)(T-1)q^2(q-2)(q+1)}\left(A_1(\mathrm{Tr}(B_T^{-1}))^2 + A_2\,\mathrm{Tr}(B_T^{-2})\right) \tag{5.3.14}$$

where

$$A_1 = 2T^2q - 2Tq^2 + 2T^2 + 2T + 2q^2 - 2q - 4$$

and

$$A_2 = Tq(2T - 2q + 2Tq - 2).$$

Proof:

$$\begin{aligned} Var(Q) &= \frac{1}{(\sum_{i,j=1}^{n} \sigma_{ij}^{(-1)})^2}\left(E((\sum_{i,j=1}^{n} \hat{\sigma}_{ij}^{(-1)})^2) - (E(\sum_{i,j=1}^{n} \hat{\sigma}_{ij}^{(-1)}))^2\right) \\ &= \frac{1}{(\sum_{i,j=1}^{n} \sigma_{ij}^{(-1)})^2}\left(\sum_{i_1,i_2,i_3,i_4=1}^{n} E(\hat{\sigma}_{i_1 i_2}^{(-1)}\hat{\sigma}_{i_3 i_4}^{(-1)}) - \left(\sum_{i,j=1}^{n} E(\hat{\sigma}_{ij}^{(-1)})\right)^2\right) \end{aligned}$$

Substitute from (5.3.1) to get

$$Var(Q) = \frac{(\mathrm{Tr}(B_T))^2}{(\sum\limits_{i,j=1}^{n} \sigma_{ij}^{(-1)})^2} \left(E(\sum_{i_1,i_2,i_3,i_4=1}^{n} w_{i_1 i_2}^{(-1)} w_{i_3 i_4}^{(-1)}) - (E(\sum_{i,j=1}^{n} w_{ij}^{(-1)}))^2 \right) \qquad (5.3.15)$$

where $W = (w_{ij})$ is an $n \times n$ compound Wishart matrix with scale parameter Σ and shape parameter B. By applying Theorem 4.6.13, we get

$$\begin{aligned} E(w_{ij}^{(-1)}) &= (-1) \sum_{\sigma,\rho \in M_2} \mathrm{Tr}_\sigma(B_T^{-1}) \mathrm{Wg}^{\mathrm{O}}(\sigma^{-1}\rho; T, -q) \prod_{\{u,v\} \in \rho} \sigma_{i_u i_v}^{(-1)} \\ &= (-1)\, \mathrm{Tr}(B_T^{-1}) \mathrm{Wg}^{\mathrm{O}}(\{1,2\}; T, -q) \sigma_{ij}^{(-1)} \end{aligned}$$

where $q = T - n - 1 \geq 1$. By using the values of Wg in [CM], we get

$$E(w_{ij}^{(-1)}) = \frac{1}{Tq} \mathrm{Tr}(B_T^{-1}) \sigma_{ij}^{(-1)}. \qquad (5.3.16)$$

By applying Theorem 4.6.13 again, then for $q \geq 3$ we get

$$E(w_{i_1 i_2}^{(-1)} w_{i_3 i_4}^{(-1)}) = \sum_{\rho \in M_4} \Big((\mathrm{Tr}(B_T^{-1}))^2 \mathrm{Wg}^{\mathrm{O}}(\rho; T, -q) + \mathrm{Tr}(B_T^{-2}) [\mathrm{Wg}^{\mathrm{O}}(\pi_1 \rho; T, -q) + \mathrm{Wg}^{\mathrm{O}}(\pi_2^{-1} \rho; T, -q)] \Big) \prod_{\{u,v\} \in \rho} \sigma_{i_u i_v}^{(-1)} \qquad (5.3.17)$$

where $\pi_1 = \{\{1,3\},\{2,4\}\}$ and $\pi_2 = \{\{1,4\},\{2,3\}\}$.

From direct computations using (4.3.4) and the values of Wg in [CM], we obtain the following equations:

$$\sum_{\rho \in M_4} \mathrm{Wg}^{\mathrm{O}}(\rho; T, -q) \prod_{\{u,v\} \in \rho} \sigma_{i_u i_v}^{(-1)} = \frac{1}{T(T+2)(T-1)q(-q+2)(q+1)} \Big(((T+1)(-q+1)+2) \sigma_{i_1 i_2}^{(-1)} \sigma_{i_3 i_4}^{(-1)} + (q-T-1) \sigma_{i_1 i_3}^{(-1)} \sigma_{i_2 i_4}^{(-1)} + (q-T-1) \sigma_{i_1 i_4}^{(-1)} \sigma_{i_2 i_3}^{(-1)} \Big), \qquad (5.3.18)$$

$$\sum_{\rho \in M_4} \mathrm{Wg}^{\mathrm{O}}(\pi_1\rho; T, -q) \prod_{\{u,v\}\in\rho} \sigma^{(-1)}_{i_u i_v} = \frac{1}{T(T+2)(T-1)q(-q+2)(q+1)}$$
$$\Big((T+1)(q-T-1)\sigma^{(-1)}_{i_1i_2}\sigma^{(-1)}_{i_3i_4} + ((T+1)(-q+1)+2)\sigma^{(-1)}_{i_1i_3}\sigma^{(-1)}_{i_2i_4} + (q-T-1)\sigma^{(-1)}_{i_1i_4}\sigma^{(-1)}_{i_2i_3}\Big), \tag{5.3.19}$$

and,

$$\sum_{\rho \in M_4} \mathrm{Wg}^{\mathrm{O}}(\pi_2^{-1}\rho; T, -q) \prod_{\{u,v\}\in\rho} \sigma^{(-1)}_{i_u i_v} = \frac{1}{T(T+2)(T-1)q(-q+2)(q+1)}$$
$$\Big((q-T-1)\sigma^{(-1)}_{i_1i_2}\sigma^{(-1)}_{i_3i_4} + (q-T-1)\sigma^{(-1)}_{i_1i_3}\sigma^{(-1)}_{i_2i_4} + ((T+1)(-q+1)+2)\sigma^{(-1)}_{i_1i_4}\sigma^{(-1)}_{i_2i_3}\Big). \tag{5.3.20}$$

Substitute from (5.3.18), (5.3.19), and (5.3.20) into (5.3.17) to obtain

$$E(w^{(-1)}_{i_1i_2} w^{(-1)}_{i_3i_4}) = \frac{1}{T(T+2)(T-1)q(q-2)(q+1)} \left((\mathrm{Tr}(B_T^{-1}))^2 I_1 + \mathrm{Tr}(B_T^{-2}) I_2\right), \tag{5.3.21}$$

where $q > 2$ and

$$I_1 = \Big(((T+1)(q-1)-2)\sigma^{(-1)}_{i_1i_2}\sigma^{(-1)}_{i_3i_4} + (T-q+1)\sigma^{(-1)}_{i_1i_3}\sigma^{(-1)}_{i_2i_4} + (T-q+1)\sigma^{(-1)}_{i_1i_4}\sigma^{(-1)}_{i_2i_3}\Big),$$

and

$$I_2 = \Big(2(T-q+1)\sigma^{(-1)}_{i_1i_2}\sigma^{(-1)}_{i_3i_4} + (Tq-2)\sigma^{(-1)}_{i_1i_3}\sigma^{(-1)}_{i_2i_4} + (Tq-2)\sigma^{(-1)}_{i_1i_4}\sigma^{(-1)}_{i_2i_3}\Big).$$

By substituting from (5.3.16) and (5.3.21) into (5.3.15), the proof is complete. ∎

If $B = I_T$, then $\widehat{\Sigma}$ in (5.3.1) is the MLE of the covariance matrix Σ. For this case, Proposition 5.3.9 reduces to the following interesting corollary.

Corollary 5.3.10 *Let $\widehat{\Sigma}$ be as defined in (5.3.1). If $B_T = I_T$, then for $q > 2$*

$$Var(Q) = \frac{2T^2}{q^2(q-2)}. \tag{5.3.22}$$

Remark 5.3.11 *Note that if $B = I_T$, then Corollary 5.3.10 implies that as $n, T \to \infty$ such that $\frac{n}{T} \to r (r < 1)$, $Var(Q) \to 0$.*

Now, we are going to prove Theorem 5.3.7.

Proof: Let

$$Z_{n,T} = Q - E(\mathrm{tr}((\frac{1}{\mathrm{Tr}(B_T)} X^t B_T X)^{-1})).$$

The proof is divided into two parts. First, we show that

$$E(Q) = \mathrm{Tr}(B_T) E(\mathrm{tr}((X^t B_T X)^{-1})),$$

then we will prove that for $T > n + 3$

$$Var(Z_{n,T}) \to 0 \quad \text{as} \quad n, T \to \infty \quad \text{such that} \quad n/T \to r < 1.$$

For the first part, apply Proposition 5.3.4 to (5.3.7) to get:

$$E(Q) = \frac{E(\mathrm{Tr}(\widehat{\Sigma}^{-1}))}{\mathrm{Tr}(\Sigma^{-1})}. \tag{5.3.23}$$

From (5.3.3), we have

$$\begin{aligned} E(Q) &= \frac{\mathrm{Tr}(B_T) E(\mathrm{Tr}(\Sigma^{-\frac{1}{2}} (X^t B_T X)^{-1} \Sigma^{-\frac{1}{2}}))}{\mathrm{Tr}(\Sigma^{-1})} \\ &= \frac{\mathrm{Tr}(B_T) \mathrm{Tr}(\Sigma^{-1} E((X^t B_T X)^{-1}))}{\mathrm{Tr}(\Sigma^{-1})}. \end{aligned}$$

Since $X^t B X$ is orthogonally invariant then by Lemma 5.3.5, we obtain

$$E(Q) = \frac{\beta \, \mathrm{Tr}(B_T) \, \mathrm{Tr}(\Sigma^{-1})}{\mathrm{Tr}(\Sigma^{-1})},$$

where $\beta = E(tr((X^t B_T X)^{-1}))$ which prove that $E(Z_{n,T}) = 0$. This concludes the first part of the proof.

To complete the proof of the theorem, it is enough to show that for $T > n + 3$ and as $T, n \to \infty$ such that $\frac{n}{T} \to r (r < 1)$, $Var(Q) \to 0$. By Proposition 5.3.9, for $q > 2$

$$Var(Q) = \frac{(\mathrm{Tr}(B_T))^2}{T^2 (T+2)(T-1) q^2 (q-2)(q+1)} \left(A_1 (\mathrm{Tr}(B_T^{-1}))^2 + A_2 \, \mathrm{Tr}(B_T^{-2}) \right) \tag{5.3.24}$$

where

$$A_1 = 2T^2q - 2Tq^2 + 2T^2 + 2T + 2q^2 - 2q - 4$$

and,

$$A_2 = Tq(2T - 2q + 2Tq - 2).$$

Suppose that

$$\lim_{T\to\infty} \frac{1}{T^4}(\mathrm{Tr}(B_T))^2\,\mathrm{Tr}(B_T^{-2}) = 0. \tag{5.3.25}$$

By the Cauchy-Schwarz inequality,

$$(\mathrm{Tr}(B_T^{-1}))^2 \leq T\,\mathrm{Tr}(B_T^{-2}). \tag{5.3.26}$$

From (5.3.25) and (5.3.26), we get

$$\lim_{T\to\infty} \frac{1}{T^4}(\mathrm{Tr}(B_T))^2(\mathrm{Tr}(B_T^{-1}))^2 = 0. \tag{5.3.27}$$

Since $q = T - n - 1$ then, for $T - n > 3$, (5.3.24) can be written as:

$$Var(Q) = \frac{(\mathrm{Tr}(B_T))^2 \cdot (A_1^*(\mathrm{Tr}(B_T^{-1}))^2 + A_2^*\,\mathrm{Tr}(B_T^{-2}))}{T^2(T+2)(T-1)\cdot S(T,n)} \tag{5.3.28}$$

where

$$A_1^* = 2(T^2n - Tn^2 + 3T^2 + n^2 - 4Tn - 3T + 3n),$$

$$A_2^* = 2T^4 - 4T^3n + 2T^2n^2 - 2Tn^2 + 6T^2n - 4T^3 + 2T^2 - 2Tn,$$

and

$$\begin{aligned} S(T,n) &= T^4 - 4T^3n + 6T^2n^2 - 4Tn^3 + n^4 - 5T^3 + 15T^2n - 15Tn^2 + 5n^3 \\ &+ 7T^2 - 14Tn + 7n^2 - 3T + 3n. \end{aligned}$$

Let $R = \frac{n}{T}$ then, from (5.3.28) and for $T > n + 3$,

$$Var(Q) = \frac{(\mathrm{Tr}(B_T))^2\big(A_1^{**}(\mathrm{Tr}(B_T^{-1}))^2 + A_2^{**}\,\mathrm{Tr}(B_T^{-2})\big)}{T^3(T+2)(T-1)((1-R^3)T^3 - 5(1-R)^2T^2 + 7(1-R)T - 3)} \tag{5.3.29}$$

where

$$A_1^{**} = 2T(RT^2 - (3-R)T - 3),$$

and

$$A_2^{**} = 2T^2(T^2(1-R) - (2-R)T + 1).$$

From (5.3.29), (5.3.25) and (5.3.27),

$$Var(Q) \to 0 \quad \text{as} \quad n, T \to \infty \quad \text{such that} \quad R \to r < 1.$$

∎

Remark 5.3.12 *For $T > n+3$ and for the case $r = 1$, (5.3.29) implies that as n, T tend to infinity such that $n/T \to 1$, Theorem 5.3.7 still holds if $\frac{1}{T^4}(\operatorname{Tr}(B_T))^2 \operatorname{Tr}(B_T^{-2})$ converges to 0 faster than the convergence of $(1 - \frac{n}{T})^2$ to 0. Under this condition, our simulation shows that this result still works for $T > n$.*

Remark 5.3.13 *For the case $T \leq n$, we need to compute the moments of the inverted Wishart matrices when $T < n+3$.*

Remark 5.3.14 *For*

$$B_T = \begin{pmatrix} e^{-1} & 0 & \dots & \dots & 0 \\ 0 & e^{-2} & 0 & \dots & 0 \\ 0 & 0 & \ddots & 0 & \dots \\ \vdots & 0 & \dots & \ddots & 0 \\ 0 & \dots & \dots & 0 & e^{-T} \end{pmatrix}$$

the condition in (5.3.12) is not satisfied and our simulation shows that Theorem 5.3.7 *is not valid either. On the other hand side, if*

$$B_T = \begin{pmatrix} 1 & 0 & \dots & \dots & 0 \\ 0 & 2 & 0 & \dots & 0 \\ 0 & 0 & 3 & 0 & \dots \\ \vdots & 0 & \dots & \ddots & \vdots \\ 0 & \dots & \dots & 0 & T \end{pmatrix}$$

then, the condition in (5.3.12) is not satisfied while Theorem 5.3.7 holds. From that we conclude that the condition in (5.3.12) is a sufficient condition for Theorem 5.3.7 but not a necessary one.

According to Theorem 5.3.7, to know the asymptotic value of Q we need to study the asymptotic behavior of the term $\operatorname{Tr}(B)\operatorname{tr}((X^t B_T X)^{-1})$. In the following lemma, we study the distribution of the matrix $X^t B_T X$.

Lemma 5.3.15 *[Coch] Let X be a Gaussian matrix with i.i.d. standard normal entries. For a $T \times T$ real matrix B_T, $X^t B_T X$ has the same distribution as a weighted sum of independent white Wishart matrices such that the weights are the eigenvalues of the matrix B_T.*

Proof: Since X is left-right orthogonally invariant, then by the spectral decomposition of the matrix B_T,

$$X^t B_T X \stackrel{\mathcal{L}}{=} X^t \Lambda_B X$$

where, for the eigenvalues $(\lambda_1, \lambda_2, \dots, \lambda_T)$ of the matrix B_T,

$$\Lambda_B = \begin{pmatrix} \lambda_1 & 0 & \dots & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ \vdots & 0 & \ddots & 0 & \vdots \\ 0 & \dots & & \dots & \lambda_T \end{pmatrix}$$

Then,

$$X^t B_T X \overset{\mathcal{L}}{=} \sum_{i=1}^{T} \lambda_i X_i^t X_i$$

where X_i is an $n-$dimensional vector such that its entries are those of the i^{th} row of the matrix X. For $i = 1, \ldots, T$, $X_i^t X_i$ is a white Wishart matrix of rank one. Hence, the proof is complete. ∎

As shown in Lemma 5.3.15, the distribution of the matrix $X^t B_T X$ depends essentially on the eigenvalues of the matrix B_T. By applying Theorem 4.6.13 to Theorem 5.3.7, we obtain the following interesting corollary.

Corollary 5.3.16 *Let B_T be a $T \times T$ real matrix and let $\widehat{\Sigma}$ be as defined in (5.3.1). If $T > n+3$, and $\lim_{T\to\infty} \frac{1}{T}(\mathrm{tr}(B_T))^2 \,\mathrm{tr}(B_T^{-2}) = 0$ then, as T and n tend to infinity such that $\frac{n}{T} \to r < 1$ we have*

$$Q - \frac{\mathrm{Tr}(B_T)\,\mathrm{Tr}(B_T^{-1})}{T(T-n-1)} \xrightarrow{\mathrm{P}} 0. \qquad (5.3.30)$$

In the next section we are going to consider the case of independent observations.

5.4 The case where B_T is an idempotent

In the following, we are going to consider an important case of the matrix B_T. Let B_T be an idempotent i.e., $B_T = B_T^2$. If B_T has rank $m \leq T$ then, B_T has m nonzero eigenvalues and each eigenvalue equals one. In this case, Lemma 5.3.15 implies that $X^t B_T X$ is a white Wishart matrix with m degrees of freedom (a sum of m independent white Wishart matrices). Also,

$$\mathrm{Tr}(B_T) = \mathrm{Tr}(B_T^{-1}) = m,$$

and from Corollary 5.3.16, we get the following important result:

Corollary 5.4.1 *For $\widehat{\Sigma}$ as defined in (5.3.1), if the matrix B_T is an idempotent of rank $m = T - k$ for some $k \geq 0$ and $T > n+3$, then as T and n tend to infinity and $\frac{n}{T} \to r < 1$, we have*

$$Q \xrightarrow{\text{P}} 1/(1-r).$$

In the following, we are going to discuss an important example of such case.

5.4.1 Example: Maximum Likelihood Estimator (MLE)

$\widehat{\Sigma}$ in (5.3.1) is the maximum likelihood estimator of the covariance matrix Σ if $B_T = I_T$. By applying Corollary 5.4.1, we get the following corollary

Corollary 5.4.2 *For the MLE of the covariance matrix, if $T > n+3$ then, as T and n tend to infinity and $\frac{n}{T} \to r < 1$, we have*

$$Q \xrightarrow{\text{P}} 1/(1-r).$$

Remark 5.4.3 *This result coincides with the result of Pafka and Kondor in [PK].*

Now, let us simulate the result in Corollary 5.4.2 using the following algorithm:

Algorithm 1 Algorithm for simulating the risk of the optimal portfolio of MLE before and after scaling.

Choose n, T such that $T > n+3$.

Choose some $\Sigma \in \mathbb{M}_n^+$ ($n \times n$ positive definite matrix).

Find the matrix Σ^{-1} and compute the True risk.

Construct a data matrix from the normal distribution $N(0, \Sigma)$.

Estimate the covariance matrix using the MLE.

Find the inverse of the covariance estimator and compute the Predicted risk.

Scale the Predicted risk by the ratio $\frac{1}{\sqrt{1-\frac{n}{T}}}$.

Plot the histogram of the ratio between the risks before and after scaling.

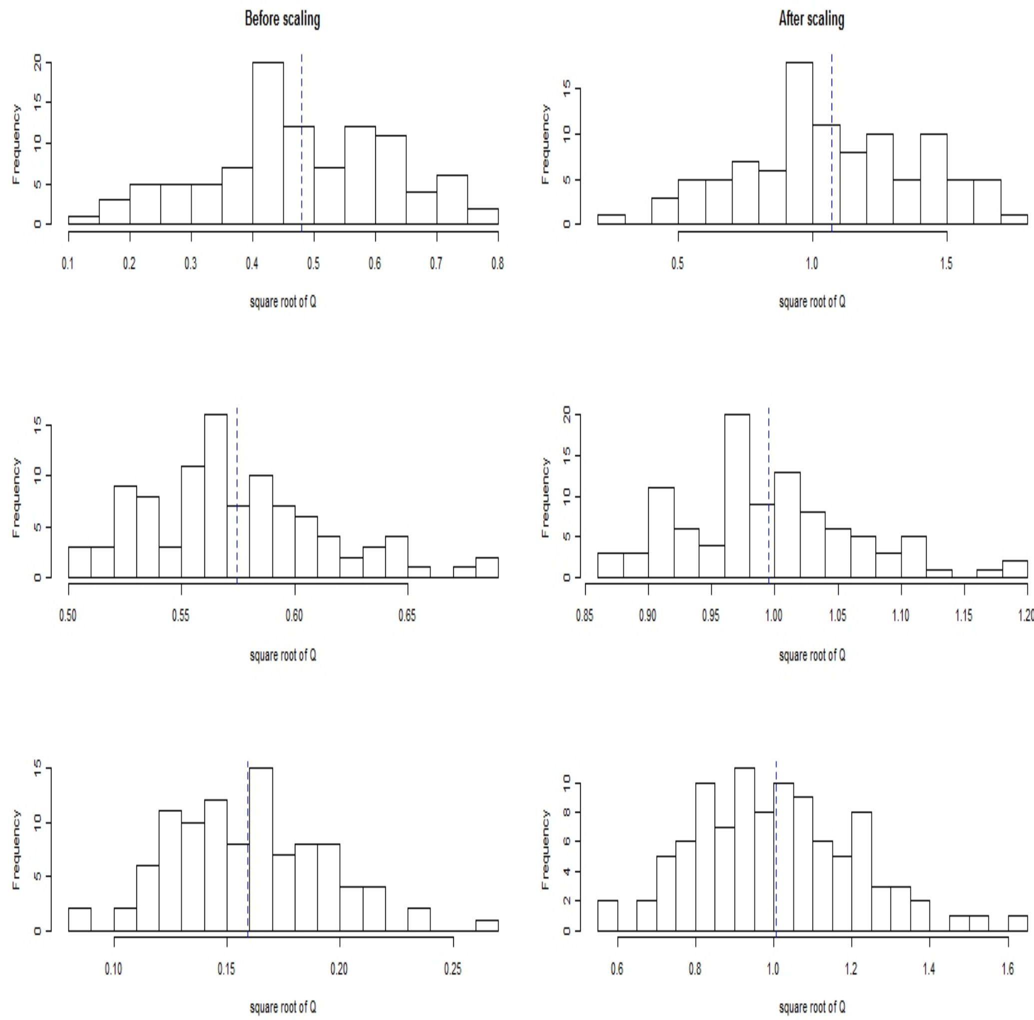


Figure 5.1: The figure illustrates the ratio between the Predicted and the True risks for the MLE before and after scaling using Corollary 5.4.2. The left side of the figure represents the ratio between the two risks before scaling while the graphs on the right hand side of the figure describe the histogram of the ratio between the risks after scaling by the factor $\frac{1}{\sqrt{1-r}}$. The middle part of the figure illustrates the ratio between the risks when $n = 200$ and $T = 250$. In the upper part of the figure, we focus on the case of small values of n and T ($n = 20, T = 25$) while in the lower graphs, we choose n and T with close values ($n = 390$ and $T = 400$). The mean of the ratio between the Predicted and the True risks, represented by a dotted line in each histogram, shows a valuable improvement in estimating the Predicted risk after scaling the Predicted risk using Corollary 5.4.2.

In Figure (5.1), simulations show that we get a remarkable improvement in estimating the risk for MLE after scaling the Predicted risk using the factor $\frac{1}{\sqrt{1-\frac{n}{T}}}$ in Corollary 5.4.2. The figure illustrates the ratio between the Predicted and the True risks before and after applying Corollary 5.4.2. The dotted line in each histogram represents the mean of the ratio between the two risks. For the middle graphs of the figure, we take $n = 200$ and $T = 250$ and for these values the mean of the ratio between the risks before and after scaling equals 0.575 and 0.996, respectively which shows a remarkable improvement in computing the Predicted risk.
To study the validity of the Scaling technique for small values of n, T, we take $n = 20$ and $T = 25$ and as shown in the upper graphs of Figure 1 the mean of the ratio between the risks before and after scaling is 0.464 and 1.037. So, Scaling technique is still valid for small dimensions and small observations situations.
In the lower graphs of Figure 1, we choose closed values for n and T ($n = 390$ and $T = 400$) and the mean of the ratio between the risks equals 0.159 and 1.007 before and after scaling, respectively. From the simulations, we conclude that "for the MLE, the Scaling technique is a real improvement in estimating the risk". Also, note that the reduction in the standard deviation of the ratio of the Predicted and the True risks from the upper graph to the middle graph as n and T increases from $n = 20$ and $T = 25$ t0 $n = 200$ and $T = 250$. In theory, the standard deviation goes to zero an n and T tend to infinity such that $n/T \to r$ ($r < 1$) by Corollary 5.4.2.

5.4.2 Sample Covariance Matrix (SCM) (with unknown mean)

In the case of unknown returns' expected means, the sample covariance matrix (the unbiased estimator of the covariance matrix) is given by

$$\widehat{\Sigma} = \frac{1}{T-1} Y^t Y.$$

The sample covariance estimator can be obtained from (5.3.1) by considering the matrix B_T as follows:

$$B_T = \begin{pmatrix} 1-\frac{1}{T} & -\frac{1}{T} & \dots & \dots & -\frac{1}{T} \\ -\frac{1}{T} & 1-\frac{1}{T} & -\frac{1}{T} & \ddots & \vdots \\ \vdots & -\frac{1}{T} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & -\frac{1}{T} \\ -\frac{1}{T} & -\frac{1}{T} & \dots & -\frac{1}{T} & 1-\frac{1}{T} \end{pmatrix}$$

In this case, B is an idempotent of rank $T-1$. In [El-K], El-Karoui shows that the asymptotic behavior of the noise resulting from estimating the covariance matrix using the sample covariance estimator (with unknown expected means of the returns) is $\frac{1}{\sqrt{1-\frac{n-1}{T-1}}}$ which is still coincides with our result in Corollary 5.4.1 although in our case we assume the returns are centered. This similarity between the two cases due to the independence between the estimators $\hat{\mu}$ and $\widehat{\Sigma}$. To simulate this case, we randomly choose certain values to define the mean vector μ and the covariance matrix Σ. Using these values, we compute the True risk. Now, we generate a set of observations from the distribution $N(0,\Sigma)$ and estimate μ and Σ using these observations. Finally, we compute the Predicted risk using the estimators $\hat{\mu}$ and $\widehat{\Sigma}$ and compare the Predicted and the True risks. As shown in Figure (5.2), the ratio between the scaled Predicted risk and the True risk is very close to one and there is a valuable improvement in estimating the Predicted risk after using the Scaling technique.

In the next section, we are going to study an important estimator of the covariance matrix which plays a great role in many fields, specially in finance.

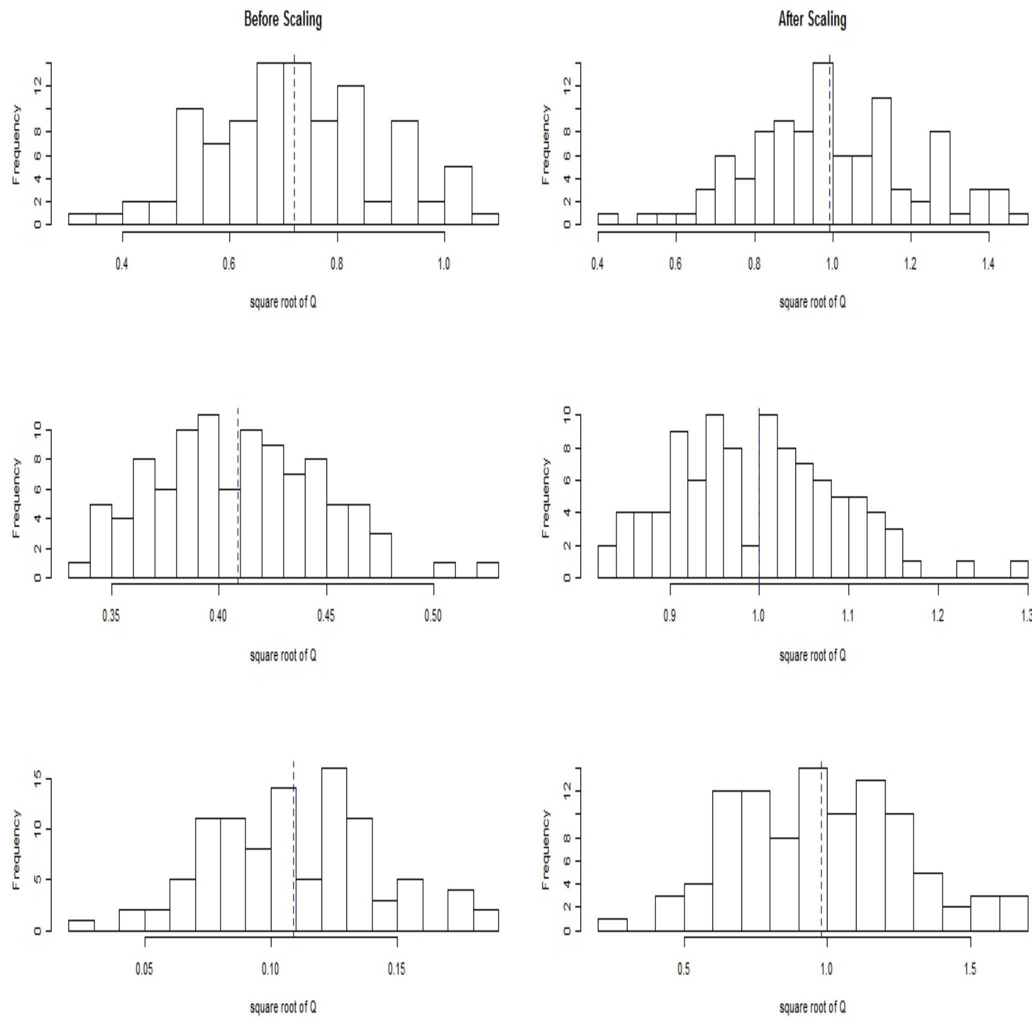


Figure 5.2: The figure describes the ratio between the *Predicted* and *True* risks for the Sample covariance matrix (the Standard estimator) before and after scaling by the factor $\frac{1}{\sqrt{1-\frac{n}{T}}}$. The left side of the figure represents the ratio between the two risks before scaling while the graphs on the right hand side of the figure describe the histogram of the ratio between the risks after scaling. In each histogram, the dotted line represents the mean value of the histogram. For the first part of the figure, $n = 10$, $T = 20$ and the mean of the ratio between the two risks equals 0.696 (before scaling) and 0.959 (after scaling). In the middle graphs, we talk $n = 250$ and $T = 300$ and for these values of n and T, the mean of the middle histograms before and after scaling equals 0.407 and 0.996, respectively. The lower graphs of the figure describe the case of closed values of n and T, we take $n = 400$ and $T = 405$. As shown in the figure, there is a remarkable improvement in predicting the risk of the optimal portfolio.

5.5 Exponentially Weighted Moving Average (EWMA) Covariance Matrix

Using equally weighted data doesn't accurately exhibit the current state of the market. It reflects market conditions which are no longer valid by assigning equal weights to the most recent and the most distant observations. To express the dynamic structure of the market, it is better to use exponentially weighted variances.

Exponentially weighted data gives greater weight to the most recent observation. Thus, current market conditions are taken into consideration more accurately. The EWMA model is proposed by Bollerslev [Bol]. Related studies ([F], [T], [RN]) are made in the equity market and using exponentially weighted moving average techniques (weighting recent observations more heavily than older observations). In [Ak], Akgiray shows that using EWMA techniques are more powerful than the equally weighted scheme.

In EWMA technique, returns of recent observations to distant ones are weighted by multiplying each term by an exponential factor $\lambda^0, \lambda^1, \lambda^2, \ldots$ $(0 < \lambda < 1)$, respectively. In common, λ is called the decay factor. In [PB], Penza et al. choose the values of the decay factor to be 0.97 for the daily data set and 0.94 for the monthly data set. For the EWMA covariance matrix, the weighted matrix $B_T = (b_{ij})_{i,j=1}^T$ is a diagonal matrix such that $b_{ii} = \lambda^{i-1}$ for $(i = 1, \ldots, T)$ i.e.,

$$B_T = \begin{pmatrix} 1 & 0 & 0 & \ldots & \ldots & \ldots & 0 \\ 0 & \lambda & 0 & \ldots & \ldots & \ldots & \vdots \\ \vdots & 0 & \lambda^2 & 0 & \ldots & \ldots & \vdots \\ \vdots & \vdots & 0 & \ddots & 0 & \ldots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \ldots & \ldots & 0 & 0 & \lambda^{T-1} \end{pmatrix}$$

In this case, we have

$$\operatorname{Tr}(B_T)\operatorname{Tr}(B_T^{-1}) = \frac{(1-\lambda^T)^2}{\lambda^{T-1}(1-\lambda)^2}.$$

If $\lambda \to 1$ then,

$$\lim_{T\to\infty} \frac{1}{T}(\operatorname{tr}(B_T))^2 \operatorname{tr}(B_T^{-2}) = 0.$$

Now, let us apply Theorem 5.3.7 to the EWMA estimator and obtain the following corollary.

Corollary 5.5.1 *Let $\widehat{\Sigma}$ be the EWMA estimator of the covariance matrix Σ with decay factor $0 < \lambda < 1$. If $T > n+3$ then, as λ tend to 1 and as T, n tend to infinity such that $(1-\lambda)T = c$ (for some positive constant c) and $n/T \to r < 1$, we have*

$$Q \xrightarrow{\text{P}} (e^c - 1)^2 / c^2(1-r)e^c.$$

Now, let us simulate the result in Corollary 5.5.1 using the following algorithm:

Algorithm 2 Algorithm for simulating the risk of the optimal portfolio of EWMA before and after scaling.

Define n, T and $\lambda < 1$ such that $T > n+3$ and $(1-\lambda)T = c$

Choose $\Sigma \in \mathbb{M}_n^+$ ($n \times n$ positive definite matrix).

Find the matrix Σ^{-1} and compute the True risk.

Construct a data matrix from the normal distribution $N(0, \Sigma)$.

Estimate the covariance matrix using the EWMA.

Find the inverse of the covariance estimator and compute the Predicted risk.

Scale the Predicted risk by the ratio $(\exp(c) - 1)/(c\sqrt{(1-r)\exp(c)})$.

Plot the histogram of the ratio between the Predicted and the True risks before and after scaling.

As shown in Figure (5.3), for the EWMA covariance matrices, scaling the Predicted risk using Corollary 5.5.1 gives a great improvement to estimate the risk of the

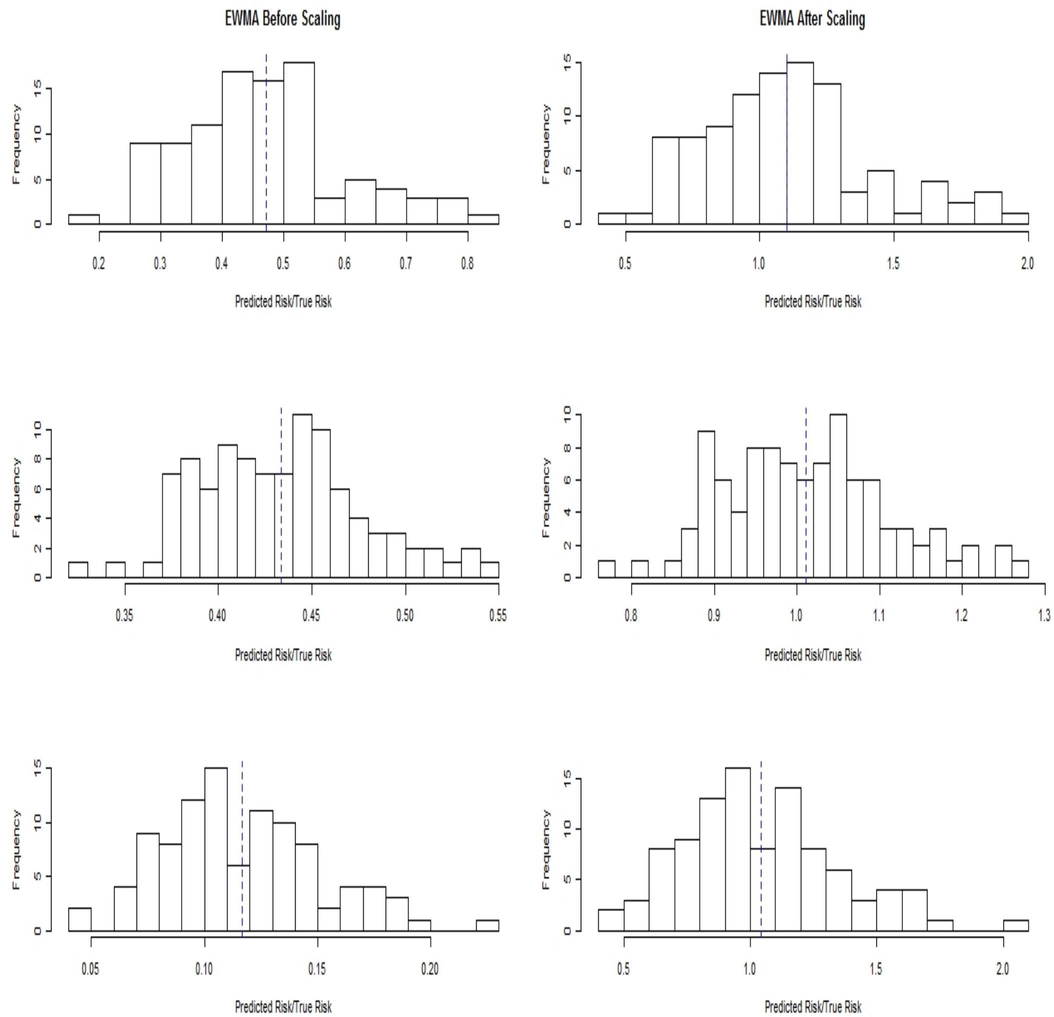


Figure 5.3: The figure describes the ratio between the *Predicted* and the *True* risks for the EWMA covariance estimator before and after scaling using Corollary 5.5.1. In the first row, we take small values for n and T ($n = 20$, $T = 25$, and $\lambda = 0.96$). The means of the histograms of the upper graphs, represented by the dotted line in each histogram, equal 0.47 (before scaling) and 1.099 (after scaling). In the second row, we take $n = 200$, $T = 250$, and $\lambda = 0.996$. The means of the histograms before and after scaling are 0.43 and 1.01, respectively. In the lower graphs, $n = 395$, $T = 400$, and $\lambda = 0.9996$, and the means of the histograms equal 0.12 (before scaling) and 1.04 (after scaling). Comparing the graphs before scaling (on the left) and the graphs on the right (after scaling), it is clear that the ratio between the Predicted and the True risks becomes closer to one after using the Scaling technique.

optimal portfolio. Before scaling as illustrated in the graphs on the left hand side of Figure (5.3), the ratio between the two risks is far from 1 specially for close values of n and T ($n = 395, T = 400$) as shown in the lower left graph of the figure. After scaling the Predicted risk by the factor $\frac{(\exp(c)-1)}{c\sqrt{(1-r)\exp(c)}}$ as in Corollary 5.5.1, the ratio between the Predicted and the True risks becomes very close to 1 as in the right hand sides graphs of the figure. For small values of n and T, as in the upper graphs of the figure, $n = 20$ and $T = 25$, the means of the histograms of the upper graphs, represented by the dotted line in each histogram, equal 0.47 (before scaling) and 1.099 (after scaling). So, the Scaling technique still works and improves the estimation of the Predicted risk. Again note the reduction in the standard deviation of the ratio of the Predicted and the True risks from the upper graph to the middle graph as n and T increases from $n = 20$ and $T = 25$ t0 $n = 200$ and $T = 250$.

5.6 Comparison Between Filtering and Scaling Techniques

In this section, we would like to make a comparison between the Filtering and the Scaling techniques to show which one provides a better prediction of the optimal portfolio's risk. As shown before, the Filtering technique deals with correlation matrices while the scaling technique depends on the covariance matrices. To apply the Filtering technique for $\widehat{\Sigma} = \frac{1}{T}Y^tY$ where Y is a $T \times n$ data matrix whose rows are $n-$dimensional vectors of centered returns which are taken sequentially in time: $Y_1, Y_2, \dots, Y_T$. We assume that these vectors are i.i.d. with distribution $N(0, \Sigma)$, we need to convert the estimated covariance matrix to the corresponding correlation matrix $C = \frac{1}{T}X^tX$ where X is a $T \times n$ Gaussian matrix with i.i.d. standard normals entries. According to the work of Marčenko and Pastur [MP] (as discussed in Chapter 3), if both sample size T and data dimension n proportionally grow to ∞ such that $\lim n/T = r$ for some positive $r > 0$, the empirical spectral distribution of the correlation matrix C converges to a nonrandom distribution. This limiting spectral distribution, the Marčenko-Pastur distribution of index r, has a density function

$$\mu(x) = \frac{\sqrt{(b-x)(x-a)}}{2\pi r x} \qquad a \leq x \leq b$$

with $a = (1-\sqrt{r})^2$ and $b = (1+\sqrt{r})^2$. The Filtering technique relies on the Marcenko-Pastur distribution to remove the noisy eigenvalues (these eigenvalues fall in the region [a,b]).

Using Algorithm 3, we simulate the ratio between the Predicted and the True risks using the Filtering and the Scaling techniques. As shown in Figure 5.4, for different values of n and T, the Scaling technique (left graphs) provides a better prediction of the optimal portfolio's risk than the Filtering technique (in the right graphs). For $n = 50$ and $T = 100$ in the upper graphs of Figure 5.4, the means of the ratio between the risks are 0.715 and 1.01 after filtering and scaling, respectively. It

Algorithm 3 Algorithm for simulating the risk of the optimal portfolio using the Filtering and the Scaling techniques

For n (no. of the assets) and T (no. of observations of the returns), choose some positive definite matrix to be the covariance matrix Σ.

Find the Inverse of Σ and use it to compute the True risk.

Construct a data matrix from the distribution Normal$(0, \Sigma)$.

Estimate the covariance matrix using the MLE estimator and find its inverse to compute the Predicted risk.

Apply the Scaling technique

Scale the Predicted risk by the factor $1/\sqrt{1-\frac{n}{T}}$ and call it the "scaled" Predicted risk.

Apply the Filtering technique:

for m = 1 . . . 100 **do**

use the estimated covariance matrix to get the corresponding correlation matrix.

find the average value of the noisy eigenvalues of the estimated correlation matrix:

$Average = 0, \; s = 0$

for $1 \le i \le n$ **do**

if $(1-\sqrt{\frac{n}{T}})^2 < eigenvalue < (1+\sqrt{\frac{n}{T}})^2$ **then**

$Average = Average + eigenvalue, \; s = s + 1$

end if

end for $\quad Average = Average/s$

Clean the eigenvalues of the correlation matrix by replace the noisy eigenvalues by their average value

for $1 \le i \le n$ **do**

if $(1-\sqrt{\frac{n}{T}})^2 < eigenvalue < (1+\sqrt{\frac{n}{T}})^2$ **then**

$eigenvalue = Average$

end if

end for

end for

From the cleaned Correlation matrix, find the corresponding cleaned Covariance matrix and get the inverse of the cleaned Covariance matrix to compute the "filtered" Predicted risk.

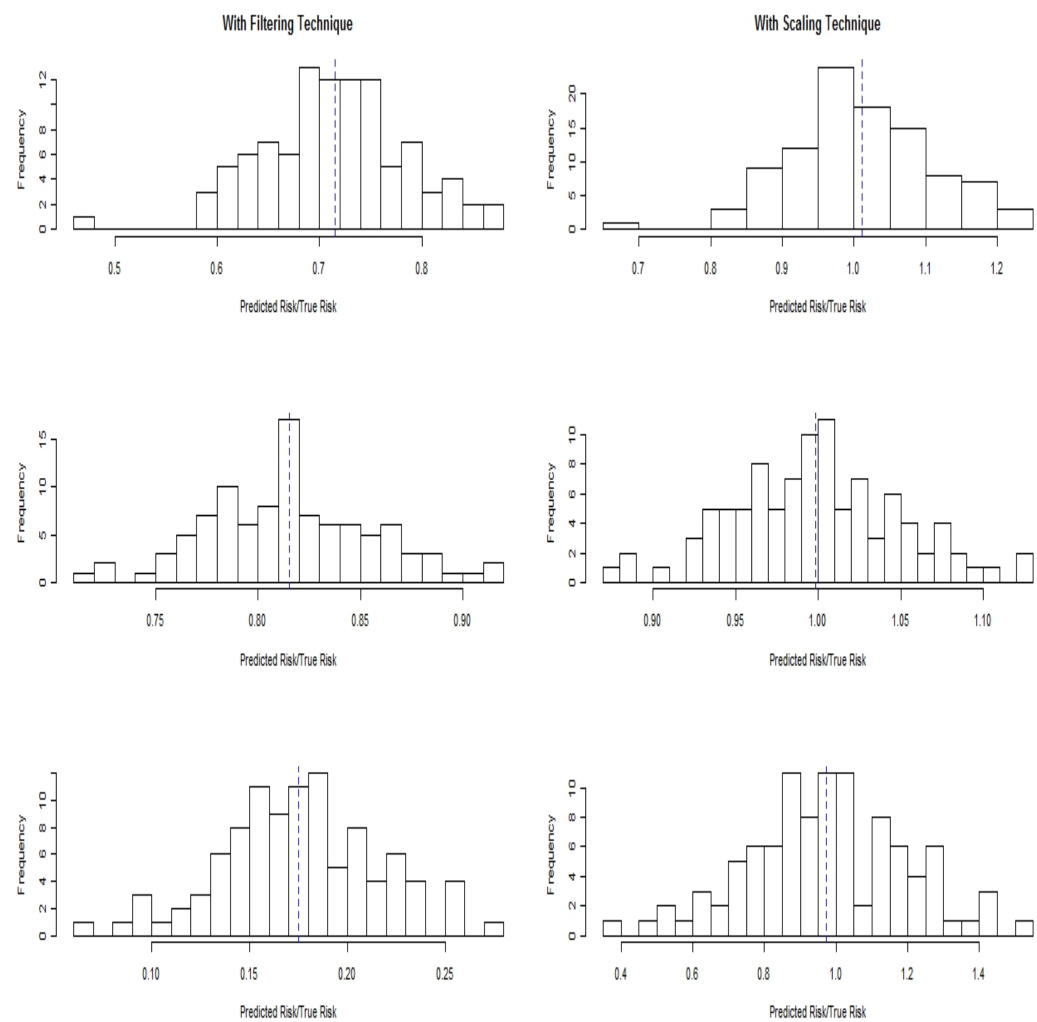


Figure 5.4: The figure describes the ratio between the *Predicted* and the *True* risks using the Filtering and the Scaling techniques. In the first row, we take $n = 50$ and $T = 100$. The means of the histograms, represented by the dotted line in each histogram, show that the Scaling technique provides a better estimation of the risk's optimal portfolio than the Filtering technique. In the second row, we take $n = 100$ and $T = 300$. The means of the histograms are 0.82 (after filtering) and 0.998 (after scaling). For close values of n and T, we take $n = 300$ and $T = 310$ and from the lower graphs, the Scaling technique admits a real estimator of the risk for this case. The means of the histograms equal 0.175 and 0.973 after filtering and scaling, respectively.

is clear that the ratio between the two risks becomes closer to one after scaling than after filtering, specially when the dimensions n and T are close as shown in the lower histograms of the figure where $n = 300$ and $T = 310$ and the means of the histograms equal 0.175 (after filtering) and 0.973 (after scaling).

5.7 Optimal Weights

As shown in Chapter 2, the optimal weights are given by:

$$\omega_i = \frac{\sum_{j=1}^{n} \sigma_{ij}^{(-1)}}{\sum_{j,k=1}^{n} \sigma_{jk}^{(-1)}} \quad (i = 1, \ldots, n). \tag{5.7.1}$$

It is clear that the weights of the optimal portfolio depends essentially on the covariance matrix of returns. Again, the covariance matrix of the returns is unknown in practice and we are dealing with an estimator of the covariance matrix. So, we need to study the effect of the noise resulting from estimating the covariance matrix on computing the optimal weights. To do that, we need to define the vectors:

$$\mathbf{w} = (\omega_1, \omega_2, \ldots, \omega_n)^t,$$

the vector of the true optimal weights and

$$\hat{\mathbf{w}} = (\hat{\omega}_1, \hat{\omega}_2, \ldots, \hat{\omega}_n)^t,$$

the vector of the optimal weights using the empirical covariance matrix $\widehat{\Sigma}$ defined in (5.3.1). Define

$$\breve{\mathbf{w}} = (\breve{\omega}_1, \breve{\omega}_2, \ldots, \breve{\omega}_n)^t$$

where

$$\breve{\omega}_i = \frac{\hat{\omega}_i}{\omega_i}. \tag{5.7.2}$$

Theorem 5.7.1 *Let B_T be a $T \times T$ real matrix such that*

$$\lim_{T\to\infty} \frac{1}{T}(\mathrm{tr}(B_T))^2 \,\mathrm{tr}(B_T^{-2}) = 0 \tag{5.7.3}$$

Let $\widehat{\Sigma}$ be as defined in (5.3.1). If $T > n+3$, then as n and T tends to ∞ such that $\frac{n}{T} \to r$ $(r < 1)$ we have

- *For an n-dimensional vector $\mathbf{e}$ whose entries are ones,*

$$\|E(\breve{\mathbf{w}}) - \mathbf{e}\| = 0.$$

  where $\|.\|$ is the l^2-norm. So, $\hat{\mathbf{w}}$ is an asymptotically unbiased estimator of the optimal weights.

- *Let $c_{in} = \frac{\sigma_{ii}^{(-1)} \sum\limits_{j_1,j_2=1}^{n} \sigma_{j_1 j_2}^{(-1)}}{(\sum\limits_{j=1}^{n} \sigma_{ij}^{(-1)})^2}$. If $\frac{c_{in}\,\mathrm{Tr}(B^{-2})}{(\mathrm{Tr}(B^{-1}))^2} \to 0$, then*

$$Var(\breve{\omega}_i) \to 0 \quad \text{for } i = 1, \ldots, n.$$

Proof: From (5.7.2) and (5.7.1), we get

$$\breve{\omega}_i = \frac{\sum\limits_{j=1}^{n} \hat{\sigma}_{ij}^{(-1)} / \sum\limits_{j=1}^{n} \sigma_{ij}^{(-1)}}{\sum\limits_{j,k=1}^{n} \hat{\sigma}_{jk}^{(-1)} / \sum\limits_{j,k=1}^{n} \sigma_{jk}^{(-1)}}. \tag{5.7.4}$$

By Corollary 5.3.16 and Corollary 2 (page 334 in [B]), if $T > n+3$ and $\lim\limits_{T\to\infty} \frac{1}{T}(\mathrm{tr}(B_T))^2\,\mathrm{tr}(B_T^{-2}) = 0$, then as T and n tend to infinity such that $\frac{n}{T} \to r < 1$ we have

$$\begin{aligned} E(\breve{\omega}_i) &= \frac{E(\sum\limits_{j=1}^{n} \hat{\sigma}_{ij}^{(-1)} / \sum\limits_{j=1}^{n} \sigma_{ij}^{(-1)})}{E(\sum\limits_{j,k=1}^{n} \hat{\sigma}_{jk}^{(-1)} / \sum\limits_{j,k=1}^{n} \sigma_{jk}^{(-1)})} \\ &= \frac{\sum\limits_{j,k=1}^{n} \sigma_{jk}^{(-1)} E(\sum\limits_{j=1}^{n} \hat{\sigma}_{ij}^{(-1)})}{\sum\limits_{j=1}^{n} \sigma_{ij}^{(-1)} E(\sum\limits_{j,k=1}^{n} \hat{\sigma}_{jk}^{(-1)})} \end{aligned}$$

$$= \frac{\sum_{j,k=1}^{n} \sigma_{jk}^{(-1)} \sum_{j=1}^{n} E(\hat{\sigma}_{ij}^{(-1)})}{\sum_{j=1}^{n} \sigma_{ij}^{(-1)} \sum_{j,k=1}^{n} E(\hat{\sigma}_{jk}^{(-1)})} \tag{5.7.5}$$

By Theorem 4.6.13,

$$E(\hat{\sigma}_{ij}^{(-1)}) = \frac{1}{Tq} \operatorname{Tr}(B_T^{-1}) \operatorname{Tr}(B_T) \sigma_{ij}^{(-1)}. \tag{5.7.6}$$

Substitute from (5.7.6) into (5.7.5) to get that as T and n tend to infinity such that $\frac{n}{T} \to r < 1$ we have

$$E(\breve{\omega}_i) = 1 \quad (i = 1 \ldots, n). \tag{5.7.7}$$

Hence,

$$\begin{aligned} \|E(\breve{\mathbf{w}}) - \mathbf{e}\|^2 &= \sum_{i=1}^{n} (E(\breve{\omega}_i) - 1)^2 \\ &= 0. \end{aligned}$$

This completes the proof of the first part. For the second part, By Corollary 5.3.16 and Corollary 2 (page 334 in [B]), if $T > n+3$ and $\lim_{T\to\infty} \frac{1}{T}(\operatorname{tr}(B_T))^2 \operatorname{tr}(B_T^{-2}) = 0$, then as T and n tend to infinity such that $\frac{n}{T} \to r < 1$ we have

$$E(\breve{\omega}_i^2) = \frac{(\sum_{j,k=1}^{n} \sigma_{jk}^{(-1)})^2 \sum_{j_1,j_2=1}^{n} E(\hat{\sigma}_{ij_1}^{(-1)} \hat{\sigma}_{ij_2}^{(-1)})}{(\sum_{j=1}^{n} \sigma_{ij}^{(-1)})^2 \sum_{j_1,j_2,k_1,k_2} E(\hat{\sigma}_{j_1k_1}^{(-1)} \hat{\sigma}_{j_2k_2}^{(-1)})} \tag{5.7.8}$$

Since $q = T - n - 1 > 2$, then from (5.3.1) and (5.3.21) we get

$$E(\hat{\sigma}_{ij_1}^{(-1)} \hat{\sigma}_{ij_2}^{(-1)}) = \frac{(\operatorname{Tr}(B))^2}{T(T+2)(T-1)q(q-2)(q+1)} \left((\operatorname{Tr}(B^{-1}))^2 S_1 + \operatorname{Tr}(B^{-2}) S_2 \right) \tag{5.7.9}$$

where

$$S_1 = (Tq-2)\sigma_{ij_1}^{(-1)} \sigma_{ij_2}^{(-1)} + (T - q + 1)\sigma_{ii}^{(-1)} \sigma_{j_1j_2}^{(-1)}$$

and

$$S_2 = (2T - 2q + Tq)\sigma_{ij_1}^{(-1)} \sigma_{ij_2}^{(-1)} + (Tq - 2)\sigma_{ii}^{(-1)} \sigma_{j_1j_2}^{(-1)}$$

Also, from (5.3.1) and (5.3.21) we get

$$E(\hat{\sigma}_{ij_1}^{(-1)}\hat{\sigma}_{ij_2}^{(-1)}) = \frac{(\mathrm{Tr}(B))^2}{T(T+2)(T-1)q(q-2)(q+1)}\left((\mathrm{Tr}(B^{-1}))^2 S_1^* + \mathrm{Tr}(B^{-2})S_2^*\right) \tag{5.7.10}$$

where

$$S_1^* = (Tq - T + q - 3)\sigma_{j_1k_1}^{(-1)}\sigma_{j_2k_2}^{(-1)} + (T-q+1)\sigma_{j_1j_2}^{(-1)}\sigma_{k_1k_2}^{(-1)} + (T-q+1)\sigma_{j_1k_2}^{(-1)}\sigma_{j_2k_1}^{(-1)}$$

and

$$S_2^* = 2(T-q+1)\sigma_{j_1k_1}^{(-1)}\sigma_{j_2k_2}^{(-1)} + (Tq-2)\sigma_{j_1j_2}^{(-1)}\sigma_{k_1k_2}^{(-1)} + (Tq-2)\sigma_{j_1k_2}^{(-1)}\sigma_{j_2k_1}^{(-1)}$$

By substituting from (5.7.9) and (5.7.10) into (5.7.8), we get

$$E(\breve{\omega}_i^2) = \frac{(Tq + c_{in}(T-q+1) - 2)(\mathrm{Tr}(B^{-1}))^2 + ((1+c_{in})Tq + 2T - 2q - 2c_{in})\,\mathrm{Tr}(B^{-2})}{(Tq + T - q - 1)(\mathrm{Tr}(B^{-1}))^2 + (2Tq + 2T - 2q - 6)\,\mathrm{Tr}(B^{-2})} \tag{5.7.11}$$

Since $q = T - n - 1$, then

$$E(\breve{\omega}_i^2) = \tag{5.7.12}$$

$$\frac{(T^2 - Tn - T + c_{in}(n+2) - 2)(\mathrm{Tr}(B^{-1}))^2 + (c_{in}(T^2 - Tn - T - 2) + 2n + 3)\,\mathrm{Tr}(B^{-2})}{(T^2 - Tn - T + n)(\mathrm{Tr}(B^{-1}))^2 + (2T^2 - 2Tn - 2T + 2n - 4)\,\mathrm{Tr}(B^{-2})}. \tag{5.7.13}$$

Hence, if $\frac{c_{in}\,\mathrm{Tr}(B^{-2})}{(\mathrm{Tr}(B^{-1}))^2} \to 0$, then $E(\breve{\omega}_i^2) = 1$ and the proof is complete. ∎

Remark 5.7.2

- *From Theorem 5.7.1, it is clear that $Var(\breve{\omega}_i)$ depends not only on B_T but also on Σ. For the MLE, the condition in the second part of Theorem 5.7.1 reduces to $\frac{c_{in}}{T} \to 0$.*

- *If $\Sigma = I_n$, then $c_{in} = n$ ($i = 1, \ldots, n$). Hence, Theorem 5.7.1 implies that for the MLE and as n and T tend to infinity such that $\frac{n}{T} \to r < 1$, then*

$$Var(\breve{\omega}_i) \to \frac{r}{1-r} \; (i = 1, \ldots, n).$$

It follows that $\hat{\omega}_i$ is an asymptotically unbiased consistent estimator of ω_i (for $i = 1, \ldots, n$) if $r = 0$.

In Figure (5.5), we take $\Sigma = I_n$ and for a certain asset, we simulate the ratio between the predicted and the true weights when $\widehat{\Sigma}$ is the MLE and $r = 1/2$. We take the following values of n and T, $(n = 30, T = 60)$, $(n = 100, T = 200)$, and $(n = 200, T = 400)$, respectively. As shown in the figure the variance of the ratio between the predicted and the true weights tends to a constant while the mean of the ratio tends to one.

In Figure (5.6), we take $\Sigma = I_n$ and for a certain asset, we simulate the ratio between the predicted and the true weights when $\widehat{\Sigma}$ is the MLE. The figure shows that as n, T tend to infinity such that $\frac{n}{T} \to 0$ the ratio $\breve{\omega}$ becomes closer to one. In the figure, we take the following values of n and T: $(n = 30, T = 60)$, $(n = 100, T = 300)$ and $(n = 150, T = 600)$, respectively.

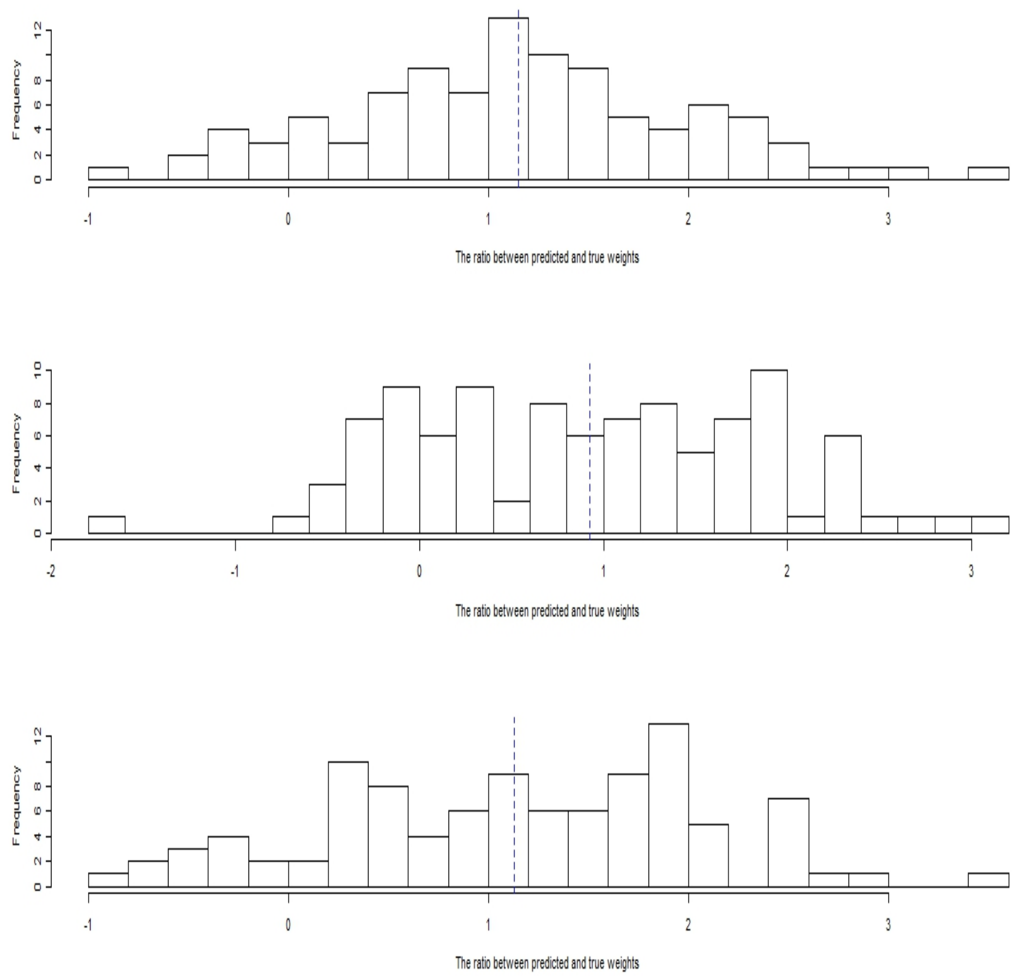


Figure 5.5: For the MLE, the figure describes the ratio between the *Predicted* and the *True* weights of a certain asset when $\Sigma = I_n$ and as n and T tend to infinity such that $\frac{n}{T} \to 1/2$.

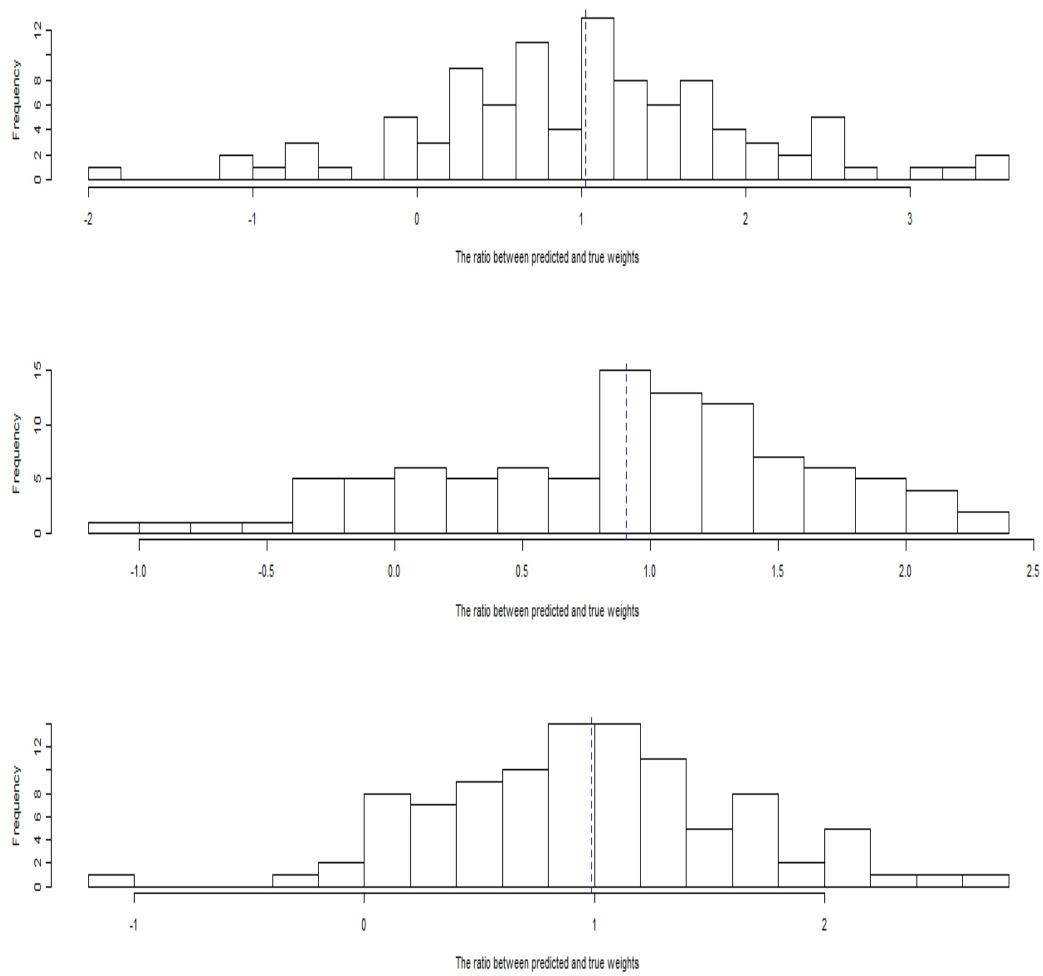


Figure 5.6: For the MLE, the figure describes the ratio between the *Predicted* and the *True* weights of a certain asset when $\Sigma = I_n$ and as n and T tend to infinity such that $\frac{n}{T} \to 0$.

5.8 Conclusion

For a general estimator of the covariance matrix and using our results concerning the moments of the inverse of the compound Wishart matrices in Chapter 4, we are able to get the asymptotic effect of the noise induced by estimating the covariance matrix of the returns on the risk of the optimal portfolio. As an application, we get a new approach for estimating the risk based on estimating the covariance matrices of stocks returns using the exponentially weighted moving average. Simulations show a remarkable improvement in estimating the risk of the optimal portfolio using the Scaling technique which outperforms the improvement obtained by using the Filtering technique.

We believe that the effect of noise on computing the risk and the weights of the optimal portfolio results from estimating the inverse of the covariance matrix (using the inverse of the estimator of the covariance matrix) not from estimating the covariance matrix itself. Improving the estimator of the inverse of the covariance matrix is an interesting topic which we pursue in our future work.

Chapter 6

Numerical Observations and Future Work

In Section 6.1, we present some numerical observations. We illustrate some simulations which discuss the underestimation of the risk of MV model when $T < n + 3$ and the errors in estimating the inverse of the covariance matrix. These simulations, together with some other topics, represent our current ideas for future work on this broad topic.

6.1 Numerical Observations

6.1.1 The Risk of the Optimal Portfolio ($T < n + 3$)

In this thesis, we were interested in studying the effect of estimating the covariance matrix on measuring the risk and the weights of the optimal portfolios. We cover the case $T > n+3$ (where, n denotes the number of the assets and T denotes the number of observations of the returns of the assets) and obtain the asymptotic behavior of Q, the ratio between the Predicted and the True risks. This is shown in applying Theorem 5.3.7 and Corollary 5.3.16 to different estimators of the covariance matrix in Corollary 5.4.2 and Corollary 5.5.1. As future work, we plan to study the case $T < n+3$. Since the result in Proposition 5.3.4 is valid for any values of n and T then, $\forall n, T$

$$E(Q) = \mathrm{Tr}(B_T)E(\mathrm{tr}((X^t B_T X)^{-1})).$$

To study the asymptotic behavior of Q when $T < n+3$, we need to extend the result of Matsumoto [M2] concerning the moments of the inverse of the Wishart matrices. Actually to solve this financial problem, it is enough to find the formula for the first two local moments of the inverse of real Wishart matrices when $T < n+3$. At this point, we can use the orthogonal invariance of the real Ginibre matrices to get its moments as shown in Theorem 4.6.5. Then, as in Theorem 4.6.13, we will be able to obtain the first two local moments of the real compound Wishart matrices when $T < n+3$.

We made some simulations to study this case using the following algorithm:

Algorithm 4 Algorithm for simulating the ratio between the Predicted and the True risks when $T < n+3$

Choose values for n and T such that $T < n+3$.

Choose some positive definite matrix to be the true covariance matrix Σ.

Find Σ^{-1}, the inverse of the true covariance matrix Σ, and compute the True risk of the optimal portfolio.

For 100 times:

Generate data set from the normal distribution $N(0, \Sigma)$.

Find the corresponding empirical covariance matrix $\widehat{\Sigma}$.

Find the Pseudo inverse $\widehat{\Sigma}^{-1}$ of the empirical covariance matrix $\widehat{\Sigma}$.

Compute the Predicted risk.

Scale the Predicted risk using the ratio $\frac{\mathrm{Tr}(\widehat{\Sigma}^{-1})}{\mathrm{Tr}(\Sigma^{-1})}$.

Histogram the ratio between the Predicted and the True risks before and after scaling by the ratio $\frac{\mathrm{Tr}(\widehat{\Sigma}^{-1})}{\mathrm{Tr}(\Sigma^{-1})}$.

In Figure (6.1), we simulate the case $T = n$. The left and right graphs illustrate the ratio between the Predicted and the True risks before and after scaling the Predicted risk by the factor $\frac{\mathrm{Tr}(\widehat{\Sigma}^{-1})}{\mathrm{Tr}(\Sigma^{-1})}$, respectively. In the upper graphs, we take $n = T = 50$. For these value of n and T, the means of the histograms before and after scaling equal 0.1 and 1.5, respectively. In the lower part of the figure, $n = T = 150$ and the means of the histograms before and after scaling equal 0.06 and 1.2, respectively. It is clear that there is a real improvement in estimating the risk when it is scaled by the ratio $\frac{\mathrm{Tr}(\widehat{\Sigma}^{-1})}{\mathrm{Tr}(\Sigma^{-1})}$.

In Figure (6.2), we simulate the case $T < n$. The left and right graphs illustrate the ratio between the Predicted and the True risks before and after scaling the Predicted risk by the factor $\frac{\mathrm{Tr}(\widehat{\Sigma}^{-1})}{\mathrm{Tr}(\Sigma^{-1})}$, respectively. In the upper graphs, we take $n = 55$ and $T = 50$. For these value of n and T, the means of the histograms before and after scaling equal 4×10^{-6} and 0.9, respectively. In the lower part of the figure, $n = 150$ and $T = 140$ and the means of the histograms before and after scaling equal 9×10^{-6} and 0.64, respectively. Again, there is a real improvement in estimating the risk when it is scaled by the ratio $\frac{\mathrm{Tr}(\widehat{\Sigma}^{-1})}{\mathrm{Tr}(\Sigma^{-1})}$.

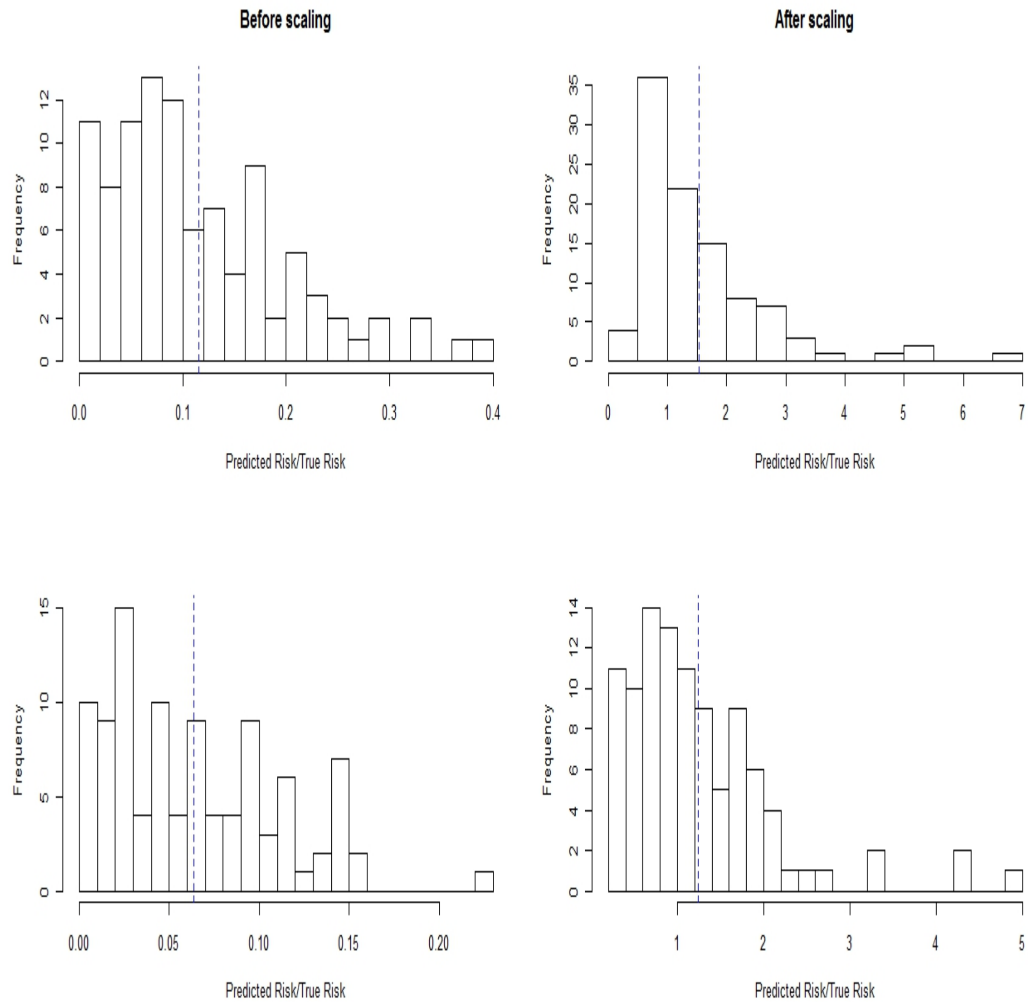


Figure 6.1: In the figure, for the MLE covariance estimator, we study the ratio between the Predicted and the True risks before and after scaling the Predicted risk using the ratio $\frac{\mathrm{Tr}(\widehat{\Sigma}^{-1})}{\mathrm{Tr}(\Sigma^{-1})}$ when n and T have the same value. For each row of the figure, it is clear that the right graph (represent the ratio between the risks after scaling) provides a better estimator of the optimal risk than the left one (represent the ratio between the risks before scaling). In the first row, we take $n = T = 50$ and the mean values of the left and right histograms equal 0.1154792 and 1.53666, respectively. For the lower graphs of the figure, $n = T = 150$ and the mean of the right histogram is 1.242206 which is still closer to 1 than the mean of the left histogram 0.06377931. Hence, the scaling of the Predicted risk provides a remarkable improvement in estimating the optimal risk.

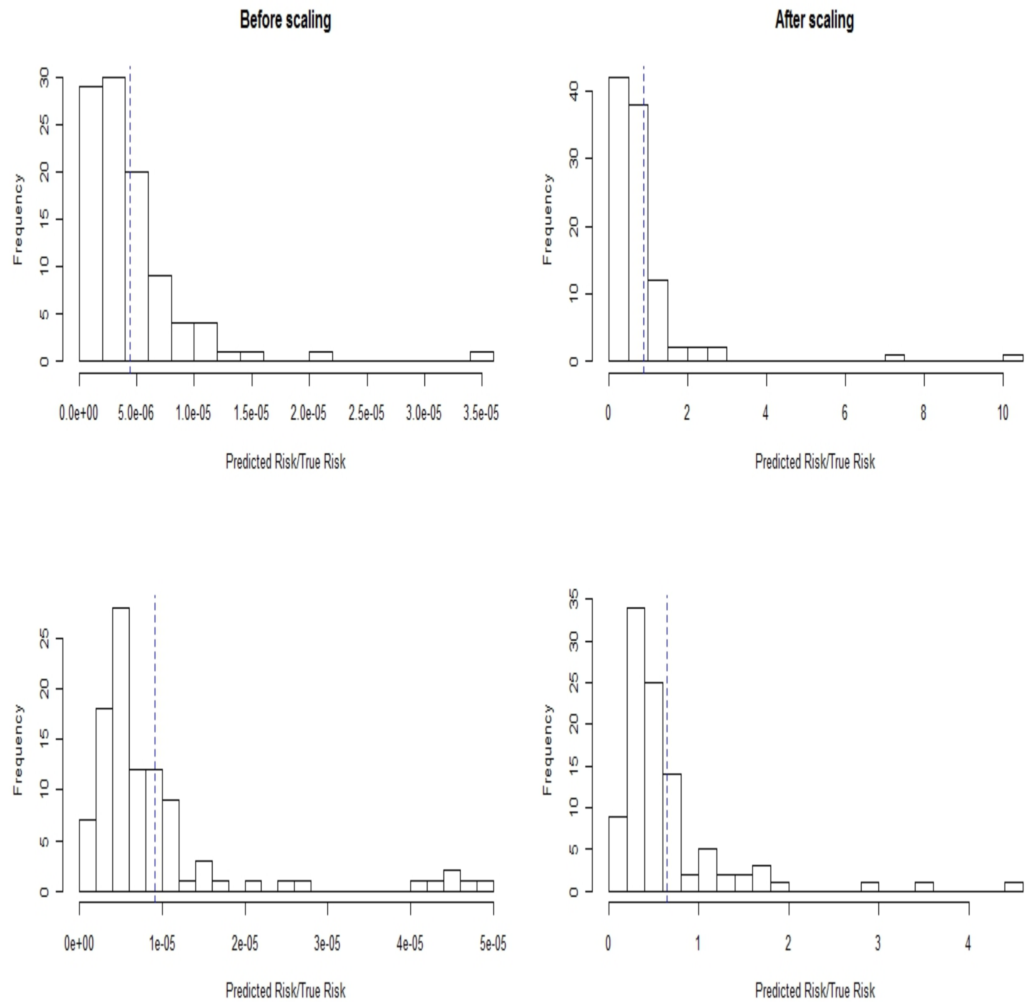


Figure 6.2: In the figure, for the MLE covariance estimator, we study the ratio between the Predicted and the True risks before and after scaling the Predicted risk using the ratio $\frac{\mathrm{Tr}(\widehat{\Sigma}^{-1})}{\mathrm{Tr}(\Sigma^{-1})}$ when $n > T$. For each row of the figure, it is clear that the right graph (represent the ratio between the risks after scaling) provides a better estimator of the optimal risk than the left one (represent the ratio between the risks before scaling). In the first row, we take $n = 55$ and $T = 50$ and the mean values of the left and right histograms equal $4.436919e - 06$ and 0.8851772, respectively. For the lower graphs of the figure, $n = 150$ and $T = 140$ and the mean of the right histogram is 0.6423968 which is still closer to 1 than the mean of the left histogram $9.146274e - 06$. Hence, the scaling of the Predicted risk still provides a remarkable improvement in estimating the optimal risk.

6.1.2 Estimation of the Inverse of the Covariance Matrix

As shown in Chapter 2, the weights and the risk of the optimal portfolio depend essentially on the entries of the precision matrix, i.e. the inverse of the covariance matrix Σ. The estimation of the precision matrix of a multivariate normal distribution has been an important issue in practical situations and is an important theoretical question. The estimation of the precision matrix is important in various statistical inference problems including the Fisher linear discriminant analysis, confidence regions based on the Mahalanobis distance and others. A standard estimator is the inverse of the sample covariance matrix, but it may be unstable or can not be defined in the high dimension. When the dimension n is smaller than the number of observations T, Efron and Morris [EM] considered this problem. But, when $T < n$, the Wishart matrix is singular, and thus many estimators can be constructed by using a generalized inverse of the sample covariance matrix. However, Srivastava [S] proposed the unique Moore-Penrose inverse of the sample covariance matrix as it uses the sufficient statistic for Σ. In this section and using the results of the Scaling technique we try to improve the estimator of the inverse of the covariance matrix.

The ratio Q between the True risk and the Predicted risk can be written as the ratio between the two quadratic forms $\mathbf{e}^t\widehat{\Sigma}^{-1}\mathbf{e}$ and $\mathbf{e}^t\Sigma^{-1}\mathbf{e}$ (where $\mathbf{e}$ is an $n-$ dimensional vector whose entries are ones). We want to study the performance of using our result in Corollary 5.3.16 concerning the asymptotic behavior of Q to improve the estimation of the inverse $\widehat{\Sigma}^{-1}$ of the covariance matrix. We make some simulations to see the effect of scaling the inverse of the empirical covariance matrix by the factor f given by

$$f = \lim_{n,T\to\infty} \frac{T(T-n-1)}{\operatorname{Tr}(B)\operatorname{Tr}(B^{-1})}, \tag{6.1.1}$$

on improving the estimator of the inverse of the covariance matrix. In these simulations, we define

$$D_1 = \|\widehat{\Sigma}^{-1} - \Sigma^{-1}\|, \tag{6.1.2}$$

and

$$D_2 = \|f\widehat{\Sigma}^{-1} - \Sigma^{-1}\|. \tag{6.1.3}$$

where $\|.\|$ denotes the Hilbert Schmidt norm. So,

$$D_1 = \sqrt{\mathrm{Tr}((\widehat{\Sigma}^{-1} - \Sigma^{-1})^t(\widehat{\Sigma}^{-1} - \Sigma^{-1}))}, \tag{6.1.4}$$

and

$$D_2 = \sqrt{\mathrm{Tr}((f\widehat{\Sigma}^{-1} - \Sigma^{-1})^t(f\widehat{\Sigma}^{-1} - \Sigma^{-1}))}. \tag{6.1.5}$$

Remark 6.1.1 *D_1 and D_2 represent the errors in estimating the inverse of the covariance matrix using the estimators $\widehat{\Sigma}^{-1}$ and $f\widehat{\Sigma}^{-1}$, respectively.*

Remark 6.1.2 *According to Corollary 5.4.2, if we are dealing with the MLE of the covariance matrix then, $f = 1 - n/T$. For the EWMA, the factor f equal $(e^c - 1)^2/c^2(1-r)e^c$ as in Corollary 5.5.1.*

To simulate the error in estimating the inverse of the covariance matrix before and after scaling the inverse of the empirical covariance matrix by the factor f, let us use the following algorithm:

Algorithm 5 Simulation of the errors D_1 and D_2 in estimating the inverse of the covariance matrix.

Choose some values for n and T such that $T > n+3$ and define f according to the definition of the estimator of the covariance matrix.

Define Σ as some positive definite matrix and find its inverse Σ^{-1}.

for $1 \leq m \leq 100$ **do**

Construct T random vectors each of n dimensions and from the distribution $N(0, \Sigma)$.

Using the data, find the empirical covariance matrix $\widehat{\Sigma}$ and then find its inverse $\widehat{\Sigma}^{-1}$

Evaluate $D_1 = \sqrt{\text{Tr}((\widehat{\Sigma}^{-1} - \Sigma^{-1})^t(\widehat{\Sigma}^{-1} - \Sigma^{-1}))}$

let $f = 1 - rr$ and evaluate

$D_2 = \sqrt{\text{Tr}((f * \widehat{\Sigma}^{-1} - \Sigma^{-1})^t(f * \widehat{\Sigma}^{-1} - \Sigma^{-1}))}$

end for

Histogram the errors D_1 and D_2.

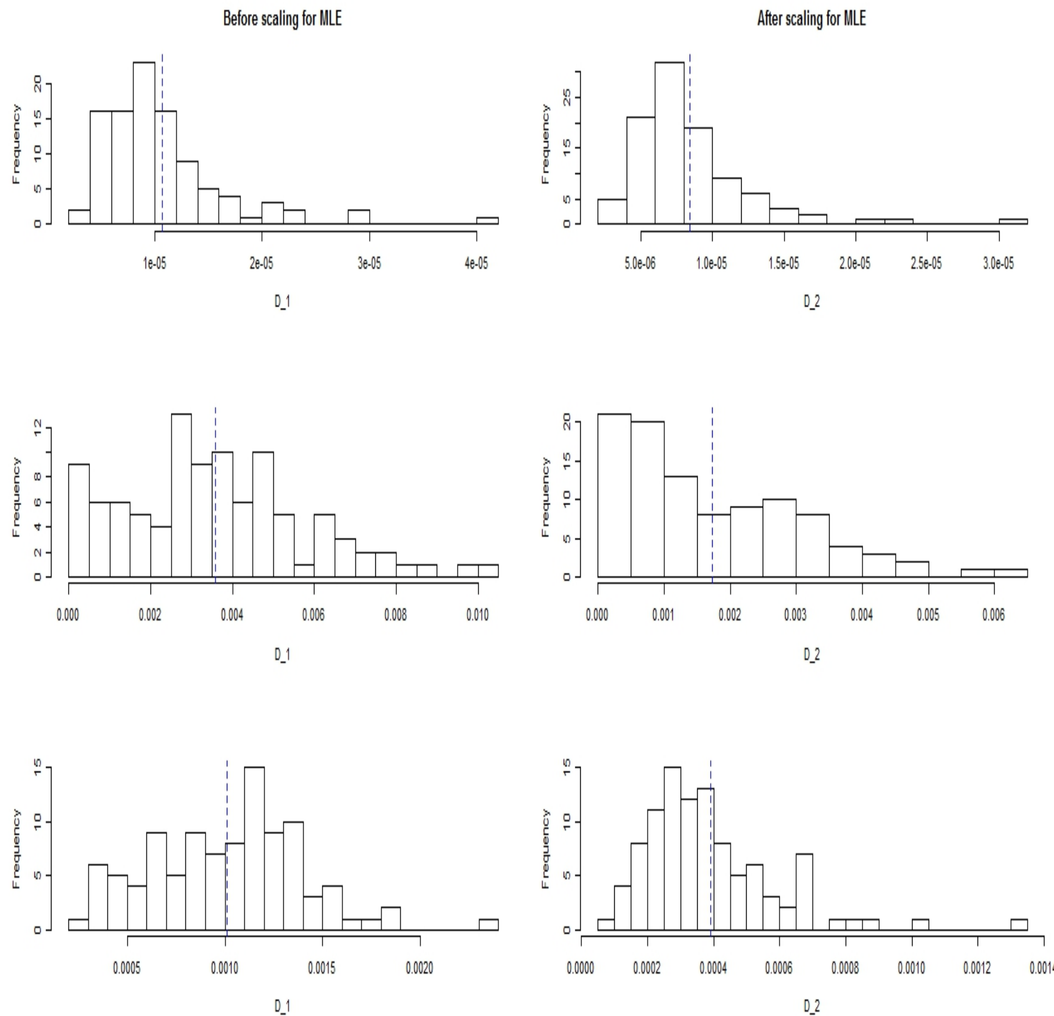


Figure 6.3: In the figure, for the MLE $\widehat{\Sigma}$, we study the errors D_1 and D_2 in estimating the inverse of the covariance matrix before and after scaling the inverse of the MLE covariance by the factor $1 - n/T$. For each row of the figure, it is clear that the right graph provides a smaller error D_2 (after scaling) than the left one D_1 (before scaling). In the first row, we take $n = 10$ and $T = 100$ and the mean values of D_1 and D_2 equal $1.071567e - 05$ and $8.431632e - 06$, respectively. In the middle graphs, $n = 50$ and $T = 500$ and the mean of $D_1 = 0.003579674$ while the mean of $D_2 = 0.001729605$, For the lower graphs of the figure, $n = 100$ and $T = 1000$ and the mean of D_1 is 0.001011469 which is still greater than the mean of D_2 which equals 0.0003901861.

As shown in Figure (6.3), the error D_1 in estimating Σ^{-1} using the inverse of the MLE covariance, represented in the left graphs of the figure, is greater than the error D_2 in estimating Σ^{-1} after scaling the inverse of the MLE covariance by the factor $1 - n/T$.

In Figure (6.4), for the EWMA covariance estimator, if n has large values $n = 50$ or $n = 100$, then the error D_2 is smaller than the error D_1. While for small value of n for example $n = 10$, the error D_1 is less than the error D_2.

These observations will be interesting to study in our future work.

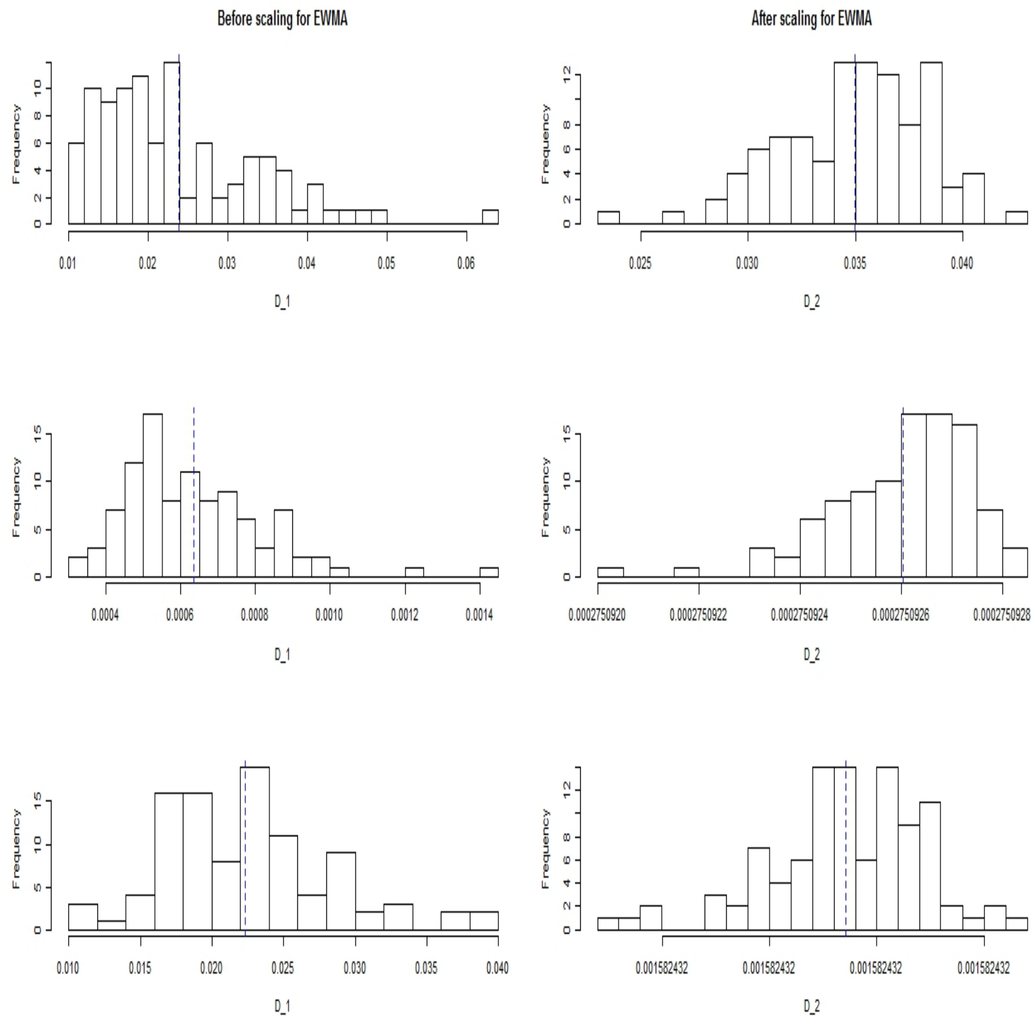


Figure 6.4: In the figure, we study the errors D_1 and D_2 in estimating the inverse of the covariance matrix when $\widehat{\Sigma}$ is the EWMA. For each row of the figure, it is clear that the right graphs (representing the error after scaling) provide a smaller error than the left ones. In the first row, we take $n = 10$ and $T = 100$ and the means of the errors are 0.02384064 (before scaling) and 0.03496347 (after scaling). In the middle graphs, $n = 50$ and $T = 500$. The means of the errors D_1 and D_2 equal 0.0006338242 and 0.0002750926, respectively. The lower graphs of the figure illustrate the values $n = 100$ and $T = 1000$. The means of the errors D_1 and D_2 equal 0.02235427 and 0.001582432, respectively.

6.2 Future Work

As future work, we first would like to analyze the above simulations using techniques from RMT. We strongly aim to focus on our results concerning the optimal weights. We would like to exploit these results to obtain a formula of the asymptotically unbiased consistent estimators of the optimal weights as the number of the assets n and the number of observations T tend to infinity.

There are many other topics that we plan to study. We aim to apply the random matrix results of Chapter 4 to study other models of covariance matrices such as the GARCH model which plays an important role in Finance. It would of course be of interest to extend this work to the return constrained Markowitz formulation. We'd begin that study in its simplest, no short sale constraint, form.

Also, we hope to get the exact distribution of the ratio Q between the Predicted and the True risks. To do that we will try to extend the work of El Karoui in [El-K] by extending some properties of the Wishart matrices to the compound Wishart matrices.

On the other hand side, it is also interesting to use RMT to study the effect of the noise induced by estimating the mean of the returns on estimating the risk and the weights of the optimal portfolio.

As an application to the results of Chapter 4, we aim to extend the result of Marceko and Pastur in [MP] and study the limiting spectral distribution of the compound Wishart matrices and of the inverse compound Wishart matrices.

Appendix A

R Language Codes for Algorithms

R Language code to Simulate the ratio between the Predicted and the True risks before and after applying the Scaling technique for the MLE covariance matrix:

```
X<-matrix(c(rep(0,100)),100,1)
Y<-matrix(c(rep(0,100)),100,1)

n<-200  #number of assets
T<-250  #number of observations
r<-n/T


# Construct the weighted matrix B and compute its trace.
B<-diag(c(rep(1,T)),T,T)
TrB<-0
for(i in 1:T){
TrB<-TrB+B[i,i]}


# Construct some covariance matrix and its inverse.
S<-matrix(c(rep(1:n^2)),n,n)
for(i in 1:n)
{
S[i,i]<-runif(1,min=-400,max=1000)
}
Sigma<-S%*%t(S)
Inv1<-solve(Sigma)


# Compute the True risk.
True<-0
for(i in 1:n){
for(j in 1:n){
True<-True+Inv1[i,j]}}
True<-1/sqrt(True)


for(m in 1:100){


# Generate data set.
library(MASS)
D<-mvrnorm(T,rep(0,n),Sigma)
data<-t(D)


# Construct the MLE of the covariance matrix and its inverse.
Hat<-(1/(TrB))*(data%*%B%*%t(data))
Inv2<-solve(Hat)


# Compute the Predicted risk.
Predict=0
for(j in 1:n)
{
for(u in 1:n)
{
Predict=Predict+Inv2[j,u]
```

```
}
}
Predict=1/sqrt(Predict)


#Compute the Predicted risk after scaling.
Q<-1/sqrt(1-r)
Scale<-Q*Predict

X[m,1]<-Predict/True
Y[m,1]<-Scale/True

}


# Histogram the ratio between the Predicted and the True risks before and after scaling and draw the means of the histograms.
par(mfrow=c(1,2))
hist(X,breaks=20,main=paste("Before scaling"),xlab="square root of Q")
a<-mean(X)
abline(v=c(a),lty=2,col="blue")
hist(Y,breaks=20,main=paste("After scaling"),xlab="square root of Q")
aa<-mean(Y)
abline(v=c(aa),lty=2,col="blue")
```

R Language code to Simulate the ratio between the Predicted and the True risks before and after applying the Scaling technique for the SCM:

```
X<-matrix(c(rep(0,100)),100,1)
Y<-matrix(c(rep(0,100)),100,1)
n<-300
T<-320
rr<-(n-1)/(T-1)


# Construct the weighted matrix B and compute its trace.
B<-diag(c(rep(1,T)),T,T)
for(i in 1:T){
for(j in 1:T){
B[i,j]=-1/T}}
for(i in 1:T){
B[i,i]<-1-(1/T)}
TrB<-0
for(i in 1:T){
TrB<-TrB+B[i,i]}


# Construct some covariance matrix and its inverse.
S<-matrix(c(rep(1:n^2)),n,n)
for(i in 1:n)
{
S[i,i]<-runif(1,min=-1000,max=1000)
}
Sigma<-S%*%t(S)
Inv1<-solve(Sigma)


# Compute the True risk.
True<-0
for(i in 1:n){
for(j in 1:n){
True<-True+Inv1[i,j]}}
True<-1/sqrt(True)


for(m in 1:100){


# Generate data set.
library(MASS)
D<-mvrnorm(T,rep(0,n),Sigma)
data<-t(D)


# Construct the SCM and its inverse.
Hat<-(1/(TrB))*(data%*%B%*%t(data))
Inv2<-solve(Hat)


# Compute the Predicted risk.
Predict=0
for(j in 1:n)
```

```
{
for(u in 1:n)
{
Predict=Predict+Inv2[j,u]
}
}
Predict=1/sqrt(Predict)


#Compute the Predicted risk after scaling.
Q<-1/sqrt(1-rr)
Scale<-Q*Predict

X[m,1]<-Predict/True
Y[m,1]<-Scale/True

}


# Histogram the ratio between the Predicted and the True risks before and after scaling and draw the means of the histograms.
par(mfrow=c(1,2))
hist(X,breaks=20,main=paste("Before scaling"),xlab="square root of Q")
a<-mean(X)
abline(v=c(a),lty=2,col="blue")
hist(Y,breaks=20,main=paste("After scaling"),xlab="square root of Q")
aa<-mean(Y)
abline(v=c(aa),lty=2,col="blue")
```

R Language code to Simulate the ratio between the Predicted and the True risks before and after applying the Scaling technique for the EWMA covariance matrix:

```
X<-matrix(c(rep(0,100)),100,1)
Y<-matrix(c(rep(0,100)),100,1)
n<-400
T<-410
l<-0.96
r=p/n
h<-(1-l)*T


# Construct the weighted matrix B and compute its trace.
B<-diag(c(rep(1,T)),T,T)
for(i in 1:T)
{
B[i,i]<-l^(i-1)
}
TrB<-0
for(i in 1:T){
TrB<-TrB+B[i,i]}


# Construct some covariance matrix and its inverse.
S<-matrix(c(rep(1:n^2)),n,n)
for(i in 1:n)
{
S[i,i]<-runif(1,min=-1000,max=1000)
}
Sigma<-S%*%t(S)
Inv1<-solve(Sigma)


# Compute the True risk.
True<-0
for(i in 1:n){
for(j in 1:n){
True<-True+Inv1[i,j]}}
True<-1/sqrt(True)


for(m in 1:100){


# Generate data set.
library(MASS)
D<-mvrnorm(T,rep(0,n),Sigma)
data<-t(D)


# Construct the EWMA covariance matrix and its inverse.
Hat<-(1/(TrB))*(data%*%B%*%t(data))
Inv2<-solve(Hat)


# Compute the Predicted risk.
Predict=0
```

```
for(j in 1:n)
{
for(u in 1:n)
{
Predict=Predict+Inv2[j,u]
}
}
Predict=1/sqrt(Predict)


#Compute the Predicted risk after scaling.
Q<-((exp(h)-1)^2)/((h^2)*(1-r)*exp(h))
Scale<-Q*Predict
X[m,1]<-Predict/True
Y[m,1]<-Scale/True

}


# Histogram the ratio between the Predicted and the True risks before and after scaling and draw the means of the histograms.
par(mfrow=c(1,2))
hist(X,breaks=20,main=paste("Before scaling"),xlab="square root of Q")
a<-mean(X)
abline(v=c(a),lty=2,col="blue")
hist(Y,breaks=20,main=paste("After scaling"),xlab="square root of Q")
aa<-mean(Y)
abline(v=c(aa),lty=2,col="blue")
```

R Language code to compare between the Filtering and the Scaling techniques:

```
X<-matrix(c(rep(0,100)),100,1)
Y<-matrix(c(rep(0,100)),100,1)

n<-50
T<-100
r<-n/T
max<-(1+sqrt(r))^2  #the maximum noisy eigenvalue
min<-(1-sqrt(r))^2  #the minimum noisy eigenvalue


# Define the weighted matrix for the MLE covariance and find its trace.
B<-diag(c(rep(1,n)),n,n)
TrB<-0
for(i in 1:T){
TrB<-TrB+B[i,i]}


# Choose a covariance matrix and find its inverse.
Sigma<-diag(c(rep(1,n)),n,n)
Inv1<-solve(Sigma)


# Compute the True risk.
True<-0
for(i in 1:n){
for(j in 1:n){
True<-True+Inv1[i,j]}}
True<-1/(True)


# Generate data set, then find the empirical covariance matrix (using MLE) and its inverse.
for(m in 1:100){
library(MASS)
D<-mvrnorm(n,rep(0,p),Sigma)
data<-t(D)
Hat<-(1/(TrB))*(data%*%B%*%t(data))
Inv2<-solve(Hat)


# Compute the Predicted risk.
Predict=0
for(j in 1:p)
{
for(u in 1:p)
{
Predict=Predict+Inv2[j,u]
}
}
Predict=1/Predict


# Compute the Predicted risk after using the Scaling technique.
Q<-1/(1-r)
Scale<-Q*Predict
Y[m,1]<-sqrt(Scale)/sqrt(True)
```

```
# Find the corresponding empirical correlation matrix and its eigenvalues.
Cov1<-Hat
Corr1<-Cov1
for(i in 1:n){
for(j in 1:n){
Corr1[i,j]<-Corr1[i,j]/(sqrt(Cov1[i,i]*Cov1[j,j]))}}
ceval<-eigen(Corr1)$values
V<-eigen(Corr1)$vectors


# Replace the noisy eigenvalues by its average value and construct the cleaning correlation matrix.
nsum<-0
for(i in 1:n)
{
if(min>ceval[i]|ceval[i]>max){nsum=nsum}else{nsum=nsum+ceval[i]}
}
s=0
for(i in 1:n)
{
if(min>ceval[i]|ceval[i]>max){s=s}else{s=s+1}
}
nsum<-nsum/s
for(i in 1:n)
{
if(min>ceval[i]|ceval[i]>max){ceval[i]=ceval[i]}else{ceval[i]=nsum}
}
A<-diag(ceval,n,n)
Corr2<-V%*%A%*%t(V)


# Constuct the cleaning covariance matrix which corresponds to the cleaning correlation matrix:
Cov2<-Corr2
for(i in 1:n){
for(j in 1:n){
Cov2[i,j]<-Corr2[i,j]*sqrt(Cov1[i,i]*Cov1[j,j])}}
Inv3<-solve(Cov2)


# Compute the Predicted risk after using the Filtering technique.
Predictf=0
for(j in 1:p)
{
for(u in 1:p)
{
Predictf=Predictf+Inv3[j,u]
}
}
Predictf=1/Predictf
filtered<-Predictf

X[m,1]<-sqrt(filtered)/sqrt(True)
}


# Histogram the ratio between the Predicted and the True risks after using the Filtring and the Scaling techniques then draw the means of the histograms.
```

```
par(mfrow=c(1,2))
hist(X,breaks=20,main=paste("With Filtering Technique"),xlab="Predicted Risk/True Risk")
a<-mean(X)
abline(v=c(a),lty=2,col="blue")

hist(Y,breaks=20,main=paste("With Scaling Technique"),xlab="Predicted Risk/True Risk")
aa<-mean(Y)
abline(v=c(aa),lty=2,col="blue")
```

R Language code to simulate the ratio between the Predicted risk and the True risk when (T<n+3) for the MLE before and after scaling:

```
X<-matrix(c(rep(0,100)),100,1)
Y<-matrix(c(rep(0,100)),100,1)

n<-150  #number of assets
T<-140  #number of observations


# Construct some positive definite matrix to be the true covariance matrix of the returns.
S<-matrix(c(rep(1:n^2)),n,n)
for(i in 1:n)
{
S[i,i]<-runif(1,min=-400,max=1000)
}
Sigma<-S%*%t(S)


# Find the inverse of the true covariance matrix and compute its trace.
Inv1<-solve(Sigma)

Tr1<-0
for(i in 1:n){
Tr1<-Tr1+Inv1[i,i]}


# Compute the True risk.
True<-0
for(i in 1:n){
for(j in 1:n){
True<-True+Inv1[i,j]}}
True<-1/(True)



for(m in 1:100){
library(MASS)

# Generate data set and use the MLE to construct the corresponding empirical covariance matrix.
D<-mvrnorm(n,rep(0,p),Sigma)
data<-t(D)
B<-diag(c(rep(1,T)),T,T)
TrB<-0
for(i in 1:T){
TrB<-TrB+B[i,i]}
Hat<-(1/(TrB))*(data%*%B%*%t(data))


# Find the Pseudo inverse of the empirical covariance matrix and compute its trace.
a2<-eigen(Hat)$values
D2<-diag(a2,p,p)
U2<-eigen(Hat)$vectors
for(i in 1:p)
{
if(D2[i,i]<0|D2[i,i]>0){D2[i,i]<-1/D2[i,i]}
}
```

```
Inv2<-U2%*%D2%*%t(U2)

Tr2<-0
for(i in 1:p){
Tr2<-Tr2+Inv2[i,i]}


# Compute the Predicted risk before and after scaling using the ratio between the traces of the inverses of the empirical and the true covariance matrices, respectively.
Predict=0
for(j in 1:p)
{
for(u in 1:p)
{
Predict=Predict+Inv2[j,u]
}
}
Predict=1/abs(Predict)
Q<-Tr2/Tr1
Scale<-abs(Q*Predict)
X[m,1]<-sqrt(Predict)/sqrt(True)
Y[m,1]<-sqrt(Scale)/sqrt(True)
}


# Histogram the ratio between the Predicted and the True risks before and after scaling.
par(mfrow=c(1,2))

hist(X,breaks=20,main=paste("Before scaling"),xlab="Predicted Risk/True Risk")
b<-mean(X)
abline(v=c(b),lty=2,col="blue")

hist(Y,breaks=20,main=paste("After scaling"),xlab="Predicted Risk/True Risk")
bb<-mean(Y)
abline(v=c(bb),lty=2,col="blue")
```

R Language code to simulate the error in estimating the inverse of the covariance matrix for the MLE before and after scaling:

```
X<-matrix(c(rep(0,100)),100,1)
Y<-matrix(c(rep(0,100)),100,1)


n<-50
T<-1000
r<-n/T


# Construct the weighted matrix and find its trace.
B<-diag(c(rep(1,T)),T,T)
TrB<-0
for(i in 1:T){
TrB<-TrB+B[i,i]}


# Choose a covariance matrix and find its inverse.
S<-matrix(c(rep(1:n^2)),n,n)
for(i in 1:n)
for(j in 1:n)
{
S[i,j]<-runif(1,min=-400,max=1000)
}
Sigma<-S%*%t(S)
Inv1<-solve(Sigma)


# Generate the data set and the MLE empirical covariance matrix and its inverse.
for(m in 1:100){
library(MASS)
D<-mvrnorm(T,rep(0,n),Sigma)
data<-t(D)
Hat<-(1/TrB)*(data%*%B%*%t(data))
Inv2<-solve(Hat)


# Compute the error in estimating the inverse of the covariance matrix.
H2<-Inv1-Inv2
J2<-H2%*%t(H2)
Inorm2<-0
for(i in 1:n)
{
Inorm2<-Inorm2+J2[i,i]
}
Inorm2<-sqrt(Inorm2)
X[m,1]<-Inorm2


# Scale the inverse of the empirical MLE covariance matrix and comute the error in estimating the inverse of the covariance
matrix after Scaling:
f<-1-r
Inv22<-f*Inv2

HH2<-Inv1-Inv22
```

```
JJ2<-HH2%*%t(HH2)
Inorm22<-0
for(i in 1:n)
{
Inorm22<-Inorm22+JJ2[i,i]
}
Inorm22<-sqrt(Inorm22)

Y[m,1]<-Inorm22

}


# Histogram the errors in estimating the inverse of the covariance matrix before and after Scaling:
par(mfrow=c(1,2))

hist(X,breaks=20,main=paste("Before scaling for MLE"),xlab="D_1")
a<-mean(X)
abline(v=c(a),lty=2,col="blue")

hist(Y,breaks=20,main=paste("After scaling for MLE"),xlab="D_2")
aa<-mean(Y)
abline(v=c(aa),lty=2,col="blue")
```

R Language code to simulate the error in estimating the inverse of the covariance matrix for the SCM before and after scaling:

```
X<-matrix(c(rep(0,100)),100,1)
Y<-matrix(c(rep(0,100)),100,1)


n<-100  #number of variables
T<-1000 #number of observations
rr<-(n-1)/(T-1)


# Construct the weighted matrix and find its trace.
B<-diag(c(rep(1,T)),T,T)
for(i in 1:T){
for(j in 1:T){
B[i,j]=-1/T}}
for(i in 1:T){
B[i,i]<-1-(1/T)}
TrB<-0
for(i in 1:T){
TrB<-TrB+B[i,i]}


# Construct some positive definite matrix to be the true covariance matrix and find its inverse.
S<-matrix(c(rep(1,n^2)),n,n)
for(i in 1:n)
for(j in 1:n)
{
S[i,j]<-runif(1,min=-400,max=1000)
}
Sigma<-S%*%t(S)
Inv1<-solve(Sigma)


# Generate a data set from the distribution N(0,Sigma) and find the empirical SCM covariance matrix and its inverse.
for(m in 1:100){
library(MASS)
D<-mvrnorm(n,rep(0,p),Sigma)
data<-t(D)
Hat<-(1/TrB)*(data%*%B%*%t(data))
Inv2<-solve(Hat)


# Compute the error in estimating the inverse of the covariance matrix
H2<-Inv1-Inv2
J2<-H2%*%t(H2)
Inorm2<-0
for(i in 1:n)
{
Inorm2<-Inorm2+J2[i,i]
}
Inorm2<-sqrt(Inorm2)
X[m,1]<-Inorm2
```

```
# Scale the inverse of the SCM empirical covariance matrix and compute the error in estimating the inverse of the covariance matrix after scaling.
f<-1-rr
Inv22<-f*Inv2
HH2<-Inv1-Inv22
JJ2<-HH2%*%t(HH2)
Inorm22<-0
for(i in 1:n)
{
Inorm22<-Inorm22+JJ2[i,i]
}
Inorm22<-sqrt(Inorm22)
Y[m,1]<-Inorm22

}


# Histogram the errors in estimating the inverse of the covariance matrix before and after scaling
par(mfrow=c(1,2))

hist(X,breaks=20,main=paste("Before scaling for SCM"),xlab="Error D_1")
a<-mean(X)
abline(v=c(a),lty=2,col="blue")

hist(Y,breaks=20,main=paste("After scaling for SCM"),xlab="Error D_2")
aa<-mean(Y)
abline(v=c(aa),lty=2,col="blue")
```

R Language code to simulate the error in estimating the inverse of the covariance matrix for the EWMA before and after scaling:

```
X<-matrix(c(rep(0,100)),100,1)
Y<-matrix(c(rep(0,100)),100,1)


n<-100  #number of variables
T<-1000 #number of observations
l<-0.96 #decay factor
r=p/n
h<-(1-l)*T


# Construct the weighted matrix and find its trace.
B<-diag(c(rep(1,n)),n,n)
for(i in 1:n)
{
B[i,i]<-l^(i-1)
}
for(i in 1:T){
B[i,i]<-1-(1/T)}
TrB<-0
for(i in 1:T){
TrB<-TrB+B[i,i]}


# Construct some positive definite matrix to be the true covariance matrix and find its inverse.
S<-matrix(c(rep(1,n^2)),n,n)
for(i in 1:n)
for(j in 1:n)
{
S[i,j]<-runif(1,min=-400,max=1000)
}
Sigma<-S%*%t(S)
Inv1<-solve(Sigma)


# Generate a data set from the distribution N(0,Sigma) and find the empirical SCM covariance matrix and its inverse.
for(m in 1:100){
library(MASS)
D<-mvrnorm(n,rep(0,p),Sigma)
data<-t(D)
Hat<-(1/TrB)*(data%*%B%*%t(data))
Inv2<-solve(Hat)


# Compute the error in estimating the inverse of the covariance matrix
H2<-Inv1-Inv2
J2<-H2%*%t(H2)
Inorm2<-0
for(i in 1:n)
{
Inorm2<-Inorm2+J2[i,i]
}
Inorm2<-sqrt(Inorm2)
X[m,1]<-Inorm2
```

```
# Scale the inverse of the SCM empirical covariance matrix and compute the error in estimating the inverse of the covariance matrix after scaling.
f<-((h^2)*(1-r)*exp(h))/((exp(h)-1)^2)
Inv22<-f*Inv2
HH2<-Inv1-Inv22
JJ2<-HH2%*%t(HH2)
Inorm22<-0
for(i in 1:n)
{
Inorm22<-Inorm22+JJ2[i,i]
}
Inorm22<-sqrt(Inorm22)
Y[m,1]<-Inorm22

}


#Histogram the errors in estimating the inverse of the covariance matrix before and after scaling
par(mfrow=c(1,2))

hist(X,breaks=20,main=paste("Before scaling for EWMA"),xlab="Error D_1")
a<-mean(X)
abline(v=c(a),lty=2,col="blue")

hist(Y,breaks=20,main=paste("After scaling for EWMA"),xlab="Error D_2")
aa<-mean(Y)
abline(v=c(aa),lty=2,col="blue")
```

Bibliography

[Ak] V. Akgiray, 1989. *Conditional Heteroscedasticity in Time Series of Stock Returns: Evidence and Forecasts,* Journal of Business, 62: pp. 55-80.

[Akh] N. I. Akhiezer, 1965. *The Classical Moment Problem and Some Related Questions in Analysis,* Oliver & Boyd.

[AGZ] G. W. Anderson, A. Guionnet, O. Zeitouni, 2009. *An Introduction to Random Matrices,* Cambridge University Press.

[Bai] Z. D. Bai, 1999. *Methodologies in Spectral Analysis of Large Dimensional Random Matrices,* a review. Statistica Sinica, 9: pp. 611-677.

[B] P. Billingsley, 1995. *Probability and Measure,* 3rd edn, J. Wiley & Sons, Inc.

[BiBouP] G. Biroli, J.-P. Bouchaud, M. Potters, 2007. *The Student Ensemble of Correlation Matrices,* Acta Phys. Pol., B 38.

[Bol] T. Bollerslev, 1986. *Generalised Autoregressive Conditional Heteroscedasticity,* Journal of Econometrics, 31: pp. 307-327.

[BJJNPZ] Z. Burda, A. Jarosz, J. Jurkieewicz, M. A. Nowak, G. Papp, and I. Zahed, 2011. *Applying Free Random Variables to Random Matrix Analysis of Financial Data,* Quant. Finance, 11 (7): pp. 1103-1124.

[CW] C. M. Carvalho, M. West, 2007. *Dynamic Matrix-Variate Graphical Models,* Bayesian Analysis, 2: pp. 69-98.

[CMW] C. M. Carvalho, H., Massam, M. West, 2007. *Simulation of Hyper-Inverse Wishart Distributions in Graphical Models,* Biometrika, 94: pp. 647-659.

[Coch] W. G. Cochran, 1934. *The Distribution of Quadratic Forms in a Normal System with Applications to the Analysis of Variance,* Proc. Camb. Phil. Soc., 30: pp. 178-191.

[C] B. Collins, 2003. *Moments and Cumulants of Polynomial Random Variables on Unitary Groups, the Itzykson-Zuber Integral, and Free Probability,* Int. Math. Res. Not., 17: pp. 953-982.

[CMcS] B. Collins, D. McDonald, N. Saad. *Compound Wishart Matrices and Noisy Covariance Matrices: Risk Underestimation,* Preprint.

[CM] B. Collins and S. Matsumoto, 2009. *On Some Properties of Orthogonal Weingarten Functions,* J. Math. Phys., 50 (11).

[CMS] B. Collins, S. Matsumoto and N. Saad, 2012. *Integration of Invariant Matrices and Application to Statistics,* http://arxiv.org/abs/1205.0956.

[CS] B. Collins and P. Śniady, 2006. *Integration With Respect to the Haar Measure on Unitary, Orthogonal and Symplectic Group,* Comm. Math. Phys., 264 (3): pp. 773-795.

[D] M. Davison, Private Communications.

[El-K] N. El Karoui, 2009. *High Dimensionality Effects in The Markowitz Problem and Other Quadratic Programs with Linear Equality Constraints: Risk Underestimation,* Technical Report 781, Department of Statistics, UC Berkeley.

[EG1] E.J. Elton, M.J. Gruber, 1974. *Portfolio Theory when Investment Relatives are Lognormally Distributed,* Journal of Finance, 29: pp. 1265-1273.

[EG2] E.J. Elton, M.J. Gruber, 1995. *Modern Portfolio Theory and Investment Analysis,* J. Wiley, New York.

[EG3] E.J. Elton, M.J. Gruber, 1997. *Modern Portfolio Theory, 1950 to Date,* Journal of Banking & Finance, 21: pp. 1743-1759.

[EM] Efron, B. and C. Morris, 1976. *Multivariate Empirical Bayes Estimation of Covariance Matrices,* Ann. Statist., 4: pp. 22-32.

[F] E. Fama, 1965. *The Behaviour of Stock Market Prices,* Journal of Business, 38: pp. 34-105.

[GBP] G.Galluccio, J.-P. Bouchaud, and M. Potters, 2000. *Random Matrix Theory and Financial Correlations,* International Journal of Theoretical and Applied Finance, 3 (3): pp. 391-397.

[GLM] P. Graczyk, P. Letac, and H. Massam, 2003. *The Complex Wishart Distribution and the Symmetric Group,* Ann. Statist., 31: pp. 287-309.

[HP] F. Hiai, and D. Petz, 2000. *The Semicircle Law, Free Random Variables and Entropy,* American Mathematical Society, Providence, RI, vol. 77 of Mathematical Surveys and Monographs.

[J] I. M. Johnstone, 2006. *High Dimensional Statistical Inference and Random Matrices,* Proc ICM, 1: pp. 307333.

[K] H. Konno, 1990. *Piecewise Linear Risk Functions and Portfolio Optimization,* Journal of Operations Research Society of Japan, 33: pp. 139-159.

[KSY] H. Konno, H. Shirakawa, H. Yamazaki, 1993. *A Mean-Absolute Deviation-Skewness Portfolio Optimization Model,* Annals of Operations Research, 45: pp. 205-220.

[LCBP1] L. Laloux, P. Cizeau, J.-P. Bouchaud and M. Potters, 1999. *Noise Dressing of Financial Correlation Matrices* Phys. Rev. Lett., 83(7): pp. 14671470.

[LCBP2] L. Laloux, P. Cizeau, J.-P. Bouchaud and M. Potters, 2000. *Random Matrix Theory and Financial Correlations* International Journal of Theoretical and Applied Finance, 3 (3): pp. 391397.

[Lu] D. G. Luenberger, 1997. *Investment Science,* Oxford Univ. Press.

[Mac] I. G. Macdonald, 1995. *Symmetric Functions and Hall Polynomials,* 2nd ed., Oxford University Press, Oxford.

[MP] V. Marčenko, L. Pastur, 1967. *Distribution of Eigenvalues for Some Sets of Random Matrices* Math. USSR-Sb, 1: pp. 457-483.

[MKB] K. V. Mardia, J. T. Kent and J.M. Bibby, 1979. *Multivariate analysis* Academic Press [Harcourt Brace Jovanovich Publishers], London. Probability and Mathematical Statistics: A Series of Monographs and Textbooks.

[Mark] H. Markowitz, 1952. *Portfolio Selection,* Journal of Finance, 7: pp. 77-91.

[Mark2] H. Markowitz, 1959. *Portfolio Selection: Efficient Diversification of Investments,* J. Wiley, New York.

[M1] S. Matsumoto, 2001. *Jucys-Murphy Elements, Orthogonal Matrix Integrals, and Jack Measures,* The Ramanujan J., 26: pp. 69-107.

[M2] S. Matsumoto, *General Moments of the Inverse Real Wishart Distribution and Orthogonal Weingarten Functions,* arXiv1004.4717v3.

[M3] S. Matsumoto, *General Moments of Matrix Elements from Circular Orthogonal Ensembles,* Random Matrices: Theory and Applications, 1 (3).

[MN] S. Matsumoto and J. Novak, 2012. *Jucys-Murphy Elements and Unitary Matrix Integrals,* International Mathematics Research Notices, rnr 267.

[NiSp] A. Nica, R. Speicher 2006. *Lectures on the Combinatorics of Free Probability,* London Mathematical Society Lecture Note Series, 335. Cambridge University Press, Cambridge.

[PB] P. Penza and V. K. Bansal, 2001. *Measuring Market Risk with Value at Risk,* John Wiley & Sons, Canada.

[PK] S. Pafka, and I. Kondor, 2003. *Noisy Covariance Matrices and Portfolio Optimization II,* Physica A 319, pp. 487-494.

[PGRAS] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral and H. E. Stanley, 1999. *Universal and Nonuniversal Properties of Cross Correlations in Financial Time Series,* Phys. Rev. Lett., 83 (7).

[PGRAGS] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, T. Guhr, and H.E. Stanley, 2002. *Random Matrix Approach to Cross Correlations in Financial Data.* Phys.Rev. E, 65 (6).

[RN] K. Ray and D. Nawrocki, 1996. *Linear Adaptive Weights and Portfolio Optimization,* http://www.handholders.com/old/raylam.html.

[RU] R.T. Rockafellar and S. Uryasev, 2000. *Optimization of Conditional Value-at-Risk*, Journal of Risk, 2(3): pp. 1-21.

[RGPS] B. Rosenow, P. Gopikrishnan, V. Plerou, and H.E. Stanley, 2003. *Dynamics of CrossCorrelations in the Stock Market* Physica A ,324.

[RGPS] B. Rosenow, P. Gopikrishnan, V. Plerou, and H. E. Stanley, 2002. *Portfolio Optimization and the Random Magnet Problem* Europhys. Lett, 59: pp. 500-506.

[Sp] R. Speicher, 1998. *Combinatorial Theory of the Free Product with Amalgamation and Operator-Valued Free Probability Theory.* Mem. Amer. Math. Soc., 132.

[S] M. S. Srivastava, 2007. *Multivariate Theory for Analyzing High Dimensional Data,* Journal of the Japanese Statistical Society, 37 (1): pp. 53-86.

[T] Y. Tse, 1991. *Stock Return Volatility in the Tokyo Stock Exchange,* Japan and the World Economy, 3: pp. 285-298.

[We] D. Weingarten, 1978. *Asymptotic Behavior of Group Integrals in the Limit of Infinite Rank,* J. Mathematical Phys., 19 (5): pp. 999-1001.

[W] E. Wigner, 1955. *Characteristic Vectors of Bordered Matrices with Infinite Dimensions,* Ann. Of Math., 62 (3).

[Wi] J. Wishart, 1928. *The Generalised Product Moment Distribution in Samples from a Normal Multivariate Population,* Biometrika, 20A: pp. 32-52.