Building a statistical linear factor model and a global minimum variance portfolio using estimated covariance matrices

Lundi Matoti, BSc. (Hons)

Supervisors:

Dr Wilcox, D.

and

Dr Gebbie, T.

Dissertation:

presented to the Faculty of Science, University of Cape Town
in part fulfillment for the requirements of the degree of
Master of Science in Mathematics of Finance

June 3, 2009

Abstract

Simulated data is used to investigate various covariance matrix estimation techniques that can be applied to empirical data of emerging markets. A Statistical factor based model is used in the expected returns prediction, using common factors constructed from various covariance matrix estimators. These covariance estimators are also used in constructing a global minimum variance portfolio.

The aim of this work is to choose the best covariance estimator by building a statistical factor based risk model, where both inputs of the Markowitz mean-variance optimal portfolio selection problem are estimated from a simulated asset returns data. The results from the simulated asset returns data are then used in modeling the expected performance of stocks in the Johannesburg Stock Exchange.

Keywords and Phrases: Missing data, covariance estimation, statistical factors, portfolio optimization

Contents

Acknowledgements

I would like to thank my supervisors Dr D. Wilcox and Dr T. Gebbie for guiding me through the process of producing this thesis, and the helpful suggestions they have given me. I would like to thank the University of Cape Town's postgraduate funding office and Futuregrowth asset management for the financial support they provided, and extend thanks to my presentation group for the contribution they made through our Saturday mornings discussions. A special thanks goes to my proof-reader, Lindelwa Komna, who steered me in the right direction and made sure that I complete this project.

I would also like to thank my examiners for the constructive comments and criticisms that they provided to this write-up. Some of their comments were on improving the current write-up - and I have addressed those comments - and some were a contribution to future research on this topic - and I have included them at the end of the last chapter.

Chapter 1

Introduction

The mean-variance (MV) theory of Markowitz (1952, 1959) [23] [24] is a widely used method for selecting assets that constitute an optimal portfolio, nevertheless, it leaves several questions unanswered for the practitioner [2]. The theory is based on the assumption that the true population parameters of the portfolio, the first and second moments, are known [2]. In practice, these parameters need to be estimated from the available information set, for example the asset returns data, before the portfolio weights can be estimated [2].

Some authors' contributions in the financial economics literature estimate only one of the inputs of the mean-variance portfolio selection. For example authors who model either the first or the second moment (but not both) are [20], [26] and [27] who model only the risk of a portfolio (or the second moment), while [14], [15] and [25] concentrates on modeling the expected asset returns (or the first moment). Most of these studies use empirical data from developed markets like the USA and therefore, their results may not necessarily be directly applicable to developing markets like South Africa [36]. In this project, both the expected return and risk of a portfolio are modeled using simulated data first, and later empirical data.

The purpose of this project is to compare various estimated covariance matrices using a statistical factor based expected returns model and a global minimum variance portfolio. Covariance estimation techniques developed in simulated data are applied to the Johannesburg Stock Exchange(JSE) asset returns data to recover the true underlying covariances. The estimated covariances are then used to construct a global minimum variance portfolio and the common factors used in the expected returns estimation. A statistical factor based risk model is built from a simulated data set and the simulation model applied to empirical data from the JSE.

The results and findings from this work can be used to model empirical data from emerging markets in general because the simulated data has anomalies resembling those found on data sets from emerging markets.

This project is divided into two parts. The first part, consisting of chapters 3, 4, 5 and 6, reviews a simulation environment where a data set randomly

sampled from a multivariate normal distribution is used in the covariance estimation procedure, with the aim of constructing a global minimum variance portfolio and in the development of a statistical factor based expected return model. The second part, consisting of chapter 7, applies the model developed with simulated data to the analysis of empirical data from the Johannesburg Stock Exchange.

The contents of the individual chapters are as follows: chapter 2 provides definitions of some of the concepts used later in the text. Chapter 3 discusses the covariance estimation techniques used in this work and compares them by the Euclidean distance. A statistical linear factor model, used in the estimation of expected asset returns, is built in chapter 4 and a global minimum variance portfolio constructed in chapter 5. Chapter 6 concludes the simulation part and gives a summary of results of the simulation part. In chapter 7, the model investigated via the simulated data is applied to the empirical data from the JSE.

Chapter 2

Definitions

This chapter reviews basic definitions of concepts used in the chapters to follow and introduces the notation that will be used throughout this text. As a way of introducing the notation, matrices will be represented by bold capital letters, for example **A** and **M**, while vectors will be represented by bold small letters, for example **v** and **r**. Elements of matrices and vectors are represented by small italicised letters like $a_{12},\ m_{53},\ v_1$ and r_n, for example. Random variables are represented by italicised capital letters, for example X or Y, and observations of random variables are represented by italicised small letters, for example x or y. Functions are represented by italicised small letters, for example f or g.

All the definitions in this chapter are sample statistics, since they are based on sampled data, and are therefore estimates of true population statistics, for examples $\hat{\mu}$ (sample mean) for μ (true population mean), $\hat{\Omega}$ (sample covariance matrix) for Ω (true population covariance matrix) and $\hat{\mathbf{C}}$ (sample correlation matrix) for $\mathbf{C}$ (true population correlation matrix).

2.1 Multivariate processes

A stock price process represented by a $(p\ +\ 1)\ \times\ n$ matrix $\mathbf{S}$, where $\mathbf{s(t)},\ \mathbf{s(t+1)}, ...,\ \mathbf{s(p+1)}$ is a sequence of row vectors of the matrix $\mathbf{S}$ such that $\mathbf{s(t)} = \{s_{t1},\ s_{t2},\ \ldots,\ s_{tn}\}$, written as follows:

$$\mathbf{S} = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pn} \\ s_{(p+1)1} & s_{(p+1)2} & \cdots & s_{(p+1)n} \end{pmatrix} \qquad (p+1) \times n \ \text{ matrix}$$

is used. The price process is used to define asset returns and only asset returns data will be used in defining most concepts in this chapter. The stock returns process is represented by a $p \times n$ matrix $\mathbf{R}$ and stock returns at time t are defined by

Definition $r(t) = \ln\left(\frac{s(t)}{s(t-1)}\right) = \ln(s(t)) - \ln(s(t-1))$. In this definition, $\mathbf{s} = (s(1), s(2), ..., s(t))$ is the asset price process

and

$$\mathbf{R} = \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ r_{p1} & r_{p2} & \dots & r_{pn} \end{pmatrix} \qquad p \times n \text{ matrix}$$

is the asset returns process. Note that the above definition of asset returns assumes a continuous asset price process. Assume that, at least, the first two moments of the asset returns data matrix exist. That is, $E(\mathbf{R}) < \infty$ and $E(\mathbf{R} - \hat{\mu})^T(\mathbf{R} - \hat{\mu}) < \infty$, where $E(\mathbf{R}) = \hat{\mu}$, and cov($\mathbf{R}$)=$(E(\mathbf{R}) - \hat{\mu})^T(E(\mathbf{R}) - \hat{\mu}) = \hat{\Omega}$

Definition In general, the *mean* vector of the stock returns data matrix $\mathbf{R}$ is defined as

$$E(\mathbf{r}_i) = \mu_i = \frac{1}{p}\sum_{j=1}^{p} r_{ij} \quad i = 1, 2, ..., n \tag{2.1}$$

and if the columns of $\mathbf{R}$ are *i.i.d* from a distribution $f(\mathbf{R})$ say, then the mean is defined as

$$E(\mathbf{r}_i) = \mu_i = \begin{cases} \sum_{j=1}^{p} r_{ij} f(\mathbf{r}_i) & \text{for } \mathbf{r}_i \text{ discrete} \\ \int_{-\infty}^{+\infty} \mathbf{r}_i f(\mathbf{r}_i) d\mathbf{r}_i & \text{for } \mathbf{r}_i \text{ continuous} \end{cases}$$

The *expectation* or the *mean* of $\mathbf{R}$, $E(\mathbf{R})$, is also called the *expected returns* or the *first moment* of the returns process.

Definition The *covariance* between the returns of two stocks $\mathbf{R}_i$ and $\mathbf{R}_j$ is generally defined as follows:

$$\text{cov}(\mathbf{R}_i, \mathbf{R}_j) = \hat{\Omega} = \frac{1}{p-1}\sum_{i=1}^{p}(\mathbf{R}_i - \hat{\mu}_i)^T(\mathbf{R}_j - \hat{\mu}_j) \tag{2.2}$$

and if the columns of $\mathbf{R}$ are *i.i.d* from a distribution, $f(\mathbf{R})$ say, then the covariance is defined as

$$\text{cov}(\mathbf{R}_i, \mathbf{R}_j) = \hat{\Omega} = \begin{cases} \sum_{i=1}^{p-1}(\mathbf{R}_i - \hat{\mu}_i)^T(\mathbf{R}_j - \hat{\mu}_j) f(\mathbf{R}) & \text{for } \mathbf{R} \text{ discrete} \\ \int_{-\infty}^{+\infty}(\mathbf{R}_i - \hat{\mu}_i)^T(\mathbf{R}_j - \hat{\mu}_j) f(\mathbf{R}_i, \mathbf{R}_j) d\mathbf{R}_i d\mathbf{R}_j & \text{for } \mathbf{R} \text{ continuous} \end{cases}$$

When $i = j$, the covariance is called the variance. The standard deviation is defined as the square root of the variance denoted:

$$\hat{\sigma} = \sqrt{\hat{\Omega}}$$

Having defined the covariance between random variables, the next related concept to define is the correlation between random variables which is crudely defined as the standardized covariance matrix, where the covariance matrix is standardized by the standard deviations. The mathematical definition is given below.

Definition The correlation between the returns of two stocks $\mathbf{R}_i$ and $\mathbf{R}_j$ is defined as:

$$\mathbf{C}(\mathbf{R}_i, \mathbf{R}_j) = \frac{1}{\hat{\sigma}_i \hat{\sigma}_j} \text{cov}(\mathbf{R}_i, \mathbf{R}_j) = \frac{1}{\hat{\sigma}_i \hat{\sigma}_j} \hat{\Omega}$$

In matrix notation, the relationship between a covariance matrix, denoted by Ω, and a correlation matrix, denoted by $\mathbf{C}$ is:

$$\Omega = \mathbf{SCS} \quad \text{or} \quad \mathbf{C} = \mathbf{S}^{-1} \Omega \mathbf{S}^{-1}$$

where $\mathbf{S}$ is a diagonal matrix of standard deviations given by:

$$\begin{pmatrix} s_{11} & 0 & \dots & 0 \\ 0 & s_{22} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & s_{nn} \end{pmatrix}$$

The correlation and the covariance matrices are not the only statistics used in modeling dependencies among random or stochastic variables. Other measures of dependence used are *Kendall's* τ, *Spearman's* ρ and *copula*. The definition of the copula is given below.

Definition Suppose $X \in \mathbf{R}$ where each x_i has continuous marginal distribution functions F_i such that each F_i^{-1} exists on the support of x_i for $i = 1, 2, 3, ..., d$. Then

$$\begin{aligned} P(X_1 \leq F_1^{-1}(x_1), ..., X_d \leq F_d^{-1}(x_d)) &= P(F_1(X_1) \leq x_1, ..., F_d(X_d) \leq x_d) \\ &= C(x_1, x_2, ..., x_d) \end{aligned}$$

where the distribution function C has support $[0, 1]^d$ and each marginal distribution $C_i(x_i)$ (where $C_i(x_i)$ is the marginal distribution of the random variable x_i) is uniformly distributed. C is then called the copula for the vector $X = (x_1, x_2, ..., x_d)$. In general, any distribution function with support on $[0, 1]^d$ and uniform marginals is called a copula.

2.2 Positive definiteness

The population covariance and the population correlation matrices have some interesting properties which are useful in different areas of statistical finance. Examples of these properties are symmetry and positive definiteness. The entries of the correlation matrix all lie in the interval $[1, -1]$ with the diagonal elements being all ones. The sample covariance and the sample correlation matrices may not necessarily preserve all the properties of the population covariance

and correlation matrices. One such property that the sample covariance matrix loses is positive definiteness. The property of positive definiteness implies other properties like invertibility as the next definition suggests.

Definition There are several equivalent definitions of positive definiteness of a matrix. The list below gives a definition of a positive definite matrix and lists some properties of positive definite matrices.

- Formally, an $n \times n$ symmetric matrix $\mathbf{M}$ is positive definite $\iff$ for any vector $\mathbf{v} \neq 0,\ \mathbf{v}^T\mathbf{M}\mathbf{v} > 0$.
- All the eigenvalues of a positive definite matrix are positive.
- A positive definite matrix is invertible.

2.3 Distributions

There are many different kinds of distributions which model different random phenomena like continuous or discrete random behaviour, like binary response or multiple response variables, like counts or time between counts. Examples of specific distributions are the binomial distribution, the poisson distribution, the exponential distribution, the normal distribution, the gamma distribution, and the beta distribution. Only one distribution will be defined (used in chapter 5 in the *expected utility function* derivation) and that is the Gaussian or the normal distribution.

Definition Mathematically, the probability density function of a normally distributed random variable is:

$$f(r) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{\left(\frac{(r-\mu)^2}{-2\sigma^2}\right)} \quad \text{where } \mu = E(r) \text{ and } \operatorname{var}(r) = \sigma^2 \tag{2.3}$$

The normal probability distribution has the following important characteristics:

1. The curve has a single peak.
2. It is bell-shaped.
3. The mean (average) lies at the center of the distribution, and the distribution is symmetrical around the mean.
4. The two tails of the distribution extend indefinitely and never touch the horizontal axis.
5. The shape of the distribution is determined by its mean (μ) and standard Deviation (σ). These parameters are both well-defined and finite in the given sample space.

Chapter 3

Covariance estimation

The covariance matrix estimation techniques used in this project can be applied to the analysis of empirical data from emerging markets [36]. The many different covariance estimation techniques which exist were developed and tested on empirical data from developed markets like the US, Europe or Japan. These covariance estimation techniques may not necessarily be directly applicable to financial time series data from emerging markets because of data anomalies such as missing data points. Also, in emerging markets data, there may not be enough time series data to estimate all of the $N(N+1)/2$ entries of the covariance matrix, where the covariance matrix is an $N \times N$ matrix. A more realistic estimation of the covariance matrix in an emerging market can be achieved by using a measured-data sample covariance matrix [36]. That is, a sample covariance matrix or an N-factor model is used. See Appendix A for the discussion of a one factor model and a $k-$factor model.

The covariance estimation procedure is done in two steps. Firstly, a non-positive definite sample covariance matrix is constructed from a simulated asset returns data sampled from a multivariate normal distribution. This data set is sampled such that some data points are missing at randomly chosen time points. A non-positive definite sample covariance matrix, denoted by $\hat{\Omega}_{\text{npd}}$, is calculated from this data set labeled as data matrix $\mathbf{M}_r$, using a statistical method of dealing with missing data called pairwise deletion [12] (see Appendix B for a list of other statistical methods used in dealing with missing data). Secondly, four transformation methods that restore the positive definiteness property in the non-positive definite sample covariance matrix ($\hat{\Omega}_{\text{npd}}$) are applied, resulting in four sample covariance matrix estimators labeled by $\tilde{\Omega}_i,$ for $i \in \{1, 2, 3, 4\}$.

3.1 The construction of a non-positive definite (NPD) covariance matrix

The covariance estimation model building (and other models in later chapters) is done on a 1200 x 300 simulated asset returns data matrix which will be labeled by $\mathbf{M}$, sampled from a multivariate normal distribution with known sample mean $E(\mathbf{M}) = \hat{\mu}$ and known positive definite sample covariance matrix $E(\mathbf{M} - \hat{\mu})^T(\mathbf{M} - \hat{\mu}) = \hat{\Omega}$ or $\mathbf{M} \sim N(\hat{\mu}, \hat{\Omega})$. The

data matrix $\mathbf{M}$ represents an ideal market, where for each data point $\mathbf{M}_i(t),\ \exists\ \mathbf{M}_j(t)\ \forall\ t \in \{1, 2, ..., 1200\}$ and $\forall\ i, j \in \{1, 2, ..., 300\}$ (that is, no missing data points), where t represents time points and $i,\ j$ represent assets or stocks. That is, $\mathbf{M}$ is a $T \times N$ simulated asset returns data matrix where the length of the time series data $T = 1200$ is enough to estimate the $N(N+1)/2$ entries of $\hat{\Omega}$ and $N = 300$ is the number of assets. The columns of $\mathbf{M}$ are not necessarily independent and neither are they mutually uncorrelated.

The data matrix $\mathbf{M}$ is then deliberately tampered with by randomly removing exactly 30%[1] data points to simulate a time series data set of an emerging market. The removal of these data points is a way of simulating a market where thin trading, asynchronous trading or various other data anomalies are prevalent in the time series data. Let $\mathbf{M}_r$ be the data set with holes in it. The sample covariance matrix of $\mathbf{M}_r$ which is usually called a pseudo-covariance matrix [21], given by $E(\mathbf{M}_r - \hat{\mu}_r)^T(\mathbf{M}_r - \hat{\mu}_r) = \hat{\Omega}_{\text{npd}}$, is non-positive definite by construction. One of the aims of this work is to get a positive definite covariance matrix $\tilde{\Omega}$ from $\hat{\Omega}_{\text{npd}}$ which estimates the sample covariance matrix $\hat{\Omega}$. All the sample data statistics that are estimated, predicted or otherwise calculated are done on the matrix $\mathbf{M}_r$ with the results compared to the true sample results from the data matrix $\mathbf{M}$. The data matrix $\mathbf{M}_r$ represents a noisy market or time series data from an emerging market, where there are some randomly missing data points. Note that parameters calculated from $\mathbf{M}$ are indicated by hat($\hat{}$) and parameters calculated from $\mathbf{M}_r$ are indicated by tilde($\tilde{}$)

The purpose of $\mathbf{M}_r$ is to simulate emerging market data problems, such that the resulting 300×300 sample covariance matrix $\text{cov}(\mathbf{M}_r) = \hat{\Omega}_{\text{npd}}$, is non-positive definite, where $\hat{\Omega}_{\text{npd}}$ is calculated using a method of dealing with missing data called pairwise deletion. The question is whether $\hat{\Omega}_{\text{npd}}$ is always non-positive definite or it depends on the positions of the 30% missing data points. In answering this question, singular value decomposition (SVD) is used to write $\hat{\Omega}_{\text{npd}}$ as a product of three matrices as follows:

$$\hat{\Omega}_{\text{npd}} = \mathbf{V}^T \Lambda \mathbf{V}$$

where $\mathbf{V}$ is an orthonormal matrix of eigenvectors of $\hat{\Omega}_{\text{npd}}$ and Λ is a diagonal matrix of eigenvalues of $\hat{\Omega}_{\text{npd}}$. That is, $\lambda_{ij} \in \Lambda\ \forall\ i, j = 1, 2, 3, ..., 300$ is such that $\lambda_{ij} = 0,\ \forall\ i \neq j,\ \lambda_{ij} \neq 0,\ \forall\ i = j$ and $\lambda_{11} \geq \lambda_{22} \geq \lambda_{33} \geq .... \geq \lambda_{nn},$ where $n = 300$. If $\exists\ \lambda_{ii} < 0$ for at least one $i \in \{1, 2, ..., 300\}$, then $\hat{\Omega}_{\text{npd}}$ is non-positive definite, and fortunately, this is the case for the few constructions of $\hat{\Omega}_{\text{npd}}$ that were looked at.

As a way of visualising the process of successfully constructing $\hat{\Omega}_{\text{npd}}$, five $\mathbf{M}_r$ data matrices were constructed from $\mathbf{M}$, by randomly changing the positions of the 30% missing data points, and for each $\mathbf{M}_r$, $\hat{\Omega}_{\text{npd}}$ was calculated. Figure 3.1 shows a plot of (i, λ_i) where i is the position of the eigenvalue and λ_i is the i-th eigenvalue, for each of the five $\hat{\Omega}_{\text{npd}}$, where $i \in \{1, 2, ..., 300\}$.

[1]The 30% is based on the analysis that was done on the data set from the JSE, where it was concluded that there are approximately 30% data points which can be classified as missing data

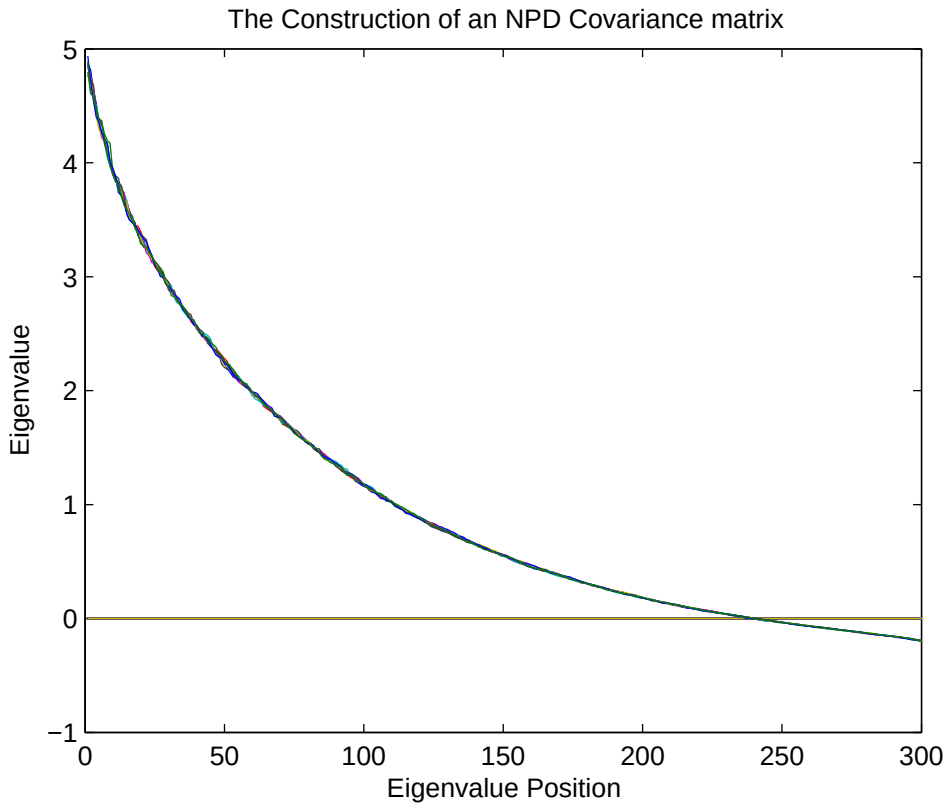

Figure 3.1: *Shows the eigenvalues of five constructed non-positive definite(NPD) matrices. More than 50 eigenvalues are less than zero proving the successful construction of an NPD matrix*

.

Since all the graphs on the diagram cross the $X-$axis, all five constructions of covariance matrices $\hat{\Omega}_{\text{npd}}$ are non-positive definite. These plots agree with each other very closely. This suggests that the eigenvalues of $\hat{\Omega}_{\text{npd}}$ are insensitive to the positions of the missing data.

This completes step 1 of the covariance estimation process - the successful construction of a non-positive definite matrix. Next the transformation methods, sourced from [21] and [8] will be applied in order to transform $\hat{\Omega}_{\text{npd}}$ into a positive definite matrix. This will then complete the covariance matrix estimation process.

3.2 The transformation of a NPD into a positive definite (PD) covariance matrix

Four transformation methods are used to transform the constructed non-positive definite covariance matrix $\hat{\Omega}_{\text{npd}}$ into a positive definite covariance matrix $\tilde{\Omega}$.

Three of these transformation algorithms take as an input a non-positive definite *correlation* matrix, which will be labeled by $\hat{\mathbf{C}}_{\text{npd}}$, and output a positive definite *correlation* matrix, which will be labeled by $\tilde{\mathbf{C}}$, while the fourth takes as an input $\hat{\Omega}_{\text{npd}}$ and outputs a positive definite covariance matrix $\tilde{\Omega}$. Therefore $\hat{\Omega}_{\text{npd}}$ (constructed in the previous section) must be converted to $\hat{\mathbf{C}}_{\text{npd}}$ for the input of the first three transformation techniques and that is done as follows:

$$\hat{\mathbf{C}}_{\text{npd}}(i,j) = \frac{1}{\hat{\sigma}_i \hat{\sigma}_j} \hat{\Omega}_{\text{npd}}(i,j)$$

and the transformation algorithms are then applied to $\hat{\mathbf{C}}_{\text{npd}}$ resulting in a positive definite correlation matrix $\tilde{\mathbf{C}}$. Note that $\hat{\sigma}_i$ and $\hat{\sigma}_j$ are the square roots of the variances of variable i and variable j respectively, and $\hat{\Omega}_{\text{npd}}(i,j)$ is the covariance between variables i and j. The positive definite covariance matrix $\tilde{\Omega}$ will then be calculated from $\tilde{\mathbf{C}}$ as follows:

$$\tilde{\Omega}(i,j) = \tilde{\sigma}_i \tilde{\sigma}_j \tilde{\mathbf{C}}(i,j)$$

Since these transformation methods are an integral part of the covariance estimation techniques used in this work, defining them therefore, forms an integral part of this work. The transformation methods used are:

- the eigenvalue method resulting in $\tilde{\Omega}_1$
- the shrinkage by arctan method resulting in $\tilde{\Omega}_2$
- the shrinkage by tanh method resulting in $\tilde{\Omega}_3$
- the area minimization method resulting in $\tilde{\Omega}_4$.

Note that, throughout this project, any estimator or predictor referenced by 1, for example $\tilde{\Omega}_1$, $\tilde{\mathbf{r}}_1$ or $\tilde{\mathbf{p}}_1$, means that the eigenvalue transformation method was used in estimating it, 2 means that the shrinkage by arctan transformation method was used, 3 refers to the shrinkage by tanh method and 4 is the area minimisation method. The mathematical description of the above transformation methods is given below, with the actual method of transforming $\hat{\Omega}_{\text{npd}}$ into $\tilde{\Omega}_i$ scripted in MatLab, where the transformations are performed.

3.2.1 The eigenvalue transformation method

This method takes a non-positive definite correlation matrix $\hat{\mathbf{C}}_{\text{npd}}$ and decomposes it, by the method of singular value decomposition (SVD), into a diagonal matrix Λ, containing the eigenvalues of $\hat{\mathbf{C}}_{\text{npd}}$, and an orthonormal matrix $\mathbf{P}$ of eigenvectors of $\hat{\mathbf{C}}_{\text{npd}}$. That is $\mathbf{P}\mathbf{P}^T = \mathbf{P}^T\mathbf{P} = \mathbf{I}$, where

$$\hat{\mathbf{C}}_{\text{npd}} = \mathbf{P}\Lambda\mathbf{P}^T$$

Since $\hat{\mathbf{C}}_{\text{npd}}$ is non-positive definite, some of the entries of Λ will be negative and some positive, that is, $\lambda_{ii} < 0$ for at least one $i \in \{1, 2, ..., 300\}$, where λ_{ii} is the $i-$th diagonal element of Λ. An approach used by Lindskog is to replace the negative eigenvalues by some constant $\delta \geq 0$ and then calculate $\tilde{\mathbf{R}} = \mathbf{P}\tilde{\Lambda}\mathbf{P}^T$, where $\tilde{\Lambda}$ is the diagonal matrix containing non-negative diagonal

elements. The diagonal elements of $\tilde{\Lambda}$ are now the eigenvalues of $\tilde{\mathbf{R}}$. That is $\tilde{\mathbf{R}}$ is a positive definite matrix by construction.

A closer look at $\tilde{\mathbf{R}}$ shows that it is a positive definite matrix which is not a correlation matrix because some of the diagonal elements are not equal to unity. Then $\tilde{\mathbf{C}}_1 = \tilde{\Lambda}\tilde{\mathbf{R}}\tilde{\Lambda}$ is a positive definite correlation matrix from $\hat{\mathbf{C}}_{\text{npd}}$ by the eigenvalue transformation method, where $\tilde{\Lambda}$ is a diagonal matrix with diagonal elements $\frac{1}{\sqrt{\tilde{r}_{ii}}}$. The eigenvalue method covariance matrix estimator $\tilde{\Omega}_1$, is given by:

$$\tilde{\Omega}_1(i,j) = \tilde{\sigma}_i \tilde{\sigma}_j \tilde{\mathbf{C}}_1(i,j)$$

3.2.2 The shrinkage transformation methods

This method transforms a non-positive definite correlation matrix $\hat{\mathbf{C}}_{\text{npd}}$ into a positive definite correlation matrix $\tilde{\mathbf{C}}_2$ by shrinking the off-diagonal elements of $\hat{\mathbf{C}}_{\text{npd}}$ until the resulting matrix is either a positive definite correlation matrix $\tilde{\mathbf{C}}_2$ or the off-diagonal elements are all zero, that is, resulting in the identity matrix $\mathbf{I}$. Each non-diagonal element of $\hat{\mathbf{C}}_{\text{npd}}$ is shrunk to zero by the following transformation, where $\hat{c}_{ij} \in \hat{\mathbf{C}}_{\text{npd}}$ for $i, j \in \{1, 2, ..., 300\}$

$$g(\hat{c}_{ij}) = \begin{cases} f^{-1}(f(\hat{c}_{ij})) + \Delta) & \text{if } \hat{c}_{ij} < -f^{-1}(\Delta) \\ 0 & \text{if } -f^{-1}(\Delta) < \hat{c}_{ij} < f^{-1}(\Delta) \\ f^{-1}(f(\hat{c}_{ij}) - \Delta) & \text{if } \hat{c}_{ij} > f^{-1}(\Delta) \end{cases}$$

where $\Delta > 0$ and f can be any function satisfying all of the following three conditions:

- f must be a strictly increasing function: $f(x_1) < f(x_2),\ \forall\ x_1 < x_2$
- f must be an odd function: $f(x) = -f(-x)\ \forall\ x$
- $f(0) = 0$

Note that as $\Delta \to 0,\ g(\hat{\mathbf{C}}_{\text{npd}}) \to \tilde{\mathbf{C}}_2$. That is

$$g(\hat{\mathbf{C}}_{\text{npd}}) = \begin{cases} \tilde{\mathbf{C}}_2 & \text{if } \tilde{\mathbf{C}} - 2 \text{ is positive definite for some } \Delta > 0 \\ \mathbf{I} & \text{if } \tilde{\mathbf{C}} - 2 \text{ is non positive definite } \forall\ \Delta > 0 \end{cases}$$

This means that the nonlinear shrinking of $\hat{\mathbf{C}}_{\text{npd}}$ will always result in a positive definite correlation matrix $\tilde{\mathbf{C}}_2$, because $g(\hat{\mathbf{C}}_{\text{npd}}) \to \mathbf{I}$ and this convergence is guaranteed from the way the function f is defined.

Four possible choices of f (all of which are scripted in MatLab) can be used in shrinking $\hat{\mathbf{C}}_{\text{npd}}$ into $\tilde{\mathbf{C}}_{f_1}, \tilde{\mathbf{C}}_{f_2}, \tilde{\mathbf{C}}_{f_3}, \tilde{\mathbf{C}}_{f_4}$. The functions are:

- $f_1(x) = \tanh(x)$
- $f_2(x) = \tanh^{-1}(x)$

- $f_3(x) = \frac{2}{\pi}\arctan(x)$
- $f_4(x) = \tan\left(\frac{\pi x}{2}\right)$

For example,

$$\tilde{\mathbf{C}}_{f_1} = \begin{cases} \tanh^{-1}(\tanh(\hat{c}_{ij}) + \Delta) & \text{if } \hat{c}_{ij} < -\tanh^{-1}(\Delta) \\ 0 & \text{if } -\tanh^{-1}(\Delta) \leq \hat{c}_{ij} \leq \tanh^{-1}(\Delta) \\ \tanh^{-1}(\tanh(\hat{c}_{ij}) - \Delta) & \text{if } \hat{c}_{ij} > \tanh^{-1}(\Delta) \end{cases}$$

The above four shrinkage functions were compared amongst each other using Euclidean distance and $f_1(x)$ and $f_3(x)$ gave the shortest distances. Thus, in all the calculations and comparisons that follow, only $f_1(x)$ and $f_3(x)$ are used. That is, the shrinkage by arctan ($f_3(x)$) covariance estimator and the shrinkage by tanh ($f_1(x)$) covariance estimator are respectively given by:

$$\tilde{\Omega}_2 = \tilde{\sigma}_i\tilde{\sigma}_j\tilde{\mathbf{C}}_2 \qquad \tilde{\mathbf{C}}_{f_3} \equiv \tilde{\mathbf{C}}_2$$
$$\tilde{\Omega}_3 = \tilde{\sigma}_i\tilde{\sigma}_j\tilde{\mathbf{C}}_3 \qquad \tilde{\mathbf{C}}_{f_3} \equiv \tilde{\mathbf{C}}_2$$

3.2.3 The area minimising transformation method

Estimation of symmetric and positive definite matrices problems can be formulated into finding an optimal solution of a set of linear equations $\mathbf{A}X \approx \mathbf{B}$, where $\mathbf{A}, \mathbf{B} \in \mathbf{R}^{m\times n}$ are given, $X \in \mathbf{P}$ is called a fitting matrix. There are many different methods of solving these systems of linear equations. One of these methods is the least squares approach, where the optimal solution is:

$$\min||\mathbf{A}X - \mathbf{B}||_F^2.$$

Although both $\mathbf{A}$ and $\mathbf{B}$ are measurement matrices, the least squares approach assumes that $\mathbf{A}$ is error free and that all the errors occur in $\mathbf{B}$.

Chen et al [8] developed a global optimizer which solves $\mathbf{A}X \approx \mathbf{B}$ subject to both $\mathbf{A}$ and $\mathbf{B}$ containing errors, X symmetric and positive definite. They construct the following area criterion:

$$\mathrm{Tr}\left[(\mathbf{A}X - \mathbf{B})^T\left(\mathbf{A} - \mathbf{B}X^{-1}\right)\right]$$

where Tr stands for trace, $\mathbf{A}X - \mathbf{B}$ represents the errors in $\mathbf{B}$ from the predictions based on $\mathbf{A}$ and $\mathbf{A} - \mathbf{B}X^{-1}$ represents the errors in $\mathbf{A}$ from the predictions based on $\mathbf{B}$.

Theorem 3.2.1 *Symmetric Positive Definite Estimation Problem (SPDE). For an over-determined set of m linear equations $\boldsymbol{A}X \approx \boldsymbol{B}$, where $\boldsymbol{A}, \boldsymbol{B} \in \boldsymbol{R}^{m\times n}$ are given, $X \in \boldsymbol{P}$ is called a fitting matrix, let the area criterion, $f : I \mapsto \boldsymbol{R}$ be defined as $f(Y) = ||\boldsymbol{A}Y - \boldsymbol{B}Y^{-T}||_F^2$, where $X = YY^T$. The symmetric positive definite estimate X^* is given by $X^* = Y^*Y^{*T}$, where Y^* is the minimizer of $f(Y)$.*

Outline of proof: Let $\mathbf{P} = \mathbf{A}^T\mathbf{A}$ and $\mathbf{Q} = \mathbf{B}^T\mathbf{B}$. This suggests that any symmetric and positive definite estimate X^* of the SPDE problem must satisfy $X^*\mathbf{P}X^* = \mathbf{Q}$, where $X^* = Y^*Y^{*T}$, Y^* is a solution of the SPDE problem. The unique minimizer of $X^*\mathbf{P}X^* = \mathbf{Q}$ is:

$$X^* = \mathbf{U}_P \sum_P^{-1} \mathbf{U}_{\tilde{\mathbf{Q}}} \sum_{\tilde{\mathbf{Q}}} \mathbf{U}_{\tilde{\mathbf{Q}}}^T \sum_P^{-1} \mathbf{U}_P^T$$

where $\mathbf{P} = \mathbf{U}_P \sum_P^2 \mathbf{U}_P^T$ and $\tilde{\mathbf{Q}} = \mathbf{U}_{\tilde{\mathbf{Q}}} \sum_{\tilde{\mathbf{Q}}}^2 \mathbf{U}_{\tilde{\mathbf{Q}}}^T$ are the Schur decomposition of $\mathbf{P}$ and $\tilde{\mathbf{Q}}$ respectively.

The area minimisation covariance estimator is given by $\tilde{\Omega}_4 = X^*$

3.2.4 The test of the transformation methods using eigenvalues

Throughout this project, the phrases transformation methods and covariance matrix estimators will be used interchangeably. Any statistic calculated from the complete sample data matrix $\mathbf{M}$ is called the true sample statistic, for example $\hat{\Omega}$ is called the true sample covariance matrix estimator, and any statistic calculated from $\mathbf{M}_r$ is called an estimator of the one calculated from $\mathbf{M}$, for example $\tilde{\Omega}$ is a sample covariance matrix estimator.

In this investigation all the true sample statistics are known by construction. Here, $\hat{\Omega}$ is the true sample covariance matrix obtained from observations of $\mathbf{M}$ and $\tilde{\Omega}_1,\ \tilde{\Omega}_2,\ \tilde{\Omega}_3,\ \tilde{\Omega}_4$ obtained from $\mathbf{M}_r$ are called the sample covariance matrix estimators of $\hat{\Omega}$. Since $\tilde{\Omega}_1,\ \tilde{\Omega}_2,\ \tilde{\Omega}_3,\ \tilde{\Omega}_4$ are calculated from $\hat{\Omega}_{\text{npd}}$, which is calculated from $\mathbf{M}_r$, are these covariance matrix estimators always positive definite? Now that the transformation methods have been applied to $\hat{\Omega}_{\text{npd}}$ and resulted in $\tilde{\Omega}_i$, has step 2 been successfully completed. This question is answered numerically by looking at the eigenvalues of $\tilde{\Omega}_1,\ \tilde{\Omega}_2,\ \tilde{\Omega}_3,$ and $\tilde{\Omega}_4$ as follows:

- simulate 1000 data matrices $\mathbf{M}_r$ from $\mathbf{M}$, where in each simulation, only the positions of the randomly missing points are changed
- in each simulation run, calculate the eigenvalues (as a 1×300 vector sorted in descending order) of each of $\tilde{\Omega}_1,\ \tilde{\Omega}_2,\ \tilde{\Omega}_3,$ and $\tilde{\Omega}_4$ and collect them into a matrix
- there will be four matrices of size 1000×300 of eigenvalues of $\tilde{\Omega}_1,\ \tilde{\Omega}_2,\ \tilde{\Omega}_3,$ and $\tilde{\Omega}_4$
- take the minimum of the last column of each of these four 1000×300 matrices. If for any of these matrices the minimum is less than zero, then in one simulation run a particular covariance estimation technique or transformation method failed. In all of the four sample covariance estimators, the minimum has been greater than zero, establishing that the transformation techniques do work and this completes step 2

Figure 3.2 is a plot of the eigenvalues of $\tilde{\Omega}_1,\ \tilde{\Omega}_2,\ \tilde{\Omega}_3, \tilde{\Omega}_4$ and $\hat{\Omega}$, where the eigenvalues are averages of the 1000 simulations. The solid line starting at below 5

represents the eigenvalues of the true sample covariance matrix. Note that the eigenvalues of the eigenvalue transformation method follow this line very closely.

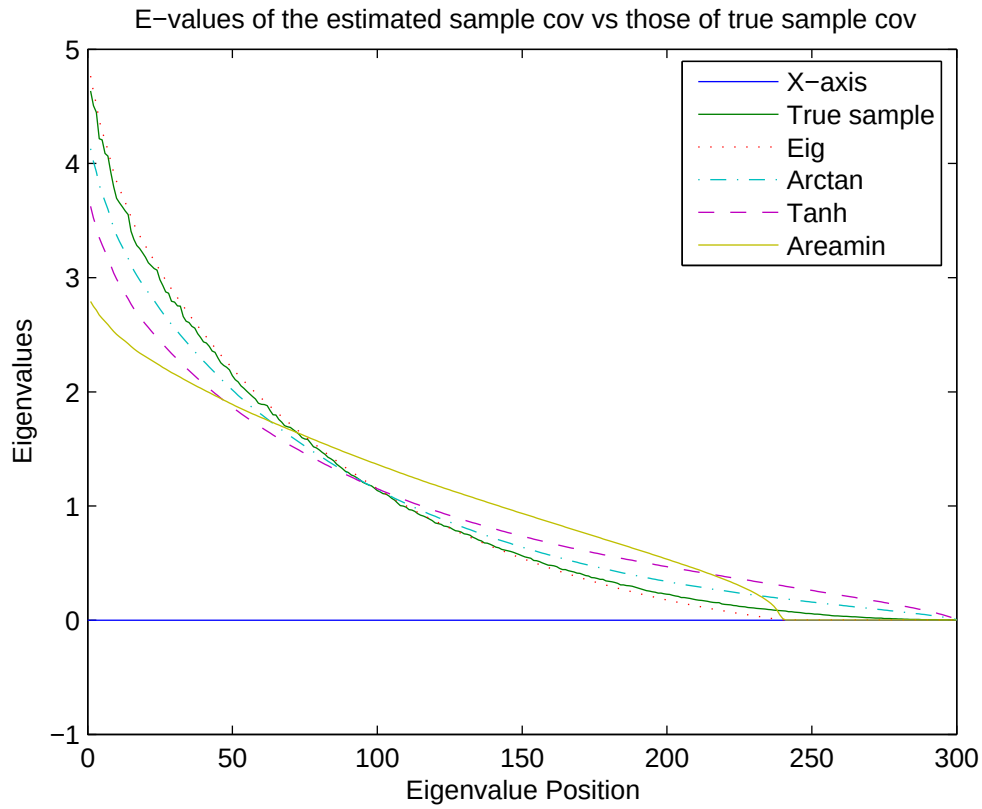

Figure 3.2: *Shows that the eigenvalues of the sample covariance estimators and those of the true sample covariance are all non-negative, where the former implies that the covariance estimation techniques work. The solid line starting at below 5 represents the eigenvalues of the true sample covariance matrix. Note that the eigenvalues of the eigenvalue transformation method follow this line very closely. The eigenvalues of* $\tilde{\Omega}_1$ *and* $\tilde{\Omega}_4$ *start to become zero at the same point on the* $X-$*axis but differ in how they approach that point.*

3.3 Comparing the covariance estimators

Although these transformation algorithms transform a non-positive definite matrix into a positive definite matrix, they differ in how they do it and the resulting matrices which are called covariance matrix estimators also differ. The true sample covariance matrix $\hat{\Omega}$ is known, from which $\hat{\Omega}_{\text{npd}}$ is constructed, so that the positive definite transformation methods can be applied to $\hat{\Omega}_{\text{npd}}$ resulting in the following estimators: $\tilde{\Omega}_1$, $\tilde{\Omega}_2$, $\tilde{\Omega}_3$, and $\tilde{\Omega}_4$ of $\hat{\Omega}$. The question to ask now is which estimator best estimates the sample covariance matrix? In the absence

of knowing $\hat{\Omega}$ (as in the case for real data), it may be difficult to choose the transformation method that can be applied to $\hat{\Omega}_{\text{npd}}$, which best estimates the true sample covariance matrix $\hat{\Omega}$. Fortunately, this is a simulation environment where $\hat{\Omega}$ is known and the Euclidean distance, defined below, is used to compare the estimators to the true sample covariance matrix:

$$\mathbf{d}\left(\hat{\Omega}, \tilde{\Omega}_i\right) = \sum_{j}^{300}\sum_{k}^{300}\left([\hat{\sigma}_{jk}] - [\tilde{\sigma}_{jk}]_i\right)^2 \quad \text{for } i \in \{1,2,3,4\}$$

Lindskog [21], from which the above analysis is adopted, simulates the data set that results in a non-positive definite correlation matrix directly from the uniform distribution, and then calculates the Euclidean distance from the non-positive definite correlation matrix $\hat{\mathbf{C}}_{\text{npd}}$ to each of the positive definite correlation matrices as follows:

$$\mathbf{d}\left(\hat{\mathbf{C}}_{\text{npd}}, \tilde{\mathbf{C}}_i\right) = \sum_{j}\sum_{k}\left([\hat{c}_{jk}]_{\text{npd}} - [\tilde{c}_{jk}]_i\right)^2$$

Lindskog's Euclidean distance calculation is between a non-positive definite matrix and a positive definite matrix (he is not comparing like with like) while the comparison made in this project is between a positive definite matrix and a PD matrix. This Euclidean distance comparison can be visualised in the following two subsections.

3.3.1 Euclidean distances

Figure 3.3 shows a plot of the squares of the Euclidean distances against the number of simulations s. As one can infer from these graphs, there is a clear distinction between the mean distances (where the mean is with respect to the number of simulations) of these sample covariance estimators.

The mean distance of the arctan shrinkage sample covariance matrix estimator, $\mathbf{d}\left(\hat{\Omega}, \tilde{\Omega}_2\right) \approx 56$ square units, is the lowest followed by the eigenvalue covariance matrix estimator at $\mathbf{d}\left(\hat{\Omega}, \tilde{\Omega}_1\right) \approx 58$ square units, then the tanh covariance matrix estimator at $\mathbf{d}\left(\hat{\Omega}, \tilde{\Omega}_3\right) \approx 85$ square units. The mean distance of the area minimisation method is the highest at $\mathbf{d}\left(\hat{\Omega}, \tilde{\Omega}_4\right) \approx 100$ square units.

The variability of these graphs suggests that the eigenvalue transformation method, the arctan shrinkage transformation method and the area minimisation transformation method are not particularly sensitive to the positions of the missing data. By inspection these three methods have similar variances. The tanh shrinkage transformation method on the other hand is very sensitive to the positions of the missing data, with a variance of approximately 30 times more than the other three methods.

The observations from the simulated data is that the sample covariance matrix estimator that is closer to $\hat{\Omega}$ is $\tilde{\Omega}_2$ followed by $\tilde{\Omega}_1$, and the difference between these two is small. Other properties of $\mathbf{d}\left(\hat{\Omega}, \tilde{\Omega}_2\right)$ and $\mathbf{d}\left(\hat{\Omega}, \tilde{\Omega}_1\right)$ need to be

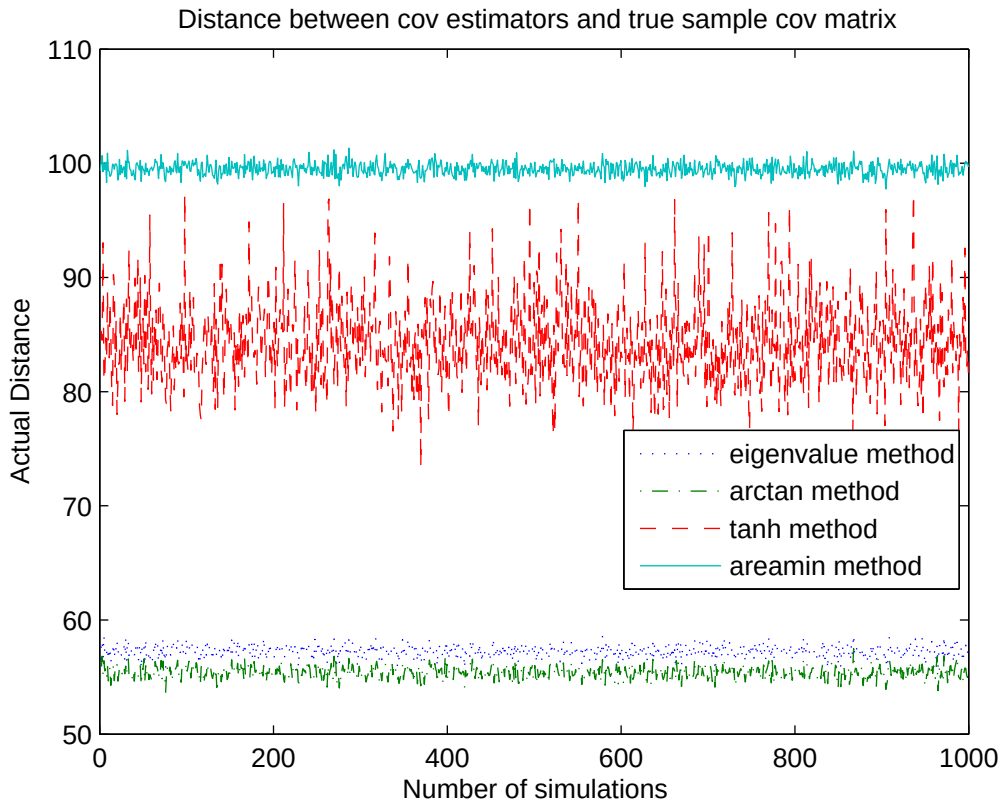

Figure 3.3: *Shows the square of the actual distance from each of the estimated sample covariance matrices to the true sample covariance matrix.*

looked at to see if one can choose an optimal estimator. In the next section, the cumulative distances are considered.

3.3.2 Cumulative Euclidean distances

The cumulative Euclidean distances are calculated as follows:

$$\mathbf{d}_{\text{cum}}\left(\hat{\mathbf{C}}_{\text{npd}}, \tilde{\mathbf{C}}_i\right) = \sum_{s=1}^{1000} \mathbf{d}\left(\hat{\Omega}, \tilde{\Omega}_{si}\right) \quad i \in \{1,2,3,4\},$$

where s is the number of simulations and i represents a transformation method.

A plot of the cumulative distances shown in figure 3.4 shows that $\mathbf{d}_{\text{cum}}\left(\hat{\Omega}, \tilde{\Omega}_3\right)$ and $\mathbf{d}_{\text{cum}}\left(\hat{\Omega}, \tilde{\Omega}_4\right)$ are moving away from $\mathbf{d}_{\text{cum}}\left(\hat{\Omega}, \tilde{\Omega}_1\right)$ and $\mathbf{d}_{\text{cum}}\left(\hat{\Omega}, \tilde{\Omega}_2\right)$. This graph verifies the conclusions reached in the actual distances comparisons. Here the cumulative distances of $\mathbf{d}_{\text{cum}}\left(\hat{\Omega}, \tilde{\Omega}_1\right)$ and $\mathbf{d}_{\text{cum}}\left(\hat{\Omega}, \tilde{\Omega}_2\right)$ are close, with $\tilde{\Omega}_2$ performing marginally better than $\tilde{\Omega}_1$.

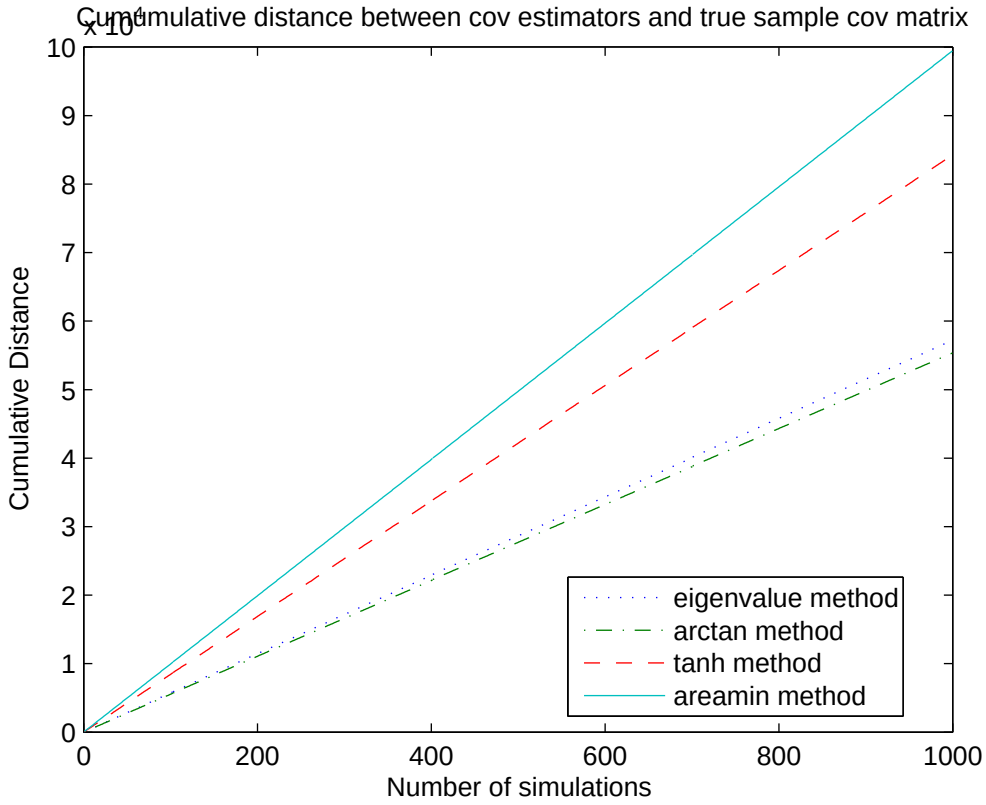

Figure 3.4: *Shows the cumulative squares of the distance from each of the estimated sample covariance matrices to the true sample covariance matrix.*

3.3.3 Average Euclidean distances

Table 3.1 gives the average distances calculated over 1000 simulations.

Method	**Mean distance**	**Standard deviation**
eigenvalue	7.57	0.52
arctan shrinkage	7.44	0.53
tanh shrinkage	9.18	3.69
area minimisation	9.97	0.56

Table 3.1: *Square roots of the average distances and their standard deviations*

As one can see, the average distances of $\mathbf{d}\left(\hat{\Omega}, \tilde{\Omega}_1\right)$ and $\mathbf{d}\left(\hat{\Omega}, \tilde{\Omega}_2\right)$ are not significantly different because 7.57 ± 0.52 and 7.44 ± 0.53 overlap significantly.

3.4 Conclusion

A complete data set $\mathbf{M}$ was sampled from a multivariate normal distribution from which emerging markets data anomalies (which result in a non-positive definite covariance matrix) were constructed, and the resulting data matrix was labeled by $\mathbf{M}_r$. Four transformation methods were applied to the non-positive definite covariance matrix calculated from $\mathbf{M}_r$ to transform it into a positive definite covariance matrix. Euclidean distance was used to compare the four positive definite covariance matrices which were labeled by $\tilde{\Omega}_1$, $\tilde{\Omega}_2$, $\tilde{\Omega}_3$ and $\tilde{\Omega}_4$, to the true sample covariance matrix labeled by $\hat{\Omega}$.

There was no significant difference between $\mathbf{d}\left(\hat{\Omega}, \tilde{\Omega}_1\right)$ and $\mathbf{d}\left(\hat{\Omega}, \tilde{\Omega}_2\right)$, and these can therefore be regarded as possibly equally close to $\hat{\Omega}$. If one wanted to choose a covariance estimator based on Euclidean distance, one may be indifferent between $\tilde{\Omega}_1$ and $\tilde{\Omega}_2$.

In the next chapter, the sample covariance estimators are put in another contest [3]. They will all be used to construct the common factors of a statistical factor based linear model, with the aim of modeling the asset returns. The variance of the leading factor constructed from each of the covariance estimators, the number of common factors k, the correlations and the mean-squared errors are statistics that are used to compare the sample covariance estimators to the true sample covariance matrix.

Chapter 4

A statistical linear factor model of expected returns

Some financial economists argue that estimating the expected asset returns vector is more important than estimating the covariance matrix since the portfolio weights are more sensitive to changes in the expected returns vector [5] [3]. Others believe that, due to the difficulty of estimating expected asset returns, the important improvement that can be made on a portfolio is through an improved covariance estimation technique [5] [3].

Now that the covariance matrix has been estimated, the other input of the mean-variance portfolio selection, the expected asset returns vector, will be estimated next. The purpose of this chapter is to predict the expected asset returns using a statistical factor based risk model, where the common factors of the model are constructed from the estimated sample covariance matrices $\tilde{\Omega}_1$, $\tilde{\Omega}_2$, $\tilde{\Omega}_3$ and $\tilde{\Omega}_4$. The true sample covariance $\hat{\Omega}$ is also used to build the statistical factor model so that each of the models built from $\tilde{\Omega}_1$, $\tilde{\Omega}_2$, $\tilde{\Omega}_3$ and $\tilde{\Omega}_4$ can be compared to the model built from $\hat{\Omega}$.

In this chapter, two differently simulated time series data sets are used. The random data sets $\mathbf{M}$ and $\mathbf{M}_r$ (described in chapter 3), sampled from a multivariate gaussian process are used for continuity. Time series data sets sampled from a GARCH process, which will be labeled by $\mathbf{G}$(simulated perfect market with no missing data points) and $\mathbf{G}_r$(simulated imperfect market with 30% missing data points), simulated in MatLab using the UGARCHSIM function are introduced to test the model's ability to make predictions on a simulated returns data with some signal.

The model is built at time t, where all the model parameters are estimated and hence used to build the asset returns predictive model, and then tested at both times t(to choose the best estimator) and $t+1$(to test the predictive power of the model). Let $\mathbf{r}_t$ be the true observed returns at time t on the data matrix $\mathbf{M}$, $\hat{\mathbf{r}}_t$ (an estimator of $\mathbf{r}_t$ using $\hat{\Omega}$) be the returns predicted from the sample covariance matrix $\hat{\Omega}$ at time t, and $\tilde{\mathbf{r}}_{t_i}$ (an estimator of $\hat{\mathbf{r}}_t$ using each $\tilde{\Omega}_i$) be the returns predicted from each $\tilde{\Omega}_i$ for $i \in \{1,2,3,4\}$ at time t. At each time

point t, a comparison between the true sample covariance and its estimators is done, where the the best covariance estimator, $\tilde{\Omega}_i$, will be chosen based on the following statistic:

$$\max \left| \rho \left(\hat{\mathbf{r}}_t, \tilde{\mathbf{r}}_{t_i} \right) \right| .$$

That is, the best covariance matrix estimator will be chosen based on the strength of the relationship between $\hat{\mathbf{r}}_t$ and $\tilde{\mathbf{r}}_{t_i}$. At time $t+1$, the best predictor will be chosen based on the following two statistics:

$$\max \left| \rho \left(\mathbf{r}_{(t+1)}, \tilde{\mathbf{r}}_{(t+1)_i} \right) \right|$$

$$\tilde{\epsilon}^2_{(t+1)} = \left(\mathbf{r}_{(t+1)} - \tilde{\mathbf{r}}_{(t+1)_i} \right)^2$$

Note that i refers to each of the four covariance estimators. Since a statistical linear factor model is used to build a predictive model of expected asset returns, the next section is dedicated to general linear factor models and their properties, the next one after that is then dedicated to statistical factor models.

4.1 Linear factor models

This section gives the three types of linear factor models, a general overview of linear factor models and mentions, as a factor model example, the Fama and French (1992) three factor model including some of the different views around it, and the linear factor model representations.

4.1.1 Types of linear factor models

There are three main types of linear factor models. These factor models differ in terms of their inputs, techniques used in estimating their parameters and their outputs. When there is no limit on data availability, that is no estimation error, these three factor models are simply restatements of one another [9]. The three types of linear factor models are: statistical, fundamental (or characteristic) and macro-economic factors.

Macro-economic factor models use observable economic time series as measures of common factors, for example, interest rates, inflation and the percentage change in industrial production [9]. Fundamental or characteristic factor models use observable firm or asset specific attributes such as firm size, dividend yield, book to market ratio and industry classification as common factors [9]. Statistical factor models use unobservable variables which are usually regarded as having no economic interpretation with the exception of the leading factor, which is taken as representing the market index.

4.1.2 Uses of linear factor models and some linear factor model examples

The original use of linear factor models was in explaining (a cross-sectional variation in) asset returns. Linear factor models have many other uses in

empirical finance, and are used extensively in (amongst other uses) predicting returns, generating estimates of abnormal returns, identifying risk sensitivities or estimating the variability and covariability of returns [37]. In this work, linear factor models are used to predict the expected asset returns. Examples of some of the well-known linear factor models used are: the capital asset pricing model or the single index factor model of Sharpe(1964) [30], Lintner(1965) [22] and Black(1972) [4][13] and the arbitrage pricing theory factor model of Ross(1976) [29].

Generally, factor models decompose asset returns into factors common to all assets and asset specific factors [37]. Linear factor models can be written as risk models, for example the successful Fama and French (1992) [13] three factor risk model of average returns, or can be written as characteristics models, for example the Daniel and Titman (1997) characteristic model.

The Fama and French (1992) [13] three factor model is based on the premise that firm characteristics such as the market index, size and book-to-market value could be used to model the cross-section of expected asset returns because of high correlations between these characteristics and the expected asset returns. They created three factor-mimicking portfolios which represent a value factor, a size factor and a market factor, where the correlation between a value factor and a size factor was removed [13]. Fama and French (1992) [13] concluded, in their three factor model, that a firm's book-to-market ratio and size are proxies for the firm's loading on priced risk factors [10].

Using intercept tests on triple sorted portfolios, Daniel and Titman (1997) argued that characteristic based models do not have linear loadings on the factor mimicking portfolios and therefore cannot have a risk interpretation. Daniel and Titman's (1997) characteristics argument about the book-to-market anomaly is disputed by Davis, Fama and French (1999) as not general but as only true to the data set used [11].

Van Rensburg and Robertson (2004) [33] created two factor mimicking portfolios - a size factor and a price-to-earnings factor - on the Johannesburg Stock Exchange empirical data, where the two factor mimicking portfolios they constructed were not adjusted for the effect of one on another as in the Fama and French (1992) model. Although the theoretical context of Van Rensburg and Robertson's (2004) model [33] was not that of discriminating between risk and characteristic-based models, their findings were that their results were well explained by the characteristics arguments. It then follows, using the reasoning of Daniel and Titman (1997), that their result did not fully support a risk interpretation premised on the use of factor mimicking portfolios.

4.1.3 Linear factor model representations

General linear multi-factor models have the following functional forms:

$$\mathbf{r}_j(t) = \sum_{i=1}^{N} \delta_{ij}(t)\mathbf{c}_i(t) + \epsilon_j(t) \quad \text{characteristic model representation}$$

- where $\mathbf{r}_j(t)$ = (average) return or excess return on asset j, for $j \in \{1, 2, ..., N\}$
- $\mathbf{c}_i(t) = i-$th common factor, for $i \in \{1, 2, ..., k\}$
- $\delta_{ij}(t)$ = factor payoff on asset j for $i-$th factor
- $\epsilon_j(t)$ = asset specific factor for asset j

or

$$\mathbf{r}_j(t) = \sum_{i=1}^{N} \beta_{ij}(t)\mathbf{f}_i(t) + \epsilon_j(t) \quad \text{risk model representation}$$

- where $\mathbf{r}_j(t)$ = (average) return or excess return on asset j, for $j \in \{1, 2, ..., N\}$
- $\mathbf{f}_i(t) = i-$th common factor, for $i \in \{1, 2, ..., k\}$
- $\beta_{ij}(t)$ = factor loading on asset j for $i-$th factor
- $\epsilon_j(t)$ = asset specific factor for asset j

Note that $\mathbf{r}(t)$ is a centred vector of returns. The following are some of the assumptions that are made about the common factors and the asset specific factors:

- $E(f_{ij}(t)) = \mu_f$ and $\text{cov}(f_{ij}(t)) = \text{E}(f_{ij}(t) - \mu_f)^T(f_{ij}(t) - \mu_f) = \Omega_f$. If $\mu_f = 0$, then $\Omega_f = \mathbf{I}_k$, where k is the number of common factors, and then the factor model is called an orthogonal factor model.
- $\text{cov}(\epsilon_i(t), \epsilon_j(s)) = \Psi = \text{diag}(\Psi_1, \Psi_2, ..., \Psi_N)$
- $\text{cov}(\epsilon_i(t), \mathbf{f}_k(t)) = 0 \ \forall \ k, \ \forall \ i, \text{and} \ \forall \ t$

Some factor models assume that the common factors are given and, they are then written such that their aim is to estimate the payoffs or loadings to the given factors, while others assume that the factor payoffs or loadings are given and then estimate the common factors, and some estimate both the common factors and the factor payoffs [37]. Based on what is given and what is to be estimated, general multi-factor models can be represented as either a cross-sectional regression model at time t as follows:

$$\mathbf{r}(t) = \alpha + \mathbf{B}\mathbf{f}(t) + \epsilon(t) \ \text{ for } t = 1, 2, ..., T$$

where $\mathbf{r}(t)$ is an $N \times 1$ vector of asset returns, $\mathbf{B}$ is an $N \times k$ matrix of factor loadings, $\mathbf{f}(t)$ is a $k \times 1$ vector of common factors and $\epsilon(t)$ is an $N \times 1$ vector of asset specific factors; or they can be represented as a time series regression as follows:

$$\mathbf{r}(i) = \alpha + \mathbf{F}\beta_i + \epsilon(i) \ \text{ for } i = 1, 2, ..., N$$

where $\mathbf{r}(i)$ is a $T \times 1$ vector of asset i's returns, $\mathbf{F}$ is a $T \times k$ matrix, β_i is a $k \times 1$ vector and $\epsilon(i)$ is a $T \times 1$ vector [37].

One can see from the two representations of linear factor models that, to build a model one needs at least the common factors and the factor loadings, and these will be calculated after the description of statistical factor models.

4.2 A statistical linear factor model

A cross-sectional representation of a statistical factor model is used in this project to model returns. The main advantage of a statistical factor model is that it is easy to build a model - one only needs the asset returns data [7]. Since the common factors of a statistical linear factor model are usually regarded as having no economic interpretation, with the exception of the leading factor, some practitioners do not favour this model. Since in this work, the statistical factors are used in the expected asset returns prediction, a lack of interpretability is therefore not an important issue [7]. The statistical risk factors become an issue when used for example in backtesting, benchmarking, performance attribution and related analysis [7].

A statistical linear factor model used is represented by:

$$\tilde{\mathbf{r}}_j(t) = \sum_{i=1}^{N} \hat{\beta}_{ij}(t)\hat{\mathbf{f}}_i(t-1) + \epsilon_j(t).$$

In this model both the common factors and the factor loadings are estimated from the data. The sample covariance matrix estimators calculated from the simulated asset returns data matrix $\mathbf{M}_r$ are used to construct the common factors using principal components based methods.

4.2.1 The construction of the common factors

The common factors are constructed from the correlation matrices $\hat{\mathbf{C}}$, $\tilde{\mathbf{C}}_1$, $\tilde{\mathbf{C}}_2$, $\tilde{\mathbf{C}}_3$, $\tilde{\mathbf{C}}_4$, which are calculated from the estimated sample covariance matrices $\hat{\Omega}$, $\tilde{\Omega}_1$, $\tilde{\Omega}_2$, $\tilde{\Omega}_3$, $\tilde{\Omega}_4$ respectively, by maximising the fit of the linear factor model [9], using principal component analysis and singular value decomposition.

Principal component analysis is a data transformation or data reduction algorithm, where the original data matrix is rotated to a new set of axis, with the columns of the new data arranged in descending order of their contribution to the total variance that exists in the original data. The following theorem gives a mathematical definition of principal components, and shows how to calculate them from the sample covariance matrix and hence from the data [3]:

Theorem 4.2.1 *Let $\boldsymbol{S}$ be the (sample) covariance matrix for the stochastic vector $\boldsymbol{r} = [r_1, r_2, ..., r_N]^T$ and $\boldsymbol{S}$ have eigenvector-eigenvalue pairs $(\boldsymbol{e}_1, \lambda_1)$, $(\boldsymbol{e}_2, \lambda_2)$, ..., $(\boldsymbol{e}_N, \lambda_N)$, where $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N > 0$ and $\boldsymbol{e}_j = [e_1, e_2, ..., e_N]$. Then the $i-th$ (sample) principal component is given by:*

$$\boldsymbol{f}_i = \boldsymbol{e}_i^T \boldsymbol{R} = \sum_{n=1}^{N} \boldsymbol{E}_{ni}\boldsymbol{r}_n \quad \text{for } i = 1, 2, ..., N$$

and

$$var(\boldsymbol{f}_i) = \boldsymbol{e}_i^T \boldsymbol{S}\boldsymbol{e}_i = \lambda_i \quad i = 1, 2, ..., N$$

$$cov(\boldsymbol{f}_j, \boldsymbol{f}_i) = \boldsymbol{e}_i^T \boldsymbol{S}\boldsymbol{e}_i = 0 \ \forall \quad i \neq j$$

Although the above theorem uses a sample covariance matrix in constructing common factors, estimated sample correlation matrices will be used in the construction of the common factors. In line with the notation in the theorem, a standardised $\mathbf{M}_r$ (mean vector of zero and a covariance matrix of $\mathbf{I}$, where $\mathbf{I}$ is the identity matrix) is used in the place of $\mathbf{r}$ and $\hat{\mathbf{C}}$, $\tilde{\mathbf{C}}_1$, $\tilde{\mathbf{C}}_2$, $\tilde{\mathbf{C}}_3$, $\tilde{\mathbf{C}}_4$ are each used in the place of $\mathbf{S}$. Let $\left(\mathbf{V}_1^{\hat{\mathbf{C}}}, \lambda_1^{\hat{\mathbf{C}}}\right), \left(\mathbf{V}_2^{\hat{\mathbf{C}}}, \lambda_2^{\hat{\mathbf{C}}}\right), ..., \left(\mathbf{V}_N^{\hat{\mathbf{C}}}, \lambda_N^{\hat{\mathbf{C}}}\right)$ be the eigenvector-eigenvalue pairs of $\hat{\mathbf{C}}$ and $\left(\mathbf{V}_1^{\tilde{\mathbf{C}}_i}, \lambda_1^{\tilde{\mathbf{C}}_i}\right), \left(\mathbf{V}_2^{\tilde{\mathbf{C}}_i}, \lambda_2^{\tilde{\mathbf{C}}_i}\right), ..., \left(\mathbf{V}_N^{\tilde{\mathbf{C}}_i}, \lambda_N^{\tilde{\mathbf{C}}_i}\right)$ be the eigenvector-eigenvalue pairs of each $\tilde{\mathbf{C}}_i$. The matrix of common factors $\mathbf{F}$, is thus calculated as follows:

$$\mathbf{F}_{\hat{\mathbf{C}}} = \mathbf{M}\mathbf{V}^{\hat{\mathbf{C}}} \quad \text{for true sample correlations and}$$

$$\mathbf{F}_{\tilde{\mathbf{C}}_i} = \mathbf{M}_r\mathbf{V}^{\tilde{\mathbf{C}}_i} \quad \text{for estimated sample correlations}$$

Note that the matrix of common factors $\mathbf{F}$ is a $T \times N$ matrix and that:

$$E(\mathbf{F}_{\hat{\mathbf{C}}}\mathbf{F}_{\hat{\mathbf{C}}}^T) = \Lambda_{\hat{\mathbf{C}}} = \begin{pmatrix} \lambda_{11}^{\hat{\mathbf{C}}} & 0 & \dots & 0 \\ 0 & \lambda_{22}^{\hat{\mathbf{C}}} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & \lambda_{nn}^{\hat{\mathbf{C}}} \end{pmatrix}$$

where $\lambda_{11}^{\hat{\mathbf{C}}} > \lambda_{22}^{\hat{\mathbf{C}}}, \ ..., \ > \lambda_{nn}^{\hat{\mathbf{C}}}$ and

$$E(\mathbf{F}_{\tilde{\mathbf{C}}}\mathbf{F}_{\tilde{\mathbf{C}}}^T) = \Lambda_{\tilde{\mathbf{C}}} = \begin{pmatrix} \lambda_{11}^{\tilde{\mathbf{C}}} & 0 & \dots & 0 \\ 0 & \lambda_{22}^{\tilde{\mathbf{C}}} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & \lambda_{nn}^{\tilde{\mathbf{C}}} \end{pmatrix}$$

where $\lambda_{11}^{\tilde{\mathbf{C}}} > \lambda_{22}^{\tilde{\mathbf{C}}}, \ ..., \ > \lambda_{nn}^{\tilde{\mathbf{C}}}$. In both of the above matrices, the eigenvalues (diagonal elements) of the common factors are also called the variances of the common factors. The average variance (using $\mathbf{M}$ and 1000 simulations of $\mathbf{M}_r$) of the leading factor in each of the above constructed common factors is (the common factors are constructed from $\hat{\mathbf{C}}$ ($\hat{\Omega}$) and $\tilde{\mathbf{C}}_i$ ($\tilde{\Omega}_i$)):

- $\lambda_1^{\hat{\mathbf{C}}} = 4.8564$
- $\lambda_1^{\tilde{\mathbf{C}}_1} = 4.9457$
- $\lambda_1^{\tilde{\mathbf{C}}_2} = 4.2806$
- $\lambda_1^{\tilde{\mathbf{C}}_3} = 3.7187$
- $\lambda_1^{\tilde{\mathbf{C}}_4} = 2.8394$

The leading factors or the market index constructed from $\tilde{\mathbf{C}}_1$ ($\tilde{\Omega}_1$) and $\tilde{\mathbf{C}}_2$ ($\tilde{\Omega}_2$) have the closest variance (given by the leading eigenvalue) to that of the true leading factor constructed from $\hat{\mathbf{C}}$ ($\hat{\Omega}$).

In the statistical factor model (built in section 4.3), standardised matrices of common factors (one for each sample covariance matrix estimator) called factor scores, are used. The factor scores calculated as follows:

$$\mathbf{F}_{\hat{\mathbf{C}}} = \mathbf{M}\mathbf{V}^{\hat{\mathbf{C}}}\Lambda_{\hat{\mathbf{C}}}^{-\frac{1}{2}}$$

$$\mathbf{F}_{\tilde{\mathbf{C}}_i} = \mathbf{M}_r\mathbf{V}^{\tilde{\mathbf{C}}_i}\Lambda_{\tilde{\mathbf{C}}}^{-\frac{1}{2}}$$

have a covariance matrix of

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

with $\mathbf{V}^{\tilde{\mathbf{C}}_i}\Lambda_{\tilde{\mathbf{C}}}^{-\frac{1}{2}}$ referred to as a matrix of factor score coefficients and $\mathbf{V}^{\tilde{\mathbf{C}}_i}\Lambda_{\tilde{\mathbf{C}}}^{\frac{1}{2}}$ referred to as a matrix of factor loadings. Note that this is one way of calculating factor loadings of a statistical factor model. The method of calculating factor loadings used in the statistical factor model built in section 4.3 is explained in section 4.2.3.

The columns of $\mathbf{F}_{\tilde{\mathbf{C}}_i}$ are arranged in descending order of their contribution to the total variance present in the data. Some of the columns of $\mathbf{F}_{\tilde{\mathbf{C}}_i}$ that are on the far right can be dropped, where only the first k columns are used, because their contribution to the total variation is minimal. That is, a method of choosing the number of common factors k to use in the model is needed and that method is found in [3] and [27] as discussed in the next section.

4.2.2 Calculating the number of common factors k

There are many methods that can be used to calculate k, examples of which are: k can be the number of eigenvalues greater than one or k can be the number of eigenvalues that explain a certain percentage (which is known a priori) of the total variation present in the data. In this work, random matrix theory (RMT) is used in calculating the number of common factors k that will be used in the model [3] [27]. The method (RMT) of Plerou et al [27] has been used by many researchers to filter out noise in empirically measured cross-correlation matrices in a wide variety of applications [3].

A brief description of RMT and how it is used to choose the number of common factors k is given below.

Construct a random data matrix $\mathbf{A}$, consisting of N mutually uncorrelated columns such that $E(\mathbf{A}) = \mathbf{0}$ and $\mathbf{A}\mathbf{A}^T = \mathbf{I}$. The random correlation matrix

$\mathbf{R}$ (that is, the correlations present in the random matrix $\mathbf{A}$) is calculated as follows:

$$\mathbf{R} = \frac{1}{N}\mathbf{A}^T\mathbf{A}$$

Let $\lambda_1 \geq \lambda_2 \geq ..., \geq \lambda_n > 0$ be the eigenvalues of $\mathbf{R}$, where $\lambda_n = \lambda_{\min}$ and $\lambda_1 = \lambda_{\max}$. The probability density function of the eigenvalues of a random matrix $\mathbf{R}$ is given by:

$$P(\lambda) = \frac{Q}{2\pi}\frac{\sqrt{(\lambda_{max} - \lambda)(\lambda_{min} + \lambda)}}{\lambda}$$

as $T \to \infty$, $N \to \infty$ where $Q = \frac{T}{N}(> 1)$ is fixed, for λ within the bounds $\lambda_{min} \leq \lambda \leq \lambda_{max}$, $\lambda_{min} = 1 + \frac{1}{Q} - 2\sqrt{\frac{1}{Q}}$ and $\lambda_{max} = 1 + \frac{1}{Q} + 2\sqrt{\frac{1}{Q}}$.

Let $\mathbf{X}$ be a $T \times N$ asset returns data matrix. Empirical correlation present in the returns data $\mathbf{X}$ are given by:

$$\mathbf{C} = \frac{1}{N}\mathbf{X}^T\mathbf{X}$$

Plerou et al [27] compared the properties of $\mathbf{C}$ to those of $\mathbf{R}$. One of the important observations they made was that the properties of the eigen structure of $\mathbf{C}$ that deviate from those of $\mathbf{R}$ must be the part that contains information present in the data - the non-random part of the data.

The number of eigenvalues of $\mathbf{C}$ that are greater than $\lambda_{\max}$, where $\lambda_{\max}$ is the largest eigenvalue of the random correlation matrix $\mathbf{R}$, is the number of eigenvectors containing information present in the data and this corresponds to the number of common factors k to be used in the model. For each eigenvalue greater than $\lambda_{\max}$, the corresponding eigenvector is used as one of the common factors. The number of common factors to use in the model is the same as the number of eigenvalues that are greater than $\lambda_{\max}$.

Generally, however, RMT suggests that one should look at the eigenvalue spectrum of $\mathbf{C}$ that deviates significantly from $\lambda_{\max}$ (not consider all the eigenvalues that are greater than $\lambda_{\max}$) [3]. In this project, the eigenvalue spectrum of $\mathbf{C}$ will not be separated into those that deviate significantly from $\lambda_{\max}$ and those that do not deviate significantly from $\lambda_{\max}$ - all the eigenvalues of $\mathbf{C}$ that deviate from $\lambda_{\max}$ are used, and this is the number of common factors k.

The number of common factors k, were calculated from $\hat{\Omega}$ and from each $\tilde{\Omega}_i$ using 1000 constructions of $\mathbf{M}_r$. Remember that 1000 constructions of $\mathbf{M}_r$ means randomly changing the 30% data points that are missing from $\mathbf{M}$ a thousand times. For each estimated sample covariance matrix, the average (over the 1000 constructions of $\mathbf{M}_r$) of the number of eigenvalues that are inside the information band is calculated and compared to those calculated from the true sample covariance matrix. The calculated average number of common factors are given as $k_{\tilde{\Omega}_1} = 40.549$, $k_{\tilde{\Omega}_2} = 40.300$, $k_{\tilde{\Omega}_3} = 31.790$ and $k_{\tilde{\Omega}_4} = 39.325$ while the true sample k is $k_{\hat{\Omega}} = 46$. That is, $\tilde{\Omega}_1$ and $\tilde{\Omega}_2$ gave the closest estimated k to the true sample k.

The last parameter of the model, the factor payoffs, is estimated next, after which the statistical factor based predictive model is built.

4.2.3 Estimation of factor loadings

The common methods of estimating the factor payoffs or factor loadings of a statistical linear factor model are the principal component analysis, time series multiple regression and the maximum likelihood methods. The principal component estimation method (discussed in section 4.2.1) is based on the spectral decomposition of the sample covariance matrix (and the factor loadings were given as $\mathbf{V}^{\tilde{\mathbf{C}}_i}\Lambda_{\tilde{\mathbf{C}}}^{\frac{1}{2}}$ in section 4.2.1), the time series multiple regression is the application of multiple regression on a time series representation of the factor model and the maximum likelihood method is based on the most likely value but requires the normality assumption and a prespecification of the number of common factors [3].

In estimating factor payoffs/loadings, a moving $1200 - h \times 300$ data matrix (there will be h of these moving data matrices making up the whole estimation period) is taken from the simulated 1200×300 returns data matrix (described in chapter 3) shown below:

$$\begin{array}{c} t = 1200 \\ t = 1199 \\ \vdots \\ t = 1200 - h \\ \vdots \\ \vdots \\ t = 1 \end{array} \begin{pmatrix} m_{1,1} & m_{1,2} & \dots & m_{1,n} \\ m_{2,1} & m_{2,2} & \dots & m_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ m_{h,1} & m_{h,2} & \dots & m_{h,n} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ m_{1200,1} & m_{1200,2} & \dots & m_{1200,n} \end{pmatrix}$$

where h is the number of periods in which the factor payoffs will be estimated and the estimated payoffs at time $t = 1200$ will be the average of these h estimated payoffs. The variable h usually depends on the frequency of the data, for example for monthly data $h = 12$ is used.

In calculating the factor loadings, one needs to calculate the common factors and the number of common factors. In the first estimation window, the common factors and the number of common factors are calculated from the first $1200-h\times 300$ matrix (as discussed in the previous two sections), resulting in a $1200-h\times k$ matrix of common factors. Each of the $1200 - h \times 1$ returns data vectors (and there are 300 of them) in the $1200 - h \times 300$ data matrix, are regressed against the $1200 - h \times k$ matrix of common factors, resulting in a $1 \times k$ vector of factor payoffs or factor loadings. That is, the first estimation results in 300 vectors of factor loadings of size $1 \times k$ each. For example, let $\mathbf{y}$ be the $1200 - h \times 1$ vector of returns and $\mathbf{X}$ be a $1200 - h \times k$ matrix of common factors. A $k \times 1$ vector of factor loadings, denoted by β, is calculated by solving the following linear regression equation:

$$\mathbf{y} = \mathbf{X}\beta \quad (\beta \text{ is called a vector of regression coefficients})$$

In the second estimation window, another $1200 - h \times 300$ matrix is created by dropping in the 1200×300 data matrix the vector at time $t = 1$ and including the vector at time $t = 1200 - h + 1$ (see the above matrix to visualise). Then the above process of calculating the common factors, the number of common factors and performing the multiple regression, is repeated, resulting in the second $1 \times k$ vector of estimated loadings for each of the 300 assets. The vector from the first estimation window is appended by this vector from the second estimation window resulting in a $2 \times k$ matrix of estimated factor payoffs for each asset.

Next the $t = 2$ vector is dropped and the $t = 1200 - h + 2$ vector becomes part of the estimation data giving the third estimation window of size $1200 - h \times 300$ and the whole process repeated. That is, the factor payoffs are estimated using moving windows of equal length $(1200 - h \times 300$ each) of data by estimating the common factors, the number of common factors and then performing multiple regression. The last estimation window uses the data matrix which is between the time points $t = h$ and $t = 1200$.

At the end there will be an $h \times k$ matrix of factor loadings for each of the 300 assets in the data. The estimated payoffs to be used in the model is the average of each $h \times k$ matrix along the h axis, resulting in 300 vectors of size $1 \times k$ of factor payoffs or a $k \times 300$ matrix of factor payoffs.

Since a matrix of common factors and a matrix of factor loadings have been estimated, the statistical factor model can now be built using the simulated returns data matrix $\mathbf{M}_r$ constructed from $\mathbf{M}$ defined in the covariance estimation chapter. The following equation is used in predicting returns

$$E(\tilde{\mathbf{r}}_j(t+1)) = \sum_i E(\hat{\beta}_{ij}(t+1))\mathbf{f}_i(t)$$

$$\text{or } E(\tilde{\mathbf{R}}(t+1)) = E(\hat{\beta}(t+1))\mathbf{F}(t)$$

4.3 Building a statistical factor model

4.3.1 The model

The asset returns model used in this project is a refinement of the model framework used by Haugen and Baker [15]. Their model is given by the following equation:

$$\tilde{\mathbf{r}}_j(t) = \sum_{i=1}^{N} \hat{\beta}_i(t)\tilde{\mathbf{F}}_{ij}(t-1) + \alpha_j(t) + \epsilon_j(j)$$

This model has been refined to the following form:

$$\tilde{\mathbf{r}}_j(t) = \sum_{i=1}^{N} \hat{\beta}_{ij}(t)\tilde{\mathbf{f}}_i(t-1) + \alpha_j(t) + \epsilon_j(j)$$

The primary difference in these two models is that, in the first model, the common factors are unique to each stock and the factor payoffs are common to all stocks while in the second model the common factors are common to all stocks and the factor loadings are unique to each stock. Another difference is that, the former model used fundamental or characteristic factors while the latter uses statistical factors.

4.3.2 The data

In building the model, the 1200×300 simulated data matrix $\mathbf{M}_r$ is split into two matrices of sizes 1180×300 (from $t = 1$ to $t = 1180$) called the in-sample period labeled by $\mathbf{M}_r^{in}$, and 20×300 (from $t = 1181$ to $t = 1200$) called the out-of-sample period labeled by $\mathbf{M}_r^{out}$. The split of $\mathbf{M}_r$ into $\mathbf{M}_r^{in}$ and $\mathbf{M}_r^{out}$ is shown below:

$$\begin{array}{c} t = 1200 \\ \vdots \\ t = 1181 \\ t = 1180 \\ \vdots \\ \vdots \\ \vdots \\ t = 1 \end{array} \begin{pmatrix} m_{1,1}^{out} & m_{1,2}^{out} & \dots & m_{1,n}^{out} \\ \vdots & \vdots & \vdots & \vdots \\ m_{20,1}^{out} & m_{20,2}^{out} & \dots & m_{20,n}^{out} \\ m_{21,1}^{in} & m_{21,2}^{in} & \dots & m_{21,n}^{in} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ m_{1200,1}^{in} & m_{1200,2}^{in} & \dots & m_{1200,n}^{in} \end{pmatrix}$$

The model is built using the in-sample data matrix $\mathbf{M}_r^{in}$ and tested out of sample using data matrix $\mathbf{M}_r^{out}$.

In-sample

The in-sample period moves forward in time but is always the same length (1180×300). For example, the first in-sample period is the data matrix between time points $t = 1$ and $t = 1180$. The second 1180×300 in-sample period is constructed as follows: the data at time $t = 1$ or the row vector $\left(m_{1200,1}^{in}, m_{1200,2}^{in}, \dots, m_{1200,n}^{in}\right)$ is dropped and the data at time $t = 1181$ or the row vector $\left(m_{20,1}^{out}, m_{20,2}^{out}, \dots, m_{20,n}^{out}\right)$ becomes part of $\mathbf{M}_r^{in}$. That is, the first $\mathbf{M}_r^{in}$ is between times $t = 1$ and $t = 1180$, the second is between times $t = 2$ and $t = 1181$,...., and the last is between times $t = 20$ and $t = 1199$.

Out-of-sample

The out-of sample period decreases every time the model is built as follows, $20 \times 300, 19 \times 300, ..., 1 \times 300$, and is always non-overlapping with the in-sample period. The aim is to estimate all the returns that are in $\mathbf{M}_r^{out}$ using information

available in $\mathbf{M}_r^{in}$. That is, the returns at $t = 1181,\ t = 1182,\ ...,\ t = 1200$ which are all in $\mathbf{M}_r^{out}$ are estimated using information available up to and including $t = 1180,\ t = 1181,\ ...,\ t = 1199$ respectively which are all in $\mathbf{M}_r^{in}$. At the end, there will be a new 20×300 matrix of returns which are estimates of $\mathbf{M}_r^{out}$.

4.3.3 Building the model

In the first estimation cycle, the way the model is built is that, the in-sample period (of size 1180×300) is split into h equal estimation windows of size $1180 - h \times 300$ each, in a manner similar to the $1200 - h \times 300$ data matrix used to estimate factor payoffs in the previous section. In this case, $h = 12$ is used. The number of common factors used is constant throughout the estimation periods or windows and is the average of the 1000 simulations given by $k_{\tilde{\Omega}_1} = 41,\ k_{\tilde{\Omega}_2} = 40,\ k_{\tilde{\Omega}_3} = 32$ and $k_{\tilde{\Omega}_4} = 39$ while the true sample k is $k_{\hat{\Omega}} = 46$. Since there are h windows, then there will be h matrices of common factors of size $1180 - h \times k$ each. The common factors used in the model is the top $1 \times k$ vector of the $1180 - h \times k$ matrix of common factors calculated in the last estimation window. The factor loadings used are the 300 vectors of size $1 \times k$ or the $k \times 300$ matrix which is calculated as explained in the previous section.

At time $t = 1180$, the following equation:

$$E(\tilde{\mathbf{r}}_j(1181)) = \sum_i E(\hat{\beta}_{ij}(1181))\mathbf{f}_i(1180)$$

$$\text{or } E(\tilde{\mathbf{R}}(1181)) = E(\hat{\beta}(1181))\mathbf{F}(1180)$$

is used to estimate the returns at time $t = 1181$, where $\hat{\beta}(1181)$ is the $k \times 300$ matrix of factor loadings and $\mathbf{f}(1180)$ is the $1 \times k$ vector of common factors, resulting in the 1×300 vector of expected returns $\tilde{\mathbf{r}}_j(1181)$. The whole process is done for each of $\tilde{\Omega}_1,\ \tilde{\Omega}_2,\ \tilde{\Omega}_3,\ \tilde{\Omega}_4$ and $\hat{\Omega}$ (the latter is for comparison)..

In the next estimation cycle, the out-of-sample period loses the 1×300 vector of returns at $t = 1180$, and this vector becomes part of the in-sample period while the in-sample period drops the 1×300 vector of returns at time $t = 1$. That is, the next round of predictions is done using $\mathbf{M}_r^{\text{in}}$ given by $t = 2$ to $t = 1181$, where the returns that are predicted are at $t = 1182$. All of the processes mentioned above are then repeated for this new data matrix:

- split the 1180×300 in-sample period into h moving estimation windows of equal length
- in each estimation window, calculate the common factors
- in each estimation window, k is constant
- perform regression to calculate the factor loadings
- repeat the first four bullets for each of the h moving estimation windows
- at the end, there will be h matrices of common factors and h matrices of factor loadings

- take the top row vector from the matrix of common factors calculated in the last estimation window, take the average of the h matrices of factor loadings and apply the following equation

$$E(\tilde{\mathbf{r}}_j(t+1)) = \sum_i E(\hat{\beta}_{ij}(t+1))\mathbf{f}_i(t)$$

$$\text{or } E(\tilde{\mathbf{R}}(t+1)) = E(\hat{\beta}(t+1))\mathbf{F}(t)$$

At the end, there will be five 20×300 matrices of predicted returns which estimate the $\mathbf{M}_r^{out}$ data matrix. Which of these estimators is close to the $\mathbf{M}_r^{out}$ data matrix is a question that will be considered in the following sections.

4.4 Results of the model

Simulated time series data sets $\mathbf{M}$ from which $\mathbf{M}_r$ is constructed (the process of which is described in detail in chapter 3), and $\mathbf{G}$ from which $\mathbf{G}_r$ is constructed are used. These data sets are simulated in MatLab using the functions MVNRND and UGARCHSIM respectively. The reason for using $\mathbf{M}(\mathbf{M}_r)$ is for continuity from the previous chapter, while that for using $\mathbf{G}(\mathbf{G}_r)$ is to test the predictive power of the model on a time series data that has some signal.

There are two levels at which the results are compared. The first level of comparison is for continuity from the covariance estimation techniques, where the quest is to find the covariance matrix estimator that results in statistics estimated from $\mathbf{M}_r$, that are closer to the true sample statistics calculated from $\mathbf{M}$. In this first level, the model is built using each $\tilde{\Omega}_i$ resulting in $\tilde{\mathbf{r}}_{t_i}$, and on $\hat{\Omega}$ resulting in $\hat{\mathbf{r}}_t$, and then $|\rho(\tilde{\mathbf{r}}_{t_i}, \hat{\mathbf{r}}_t)|$ is used as a closeness measure (remember that, in the covariance estimation, Euclidean distance was used as a measure of closeness).

The second level of comparison, tests the fit of the model in the given data sets, using $\left(\tilde{\epsilon}_{(t+1)}\right)^2 = \left(\mathbf{r}_{(t+1)} - \tilde{\mathbf{r}}_{(t+1)_i}\right)^2$ and $\tilde{\rho}_{(t+1)} = |\rho(\mathbf{r}_{(t+1)}, \tilde{\mathbf{r}}_{(t+1)_i})|$, where $\mathbf{r}_{(t+1)}$ is a 1×300 vector of observed returns in the $\mathbf{M}_r^{out}$ data matrix. The asset returns vector $\mathbf{r}_{(t+1)}$ is a vector of true observations that each of $\tilde{\mathbf{r}}_{(t+1)_i}$ are estimating.

4.4.1 Simulation results using data randomly sampled from a multivariate Gaussian process

An important advantage of a simulation environment is that the complete or true sample data set is known and therefore backtesting of the model is possible - in reality this is not the case. Building the statistical factor model from the complete data set $\mathbf{M}$ helps one know what the returns predicted from the incomplete data set $\mathbf{M}_r$ should approximately be. That is why the first comparison is between $\hat{\mathbf{r}}_t$(model parameters calculated from $\mathbf{M}$ and hence $\hat{\Omega}$) and $\tilde{\mathbf{r}}_{t_i}$(model parameters calculated from $\mathbf{M}_r$ and hence $\tilde{\Omega}_i$) - the best covariance matrix estimator needs to be chosen.

Choosing the best covariance matrix estimator

At time $t = 1180$, the model results in $\hat{\mathbf{r}}_{1181}$ calculated from $\hat{\Omega}$, and $\tilde{\mathbf{r}}_{1181_1}$, $\tilde{\mathbf{r}}_{1181_2}$, $\tilde{\mathbf{r}}_{1181_3}$ and $\tilde{\mathbf{r}}_{1181_4}$, which are calculated from $\tilde{\Omega}_1$, $\tilde{\Omega}_2$, $\tilde{\Omega}_3$ and $\tilde{\Omega}_4$ respectively. Note that, $\hat{\Omega}$ and $\hat{\mathbf{r}}_{1181}$ represents true sample statistics from an ideal market with a time series data matrix $\mathbf{M}$, and that $\tilde{\Omega}_i$ and $\tilde{\mathbf{r}}_{1181_i}$ are their estimates from the data set $\mathbf{M}_r$ using transformation method i. The correlations between each $\tilde{\mathbf{r}}_{t_i}$ and $\hat{\mathbf{r}}_t$ for $t \in \{1181, 1182, ..., 1200\}$ are calculated as follows

$$\rho(\hat{\mathbf{r}}_t, \tilde{\mathbf{r}}_{t_i}) = \left| \frac{\text{cov}(\hat{\mathbf{r}}_t, \tilde{\mathbf{r}}_{t_i})}{\sqrt{\text{var}(\hat{\mathbf{r}}_t)\text{var}(\tilde{\mathbf{r}}_{t_i})}} \right|$$

The correlations were calculated from $t = 1181$ to $t = 1200$ using equal moving windows of data. The results below give the average of the 20 absolute correlations:

- $\rho(\hat{\mathbf{r}}, \tilde{\mathbf{r}}_1) = 10.01\%$
- $\rho(\hat{\mathbf{r}}, \tilde{\mathbf{r}}_2) = 10.81\%$
- $\rho(\hat{\mathbf{r}}, \tilde{\mathbf{r}}_3) = 13.38\%$
- $\rho(\hat{\mathbf{r}}, \tilde{\mathbf{r}}_4) = 11.32\%$.

Since these correlations are significantly different to zero, this implies that the model has potential. This suggests that transformation methods 3 and 4 are the best transformation methods because $\tilde{\Omega}_3$ and $\tilde{\Omega}_4$ results in predicted returns that are relatively strongly correlated to the returns predicted from $\hat{\Omega}$. These are strong correlations considering that the data set used is random.

Testing the predictive power of the model

The comparison done in the previous section was between estimated returns $\hat{\mathbf{r}}_t$ and $\tilde{\mathbf{r}}_{t_i}$ and it was to test which covariance matrix estimator results in statistics that are close to those calculated/estimated using the true sample covariance. In this section, the comparison is between the true observations $\mathbf{r}_t$, which are in $\mathbf{M}_r^{out}$, and the estimated returns $\tilde{\mathbf{r}}_{t_i}$ calculated from $\mathbf{M}_r^{in}$. At each time point starting at time $t = 1181$ and ending at time $t = 1200$, the following two statistics are calculated:

$$\tilde{\rho}_{(t+1)_i} = \left| \rho \left(\mathbf{r}_{(t+1)}, \tilde{\mathbf{r}}_{(t+1)_i} \right) \right|$$
$$\tilde{\epsilon}^2_{(t+1)_i} = \left(\mathbf{r}_{(t+1)} - \tilde{\mathbf{r}}_{(t+1)_i} \right)^2$$

and used as a basis of comparison. That is, at each time point, the strength of the relationship between the observed and the estimated returns, and the deviations of the predictors from the true observed returns are calculated. The length of the out-of-sample period is 20 and therefore, there will be 20 time points. At time $t = 1200$, there will be four 20×1 vectors of correlations and four 20×300 matrices of error terms. The average correlations, calculated as follows:

$$\tilde{\rho}_i = \frac{1}{20} \sum_{t=1180}^{1199} \left| \rho \left(\mathbf{r}_{(t+1)}, \tilde{\mathbf{r}}_{(t+1)i} \right) \right|$$

are: $\tilde{\rho}_1 = 6.71\%$, $\tilde{\rho}_2 = 8.79\%$, $\tilde{\rho}_3 = 9.83\%$ and $\tilde{\rho}_4 = 8.64\%$. These correlations are all lower than those calculated in the previous section.

Plots of the actual absolute correlations and the cumulative absolute correlations are given in figures 4.1 and 4.2 respectively. This is to see whether there is a method that consistently had the strongest absolute correlations throughout the 20 time points. The reason for plotting cumulative correlations is that, a plot of the actual correlations results in a graph that looks very noisy(as can be seen in figure 4.1), with different correlation curves overlapping and is difficult to visualise or interpret. The cumulative correlations graph show that there is no significant difference between $\tilde{\rho}_2$ and $\tilde{\rho}_4$, while $\tilde{\rho}_3$ is consistently higher.

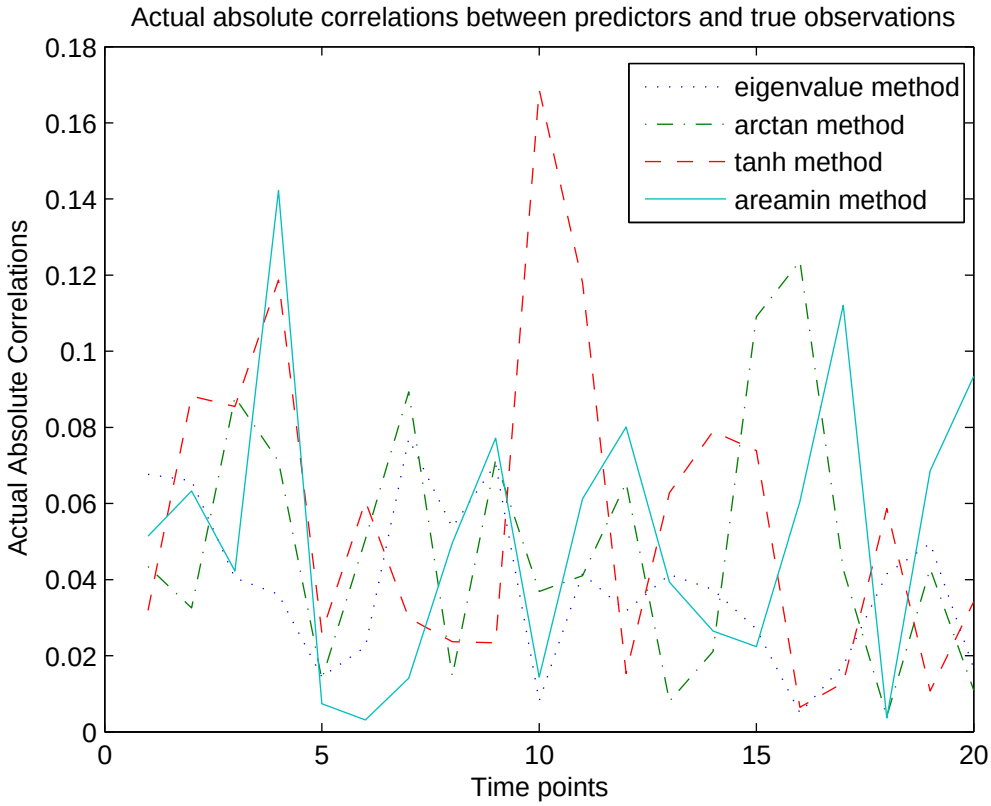

Figure 4.1: *Shows the absolute values of the correlations between* $\boldsymbol{r}_{(t+1)}$ *(true observations) and* $\tilde{\boldsymbol{r}}_{(t+1)_i}$ *(predictors) for* $t \in \{0,\ 1,\ 2, ...,\ 19\}$*. There is no clear pattern in the correlations*

The above correlations show that the model has potential and may be used to predict returns. The tanh covariance estimator is doing consistently well

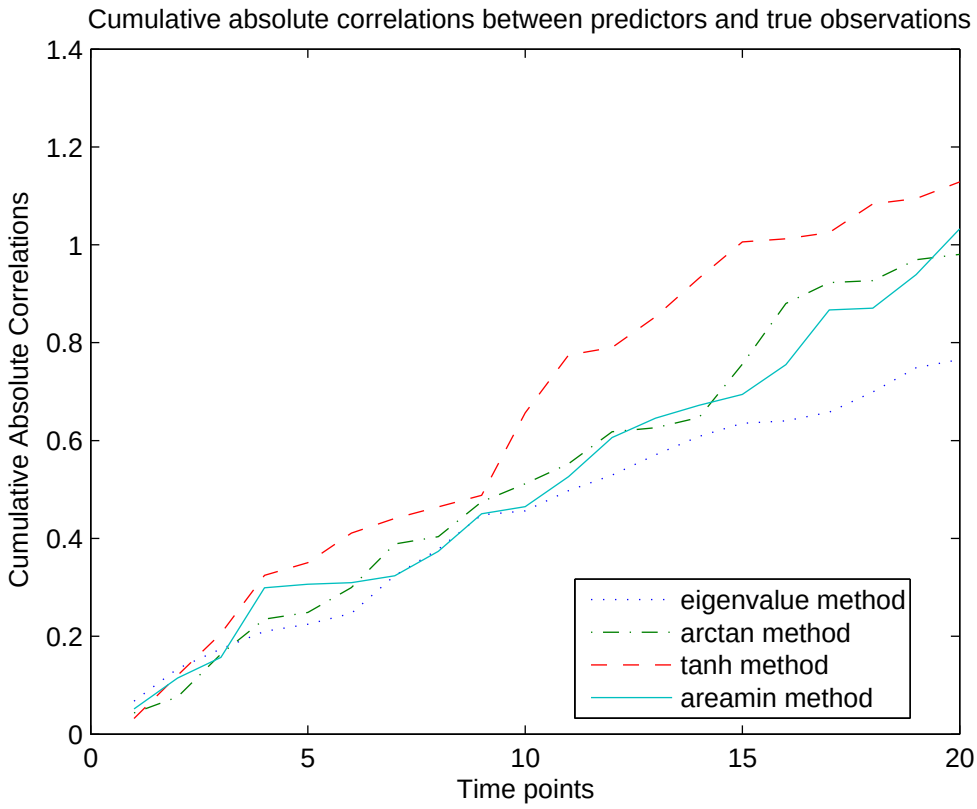

Figure 4.2: *Shows cumulative absolute values of the correlations between* $\boldsymbol{r}_{(t+1)}$ *(true observations) and* $\tilde{\boldsymbol{r}}_{(t+1)_i}$ *(predictors) for* $t \in \{0,\ 1,\ 2, ..., 19\}$. *The tanh covariance estimator gives consistently high correlations and the eigenvalue covariance estimator gives consistently low correlations using data matrix* $\boldsymbol{M}_r$.

while the eigenvalue covariance estimator is doing consistently poorly. These are good results considering that the data has been sampled randomly.

The deviations of the predicted observations from the true observed returns are considered next. At each time point t, the deviations or the error terms are calculated by the following equation:

$$\tilde{\epsilon}^2_{(t+1)_i} = \left(\mathbf{r}_{(t+1)} - \tilde{\mathbf{r}}_{(t+1)_i}\right)^2$$

where $\tilde{\epsilon}^2_{(t+1)_i}$ is a 1×300 vector of errors at each time point t. At the end of the out-of-sample period, there will four (one for each method $i \in \{1\ ,\ 2,\ 3,\ 4\}$) 20×300 matrices of error terms. How can these matrices be compared so as to choose the method that gives reasonably low errors? One way that this can be done is by calculating the average mean-squared errors as follows:

$$\tilde{\epsilon}^2_i = \frac{1}{20}\sum_{t=1}^{20}\left(\mathbf{r}_{(t+1)} - \tilde{\mathbf{r}}_{(t+1)_i}\right)^2 \quad \text{for } i \in \{1,\ 2,\ 3,\ 4\}$$

Method	Correlations	Mean-Squared Errors
eigenvalue	6.71%	112
arctan shrinkage	8.79%	257
tanh shrinkage	9.83%	123
area minimisation	8.64%	242

Table 4.1: *Summary of results: average of the absolute values of the actual correlations between* $\boldsymbol{r}_{(t+1)}$ *(true observations) and* $\tilde{\boldsymbol{r}}_{(t+1)_i}$ *(predictors) for* $t \in \{0,\ 1,\ 2, ...,\ 19\}$*; and the average of the square roots of the Mean-Squared Errors between* $\boldsymbol{r}_{(t+1)}$ *(true observations) and* $\tilde{\boldsymbol{r}}_{(t+1)_i}$ *(predictors) for* $t \in \{0,\ 1,\ 2, ...,\ 19\}$ *using data matrix* $\boldsymbol{M}_r$

and the method which best predicts the true observed returns will be the one with the lowest error, $\tilde{\epsilon} = \min\{\tilde{\epsilon}_1,\ \tilde{\epsilon}_2,\ \tilde{\epsilon}_3,\ \tilde{\epsilon}_4\}$.

A summary of the results of the returns prediction using data set $\mathbf{M}_r$ is given in table 4.1, where both the correlations and the square-roots of the mean-squared errors of each method are given. The lowest error terms are given by $\tilde{\epsilon}_1$ and $\tilde{\epsilon}_3$. These errors are large considering that this is an asset returns data and not asset prices, but maybe the randomness of the data is to blame.

At the end of the out-of-sample period, there will four 20×300 matrices of errors, that is, each asset has a 20×1 vector of errors calculated at each of the 20 time points. In theory, the mean of each of these 20×1 vectors (and there are 300 of them) calculated as follows $\tilde{\epsilon}_i^2 = \frac{1}{20}\sum_{t=1}^{20}\left(\mathbf{r}_{(t+1)} - \tilde{\mathbf{r}}_{(t+1)_i}\right)^2$ is suppose to be (close to) zero. Is this what is happening in this case? Figure 4.3 plots these errors in an attempt to visualise their evolution over the 20 time points. The mean error of each asset is calculated, resulting in four 1×300 vectors. Plotting these vectors (each of the 1×300 vectors) of means result in very noisy curves due to the different sizes of these errors. Since these mean errors have different sizes, for ease of visualization, the mean errors of the first 10 assets were plotted on the same set of axis (fig 4.3).

The different scales of the errors make it difficult to see what is going on in fig 4.3. One thing for sure though is that, the arctan method is breaking somewhere, resulting in abnormally high errors. The graph of the first 10 assets were then plotted on different axis (fig 4.4). As one can see from fig 4.4, some assets are predicted reasonably well, with errors close to zero, while some are not.

Again, the model is doing reasonably well with the random data. In the next section, the data matrix $\mathbf{G}_r$(data sampled from a GARCH process) is used.

4.4.2 Simulation results using data sampled from a GARCH process

The data matrix $\mathbf{G}_r$, simulated from a GARCH process is used in this section to test the fit of the model on a data set that has some signal. A 1200×300 time series data $\mathbf{G}_r$ is simulated from a GARCH process in MatLab, using a function called UGARCHSIM. There is only one level of comparison done in

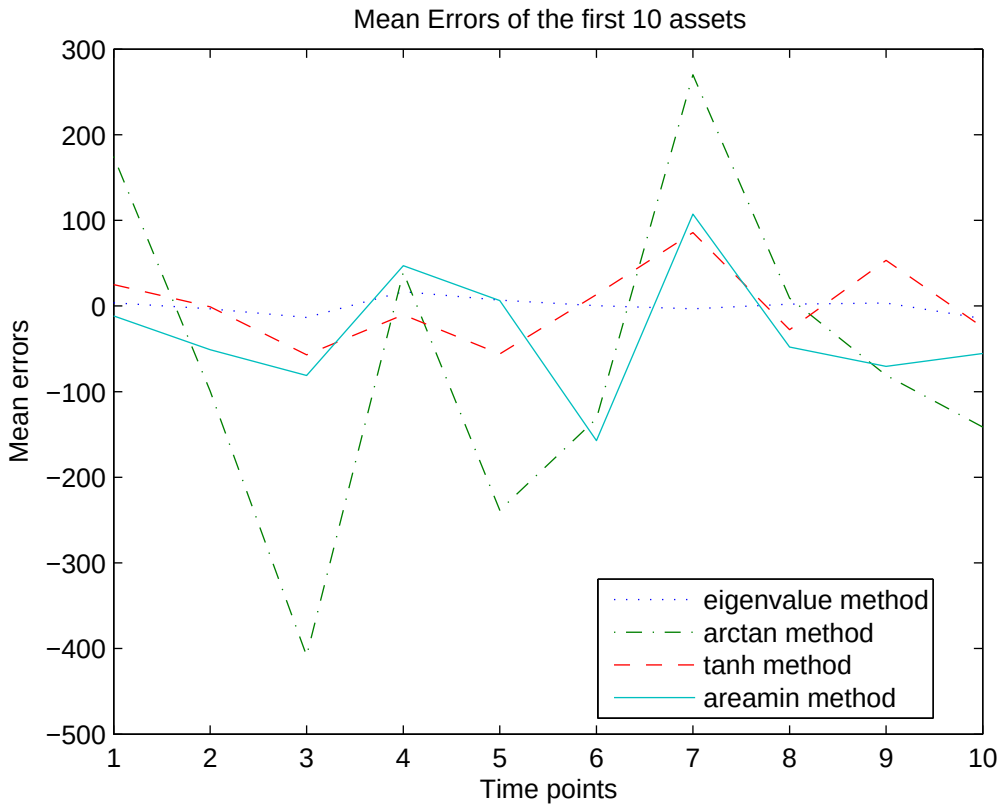

Figure 4.3: *Shows the average over the 20 estimation points of the mean errors between* $\boldsymbol{r}_{(t+1)}$ *(true observations) and* $\tilde{\boldsymbol{r}}_{(t+1)_i}$ *(predictors) of the first 10 assets using data matrix* $\boldsymbol{M}_r$ *- same axis.*

Method	**Correlations**	**Mean-Squared Errors**
eigenvalue	15.14%	6.6
arctan shrinkage	17.37%	5.7
tanh shrinkage	13.91%	7.1
area minimisation	14.04%	8.7

Table 4.2: *Summary of results: average of the absolute values of the actual correlations between* $\boldsymbol{r}_{(t+1)}$ *(true observations) and* $\tilde{\boldsymbol{r}}_{(t+1)_i}$ *(predictors) for* $t \in \{0,\ 1,\ 2, ...,\ 19\}$*; and the average of the square roots of the Mean-Squared Errors between* $\boldsymbol{r}_{(t+1)}$ *(true observations) and* $\tilde{\boldsymbol{r}}_{(t+1)_i}$ *(predictors) for* $t \in \{0,\ 1,\ 2, ...,\ 19\}$ *using data matrix* $\boldsymbol{M}_r$

the data set $\mathbf{G}_r$ - comparing the predictors to the true observations, using the two statistics $\tilde{\rho}$ and $\tilde{\epsilon}$.

On average, the results seem to have improved in both the correlations and the mean-squared errors. A summary of results is given in table 4.2. All correlations are at least 14%, and the square root of the mean-squared errors

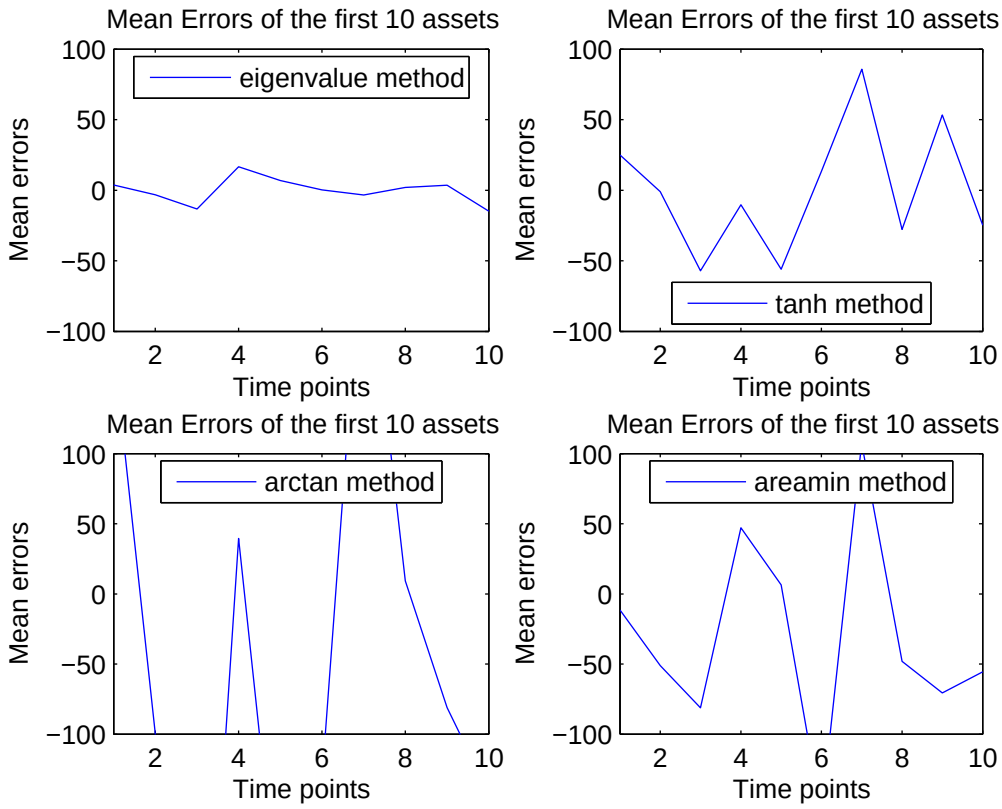

Figure 4.4: *Shows the average over the 20 estimation points of the mean errors between* $\boldsymbol{r}_{(t+1)}$ *(true observations) and* $\tilde{\boldsymbol{r}}_{(t+1)_i}$ *(predictors) of the first 10 assets using data matrix* $\boldsymbol{M}_r$ *- different axis.*

are less than those calculated using data matrix $\mathbf{M}_r$. In the above simulation results with time series data $\mathbf{M}_r$ (section 4.4.1), arctan transformation method gave extremely high errors, while with time series data $\mathbf{G}_r$, all the errors are reasonable and better than those found in data matrix $\mathbf{M}_r$, with the arctan shrinkage estimator giving the lowest (and the highest correlations).

The improvement in the two statistics, the error terms and the correlations, (summarised in table 4.2) on the data set $\mathbf{G}_r$ can be visualised in figures 4.5 and 4.6 respectively, where the mean errors of the first 10 assets are plotted on different set of axis, and cumulative absolute correlations as opposed to actual absolute correlations are plotted. In this simulation, where $\mathbf{G}_r$ is used, $\tilde{\Omega}_1$ and $\tilde{\Omega}_2$ are more correlated to the true observations than the other two transformation methods. The lowest mean-squared errors are also given by $\tilde{\Omega}_1$ and $\tilde{\Omega}_2$.

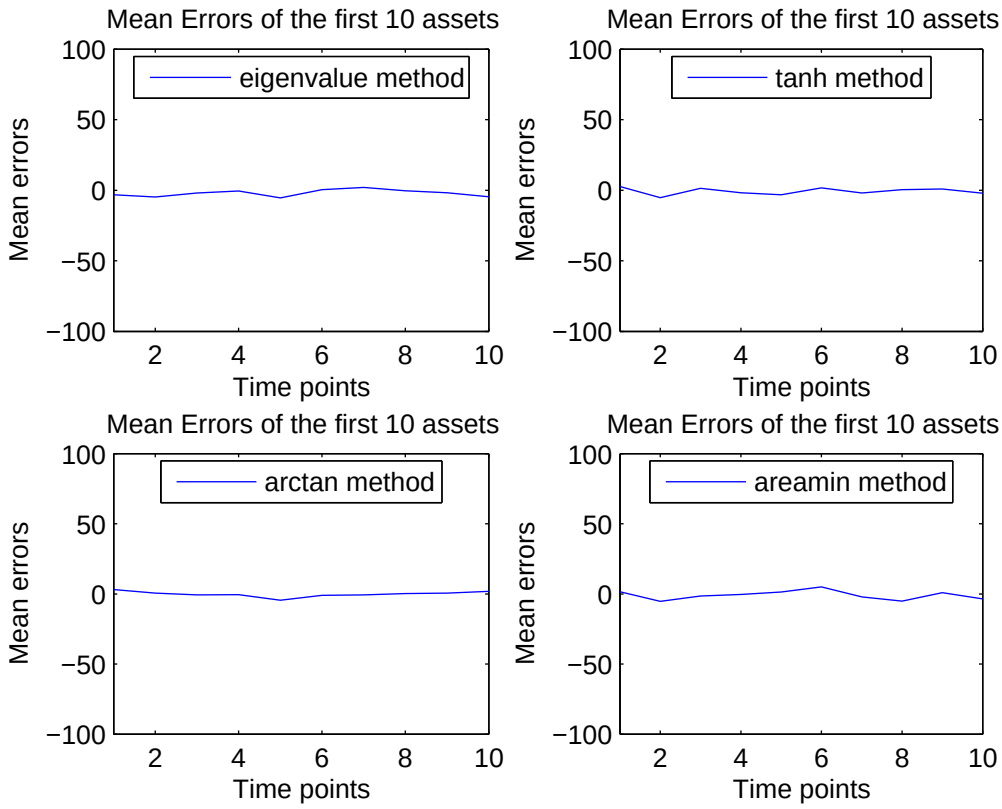

Figure 4.5: *Shows the average over the 20 estimation points of the mean errors between* $\boldsymbol{r}_{(t+1)}$ *(true observations) and* $\tilde{\boldsymbol{r}}_{(t+1)_i}$ *(predictors) of the first 10 assets using data matrix* $\boldsymbol{G}_r$ *- different axis.*

4.5 Summary of the results

This chapter was concerned with building a predictive model of asset returns and then testing the accuracy of these predictions with the true sample returns and the true observed returns of simulated time series data sets $\mathbf{M}_r$ and $\mathbf{G}_r$, with the aim of choosing the best transformation method. A statistical linear factor model was used, where all the parameters of the model were estimated from the simulated data. All the parameters and statistics calculated or estimated could be compared to the true sample parameters or statistics because they are known.

The common factors of the model constructed from each $\tilde{\Omega}_i$ were compared to the true sample common factors constructed from $\hat{\Omega}$ using the variance of the leading factor which represents the market mode. The eigenvalue ($\tilde{\Omega}_1$) and the arctan ($\tilde{\Omega}_2$) transformation methods resulted in the market index with variances that were close to the variance of the market index constructed from the true sample covariance matrix. $\tilde{\Omega}_1$ and $\tilde{\Omega}_2$ also gave the closest number of common factors, to the number of common factors calculated from the true

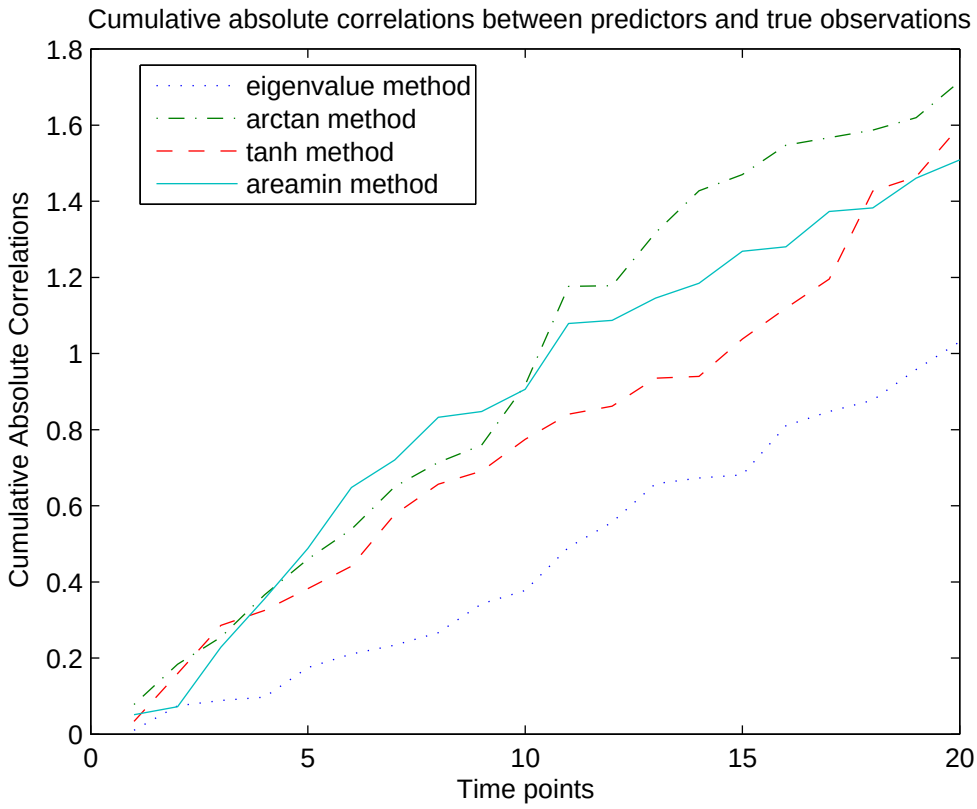

Figure 4.6: *Shows cumulative absolute values of the correlations between* $\boldsymbol{r}_{(t+1)}$ *(true observations) and* $\tilde{\boldsymbol{r}}_{(t+1)_i}$ *(predictors) for* $t \in \{0,\ 1,\ 2, ...,\ 19\}$. *The tanh covariance estimator gives consistently high correlations and the eigenvalue covariance estimator gives consistently low correlations using data matrix* $\boldsymbol{G}_r$

sample covariance matrix.

When the model is built, each $\tilde{\Omega}_i$ and $\hat{\Omega}$ are used in estimating the observations. Time series data $\mathbf{M}_r$ is used, and the asset returns predicted from each $\tilde{\Omega}_i$ are compared to the asset returns predicted from the true sample covariance matrix $\hat{\Omega}$ (calculated from data matrix $\mathbf{M}$) using correlations. The tanh transformation method and the area minimisation transformation method gave the highest correlations. Since correlations are used as a closeness measure of the predictors to the statistics calculated from the true sample covariance, this means that $\tilde{\Omega}_3$ and $\tilde{\Omega}_4$ are then closer to $\hat{\Omega}$ than the other two transformation methods.

The next comparison of the model was between the true observations and the predictors using correlations and mean-squared errors. The tanh transformation method and the arctan transformation method gave the highest correlations. The lowest mean-squared errors were given by the eigenvalue and the tanh transformation methods (results are summarised in table 4.1).

When a different data set $\mathbf{G}_r$, sampled using a Matlab function called UGARCHSIM is used, the results were significantly different in the sense that they have improved. Only the comparison between the predictors and the true observations was done, and the arctan and the eigenvalue gave the highest correlations and the lowest mean squared-errors (results are summarised in table 4.2).

In the simulated time series data $\mathbf{G}_r$, although the eigenvalue and the arctan transformation methods failed to be close to the asset returns predicted using $\hat{\Omega}$ (the true sample covariance matrix), they predicted the true observations better than the other two transformation methods.

The final comparison of these transformation methods in the simulated data is done in the next chapter, which is on portfolio construction, where a global minimum variance portfolio is used.

Chapter 5

Building a global minimum variance portfolio

The purpose of this chapter is to select an optimal portfolio from the simulated universe of $N = 300$ assets and length of data $T = 1200$. The different sample covariance estimators will be used in the construction of a global minimum variance portfolio (GMVP). This portfolio will then be used to compare the different sample covariance estimators.

Both inputs of the mean-variance portfolio selection are estimated from the simulated data matrix $\mathbf{M}_r$. Estimated parameters usually suffer from two sources of errors - the estimation and the specification errors [3]. In this work, no exogenous factors (like the market index) are used and no structural assumptions (as in the $k-$factor model) are imposed onto the estimation techniques used and therefore the impact of the specification error is minimal if it exists [21]. The error that may be applicable is the estimation error which is due to the finiteness of the simulated time series data. Its impact has been minimised by the length of the simulated time series data relative to the number of assets, that is $T = 4 * N = 1200$. It is for the reason of these errors which work in different directions [20], that only one estimated parameter, the second moment, is used in the construction of an optimal portfolio.

The same universe of $N = 300$ assets in a time series data of length $T = 1200$ described by the simulated asset returns data $\mathbf{M}_r$ constructed from $\mathbf{M}$ is used (see chapter 3 for a detailed description). As discussed in chapter 4, $\tilde{\mathbf{r}}_{t_i}$ is the predicted returns from $\mathbf{M}_r$ using estimation (or transformation) method i, $\tilde{\Omega}_i$ is the estimated covariance matrix from $\mathbf{M}_r$ using method i, and then let $\tilde{\mathbf{p}}_i$ be an optimal portfolio of $N = 300$ assets constructed from parameters estimated by method i. $\hat{\mathbf{p}}$, $\hat{\mathbf{r}}$ and $\hat{\Omega}$ represents an optimal portfolio, the predicted returns and the covariance matrix respectively, calculated from the simulated perfect market time series data $\mathbf{M}$.

The expected return and risk of portfolio $\tilde{\mathbf{p}}_i$, constructed from the imperfectly simulated time series data matrix $\mathbf{M}_r$ is defined by:

$$E(\tilde{\mathbf{p}}_i) = \tilde{\mathbf{w}}_i^T \tilde{\mathbf{r}}_i$$

$$\text{var}(\tilde{\mathbf{p}}_i) = \tilde{\mathbf{w}}_i^T \tilde{\Omega}_i \tilde{\mathbf{w}}_i$$

respectively, where $\tilde{\mathbf{r}}_i$ is a vector of average returns estimated using method i, $\tilde{\mathbf{w}}_i = [\omega_i^1, \omega_i^2, ..., \omega_i^n]^T$ is a vector of weights calculated using transformation method i, with the j-th weight ω_i^j representing the fraction of the total amount of invested capital that is placed on asset j and $\sum_j^n \omega_i^j = 1$. For example, $\tilde{\mathbf{p}}_1$ is a portfolio constructed from $\tilde{\Omega}_1$ and $\tilde{\mathbf{r}}_i$, using the eigenvalue transformation technique.

Amongst the different methods used in constructing optimal portfolios from a universe of N assets, the expected utility function [34] is used in this project. Since optimal portfolio construction is usually associated with the mean-variance framework of Markowitz [23], where one random variable is fixed and the portfolio is constructed by optimising the unfixed random variable, a brief description of the mean-variance framework is given in the next section, with the expected utility function discussed afterwards.

5.1 Mean-variance framework

The mean-variance optimisation problem is finding an optimal combination of return and risk, that an investor can consume under limited wealth. Since the return and risk of a portfolio are random variables that are difficult to control simultaneously, investors tend to optimise one variable while holding the other constant. For example, one investor might want to cap risk at a certain level, say $\sigma_{\mathbf{p}}$, and then construct a portfolio that gives the maximum return. This investor's portfolio optimisation problem will, for example, be formulated as follows, where $\mathbf{p}$ represents the portfolio:

$$\begin{aligned} \max E(\mathbf{p}) = \mathbf{w}^T \mu = \sum_{i=1}^{n} \omega_i \mu_i \\ \text{subject to } \mathbf{w}^T \Omega \mathbf{w} = \sigma_{\mathbf{p}}^2 \text{ and} \\ \text{subject to } \sum_{i=1}^{n} \omega_i = 1 \end{aligned}$$

Another investor might want to fix the return of the portfolio at $\mu_{\mathbf{p}}$ and construct a portfolio that will have the lowest risk given that $E(\mathbf{p}) = \mu_{\mathbf{p}}$. This investor's portfolio optimisation problem is then:

$$\min \text{var}(\mathbf{p}) = \mathbf{w}^T \Omega \mathbf{w} = \sum_{i=1}^{n} \sum_{j=1}^{n} \omega_i \sigma_{ij} \omega_j$$
$$\text{subject to } E(\mathbf{p}) = \mathbf{w}^T \mu = \mu_{\mathbf{p}} \text{ and}$$
$$\text{subject to } \sum_{i=1}^{n} \omega_i = 1$$

Note that both of these methods have two constraints - the fixed parameter (either $\mu_{\mathbf{p}}$ or $\sigma_{\mathbf{p}}$ but not both) and the weights (which must sum to unity). The above two approaches are well suited to studies that estimate one input (which is the unfixed parameter) of the mean-variance optimal portfolio framework, where the aim is to test the impact of the estimated parameter on whether it improves the portfolio or not. As an example [20] and [3] estimate the covariance matrix and use the global minimum variance portfolio, which does not require returns as an input, to test its impact.

In this project, none of the two approaches are used. A function of the expected return and risk of a portfolio, called the expected utility function, is used to formulate the portfolio optimisation problem. The definition of the expected utility function is given in the next section.

5.2 Expected utility function

Utility is used (amongst other uses) in economics to quantify satisfaction derived from the consumption of a combination of goods and/or services, subject to some constraint which is usually wealth. If W denotes the wealth of an individual, then W can be written as a function of the goods and/or services this individual can consume. Suppose that the wealth W can take on values w_i, for $i = 1, 2, ..., n$, each with probability π_i. Denote the utility derived from any wealth outcome w_i by $U(w_i)$. The expected utility from the risky outcome is

$$E(U(w_i)) = \begin{cases} \sum_i \pi(w_i) U(w_i) & \text{using a discrete density function} \\ \int \pi(w_i) U(w_i) dw_i & \text{using a continuous density function} \end{cases}$$

There are different kinds of utility functions that are used to model various kinds of (random) variables. Examples of utility functions used are:

1. The Power Utility Function:
$$U(W) = \frac{1}{1-\gamma} W^{1-\gamma} \quad , \ \gamma > 0 \ , \ \gamma \neq 0$$
2. The Logarithmic Utility Function:
$$U(W) = \ln(W)$$
3. The Quadratic Utility Function:
$$U(W) = W - \frac{b}{2} W^2 \quad , \ b > 0$$

4. The Negative Exponential Utility Function:

$$U(W) = a - be^{-cW} \quad , \; c > 0 \text{ is a risk aversion constant}$$

The individual's utility or the consumption strategy of risk and return is modeled by the negative exponential utility function, together with the assumption of normality of the (random) data [34]. Since $\mathbf{M}$ is sampled from the multivariate normal distribution, the simulated time series data matrix $\mathbf{M}_r$ is normally distributed. Consider the negative utility function given by:

$$U(W) = a - be^{-cW}$$

and let $a = 0,\ b = 1$ and $c = \gamma$. The negative utility function then becomes:

$$U(W) = -\exp(-\gamma W)$$

where $\gamma > 0$ is a risk aversion constant. The expected utility with respect to the distribution of the random data matrix (which is the normal distribution in this case) is calculated as follows:

$$\begin{aligned} E(U(W)) &= \int_{-\infty}^{\infty} U(W)\phi(W)dW \\ &= \int_{-\infty}^{\infty} (-\exp(-\gamma W))\phi(W)dW \\ &= -\frac{1}{\sigma_p\sqrt{2\pi}} \exp\left[-\gamma W - \frac{1}{2}\left(\frac{W-\mu_p}{\sigma_p}\right)^2\right] \\ &= -\exp\left(-\gamma\mu_p - \frac{1}{2}\gamma^2\sigma_p^2\right) \end{aligned}$$

where $\phi(W)$ is the normal density function defined in chapter 2. This is the function used in the formulation of the optimal portfolio problem in the next section.

5.3 Deriving the optimal portfolio weights

The expected utility function has been written as a function of the return and risk of a portfolio as shown in the following equation[34]:

$$E(U(W)) = -\exp\left(-\gamma\left(\mu_p - \frac{\gamma}{2}\sigma_p^2\right)\right)$$

where γ is the constant relative risk aversion coefficient. Since the aim is to maximise expected utility, maximising $E(U(W))$ therefore, means the same thing as maximising:

$$\mu_{\mathbf{p}} - \frac{\gamma}{2}\sigma_{\mathbf{p}}^2 = \mathbf{w}^T\mu - \frac{\gamma}{2}\mathbf{w}^T\Omega\mathbf{w}$$

$$\text{subject to } \sum_i \omega_i = 1$$

That is, one needs to solve the following quadratic programming problem:

$$\max\{\mathbf{w}^T\mu - \frac{\gamma}{2}\mathbf{w}^T\Omega\mathbf{w}\} \text{ subject to } \mathbf{w}^T\mathbf{1} = 1$$

This is solved using the Lagrangian multiplier λ, where the problem is transformed into the following equation:

$$\max\, L = \mathbf{w}^T\mu - \frac{\gamma}{2}\mathbf{w}^T\Omega\mathbf{w} - \lambda\left(\mathbf{w}^T\mathbf{1} - 1\right)$$

The above optimisation problem will be solved in the next section, where $E(\mathbf{r})$ will be used in the place of μ to stick to Wai Lee's notation.

5.3.1 Solving the optimization problem

The aim is to find $\mathbf{w}$ that will give an optimal L in the following equation:

$$L = \mathbf{w}^T E(\mathbf{r}) - \frac{\gamma}{2}\mathbf{w}^T\Omega\mathbf{w} - \lambda\left(\mathbf{w}^T\mathbf{1} - 1\right)$$

Optimization problems of this kind are usually solved by taking the first derivative and equating it to zero as follows [34]:

$$\frac{\partial L}{\partial \omega_n} = E(\mathbf{r})_n - \gamma\sum_{j=1}^{n}\omega_j\sigma_{ij} - \lambda = 0 \text{ and}$$

$$\frac{\partial L}{\partial \lambda} = 1 - \sum_{j=1}^{n}\omega_j = 0.$$

Re-arranging the partial derivative with respect to ω equation gives $E(\mathbf{r})_n - \gamma\sum_{j=1}^{n}\omega_j\sigma_{ij} - \lambda = 0 \implies \gamma\sum_{j=1}^{n}\omega_j\sigma_{ij} = E(\mathbf{r})_n - \lambda$, which can be written in matrix notation as follows:

$$\gamma\begin{pmatrix}\sigma_{11} & \sigma_{12} & \dots & \sigma_{1n}\\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n}\\ \vdots & \vdots & \vdots & \vdots\\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn}\end{pmatrix}\begin{pmatrix}\omega_1\\ \omega_2\\ \vdots\\ \omega_n\end{pmatrix} = \begin{pmatrix}E(r)_1\\ E(r)_2\\ \vdots\\ E(r)_n\end{pmatrix} - \begin{pmatrix}\lambda\\ \lambda\\ \vdots\\ \lambda\end{pmatrix}$$

which is then simplified to the following equations

$$\begin{aligned}\gamma\Omega\mathbf{w} &= E(\mathbf{r}) - \lambda\mathbf{.1}\\ \implies \Omega\mathbf{w} &= \frac{1}{\gamma}\left(E(\mathbf{r}) - \lambda\mathbf{.1}\right)\\ \Omega^{-1}\Omega\mathbf{w} &= \Omega^{-1}.\frac{1}{\gamma}\left(E(\mathbf{r}) - \lambda\mathbf{.1}\right)\\ \therefore \mathbf{w} &= \Omega^{-1}.\frac{1}{\gamma}\left(E(\mathbf{r}) - \lambda\mathbf{.1}\right).\end{aligned}$$

The partial derivative with respect to λ can be simplified as follows:

$$\begin{aligned}\frac{\partial L}{\partial \lambda} &= -(\mathbf{w}^T\mathbf{1} - 1) = 0\\ \implies \mathbf{w}^T\mathbf{1} &= 1\end{aligned}$$

There are two equations (rewritten below as equations 5.1 and 5.2) and two unknowns, $\mathbf{w}$ and λ,

$$\mathbf{w} = \Omega^{-1}.\frac{1}{\gamma}\left(E(\mathbf{r}) - \lambda\mathbf{.1}\right) \tag{5.1}$$

$$\mathbf{w}^T\mathbf{1} = 1 \tag{5.2}$$

which can be solved simultaneously. Substituting equation 5.1 in equation 5.2 results in:

$$\begin{aligned}
\left(\frac{1}{\gamma}\Omega^{-1}\left(E(\mathbf{r}) - \lambda\mathbf{1}\right).\mathbf{1}\right)^T &= 1 \\
\mathbf{1}^T\Omega^{-1}E(\mathbf{r}) - \mathbf{1}^T\Omega^{-1}\mathbf{1}\lambda &= \gamma \\
\mathbf{1}^T\Omega^{-1}\mathbf{1}\lambda &= \mathbf{1}^T\Omega^{-1}E(\mathbf{r}) - \gamma \\
\therefore \lambda &= \frac{\mathbf{1}^T\Omega^{-1}E(\mathbf{r})}{\mathbf{1}^T\Omega^{-1}\mathbf{1}} - \frac{\gamma}{\mathbf{1}^T\Omega^{-1}\mathbf{1}}
\end{aligned}$$

Since the aim is to calculate an optimal portfolio $\mathbf{w}$, find it by substituting λ into equation 5.1

$$\begin{aligned}
\mathbf{w} &= \frac{\Omega^{-1}}{\gamma}\left(E(\mathbf{r}) - \left(\frac{\mathbf{1}^T\Omega^{-1}E(\mathbf{r})}{\mathbf{1}^T\Omega^{-1}\mathbf{1}} - \frac{\gamma}{\mathbf{1}^T\Omega^{-1}\mathbf{1}}\right).\mathbf{1}\right) \\
&= \frac{\Omega^{-1}E(\mathbf{r})}{\gamma} - \frac{\Omega^{-1}}{\gamma}\left(\frac{\mathbf{1}^T\Omega^{-1}E(\mathbf{r})}{\mathbf{1}^T\Omega^{-1}\mathbf{1}} - \frac{\gamma}{\mathbf{1}^T\Omega^{-1}\mathbf{1}}\right).\mathbf{1} \\
&= \frac{\Omega^{-1}E(\mathbf{r})}{\gamma} - \frac{\Omega^{-1}.\mathbf{1}}{\gamma}\frac{\mathbf{1}^T\Omega^{-1}E(\mathbf{r})}{\mathbf{1}^T\Omega^{-1}\mathbf{1}} + \frac{\Omega^{-1}.\mathbf{1}}{\mathbf{1}^T\Omega^{-1}\mathbf{1}} \\
&= \left(\frac{\Omega^{-1}E(\mathbf{r})}{\gamma}.\frac{\mathbf{1}^T\Omega^{-1}E(\mathbf{r})}{\mathbf{1}^T\Omega^{-1}E(\mathbf{r})}\right) - \left(\frac{\mathbf{1}^T\Omega^{-1}E(\mathbf{r})}{\gamma} - 1\right)\frac{\Omega^{-1}.\mathbf{1}}{\mathbf{1}^T\Omega^{-1}\mathbf{1}} \\
&= \left(1 - \frac{\mathbf{1}^T\Omega^{-1}E(\mathbf{r})}{\gamma}\right)\frac{\Omega^{-1}.1}{\mathbf{1}^T\Omega^{-1}\mathbf{1}} + \left(\frac{\mathbf{1}^T\Omega^{-1}E(\mathbf{r})}{\gamma}\right).\frac{\Omega^{-1}E(\mathbf{r})}{\mathbf{1}^T\Omega^{-1}E(\mathbf{r})}
\end{aligned}$$

and the latter equation gives the general form of an optimal portfolio, split into a part that has $E(\mathbf{r})$ and a part that doesn't.

5.4 Global minimum-variance portfolio

The above optimal weights equation is known as the Mutual Fund Separation Theorem because the optimal portfolio $\mathbf{w}$ is made up of two distinct terms or portfolios - the one with expected returns vector and the other without the expected returns vector. That is, $\mathbf{w}$ is a weighted average of the following two quantities[34]: $\frac{\Omega^{-1}.\mathbf{1}}{\mathbf{1}^T\Omega^{-1}\mathbf{1}}$ and $\frac{\Omega^{-1}E(\mathbf{r})}{\mathbf{1}^T\Omega^{-1}E(\mathbf{r})}$, where the weights are $\left(1 - \frac{\mathbf{1}^T\Omega^{-1}E(\mathbf{r})}{\gamma}\right)$ and $\left(\frac{\mathbf{1}^T\Omega^{-1}E(\mathbf{r})}{\gamma}\right)$, with the weights summing to unity [34].

Represent the different terms by ω_G, ω and α as shown below:

$$\omega_G = \frac{\Omega^{-1}.\mathbf{1}}{\mathbf{1}^T\Omega^{-1}\mathbf{1}}$$
$$\omega = \frac{\Omega^{-1}E(\mathbf{r})}{\mathbf{1}^T\Omega^{-1}E(\mathbf{r})}$$
$$\alpha = \left(\frac{\mathbf{1}^T\Omega^{-1}E(\mathbf{r})}{\gamma}\right)$$

The optimal portfolio can then be written as:

$$\mathbf{w} = (1-\alpha)\omega_G + \alpha\omega = \omega_G + \alpha(\omega - \omega_G)$$

Wai Lee does some interesting matrix algebra and further separates $\alpha(\omega - \omega_G)$ into ω_S and ω_T, with the optimal weights equation becoming (see [34] or Appendix C for a full derivation)

$$\mathbf{w} = \omega_G + \omega_S + \omega_T,$$

where ω_G is called the global minimum-variance portfolio (GMVP) and is located at the apex of the standard mean-variance frontier, the subscripts G, S and T are for the global minimum variance, strategic and tactical portfolios [34].

Since the purpose of this chapter is to construct an optimal portfolio from estimated parameters, and it is known that estimated parameters suffer from estimation and specification errors, ω_G is used as an optimal portfolio, and not $\mathbf{w}$. As one can see from above, the general optimal weights are a function of the expected returns and the covariance matrix. It is said that portfolio weights are very sensitive to the expected returns vector and therefore using $\mathbf{w}$ as an optimal portfolio will mean putting more weight on $\tilde{\mathbf{r}}_i$ [3]. The global minimum variance portfolio, ω_G, will be used because it is independent of the expected returns and since the purpose is to test the covariance matrix estimators.

5.5 Building the risk model

The risk model is built using the global minimum-variance portfolio given by:

$$\tilde{\omega}_G^i = \frac{\tilde{\Omega}_i^{-1}\mathbf{1}}{\mathbf{1}^T\tilde{\Omega}_i^{-1}\mathbf{1}}$$

where each $\tilde{\omega}_G^i$, for $i \in \{1,\ 2,\ 3,\ 4\}$, is calculated from the different sample covariance estimators. Since each of the global minimum variance portfolios, $\tilde{\omega}_G^i$, is used in building the risk model, the returns predicted in chapter 4 will not be used in the portfolio construction for reasons of the different errors (estimation and specification errors) mentioned at the beginning of this chapter [3].

The model is built on the same simulated data matrix $\mathbf{M}_r$ constructed from $\mathbf{M}$ used in the previous two chapters. The reason for using the data set $\mathbf{M}_r$ is for continuity and consistency across the different chapters of this project. A time period from t to $t+1$ (one period model) is used, where the covariance

matrix estimation is done at time t. The portfolio risk is used to choose the best covariance estimator (Euclidean distance was used in chapter 3 and correlations were used in chapter 4). The portfolio risk at time t or ex-ante risk calculated using transformation method i or estimated covariance matrix $\tilde{\Omega}_i$, denoted by $\tilde{\sigma}^2_{t_i}$, is calculated by the following equation:

$$\tilde{\sigma}^2_{t_i} = \left(\left(\tilde{\omega}^i_G\right)^T \tilde{\Omega}_i^{-1} \tilde{\omega}^i_G \right)_t$$

where the $\tilde{\omega}^i_G$ are the portfolio weights calculated from the estimated sample covariance matrices $\tilde{\Omega}_i$ at time t. The portfolio risk at time $t+1$ or ex-post risk, denoted by $\tilde{\sigma}^2_{(t+1)_i}$, is calculated by the following equation:

$$\tilde{\sigma}^2_{(t+1)_i} = \left(\left(\tilde{\omega}^i_G\right)^T_{\mathrm{W}} \tilde{\Omega}_i^{-1} \left(\tilde{\omega}^i_G\right)_{\mathrm{W}} \right)_{t+1}$$

where $\tilde{\Omega}_i$ is the same covariance matrix estimated at time t and the $\left(\tilde{\omega}^i_G\right)_{\mathrm{W}}$ are the wealth adjusted weights. That is, the same weights calculated at time t are adjusted by wealth and used as portfolio weights at time $t+1$.

The simulation results are given in the next section, where 1000 simulations were used.

5.6 Results of the model

As in the expected returns prediction chapter, there are two levels of comparing the results. The first level of comparison is for continuity from the covariance estimation techniques and the returns prediction chapters, where the aim is to find the best transformation method or the covariance estimation technique that results in statistics estimated from $\mathbf{M}_r$ that are closer to the true sample statistics calculated from $\mathbf{M}$. In this first level, the model is built from each covariance matrix estimator $\tilde{\Omega}_i$, resulting in $\tilde{\omega}^i_G$, from which $\tilde{\sigma}^2_{t_i}$ (the ex-ante risk of portfolio $\tilde{\omega}^i_G$ at time t) is calculated, and from the true sample covariance matrix $\hat{\Omega}$ resulting in $\hat{\omega}_G$ from which $\hat{\sigma}^2_t$ (the ex-ante risk of portfolio $\hat{\omega}_G$ at time t) is calculated. The following statistic:

$$|\hat{\sigma}^2_t - \tilde{\sigma}^2_{t_i}| \quad \text{for } i \in \{1,\ 2,\ 3,\ 4\}$$

is used as a closeness measure of each $\tilde{\Omega}_i$(estimators) to $\hat{\Omega}$(true sample).

The second level of comparison calculates $\tilde{\sigma}^2_{(t+1)_i}$ (the ex-post risk of portfolio $\tilde{\omega}^i_G$ at time t+1) and looks at the difference between the ex-ante and the ex-post risks of each method i by calculating the following statistic:

$$|\tilde{\sigma}^2_{(t+1)_i} - \tilde{\sigma}^2_{t_i}|$$

The assumption made is that, the risk of a portfolio shouldn't change by much in one time period. That is, there shouldn't be much difference between $\tilde{\sigma}^2_{t_i}$ and $\tilde{\sigma}^2_{(t+1)_i}$ in one time period. This implies that the stable estimator will be the one that gives the minimum $|\tilde{\sigma}^2_{(t+1)_i} - \tilde{\sigma}^2_{t_i}|$. The second level also looks at short interests, defined as the sum of the negative weights, of the wealth adjusted weights, calculated from the four sample covariance estimators. The

Method	Risk	Std deviation
Sample	0.0000	0.0000
Eigenvalue	0.0000	0.0000
Arctan Shrinkage	0.0209	0.0043
Tanh Shrinkage	0.0163	0.0075
Area Minimisation	0.0000	0.0000

Table 5.1: *Ex-ante risk of each method i (and standard deviations across simulations), compared to the ex-ante risk of the true sample covariance matrix at time t*

estimators are compared against each other and the estimator with the lowest short selling (short interest) is the best.

5.6.1 Simulation results

In this simulation environment, the complete or true sample data set is known and therefore all the true sample parameters are known. Building the risk model from the complete data set $\mathbf{M}$ helps one know what the true sample portfolio should be. Parameters estimated from the incomplete data set $\mathbf{M}_r$ should approximately be close to those calculated from the complete data set $\mathbf{M}$ since the transformation methods are attempting to restore properties of $\mathbf{M}$ that are due to missing data points in $\mathbf{M}_r$. That is why the first comparison is between $\hat{\sigma}_t^2$ and $\tilde{\sigma}_{t_i}^2$ - the best transformation method needs to be chosen.

Choosing the best covariance matrix estimator

At time t, the risk model results in the following ex-ante risks: $\hat{\sigma}_t^2,\ \tilde{\sigma}_{t_1}^2,\ \tilde{\sigma}_{t_2}^2,\ \tilde{\sigma}_{t_3}^2,\ \tilde{\sigma}_{t_4}^2$, which are calculated from $\hat{\Omega},\ \tilde{\Omega}_1,\ \tilde{\Omega}_2,\ \tilde{\Omega}_3,\ \tilde{\Omega}_4$ respectively, where $\hat{\sigma}_t^2$ and $\hat{\Omega}$ are the true sample statistics from the simulated perfect market represented by $\mathbf{M}$, and $\tilde{\sigma}_{t_i}^2$ and $\tilde{\Omega}_i$ are their estimates from the data set $\mathbf{M}_r$. The true sample risk is calculated as follows:

$$\hat{\sigma}_t^2 = \left((\hat{\omega}_G)^T \hat{\Omega}^{-1} \hat{\omega}_G\right)_t$$

where $\hat{\omega}_G = \frac{\hat{\Omega}^{-1}.\mathbf{1}}{\mathbf{1}^T\hat{\Omega}^{-1}\mathbf{1}}$. The estimators of the true sample risk are calculated by:

$$\tilde{\sigma}_{t_i}^2 = \left((\tilde{\omega}_G^i)^T \tilde{\Omega}_i^{-1} \tilde{\omega}_G^i\right)_t$$

where $\tilde{\omega}_G^i = \frac{\tilde{\Omega}_i^{-1}}{\mathbf{1}^T\tilde{\Omega}_i^{-1}\mathbf{1}}$. Since $\tilde{\Omega}_i$ is known (estimated in chapter 3) and $\tilde{\omega}_G^i$ is a function of $\tilde{\Omega}_i$ only, this means that $\tilde{\omega}_G^i$ can be calculated. Since $\tilde{\Omega}_i$ and $\tilde{\omega}_G^i$ are both known and $\tilde{\sigma}_{t_i}^2$ is a function of both of them, this implies that $\tilde{\sigma}_{t_i}^2$ can be calculated.

The following statistic $|\hat{\sigma}_t^2 - \tilde{\sigma}_{t_i}^2|$ is used as a basis of comparison and the simulation results are given in table 5.1.

The GMVP calculated from the true sample covariance $\hat{\Omega}$ results in $\hat{\sigma}_t^2 = 0$. Therefore, $|\hat{\sigma}_t^2 - \tilde{\sigma}_{t_1}^2| = |\hat{\sigma}_t^2 - \tilde{\sigma}_{t_4}^2| = 0$, that is, $\tilde{\sigma}_{t_1}^2 = \tilde{\sigma}_{t_4}^2 = 0$. This means that $\tilde{\Omega}_1$ and $\tilde{\Omega}_4$ are closer to $\hat{\Omega}$ than the other two transformation methods.

Wealth adjusted weights

The ex-post risk depends on the portfolio calculated at time t. This portfolio is then adjusted by wealth and so it is worth dedicating a section into calculating the wealth adjusted weights before calculating the ex-post risk.

Assume the following information is available at time t

- initial wealth $W(t)$ (scalar)
- initial stock prices $\mathbf{s}(t)$ (a $1 \times N$ vector)
- portfolios $\tilde{\omega}_G^i(t)$ for each estimator, which have been calculated above (a $1 \times N$ vector)

From the above information, one can calculate the fraction of each stock that form the portfolio. This is represented by the vector $\Pi(t) = \{\pi_1(t), \pi_2(t), ..., \pi_n(t)\}$, which is calculated by the following formula:

$$\pi_i(t) = \frac{W(t) * \omega_i(t)}{s_i(t)}$$

At time $t+1$, assume that the evolution of the price process is known, that is, assume that the returns for time $t+1$ are given as $\mathbf{r}(t+1) = \{r_1(t+1),\ r_2(t+1), ...,\ r_n(t+1)\}$ and therefore

$$s_i(t+1) = s_i(t) \exp^{r_i(t+1)}$$

The wealth at time $t+1$ is $W(t+1) = \Pi \mathbf{s}(t+1)$. The portfolio that was held at time t is no longer applicable and therefore needs to be adjusted with the new information. The i-th adjusted weight at time $t+1$ is

$$w_W^i(t+1) = \frac{\pi_i(t+1) * s_i(t+1)}{\Pi \mathbf{s}(t+1)} \text{ for } i \in \{1,\ 2, ...,\ N\}$$

These wealth adjusted weights are used in the simulation part only. In the empirical part (chapter 7), return adjusted weights calculated as follows:

$$\tilde{w}_{\text{ret}}^i(t+1) = \tilde{w}_G^i(t) \exp^{r(t+1)}$$

are used in the analysis. The wealth adjusted weights are then used to calculate the ex-post risk in the next section.

Method	**Risk**	**Std deviation**
Eigenvalue	0.1057	0.0016
Arctan Shrinkage	0.1035	0.0014
Tanh Shrinkage	0.1012	0.0013
Area Minimisation	0.1007	0.0011

Table 5.2: *Ex-post risk of the different covariance matrix estimation techniques (and standard deviations across simulations) compared against each other at time* $t+1$

One time period comparison

At time $t+1$, the risk model results in the following post risks: $\tilde{\sigma}^2_{(t+1)_1}$, $\tilde{\sigma}^2_{(t+1)_2}$, $\tilde{\sigma}^2_{(t+1)_3}$, $\tilde{\sigma}^2_{(t+1)_4}$, which are calculated from $\tilde{\Omega}_1$, $\tilde{\Omega}_2$, $\tilde{\Omega}_3$, $\tilde{\Omega}_4$ respectively, where each $\tilde{\Omega}_i$ was calculated at time t. That is, the same covariance matrices calculated at time t are also used at time $t+1$. These ex-post risks are calculated using the following equation:

$$\tilde{\sigma}^2_{(t+1)_i} = \left(\left(\tilde{\omega}^i_G\right)^T_{\mathrm{W}} \tilde{\Omega}_i^{-1} \left(\tilde{\omega}^i_G\right)_{\mathrm{W}} \right)_{t+1}$$

Table 5.2 gives the average ex-post risk over 1000 simulations and the shrinkage by tanh and the area minimisation give the lowest post risks with minimal variability over the simulations.

In one time period, the risk calculated at time t is not suppose to be very different to the one calculated at time $t+1$. The following statistic $|\tilde{\sigma}_{(t+1)_i} - \tilde{\sigma}_{t_i}|$, called the absolute actual risk difference, is used as a basis of comparing the two risks and the simulation results are given in table 5.3. Both shrinkage transformation methods seem to be giving the post risk that is closer to their respective ante risks.

The results of the absolute cumulative risk difference between times t and $t+1$ for each method i calculated by:

$$\sum_{j=1}^{1000} \left| \tilde{\sigma}_{(t+1)_i} - \tilde{\sigma}_{t_i} \right|_j$$

are plotted in figure 5.1. The lines that represent the eigenvalue and the area minimisation transformation methods are higher than the lines representing the two shrinkage transformation methods. The results of the average of the absolute cumulative risk difference between times t and $t+1$ are summarised in table 5.3.

Fig 5.2 plots the wealth adjusted weights on separate sets of axis. Comparing the wealth adjusted portfolios using a graph, one can see that the area minimisation method has spikes of huge short positions than the other three transformation methods, as evident in figure 5.2. Nothing conclusive can be said by looking at the graphs only. Need to look at the numbers given by these portfolios.

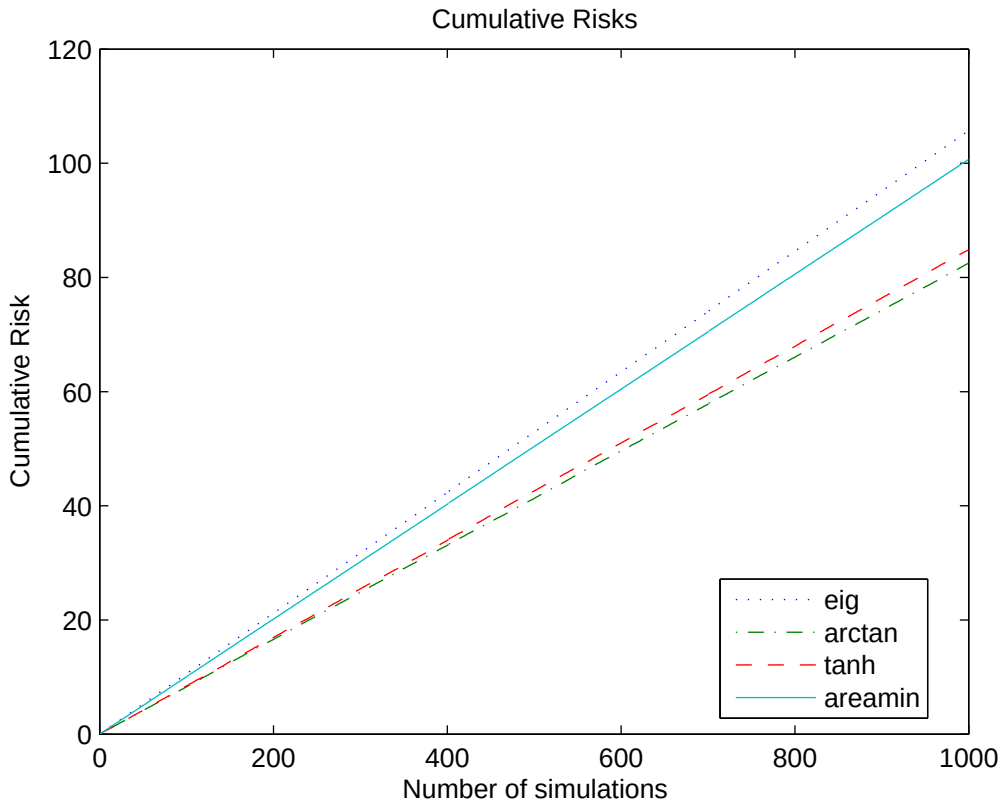

Figure 5.1: *Shows the cumulative post risk calculated across the 1000 simulations using each of the covariance matrix estimators*

Method	**Risk**	**Std deviation**
Eigenvalue	0.1057	0.0016
Arctan Shrinkage	0.0826	0.0050
Tanh Shrinkage	0.0849	0.0089
Area Minimisation	0.1007	0.0010

Table 5.3: *The average of the absolute value of the difference between ante and post risks (and standard deviations across simulations)*

Short selling is hard to carry out in practice and in some markets impossible [3]. Attractive portfolios are therefore those that require the least amount of short selling and, similarly, require reasonable (sum of long positions close to unity) amounts of long positions since the portfolio weights must sum to unity.

Summing all the negative weights, the sum of which is called short interest [20], results in $\tilde{\Omega}_1$ and $\tilde{\Omega}_2$ giving portfolios with the least amount of short selling. The results are summarised in table 5.4. The short interest of the area minimisation transformation method, as observed in figure 5.2, is extremely huge at approximately 286.49%. A short interest of 13.41%, for example, means that for

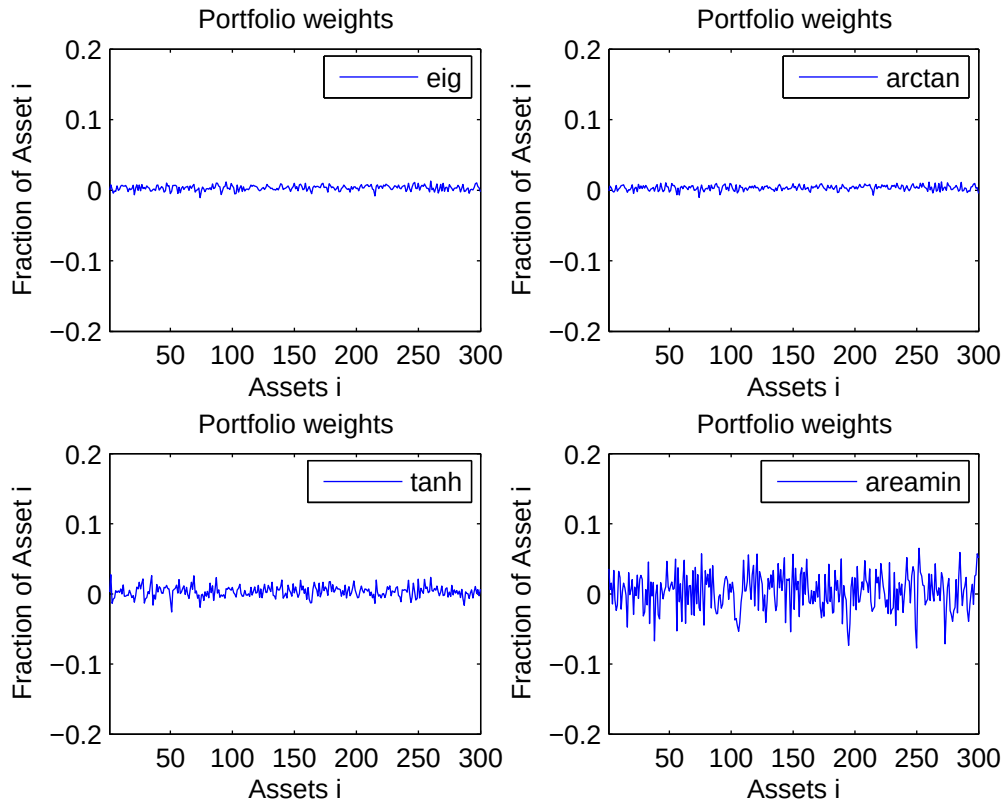

Figure 5.2: *Shows the fraction of asset i in the global minimum variance optimal portfolio for covariance matrix estimator on separate sets of axes*

Method	**Short interest**
Eigenvalue	13.41%
Arctan Shrinkage	12.22%
Tanh Shrinkage	51.89%
Area Minimisation	286.49%

Table 5.4: *Shows the total amount of short selling required from portfolios constructed from each covariance estimator*

every rand invested in the portfolio, an amount of 13.41 cents worth of stocks is short, while buying R1.1341 worth of other stocks [20]. Therefore, $\hat{\Omega}_1$ and $\hat{\Omega}_2$ win in this contest by giving the least amounts of short selling, that is, they result in more reasonable portfolios than the other two transformation methods.

Chapter 6

Simulation concluded

The same 1200×300 data matrix sampled from a multivariate normal distribution has been used throughout the simulation part. This project is split into two parts - the simulation section (being concluded) dealt with in chapters 3, 4 and 5, and the empirical section (to be tackled in chapter 7). The purpose of the simulation part was:

1. to introduce the covariance estimation techniques
2. to choose the best covariance estimator to tackle measurement error in emerging markets empirical data
3. to build a statistical linear factor model
4. to build a global minimum variance portfolio

The four competing covariance estimation techniques are:

1. The eigenvalue covariance estimator $\left(\tilde{\Omega}_1\right)$
2. The arctan shrinkage covariance estimator $\left(\tilde{\Omega}_2\right)$
3. The tanh shrinkage covariance estimator $\left(\tilde{\Omega}_3\right)$
4. The area minimisation covariance estimator $\left(\tilde{\Omega}_4\right)$

These covariance estimators are named based on the method used to transform a non-positive definite covariance matrix into a positive definite covariance matrix.

The four covariance matrix estimators are compared against each other to choose the best estimator using:

1. Euclidean distance from each estimated covariance matrix calculated from the incomplete data set $\mathbf{M}_r$ to the true sample covariance matrix calculated from the complete data set $\mathbf{M}$ (chapter 3)

2. correlations and mean-squared errors between the estimated returns (calculated from $\tilde{\Omega}_i$) and the true sample returns (calculated from $\hat{\Omega}$) both of which are estimated from a statistical linear factor model (chapter 4)

3. risk (ex-ante and ex-post) and short interest calculated from the portfolio weights estimated from $\tilde{\Omega}_i$ and the true sample portfolio weights estimated from $\hat{\Omega}$. These are both global minimum variance portfolio weights (chapter 5).

In each of these three areas in which the estimated covariance matrices are compared, the calculated or estimated statistic (using simulated time series data $\mathbf{M}_r$) was compared to the true sample statistic (using simulated time series data $\mathbf{M}$).

In the covariance estimation techniques, Euclidean distance and its variability across simulations were used to compare each of the sample covariance matrix estimators to the true sample covariance matrix. The actual and the cumulative Euclidean distance both gave the same results. The two covariance estimation techniques which resulted in the lowest Euclidean distance that varied the least across 1000 simulations were the eigenvalue covariance estimator and the arctan shrinkage covariance estimator. There was no significant difference between the two but the arctan performed marginally better than the eigenvalue.

A statistical linear factor model was then built using the four covariance matrix estimators to model expected asset returns. Correlations between the asset returns predicted from the estimated sample covariance matrices $\tilde{\Omega}_1$, $\tilde{\Omega}_2$, $\tilde{\Omega}_3$, $\tilde{\Omega}_4$ and the asset returns predicted from the true sample covariance matrix $\hat{\Omega}$, were used as one method of choosing the best transformation method. Other comparison methods used in the statistical model chapter were the mean-squared errors, the variance of the leading factor (market mode) constructed from each method i and the number of common factors k to be used in the model. The results are summarised in table 6.1. The best estimator must give the highest correlations, the lowest square root of the mean-squared error (SQRT MSE), a variance of the leading factor closer to 4.8564 (the variance of the leading factor from data matrix $\mathbf{M}$) and a k closer to 46 (the number of common factors calculated from data matrix $\mathbf{M}$). The eigenvalue and the arctan shrinkage covariance estimators did better than the other two estimators in most of the areas used to compare them in the returns prediction chapter, with the arctan performing better in the two important areas - the correlations and the SQRT MSE.

An optimal portfolio was constructed using the global minimum variance portfolio, where only the second moment was used as an input. Portfolio risk (difference between ante and post) and the amount of short interest, were used in comparing the estimators to the true sample statistic. The arctan shrinkage covariance estimator gave the lowest risk difference and the lowest amount of short interest. Although the eigenvalue covariance estimator gave the highest risk difference, it results in a reasonable short interest than the other two estimators and for this reason is therefore the best relative to the tanh shrinkage and area minimisation estimators.

Method	**Correlations**	**SQRT MSE**	**Variance(4.8564)**	k **(46)**
Eigenvalue	15.14%	6.6	4.9457	40.549
Arctan Shrinkage	17.37%	5.7	4.2806	40.300
Tanh Shrinkage	13.91%	7.1	3.7187	31.790
Area Minimisation	14.04%	8.7	4.6507	39.325

Table 6.1: *The variances of the leading factors and the number of common factors are included as a way of comparing the different covariance estimators with the true sample covariance estimator. Also, the model's predictive power is tested using absolute mean correlations and the square root of the mean squared errors*

Method	**Ante risk**	**Post risk**	**Post-Ante**	**Short Interest**
Eigenvalue	0.0000	0.1057	0.1057	13.41%
Arctan Shrinkage	0.0209	0.1035	0.0826	12.22%
Tanh Shrinkage	0.0163	0.1012	0.0849	51.89%
Area Minimisation	0.0000	0.1007	0.1007	286.49%

Table 6.2: *Comparing the estimators using the risk difference and short interest*

Note that the results summarised in this chapter are based on a simulated 1200×300 data matrix sampled from a multivariate normal distribution, where the method of dealing with missing data points was pairwise deletion. Although $\mathbf{M}_r$ is a random data set, the results may not be generalised to, for example, any size data matrix sampled from any distribution like the uniform or the gamma, where the method of dealing with missing data is maybe listwise or casewise deletion.

Although the arctan shrinkage covariance estimator did marginally better than the eigenvalue covariance estimator in the simulation part, it will be premature to choose it and discard the eigenvalue estimator (and the other two estimators) because the empirical data might have different properties to those of a simulated data set. The suggestion, therefore, is to use all four estimators in modeling empirical data from the JSE. As to the simulation data set used, there's minimal difference between the arctan shrinkage and the eigenvalue estimators.

Chapter 7

Empirical results

The Johannesburg Stock Exchange (JSE) is one of the 20 largest national securities markets in the world and the largest in Africa [36]. Analysing empirical time series data of emerging markets requires different approaches than those required in analysing empirical time series data from developed markets because some data anomalies that exist in emerging markets time series data are not found in empirical data from developed markets. This chapter presents empirical evidence on the performance of the statistical factor model and the risk model developed in previous chapters from simulated data. Although different methods of testing the model, like correlations, the mean-squared errors, risk or short selling are used, a better test of the usefulness of any model is its ability to explain performance out-of-sample [31].

The modeling of expected returns done in this project depends on principal component analysis. Principal component analysis is a data reduction algorithm designed for systems which have a high degree of built in correlation, with the first three components accounting for most of the total variation present in the original data set [35]. Since there is no inherent correlation structure in equity market time series data, principal component analysis may not be an appropriate modeling tool in the expected return estimation [35]. That is, the results of the statistical model may not be as good as one might hope.

There are authors who have previously analysed the South African market time series data using different analytic tools. Some of the work of these authors is general enough to be applicable to all markets and some is applicable only to emerging markets. Examples of the authors who have modeled the JSE time series data include Barr and Bradfield who propose the use of Capital Asset Pricing Model with a thinly traded beta estimator to generate estimates of the expected asset returns vector [2], Wilcox and Gebbie test the agreement of the measured empirical data correlation matrices to the random matrix theory predictions, van Rensberg and Robertson investigate a factor model in the spirit of the characteristic based model arguments of Daniel and Titman (1997) [33] and Haugen and Baker (1996) [15]. In this project, an analytic tool based on the works of Lindskog [21], Chen and McInroy [8], Haugen and Baker [15], Ledoit and Wolf [20] and Bengtsson and Holst [3] is used to analyse the JSE time series data.

	epoch 1	**epoch 2**	**epoch 3**	**epoch 4**	**epoch 5**	**Epoch 6**
start date	19930101	19940101	19950101	19960101	19970101	19980101
end date	19971231	19981231	19991231	20001231	20011231	20030630
trading days	1303	1303	1304	1304	1303	1431
no. of stocks	442	442	442	442	442	442
no. of filtered stocks	129	154	172	186	208	227
% of stocks used	29%	35%	39%	42%	47%	51%

Table 7.1: *Shows the 6 demarcations of the JSE time series data used in the analysis, and the percentage of assets in each epoch*

7.1 Data

Daily data consisting of 2737 trading days (for some stocks), from 1 January 1993 to 30 June 2003, or an equivalent of 10.5 years and 442 stocks is used. This data set is windowed to create six sets of overlapping time series data, each consisting of five years of daily data [36] with the exception of the last window which contains 5.5 years of daily data (for a more detailed qualitative description of the JSE see [36]).

The different windows are 1993 to 1997, 1994 to 1998, 1995 to 1999, 1996 to 2000, 1997 to 2001 and 1998 to June 2003 and are labeled epoch 1 to epoch 6 [36]. Since the JSE is an emerging market, not all listed assets have the same length of history in the period under consideration and neither do they trade in the same frequency. Based on the trading frequency of the stocks, infrequently trading assets were filtered out, where the filtering condition was that stocks must trade at least 80% in the relevant epoch.

A summary of each of the different epochs is given in table 7.1[1]. Note that as one moves from epoch 1 to epoch 6, the number of filtered stocks increase. This shows that the JSE is a developing market and that more and more companies have started to list in later years.

7.2 Methodology

Each epoch is analysed as a separate time series data, where the four sample covariance matrix estimators are calculated, the asset returns predicted and the global minimum variance portfolio constructed. An individual epoch is split into an in-sample period, where all the parameters of the model are estimated and the model built, and an out-of-sample period, where the model is tested. The length of the out-of-sample period used is 75 days, and $h = 100$ days moving estimation windows of equal length are used in the in-sample period. Note that the in-sample period, the out-of-sample period and the modeling are as described in section 4.3 and section 5.5. The only difference is the length of the respective periods and the number of moving estimation windows h.

[1] adopted from table 1, page 8 of [36]

	epoch 1	**epoch 2**	**epoch 3**	**epoch 4**	**epoch 5**	**epoch 6**
eigenvalue	5.78%	4.88%	5.97%	5.46%	5.44%	5.93%
arctan	7.42%	7.41%	6.24%	5.61%	5.63%	5.56%
tanh	6.01%	5.96%	5.50%	5.73%	5.47%	5.49%
area min	N/A	7.26%	6.58%	7.72%	6.28%	5.97%
best	**arctan**	**arctan**	**area min**	**area min**	**area min**	**area min**

Table 7.2: *Shows the average absolute correlations between the estimated returns and the true observed returns across the different epochs*

Using epoch 1 as an example to describe the methodology, the models are built on the in-sample period data which consists of 1228 (1303 -75) days of data. As mentioned in the simulation part, the variable h is chosen depending on the frequency of the data and since the data set used is daily, $h = 100$ days is used (longer h for example 200 or 250 results in ill-conditioned covariance matrix estimators) in building a statistical factor based linear model and the minimum risk portfolio model.

In the statistical linear factor model for example, if the first estimation is done on 30 September 1997, then the resulting vectors of predicted returns are estimators of the true sample returns vector observed on 1 October 1997 (or the next trading day after 30 September 1997). The correlations and the mean-squared errors between the returns estimated at 30 September 1997 and the true observed returns at 1 October 1997 are calculated and kept on a vector.

The second estimation window uses a data set that has the same length as the data used in the first estimation window, where the returns at 1 January 1993 are dropped and those at 1 October 1997 become part of the in-sample period that will be used to estimate returns at 2 October 1997. That is, the second estimation is done at 1 October 1997 to estimate returns at 2 October 1997. Again the correlations and mean-squared errors between the predicted statistics and the observed returns are calculated, where the previous vectors (of each method in the first estimation window) are appended with the latter vectors (one vector for each transformation method).

This process is repeated, where all the 75 returns on the out-of-sample period, are estimated. The last epoch 1 estimation is done on 30 December 1997 to estimate returns at 31 December 1997. At the end, there will be three 1×75 vectors of correlations and three 75×129 vectors of errors in epoch 1. The area minimisation covariance estimator is excluded in epoch 1, but included in all other epochs because it results in a very ill-conditioned covariance matrix. The averages of the results for each epoch, are given in table 7.2 for the correlations and table 7.3 for the squares of the mean-squared errors.

In the minimum risk portfolio (GMVP) for example, the first portfolio is constructed on 30 September 1997 with the resulting portfolio weights used to calculate the return adjusted weights. Then the ante-risk and the ex-post risk are calculated and their absolute difference taken. The short interest of the portfolio weights (GMVP) is also calculated. The whole process is repeated

	epoch 1	**epoch 2**	**epoch 3**	**epoch 4**	**epoch 5**	**epoch 6**
eigenvalue	72.20	13.57	6.58	174.68	20.18	7.13
arctan	16.70	4.87	3.87	113.94	7.69	9.84
tanh	32.59	13.72	4.30	127.11	9.52	6.19
area min	N/A	496180	2086	592.99	16.38	300.65
best	**arctan**	**arctan**	**arctan**	**arctan**	**arctan**	**tanh**

Table 7.3: *Shows the average of the squares of the mean squared errors between the predicted returns and the true observed returns*

on 1 October 1997 - a GMVP is constructed, from which the return adjusted weights are calculated, then the difference of the ante and the ex-post risks and the short interest are calculated.

This process continues until 75 portfolios are constructed for each of the four sample covariance estimators. That is, the last epoch 1 portfolio is constructed on 30 December 1997. At the end, there will be four 1×75 vectors of the absolute difference of the ex-post and the ante risks, and four 1×75 vectors of short interest of the GMVP. A summary of the results is given in table 7.4 for the absolute difference of the risks and 7.5 for the short interest.

7.3 An analysis of the results

The different sample covariance matrix estimators are not compared directly in this chapter, as done in the simulation part (section 3.4), because only the imperfect time series data is available. Since the sample covariance estimators are the foundations for building the model, their comparison is done on the returns prediction and portfolio construction. The correlations between the leading factors, called temporal stability [36], constructed from the different sample covariance estimators is also used as a comparison method [36] [27].

The area minimisation covariance estimator had to be dropped out of the contest in epoch 1 because it gave many error messages. It would seem that the epoch 1 data resulted in a very ill-conditioned non-positive definite sample covariance matrix such that the area minimisation technique outputs a positive definite matrix which is not symmetric (not a covariance matrix), with some negative diagonal elements (negative variances). In some iterations the area minimisation estimator just fails to output a positive definite matrix.

The sample covariance matrix estimators constructed from the eigenvalue and the two shrinkage estimators were then used in building the model in epoch 1. This suggests that these three covariance estimators are more robust than the area minimisation covariance estimator in terms of restoring positive definiteness to a non-positive definite symmetric matrix. For all other epochs, the area minimisation transformation method was included in the model building and analysis.

	epoch 1	**epoch 2**	**epoch 3**	**epoch 4**	**epoch 5**	**epoch 6**
eigenvalue	0.4897	0.4415	0.4366	0.4825	0.3251	0.3950
arctan	0.4910	0.4323	0.4276	0.4806	0.3219	0.3954
tanh	0.4879	0.4396	0.4347	0.4809	0.3237	0.3932
area min	N/A	0.5357	0.5555	1.7190	1.2186	0.8384
results	**tanh**	**arctan**	**arctan**	**arctan**	**arctan**	**tanh**

Table 7.4: *Shows the squares of the absolute difference between the ante and the post risks of the different covariance estimators across the epochs*

7.3.1 Expected asset returns prediction

The results of the expected asset returns prediction are given in table 7.2 (correlations) and 7.3 (mean-squared errors). The correlations look sporadic across epochs, where none of the epochs and none of the estimators had a consistent performance. The arctan shrinkage estimator had the highest correlations in both epochs 1 and 2, while the area minimisation covariance estimator had the highest correlations in all the remaining epochs. These correlations are very low at between 5% and 8%.

At least some pattern is present across epochs in the mean-squared errors results. In epochs 1, 2 and 3, all the mean-squared errors calculated from the different covariance estimators are decreasing, implying that the estimation is improving from epoch 1 to epoch 3. Then they become extremely high (reaching their highest errors) in epoch 4 but then decrease again, with the exception of the area minimisation method, from epochs 4, 5 and 6. That is, the estimation becomes bad between epochs 3 and 4 and then improves from epochs 4, 5 and 6. The arctan covariance estimator had the lowest mean-squared errors in 5 out of the 6 epochs. The returns predictions was good in epoch 3 because all the estimators, except the area minimisation method, reached their minimum in this epoch. The area minimisation covariance estimator resulted in relatively abnormally high errors in all the epochs.

7.3.2 Global minimum variance optimal portfolio

The results of the optimal portfolio construction are given in tables 7.4 (squares of the absolute difference between the ante and the post risks) and 7.5 (short interest). The absolute difference between the ante and the post risks squared follows a pattern similar to the one followed by the mean-squared errors except that the estimators reach their minimum risks in epoch 6 (with the exception of the area minimisation). The arctan shrinkage covariance estimator gave the lowest risk in four of the six epochs, with the other two epochs taken by the tanh shrinkage covariance estimator. The area minimisation estimator gave the highest risks in all the five epochs in which it participated.

The portfolio weights of the eigenvalue, the arctan and the tanh all had no short sales, that is, they were all non-negative. The area minimisation estimator resulted in extremely high short selling amounts which are given in table 7.5.

epochs	short interest
1	N/A
2	641.4997
3	1.6049
4	25.1656
5	3.6589
6	7.0179

Table 7.5: *Shows the short interest of the area minimisation covariance estimators across all epochs. The other estimators had no short sales.*

	epoch 1	epoch 2	epoch 3	epoch 4	epoch 5	epoch 6
epoch 1	1.0000	0.3598	-0.3138	-0.0542	-0.0172	-0.1220
epoch 2	0.3598	1.0000	-0.6009	-0.2299	0.1800	-0.0272
epoch 3	-0.3138	-0.6009	1.0000	0.4899	-0.2596	0.1167
epoch 4	-0.0542	-0.2299	0.4899	1.0000	-0.7251	-0.2944
epoch 5	-0.0172	0.1800	-0.2596	-0.7251	1.0000	0.6106
epoch 6	-0.1220	-0.0272	0.1167	-0.2944	0.6106	1.0000

Table 7.6: *Shows correlations amongst the leading factors calculated across all epochs using the eigenvalue covariance estimator*

	epoch 1	epoch 2	epoch 3	epoch 4	epoch 5	epoch 6
epoch 1	1.0000	-0.4120	0.3284	0.0764	0.0032	0.1278
epoch 2	-0.4120	1.0000	-0.5955	-0.2395	0.1775	-0.0476
epoch 3	0.3284	-0.5955	1.0000	0.4750	-0.2226	0.1560
epoch 4	0.0764	-0.2395	0.4750	1.0000	-0.7230	-0.2850
epoch 5	0.0032	0.1775	-0.2226	-0.7230	1.0000	0.6031
epoch 6	0.1278	-0.0476	0.1560	-0.2850	0.6031	1.0000

Table 7.7: *Shows correlations amongst the leading factors calculated across all epochs using the arctan covariance estimator*

7.3.3 Stability of the leading eigenvectors

The leading factor of the statistical common factors constructed in each of the different epochs using the eigenvalue, arctan shrinkage and tanh shrinkage covariance estimators were tested for stability over time. The leading factor represents genuine correlations because it deviates significantly from the random matrix theory predictions and genuine correlations should remain stable in time [27]. The cross-correlations among the leading factors constructed from each of the three estimators across epochs were calculated[2].

The correlations across all epochs amongst the leading factors constructed from the eigenvalue, the arctan and the tanh covariance estimators are given in tables 7.6, 7.7 and 7.8 respectively. The area minimisation estimator behaved badly as a result it was omitted.

[2]Based on the papers by Wilcox et al [36] and Plerou et al [27]

	epoch 1	epoch 2	epoch 3	epoch 4	epoch 5	epoch 6
epoch 1	1.0000	0.3597	-0.3136	-0.0541	-0.0172	-0.1220
epoch 2	0.3597	1.0000	-0.6008	-0.2299	0.1801	-0.0273
epoch 3	-0.3136	-0.6008	1.0000	0.4898	-0.2596	0.1167
epoch 4	-0.0541	-0.2299	0.4898	1.0000	-0.7251	-0.2944
epoch 5	-0.0172	0.1801	-0.2596	-0.7251	1.0000	0.6105
epoch 6	-0.1220	-0.0273	0.1167	-0.2944	0.6105	1.0000

Table 7.8: *Shows correlations amongst the leading factors calculated across all epochs using the tanh covariance estimator*

In looking at temporal stability of the eigenvectors, Wilcox and Gebbie [36] use the 15 leading eigenvectors, while Plerou et al [27] use all the eigenvectors that deviate from the random matrix theory predictions. All the 442 stocks are used in each epoch in this work, where a value of zero is assigned to stocks that have missing values in a particular epoch. Wilcox and Gebbie used the filtered assets and only expand one eigenvector to be the same size as the longest eigenvector in that particular epoch's correlation calculation.

The results show that the highest correlations are at lag 1. That is, epoch 1 is highly correlated to epoch 2 than epochs 3 to 6, epoch 2 is highly correlated to epoch 3 or epoch 1 than epochs 4 to 6, etc. The eigenvalue, the arctan and the tanh covariance estimators gave very similar results. The highly stable genuine correlations were in time periods between 1996-2000 (epoch 4) and 1997-2001 (epoch 5), with a correlation of around 72% by all the three estimators, followed by 1997-2001 (epoch 5) and 1998- 30 June 2003 (epoch 6) with a correlation of around 61%.

7.4 Conclusion

The purpose of this project is to choose the best sample covariance estimator, with the aim of using it in the analysis of empirical data from emerging markets, taking the JSE as an example. Four transformation methods namely, the eigenvector, the arctan shrinkage, the tanh shrinkage and the area minimisation methods, were used in the covariance estimation technique, resulting in four sample covariance estimators. These estimated sample covariance matrices were used in the asset returns prediction and in constructing a global minimum variance optimal portfolio, where different comparison techniques were used to compare them.

The results of the simulation part and those of the empirical part are in agreement. The arctan transformation method is the best transformation method to use in restoring the positive definiteness property of the sample covariance matrix when analysing empirical data from an emerging market, followed by the eigenvalue transformation method.

The statistical linear factor model didn't do well in estimating the expected asset returns in the JSE empirical data. For other alternative models to use in

the empirical data from the JSE see [2] and [33].

7.4.1 Contribution to future research

Since there is a rich literature on converting non-positive definite matrices into positive definite matrices, one examiner felt that a comparison between the transformations used in this text and the norm minimisation (which is the standard method used) would have added some value and referred me to the work of Higham (2002) [17]. The other examiner questioned the realistic nature of the way the thinly-traded environment was emulated and referred me to an unpublished PhD thesis of Bowie (1994) [6].

Appendix A

Other covariance estimation techniques

In this project covariance estimation includes restoring the positive definiteness property to a non-positive definite sample covariance matrix calculated from a data set with missing data points by the method of dealing with missing data called pairwise deletion[1]. The statistical finance literature is rich with different kinds of covariance estimation techniques, all of which can be broadly grouped into one-factor, k-factor and $N-$factor models. The estimators in this project are examples of $N-$factor models. Brief descriptions of an example of a one-factor model and an example of a $k-$factor model are given below.

A.1 The single-index covariance estimator

Sharpe's single index model, an example of a one-factor model, estimates stock returns using the market as a single factor, where all stocks are regressed against the market. For a single stock, the factor model at time t is given by:

$$\mathbf{r}_{it} = \alpha_i + \beta_i \mathbf{r}_{Mt} + \epsilon_{it}$$

with the following assumptions:

$$E\left(\epsilon_{it}\right) = 0$$

$$E\left(\epsilon_{it}\epsilon_{is}\right) = \begin{cases} 0 \text{ for } t \neq s \\ \sigma^2 \text{ for } t = s \end{cases}$$

$$E(\epsilon_{it} r_{Mt}) = 0$$

where r_{Mt} is the market index, β_i is the measure of the volatility of the security relative to the market and ϵ_{it} are the residual terms. This model gives the covariance as:

[1] this section ia based on [3]

$$\begin{aligned}
\sigma_{ij} &= E\left((r_{it} - E(r_{it}))(r_{jt} - E(r_{jt}))\right) \\
&= E\left(r_{it} r_{jt}\right) - E\left(r_{it}\right) E\left(r_{jt}\right) \\
&= \beta_{it}\beta_{jt} E\left(r_{Mt}^2\right) + E\left(\epsilon_i \epsilon_j\right) \\
&= \beta_{it}\beta_{jt}\sigma_M^2 + \sigma_\epsilon^2 \lambda_{ij} \iff E\left(r_{Mt}^2\right) = \sigma_M^2 \text{ and } E\left(\epsilon_i \epsilon_j\right) = \sigma_\epsilon^2 \lambda_{ij}
\end{aligned}$$

In matrix notation, the last equation can be written as:

$$\Sigma = \overrightarrow{\beta}\,\overrightarrow{\beta}^T \sigma_M^2 + \Sigma_\epsilon$$

where Σ is the true population covariance of asset returns, σ_M^2 is the variance of the market return, $\overrightarrow{\beta}$ is the vector of slopes, and Σ_ϵ is the diagonal matrix containing residual variances. This model can be estimated by running a regression on each asset's return against the market. If b_i represents the slope estimate (estimates $\overrightarrow{\beta}$), s_M^2 represents an estimate of σ_M^2 and d_{ii} represents the residual variance estimate (estimates Σ_ϵ), then the single-index model estimator for the covariance matrix of asset returns is:

$$\mathbf{F} = \mathbf{b}\mathbf{b}^T s_M^2 + \hat{\Sigma}_\epsilon$$

A.2 The shrinkage estimator

Many estimates suffer from at least two sources of errors, the estimation error and the specification error[2]. At one extreme, the single-index covariance matrix comes from a one-factor model with all the errors attributed to specification, while at the other extreme, the sample covariance matrix can be interpreted as an N-factor model with only estimation error (depending on the sample size). The shrinkage estimator is a compromise between these two extremes - it suggests a k-factor model where $1 < k < N$. In one extreme there's an N-factor model with an unbiased estimator full of estimation error and in the other extreme there's a one-factor model with a bias estimator that is full of specification error.

"A fundamental principle of statistical decision theory is that there exists an interior optimum in the trade-off between bias and estimation error. Since Stein's (1956) seminal work, we know that one way of attaining this optimal trade-off is simply to take a properly weighted average of the biased and unbiased estimators. This is called *shrinking* the unbiased estimator full of estimation error towards a fixed target represented by the biased estimator" [3].

The shrinkage estimator for the covariance matrix of stock returns, as developed by Ledoit et al, is:

$$\hat{\mathbf{S}} = \alpha\mathbf{F} + (1 - \alpha)\mathbf{S}$$

[2] this section ia based on [3]

where α is called the shrinkage intensity.

Ledoit et al derived a mathematical formula for α by defining a loss function which if they optimise would result in a formula for the shrinkage intensity α. A short coming of this model was the number of factors k to use. Bengtsson and Holst improved on this shrinkage method by mathematically solving for k using the work of Plerou et al of random matrix theory.

Appendix B

Methods of dealing with missing data

There are different methods of dealing with missing data, with each method having its own advantages and disadvantages. Examples of the methods are [12]:

- listwise or casewise deletion
- pairwise deletion
- mean substitution
- imputation by regression
- hot deck imputation
- expectation maximisation algorithm
- raw maximum likelihood or full information maximum likelihood method
- multiple imputations

Appendix C

Separating the optimal portfolio weights

The Mutual Fund Separation Theorem is given by the following equation:

$$\begin{aligned}\mathbf{w} &= \left(1 - \frac{\mathbf{1}^T\Omega^{-1}\mu}{\gamma}\right)\frac{\Omega^{-1}.1}{\mathbf{1}^T\Omega^{-1}\mathbf{1}} + \left(\frac{\mathbf{1}^T\Omega^{-1}\mu}{\gamma}\right).\frac{\Omega^{-1}\mu}{\mathbf{1}^T\Omega^{-1}\mu} \\ &= \frac{\Omega^{-1}\mathbf{1}}{\mathbf{1}^T\Omega^{-1}\mathbf{1}} + \frac{\Omega^{-1}}{\gamma}\left(E(R) - \mathbf{1}.\frac{\mathbf{1}^T\Omega^{-1}E(R)}{\mathbf{1}^T\Omega^{-1}\mathbf{1}}\right)\end{aligned}$$

let $\bar{R}$ be a vector of equilibrium returns. Then

$$\begin{aligned}\mathbf{w} &= \frac{\Omega^{-1}\mathbf{1}}{\mathbf{1}^T\Omega^{-1}\mathbf{1}} + \frac{\Omega^{-1}}{\gamma}\left(E(R) - \bar{R} + \bar{R} - \mathbf{1}.\frac{\mathbf{1}^T\Omega^{-1}(E(R) - \bar{R} + \bar{R})}{\mathbf{1}^T\Omega^{-1}\mathbf{1}}\right) \\ &= \frac{\Omega^{-1}\mathbf{1}}{\mathbf{1}^T\Omega^{-1}\mathbf{1}} + \frac{\Omega^{-1}}{\gamma}\left(\bar{R} - \mathbf{1}\frac{\mathbf{1}^T\Omega^{-1}\bar{R}}{\mathbf{1}^T\Omega^{-1}\mathbf{1}}\right) + \frac{\Omega^{-1}}{\gamma}\left(E(R) - \bar{R} - 1\frac{\mathbf{1}^T\Omega^{-1}(E(R) - \bar{R})}{\mathbf{1}^T\Omega^{-1}\mathbf{1}}\right)\end{aligned}$$

Using the identity of $\mathbf{1}^T ZG = G^T Z\mathbf{1}$, where Z is any $n \times n$ symmetric matrix and G is an $n \times 1$ vector, it can be shown that:

$$\begin{aligned}\mathbf{w} &= \frac{\Omega^{-1}\mathbf{1}}{\mathbf{1}^T\Omega^{-1}\mathbf{1}} + \frac{1}{\gamma}\frac{\Omega^{-1}\mathbf{1}}{\mathbf{1}^T\Omega^{-1}\mathbf{1}}.\left(\bar{R}\mathbf{1}^T\Omega^{-1}\mathbf{1} - \mathbf{1}\bar{R}\Omega^{-1}\mathbf{1}\right)\frac{1}{\gamma}\frac{\Omega^{-1}\mathbf{1}}{\mathbf{1}^T\Omega^{-1}\mathbf{1}}\ldots\ldots\ldots\ldots\ldots\ldots \\ &\ldots\ldots\ldots\ldots \left((E(R) - \bar{R})\mathbf{1}^T\Omega^{-1}\mathbf{1} - \mathbf{1}(E(R) - \bar{R})^T\Omega^{-1}\mathbf{1}\right) \\ &= \frac{\Omega^{-1}\mathbf{1}}{\mathbf{1}^T\Omega^{-1}\mathbf{1}} + \frac{1}{\gamma}\frac{\Omega^{-1}\left(\bar{R}\mathbf{1}^T - \mathbf{1}\bar{R}^T\right)\Omega^{-1}}{\mathbf{1}^T\Omega^{-1}\mathbf{1}}\mathbf{1} + \frac{1}{\gamma}\frac{\left((E(R) - \bar{R})\mathbf{1}^T - \mathbf{1}(E(R) - \bar{R})^T\right)\Omega^{-1}}{\mathbf{1}^T\Omega^{-1}\mathbf{1}}\mathbf{1} \\ &= \omega_g + \omega_S + \omega_T\end{aligned}$$

where $\omega_g = \frac{\Omega^{-1}\mathbf{1}}{\mathbf{1}^T\Omega^{-1}\mathbf{1}}$, $\omega_S = \frac{1}{\gamma}\frac{\Omega^{-1}\left(\bar{R}\mathbf{1}^T - \mathbf{1}\bar{R}^T\right)\Omega^{-1}}{\mathbf{1}^T\Omega^{-1}\mathbf{1}}\mathbf{1}$ and $\omega_T = \frac{1}{\gamma}\frac{\left((E(R) - \bar{R})\mathbf{1}^T - \mathbf{1}(E(R) - \bar{R})^T\right)\Omega^{-1}}{\mathbf{1}^T\Omega^{-1}\mathbf{1}}\mathbf{1}$

Bibliography

[1] Amihud Y., *Illiquidity and stock returns: Cross-section and timeseries effects.* Journal of Financial Markets, Vol. 5 No. 1 (2002) 31 - 56.

[2] Barr G.D.I and Bradfield D.J., *Portfolio Selection in Thinly Traded Environments - a case study*, Managerial and Decision Economics Vol. 9, (1998) pp. 287-290

[3] Bengtsson C. and Holst J., *On portfolio selection: Improved covariance matrix estimation for Swedish asset returns*, Technical report, Department of Economics, Lund University, Sweden, 2002

[4] Black F. *The capital market equilibrium with restricted borrowing.* Journal of Business, Vol. 45 (1972) 444 - 455.

[5] Black F. and Litterman R., *Global Portfolio Optimization*, Financial Analysts Journal, September-October, (1992) pp. 28-43.

[6] Bowie D.C., *Thin trading, non-normality and the estimation of systematic risk on small stock markets.* Unpublished Ph.D. thesis, University of Cape Town (1994).

[7] Burns P., *On Using Statistical Factor Models in Optimizing Long-Only Portfolios*, Working Papers (2003): http://www.burns-stat.com/

[8] Chen Y. and McInroy J.E., *Estimation of symmetric positive-definite matrices from imperfect measurements*, IEEE Transactions on Automatic Control, Vol. 47, No. 10 (2002), pp. 1721-1725 (ISSN: 0018-9286)

[9] Connor G., *The Three Types of Factor Models: A Comparison of Their Explanatory Power*, Financial Analysts Journal, Vol. 51, No. 3 (1995) pp. 42-46

[10] Daniel K. and Titman S., *Evidence on The Characteristics of Cross Sectional Variation in Stock Returns*, The Journal of Finance, Vol. LII, No. 1 (1997), pp. 1-33

[11] Davis J.L., Fama E.F. and French K.R., *Characteristics, Covariances, and Average Returns: 1929-1997*, Journal of Finance, Vol. 55 (2000), 389-406.

[12] Different methods of dealing with missing data, http://www.secondmoment.org/etal-column/ea-missingdata.php

[13] Fama E.F. and French K.R., *The Cross-section of Expected Stock Returns*, The Journal of Finance, Vol. XLVII, No. 2 (1992), pp. 427-465

[14] Fama E.F. and French K.R. *Common Risk Factors in the Returns on Stock and Bonds*, Journal of Financial Economics, Vol. 33 (1993), pp. 3-56

[15] Haugen R.A. and Baker N.L., *Commonality in the determinants of expected stock returns*, Journal of Financial Economics, Vol. 41 (1996), pp. 401-439

[16] Hawanini and Keim, *The Cross-section of Common Stock returns: A Review of the Evidence and Some New Findings, Security market imperfections in worldwide equity markets*, Cambridge University Press, New York, (2000) 3-43.

[17] Higham N.J., *Computing the Nearest Correlation Matrix: A Problem from Finance.*, IMA Journal of Numerical Analysis, Vol. 22 No. 3. (2002), pp. 329-343. ISSN 0272-4979

[18] Knez P.J. and Ready M.J., *On the Robustness of Size and Book-to-market in Cross-sectional Regressions*, Journal of Finance, Vol. 52 (1997), 1355-1382.

[19] Laloux L., Cizeau P., Bouchaud J. and Potters M., *Noise Dressing of Financial Correlation Matrices*, Phys. Rev. Lett., Vol. 83 (1999), pp. 1467-1470

[20] Ledoit O. and Wolf M., *Improved Estimation of the Covariance Matrix of Stock Returns With an application to Portfolio Selection*, Journal of Empirical Finance Vol. 10, (2003) pp. 603621

[21] Lindskog F., *Linear Correlation Estimation, 2000*: http://www.risklab.ch/Papers.html#LCELindskog

[22] Lintner J. *The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets.* Review of Economics and Statistics, Vol. 47 (1965) 13 - 37.

[23] Markowitz, H. *Portfolio selection.* Journal of Finance, Vol. 7, No. 1 (1952) 77 - 91

[24] Markowitz H.M., *Portfolio Selection: Efficient Diversification of Investments.* New York: John Wiley and Sons, (1959)

[25] Monch E. *Explaining the Cross-section of Returns with Macro-economic Factors,* Working Paper: Humboldt University of Berlin - School of Business and Economics; University of Pennsylvania (2004)

[26] Pafka S., Potters M. and Kondor I. *Exponential Weighting and Random-Matrix-Theory-Based Filtering of Financial Covariance Matrices for Portfolio Optimization*, Science and Finance, Capital Fund Management, eprint arXiv:cond-mat/0402573 Vol. 1 (2004)

[27] Rosenow B., Plerou V., Gopikrishnan P., Amaral L.A.N., Guhr T. and Stanley H.E. *Random Matrix Approach to Cross Correlations in Financial Data* Physical Review E, Issue 6, Vol. 65 (2002), 066126

[28] Rosenow B., Plerou V., Gopikrishnan P., and Stanley H.E. *Portfolio Optimization and the Random Magnet Problem*, Eur. Phys. Lett. 59, 500 (2002) eprint cond-mat/0111537

[29] Ross S. A., *The Arbitrage Theory of Capital Asset Pricing* Journal of Economic Theory, Vol. 13 (1976) 341360.

[30] Sharpe W.F. *Capital asset prices: a theory of market equilibrium under conditions of risk.* Journal of Finance, Vol. 19, No. 3 (1964) 425 - 442.

[31] Sharpe W.F., *Asset Allocation: Management Style and Performance Measurement*, Journal of Portfolio Management, Vol. 18 (1992), pp.7-19.

[32] Troskie C.G., *Financial Mathematics: Multivariate Analysis Lecture Notes* (2002)

[33] van Rensburg P. and Robertson M., *Explaining the Cross-section of Returns in South Africa: Attributes or Factor Loadings?*, Journal of Asset Management, Vol. 4, No. 5 (2004) , pp. 334-347

[34] Wai L., *Theory and Methodology of Tactical Asset Allocation, Wiley, New York (2000)*

[35] West G. (Graeme West Financial Modeling Agency), *Principal Component Analysis*: Financial Mathematics Lecture Notes (2006)

[36] Wilcox D. and Gebbie T., *An analysis of Cross-correlations in an emerging market*, Physica A, Vol. 375, No. 2, (2007), pp. 584-598

[37] Zivot E., *Financial Econometrics Lecture Notes (2002)*