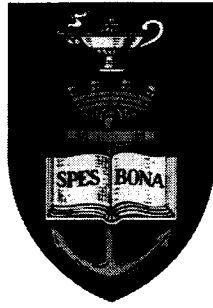


The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.



Multivariate Volatility Modelling in Modern Finance

Martin B. Bongers

Submitted to the Department of Statistical Sciences
in fulfillment of the requirements for the degree of

Master of Science in Mathematical Statistics
at the
UNIVERSITY OF CAPE TOWN

May 23, 2008

The author hereby grants to the University of Cape Town permission to reproduce and to distribute copies of this thesis document in whole or in part

Supervisor: Professor Linda Haines

Abstract

It is common in the financial industry to measure the risk of a stock with univariate volatility models and to use the theory of the Capital Asset Pricing Model (CAPM) to generate the associated covariance matrix between the different stocks. The CAPM model however does not appear to capture the true covariance between the different stock returns. By relaxing one of the assumptions of CAPM, multivariate volatility models can be used to more appropriately model the covariance between the stock returns. Multivariate volatility models are traditionally, however, parameter-heavy and fail in practically modelling financial time-series. In this thesis a class of multivariate volatility models are discussed which use matrix decomposition methods on the covariance matrix of a portfolio of stocks to create more parsimonious and practically applicable models. The aim of the study is to ascertain whether the information gained from the more complicated multivariate matrix decomposition models can be used to better forecast the covariance matrix and produce a Value at Risk estimate which more appropriately describes fat-tailed financial time-series.

I hereby grant the University of Cape Town permission to copy and disseminate this work, or any part thereof, for the purposes of study and research.

Plagiarism Declaration

1. This dissertation is my own work. It has not been submitted before for any degree or examination to any other University.
2. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.
3. Each significant contribution to, and quotation in, this dissertation from the work of other people has been cited and referenced.

Signature

Date

Acknowledgements

I would like to thank my supervisor, Professor Linda Haines, for her effort and guidance throughout this Masters thesis. Her patience, wisdom and experience helped formulate many of the ideas throughout this work.

I also thank my parents for investing in my studies both financially and emotionally. Their support in my academic career is boundless. Thank you for always believing in my abilities.

Finally, I thank my fiancé, Candyce Hall, for all the help with research, restructuring, spelling and grammar. For her sacrifice and effort I am eternally grateful.

University of Cape Town

Contents

List of Figures	vii
List of Tables	viii
1 Introduction	1
2 Modern Financial Theory	3
2.1 The efficient frontier	5
2.1.1 Quadratic programming	5
2.1.2 Lagrangian multiplier technique	6
2.2 Tangency portfolio and the risk-free asset	7
2.3 The effect of correlation between stocks	8
2.4 The Capital Asset Pricing Model	9
2.5 The Securities Market Line	11
2.6 Index models	12
2.6.1 The Single Index Model	12
2.6.2 Applications of the Single Index Model	13
2.6.3 Factor Models	14
2.7 Diversification	15
3 Objectives and Aims of the Thesis	17
3.1 The Single Index as opposed to the Factor Model	17
3.2 Univariate modelling of expected return and variance	18
3.3 Multivariate volatility models	18
3.4 Parsimony	19
3.5 Aim of the thesis	20
4 Univariate Volatility Model Theory	21
4.1 Introduction	21
4.2 Univariate GARCH Model	21
4.2.1 Simultaneous and two-tier parameter estimation	22
4.2.2 Unconditional variance for a GARCH(m,s) model	23
4.2.3 Estimation for a GARCH(m,s) model	24

4.2.4	Forecasting	25
4.3	Integrated GARCH Model	27
4.4	Exponential GARCH Model	28
5	Application of Univariate Volatility Models	30
5.1	Selection of time series length	30
5.2	Portfolio selection	31
5.3	Model fitting	33
5.4	Volatility model specification	35
5.4.1	GARCH	35
5.4.2	IGARCH	37
5.4.3	EGARCH	38
5.5	Creating mean and covariance forecasts	41
5.5.1	Conditional mean forecasts	41
5.5.2	Conditional variance forecasts	41
5.5.3	Efficient frontier analysis	43
5.5.4	Sum of squares forecasting error	43
6	Multivariate Volatility Model Theory	46
6.1	The BEW and BEKK model generalizations	47
6.2	Constant Conditional Correlation	48
6.3	Time-varying correlation models	49
7	Matrix Decomposition Methods in Multivariate Volatility Models	51
7.1	Principal component analysis	51
7.2	Orthogonal-GARCH model	52
7.2.1	Basic O-GARCH volatility model	53
7.2.2	Alexander O-GARCH volatility model	54
7.2.3	Modified O-GARCH volatility model	56
7.2.4	Parameter simulation on the O-GARCH formulations	58
7.3	Cholesky decomposition theory	59
7.4	Cholesky decomposition and volatility modelling	60
7.4.1	Basic Cholesky decomposition model	60
7.4.2	Tsay Cholesky decomposition volatility models	61
7.4.3	Constant L models	62
7.4.4	The problem of order	65
8	Constant Conditional Correlation Simulations	66
8.1	Simulation algorithm	66
8.2	Methodology	67
8.3	Simulation results	68
8.3.1	Parameter estimates	68

8.3.2	Simulated Forecast Error (Total SFE)	69
8.4	Remarks and conclusions	72
9	Application of Matrix Decomposition Methods in Multivariate Volatility Models	74
9.1	Introduction	74
9.2	Volatility model specification for O-GARCH models	75
9.3	Model selection for O-GARCH	77
9.3.1	Efficient frontier with dynamic covariance estimates	77
9.3.2	Sum of square variance forecasting error	77
9.3.3	The Frobenius norm error	81
9.4	Modified Cholesky decomposition volatility model	83
9.4.1	Volatility model specification for the Modified Cholesky decomposition	83
9.4.2	Effect of various stock orderings on Portfolio 2	84
9.4.3	Order determination for the covariance matrix	85
9.4.4	Model selection for the Modified Cholesky decomposition volatility model	86
10	Value at Risk	90
10.1	Backtesting	91
11	Conclusions	96

List of Figures

2.1	Tangency Portfolio and the Capital Allocation Line	8
2.2	Effect of correlation on the efficient frontier	9
5.1	Simulation experiment for GARCH(1,1) model	31
5.2	Forecast error per stock for univariate models: Portfolio 1	44
5.3	Forecast error per stock for univariate models: Portfolio 2	45
7.1	Parameter square error comparison for O-GARCH formulations using simulation	58
9.1	Efficient frontiers using O-GARCH and univariate EGARCH and GARCH volatility models: Portfolio 1	79
9.2	Efficient frontiers using O-GARCH and univariate volatility mod- els: Portfolio 2	80
9.3	Forecast error per stock using Alexander O-GARCH, Modified O- GARCH and EGARCH volatility models: Portfolio 1	81
9.4	Forecast error per stock using Alexander O-GARCH and Modified O-GARCH volatility models: Portfolio 2	82
9.5	Forecast error per stock using different volatility models: Portfolio 1	89
9.6	Forecast error per stock using different volatility models: Portfolio 2	89

List of Tables

5.1	Correlation between the log-returns of the five stocks in Portfolio 1 (df=366)	32
5.2	Correlation between the log returns of five stocks in Portfolio 2 (df=366)	33
5.3	Mean equations for Portfolio 1 using GARCH specification for residuals (standard errors in brackets)	35
5.4	Beta coefficients for Portfolio 1 stocks	36
5.5	Mean equations for Portfolio 2 using GARCH specification for residuals (standard errors in brackets)	36
5.6	Beta coefficients for the stocks of Portfolio 2	36
5.7	GARCH parameter estimates (standard errors in brackets)	37
5.8	GARCH standard deviation forecasts	38
5.9	IGARCH parameter estimates for volatility equations	39
5.10	IGARCH standard deviation forecasts	39
5.11	Mean equations for Portfolio 1 using EGARCH specification for residuals (standard errors in brackets)	40
5.12	EGARCH parameter estimates (standard errors in brackets), $\gamma \neq 0$	40
5.13	EGARCH standard deviation forecasts (*GARCH model)	41
5.14	Expected returns using mean equations for univariate models (*Not fitted due to insignificant leverage effect)	42
5.15	ALSI market index parameter estimates for univariate models (standard errors in brackets)	42
5.16	Sum of forecast square errors for univariate models	44
8.1	Average parameter estimates over 1000 three-stock CCC(1,1) simulated series: Weak correlation (between-simulation parameter standard deviation in brackets)	70
8.2	Average parameter estimates over 1000 three-stock CCC(1,1) simulated series: Strong correlation (between-simulation parameter standard deviation in brackets)	71
8.3	Total SFE for simulated CCC(1,1) sets	73
9.1	Alexander O-GARCH and Modified O-GARCH parameter estimates for Portfolio 1 (standard errors in brackets)	78

9.2	Alexander O-GARCH and Modified O-GARCH parameter estimates for Portfolio 2 (standard errors in brackets)	78
9.3	Alexander O-GARCH and Modified O-GARCH standard deviation forecasts of unique risk	79
9.4	Sum of forecast square errors for different volatility models	81
9.5	Sum of Frobenius norms for the difference between actual and forecast covariance matrices	83
9.6	Modified Cholesky decomposition parameter set for coefficient of determination and maximum variance orderings: Portfolio 1 (standard errors in brackets)	87
9.7	Modified Cholesky decomposition parameter set for coefficient of determination and maximum variance orderings: Portfolio 2 (standard errors in brackets)	87
9.8	Modified Cholesky decomposition standard deviation forecast of unique risk for different input orderings	88
9.9	Sum of forecast square error of variance for different volatility models	89
10.1	Value at Risk estimated as 5% quantile for one-week loss in Rands: Portfolio 1	91
10.2	Value at Risk estimated as 1% quantile for one-week loss in Rands: Portfolio 1	92
10.3	Value at Risk estimated as 5% quantile for one-week loss in Rands: Portfolio 2	92
10.4	Value as Risk estimated as 1% quantile for one-week loss in Rands: Portfolio 2	92
10.5	Number of log-return exceedences per stock from the $VaR_{0.05}$ level: Portfolio 1	93
10.6	Number of log-return exceedences per stock from the $VaR_{0.01}$ level: Portfolio 1	94
10.7	Number of log-return exceedences from the $VaR_{0.05}$ level per stock: Portfolio 2	94
10.8	Number of log-return exceedences from the $VaR_{0.01}$ level per stock: Portfolio 2	94
10.9	Sum of squares of (observed exceedences-expected exceedences)/expected exceedences for different VaR levels	95

Chapter 1

Introduction

Stock portfolios are said to be functions of expected return and expected risk [1]. The concept of risk is an important factor in the modern economy, as scarce resources must be allocated as efficiently as possible to the best-perceived factors of production.

Volatility is the unobservable mathematical measure of investor uncertainty in realizable return on investment, the best available proxy for which is realized variance [2]. Traditionally, the volatility of an equity was modelled individually, assuming the only risk common to all assets is that of the market in which the equity is traded. However, the interconnectedness of the modern financial market makes the relationship between equities much more complex.

Tsay [3] lists the financial functions of volatility modelling as including roles in risk management such as calculation of value at risk of an investor position, asset allocation based on mean-variance portfolio optimization, and improvement of parameter estimation and interval forecast efficiency. All of these applications will be discussed in the body of this thesis.

As traditional univariate volatility models appear to neglect some form of covariance risk, multivariate volatility techniques have been developed to explore this extra source of covariance within portfolios of equities. Traditional multivariate volatility models try to parameterize the entire covariance matrix, rendering such approaches unparsimonious and impractical for fitting to large portfolios or for portfolio selection in real time. The simpler univariate models are therefore often favoured for reliable volatility estimation in real-life contexts.

A dynamic estimate of the covariance matrix of a portfolio of risky assets is integral to many areas of finance, with both past and present financial disasters emphasizing the importance of proper risk management. A deeper understanding of the dynamic relationship between assets, particularly in times of market up-

heaval, would be invaluable information for financial decision-makers. Creating time-varying covariance matrices substantially improves portfolio allocation accuracy, a forecast of future correlations and volatilities is the basis of any option pricing formula [4], financial strategies such as hedging require estimates of the correlation between asset returns, and the prices of any derivative based on a basket of assets sensitive to correlation between underlying returns.

The main objective of this thesis was to develop and implement multivariate volatility models on portfolios of real-world equities displaying different characteristics, to compare predictive performance, both among the multivariate models and with the more theoretically-established univariate volatility models. The research question is whether multivariate volatility models provide a more accurate model of covariance between stock returns than univariate, with comparisons based on volatility forecasting accuracy, Value at Risk estimation and computational practicality. Few papers exist which compare different multivariate GARCH models for the same problem and data sets [5],[6]. The multivariate volatility models are expected to fit better to smaller, highly correlated portfolios, but traditional multivariate volatility models tend to be less effective in low-correlated, large-dimension data sets.

Chapter 2 briefly summarizes the relevant modern financial theory in discussions of Markowitz portfolio selection, the Capital Asset Pricing Model (CAPM), and a dynamic Capital Asset Pricing Model required for implementation of the multivariate volatility models. Based on this background, Chapter 3 elaborates on the aims and objectives of this study. Chapter 4 reviews the theoretical development of univariate volatility estimation procedures followed, in Chapter 5, by explaining the selection and features of the chosen data set and an application of the univariate volatility models to this data. Chapter 6 gives a brief history of selected important multivariate volatility models proposed to create time-varying covariance estimates, while Chapter 7 focuses more specifically on a suite of multivariate volatility models based on matrix decomposition. Chapter 9 applies these matrix decomposition multivariate volatility models and compares covariance matrix forecasting ability among all multivariate and univariate models implemented. Chapter 8 bases the same comparison on computer-simulated volatility series. Chapter 10 discusses Value at Risk estimation for all models fitted in the thesis. Finally, some broad conclusions are drawn and further avenues of research are identified in Chapter 11.

Chapter 2

Modern Financial Theory

Stock market investors must allocate investment capital among companies in the market. With hundreds of equities listed on the Johannesburg Stock Exchange (JSE) alone, the procedures of stock selection and portfolio construction are neither trivial nor arbitrary. The established approach is to quantify the positive relationship between expected returns on assets, estimated using time series methods, simulation or scenario planning, and corresponding risk. The aim of portfolio theory is to optimally choose a trade off between risk and return.

Define the return R_t at time t of an asset as a random variable

$$R_t = \frac{\varrho_t - \varrho_{t-1}}{\varrho_{t-1}}$$

where ϱ_t is the price of the asset at time t . A log-return of an individual asset r_t at time t can now be defined as the random variable $r_t = \ln(R_t + 1)$ where \ln denotes the natural logarithm. Let the random vector \mathbf{r} of p sets of log-returns of p stocks be defined as

$$\mathbf{r}_t = \begin{pmatrix} r_{1t} \\ r_{2t} \\ \vdots \\ r_{pt} \end{pmatrix}$$

The realization of the random vector \mathbf{r}_t contains log-returns for a single time period t , but for the sake of notational convenience the time subscript t will be suppressed. Therefore, consider the vector of log-returns \mathbf{r} , with $E(\mathbf{r}) = \boldsymbol{\mu}$. It is anticipated that returns on stocks are correlated and the covariance between stock returns can be defined as

$$\Sigma = E(\mathbf{r} - \boldsymbol{\mu})(\mathbf{r} - \boldsymbol{\mu})' = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix}$$

where $\sigma_{ii} = \sigma_i^2$ is the variance of the i^{th} return and σ_{ij} is the covariance between the i^{th} and j^{th} returns for $i, j = 1, \dots, p$ at some unspecified time t .

A portfolio of stocks is an investment of a proportion w_i of investor capital in acquiring each of the p stocks for $i = 1, \dots, p$. This means that a fully invested portfolio has

$$\sum_{i=1}^p w_i = 1.$$

A sum of weights less than one implies that some of the portfolio is held as a cash equivalent instead of stock investments, while a sum of weights greater than one indicates the use of borrowed funds. Negative proportions of w_i imply selling short.

Define the vector \mathbf{w} as

$$\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{pmatrix}$$

and hence define the random variable representing a return of a portfolio at an unspecified time t as being

$$P = \mathbf{w}'\mathbf{r} = \sum_{i=1}^p w_i r_i.$$

The expected return of the portfolio is then

$$E(P) = \mathbf{w}'E(\mathbf{r}) = \mathbf{w}'\boldsymbol{\mu} = \mu_P$$

and the variance is

$$\text{var}(P) = \mathbf{w}'\Sigma\mathbf{w} = \sigma_P^2.$$

If the returns are taken to be multivariate normal, that is

$$\mathbf{r} \sim N(\boldsymbol{\mu}, \Sigma),$$

then return on the portfolio is the random variable $P \sim N(\mu_P, \sigma_P^2)$.

The above formulation shows that the value of the portfolio varies with the proportions of capital invested in each stock. The aim of optimal investment is to allocate capital proportionately to each stock so as to simultaneously maximize expected portfolio return μ_P and minimize associated risk defined here as σ_P^2 .

2.1 The efficient frontier

For the mathematical formulation of portfolio selection developed by Markowitz, the concept of efficiency must be defined [7]. A portfolio is called mean-variance efficient if, for a given amount of risk, the expected return is maximized, and, for a given amount of return, the expected risk is minimized. The efficient frontier comprises the set of all possible mean-variance efficient portfolios. The efficient frontier can therefore be found either by fixing the expected portfolio return μ_P and minimizing the portfolio variance σ_P^2 or by fixing σ_P^2 and maximizing μ_P . Most investors prefer to fix returns and then minimize variance. The specific efficient portfolio chosen by the investor is determined by risk preference.

Markowitz [7] formulated the portfolio problem as

$$\begin{aligned} \text{maximize } \mu_P &= \mathbf{w}'\boldsymbol{\mu} \\ &= \sum_{i=1}^p w_i \mu_i \\ \text{minimize } \sigma_P^2 &= \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} \\ &= \sum_{i=1}^p \sum_{j=1}^p w_i w_j \sigma_{ij} \\ \text{subject to} \\ \sum_{i=1}^p w_i &= 1. \end{aligned} \tag{2.1}$$

There is no unique solution to the problem formulated in (2.1). The feasible region of attainable portfolios is created by all vectors of weights \mathbf{w} , such that $\mathbf{w}'\mathbf{1} = 1$, where $\mathbf{1}$ represents a vector of ones, and $\mathbf{w} \geq 0$, which implies that short sales are disallowed.

Although there are many approaches to calculating efficient portfolios and drawing the efficient frontier, only the most widely used, namely the quadratic programming and Lagrangian multiplier methods, are discussed here.

2.1.1 Quadratic programming

The quadratic programming portfolio construction formulation requires (2.1) to be rewritten as

$$\begin{aligned} \text{minimize } \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} \\ \text{subject to } \mathbf{w}'\boldsymbol{\mu} &= \mu_p \end{aligned}$$

$$\begin{aligned}\mathbf{w}'\mathbf{1} &= 1 \\ w_i &\geq 0 \text{ for } i = 1, \dots, p\end{aligned}$$

where μ_P is a fixed portfolio return. The quadratic programming algorithm can be run for a grid of μ_P values to create the efficient frontier [8].

Quadratic programming allows for additional linear constraints, for example, lower and upper bounds on selected weights to prevent over- or under-investment in a stock. However, the inclusion of too many constraints diminishes the area of the feasible region and hence the set of efficient portfolios. Time-consuming computations are a disadvantage of quadratic programming.

2.1.2 Lagrangian multiplier technique

The Lagrangian multiplier method is used to derive an explicit expression for the efficient frontier by providing an analytic solution to the portfolio problem [9]. The portfolio optimization problem can be written in the form of a Lagrangian multiplier problem if only equality constraints are considered. Hence, $\mathbf{w}'\Sigma\mathbf{w}$ is minimized subject to $\mathbf{w}'\mathbf{1} = 1$ and $\mathbf{w}'\boldsymbol{\mu} = \mu_P$. The Lagrangian function

$$L(\mathbf{w}, \delta_1, \delta_2) = \mathbf{w}'\Sigma\mathbf{w} + \delta_1(\mu_P - \mathbf{w}'\boldsymbol{\mu}) + \delta_2(1 - \mathbf{w}'\mathbf{1})$$

is minimized with respect to the vector \mathbf{w} , where δ_1 and δ_2 are Lagrangian multipliers. The results of this approach can be found in Ruppert [10]. Specifically, the optimal vector of weights \mathbf{w}_{μ_P} for an expected return μ_P is

$$\mathbf{w}_{\mu_P} = \mathbf{g} + \mu_P \mathbf{h}$$

where

$$\begin{aligned}\mathbf{g} &= \frac{\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}\Sigma^{-1}\mathbf{1} - \mathbf{1}'\Sigma^{-1}\boldsymbol{\mu}\Sigma^{-1}\boldsymbol{\mu}}{\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}\mathbf{1}'\Sigma^{-1}\mathbf{1} - \boldsymbol{\mu}'\Sigma^{-1}\mathbf{1}\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu}} \\ \mathbf{h} &= \frac{\mathbf{1}'\Sigma^{-1}\mathbf{1}\Sigma^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}\Sigma^{-1}\mathbf{1}\Sigma^{-1}\mathbf{1}}{\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}\mathbf{1}'\Sigma^{-1}\mathbf{1} - \boldsymbol{\mu}'\Sigma^{-1}\mathbf{1}\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu}}\end{aligned}$$

are fixed vectors and \mathbf{w}_{μ_P} depends only on μ_P . To create the efficient frontier, the target return μ_P must only be varied between the maximum and minimum returns in the stock portfolio.

A major drawback of this technique is that only equality constraints can be accommodated. The constraint $w_i \geq 0$ for $i = 1, \dots, p$, that ensures positive weights to disallow short sales, can therefore not be applied. Since the Lagrangian multiplier approach has a closed-form solution, it is fast to compute, and can be effective when short sales are considered a legitimate trading operation.

2.2 Tangency portfolio and the risk-free asset

Theoretically, the risk-free rate of return is the rate at which money is borrowed or lent when there is no risk of default (credit risk), in other words, repayment is certain [11]. In practice, short term government bonds are considered virtually risk-free, since government can always print more money, thereby avoiding default on such bonds. The risks of inflation and interest rate fluctuation are assumed negligible over a short time period. Statistically, the term risk-free asset implies that the asset has zero variance.

Consider a portfolio of shares with return P and a basket of assets yielding risk-free rate r_f held in proportions α and $(1 - \alpha)$ respectively in a new portfolio with return Z , at an unspecified time t , where

$$Z = \alpha P + (1 - \alpha)r_f.$$

Then

$$\mu_Z = \alpha\mu_P + (1 - \alpha)r_f \quad (2.2)$$

$$\sigma_Z = \alpha\sigma_P \quad (2.3)$$

where μ_Z is the expected return of portfolio Z at time t , σ_Z is the standard deviation, and P and Z are random variables. Solving Equations (2.2) and (2.3) simultaneously for the proportion α gives

$$\mu_Z = r_f + \frac{\mu_P - r_f}{\sigma_P} \sigma_Z. \quad (2.4)$$

Fixing σ_Z , the maximum actual return on portfolio Z lies on the tangent to the efficient frontier with a slope maximizing the Sharpe ratio [12],

$$\frac{\mu_P - r_f}{\sigma_P}, \quad (2.5)$$

where the net returns in excess of the risk-free rate $\mu_P - r_f$ are risk-adjusted by dividing by the standard deviation σ_P of the portfolio. The investor therefore strives to allocate limited funds so as to maximize return per unit risk as stipulated in the Sharpe ratio. This allocation is known as the tangency portfolio. Figure 2.1 illustrates the tangent line drawn from the risk-free rate of return to the tangency portfolio which maximizes the Sharpe ratio. This line, referred to as the capital allocation line (CAL), is expressed as

$$\mu_Z = r_f + \frac{\mu_T - r_f}{\sigma_P} \sigma_Z \quad (2.6)$$

where μ_T is the expected return on a portfolio of shares at the point of tangency.

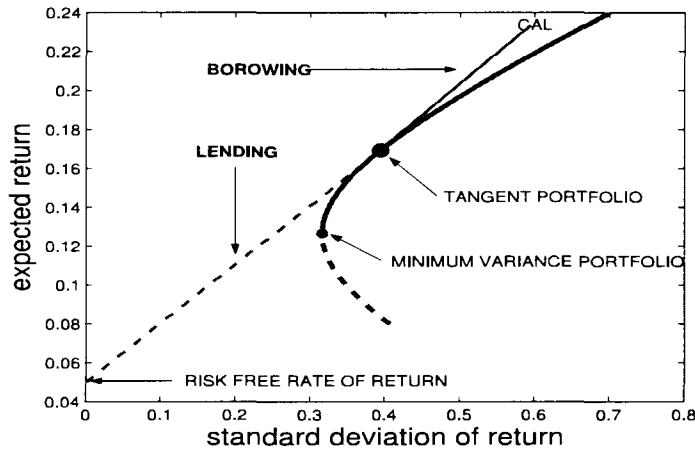


Figure 2.1: Tangency Portfolio and the Capital Allocation Line

The risk-averse investor will invest some funds in the risk-free asset and place the remainder in the tangency portfolio. This investor is lending money to the market, an action represented by the dashed component of the CAL in Figure 2.1. Alternatively, an investor could opt to take on more risk by borrowing money to invest in the tangency portfolio. This is illustrated by the solid part of the CAL in Figure 2.1. This means that Equation (2.6) enables satisfaction of any risk preference by a linear combination of the risk-free asset and the tangency portfolio. This is known as the separation property, and implies that a brokerage institution need only calculate the proportion of the client investment to be placed into the tangency portfolio, instead of individual client portfolios according to risk preference.

Whereas efficient portfolio theory assumes that individuals or companies lend and borrow money at a single risk-free rate, this is not the case, and the above formulation can accommodate this by varying borrowing and lending rates.

Furthermore, as the most salient feature of the Markowitz formulation is the importance of the efficient frontier in portfolio construction, the efficient frontier must be comprehensively researched in order for the tangency portfolio to convey the optimal investment.

2.3 The effect of correlation between stocks

In order to demonstrate the effect of correlation between stocks on the efficient frontier, consider an example using two hypothetical stocks, A and B. Stock A has expected return 15% and standard deviation 28.8%. Stock B has expected

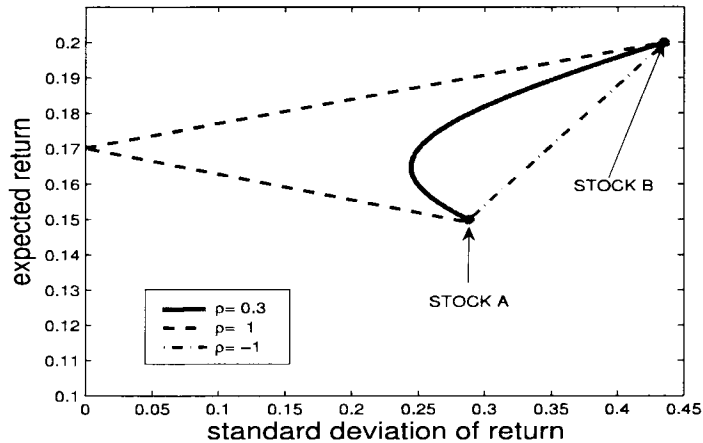


Figure 2.2: Effect of correlation on the efficient frontier

return 20% and standard deviation 43.5%. The correlations between the returns of the two stocks are varied, using correlation coefficient values $\rho = -1$, $\rho = 0.3$ and $\rho = 1$. The resulting frontier changes are displayed in Figure 2.2. Perfect positive correlation, where $\rho = 1$, results in a straight line between the assets, in other words, if one asset increases in value, so will the other. There is no benefit in diversifying between Stocks A and B in this case, since the stocks move in lockstep. Perfect negative correlation, $\rho = -1$, where returns on A and B move in opposite directions, implies that the assets could provide perfect diversification. In Figure 2.2, this is shown by the fact that the efficient frontier touches the expected return axis at 17% return, with 0% risk. This means that 17% risk-free profit could be generated by holding these two assets. All other correlations between $\rho = -1$ and $\rho = 1$ result in more traditional efficient frontiers encapsulated between the perfect correlation frontiers.

It is clear that inter-stock correlations significantly affect choice of fund allocation among assets. Therefore, correlations between assets are as important to correct modelling as variances and expected returns. Furthermore, even if correlations between assets could be estimated without sampling error, these correlations are not time-invariant. That correlations among assets change dramatically during market crises emphasizes the necessity for correct explanation thereof, particularly for investors interested in hedging positions.

2.4 The Capital Asset Pricing Model

The theory behind the Capital Asset Pricing Model (CAPM) was developed by Sharpe [13] using simple calculus and geometric reasoning to describe the

return on a risky asset by relating returns to the market return [5]. The CAPM is an equilibrium theory which states that the market portfolio coincides with the tangency portfolio of risky assets and is the most efficient to hold. When the market is in equilibrium and efficient, with all assets held to market value weights, the following CAPM equation will hold for the expected return $E(R_i)$ on a risky asset i in a market of any number of risky assets:

$$E(R_i) = r_f + \frac{\sigma_{im}}{\sigma_m^2} [E(R_m) - r_f] \quad (2.7)$$

where r_f is the rate of return on the risk-free asset, σ_{im} is the covariance between the i^{th} asset and the market, and σ_m^2 represents the market risk [10].

In reality, the market is rarely in equilibrium and few, if any, of the CAPM assumptions hold. Nevertheless, the following assumptions are practical simplifications which make the CAPM a useful tool for quantifying and pricing risk [12]:

1. Investors are price takers. No single investor has the resources to affect the market. The total assets held in the market are fixed but perfectly divisible, suggesting it is possible to own a fraction of a share. The market is perfectly competitive.
2. All investors are invested in the market for one time period that is the same length for all investors.
3. Investments are limited to publicly-traded equities and bonds. There is also a fixed risk-free asset which investors can lend to and borrow from in an unlimited amount.
4. The market is frictionless and hence there are no tax implications or costs involved with trading equities. Information has no cost and is available to all investors.
5. Investors are rational, risk-averse individuals trying to maximize wealth at the end of the investment period. Moreover, investors are mean-variance optimizers with respect to the Markowitz portfolio selection. This means each investor will hold the tangency portfolio of risky assets as well as the risk-free asset, with only the mix between the two asset classes differing among investors according to risk preference.
6. Investor expectations of risks and returns are homogeneous.

The above assumptions facilitate a link between the CAPM and portfolio selection methodology. If the above assumptions hold, all investors will select the

same tangency portfolio for risky assets and the CAL will be the same for all investors, who would be required to hold stocks in the same, efficient proportions. If summed over all the investors in the market, the tangency portfolio must be tantamount to the efficient market portfolio in equilibrium, since the sum of all investors is the market.

However, if investors have different views and expectations, the market portfolio is not necessarily efficient. The efficiency of the market portfolio and the CAPM theory are inseparable hypotheses, since inefficiency of the market portfolio implies that investors will hold portfolios other than the efficient market portfolio.

2.5 The Securities Market Line

Consider re-arranging Equation (2.7) to obtain

$$\mu_i - r_f = \beta_i(\mu_m - r_f) \quad (2.8)$$

where $\mu_i = E(R_i)$ is the expected return of stock i , $\mu_m = E(R_m)$ is the expected return of the market respectively and β_i is the beta coefficient, which describes the riskiness of an asset i relative to σ_m^2 . More specifically, the beta coefficient is defined by

$$\beta_i = \frac{\sigma_{im}}{\sigma_m^2} = \frac{\text{cov}(R_i, R_m)}{\text{var}(R_m)}.$$

An example of this relationship is shown in Ruppert [10].

The Securities Market Line (SML) refers to the linear relationship produced by graphing excess expected returns against beta coefficients for all assets, that is, $\mu_i - r_f$ against β_i , where $i = 1, \dots, p$ [10]. According to the CAPM, extra return can only be obtained by assuming more beta risk, which means taking on more market risk. All assets should, in theory, lie on the SML. If a stock lies below the SML, which means that for the same amount of beta risk more excess return could be obtained, that stock is overpriced and for the market equilibrium to be reached, the price must decrease and the return increase.

Both the CAPM and the resulting SML are equilibrium models, which implies correct asset pricing only in equilibrium. However the models do not indicate time to reach market equilibrium, or indeed if equilibrium will be reached. The beta concept therefore assigns a monetary value risk, but cannot necessarily provide a means of earning abnormal economic profit.

2.6 Index models

2.6.1 The Single Index Model

This methodology, developed by Sharpe, estimates the beta coefficient in practice using linear regression [13]. Define i now as a specific stock in a portfolio containing p stocks, so $i = 1, \dots, p$. Let R_{it} be the return stock i , R_{mt} the return on the market portfolio m and r_{ft} the risk-free rate of return, all at time t . The SML specification in Equation (2.8) can be expressed as a regression model by

$$R_{it} = r_{ft} + \beta_i(R_{mt} - r_{ft}) + \epsilon_{it} \quad (2.9)$$

where ϵ_{it} is assumed to be an independent error term distributed as $N(0, \sigma_{\epsilon_i}^2)$ for $t = 1, \dots, T$, where T represents the final time period measured [10].

The above linear model is often referred to as the security characteristic line. Taking expectations on both sides of Equation (2.9) produces the SML specified in Equation (2.8). The variance of the risk-free rate r_{ft} is zero, so the variance of the i^{th} stock can be rewritten as

$$\sigma_i^2 = \beta_i^2 \sigma_m^2 + \sigma_{\epsilon_i}^2 \quad (2.10)$$

and the covariance between the i^{th} and j^{th} stock as

$$\sigma_{ij} = \beta_i \beta_j \sigma_m^2 \text{ where } i, j = 1, \dots, p \text{ } i \neq j \quad (2.11)$$

since ϵ_{it} and ϵ_{jt} are independent. Equation (2.10) shows how total risk on an asset can be broken into systematic, or market, risk $\beta_i^2 \sigma_m^2$ and unsystematic, or unique risk $\sigma_{\epsilon_i}^2$.

Although short-term government bonds or overnight swap rates can be used to estimate the risk-free rate, the result will still be a stochastic rather than a deterministic variable, hence risk-free rate does not exist in reality. In order to regress two time series of excess returns against each other with no intercept term, a reliable risk-free rate must be assumed. Consider, therefore, the following regression model:

$$(R_{it} - r_{ft}) = \alpha_i^* + \beta_i(R_{Mt} - r_{ft}) + \epsilon_{it} \quad (2.12)$$

where α_i^* is a constant. The intercept term, commonly referred to as α^* , can be used to verify the correctness of the CAPM, which predicts this constant term to be zero. A positive α^* implies that a stock is underpriced, since the excess returns are greater on average than CAPM predictions. According to the Single Index Model, α_i^* values should average to zero over all securities, implying that, on average, securities are not overpriced.

2.6.2 Applications of the Single Index Model

Consider the revised regression given below, which uses log-return r_{it} instead of excess returns:

$$r_{it} = \alpha_i + \beta_i r_{mt} + \epsilon_{it} \quad i = 1, \dots, p, \quad (2.13)$$

where α_i and β_i are unknown parameters, r_{mt} is the log-return of the market index, and ϵ_{it} is an error term at time $t = 1, \dots, T$. The estimated beta coefficient will be the same as that of Equation (2.12), since the same linear relationship is estimated. This formulation removes the need to model a risk-free rate for the entire sample period, and will therefore be referred to in the following sections.

The basic time series model for the log-return r_{it} of the i^{th} asset can now be expressed as

$$r_{it} = \mu_{it} + a_{it} \quad (2.14)$$

where the log-return r_{it} is modelled by a conditional mean μ_{it} and an error term a_{it} , with conditional variance σ_{it}^2 , for stock i . With regards to r_{it} , the conditional mean and variance are defined as

$$\mu_{it} = E(r_{it} | F_{t-1}) \quad (2.15)$$

$$\sigma_{it}^2 = \text{var}(a_{it} | F_{t-1}) \quad (2.16)$$

where F_{t-1} is the information set up to time $t - 1$. Note that Equations (2.15) and (2.16) define the mean and variance conditional on the information set F_{t-1} respectively.

The conditional mean is usually described by an autoregressive moving average (ARMA) model, which may include some explanatory economic variables [3]. For parsimony, only the market index return series will be considered. So, the conditional mean can be modelled as follows:

$$\mu_{it} = \gamma_i + \beta_i r_{mt} + \sum_{j=1}^p \phi_{ij} r_{i,t-j} - \sum_{k=1}^q \theta_{ik} a_{i,t-k} \quad (2.17)$$

where the parameters $\phi_{i1}, \dots, \phi_{ip}$ and $\theta_{i1}, \dots, \theta_{iq}$ are ARMA(p,q) model parameters for the mean and r_{mt} is the log-return series of a market index for stocks $i = 1, \dots, p$. Note that γ_i is a constant term for the ARMA process, incorporating the constant term α_i from the Single Index Model described in Equation (2.13). The above univariate ARMA(p,q) model (2.17) will ensure that the return series is serially uncorrelated, but does not remove correlation between the log-return series of different stocks.

The value for each c_{it}^2 , the unique risk for stock $i = 1, \dots, p$ and time $t = 1, \dots, T$ is estimated by the square of

$$\hat{c}_{it} = r_{it} - \hat{\mu}_{it} = r_{it} - \hat{r}_{it} \quad (2.18)$$

where \hat{r}_{it} is the estimated return of stock i at time t using Equation (2.17).

After the removal of the the linear effect of the market and the ARMA effects of the stock from a return series, only the risk unique to the asset remains. The portfolio selection problem described previously requires the expected return and variance of each stock to be estimated. Using the time series model in (2.17), the return of a stock can be described by the regression parameters, α_i and β_i , the ARMA(p,q) parameters, and the variance by β and σ_{ic}^2 .

The ARMA(p,q) component is easily forecast using theory found in Tsay [3]. This means that the expected return of a stock can be forecast with only its β_i value together with the estimated expected market return in conjunction with the ARMA(p,q) forecasts.

Unique risks and beta values of stocks can be used in Equations (2.10) and (2.11) to construct the covariance matrix after computing the market volatility σ_m^2 . So α , β , and uncorrelated error terms ϵ_{it} are estimated p times each, and the expected market return and market volatility are also estimated. In other words, $3p+2$ quantities must be estimated, where p is the number of stocks, which is more workable than the $\frac{1}{2}p(p+1)$ estimates required for the covariance matrix in the original Markowitz portfolio selection formulation, although this simplification is at the cost of assuming no inter-stock correlation.

2.6.3 Factor Models

The Arbitrage Pricing Theory (APT) was developed by Ross in 1976 [14] using arbitrage and the law of one price to elaborate the simpler CAPM, since the market is not the only financial market risk factor. Any quantifiable risk factor can be included in the factor model, ranging from sector indexes to interest rate data or any time series that can provide a measure of a certain type of risk. This provides one option for eliminating some correlation between the log-return series of stocks. A factor model can be defined by the following multiple linear regression:

$$R_{it} = \alpha_i + \beta_{i1}Y_{1t} + \beta_{i2}Y_{2t} + \dots + \beta_{ik}Y_{kt} + \epsilon_{it}, \quad (2.19)$$

where there are k different factors Y_1, Y_2, \dots, Y_k recorded at time t , with the corresponding k unknown parameters $\beta_{i1}, \dots, \beta_{ik}$ respectively, ϵ_{it} the associated error term, and there are $i = 1, \dots, p$ stocks and with $t = 1, \dots, T$ sample time

periods.

The model in (2.19) can be estimated as a standard multiple linear regression model using maximum likelihood estimation (MLE), assuming normality. Every factor filters out some linear risk. With enough linear factors all commonalities between assets should be removed. Hence, the error series ϵ_{it} should be unique to the i^{th} asset, and there should be no correlation across assets.

2.7 Diversification

Consider a portfolio with p asset log-returns r_{1t}, \dots, r_{pt} for holding period t . Defining the portfolio weights as w_1, \dots, w_p means that the return on a portfolio r_{Pt} is given by

$$r_{Pt} = w_1 r_{1t} + \dots + w_p r_{pt}.$$

Using the market model of Equation (2.13), each return can be re-written as $r_{it} = \alpha_i + \beta_i r_{mt} + \epsilon_{it}$ for $i = 1, \dots, p$ where r_{mt} is the return on the market portfolio at time t . Hence the return r_{Pt} of the portfolio of p assets is

$$r_{Pt} = \sum_{i=1}^p w_i \alpha_i + \left(\sum_{i=1}^p w_i \beta_i \right) r_{mt} + \sum_{i=1}^p w_i \epsilon_{it} \quad (2.20)$$

The beta for a portfolio is therefore the weighted average of the individual β_i of every stock in that portfolio. As in Section 2.6.1, the total portfolio risk σ_{Pt}^2 can also be separated into systematic and unsystematic risk as follows:

$$\sigma_{Pt}^2 = \left(\sum_{i=1}^p w_i \beta_i \right)^2 \sigma_{mt}^2 + \text{var} \left(\sum_{i=1}^p w_i \epsilon_{it} \right) \quad (2.21)$$

where $\left(\sum_{i=1}^p w_i \beta_i \right)^2 \sigma_{mt}^2$ is the systematic or market risk and $\text{var} \left(\sum_{i=1}^p w_i \epsilon_{it} \right)$ represents the unique risk for a particular time t .

If the error terms $\epsilon_{1t}, \dots, \epsilon_{pt}$ for all t are assumed to be uncorrelated, the unique risk can be rewritten as

$$\sigma_{\epsilon P,t}^2 = \sum_{i=1}^p w_i^2 \sigma_{\epsilon i,t}^2$$

Now consider an equally-weighted portfolio with $w_i = 1/p$ for all $i = 1, \dots, p$. The systematic risk becomes $\left(\sum_{i=1}^p \frac{1}{p} \beta_i \right)^2 \sigma_{mt}^2 = \bar{\beta}^2 \sigma_{mt}^2$, where $\bar{\beta}$ is the average of the β_i coefficient in a portfolio over stocks $i = 1, \dots, p$ at a particular time t . As the size of a portfolio increases it will converge to the market portfolio so that the average beta will converge to one and hence market risk is always present in

a portfolio.

Unique risk, however, is represented by

$$\sigma_{\epsilon P,t}^2 = \frac{\sum_{i=1}^p \sigma_{\epsilon i,t}^2 / p}{p} = \frac{\bar{\sigma}_{\epsilon t}^2}{p},$$

where $\bar{\sigma}_{\epsilon}^2$ is the average of the $\sigma_{\epsilon i}^2$ over stocks $i = 1, \dots, p$ at a particular time t , which will converge to 0 as the number of stocks p increases. The term diversification refers to the fact that the risk unique to an individual share can be reduced to a negligible value by holding a diverse portfolio.

However, this argument becomes flawed when assets are correlated over and above the market risk or any other risk factor chosen [15]. Consider the case where linear risk factors do not remove all correlation between the errors $\epsilon_{1t}, \dots, \epsilon_{pt}$ for each time t . Then the unique risk for the portfolio becomes

$$\sigma_{\epsilon P,t}^2 = \sum_{i=1}^p w_i^2 \sigma_{\epsilon i,t}^2 + 2 \sum_{i < j} w_i w_j \text{cov}(\epsilon_{it}, \epsilon_{jt}).$$

For simplicity of illustration, assume that the shares are equally weighted, so that

$$\begin{aligned} \sigma_{\epsilon P,t}^2 &= \frac{1}{p} \sigma_{\epsilon i,t}^2 + \sum_{i < j} \frac{1}{p} \frac{1}{p} \sigma_{\epsilon ij,t} \\ &= \frac{1}{p} \sigma_{\epsilon i,t}^2 + \frac{p(p-1)}{p^2} \sum_{i < j} \left(\frac{\sigma_{\epsilon ij,t}}{p(p-1)} \right) \\ &= \frac{1}{p} \sigma_{\epsilon i,t}^2 + \frac{p-1}{p} \bar{\sigma}_{\epsilon,t}, \end{aligned} \tag{2.22}$$

where $\bar{\sigma}_{\epsilon,t}$ is the average covariance between all the ϵ_{it} and ϵ_{jt} [16] for a particular time t . Equation (2.22) shows unique risk of a portfolio cannot be completely diversified away if covariance exceeds the linear risk factors. More importantly, as p tends toward a larger number, the average covariance between the errors ϵ_{it} and ϵ_{jt} becomes dominant compared to the diminishing unique risk. This means that in a large portfolio the actual unique risk tends toward a covariance risk. It would therefore be informative to model the entire covariance matrix at each time point t to examine the effect of between-stock covariance on diversification.

Chapter 3

Objectives and Aims of the Thesis

For effective portfolio selection, the expected return on, and covariance among, a selected group of p stocks need to be estimated into the future. This thesis investigates which such method of extrapolation provides the most effective results.

3.1 The Single Index as opposed to the Factor Model

The simpler single index model, which requires forecasting of only expected market return and market variance, will be implemented in this thesis. The reason for this is that the factor model has certain disadvantages, namely a lack of default method of factor selection, and the fact that some factors influence some stocks and not others. To compare stocks within a factor model framework, the same factors must be used for all stocks analyzed. For example, the gold price is a good linear predictive factor for gold stocks, but for a clothing retailer this is likely to be irrelevant. Model building therefore requires finding the correct factors that significantly affect all the stocks in the chosen portfolio. Another concern is parsimony, as many additional factors may remove correlation, when in reality some factors may be irrelevant. In order for the factor model to be properly used in portfolio optimization, a forecast value for each factor is required to estimate the expected returns for each stock in the portfolio at the next time period. The method of forecasting interest rates is far removed from that of forecasting Gross Domestic Product (GDP), since so many specialized fields are required. Extrapolation of the multiple factors used is therefore of great concern in factor modelling.

3.2 Univariate modelling of expected return and variance

Prior to the evolution of the Single Index Model from the CAPM, individual expected returns were calculated for every stock in a portfolio. Equation (2.13) in Section 2.6.2 showed forecasting of expected return using an ARMA(p,q) model in conjunction with a Single Index Model. Assuming the α and β coefficients of the CAPM, and hence the estimates of these parameters ($\hat{\alpha}$ and $\hat{\beta}$), are stable over time, only market return needs to be forecast.

Univariate volatility models will be used to generate expected market variance. Equation (2.10) shows the breakdown of risk associated with a stock i into systematic ($\hat{\beta}_i^2 \hat{\sigma}_{mt}^2$) and unique ($\sigma_{\epsilon_i}^2$) risk. Since β_i is usually stable over time, the systematic risk of a stock can be forecast using its computed β_i and the forecast market variance as $\hat{\beta}_i \hat{\sigma}_{M,T+1}^2$ from a time series defined for time periods $t = 1, \dots, T$. The off-diagonal elements of the covariance matrix for systematic risk will be calculated with estimates inserted into $\hat{\beta}_i \hat{\beta}_j \hat{\sigma}_{T+1}^2$. Univariate volatility models can be used to forecast the unique risk of a stock i over time $t = 1, \dots, T$, specified by $\epsilon_{i1}^2, \dots, \epsilon_{it}^2$, since $\epsilon_{it} \sim N(0, \sigma_{\epsilon_i}^2)$. The ϵ_{it} values are estimated as the difference between r_{it} and the estimated value \hat{r}_{it} in Equation (2.18) in Section 2.6.2. Univariate volatility models are used, since the standard CAPM methodology for estimating the efficient frontier assumes that the market is the only risk factor and therefore unique risks have no correlation beyond that of the market.

These univariate models will therefore serve as a benchmark for more elaborate models. Since historical data were available for this thesis, the actual ('future') value of the market $\mu_{M,T+1}$ will be available for comparison with predictions.

3.3 Multivariate volatility models

The diversification Equation (2.22) indicates that covariance exceeding that of the market has a potentially large effect on unique portfolio risk, and hence on the risk of the entire portfolio. A dynamic forecast of covariance among unique risks would be particularly useful during times of market upheaval when covariances change rapidly. The best way to model covariance is to develop multivariate volatility models to simultaneously estimate unique risk and unique covariance risk.

The error terms from Equation (2.13), ϵ_{it} for $i = 1, \dots, p$ and $t = 1, \dots, T$, are assumed to be normally distributed with mean zero. From $\mathbf{e}_t = (\epsilon_{1t}, \epsilon_{2t}, \dots, \epsilon_{pt})'$

as a vector of errors for p stocks at time t ,

$$\Sigma_t = E(\mathbf{e}_t \mathbf{e}_t') = E \begin{pmatrix} \epsilon_{1t}^2 & \epsilon_{1t}\epsilon_{2t} & \cdots & \epsilon_{1t}\epsilon_{pt} \\ \epsilon_{2t}\epsilon_{1t} & \epsilon_{2t}^2 & \cdots & \epsilon_{2t}\epsilon_{pt} \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon_{pt}^2\epsilon_{1t} & \epsilon_{pt}\epsilon_{2t} & \cdots & \epsilon_{pt}^2 \end{pmatrix} \quad (3.1)$$

where Σ_t is the time-indexed covariance matrix of the error terms. The expected values of the diagonal elements of Equation (3.1), $E(\epsilon_{it}^2) = \sigma_{\epsilon, it}^2$, represent the unique risk of stock i at time t . The expectation of the off-diagonal elements of Equation (3.1), namely $E(\epsilon_{it}\epsilon_{jt}) = \sigma_{\epsilon, ij, t}$ for $i \neq j$, can now be defined as the unique covariance risk between stocks i and j for $i, j = 1, \dots, p$.

Dropping the assumption of uncorrelated error terms affects the covariance matrix implied by CAPM in Equation (2.11). The systematic risk remains the same, but the covariance between stocks i and j is modelled by

$$\text{cov}(r_{it}, r_{jt}) = \beta_i \beta_j \sigma_{mt}^2 + \sigma_{\epsilon, ij, t}$$

as given in Alexander [17]. In order to obtain a better forecast of risk, the unique covariance between shares will be modelled using a Single Index Model together with multivariate volatility models for the error terms. No attempt will be made to explain the source of the covariance.

3.4 Parsimony

The issue of parsimony is central to the multivariate volatility models explored in later chapters. The advantages of the modified CAPM suggested in this thesis outweigh the use of extra parameters. Modified CAPM uses the CAPM methodology to estimate the expected returns of p risky assets, as well as enabling the risk of holding a stock to be broken into market, unique and covariance risk. Since certain groups of stocks seem to move together, the need arises to analyze this risk statistically, with the most modern methods available.

In order to calculate efficient portfolio weights, the original Markowitz formulation requires estimation of the entire covariance matrix, with $\frac{1}{2}p(p+1)$ parameters, where p is the number of stocks. An advantage of CAPM methodology is that only $3p+2$ parameters (where $p \geq 1$) are required to model the covariance matrix and expected returns of individual shares. In this thesis, the assumption of covariance risk exceeding that of the market risk means that the covariance matrix in Equation (3.1) must be estimated as well as single index models for each of p stocks. This procedure, to be referred to as the Modified CAPM, requires $\frac{1}{2}p(p+1) + 3p+2$ parameters.

3.5 Aim of the thesis

In summary, two approaches for forecasting expected returns and variances will be considered and compared in this thesis. The first approach involves the original Single Index Model, with univariate volatility models employed to obtain a forecast of unique risk. This approach will be used as a benchmark for the second, more complicated approach, which will invoke the Single Index Model with unique covariance risk. Multivariate volatility models will be employed to model and forecast the entire unique risk covariance matrix. Although this modelling approach has been suggested by Alexander [17], it has yet to be implemented on a portfolio of risky assets.

University of Cape Town

Chapter 4

Univariate Volatility Model Theory

4.1 Introduction

Economists have suggested numerous linear and non-linear time series and regression models to explain expected return, but most of these models incorporate a static estimate of variance. Tsay [3] points out that in the real world, however, volatility, as the conditional standard deviation of underlying asset return, evolves continuously rather than by way of volatility jumps. Volatility is any measure of the variation in the range of different returns assumed by an asset, and seems to react differently to increasing and decreasing market prices. The practice of dynamically modelling time series variance is called volatility modelling.

Only a few univariate volatility models were selected for use in this thesis in order to serve as benchmarks for the multivariate models in portfolio selection and market risk assessment.

4.2 Univariate GARCH Model

The basic time series model for a log-return r_t of a single stock at time t can be defined as

$$r_t = \mu_t + a_t,$$

where μ_t is the conditional mean and a_t is the innovation or error term. The aim of univariate volatility modelling in financial portfolio theory is to model the variance of the innovation series a_t of a single stock over time $t = 1, \dots, T$.

The conditional mean μ_t and variance σ_t^2 are defined in Equations (2.15) and

(2.16) as

$$\begin{aligned}\mu_t &= E(r_t|F_{t-1}) \\ \sigma_t^2 &= \text{var}(r_t|F_{t-1}) = E((r_t - \mu_t)^2|F_{t-1}),\end{aligned}\quad (4.1)$$

where F_{t-1} is the set of information up to time $t - 1$.

The conditional mean μ_t is described by a combination of the Single Index Model and an autoregressive moving average (ARMA) model in Equation (2.17) in Section 2.6.2 as

$$\mu_t = A + Br_{mt} + \sum_{i=1}^p \phi_i r_{t-i} - \sum_{j=1}^q \theta_j a_{t-j}, \quad (4.2)$$

where A and B are the α and β parameters from the Single Index Model detailed in Section 2.6.1, and ϕ_1, \dots, ϕ_p and $\theta_1, \dots, \theta_q$ are the parameters from the ARMA(p,q) model. The ARMA(p,q) model for the return series will ensure that the series is serially uncorrelated. The conditional variance can now be described by the model

$$\begin{aligned}a_t &= \sigma_t \epsilon_t = r_t - \mu_t \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^m \alpha_i a_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2\end{aligned}\quad (4.3)$$

where ϵ_t is an independently identically distributed (iid) random variable with mean zero and variance one [3], and μ_t is the conditional mean equation at time t . Also, $\{\alpha_0, \alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_s\}$ is a set of new parameters, all of which are restricted to be positive to ensure that the conditional variance remains positive under all circumstances. The formulation in Equation (4.3) is referred to as a Generalized Autoregressive Conditional Heteroscedastic (GARCH) model [18]. A special case of the GARCH(m,s) model where s equals zero is Engle's ARCH model [1].

4.2.1 Simultaneous and two-tier parameter estimation

Define the parameter set of the conditional mean μ_t in Equation (4.2) as $\Omega = \{A, B, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q\}$. Also, define the parameter set of the conditional variance σ_t^2 in Equation (4.3) as $\Lambda = \{\alpha_0, \alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_s\}$. Equation (4.1) shows the conditional variance σ_t^2 calculated using the conditional mean μ_t , implying parameter sets Ω and Λ can be jointly estimated under a single optimization of the conditional variance since the conditional mean is defined within the conditional variance formulation.

The process of jointly estimating the conditional mean and conditional variance

parameters will be referred to as simultaneous parameter estimation. Although a more complicated optimization, this optimization technique results in better parameter estimates than the two-tier approach described below.

The parameter sets for the conditional mean μ_t and conditional variance σ_t^2 could also be optimized separately. The process of estimating the parameter set Ω of the conditional mean in the absence of the parameter set Λ of the conditional variance will be referred to as two-tier parameter estimation. In other words, after fitting the parameters of the conditional mean equation μ_t , the realized innovations \hat{a}_t , calculated as $\hat{a}_t = r_t - \hat{\mu}_t$, are modelled separately to the log-returns r_t . The two-tier method results in two smaller parameter sets estimated per parameter optimization.

Simultaneous parameter optimization is usually preferred in cases where the total number of parameters in both the conditional mean and conditional variance sets is small. More parameters in any system of equations results in a greater probability of optimization failure, where the parameter estimates fail to converge. Optimization therefore entails two-tier methodology when considering parameter-heavy multivariate volatility models, while simultaneous parameter optimization is used for some univariate volatility models.

4.2.2 Unconditional variance for a GARCH(m,s) model

The unconditional variance of a_t will be derived for a GARCH(1,1) model to illustrate the definitions and concepts necessary to derive unconditional variances for other volatility models.

Define an information set

$$F_{t-1} = \{a_1^2, \sigma_1^2, a_2^2, \sigma_2^2, \dots, a_{t-1}^2, \sigma_{t-1}^2\} \quad (4.4)$$

as all the information up to time $t-1$ used to calculate a_t . A martingale difference is defined as a process a_t with $E|a_t| < \infty$ and $E(a_t|F_{t-1}) = 0$. Two critical properties of a series which is a martingale difference are that the series has an unconditional mean of zero and is serially uncorrelated [19]. The series a_t as defined in Equation (4.3) is a martingale difference since

$$\begin{aligned} E(a_t|F_{t-1}) &= E(\sigma_t \epsilon_t|F_{t-1}) \\ &= \sigma_t E(\epsilon_t|F_{t-1}) \\ &= 0 \end{aligned}$$

where σ_t can be calculated using the set F_{t-1} and is therefore conditionally a constant. Also, ϵ_t does not depend on F_{t-1} and $E(\epsilon_t) = 0$.

Second-order stationarity must be assumed to derive the unconditional variance. Now the variance of a_t is $E(a_t^2)$ since a_t has an unconditional mean of zero by way of the martingale difference property. Hence

$$\begin{aligned}
\text{var}(a_t) &= E(a_t^2) \\
&= E(\sigma_t^2 \epsilon_t^2) \\
&= E_{F_{t-1}}\{E(\sigma_t^2 \epsilon_t^2) | F_{t-1}\} \text{ by the Law of Iterated Expectations} \\
&= E_{F_{t-1}}\{\sigma_t^2 E(\epsilon_t^2) | F_{t-1}\} \\
&= E(\sigma_t^2) \text{ since } E(\epsilon_t^2) = 1 \\
&= E\{\alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2\} \\
&= \alpha_0 + \alpha_1 E(a_{t-1}^2) + \beta_1 E(\sigma_{t-1}^2)
\end{aligned}$$

Stationarity implies that $E(a_t^2) = E(a_{t-1}^2)$ and by definition $E(a_t^2) = E(\sigma_t^2)$. So,

$$E(a_t^2) = \alpha_0 + \alpha_1 E(a_t^2) + \beta_1 E(a_t^2)$$

which implies that

$$\text{var}(a_t) = \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)}. \quad (4.5)$$

This can be generalised for a GARCH(m,s) model [3] as

$$\text{var}(a_t) = \frac{\alpha_0}{1 - \sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i)}.$$

Note that $\sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) < 1$ is a condition for stationarity for the GARCH(m,s) model to ensure that the unconditional variance remains positive and finite.

The kurtosis for a GARCH(1,1) model, derived in Tsay [20], is given by

$$\kappa = \frac{E(a_t^4)}{[\text{Var}(a_t)]^2} = \frac{3(1 - (\alpha_1 + \beta_1)^2)}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2} > 3$$

provided that the $\epsilon_t \sim N(0, 1)$. The kurtosis of greater than three is important in finance since it follows that a GARCH(1,1) model implies a fat-tailed distribution. The normal distribution does not have fat enough tails to model the extreme returns that can occur in a financial market.

4.2.3 Estimation for a GARCH(m,s) model

The joint probability density function (pdf) of the data can be written as a product of the conditional densities [3] as

$$\begin{aligned}
f(a_1, a_2, \dots, a_T | \theta) &= f(a_T, \sigma_T^2 | F_{T-1}) f(a_{T-1}, \sigma_{T-1}^2 | F_{T-2}) \times \\
&\quad \dots \times f(a_2, \sigma_2^2 | F_1) f(a_1, \sigma_1^2 | \theta)
\end{aligned} \quad (4.6)$$

where $\theta = \{\alpha_0, \alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_s\}$ denotes the set of parameters that is to be estimated. In the case of a GARCH(1,1) model, the set θ would contain $\{\alpha_0, \alpha_1, \beta_1\}$. Assume a_1 and σ_1^2 are known and equal to observed values in order to obtain the conditional likelihood

$$f(a_2, \dots, a_T, \sigma_2^2, \dots, \sigma_T^2 | \theta; a_1, \sigma_1^2) = f(a_T, \sigma_T^2 | F_{T-1}) f(a_{T-1}, \sigma_{T-1}^2 | F_{T-2}) \times \dots \times f(a_2, \sigma_2^2 | \theta; a_1, \sigma_1^2). \quad (4.7)$$

This is considered reasonable for large sample sizes [3].

In order to build a likelihood function a distributional assumption needs to be made. The normal distribution is insufficiently leptokurtic to describe financial data. Fatter tails are required to capture all the extreme events observed in the marketplace, so Student's t distribution and the generalized error distribution have been suggested as alternatives [3]. In contrast, Hansen and Lunde [6] found that a likelihood that assumes the normal distribution is a better option with respect to one-step ahead volatility forecasting. For the sake of continuity between the univariate and multivariate volatility models, the normal distribution will be used in this thesis.

Assume that $a_t | F_{t-1} \sim N(0, \sigma_t^2)$. Then the conditional log-likelihood function can be expressed as

$$\begin{aligned} l_c(\theta | a_1, \sigma_1^2) &= \ln f(a_2, \dots, a_T, \sigma_2^2, \dots, \sigma_T^2 | \theta; a_1, \sigma_1^2) \\ &= -\frac{1}{2} \sum_{t=2}^T \ln(2\pi\sigma_t^2) - \frac{1}{2} \sum_{t=2}^T \frac{a_t^2}{\sigma_t^2}. \end{aligned} \quad (4.8)$$

where σ_t^2 is defined recursively in Equation (4.3). The log-likelihood function given in Equation (4.8) can be maximized with respect to the parameter set θ , but is non-linear in these parameters. The maximization must be performed numerically with an appropriate non-linear optimization routine. In most software packages the Brendt, Hall, Hall, and Hausman (BHHH) method of optimization is favoured for GARCH parameter estimation [19].

4.2.4 Forecasting

A major application of time series analysis is forecasting future values of a given series based on information from the past. Suppose that interest lies in forecasting the volatility σ_{T+l} , where T is the forecast origin and $l \geq 1$ is the forecast horizon. Let $\hat{\sigma}_T(l)$ be the forecast of σ_{T+l} given F_T , the set of information up to and including the forecast horizon T . The forecast $\hat{\sigma}_T(l)$ is chosen such that [3]

$$E\{[\sigma_{T+l} - \hat{\sigma}_T(l)]^2 | F_T\} \leq \min_g E[(\sigma_{T+l} - g)^2 | F_T]$$

where g is some measurable function of the observations [19]. This is called the minimum square error loss function method of forecasting. Using Equation (4.3), the GARCH(1,1) model can be written as

$$\sigma_{T+1}^2 = \alpha_0 + \alpha_1 a_T^2 + \beta_1 \sigma_T^2. \quad (4.9)$$

Under minimum square error loss [3] and Equation (4.9), the point forecast of σ_{T+1}^2 given $F_T = \{a_1^2, \dots, a_T^2, \sigma_1^2, \dots, \sigma_T^2\}$ is the conditional expectation

$$\begin{aligned} \hat{\sigma}_T^2(1) &= E(\sigma_{T+1}^2 | F_T) \\ &= E(\alpha_0 + \alpha_1 a_T^2 + \beta_1 \sigma_T^2 | F_T) \\ &= \alpha_0 + \alpha_1 a_T^2 + \beta_1 \sigma_T^2 \end{aligned} \quad (4.10)$$

Equation (4.10) is a forecast that does not take into account uncertainty due to parameter estimation. This is not the case in practice where estimated parameters from the fitted model are used. Hence the parameter set for the GARCH(1,1) model is $\hat{\theta} = \{\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1\}$ and is used to calculate estimates for $\sigma_2^2, \dots, \sigma_T^2$ recursively from the equation $\hat{\sigma}_{t+1}^2 = \hat{\alpha}_0 + \hat{\alpha}_1 a_t^2 + \hat{\beta}_1 \hat{\sigma}_t^2$ for $t = 1, \dots, T$ assuming that the initial volatility σ_1^2 is known. Conventionally, σ_1^2 is assumed equal to the first squared shock a_1^2 .

When the sample size is sufficiently large, the difference in forecasts using the parameter uncertainty forecast methodology and the minimum square error loss methodology is negligible. The more complicated parameter uncertainty forecast methodology will therefore not be considered in this thesis. Similar considerations apply for the GARCH(m,s) model [3].

Although multi-step forecasts are not required in this thesis, the theoretical conditional forecast illustrates a salient feature of the GARCH family of models. Consider rewriting Equation (4.3) as

$$\sigma_{T+1}^2 = \alpha_0 + (\alpha_1 + \beta_1) \sigma_T^2 + \alpha_1 \sigma_T^2 (\epsilon_T^2 - 1), \quad (4.11)$$

since $a_t^2 = \sigma_t^2 \epsilon_t^2$. The expectation of Equation (4.11) simplifies to

$$E(\sigma_{T+1}^2 | a_T^2, \sigma_T^2, \hat{\theta}) = \sigma_T^2(1) = \alpha_0 + (\alpha_1 + \beta_1) \sigma_T^2$$

and so the l^{th} step-ahead forecast can be written as

$$\hat{\sigma}_T^2(l) = \alpha_0 + (\alpha_1 + \beta_1) \hat{\sigma}_T^2(l-1) \quad (4.12)$$

for $l > 1$ [3]. By repeatedly substituting the $(l-1)$ -step-ahead forecast into Equation (4.12), the l -step ahead volatility forecast can be written in terms of the 1-step-ahead forecast [3] as

$$\sigma_T^2(l) = \frac{\alpha_0(1 - (\alpha_1 + \beta_1)^{l-1})}{1 - (\alpha_1 + \beta_1)} + (\alpha_1 + \beta_1)^{l-1} \sigma_T^2(1). \quad (4.13)$$

It is interesting that the expression for $\sigma_T^2(l)$ in Equation (4.13) tends toward the unconditional variance of a_t as l approaches infinity. This property is called mean reversion.

4.3 Integrated GARCH Model

The Integrated GARCH (IGARCH) model is defined as a GARCH model where $\sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) = 1$. If the GARCH(1,1) model is taken as an example, with $\alpha_1 + \beta_1 = 1$, then Equation (4.5) results in a mathematical singularity in the unconditional variance of a_t , due to division by zero. This implies the unconditional variance does not exist, and therefore weak stationarity does not apply. Formally, an IGARCH(1,1) model can be written as

$$\begin{aligned} a_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \alpha_0 + (1 - \beta_1)a_{t-1}^2 + \beta_1\sigma_{t-1}^2 \end{aligned} \quad (4.14)$$

The parameters for an IGARCH(1,1) model can be estimated using the same methods as those for the GARCH models.

In practice, the parameters of GARCH models, excluding the intercept, occasionally sum to very close to one. This is referred to as persistence in volatility. The parameters of a GARCH(1,1) are constrained to sum to less than one, so risk may be understated during times of low volatility. For a risk manager the IGARCH volatility model provides the parsimony of a GARCH model with a more prudent, slightly higher, volatility estimate. This was one of the reasons the investment group JP Morgan adopted the IGARCH(1,1) model without intercept for its RiskmetricsTM methodology.

Forecasting with an IGARCH model has interesting implications in certain special cases. For example, the forecast generated for an IGARCH(1,1) model is

$$\sigma_T^2(l) = \alpha_0 + \sigma_T^2(l-1) \quad l \geq 1 \quad (4.15)$$

using Equation (4.12) and the fact that $\alpha_1 + \beta_1 = 1$. By repeated substitution into Equation (4.15), the forecast can be rewritten as

$$\sigma_{T+l}^2 = \sigma_{T+1}^2 + (l-1)\alpha_0 \quad l \geq 1. \quad (4.16)$$

The conditional volatility forecast grows linearly with respect to the constant term α_0 , which shows that the model is non-stationary. The forecast for conditional variance will also tend to infinity as l tends to infinity. The current shock persists indefinitely into the future, and therefore multi-period forecasts with an IGARCH(1,1) model are of little value.

4.4 Exponential GARCH Model

A perceived weakness of the GARCH model is the failure to model the asymmetric nature of asset returns. In financial markets, negative impacts may have a greater effect on asset volatility than positive shocks, a phenomenon called the leverage effect. Nelson [21] accommodated asymmetry by proposing the Exponential GARCH (EGARCH) model by considering the following weighted innovation:

$$g(\epsilon_t) = \begin{cases} (\theta + \gamma)\epsilon_t - \gamma E(|\epsilon_t|) & \text{if } \epsilon_t \geq 0, \\ (\theta - \gamma)\epsilon_t - \gamma E(|\epsilon_t|) & \text{if } \epsilon_t < 0, \end{cases} \quad (4.17)$$

where θ and γ are unknown parameter constants and ϵ_t is a zero-mean iid sequence with a continuous distribution, such as the standard normal distribution. It can be shown that $E[g(\epsilon_t)] = 0$. Specifically, the parameters of the innovation in Equation (4.17), namely θ and γ , model different aspects of asymmetry. The parameter θ weighs the effect of sign on volatility, while γ measures a magnitude effect for asymmetry. Note that the term $E(|\epsilon_t|)$ differs for each distribution and is equal to $\sqrt{\frac{2}{\pi}}$ for the standard normal distribution.

The EGARCH(p,q) model can now be defined as

$$\begin{aligned} a_t &= \sigma_t \epsilon_t \\ \ln(\sigma_t^2) &= \alpha_0 + \sum_{i=1}^q \alpha_i g(\epsilon_{t-i}) + \sum_{j=1}^p \beta_j \ln(\sigma_{t-j}^2) \end{aligned} \quad (4.18)$$

where α_i for $i = 0, \dots, q$ and β_j for $j = 1, \dots, p$ are unknown parameters. One difference between the EGARCH(p,q) and GARCH(m,s) models is that the logarithm of the volatility is modelled in EGARCH(p,q). This ensures volatility remains positive without restricting any of the parameters, allowing easier parameter estimation and flexibility in variance.

Consider specifically an EGARCH(1,1) model. Equation (4.18) is implemented in the EViews 3.1 time series package as

$$\ln(\sigma_t^2) = \omega + \alpha \left(\left| \frac{a_{t-1}}{\sigma_{t-1}} \right| - \sqrt{\frac{2}{\pi}} \right) + \gamma \frac{a_{t-1}}{\sigma_{t-1}} + \beta \ln(\sigma_{t-1}^2) \quad (4.19)$$

where ω is a constant and α , γ and β are model parameters. The EGARCH model aims to model the effect of leverage in financial time series via the inclusion of the γ parameter, whereas the GARCH formulation does not. Equation (4.19) allows for direct testing for the existence of the leverage effect in the innovation series using the hypotheses

$$\begin{aligned} H_o &: \gamma = 0 \\ H_a &: \gamma \neq 0. \end{aligned} \quad (4.20)$$

Estimation of the EGARCH(p,q) model in Equation (4.18) is similar to that of the GARCH(m,s) model. The likelihood function is slightly more complicated, since the error term appears within the function $g(\epsilon_t)$. The unknown parameters are non-linear with respect to the likelihood function and finding initial values for the optimization is also not straightforward. The EGARCH(m,s) model is less parsimonious than the GARCH model, since it contains two extra parameters, θ and γ .

Forecasting for an EGARCH model one time step into the future is relatively simple. Consider an EGARCH(0,1) model, with $\ln(\sigma_t^2) = \alpha_0 + \alpha_1 g(\epsilon_{t-1})$ for $t = 1, \dots, T$. The conditional variance is given by $\sigma_T^2 = \exp(\alpha_0) \exp[\alpha_1 g(\epsilon_{T-1})]$ at time T , and

$$\sigma_{T+1}^2 = \exp(\alpha_0) \exp[\alpha_1 g(\epsilon_T)]$$

at time $T+1$, where $\alpha_0, \alpha_1, \theta$ and γ are estimated from the data. The value for ϵ_T is estimated by the ratio a_T/σ_T , where a_T is observed and σ_T is estimated. The expression for a forecast two steps into the future is $\sigma_{T+2}^2 = \exp(\alpha_0) \exp[\alpha_1 g(\epsilon_{T+1})]$, which is complicated to evaluate, since the expectation depends on the underlying distribution used to estimate the EGARCH model [3]. Only one-step-ahead forecasts are considered for portfolio construction in this thesis.

Chapter 5

Application of Univariate Volatility Models

Relatively small stock portfolios (fewer than ten stocks) will be analyzed in this thesis. This is because most univariate volatility models are very parameter heavy, so if too many stocks are considered, ordinary optimization techniques may either fail to provide good parameter estimates or, in some cases, fail to converge.

5.1 Selection of time series length

In terms of quantity of data selected, there is a trade-off between too much data, where the assumption of stationarity can be violated since the expected value or mean of the time series might not remain constant throughout the sample, and too small a sample, resulting in poor estimates of relevant parameters.

GARCH models are built on conditional likelihoods, so the length of the time series selected directly influences parameter estimates. Univariate GARCH and IGARCH series were therefore simulated over various time frames to estimate the optimal amount of weekly data. Matlab routines were used to simulate GARCH(1,1) and IGARCH(1,1) series of any desired length and parameter set. The Matlab GARCH toolbox was used to estimate the GARCH parameters from the simulated series. As Matlab and EViews 3.1 lack built-in IGARCH estimation functions, code was written to estimate the parameters of an IGARCH(1,1) series with a constant.

The following simulation experiment was run to estimate the optimal amount of weekly data. One hundred GARCH(1,1) series of size fifty-two were simulated to represent one year of weekly data, a process repeated for sample sizes up to eight times fifty-two, representing eight years of weekly data. The simulated

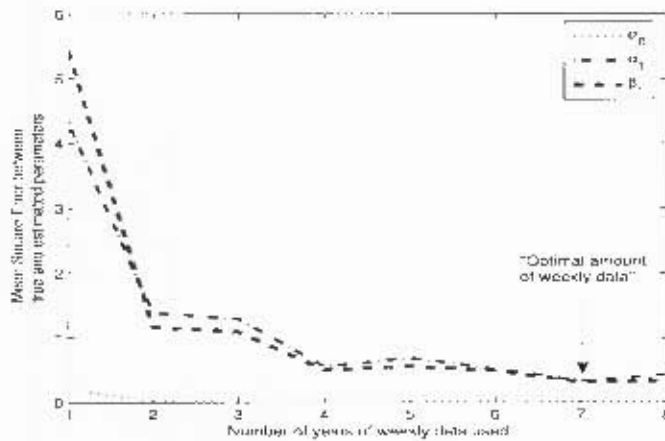


Figure 5.1: Simulation experiment for GARCH(1,1) model

GARCH series all shared a parameter set ($\alpha_0 = 0.001, \alpha_1 = 0.15, \beta_1 = 0.8$), so the true parameters were known. For each of the GARCH series generated the parameters were estimated and the squared difference between the estimated and true parameters was considered an error measure for the model.

Figure 5.1 illustrates this error summed over the 100 simulations for different time periods. Similar results were found for the IGARCH model simulations. An eight-year period of weekly data was simulated, and the minimum mean square error of the parameters occurred using seven years of weekly data to estimate the parameters of the GARCH model. This required a series of approximately 350 observations, which is longer than the recommended length for a GARCH parameter optimization, namely approximately 250 observations.

5.2 Portfolio selection

Two stock portfolios, hereafter referred to as Portfolio 1 and Portfolio 2, were selected for analysis in order to compare the performance of new multivariate models with that of established univariate volatility models. The ability of multivariate volatility models to provide additional information on the risk of a stock due to relationships with other stocks becomes more useful the higher the covariance among different stocks.

The All Share Index (ALSI), an index of the top forty shares based on market capitalization on the JSE, was chosen for modeling the market in this thesis. The stocks in the index are weighted by total market capitalization. The seven years of actual ALSI data (367 closing price values) from 31 January 1997 to 6

	ABSA	EDCON	GRINROD	GROUP5	JDGROUP
ABSA	1.0000	0.1581	0.0392	0.1399	0.4306
EDCON	0.1581	1.0000	0.0689	0.0985	0.1955
GRINROD	0.0392	0.0689	1.0000	0.1160	0.1234
GROUP5	0.1399	0.0985	0.1160	1.0000	0.1377
JDGROUP	0.4306	0.1955	0.1234	0.1377	1.0000

Table 5.1: Correlation between the log-returns of the five stocks in Portfolio 1 (df=366)

February 2004 (31/01/97 - 06/02/04) for all the stocks in Portfolio 1 and Portfolio 2 and the market index was used. This time series incorporates numerous troughs and peaks, allowing the volatility models to be compared on a data set with both upside and downside risk.

If stocks are thinly traded, the price series changes infrequently, leading to poor parameter estimates especially for regression coefficients, since the repeated values erroneously carry more weight in a statistical sense and force the model to pass through them. If the square of the return is used as an estimate for variance, zero variance could be encountered, implying the invalid situation of zero risk. In this thesis, the potential problem of thin trading was addressed by selecting return series of large companies with high trading volumes. Low frequency data (weekly instead of daily) were used to further prevent the thin trading phenomenon. The two portfolios selected for this thesis are composed as follows, with abbreviation and sector bracketed:

Portfolio 1: ABSA (ABSA; banking), Edgars Consolidated Stores (EDCON; retail), Grinrod (GRINROD; transport), Group Five (GROUP5; construction), JD Group (JDGROUP; retail), SAB-Miller (SAB; beverages) and Shoprite Holdings (SHPRITE; retail). These stocks compete in different sectors of the economy and so are influenced by different economic variables. This means that the stock log-returns are weakly related to each other and consequently have low average correlations with each other over the chosen time period, as illustrated for five of these stocks (due to limited space) in Table 5.1.

Portfolio 2: Anglo American (ANGLO; gold), Durban Roodepoort Deep (DRD; gold), Goldfields (GFIELDSD; gold), Harmony (HARMONY; gold) and Implats (IMPLATS; platinum). This portfolio consists entirely of mining stocks in very large and actively traded companies. Being in the same industry, the stocks should move under similar economic pressures, and the resulting high average correlations between the stock log-returns are shown in Table 5.2.

	ANGLO	DRD	GFIELDS	HARMONY	IMPLATS
ANGLO	1.0000	0.5417	0.6433	0.7102	0.3863
DRD	0.5417	1.0000	0.5999	0.6509	0.2107
GFIELDS	0.6433	0.5999	1.0000	0.7200	0.2559
HARMONY	0.7102	0.6509	0.7200	1.0000	0.3269
IMPLATS	0.3863	0.2107	0.2559	0.3269	1.0000

Table 5.2: Correlation between the log returns of five stocks in Portfolio 2 (df=366)

5.3 Model fitting

The GARCH, IGARCH and EGARCH univariate volatility models and respective mean equations will be estimated for Portfolios 1 and 2.

The conditional mean μ_t for all the univariate volatility models is described by a combination of the Single Index Model and an autoregressive moving average (ARMA) model described in Equation (4.2) as

$$\mu_t = A + Br_{mt} + \sum_{i=1}^p \phi_i e_{t-i} - \sum_{j=1}^q \theta_j a_{t-j}, \quad (5.1)$$

where A and B (alpha and beta coefficients) are parameters from the Single Index Model and ϕ_1, \dots, ϕ_p and $\theta_1, \dots, \theta_q$ are the parameters from the ARMA(p,q) model. The conditional variance formulation estimated for the univariate models is described by

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

for the GARCH(1,1) model, a special case of Equation (4.3), with parameters α_0, α_1 and β_1 . The conditional variance formulation for the IGARCH(1,1) model, a special case of Equation (4.14), is described by

$$\sigma_t^2 = \alpha_0 + (1 - \alpha_1) a_{t-1}^2 + \alpha_1 \sigma_{t-1}^2$$

with parameters α_0 and α_1 . Lastly, the conditional variance formulation for the EGARCH(1,1) model from Equation (4.19), a special case of Equation (4.18), is described by

$$\ln(\sigma_t^2) = \omega + \alpha \left(\left| \frac{a_{t-1}}{\sigma_{t-1}} \right| - \sqrt{\frac{2}{\pi}} \right) + \gamma \frac{a_{t-1}}{\sigma_{t-1}} + \beta \ln(\sigma_{t-1}^2)$$

with parameters ω, α, γ and β . For all the above conditional variance models, the conditional variance is defined as σ_t^2 and the innovation is defined as a_t for time t .

The GARCH(1,1) and EGARCH(1,1) volatility models above were estimated using simultaneous parameter estimation methodology described in Section 4.2.1. The conditional mean and conditional variance parameters were therefore estimated under a single log-likelihood function using the EViews 3.1 time series package. The IGARCH(1,1) model is estimated using two-tier parameter estimation methodology. The parameters for the conditional mean were estimated with a separate log-likelihood function, and the conditional variance was estimated in a separate likelihood function using the estimated conditional mean values $\hat{\mu}_t$.

The IGARCH(1,1) specification is not available in EViews 3.1 or the GARCH toolbox in Matlab. For this reason an IGARCH(1,1) estimation program was written in Matlab, in order to fit the IGARCH(1,1) model in this thesis. The standard errors of the IGARCH(1,1) parameters were not calculated, since the IGARCH(1,1) model was only implemented to test one-step-ahead forecasts and volatility persistence.

The modelling methodology employed began with choosing appropriate mean equations using the sample autocorrelation function (ACF) and partial autocorrelation function (PACF) of the relevant series in EViews 3.1 to determine the ARMA(p,q) order. The Portmanteau statistic [3] was then used to test for non-significant autocorrelation in the residual series, with a condition of close-to-zero autocorrelation indicating estimated mean equations to be considered sufficient to assume weak stationarity for the resultant time series.

The GARCH, IGARCH and EGARCH volatility models were fitted to the innovation time series a_t for all the stocks in each of Portfolios 1 and 2. Only lower-order GARCH and EGARCH models are commonly used in practice [3]. In the process of model order selection, if the sample ACF for the squared residuals a_t^2 has no significant lags, then the fitted GARCH order is considered correct. If the ACF displays significantly large values at certain lags, a higher-order model was fitted, a process repeated until no significant auto-correlations are present in the squared innovations a_t^2 . Model order was not determined for the IGARCH model, since the IGARCH(1,1) model was fitted as is.

When a parameter was encountered that was non-significant at the five percent level of significance, it was removed from the model and the parameters were subsequently re-estimated with the new parameter set.

Within a particular sample, statistical models are usually compared using the Akaike Information Criterion (AIC), or some other likelihood-based measure of fit. This thesis does not compare within-sample performance of models. Rather, it aims to compare different models solely based on forecasting ability. A sum

ABSA	$0.886(0.057)ALSI_t - 0.104(0.047)ABSA_{t-1} + a_t$
EDCON	$0.266(0.071)ALSI_t + 0.756(0.091)EDCON_{t-1} + a_t - 0.610(0.131)a_{t-1}$
GRINROD	$0.289(0.098)ALSI_t + 0.552(0.155)GRINROD_{t-1} + a_t - 0.674(0.144)a_{t-1}$
GROUP5	$0.244(0.118)ALSI_t + a_t$
JDGROUP	$0.760(0.078)ALSI_t + a_t$
LIBERTY	$0.827(0.049)ALSI_t + a_t - 0.150(0.058)a_{t-1}$
SABREW	$0.824(0.047)ALSI_t + 0.146(0.057)SABREW_{t-1} - 0.070(0.035)SABREW_{t-2} + a_t - 0.284(0.075)a_{t-1}$
SHPRITE	$0.526(0.068)ALSI_t + a_t$

Table 5.3: Mean equations for Portfolio 1 using GARCH specification for residuals (standard errors in brackets)

of squares forecasting error metric to compare model performance and select the most effective univariate model for the stocks of Portfolios 1 and 2 will be introduced later in this thesis.

5.4 Volatility model specification

5.4.1 GARCH

GARCH mean equations

The mean equations for the stocks in Portfolio 1 using simultaneous parameter estimation and the GARCH conditional variance formulation, are summarized in Table 5.3, where a_t is the innovation at time t .

Table 5.4 shows specific beta coefficients of Portfolio 1, where the beta coefficients refer to the linear regression of the risky asset and the ALSI market index, estimated by Equation (2.9). These values will become inputs for creation of the covariance matrix in later sections.

Table 5.5 and Table 5.6 show the estimated mean equations and beta coefficients for Portfolio 2 respectively. As with Portfolio 1 (Table 5.3), parameter optimization of each conditional mean equation for Portfolio 2 under its own likelihood function produced very similar parameter estimates to those in Table 5.5.

GARCH volatility equations

Parameter estimates and standard errors for GARCH(1,1) models fitted to the stocks in each portfolio are found in Table 5.7. The Liberty stock is not listed in the table since a GARCH(2,1) model, defined as $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \beta_1 \sigma_{t-1}^2$,

	beta coefficient
ABSA	0.886
EDCON	0.266
GRINROD	0.289
GROUP5	0.244
JDGROUP	0.760
LIBERTY	0.827
SABREW	0.824
SHPR	0.526

Table 5.4: Beta coefficients for Portfolio 1 stocks

ANGLO	$0.653(0.084)ALSI_t - 0.130(0.056)ANGLO_{t-2} + a_t$
DRD	$0.593(0.146)ALSI_t + a_t$
GFIELDS	$0.615(0.090)ALSI_t + 0.156(0.050)GFIELDS_{t-8} + a_t$
HARMONY	$0.764(0.111)ALSI_t - 0.108(0.055)HARMONY_{t-3} + a_t$
IMPLATS	$1.247(0.072)ALSI_t - 0.055(0.055)IMPLATS_{t-1} + a_t - 0.140(0.061)a_{t-2}$

Table 5.5: Mean equations for Portfolio 2 using GARCH specification for residuals (standard errors in brackets)

	beta coefficient
ANGLO	0.653
DRD	0.593
GFIELDS	0.615
HARMONY	0.764
IMPLATS	1.247

Table 5.6: Beta coefficients for the stocks of Portfolio 2

Portfolio 1	$\hat{\alpha}_0$ (10^{-4})	$\hat{\alpha}_1$	$\hat{\beta}_1$
ABSA	0.922 (0.627)	0.077 (0.031)	0.888 (0.046)
EDCON	1.990 (0.578)	0.104 (0.032)	0.852 (0.037)
GRINROD	1.110 (0.458)	0.058 (0.013)	0.926 (0.018)
GROUP5	2.490 (1.340)	0.041 (0.015)	0.917 (0.031)
JDGROUP	3.570 (2.020)	0.100 (0.050)	0.797 (0.010)
SABREW	0.086 (0.084)	0.052 (0.015)	0.944 (0.014)
SHPRITE	30.00 (1.040)	-0.040 (0.013)	-
Portfolio 2			
ANGLO	1.57 (1.120)	0.063 (0.030)	0.895 (0.043)
DRD	89.65 (5.980)	0.074 (0.050)	-
GFIELDS	2.820 (1.900)	0.053 (0.030)	0.888 (0.058)
HARMONY	4.230 (2.790)	0.039 (0.028)	0.893 (0.068)
IMPLATS	1.660 (0.881)	0.110 (0.026)	0.844 (0.036)

Table 5.7: GARCH parameter estimates (standard errors in brackets)

with estimated parameters $\hat{\alpha}_0 = 0.0000265$, $\hat{\alpha}_1 = 0.219$, $\hat{\alpha}_2 = -0.178$, $\hat{\beta}_1 = 0.944$ and parameter standard errors of 0.0000248, 0.086, 0.088, 0.032 respectively, was fitted.

Some of the fitted GARCH volatility models exhibit fairly strong persistence, as the estimated $\hat{\beta}_1$ parameter value is close to one, implying past volatility carries more weight than the reactive shock parameter α_1 . The fact that the estimated parameter sum ($\hat{\alpha}_1 + \hat{\beta}_1$) for many of the Portfolio 1 stocks equals almost one implies that the IGARCH family of models is likely to be appropriate for such time series.

Table 5.8 shows the one-step-ahead forecast of the innovation time series a_t , which is also the square root of the unique risk for the portfolios estimated by the GARCH volatility models with parameters specified in Table 5.7. Standard deviation was preferred to variance for simpler comparison among the volatility models.

5.4.2 IGARCH

IGARCH mean equations

The IGARCH model is estimated using two-tier parameter estimation. IGARCH mean equations were therefore estimated under likelihood functions separate from the conditional mean. Parameter estimates found for each conditional mean equation using particular likelihood functions in the two-tier estimation procedure closely resembled those found by simultaneous parameter estimation of GARCH

Portfolio 1	
ABSA	0.044
EDCON	0.054
GRINROD	0.057
GROUP5	0.057
JDGROUP	0.055
LIBERTY	0.0278
SABREW	0.024
SHPRITE	0.040
Portfolio 2	
ANGLO	0.047
DRD	0.095
GFIELDS	0.057
HARMONY	0.068
IMPLATS	0.046

Table 5.8: GARCH standard deviation forecasts

conditional mean equations. For this reason only the GARCH conditional mean equations displayed in Tables 5.3 and 5.5 for Portfolios 1 and 2 respectively are used as proxies for the IGARCH conditional mean equations.

IGARCH volatility equations

Table 5.9 shows parameter estimates for IGARCH(1,1) models fitted to the stocks of both portfolios. The α_1 parameter estimate from Table 5.9 in most cases is close to the persistence parameter β_1 of the fitted GARCH model in Table 5.7. IGARCH(1,1) without intercept is the RiskmetricsTM measure of risk due to its prudent nature, as demonstrated by the standard deviation forecast of the innovation series a_t in Table 5.10.

5.4.3 EGARCH

EGARCH mean equations

Mean equations for Portfolio 1, presented in Table 5.11, were estimated for the EGARCH volatility model using the simultaneous parameter estimation approach. When testing the null hypothesis (4.20), that the leverage parameter γ equals zero, it was found that the Liberty and SA Breweries stocks had no leverage effect at the 5% significance level. Since the EGARCH model purports to exploit the natural leverage effect in econometric data, the EGARCH model is over-specified for these particular stocks. As such, the Liberty and SA Breweries stocks were modelled using GARCH specifications from Table 5.7 instead.

Portfolio 1	$\hat{\alpha}_0 (10^{-4})$	$\hat{\alpha}_1$
ABSA	1.157	0.802
EDCON	2.477	0.766
GRINROD	1.678	0.872
GROUP5	41.49	0.934
JDGROUP	5.826	0.536
LIBERTY	1.464	0.667
SABREW	0.790	0.760
SHPRITE	0.651	0.885
Portfolio 2		
ANGLO	5.610	0.567
DRD	25.49	0.456
GFIELDS	2.260	0.813
HARMONY	22.38	0.239
IMPLATS	1.550	0.793

Table 5.9: IGARCH parameter estimates for volatility equations

Portfolio 1	
ABSA	0.051
EDCON	0.058
GRINROD	0.084
GROUP5	0.067
JDGROUP	0.063
LIBERTY	0.029
SABREW	0.029
SHPRITE	0.047
Portfolio 2	
ANGLO	0.048
DRD	0.075
GFIELDS	0.052
HARMONY	0.065
IMPLATS	0.046

Table 5.10: IGARCH standard deviation forecasts

ABSA	$0.884(0.058)ALSI_t - 0.128(0.048)ABSA_{t-1} + a_t$
EDCON	$0.275(0.077)ALSI_t + 0.719(0.010)EDCON_{t-1} + a_t - 0.601(0.126)a_{t-1}$
GRINROD	$0.359(0.068)ALSI_t + 0.571(0.085)GRINROD_{t-1} + 0.116(0.042)GRINROD_{t-3} + a_t - 0.740(0.082)a_{t-1}$
GROUP5	$0.278(0.099)ALSI_t + 0.575(0.192)GROUP5_{t-1} + a_t - 0.635(0.188)a_{t-1}$
JDGROUP	$0.737(0.079)ALSI_t + a_t$
LIBERTY	$0.827(0.049)ALSI_t + a_t - 0.150(0.058)a_{t-1}$
SABREW	$0.824(0.047)ALSI_t + 0.146(0.057)SABREW_{t-1} - 0.070(0.036)SABREW_{t-2} + a_t - 0.284(0.075)a_{t-1}$
SHPRIT	$0.568(0.075)ALSI_t + a_t$

Table 5.11: Mean equations for Portfolio 1 using EGARCH specification for residuals (standard errors in brackets)

	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\gamma}$	$\hat{\beta}$
ABSA	-0.336 (0.142)	0.104 (0.054)	-0.084 (0.035)	0.958 (0.020)
EDCON	-0.352 (0.069)	0.101 (0.041)	-0.150 (0.029)	0.950 (0.010)
GRINROD	-0.229 (0.069)	0.0691 (0.028)	-0.142 (0.034)	0.965 (0.011)
GROUP5	-9.544 (0.449)	0.082 (0.044)	0.095 (0.037)	-0.840 (0.090)
JDGROUP	-0.513 (0.167)	0.056 (0.046)	-0.164 (0.032)	0.918 (0.026)
SHPRITE	-1.014 (0.362)	-0.063 (0.047)	-0.174 (0.049)	0.824 (0.061)

Table 5.12: EGARCH parameter estimates (standard errors in brackets), $\gamma \neq 0$

None of the stocks from Portfolio 2 exhibited the EGARCH leverage effect when testing for the null hypothesis (4.20) that the leverage parameter γ equals zero at a 5% level of significance. The GARCH specifications for conditional means and volatility in Tables 5.5 and 5.7 will therefore be used for all Portfolio 2 stocks.

EGARCH volatility equations

Estimated EGARCH(1,1) coefficients for Portfolio 1 stocks, with respective standard errors, are presented in Table 5.12.

Forecast EGARCH values (or GARCH values where appropriate) for the innovation series a_t for Portfolio 1 stocks are displayed in Table 5.13.

Portfolio 1	
ABSA	0.043
EDCON	0.055
GRINROD	0.043
GROUP5	0.044
JDGROUP	0.051
LIBERTY	0.028*
SABREW	0.024*
SHPRITE	0.054

Table 5.13: EGARCH standard deviation forecasts (*GARCH model)

5.5 Creating mean and covariance forecasts

5.5.1 Conditional mean forecasts

Expected returns were estimated for the GARCH and IGARCH models by forecasting the conditional mean equations in Table 5.3 and 5.5 for Portfolios 1 and 2 respectively, and for the EGARCH model by forecasting those in Table 5.11 for Portfolio 1. Expected returns for each of the GARCH, IGARCH and EGARCH models are shown in Table 5.14.

The market return $E(r_{m,T+1})$, taken to be equal to -0.0018 , is the actual recorded weekly return for the week ending 13 February 2004 (13/02/2004) and is not estimated using any econometric modelling. This value corresponds to the recorded JSE value for the ALSI one week hence from the final value in the time series selected for this analysis (06/02/2004).

5.5.2 Conditional variance forecasts

Using equations (2.10) and (2.11), the covariance matrix of a portfolio can be expressed as

$$\Sigma_t = \begin{pmatrix} \beta_1^2 \sigma_{mt}^2 & \beta_1 \beta_2 \sigma_{mt}^2 & \cdots & \beta_1 \beta_p \sigma_{mt}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \beta_p \beta_1 \sigma_{mt}^2 & \beta_p \beta_2 \sigma_{mt}^2 & \cdots & \beta_p^2 \sigma_{mt}^2 \end{pmatrix} + \begin{pmatrix} \sigma_{\epsilon 1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{\epsilon 2}^2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma_{\epsilon p}^2 \end{pmatrix} \quad (5.2)$$

where β_i is the beta value for stock for i , σ_{mt}^2 is the market volatility at time t , and $\sigma_{\epsilon i}$ for $i = 1, \dots, p$ is the unique risk for stock $i = 1, \dots, p$. Equation (5.2) can be written in vector form as

$$\Sigma_t = \beta \beta' \sigma_{mt}^2 + D_{\epsilon t} \quad (5.3)$$

Portfolio 1	EGARCH	GARCH
ABSA	-0.00182	0.000448
EDCON	0.002237	0.002822
GRINROD	0.004249	0.000765
GROUP5	0.001307	0.000116
JDGROUP	0.000366	0.000378
LIBERTY	-0.00143	-0.00143
SABREW	-0.00259	-0.00259
SHPRITE	0.000289	0.000221
Portfolio 2		
ANGLO	-*	0.005775
DRD	-*	0.000299
GFIELDS	-*	0.0081
HARMONY	-*	-0.00383
IMPLATS	-*	0.01026

Table 5.14: Expected returns using mean equations for univariate models (*Not fitted due to insignificant leverage effect)

GARCH	$\hat{\alpha}_0$ (10^{-4}) 0.584 (0.295)	$\hat{\alpha}_1$ 0.071 (0.024)	$\hat{\beta}_1$ 0.880 (0.045)	
IGARCH	$\hat{\alpha}_0$ (10^{-4}) 1.65	$\hat{\alpha}_1$ 0.618		
EGARCH	$\hat{\omega}$ -0.401 (0.176)	$\hat{\alpha}$ 0.108 (0.049)	$\hat{\gamma}$ -0.107 (0.022)	$\hat{\beta}$ 0.953 (0.022)

Table 5.15: ALSI market index parameter estimates for univariate models (standard errors in brackets)

where β is a $p \times 1$ vector of beta values and D_{ct} is a $p \times p$ diagonal matrix of unique risks, with i^{th} diagonal elements $\sigma_{c,it}^2$ for $i = 1, \dots, p$. The portfolio variance Σ_t can be computed using the values in Table 5.4 as the vector of beta values. The squares of the standard deviation forecast values in Tables 5.8, 5.10 and 5.13 represent the the diagonal elements of matrix D_{ct} since the forecast of the innovation series a_t is the estimated unique risk at time t .

In order to estimate the portfolio covariance matrix Σ_t , expected market volatility σ_{mt}^2 must be estimated with the formula in Equation (5.3). Market volatility σ_{mt}^2 was estimated by fitting GARCH, IGARCH and EGARCH volatility models independently to the square of the log-returns of the ALSI. The volatility model parameters (with standard errors) for each model were estimated on the ALSI log-return series and presented in Table 5.15. The remarkably similar values for the one-step-ahead forecasts of the square root of market volatility $\hat{\sigma}_{m,T+1}$ (fore-

cast standard deviation) were 0.029, 0.029, and 0.026 for the GARCH, IGARCH and EGARCH volatility models respectively.

5.5.3 Efficient frontier analysis

The efficient frontier requires two inputs, namely a vector of expected stock returns and the expected covariance matrix. There is a school of thought that the higher and more to the left on the risk-return space an efficient frontier curve is, the better the model giving rise to that efficient frontier is, since higher return should be realized for a given level of risk. However, there is no means of objectively testing performance of efficient frontiers generated using various volatility models. Efficient frontier generation is merely a mathematical means of finding the correct weighting of stocks, given two inputs, and a set of inputs can always be created to elicit an efficient frontier to the north-west of the risk-return space.

It was proposed in this thesis that a preferable means of determining volatility model effectiveness is by investigating one-step-ahead forecasting performance.

5.5.4 Sum of squares forecasting error

For more accurate measurement of forecasting performance, on average, of a volatility model, it is best to estimate a number of forecast periods and then sum the forecast error. A Matlab routine was written for this process, resembling the bootstrapping methodology of Hansen and Lunde [6].

Consider a stock in a portfolio. A volatility model was fitted to a moving data set of $T - (M + 1)$ observations from the time series with a total of T observations, where $M=100$. The one-step-ahead forecast $\hat{a}_{i,T-k}$ for stock i was estimated for time $T - k$, where $k = 100, \dots, 1$. These M forecast values were then compared individually with the observed square log-returns of the stock at each appropriate time period, and a sum of squared forecasting errors was computed.

This process was repeated for each stock in the portfolio and the total sum of squared forecasting error (Total FE), almost identical to one of the the mean square forecast error measure used by Hansen and Lunde [6], was computed using

$$\sum_{i=1}^p \sum_{t=T-M+1}^T (|\hat{a}_{it}| - |a_{it}|)^2. \quad (5.4)$$

where i represents each stock within a portfolio P , \hat{a}_{it} is the forecast (predicted) square root of volatility and a_{it} is the actual observed square-root innovation for the stock. Anderson and Bollerslev [2] have shown that realized variance and \hat{a}_{it}^2 in the above Equation (5.4), is a more appropriate proxy for volatility than the

	EGARCH	GARCH	IGARCH
Total FE for Portfolio 1	0.0157	0.0164	0.0220
Total FE for Portfolio 2	-	0.0963	0.129

Table 5.16: Sum of forecast square errors for univariate models

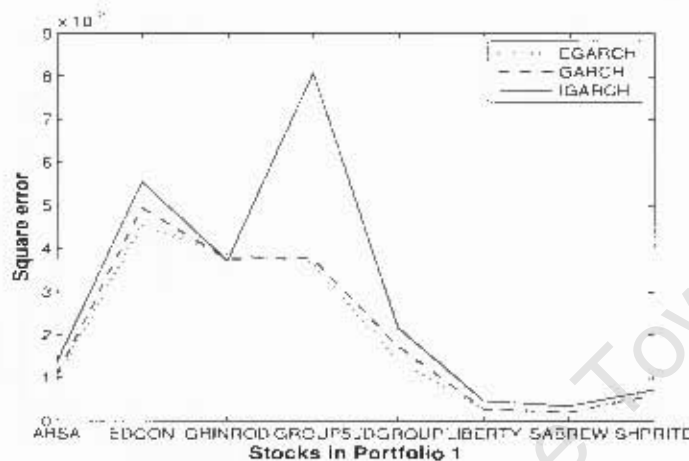


Figure 5.2: Forecast error per stock for univariate models: Portfolio 1

squared log-return r_{it}^2 .

The entire procedure was then repeated for each of the GARCH, IGARCH and EGARCH models. An identical method was followed using each stock in Portfolio 2. The Total FE criterion was used to compare unique risk forecasting ability of a volatility model.

Findings are presented in Figure 5.2, Figure 5.3, and Table 5.16. Table 5.16 shows that, for Portfolio 1, the EGARCH model has the lowest Total FE, followed by GARCH and then IGARCH, implying that the EGARCH was the best predictive model. This finding echoes that of Hansen and Lunde [6] that the one-step-ahead forecasting ability of the GARCH(1,1) model was outperformed on average by that of volatility models incorporating a leverage effect. Figure 5.2 shows that the predictive errors for EGARCH and GARCH models track each other closely, while those for the IGARCH model are much larger, specifically for the GROUP5 stock.

Figure 5.3 and Table 5.16 indicate that GARCH again outperforms the IGARCH in terms of volatility forecasting for Portfolio 2. According to the Total FE selection methodology, the EGARCH univariate volatility model performed most effectively for the stocks of Portfolio 1, while for Portfolio 2 the GARCH model

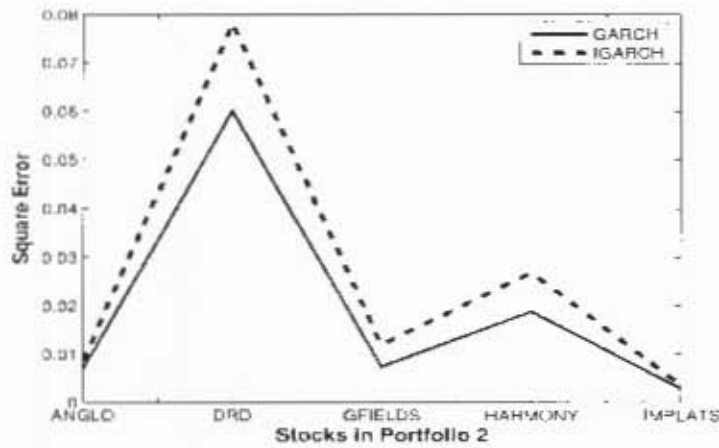


Figure 5.3: Forecast error per stock for univariate models: Portfolio 2

(in the absence of the EGARCH due to insignificant leverage effects) gave the best predictions. However, given that in Portfolio 1 the Liberty and SA Breweries stocks were fitted using only the GARCH model, it seems that a mixture of volatility models may provide the most accurate univariate volatility forecasts. In practice, time constraints limit the variety of models that can be fitted, and therefore the GARCH(1,1) model seems to be the best trade-off, since it provides very good forecasts and in most cases is easy to fit to the available data.

where A_0 is a lower triangular matrix and A_1, \dots, A_m and B_1, \dots, B_s are $p \times p$ matrices of unknown parameters. For this formulation, the matrix Σ_t in the BEKK model is almost surely positive definite provided that $A_0 A_0'$ is positive definite. The BEKK model has attractive properties as a linear system [25] and allows for dynamic dependence between volatility series, but some disadvantages prohibit practical, everyday application. The number of parameters to be estimated is $p^2(m+s) + \frac{p(p+1)}{2}$, where p is the number of assets. The All Share Index (ALSI) on the Johannesburg Stock Exchange (JSE) consists of 40 stocks, so that 4020 parameters have to be estimated for the BEKK(1,1) model. Many estimated parameters in the full model are not significantly different from zero and estimation is time-consuming, a problem minimized by assuming the matrices A_i and B_j for $i = 1, \dots, m$ and $j = 1, \dots, s$ are diagonal or even by reducing these matrices to scalars [4]. A further problem with the model is that the matrices A_i and B_j for $i = 1, \dots, m$ and $j = 1, \dots, s$ have no direct econometric interpretation.

6.2 Constant Conditional Correlation

Bollerslev assumed time-invariant correlation and formalized the Constant Conditional Correlation (CCC) model or, in other words, by holding the conditional correlations constant, the conditional covariances are proportional to the product of the corresponding conditional standard deviations, thereby substantially reducing the number of unknown parameters to be estimated [26]. This type of modelling appears similar to univariate GARCH modelling, since only the variances are modelled, but a subtle difference in model specification adds more econometric information. Consider

$$\Sigma_t = D_t C D_t$$

where C is a matrix of constant correlations, D_t is a diagonal matrix consisting of elements $\sqrt{\sigma_{ii,t}}$ for $i = 1, \dots, p$ where $\sigma_{ij,t}$ are elements from the covariance matrix Σ_t for $i, j = 1, \dots, p$ at time t . Define vector $\mathbf{v}_t = (v_{1t}, \dots, v_{pt})'$ where elements $v_{it} = \sigma_{ii,t}$ for $i = 1, \dots, p$. Then the CCC(1,1) model can be specified as

$$\mathbf{v}_t = \mathbf{a}_0 + A_1(\mathbf{a}_{t-1} \odot \mathbf{a}_{t-1}) + B_1 \mathbf{v}_{t-1} \quad (6.3)$$

where \mathbf{a}_0 is a $p \times 1$ dimensional vector, A_1 and B_1 are $p \times p$ dimensional matrices of parameters with strictly positive elements and \mathbf{v}_t is a vector of standard deviations. The more general CCC(m,s) model can be written as

$$\mathbf{v}_t = \mathbf{a}_0 + \sum_{i=1}^m A_i(\mathbf{a}_{t-i} \odot \mathbf{a}_{t-i}) + \sum_{j=1}^s B_j \mathbf{v}_{t-j}.$$

Note that by defining $\mathbf{w}_t = \mathbf{a}_t \odot \mathbf{a}_t - \mathbf{v}_t$, Equation (6.3) can be rewritten as a bivariate ARMA(1,1) model as follows [3]:

$$\mathbf{a}_t \odot \mathbf{a}_t = \mathbf{a}_0 + (A_1 + B_1)(\mathbf{a}_{t-1} \odot \mathbf{a}_{t-1}) + \mathbf{w}_t - B_1 \mathbf{w}_{t-1}.$$

Some properties of the CCC family of models are therefore directly available from the theory of the Vector ARMA (VARMA) models.

To highlight some of the salient features of the CCC modelling approach, a 2-stock case will be considered for a CCC(1,1) model. Specifically,

$$\begin{bmatrix} \sigma_{11,t} \\ \sigma_{22,t} \end{bmatrix} = \begin{bmatrix} \alpha_{10} \\ \alpha_{20} \end{bmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} a_{1,t-1}^2 \\ a_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} \sigma_{11,t-1} \\ \sigma_{22,t-1} \end{bmatrix} \quad (6.4)$$

The main interest in the CCC model lies in the interpretation of the A_i and B_j parameter matrices. If only the diagonal elements of these parameter matrices are non-zero, then the volatility series are clearly uncoupled, which implies that the different volatility series have no observable impact on each other. Now consider the off-diagonal elements of the A and B parameter matrices in Equation (6.4). If $\alpha_{12} = \beta_{12} = 0$ and $\alpha_{21} > 0, \beta_{21} > 0$ or $\alpha_{21} = \beta_{21} = 0$ and $\alpha_{12} > 0, \beta_{12} > 0$, then a unidirectional relationship is observed between stock 1 and stock 2 [3]. For example, the volatility of a small company will depend on the volatility of the market, but the market will not depend on the volatility of the small company. When all the elements of the parameter matrices are positive, a feedback relationship is observed which quantifies the extent to which different volatility series depend on each other.

6.3 Time-varying correlation models

A disadvantage of the CCC model is that the correlation matrix is assumed to be constant over the entire sample period, which seems unrealistic in empirical application.

Exact GARCH-like formulations can be used to describe conditional correlation, according to Tsay [3]. The approach models every element of the correlation matrix, namely $\rho_{ij,t}$ for $i \neq j = 1, \dots, p$ using a GARCH-like formulation, separately over time $t = 1, \dots, T$ and then re-assembles the correlation matrix element by element. This approach is problematic, since positive definiteness of the assembled correlation matrix is assured for only the $p = 2$ case.

The simplest time-varying correlation matrix approach would be the exponential smoothing technique specified by RiskmetricsTM [27], where the covariance matrix is modelled as

$$\Sigma_t = (1 - \lambda)\mathbf{a}_{t-1}\mathbf{a}'_{t-1} + \lambda\Sigma_{t-1} \quad (6.5)$$

where λ is a scalar constant. The value of λ can be optimized by minimizing forecast square error. Engle [4] proposed a generalization of the exponential smoothing methodology together with Bollerslev's CCC model, called the Dynamic Conditional Correlation (DCC) model, with Engle and Shephard developing the theoretical and empirical properties of this modified approach in a discussion paper [28].

The similarity between all the multivariate models discussed in this chapter is that the covariance matrices are modelled directly. The next chapter explores a different class of volatility models, where a transformation is applied to the covariance matrix in order to better and more parsimoniously model the dynamic relationships between assets.

University of Cape Town

Chapter 6

Multivariate Volatility Model Theory

In more recent decades, it has been established that financial volatilities vary together over time across assets and markets, an acknowledgment which improves decision-making with regards to asset and option pricing, portfolio selection, hedging, risk management, and numerous other financial areas [5].

Multivariate volatility models tend to be very parameter heavy, leading to flat likelihood functions and consequently parameter estimates which are inaccurate if available at all. Another problem is that the time varying covariance matrices have to be positive definite, so every eigenvalue of the covariance matrix must also be positive. Models where positive definiteness is not guaranteed require this condition to be checked at every step in the optimization to obtain parameter estimates. This means that multivariate volatility models have numerous implementation problems. Multivariate volatility modelling aims to provide a formulation of the covariance matrix, ensuring its positivity, in a compromise between realistic flexibility and parsimony [5].

A background of multivariate models is presented here to show the evolution to more parsimonious models. Thereafter, in the next chapter, attention will shift toward models using matrix decomposition methods to model a covariance matrix. The matrix decomposition methods are promising in accurately modelling risk in real time, but there is much research to be performed prior to suggesting widespread implementation. The application chapters of this thesis will therefore focus on decomposition models.

6.1 The BEW and BEKK model generalizations

The first multivariate extension of univariate time series models was the BEW model, proposed by Bollerslev, Engle and Woolridge in 1988 [22]. It is a basic extension of the exponentially weighted moving average methodology [3]. Consider the model for log returns written as

$$\mathbf{r}_t = \boldsymbol{\mu}_t + \mathbf{a}_t$$

where $\mathbf{r}_t = (r_{1t}, \dots, r_{pt})'$ is a vector of log returns at time t , $\boldsymbol{\mu}_t = (\mu_{1t}, \dots, \mu_{pt})'$ is a vector of conditional means at time t where μ_{it} is defined in Equation (2.17) and $\mathbf{a}_t = (a_{1t}, \dots, a_{pt})'$ is a vector of innovations at time t . Specify $\mathbf{a}_t | F_{t-1} \sim N(0, \Sigma_t)$, where F_{t-1} is the set of available information at time $t - 1$.

The BEW(m,s) model can now be defined as

$$\Sigma_t = A_0 + \sum_{i=1}^m A_i \odot (\mathbf{a}_{t-i} \mathbf{a}'_{t-i}) + \sum_{j=1}^s B_j \odot \Sigma_{t-j} \quad (6.1)$$

where m and s are integers, A_0, A_1, \dots, A_m and B_1, \dots, B_s are symmetric matrices and \odot denotes the Hadamard product, which is element by element matrix multiplication. For computation, Equation (6.1) can be rewritten in *vech* notation form, which involves stacking the upper triangular elements of a $p \times p$ symmetric matrix into a $\frac{p(p+1)}{2} \times 1$ vector, as follows:

$$vech(\Sigma_t) = vech(A_0) + \sum_{i=1}^m A_i^* vech(\mathbf{a}_{t-i} \mathbf{a}'_{t-i}) + \sum_{j=1}^s B_j^* vech(\Sigma_{t-j})$$

where $vech(A_0)$ is a $\frac{p(p+1)}{2} \times 1$ vector and A_1^*, \dots, A_m^* and B_1^*, \dots, B_s^* are $\frac{p(p+1)}{2} \times \frac{p(p+1)}{2}$ matrices [23].

A major problem with the BEW formulation is the failure to ensure a positive definite covariance matrix since negative eigenvalues of that matrix are still possible, even if all elements of the covariance matrix are positive. The variance matrix associated with the BEW model cannot be written in the form $\Sigma = H'H$, for some matrix H with full row rank; therefore the matrix is not assured to be positive definite.

In order to guarantee positive definiteness of the time-dependent covariance matrices, Engle and Kroner suggested the BEKK model in 1995 [24]. The BEKK(m,s) model is written as [3]

$$\Sigma_t = A_0 A_0' + \sum_{i=1}^m A_i (\mathbf{a}_{t-i} \mathbf{a}'_{t-i}) A_i' + \sum_{j=1}^s B_j \Sigma_{t-j} B_j' \quad (6.2)$$

Chapter 7

Matrix Decomposition Methods in Multivariate Volatility Models

Certain re-parameterizations of the positive definite covariance matrix are useful in the study of volatility, since the symmetry of covariance matrices can provide additional econometric information. Two multivariate volatility models will be discussed, using the spectral and the Cholesky decomposition.

7.1 Principal component analysis

Principal component analysis (PCA) takes a p -dimensional random variable $\mathbf{u} = (u_1, \dots, u_p)'$ and its covariance matrix Σ to find uncorrelated linear combinations $y_i = \mathbf{c}'_i \mathbf{u}$ that maximize variance $\text{var}(y_i)$ subject to the constraints $\mathbf{c}'_i \mathbf{c}_i = 1$ and $\mathbf{c}'_i \Sigma \mathbf{c}_j = 0$ for $j = 1, \dots, i - 1$ [3].

A positive definite covariance matrix Σ has a spectral decomposition, and so can be written as

$$\Sigma = H'DH \quad (7.1)$$

where H is a $p \times p$ orthogonal matrix with columns comprising p eigenvectors of Σ such that $H = (\mathbf{h}_1, \dots, \mathbf{h}_p)$, and D is a diagonal matrix consisting of p eigenvalues $\lambda_1, \dots, \lambda_p$ of Σ . The eigenvalues can be ordered as $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ to show that the i^{th} principal component of \mathbf{u} is given by $y_i = \mathbf{h}'_i \mathbf{u}$ for $i = 1, \dots, p$ [3]. Also, the variance of the i^{th} principal component is $\text{var}(y_i) = \mathbf{h}'_i \Sigma \mathbf{h}_i = \lambda_i$ for $i = 1, \dots, p$.

Consequently, $\sum_{i=1}^p \text{var}(u_i) = \sum_{i=1}^p \lambda_i = \sum_{i=1}^p \text{var}(y_i)$, and hence the ratio

$$\frac{\text{var}(y_i)}{\sum_{i=1}^p \text{var}(y_i)} = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \quad (7.2)$$

reflects the proportion of total variance in vector \mathbf{u} explained by the i^{th} principal component.

In practice PCA is applied to either the observed covariance matrix $\widehat{\Sigma}$ or the correlation matrix $\widehat{R} = S^{-1}\widehat{\Sigma}S^{-1}$, where S is a diagonal matrix of standard deviations. These are estimated from observed real-world data. In practice the sample correlation matrix may be preferred to the sample covariance matrix to avoid one random variable unduly influencing the principal components [29], since this is tantamount to standardization of the data to ensure that realizations of random variables are compared in relative rather than absolute terms.

Practically, when the observed covariance matrix $\widehat{\Sigma}$ or the observed correlation matrix \widehat{R} are invoked, the result in Equation (7.2) can be used for dimension reduction, since a principal component can be removed if a negligible percentage variance is explained by that component.

7.2 Orthogonal-GARCH model

Spectral decomposition models comprise the Orthogonal-GARCH (O-GARCH) volatility models. A general statistical population model for the O-GARCH class will be proposed, and then the practical implications for three application techniques highlighted, namely the basic O-GARCH, the Alexander O-GARCH [30], and the Modified O-GARCH volatility models. Each approach uses different estimates for the orthogonal matrix H , which has practical implications.

O-GARCH modelling of multivariate financial time series uses the orthogonality, or lack of correlation, and the dimension-reduction properties of a small number of principal components compared to the number of assets [31].

The population model underlying the O-GARCH procedure is specified as

$$\mathbf{a}_t|F_{t-1} \sim N(\mathbf{0}, \Sigma_t)$$

where $\mathbf{a}_t = (a_{1t}, \dots, a_{pt})'$ is the vector of innovations at time t , $\Sigma_t = HD_tH'$, with matrix H orthogonal and matrix D_t diagonal. Then the vector

$$\mathbf{q}_t = H'\mathbf{a}_t|F_{t-1} \sim N(0, D_t) \quad (7.3)$$

can be defined by F_{t-1} , the information set up to and including time $t - 1$. Now $\mathbf{q}_t = (q_{1t}, \dots, q_{pt})'$ is the vector of principal component scores at time t . The O-GARCH model can be regarded as a factor model $\mathbf{a}_t = H\mathbf{q}_t$ with $\mathbf{q}_t|F_{t-1} \sim N(0, D_t)$ and zero idiosyncratic variance. Equation (7.3) is an orthogonal transformation of the innovations \mathbf{a}_t , so the resultant innovations are now

orthogonal and uncorrelated.

Univariate volatility models, such as the GARCH(1,1) model, can be applied to the principal component scores q_{it} since the time series q_{it} for $i = 1, \dots, p$ are independent. This GARCH(1,1) model is written as

$$\lambda_{it} = \alpha_{0i} + \alpha_{1i}q_{i,t-1}^2 + \beta_{1i}\lambda_{i,t-1}, \quad (7.4)$$

where λ_{it} is the i^{th} diagonal element of the matrix D_t , for $i = 1, \dots, p$.

7.2.1 Basic O-GARCH volatility model

The population model in Equations (7.3) and (7.4) can be applied to compute a spectral decomposition of the overall sample covariance matrix. Define a $T \times p$ data matrix

$$A = (\mathbf{a}_1, \dots, \mathbf{a}_T)' \quad (7.5)$$

as a matrix of innovations, where vectors $\mathbf{a}_t = (a_{1t}, a_{2t}, \dots, a_{pt})'$ are observed for $t = 1, \dots, T$. The sample covariance matrix $\hat{\Sigma}$ and the corresponding spectral analysis of the innovations can be expressed as

$$\text{cov}(A) = \hat{\Sigma} = (H^\circ)'(D^\circ)(H^\circ) \quad (7.6)$$

where H° is the orthogonal $p \times p$ matrix of eigenvectors and D° is a $p \times p$ diagonal matrix of eigenvalues of the sample covariance matrix $\hat{\Sigma}$.

Using the orthogonal matrix H° , Equation (7.3) can be rewritten for basic O-GARCH as

$$\mathbf{v}_t = (H^\circ)' \mathbf{a}_t | F_{t-1} \sim N(0, D^\circ) \quad (7.7)$$

where vectors $\mathbf{v}_t = (v_{1t}, \dots, v_{pt})'$ represent the principal component scores at time t , for $t = 1, \dots, T$.

In essence, the general O-GARCH formulation is equivalent to estimating GARCH(1,1) models using the principal component scores of innovations instead of the innovations themselves as in univariate volatility models. Hence, a GARCH(1,1) model is fitted to each of the series v_{it} for $i = 1, \dots, p$ over time $t = 1, \dots, T$ as

$$\lambda_{it} = \alpha_{0i} + \alpha_{1i}v_{i,t-1}^2 + \beta_{1i}\lambda_{i,t-1} \quad (7.8)$$

where $\theta_i = \{\alpha_{0i}, \alpha_{1i}, \beta_{1i}\}$ is a set of positive GARCH parameters for $i = 1, \dots, p$ and λ_{it} is the i^{th} diagonal element of the matrix D_t° for $t = 1, \dots, T$. The initial value of matrix D_1° for the GARCH(1,1) processes can be specified as

$$d_{ii,1} = \frac{\sum_{t=1}^T (v_{it} - \bar{v}_i)^2}{T-1} \text{ where } \bar{v}_i = \frac{\sum_{t=1}^T v_{it}}{T} \text{ for } i = 1, \dots, p.$$

Time-varying matrices D_t° can be constructed recursively for $t = 2, \dots, T$ and $i = 1, \dots, p$ from the parameter estimates.

$$\hat{\lambda}_{it} = \hat{\alpha}_{0i} + \hat{\alpha}_{1i} v_{i,t-1}^2 + \hat{\beta}_{1i} \hat{\lambda}_{i,t-1}. \quad (7.9)$$

The conditional covariance matrix $\hat{\Sigma}_t$ of \mathbf{a}_t can be estimated by $(H^\circ)'(\widehat{D}_t^\circ)(H^\circ)$ where \widehat{D}_t° represents a diagonal matrix comprising estimated conditional variances $\hat{\lambda}_{it}$ of the GARCH(1,1) processes for $t = 1, \dots, T$.

An estimated forecast for the covariance matrix is therefore $\hat{\Sigma}_{T+1} = \hat{\Sigma}_T(1) = (H^\circ)'(\widehat{D}_T^\circ(1))(H^\circ)$ where $\widehat{D}_T(1)$ is the forecast of Equation (7.8) using the GARCH(1,1) forecasting methodology in Section 4.2.4 for $i = 1, \dots, p$.

7.2.2 Alexander O-GARCH volatility model

In the original approach formulated by Alexander [31], data are first standardized, implying that the PCA is performed on the sample correlation matrix rather than the sample covariance matrix $\widehat{\Sigma}$. The series x_{it} is defined as

$$x_{it} = \frac{a_{it} - \bar{a}_i}{s_i} \quad (7.10)$$

where a_{it} is the series of innovations of stock i at time $t = 1, \dots, T$, $\bar{a}_i = \frac{\sum_{t=1}^T a_{it}}{T}$ is the mean, and $s_i^2 = \frac{\sum_{t=1}^T (a_{it} - \bar{a}_i)^2}{T-1}$ is the sample variance of series a_{it} respectively for $i = 1, \dots, p$ and $t = 1, \dots, T$. Define the $T \times p$ matrix $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T)'$ as the matrix of standardized innovations. Also define vectors $\mathbf{x}_t = (x_{1t}, x_{2t}, \dots, x_{pt})'$ for $t = 1, \dots, T$.

The sample correlation matrix \widehat{R} and its spectral decomposition can be expressed as

$$\text{cov}(X) = \widehat{R} = (H^*)'(D^*)(H^*) \quad (7.11)$$

where H^* is the orthogonal $p \times p$ matrix of eigenvectors and D^* is a $p \times p$ diagonal matrix of eigenvalues of the sample correlation matrix \widehat{R} . Note that matrix H^* would contain different eigenvectors to those of the orthogonal matrix H° associated with $\widehat{\Sigma}$.

Using the orthogonal matrix H^* , Equation (7.3) can be rewritten for the Alexander O-GARCH model as

$$\mathbf{f}_t = (H^*)'\mathbf{x}_t | F_{t-1} \sim N(0, D^*) \quad (7.12)$$

where the vector $\mathbf{f}_t = (f_{1t}, \dots, f_{pt})$ represents the principal component scores at time t for $t = 1, \dots, T$. A GARCH(1,1) model is fitted to each series f_{it} for $i = 1, \dots, p$ over time $t = 1, \dots, T$ as

$$\lambda_{it} = \alpha_{0i} + \alpha_{1i}f_{i,t-1}^2 + \beta_{1i}\lambda_{i,t-1} \quad (7.13)$$

where $\theta_i = \{\alpha_{0i}, \alpha_{1i}, \beta_{1i}\}$ is a set of positive GARCH parameters for $i = 1, \dots, p$, and λ_{it} is the i^{th} diagonal element of the matrix D_t^* for $t = 1, \dots, T$. The initial value of matrix D_1^∇ can be specified as $\lambda_{i,1} = \frac{\sum_{t=1}^T (f_{it} - \bar{f}_i)^2}{T-1}$ for $i = 1, \dots, p$ where $\bar{f}_i = \frac{\sum_{t=1}^T f_{it}}{T}$.

From the parameter estimates

$$\hat{\lambda}_{it} = \hat{\alpha}_{0i} + \hat{\alpha}_{1i}f_{i,t-1}^2 + \hat{\beta}_{1i}\hat{\lambda}_{i,t-1}, \quad (7.14)$$

time-varying matrices D_t^∇ can be constructed recursively for $t = 2, \dots, T$ and $i = 1, \dots, p$. An estimate of the conditional correlation matrix \hat{R}_t can be constructed as $(H^*)'(\widehat{D}_t^\nabla)(H^*)$ where \widehat{D}_t^∇ represents a diagonal matrix comprising the conditional variances of the GARCH(1,1) processes for $t = 1, \dots, T$.

This means the forecast for the correlation matrix is estimated as $\hat{R}_{T+1} = \hat{R}_T(1) = (H^*)'(\widehat{D}_T^\nabla(1))(H^*)$ where $\widehat{D}_T^\nabla(1)$ is the forecast of Equation (7.14) using the GARCH(1,1) forecasting methodology from Section 4.2.4 for $i = 1, \dots, p$.

Alexander's technique [31] can, however, be criticized with regards to certain details. The first criticism is the use of standardized innovations as defined in Equation (7.10). Using correlation for PCA instead of covariance would avoid a single variable unduly influencing the principal components [29]. If one time series is much larger in absolute terms, the eigenvectors could be skewed in that direction. However, in the case of returns, all time series are in the same form, so scale ceases to be a problem and standardization is not necessary in financial application. Standardization can also cause theoretical inconsistencies, as a correlation matrix by definition has unit diagonal elements, which will not always be true with dynamic estimates of D_t .

A second concern regarding the Alexander O-GARCH model is that it provides time-varying estimates of the correlation matrix R_t , whereas most financial applications require the covariance matrix Σ_t . The covariance matrix $\hat{\Sigma}_t$ given in Alexander [31] is generated from the average sample variance as follows:

$$\hat{\Sigma} = S^{\frac{1}{2}}\hat{R}_tS^{\frac{1}{2}} \quad (7.15)$$

where S is a diagonal matrix with elements

$$s_{ii}^2 = \frac{\sum_{t=1}^T (a_{it} - \bar{a}_i)^2}{T-1}$$

and $\bar{a}_i = \frac{\sum_{t=1}^T a_{it}}{T}$ for $i = 1, \dots, p$. The dynamic diagonal estimates of \hat{R}_t create dynamic variance estimates in the covariance matrix $\hat{\Sigma}_t$ using the static variance matrix S from Equation (7.15). Dynamic covariance estimates required may be affected by using non-dynamic estimates of variance. Note, however, the theoretical inconsistency which produces the correlation matrices with diagonal elements not equal to one.

7.2.3 Modified O-GARCH volatility model

This incorporates a slight practical adaptation of the basic O-GARCH model described previously, and was developed for the purposes of this thesis. Consider the same $T \times p$ matrix $A = (\mathbf{a}_1, \dots, \mathbf{a}_T)'$ of innovations as defined for the O-GARCH model in Equation (7.5). The sample covariance matrix $\hat{\Sigma}$ and corresponding spectral analysis of the covariance matrix of the innovations are expressed in Equation (7.6) as

$$\hat{\Sigma} = (H^\circ)'(D^\circ)(H^\circ).$$

The principal component scores as defined in Equation (7.7) are

$$\mathbf{v}_t = (H^\circ)' \mathbf{a}_t | F_{t-1} \sim N(0, D^\circ). \quad (7.16)$$

The Modified O-GARCH developed in this thesis requires one further transformation procedure called whitening the data. Consider

$$\mathbf{w}_t = (D^\circ)^{-\frac{1}{2}} \mathbf{v}_t \quad (7.17)$$

where the vector $\mathbf{w}_t = (w_{1t}, \dots, w_{pt})'$ for $t = 1, \dots, T$ represents the whitened scores at time t and the matrix D° is the diagonal matrix of eigenvalues of the sample covariance matrix $\hat{\Sigma}$. This makes the $T \times P$ matrix of whitened scores

$$W = (\mathbf{w}_1, \dots, \mathbf{w}_T)' = (A)(H^\circ)(D^\circ)^{-\frac{1}{2}}. \quad (7.18)$$

A GARCH(1,1) model is fitted to each of the series of whitened scores w_{it} for $i = 1, \dots, p$ over time $t = 1, \dots, T$ as

$$\lambda_{w,it} = \alpha_{w,0i} + \alpha_{w,1i} w_{i,t-1}^2 + \beta_{w,1i} \lambda_{w,i,t-1}. \quad (7.19)$$

where $\theta_i = \{\alpha_{w,0i}, \alpha_{w,1i}, \beta_{w,1i}\}$ is a set of positive GARCH parameters for $i = 1, \dots, p$, and λ_{it} is the i^{th} diagonal element of the matrix D_i° for $t = 1, \dots, T$. The initial value matrix D_1° for the GARCH(1,1) process can be specified as $d_{ii,1} = \frac{\sum_{t=1}^T (w_{it} - \bar{w}_i)^2}{T-1}$ for $i = 1, \dots, p$ where $\bar{w}_i = \frac{\sum_{t=1}^T w_{it}}{T}$.

Where principal component scores \mathbf{v}_t orthogonally rotate the data to remove correlation between the innovation series $\mathbf{a}_1, \dots, \mathbf{a}_p$, whitening the data also removes the scale. The reason for this is that the diagonal matrix D° consists of eigenvalues $\lambda_1, \dots, \lambda_p$ which are the variances of the principal component scores. This is a scaling operation in the eigenspace, since the principal component scores are in essence divided by the standard deviations of the principal component scores.

Consider changing the matrix W in Equation (7.18) to

$$W^\circ = AHD^{-\alpha}$$

where α is some real number between zero and one. The value for α could be optimized by a cross-validation process. The α power value of $\frac{1}{2}$ was initially chosen for the analogy to the square-root relationship between variance and standard deviation. However, varying the value of α could indicate different strengths of scaling for different data sets, although this suggestion was not implemented in this thesis.

Note that whitened data is used only to improve parameter estimation and not for the reconstruction of the covariance matrix Σ_t . From the parameter estimates,

$$\hat{\lambda}_{it} = \hat{\alpha}_{w,0i} + \hat{\alpha}_{w,1i}v_{i,t-1}^2 + \hat{\beta}_{w,1i}\hat{\lambda}_{i,t-1}, \quad (7.20)$$

time-varying matrices D_t° can be constructed recursively for $t = 2, \dots, T$ and $i = 1, \dots, p$ using the principal component scores v_{it} to construct the time-varying matrices D_t° instead of the whitened data w_{it} for $i = 1, \dots, p$ and $t = 1, \dots, T$. Hence the conditional covariance matrix $\hat{\Sigma}_t$ of \mathbf{a}_t can be estimated by $(H^\circ)'(\hat{D}_t^\circ)(H^\circ)$ as in the basic O-GARCH approach.

The initial value of matrix D_1° for the GARCH(1,1) process can be specified as $d_{ii,1} = \frac{\sum_{t=1}^T (v_{it} - \bar{v}_i)^2}{T-1}$ where $\bar{v}_i = \frac{\sum_{t=1}^T v_{it}}{T}$ for $i = 1, \dots, p$, again as given in the basic O-GARCH model. The conditional covariance matrix $\hat{\Sigma}_t$ of \mathbf{a}_t can be estimated by $(H^\circ)'(\hat{D}_t^\circ)(H^\circ)$ where \hat{D}_t° represents a diagonal matrix comprising the conditional variances of the GARCH(1,1) processes for $t = 1, \dots, T$. An estimate of the covariance matrix forecast is $\hat{\Sigma}_{T+1} = \hat{\Sigma}_T(1) = (H^\circ)'(\hat{D}_T^\circ(1))(H^\circ)$, where $\hat{D}_T(1)$ is the Equation (7.14) forecast using GARCH(1,1) forecasting methodology from Section 7.8 for $i = 1, \dots, p$. This is almost identical to basic O-GARCH methods, but the diagonal matrix D_t is estimated slightly differently.

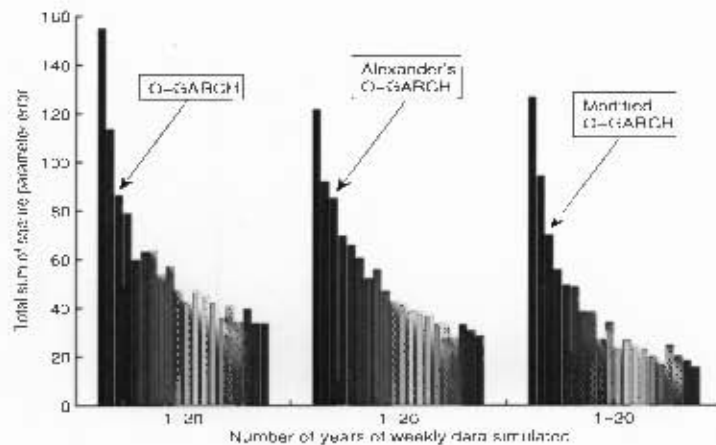


Figure 7.1: Parameter square error comparison for O-GARCH formulations using simulation

7.2.4 Parameter simulation on the O-GARCH formulations

A simulation study was performed on each of the basic, Alexander and Modified O-GARCH volatility models. Five dependent O-GARCH series were simulated with a known parameter set for different time series varying in size from one year of weekly data (52×5 observations) to twenty years of weekly data (1040×5 observations). This process was repeated a hundred times in order to generate one hundred $52i$ by 5 O-GARCH time series for $i = 1, 2, \dots, 20$.

The parameters for all O-GARCH series were estimated with each methodology. The sum of square differences between the known and estimated parameter sets, called parameter error, was computed for all i and totalled over i , the number of years of data in the particular series. Hence, the parameter error for each of the O-GARCH formulations was recorded per year of weekly data used for twenty years.

The simulation results are summarized in Figure 7.1.

Figure 7.1 indicates that the whitened data used for parameter estimation in the Modified O-GARCH model provides more accurate parameter estimates than the other two methods, which becomes increasingly clear as the number of years of data increases. The parameter error for the basic O-GARCH is consistently greater than for the Alexander O-GARCH formulation, suggesting that normalization reduces parameter error. For this reason, only the Alexander O-GARCH

and the Modified O-GARCH methods of using spectral decomposition in volatility modelling will be elaborated on later in this thesis.

7.3 Cholesky decomposition theory

The theory behind Cholesky decomposition is detailed in Tsay [3], and will be discussed briefly by considering specifically the generalized case of a volatility series of p variables.

Cholesky decomposition is the process of finding the square root of a positive definite square matrix. If a matrix Σ_t is positive definite, there exists a lower triangular matrix L_t with unit diagonal elements and a diagonal matrix G_t with positive diagonal elements such that [3]

$$\Sigma_t = L_t G_t L_t'. \quad (7.21)$$

Multiple linear regression is used only to interpret the implied orthogonality of the Cholesky decomposition. The vector $\mathbf{b}_t = (b_{1t}, \dots, b_{pt})$ is calculated after the Cholesky decomposition [3] resulting from

$$\mathbf{b}_t = L_t^{-1} \mathbf{a}_t, \quad (7.22)$$

where \mathbf{a}_t is a vector of innovations at time t . The covariance of \mathbf{b}_t is therefore

$$\begin{aligned} \text{cov}(\mathbf{b}_t) &= L_t^{-1} \text{cov}(\mathbf{a}_t \mathbf{a}_t') (L_t^{-1})' \\ &= L_t^{-1} \Sigma_t (L_t^{-1})' \\ &= G_t. \end{aligned}$$

Consider p multiple linear regressions

$$\begin{aligned} b_{1t} &= a_{1t} \\ b_{2t} &= a_{2t} - \beta_{21} b_{1t} \\ b_{3t} &= a_{3t} - \beta_{31} b_{1t} - \beta_{32} b_{2t} \\ &\vdots \\ b_{pt} &= a_{pt} - \beta_{p1} b_{1t} - \beta_{p2} b_{2t} - \dots - \beta_{p-1,t} b_{p-1,t} \end{aligned} \quad (7.23)$$

where β_{ij} are unknown regression coefficients. The elements of matrix L_t are given by $q_{ij,t} = \beta_{ij}$, the elements of matrix G_t by $g_{ii,t} = \text{var}(b_{it})$ and the orthogonality condition holds between b_{it} and b_{jt} for $i \neq j$.

7.4 Cholesky decomposition and volatility modelling

7.4.1 Basic Cholesky decomposition model

The population model underlying the basic Cholesky decomposition volatility model procedure can be specified as

$$\mathbf{a}_t | F_{t-1} \sim N(\mathbf{0}, \Sigma_t)$$

where $\mathbf{a}_t = (a_{1t}, \dots, a_{pt})'$ is the vector of innovations at time t and the covariance matrix $\Sigma_t = L_t G_t L_t'$ with L_t a lower triangular matrix and G_t a diagonal matrix at time t . Then the vector

$$\mathbf{b}_t = L_t^{-1} \mathbf{a}_t | F_{t-1} \sim N(\mathbf{0}, G_t) \quad (7.24)$$

can be defined with F_{t-1} being the set of information up to and including time $t-1$. Vector $\mathbf{b}_t = (b_{1t}, \dots, b_{pt})'$ is an orthogonal transformation of the innovations \mathbf{a}_t , making the resultant innovations orthogonal and uncorrelated.

Univariate volatility models, such as the GARCH(1,1) model, can be applied to the Cholesky scores b_{it} as

$$g_{ii,t} = \alpha_{0i} + \alpha_{1i} b_{i,t-1}^2 + \beta_{1i} g_{ii,t-1}$$

for $i = 1, \dots, p$, where $g_{ii,t}$ is the i^{th} diagonal element of the matrix G_t . Matrices L_t and G_t are modelled for $t = 1, \dots, T$, so every element σ_{ij} for $i = j = 1, \dots, p$ of the covariance matrix Σ_t is modelled over the entire sample period of length T . The number of parameters modelled in the covariance matrix Σ_t in the basic Cholesky decomposition volatility model cannot be reduced, since the matrices L_t and G_t together contain the same number of parameters as the covariance matrix Σ_t over time $t = 1, \dots, T$.

One advantage of using the Cholesky decomposition volatility models is the relatively simple resultant likelihood [3]. From the fact that $|L_t| = 1$, it can be shown that $|\Sigma_t| = |L_t G_t L_t'| = |G_t|$. If the conditional distribution of $\mathbf{a}_t | F_{t-1}$ is considered multivariate normal $N(0, \Sigma_t)$, the log-likelihood of the transformed series \mathbf{b}_t is

$$l(\mathbf{a}_t, \Sigma_t) = l(\mathbf{b}_t, G_t) = -\frac{1}{2} \sum_{i=1}^p \left\{ \ln(g_{ii,t}) + \frac{b_{it}^2}{g_{ii,t}} \right\}, \quad (7.25)$$

where the constant term is ignored and $g_{ii,t}$ is the variance of b_{it} .

Finally, an important observation is that the Cholesky decomposition differs when the covariance matrix Σ_t is specified with a different innovation ordering a_{it} for $i = 1, \dots, p$. This ordering problem affects all Cholesky decomposition volatility models due to the multiple linear regressions described in Equation (7.23).

7.4.2 Tsay Cholesky decomposition volatility models

Due to the orthogonality imposed on the transformed innovations $b_{i,t}$ for $i = 1, \dots, p$ and $t = 1, \dots, T$ by Cholesky decomposition, Tsay [3] proposes GARCH-like methodology to model the $g_{ii,t}$ parameters for $i = 1, \dots, p$ as

$$\begin{aligned}
 g_{11,t} &= \alpha_{01} + \alpha_{11}b_{1,t-1}^2 + \phi_{11}g_{11,t-1} \\
 g_{22,t} &= \alpha_{02} + \alpha_{12}b_{1,t-1}^2 + \alpha_{22}b_{2,t-1}^2 + \phi_{12}g_{11,t-1} + \phi_{22}g_{22,t-1} \\
 &\vdots \\
 g_{pp,t} &= \alpha_{0p} + \alpha_{1p}b_{1,t-1}^2 + \dots + \alpha_{pp}b_{p,t-1}^2 + \phi_{1p}g_{11,t-1} \\
 &\quad + \dots + \phi_{pp}g_{pp,t-1}
 \end{aligned} \tag{7.26}$$

where $\alpha_{01}, \dots, \alpha_{0p}, \alpha_{11}, \alpha_{12}, \dots, \alpha_{pp}$ and $\phi_{11}, \phi_{12}, \dots, \phi_{pp}$ are unknown parameters. In Equation (7.26), $g_{11,t}$ is modelled using a simple univariate GARCH(1,1), while $g_{pp,t}$ is effectively modelled as a CCC(1,1) model, with two $p \times p$ parameter matrices. No positivity constraints are imposed on the parameters, so if $g_{ii,t}$ becomes negative, the natural logarithm of $g_{ii,t}$ could be modelled instead. Tsay [3] also suggests modelling the elements of L_t , namely $q_{ij,t}$ for $j < i$ and $i = 1, \dots, p$, and $t = 1, \dots, T$ as follows:

$$\begin{aligned}
 q_{21,t} &= \gamma_{01} + \gamma_{11}q_{21,t-1} + \gamma_{21}a_{2,t-1} \\
 &\vdots \\
 q_{p,p-1,t} &= \gamma_{0,p(p-1)} + \gamma_{1,p(p-1)}q_{p,p-1,t-1} + \gamma_{2,p(p-1)}a_{p-1,t-1}
 \end{aligned} \tag{7.27}$$

where $\gamma_{01}, \dots, \gamma_{0,p(p-1)}, \gamma_{11}, \dots, \gamma_{1,p(p-1)}, \gamma_{21}, \dots, \gamma_{2,p(p-1)}$ are unknown parameters. The $q_{ij,t}$ are auto-correlated, and so a basic lag model seems appropriate. Also, the shocks a_2, \dots, a_{p-1} , which are not squared so negative values can be estimated for the elements of the matrix L_t , are added as explanatory variables.

The elements $q_{ij,t}$ for $i, j = 1, \dots, p$ of L_t in Equation (7.27) are modelled using a simple one-period lag model, so $p(p-1)$ time series of length t must be generated by performing a Cholesky decomposition at each time step. The values $b_{1,t}, \dots, b_{p,t}$ required for Equation (7.26) are obtained by combining the time-indexed L_t matrices for $t = 1, \dots, T$ and Equation (7.22).

Only once these data have been generated can the parameters in Equations (7.26) and (7.27) be estimated and used to build estimated time-varying matrices \hat{L}_t and \hat{G}_t . These in turn can be recombined into time-varying covariance matrices $\hat{\Sigma}_t$ by invoking Equation (7.21). Modelling $\ln(g_{ii,t})$ instead of $g_{ii,t}$ for $i = 1, \dots, p$ allows the parameters in the Tsay formulation to take on negative values. This ensures the covariance matrix Σ_t is positive definite with no parameter constraints. The log-likelihood function is relatively simple. Another advantage of this approach

is the manner of obtaining time-varying correlation estimates even if $q_{21,t}$ is modelled as a constant [3].

A criticism of the Tsay model is the difficulty of inference for the parameter estimates. This contention may, however, be overlooked when considering forecasting utility alone. A major disadvantage of the Tsay formulation summarized in Equations (7.26) and (7.27) is that it is not parsimonious. All parameters must be modelled using one likelihood function since some parameters depend on others in the formulation. This means that the original formulation is impractical for modelling even a few risky assets. The Tsay Cholesky decomposition volatility model will therefore not be implemented in this thesis.

7.4.3 Constant L models

As the difficulties encountered in the Tsay Cholesky decomposition volatility model result mainly from over-parameterization, a pragmatic variation would be to model the time-varying matrix L_t for $t = 1, \dots, T$ as a constant, non-time-varying matrix L . In other words, holding the conditional correlations constant means the conditional covariances are proportional to the product of the corresponding conditional standard deviations, substantially reducing the number of unknown parameters to be estimated [26]. Two approaches which employ this simplification will be explored in this thesis. The first, developed by Vrontos *et al* [32], involves estimating the matrix L using maximum likelihood. The second, developed specifically for this thesis, takes an approach resembling that of the O-GARCH model, specifying the matrix L by performing a Cholesky decomposition on the overall sample covariance matrix $\hat{\Sigma}$.

The full-factor multivariate GARCH model

This class of volatility models, first developed by Vrontos *et al* [32], also employs Cholesky decomposition. Consider

$$\mathbf{a}_t = L\mathbf{x}_t \quad (7.28)$$

where L is a $p \times p$ vector of parameters and \mathbf{x}_t is a vector of factors with elements x_{it} for $i = 1, \dots, p$. Define the conditional distribution of $\mathbf{x}_t | F_{t-1}$, where F_{t-1} is the information set up to time $t - 1$, as $N(\mathbf{0}, G_t^*)$, where G_t^* is a diagonal covariance matrix. Specifically, $G_t^* = \text{diag}(g_{1t}^2, \dots, g_{pt}^2)$ with

$$g_{it}^2 = \alpha_{0i} + \alpha_{1i}x_{i,t-1}^2 + \beta_{1i}g_{i,t-1}^2 \quad (7.29)$$

and g_{it}^2 for $i = 1, \dots, p$ is the variance of the i^{th} factor at time t for $t = 1, \dots, T$. The parameters α_{0i}, α_{1i} and β_{1i} for $i = 1, \dots, p$ are necessarily positive. This means the factors x_{it} for $i = 1, \dots, p$ follow a GARCH(1,1) model and the vector

of innovations \mathbf{a}_t is a linear combination of these factors.

The vector $\mathbf{a}_t|F_{t-1}$ is distributed as $N(\mathbf{0}, LG_t^*L' = \Sigma_t)$ [32], where

$$\Sigma_t = LG_t^*L' = (L(G_t^*)^{\frac{1}{2}})(L(G_t^*)^{\frac{1}{2}})' = (L^*)'(L^*). \quad (7.30)$$

The parameter matrix L^* can be defined as lower triangular as it is analogous to Cholesky decomposition. In practice, the parameters $\alpha_{1i} = \alpha_1$ and $\beta_{1i} = \beta_1$ for $i = 1, \dots, p$, which implies the same GARCH α_{1i} and β_{1i} parameters are used for all p factor series, thereby restricting the number of parameters in the formulation.

The likelihood function for the full-factor multivariate GARCH model can be found in Vrontos *et al* [32]. The parameter set to be estimated is $\theta = \{\alpha_{01}, \dots, \alpha_{0p}, \alpha_1, \beta_1, l_{21}, l_{31}, \dots, l_{p1}, \dots, l_{p,p-1}\}$, where $l_{21}, \dots, l_{p,p-1}$ are elements of the lower triangular matrix L . The full-factor multivariate GARCH model therefore has $\frac{p(p-1)}{2} + p + 2$ parameters to be estimated.

Despite the full-factor multivariate GARCH model of Vrontos *et al* [32] being well-developed theoretically, the large number of parameters required to be estimated renders it impractical. For example, a volatility model based on the 40 stocks in the ALSI market index, requires 862 parameters estimates using one likelihood function.

Vrontos *et al* [32] also make the simplifying assumption that every transformed innovation series x_{it} for $i = 1, \dots, p$ can be modelled using the same GARCH parameters α_1 and β_1 for reactivity and persistence in volatility. The focus of the full-factor multivariate GARCH model seems to be to model the elements of the L^* matrix to better estimate the L matrix, so that the relationships between the volatility series are modelled more accurately to the detriment of accurately modelling the individual volatility series. For the above reasons, the full-factor multivariate GARCH model of Vrontos *et al* [32] will not be implemented in this thesis.

Modified Cholesky decomposition volatility model

The underlying population model for this simplified Cholesky decomposition model can be specified as

$$\mathbf{a}_t|F_{t-1} \sim N(0, \Sigma_t)$$

where $\mathbf{a}_t = (a_{1t}, \dots, a_{pt})'$ is the vector of innovations at time t and $\Sigma_t = LG_tL'$ with G_t a diagonal matrix. Therefore,

$$\mathbf{b}_t = L^{-1}\mathbf{a}_t|F_{t-1} \sim N(0, G_t),$$

where F_{t-1} is the set of information up to and including time $t - 1$ and $b_{i,t}$ is modelled as a GARCH(1,1) process, that is

$$g_{ii,t} = \alpha_{0i} + \alpha_{1i}b_{i,t-1}^2 + \phi_{1i}g_{ii,t-1} \quad (7.31)$$

for $i = 1, \dots, p$ and where $g_{ii,t}$ is the i^{th} diagonal element of the matrix G_t .

Practically, the Cholesky decomposition on the estimated average, or unconditional, covariance matrix $\hat{\Sigma}$ is defined as

$$\hat{\Sigma} = (L^\circ)(G^\circ)(L^\circ)' \quad (7.32)$$

A time-varying vector \mathbf{b}_t is calculated as $\mathbf{b}_t = (L^\circ)^{-1}\mathbf{a}_t$, using Equation (7.22) and the L° matrix from Equation (7.32). A GARCH(1,1) model is then fitted to each of the calculated b_{ii} series for $i = 1, \dots, p$ over time $t = 1, \dots, T$ as

$$g_{ii,t} = \alpha_{0i} + \alpha_{1i}b_{i,t-1}^2 + \phi_{1i}g_{ii,t-1} \quad (7.33)$$

where $\theta_i = \{\alpha_{0i}, \alpha_{1i}, \phi_{1i}\}$ is a set of positive GARCH parameters for $i = 1, \dots, p$, and $g_{ii,t}$ is the i^{th} diagonal element of the matrix G_t° , all for $t = 1, \dots, T$. The initial value matrix \hat{G}_1 for the GARCH(1,1) process can be specified as $g_{ii,1} = \frac{\sum_{t=1}^T (b_{it} - \bar{b}_i)^2}{T-1}$ for $i = 1, \dots, p$ where $\bar{b}_i = \frac{\sum_{t=1}^T b_{it}}{T}$.

Each of the p GARCH(1,1) models in Equation (7.33) can be estimated separately, as in the O-GARCH model, in effect eliminating parameter estimation problems characteristic of traditional multivariate volatility models. From the parameter estimates

$$\hat{g}_{ii,t} = \hat{\alpha}_{0i} + \hat{\alpha}_{1i}b_{i,t-1}^2 + \hat{\phi}_{1i}\hat{g}_{ii,t-1}, \quad (7.34)$$

time-varying matrices \hat{G}_t can be constructed recursively for $t = 2, \dots, T$ and $i = 1, \dots, p$. The conditional covariance matrix $\hat{\Sigma}_t$ of \mathbf{a}_t can be estimated by $(L^\circ)\hat{G}_t(L^\circ)'$ where \hat{G}_t represents a diagonal matrix comprising the conditional variances of the GARCH(1,1) processes for $t = 1, \dots, T$.

An estimate of the forecast for the covariance matrix is therefore $\hat{\Sigma}_{T+1} = \hat{\Sigma}_T(1) = L^\circ(\hat{G}_T(1))L^\circ$, where $\hat{G}_T(1)$ is the forecast of Equation (7.14) using GARCH(1,1) forecasting methodology from Section 7.8 for $i = 1, \dots, p$.

The Modified Cholesky decomposition volatility approach outlined in Equation (7.31) is implemented in this thesis, whereby GARCH(1,1) likelihood functions are fitted separately to the transformed innovations and the matrix decomposition is used to reconstruct the covariance matrix. The new parameterization requires only $3p$ parameters since GARCH(1,1) formulations are used for all $g_{ii,t}$ in each of the p series. Moreover, the likelihood can be split into p likelihoods, which can be optimized individually, significantly reducing the risk of estimation complications.

7.4.4 The problem of order

The main problem with the Cholesky decomposition volatility models is that order is important. In the multiple linear regression interpretation of Cholesky decomposition in Equation (7.23), the first shock series a_{1t} is modelled independently, while the subsequent shock series a_{2t}, \dots, a_{pt} are modelled as having a linear dependence on the innovations a_{1t} to $a_{p-1,t}$.

For example, the volatility of a small stock could be explained by a market index, but the volatility of the index could not be explained by that of the stock. There is a unidirectional relationship between the two volatility series, so the small stock would be a poor choice of the first shock series a_{1t} if a market index is in the portfolio. The values for the elements of the matrix L_t , namely $q_{ij,t}$ for $j < i$ and $i = 1, \dots, p$, are equivalent to β_{ij} , the corresponding slopes of the multiple linear regressions in Equation (7.23). This means that poor choice of ordering could result in a lack of linear relationship between the assets, and hence smaller values for the elements of $q_{ij,t}$, which in turn may affect the forecasting ability of the model.

Although ordering should theoretically not affect a volatility estimate, this modelling approach suggests that the order would tend toward the strength of relationship between different stocks. Vrontos *et al* [32] suggest and implement a Bayesian approach in an attempt to address the problem of order selection. Ordering of the shock series will be further discussed in the application section of this thesis.

Chapter 8

Constant Conditional Correlation Simulations

In order to discuss forecasting performance with respect to both univariate and multivariate volatility models, simulated data were generated using the Constant Conditional Correlation (CCC) multivariate volatility model discussed in Section 6.2. Specifically, the time-varying covariance matrix Σ_t can be re-written as

$$\Sigma_t = D_t C D_t,$$

where C is a matrix of constant correlations, and D_t is a diagonal matrix of elements $\sqrt{\sigma_{ii,t}}$ for $i = 1, \dots, p$, where $\sigma_{ij,t}$ are elements from the covariance matrix Σ_t for $i, j = 1, \dots, p$ at time t . More specifically, the CCC(1,1) model was specified in Equation (6.3) as

$$\mathbf{v}_t = \mathbf{a}_0 + A_1(\mathbf{a}_{t-1} \odot \mathbf{a}_{t-1}) + B_1 \mathbf{v}_{t-1},$$

where \mathbf{a}_0 is a $p \times 1$ dimensional vector, A_1 and B_1 are $p \times p$ dimensional matrices of parameters with strictly positive elements, \mathbf{v}_t is a random vector of variances and \mathbf{a}_t is a $p \times 1$ vector of innovations.

8.1 Simulation algorithm

A brief description of the simulation procedure used to generate a single simulated series is provided here. The simulated series length was defined as T and the number of simulated series or assets as p . After selecting T and p , the average correlation matrix R , and parameter sets \mathbf{a}_0 , A_1 and B_1 were specified. The initial value for the innovation vector \mathbf{a}_1 was set to a vector of zeros of size $p \times 1$ and for \mathbf{v}_1 to a vector of ones of size $p \times 1$.

The following procedure was repeated for $t = 2$ to $T + 1$. The vector \mathbf{v}_t was

generated recursively using Equation (6.3) re-written as

$$\mathbf{v}_t = \mathbf{a}_0 + A_1(\mathbf{a}_{t-1} \odot \mathbf{a}_{t-1}) + B_1\mathbf{v}_{t-1}.$$

A diagonal $p \times p$ matrix D_t of standard deviations was created and defined from the vector \mathbf{v}_t , and used to re-construct the covariance matrix Σ_t using the formulation $\Sigma_t = D_t R D_t$. A vector of $p \times 1$ random variates was then drawn from a multivariate normal distribution $N(\mathbf{0}, \Sigma_t)$, where $\mathbf{0}$ is a vector of zeros and Σ_t is the created time-varying covariance matrix. The resultant $p \times 1$ vector of random variates represents the simulated values for \mathbf{a}_t .

A simulated series of length $T + 1$ was generated to record the values for $t = 2, \dots, T + 1$ in order to omit the initial values from the simulated set, resulting in a simulated set of length T .

8.2 Methodology

In order to thoroughly test forecasting performance of the univariate and multivariate volatility models, the volatility for three-stock portfolio was simulated using a CCC(1,1) model defined in Equation (6.3), using different average correlation matrices, parameter sets and time series lengths, as explained below.

A strongly correlated three-stock portfolio, with a correlation matrix C_{strong} among the volatility series of

$$C_{strong} = \begin{pmatrix} 1 & 0.4 & 0.6 \\ 0.4 & 1 & 0.5 \\ 0.6 & 0.5 & 1 \end{pmatrix} \quad (8.1)$$

was simulated using parameter matrices \mathbf{a}_0 , A_1 and B_1 given by

$$\mathbf{a}_0 = \begin{pmatrix} 1.631 \\ 2.615 \\ 0.499 \end{pmatrix}, A_1 = \begin{pmatrix} 0.1 & 0 & 0 \\ 0 & 0.05 & 0 \\ 0 & 0 & 0.25 \end{pmatrix}, B_1 = \begin{pmatrix} 0.8 & 0 & 0 \\ 0 & 0.9 & 0 \\ 0 & 0 & 0.7 \end{pmatrix}. \quad (8.2)$$

One thousand (1000) simulations were performed using the correlation matrix C_{strong} (8.1) and parameters (8.2) to generate 1000 individual three-stock volatility series of length 251. Similarly, another 1000 simulated three-stock volatility series of lengths 351 and 451 were generated.

Also, a weakly correlated three-stock portfolio, with correlation matrix C_{weak} among volatility series given by

$$C_{weak} = \begin{pmatrix} 1 & 0.07 & 0.2 \\ 0.07 & 1 & 0.3 \\ 0.2 & 0.3 & 1 \end{pmatrix} \quad (8.3)$$

was simulated using parameter matrices \mathbf{a}_0 , A_1 and B_1 , where

$$\mathbf{a}_0 = \begin{pmatrix} 1.069 \\ 1.495 \\ 0.717 \end{pmatrix}, A_1 = \begin{pmatrix} 0.1 & 0 & 0 \\ 0 & 0.05 & 0 \\ 0 & 0 & 0.25 \end{pmatrix}, B_1 = \begin{pmatrix} 0.8 & 0 & 0 \\ 0 & 0.9 & 0 \\ 0 & 0 & 0.7 \end{pmatrix}. \quad (8.4)$$

A thousand (1000) simulations were again performed using the correlation matrix C_{weak} (8.3) and parameters (8.4) to generate 1000 individual three-stock volatility series of length 251 and a further 1000 simulated three-stock volatility series of lengths 351 and 451 were generated.

The individual GARCH(1,1), Alexander O-GARCH, Modified O-GARCH and Modified Cholesky Decomposition model parameters were estimated on 1000 separate time series of lengths 250, 350 and 450 for each of the strongly and weakly correlated portfolios. A one-step-ahead forecast of the 251st, 351st and 451st values in each of the series was generated for every volatility model using the estimated parameter coefficients. The Total Simulated Forecast Error (Total SFE) was calculated by summing the 1000 differences in absolute values of model forecast values and the 251st, 351st and 451st values of each simulated Stocks 1, 2 and 3 respectively. This is a slightly modified version of the Total FE criterion specified in Equation (5.4) from Section 5.5.4, and is defined as

$$\sum_{s=1}^{1000} (|\hat{a}_{i,T+1}^{(s)}| - |a_{i,T+1}^{(s)}|)^2,$$

where $i = 1, 2, 3$ denotes an individual stock within one of the $s = 1, \dots, 1000$ different simulations, $\hat{a}_{i,T+1}^{(s)}$ is the forecast return from a particular volatility model, $a_{i,T+1}^{(s)}$ is the simulated CCC(1,1) volatility for $s = 1, \dots, 1000$ simulations and $T = 251, 351, 451$ represents the forecast value in series.

8.3 Simulation results

8.3.1 Parameter estimates

Average parameter estimates and parameter standard deviations across the 1000 simulations for each volatility model are presented in Tables 8.1 and 8.2 for the weakly and strongly correlated series respectively. For the Cholesky decomposition volatility model, parameters were estimated for all six ordering variations.

The standard deviation of parameter estimates over the simulations can be compared across the different simulations to indicate the parameter stability of the particular volatility model. This was larger for the strongly correlated than for the weakly correlated set, suggesting more accurate GARCH parameter estimation

in weakly correlated portfolios. Furthermore, the between-simulation standard deviations of the GARCH parameter estimates for all three series in both the weakly and strongly correlated portfolios were far lower than for the average parameter estimates. The GARCH parameters in each of the three series seem to underestimate, yet converge to, the CCC(1,1) parameter estimates in Equations (8.2) and (8.4).

This was in contrast to the parameter estimates of all the multivariate matrix decomposition volatility models, for which the standard deviation between the different simulations was lower for the strongly correlated than the weakly correlated set, suggesting that the multivariate matrix decomposition volatility model parameters are estimated more accurately in highly correlated portfolios. However, parameter estimates for all the multivariate matrix decomposition volatility models were not very stable, with large between-simulation standard deviations, sometimes greater than the average parameter estimates. A likely reason for this phenomenon is that matrix decomposition is performed on the average covariance for every simulation, so a slight difference in covariance specification alters matrix decomposition and therefore produces a very different squared shock series on which the parameters are estimated, and hence a very different estimated parameter set.

The GARCH model has directly interpretable parameters in the model context, as a measure of persistence, which converge to a global average with diminishing standard error. In contrast, the multivariate model parameters are less stable, do not appear to converge to any fixed point, and cannot be directly interpreted.

8.3.2 Simulated Forecast Error (Total SFE)

The results for Total SFE are summarized in Table 8.3. Since all the volatility series were fitted to all of the six simulated series of data, that is the weakly correlated series in Stocks 1, 2 or 3 or strongly correlated series in Stocks 1, 2 or 3, and simulated one-step-ahead 1000 times each, Total SFE between all simulations is comparable. Moreover, Total SFE was averaged between all six of the Cholesky decomposition orderings forecast.

Calculated Total SFE was lower for all of the weakly correlated portfolios in comparison to the strongly correlated portfolios. This suggests that increasing correlation between stocks decreases forecasting performance. The best-performing forecasting model was the Alexander O-GARCH, followed narrowly by the Modified O-GARCH. The univariate GARCH outperformed only the Modified Cholesky, which was by far the poorest model for forecasting CCC simulated data.

Series Length	Volatility Model	Stock 1		
		α_{01}	α_{11}	β_{11}
250	GARCH	2.37 (1.533)	0.106 (0.061)	0.738 (0.137)
	Modified O-GARCH	0.228 (0.301)	0.3 (0.319)	0.189 (0.177)
	Alexander O-GARCH	0.39 (0.462)	0.239 (0.24)	0.104 (0.108)
	Cholesky Decomposition	0.241 (0.26)	0.233 (0.277)	0.175 (0.18)
350	GARCH	2.298 (1.428)	0.105 (0.051)	0.747 (0.119)
	Modified O-GARCH	0.207 (0.287)	0.238 (0.269)	0.151 (0.127)
	Alexander O-GARCH	0.328 (0.378)	0.22 (0.23)	0.082 (0.083)
	Cholesky Decomposition	0.202 (0.221)	0.206 (0.258)	0.144 (0.141)
450	GARCH	2.207 (1.272)	0.106 (0.042)	0.754 (0.107)
	Modified O-GARCH	0.172 (0.246)	0.193 (0.222)	0.13 (0.094)
	Alexander O-GARCH	0.183 (0.196)	0.245 (0.271)	0.113 (0.127)
	Cholesky Decomposition	0.168 (0.177)	0.174 (0.225)	0.126 (0.115)
		Stock 2		
250	GARCH	3.281 (1.742)	0.061 (0.041)	0.872 (0.063)
	Modified O-GARCH	0.072 (0.052)	0.089 (0.061)	0.156 (0.075)
	Alexander O-GARCH	0.111 (0.063)	0.06 (0.055)	0.122 (0.074)
	Cholesky Decomposition	0.117 (0.064)	0.105 (0.058)	0.171 (0.073)
350	GARCH	3.311 (1.692)	0.057 (0.034)	0.876 (0.055)
	Modified O-GARCH	0.067 (0.042)	0.087 (0.052)	0.153 (0.062)
	Alexander O-GARCH	0.109 (0.052)	0.056 (0.046)	0.121 (0.062)
	Cholesky Decomposition	0.114 (0.054)	0.101 (0.048)	0.169 (0.06)
450	GARCH	3.362 (1.615)	0.056 (0.03)	0.877 (0.05)
	Modified O-GARCH	0.065 (0.039)	0.084 (0.045)	0.155 (0.053)
	Alexander O-GARCH	0.102 (0.043)	0.069 (0.042)	0.102 (0.046)
	Cholesky Decomposition	0.114 (0.046)	0.099 (0.041)	0.17 (0.054)
		Stock 3		
250	GARCH	0.718 (0.506)	0.258 (0.092)	0.65 (0.125)
	Modified O-GARCH	0.702 (0.3)	0.612 (0.317)	0.657 (0.204)
	Alexander O-GARCH	0.691 (0.248)	0.559 (0.378)	0.627 (0.272)
	Cholesky Decomposition	0.644 (0.268)	0.666 (0.282)	0.657 (0.201)
350	GARCH	0.658 (0.372)	0.258 (0.076)	0.664 (0.096)
	Modified O-GARCH	0.727 (0.289)	0.676 (0.274)	0.695 (0.158)
	Alexander O-GARCH	0.724 (0.206)	0.591 (0.366)	0.683 (0.216)
	Cholesky Decomposition	0.685 (0.232)	0.696 (0.263)	0.69 (0.163)
450	GARCH	0.61 (0.251)	0.258 (0.062)	0.674 (0.067)
	Modified O-GARCH	0.763 (0.249)	0.723 (0.229)	0.714 (0.124)
	Alexander O-GARCH	0.766 (0.16)	0.669 (0.293)	0.732 (0.196)
	Cholesky Decomposition	0.719 (0.19)	0.729 (0.231)	0.705 (0.137)

Table 8.1: Average parameter estimates over 1000 three-stock CCC(1,1) simulated series: Weak correlation (between-simulation parameter standard deviation in brackets)

Series Length	Volatility Model	Stock 1		
		α_{01}	α_{11}	β_{11}
250	GARCH	1.948 (1.491)	0.112 (0.066)	0.693 (0.183)
	Modified O-GARCH	0.223 (0.299)	0.209 (0.246)	0.239 (0.254)
	Alexander O-GARCH	0.272 (0.333)	0.332 (0.329)	0.166 (0.183)
	Cholesky Decomposition	0.227 (0.247)	0.223 (0.27)	0.154 (0.16)
350	GARCH	1.764 (1.348)	0.109 (0.053)	0.715 (0.165)
	Modified O-GARCH	0.177 (0.258)	0.161 (0.186)	0.196 (0.209)
	Alexander O-GARCH	0.219 (0.263)	0.283 (0.303)	0.126 (0.153)
	Cholesky Decomposition	0.185 (0.203)	0.176 (0.225)	0.125 (0.127)
450	GARCH	1.489 (1.063)	0.105 (0.045)	0.75 (0.128)
	Modified O-GARCH	0.145 (0.223)	0.135 (0.135)	0.164 (0.163)
	Alexander O-GARCH	0.183 (0.196)	0.245 (0.271)	0.113 (0.127)
	Cholesky Decomposition	0.149 (0.154)	0.147 (0.197)	0.109 (0.102)
		Stock 2		
250	GARCH	2.481 (1.688)	0.066 (0.045)	0.844 (0.091)
	Modified O-GARCH	0.085 (0.066)	0.122 (0.069)	0.123 (0.073)
	Alexander O-GARCH	0.106 (0.062)	0.076 (0.058)	0.105 (0.063)
	Cholesky Decomposition	0.127 (0.067)	0.11 (0.06)	0.192 (0.074)
350	GARCH	2.41 (1.563)	0.065 (0.037)	0.85 (0.08)
	Modified O-GARCH	0.079 (0.053)	0.119 (0.06)	0.117 (0.058)
	Alexander O-GARCH	0.104 (0.05)	0.075 (0.049)	0.098 (0.052)
	Cholesky Decomposition	0.124 (0.054)	0.107 (0.049)	0.19 (0.06)
450	GARCH	2.258 (1.477)	0.06 (0.031)	0.861 (0.071)
	Modified O-GARCH	0.073 (0.047)	0.118 (0.054)	0.115 (0.056)
	Alexander O-GARCH	0.102 (0.043)	0.069 (0.042)	0.102 (0.046)
	Cholesky Decomposition	0.12 (0.047)	0.101 (0.042)	0.188 (0.055)
		Stock 3		
250	GARCH	1.012 (0.632)	0.254 (0.081)	0.657 (0.113)
	Modified O-GARCH	0.694 (0.299)	0.672 (0.25)	0.64 (0.265)
	Alexander O-GARCH	0.699 (0.248)	0.567 (0.351)	0.653 (0.278)
	Cholesky Decomposition	0.649 (0.256)	0.671 (0.277)	0.661 (0.18)
350	GARCH	0.911 (0.484)	0.257 (0.069)	0.669 (0.085)
	Modified O-GARCH	0.746 (0.258)	0.72 (0.196)	0.689 (0.223)
	Alexander O-GARCH	0.738 (0.202)	0.623 (0.325)	0.719 (0.234)
	Cholesky Decomposition	0.693 (0.215)	0.72 (0.233)	0.69 (0.146)
450	GARCH	0.881 (0.423)	0.255 (0.066)	0.674 (0.081)
	Modified O-GARCH	0.783 (0.227)	0.748 (0.153)	0.722 (0.18)
	Alexander O-GARCH	0.766 (0.16)	0.669 (0.293)	0.732 (0.196)
	Cholesky Decomposition	0.733 (0.17)	0.754 (0.205)	0.708 (0.124)

Table 8.2: Average parameter estimates over 1000 three-stock CCC(1,1) simulated series: Strong correlation (between-simulation parameter standard deviation in brackets)

An inconclusive pattern arose when comparing forecasts based on the lengths of different series. For example, when comparing the weakly correlated sets, Total SFE for the GARCH model declines steadily with increasing series length, implying a positive relationship between forecasting ability and series length. This could be explained by the fact that longer series result in parameters that closer approximate the sample average, which may be a suitable forecasting approach. However, this tendency was not replicated using the other volatility models, making a general conclusion regarding appropriate series length impossible based on this simulation alone. The series with higher associated \mathbf{a}_0 parameters demonstrated by far the greatest Total SFE, since larger \mathbf{a}_0 implies a more volatile series.

8.4 Remarks and conclusions

The multivariate decomposition models have parameters which are difficult to interpret and which have estimates that are somewhat erratic. In contrast, the parameters associated with the univariate GARCH(1,1) models are easily understood and have stable estimates. The forecasts provided by the O-GARCH model outperform those given by the GARCH(1,1) and the Cholesky decomposition models. The forecasts associated with the Modified Cholesky decomposition model were, however, particularly poor. A possible reason for this is that the GARCH and O-GARCH models are similar in form to the CCC(1,1) model used to simulate the data, unlike the Cholesky decomposition models, which were built to accommodate a specific ordering of stocks. It would be possible to simulate data from a model which incorporates a specific ordering of the stocks, but this would bias the results too strongly in favour of the Modified Cholesky decomposition model.

Simulated data can only be used to test real world behaviour if the simulated model resembles the real world closely enough, a feat almost impossible in econometric time series. For this reason, it was decided for the purposes of this thesis that a process of drawing bootstrap samples from a long time series of existing real world data and fitting the models to real-world data was an important test of forecasting performance.

Hansen and Lunde [6] discuss how to implement a bootstrapping methodology on time-series. The one-step-ahead forecasting performance of the univariate volatility models, in section 5.5.4, is compared using a similar bootstrapping methodology. In the next chapter the same bootstrapping methodology is applied to the multivariate volatility models, providing a fair comparison with the forecasting performance of the univariate volatility models.

Weakly Correlated					
Series Length	Volatility Model	Stock 1	Stock 2	Stock 3	Total
250	GARCH	20,073.961	55,143.667	28,135.353	103,352.981
	Modified O-GARCH	11,813.193	29,420.212	16,945.206	58,178.610
	Alexander O-GARCH	11,691.443	29,625.107	16,727.690	58,044.240
	Cholesky Decomposition	29,317.387	47,167.345	35,073.697	111,558.429
350	GARCH	20,065.994	56,146.749	26,766.825	102,979.569
	Modified O-GARCH	11,477.749	30,272.460	18,290.199	60,040.408
	Alexander O-GARCH	11,363.071	30,395.665	18,299.320	60,058.056
	Cholesky Decomposition	28,935.900	47,514.882	33,681.929	110,132.711
450	GARCH	19,349.943	58,183.235	21,712.334	99,245.512
	Modified O-GARCH	11,033.316	31,171.236	13,370.569	55,575.121
	Alexander O-GARCH	10,932.295	31,368.942	13,256.185	55,557.422
	Cholesky Decomposition	27,914.475	47,723.602	30,089.131	105,727.208
Strongly Correlated					
Series Length	Volatility Model	Stock 1	Stock 2	Stock 3	Total
250	GARCH	29,635.851	100,427.042	15,700.247	145,763.140
	Modified O-GARCH	16,548.132	52,268.645	9,888.751	78,705.529
	Alexander O-GARCH	16,336.297	52,708.281	9,564.714	78,609.292
	Cholesky Decomposition	41,583.099	77,592.645	35,011.663	154,187.407
350	GARCH	30,496.519	103,129.709	16,087.043	149,713.271
	Modified O-GARCH	18,089.339	52,206.634	9,764.756	80,060.730
	Alexander O-GARCH	16,769.911	50,597.397	9,946.134	77,313.442
	Cholesky Decomposition	42,518.319	79,531.879	35,445.228	157,495.427
450	GARCH	30,974.385	102,540.402	15,740.726	149,255.513
	Modified O-GARCH	18,061.009	52,089.523	10,440.100	80,590.632
	Alexander O-GARCH	17,883.972	52,536.270	10,237.025	80,657.267
	Cholesky Decomposition	42,613.758	78,940.785	35,225.694	156,780.237

Table 8.3: Total SFE for simulated CCC(1,1) sets

Chapter 9

Application of Matrix Decomposition Methods in Multivariate Volatility Models

This chapter discusses the performance of the Alexander and Modified O-GARCH models relative to each other and to the best univariate models with respect to the real-world data in Portfolios 1 and 2. The Modified Cholesky decomposition volatility model will then be applied to the same data and its performance compared to that of the other models used in this thesis. Emphasis will be placed on the problem of a suitable ordering of share inputs into the Modified Cholesky decomposition model, using an extensive analysis of Portfolio 2 to illustrate this issue.

9.1 Introduction

The parameters for the conditional mean and conditional variance equations were fit separately using the two-tier parameter estimation methodology introduced in Section 4.2.1. The univariate conditional mean equations from Equation (4.2) were fit as follows:

$$\mu_{it} = A_{0i} + B_i r_{mt} + \sum_{k=1}^p \phi_{ik} r_{i,t-k} - \sum_{l=1}^q \theta_{il} a_{i,t-l}$$

for $i = 1, \dots, p$. Parameters $\{A_{0i}, B_i, \phi_{1i}, \dots, \phi_{pi}, \theta_{1i}, \dots, \theta_{qi}\}$ for the conditional mean equations and resultant conditional mean models for Portfolios 1 and 2 were displayed in Tables 5.3 and 5.5. The innovation series a_{it} were then estimated by

$$\hat{a}_{it} = r_{it} - \hat{\mu}_{it} = r_{it} - \hat{r}_{it}, \quad (9.1)$$

where \hat{r}_{it} is the estimated log-return of stock i at time t using Equation (2.17) and r_{it} is the actual log-return of stock i at time t for $i = 1, \dots, p$ and $t = 1, \dots, T$.

The series of innovations \hat{a}_{it} was used as the data for the multivariate volatility models. The different stocks in Portfolio 1 and 2 had different ARMA(p,q) components in the conditional mean equations μ_{it} specified in Tables 5.3 and 5.5. For multivariate models to be fitted to the estimated innovation series \hat{a}_{it} , in order to use information from all stocks for every time point, all component time series must be of equal length. The estimated innovation series \hat{a}_{it} was for this reason truncated at the point t^* , defined as the minimum length of the estimated innovation series \hat{a}_{it} for $i = 1, \dots, p$ and $t = 1, \dots, T$.

The $(T - t^* + 1) \times p$ data matrix $A = (\mathbf{a}_{t^*}, \dots, \mathbf{a}_{T-t^*+1})'$ was defined as a matrix of innovations, where vectors $\mathbf{a}_t = (\hat{a}_{1t}, \hat{a}_{2t}, \dots, \hat{a}_{pt})'$ were observed for $t = t^*, \dots, T$. Truncated matrices A of size 360×8 for Portfolio 1 and 358×5 for Portfolio 2 were relabelled, for ease of understanding, as having size $T^* \times p$ with estimated innovations \hat{a}_{it} for $i = 1, \dots, p$ and $t = 1, \dots, T^*$, where $T^* = T - t^* + 1$.

The covariance matrix forecast Σ_t was obtained from

$$\Sigma_t = \beta\beta' \sigma_{mt}^2 + \Lambda_t \quad (9.2)$$

where β is a $p \times 1$ vector of beta coefficients for stocks $i = 1, \dots, p$, σ_{mt}^2 is the market volatility at time t as defined in Equation (5.3) and Λ_t is the covariance matrix of the innovations vector \mathbf{a}_t at time t . The market volatility σ_{mt}^2 was estimated by the univariate GARCH volatility model, with parameter estimates presented in Table 5.7. The overall covariance matrix Σ_t enabled efficient frontier generation, as well as Value at Risk (VaR) estimation, as discussed in Chapter 10. Model testing and selection were performed on Λ_t , the covariance matrix of the innovation vector \mathbf{a}_t , where the univariate volatility models estimate Λ_t as a diagonal matrix (D_t) and the multivariate volatility models estimate Λ_t as a full covariance matrix.

9.2 Volatility model specification for O-GARCH models

As in the original Alexander approach [30], the data were first standardized according to Equation (7.10) and the $T^* \times p$ matrix $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{T^*})'$ was defined as the matrix of standardized innovations. The sample correlation matrix of innovations is denoted \hat{R} and its spectral decomposition was expressed as

$$\text{cov}(X) = \hat{R} = (H^*)(D^*)(H^*)$$

where H^* is the orthogonal $p \times p$ matrix of eigenvectors and D^* is a $p \times p$ matrix of eigenvalues of the sample correlation matrix \hat{R} . Using orthogonal matrix H^* ,

principal component scores at time t for the Alexander O-GARCH model were defined as $\mathbf{f}_t = (H^*)'\mathbf{x}_t$, where vector $\mathbf{f}_t = (f_{1t}, \dots, f_{pt})'$ for $t = 1, \dots, T^*$. A GARCH(1,1) model was then fitted to each series f_{it} for $i = 1, \dots, p$ over time $t = 1, \dots, T^*$ as

$$\lambda_{it} = \alpha_{0i} + \alpha_{1i}f_{i,t-1}^2 + \beta_{1i}\lambda_{i,t-1}$$

where $\theta_i = \{\alpha_{0i}, \alpha_{1i}, \beta_{1i}\}$ is a set of positive GARCH parameters for $i = 1, \dots, p$ and λ_{it} is the i^{th} diagonal element of the matrix D_t^* for $t = 1, \dots, T^*$.

The parameter estimates for the Alexander O-GARCH model fitted to the stocks of Portfolios 1 and 2 are displayed in Tables 9.1 and 9.2, and time-varying matrices D_t^∇ were constructed recursively for $t = 2, \dots, T^*$ and $i = 1, \dots, p$. An estimate of the conditional correlation matrix \widehat{R}_t was constructed as $(H^*)'(\widehat{D}_t^\nabla)(H^*)$ where \widehat{D}_t^∇ is a diagonal matrix of the conditional standard deviations of the GARCH(1,1) processes for $t = 1, \dots, T^*$.

This means that an estimate of the forecast for the correlation matrix is $\widehat{R}_{T^*+1} = \widehat{R}_{T^*}(1) = (H^*)'(\widehat{D}_{T^*}^\nabla(1))(H^*)$ where $\widehat{D}_{T^*}^\nabla(1)$ is the forecast of Equation (7.14) using the GARCH(1,1) forecasting methodology from Section 4.2.4 for $i = 1, \dots, p$. The forecast correlation matrix $\widehat{R}_{T^*}(1)$ has to be converted to a covariance matrix in order to generate a forecast of the covariance matrix of innovations Λ_t . The conversion is performed by invoking $\widehat{\Lambda} = S^{-1}\widehat{R}_t S^{-1}$ from Equation (7.15), where S is a diagonal matrix of the sample standard deviations of the innovation series. The one-step-ahead forecast of the diagonal elements of the matrix Λ_t , namely $\text{diag}(\Lambda_{T^*}(1))$, is presented in Table 9.3. This is a forecast of the unique risk only (with no covariance forecast) and is done for ease of comparison with the univariate models.

Now consider the Modified O-GARCH procedure, with the same $T^* \times p$ matrix $A = (\mathbf{a}_1, \dots, \mathbf{a}_{T^*})'$ of innovations. The sample covariance matrix $\widehat{\Lambda}$ and the corresponding spectral analysis of the covariance matrix of the innovations are expressed as

$$\widehat{\Lambda} = (H^\circ)'(D^\circ)(H^\circ).$$

Using the orthogonal matrix H° , principal component scores at time t for the Modified O-GARCH can be defined as $\mathbf{v}_t = (H^\circ)'\mathbf{a}_t$, where the vector $\mathbf{v}_t = (v_{1t}, \dots, v_{pt})'$ for $t = 1, \dots, T^*$. Also define $\mathbf{w}_t = (D^\circ)^{-\frac{1}{2}}\mathbf{v}_t$ as the vector of whitened scores at time t , where D° is a diagonal matrix of eigenvalues. Then a GARCH(1,1) model can be fitted to each of the series of whitened scores w_{it} for $i = 1, \dots, p$ over time $t = 1, \dots, T$ as

$$\lambda_{w,it} = \alpha_{w,0i} + \alpha_{w,1i}w_{i,t-1}^2 + \beta_{w,1i}\lambda_{w,i,t-1},$$

where $\theta_i = \{\alpha_{w,0i}, \alpha_{w,1i}, \beta_{w,1i}\}$ is a set of positive GARCH parameters for $i = 1, \dots, p$ and λ_{it} is the i^{th} diagonal element of the matrix D_t° for $t = 1, \dots, T$. Tables 9.1 and 9.2 also show the parameter estimates for the Modified O-GARCH model fitted to stocks in Portfolios 1 and 2 respectively.

From the estimates of the parameters the time-varying matrices D_t° can be constructed recursively for $t = 2, \dots, T^*$ and $i = 1, \dots, p$ using the principal component scores. Hence, the conditional covariance matrix $\widehat{\Lambda}_t$ of \mathbf{a}_t can be estimated by $(H^\circ)'(\widehat{D}_t^\circ)(H^\circ)$, where \widehat{D}_t° represents a diagonal matrix comprising the conditional standard deviations of the GARCH(1,1) processes for $t = 1, \dots, T^*$. This means that an estimate of the forecast for the covariance matrix is $\widehat{\Lambda}_{T^*+1} = \widehat{\Lambda}_{T^*}(1) = (H^\circ)'(\widehat{D}_{T^*}^\circ(1))(H^\circ)$ where $\widehat{D}_{T^*}^\circ(1)$ is the forecast of Equation (7.14) using the GARCH(1,1) forecasting methodology from Section 7.8 for $i = 1, \dots, p$.

The one-step ahead forecasts of the diagonal elements of the matrix Λ_t , namely $\text{diag}(\Lambda_{T^*}(1))$, are presented in Table 9.3. The Modified O-GARCH one-step-ahead forecasts of unique risk are consistently lower than those of the Alexander O-GARCH model. This could be due to the Modified O-GARCH using whitened data in comparison to the standardized data used in the Alexander O-GARCH model. This is again a forecast of the unique risk only and is done for ease of comparison with the univariate models.

9.3 Model selection for O-GARCH

9.3.1 Efficient frontier with dynamic covariance estimates

The expected returns defined for the multivariate model mean equations in Tables 5.3 and 5.5 are displayed in Tables 5.14 for both Portfolios 1 and 2. The one-step-ahead forecast of the covariance matrix for the innovations was used as the estimated covariance matrix for both the Alexander and Modified O-GARCH models respectively.

Figures 9.1 and 9.2 show the distinct efficient frontiers plotted for the Alexander O-GARCH, Modified O-GARCH and GARCH models for Portfolios 1 and 2. There is visible covariance among the innovation series, where covariances are assumed to be zero under the CAPM and univariate volatility models.

9.3.2 Sum of square variance forecasting error

The concept of Total Forecast Error (Total FE) was proposed in Section 5.5.4 to calculate and sum forecast error over a 100 observations, and was adapted in order to compare variance forecasting performance between multivariate and

	Coefficients for Alexander O-GARCH		
	$\hat{\alpha}_{0i}$	$\hat{\alpha}_{1i}$	$\hat{\beta}_{1i}$
ABSA	0.151 (0.096)	0.075 (0.038)	0.836 (0.086)
EDCON	0.006 (0.008)	0.0343 (0.011)	0.960 (0.015)
GRINROD	0.000 (0.006)	0.031 (0.007)	0.968 (0.007)
GROUP5	0.011 (0.010)	0.057 (0.022)	0.932 (0.026)
JDGROUP	0.02 (0.011)	0.057 (0.016)	0.921 (0.022)
LIBERTY	0.006 (0.008)	0.034 (0.011)	0.959 (0.017)
SABREW	0.048 (0.032)	0.109 (0.042)	0.828 (0.069)
SHPR	0.011 (0.012)	0.060 (0.022)	0.925 (0.031)
	Coefficients for Modified O-GARCH		
	$\hat{\alpha}_{0i}$	$\hat{\alpha}_{1i}$	$\hat{\beta}_{1i}$
ABSA	0.005 (0.007)	0.025 (0.008)	0.968 (0.012)
EDCON	0.020 (0.014)	0.061 (0.019)	0.918 (0.028)
GRINROD	0.067 (0.053)	0.049 (0.026)	0.886 (0.070)
GROUP5	0.067 (0.028)	0.087 (0.029)	0.846 (0.046)
JDGROUP	0.028 (0.021)	0.073 (0.027)	0.898 (0.035)
LIBERTY	0.009 (0.012)	0.027 (0.014)	0.963 (0.022)
SABREW	0.024 (0.018)	0.065 (0.025)	0.910 (0.036)
SHPR	0.006 (0.011)	0.078 (0.025)	0.919 (0.026)

Table 9.1: Alexander O-GARCH and Modified O-GARCH parameter estimates for Portfolio 1 (standard errors in brackets)

	Coefficients for Alexander O-GARCH		
	$\hat{\alpha}_{0i}$	$\hat{\alpha}_{1i}$	$\hat{\beta}_{1i}$
ANGLO	0.208 (0.201)	0.040 (0.031)	0.89 (0.091)
DRD	0.059 (0.031)	0.064 (0.027)	0.875 (0.043)
GFIELDS	0.000 (0.000)	0.006 (0.005)	0.992 (0.005)
HARMONY	0.035 (0.026)	0.067 (0.033)	0.845 (0.091)
IMPLATS	0.000 (0.001)	0.025 (0.013)	0.973 (0.014)
	Coefficients for Modified O-GARCH		
	$\hat{\alpha}_{0i}$	$\hat{\alpha}_{1i}$	$\hat{\beta}_{1i}$
ANGLO	0.069 (0.097)	0.023 (0.025)	0.908 (0.114)
DRD	0.016 (0.006)	0.025 (0.009)	0.958 (0.012)
GFIELDS	0.000 (0.000)	0.003 (0.010)	0.997 (0.010)
HARMONY	0.000 (0.000)	0.013 (0.008)	0.984 (0.007)
IMPLATS	0.000 (0.305)	0.014 (0.083)	0.984 (0.324)

Table 9.2: Alexander O-GARCH and Modified O-GARCH parameter estimates for Portfolio 2 (standard errors in brackets)

Portfolio 1	Alexander O-GARCH	Modified O-GARCH
ABSA	0.047	0.029
EDCON	0.039	0.036
GRINROD	0.039	0.024
GROUP5	0.037	0.027
JDGROUP	0.049	0.033
LIBERTY	0.043	0.027
SABREW	0.040	0.017
SHIPT	0.046	0.029
Portfolio 2	Alexander O-GARCH	Modified O-GARCH
ANGLO	0.046	0.043
DRD	0.016	0.033
GFIELDS	0.047	0.020
HARMONY	0.045	0.037
IMPLATS	0.048	0.016

Table 9.3: Alexander O-GARCH and Modified O-GARCH standard deviation forecasts of unique risk

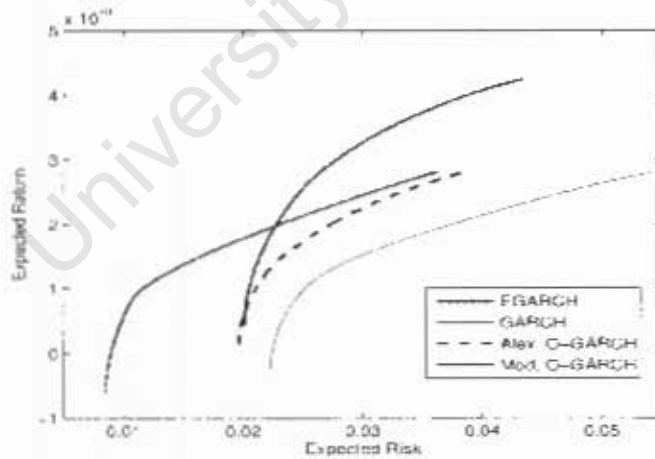


Figure 9.1: Efficient frontiers using O-GARCH and univariate EGARCH and GARCH volatility models: Portfolio I

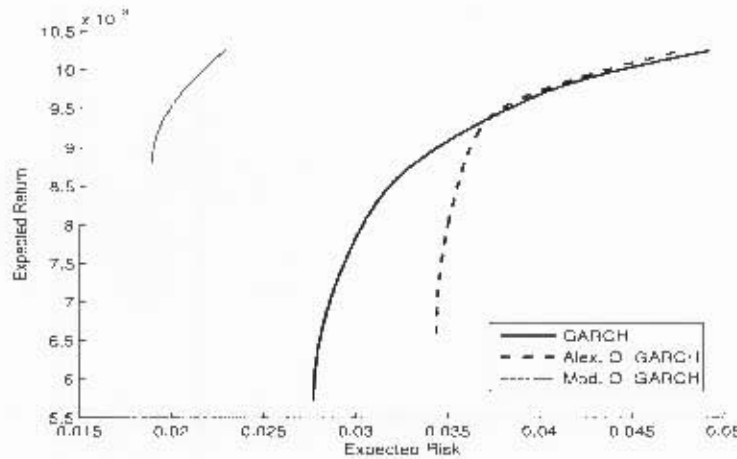


Figure 9.2: Efficient frontiers using O-GARCH and univariate volatility models: Portfolio 2

univariate volatility models. Total FE compares the diagonal elements of forecast matrix $\hat{\Lambda}_t$ to the diagonal elements of the actual covariance matrix of innovations $\Lambda_t = \text{diag}(\mathbf{a}_t \mathbf{a}_t')$ at time t . Hence, the Total FE from Equation (5.4) becomes

$$\sum_{i=1}^p \sum_{t=T}^T \sum_{M=1}^M (|\hat{a}_{ii,t} - |a_{ii,t}|)^2. \quad (9.3)$$

where $\hat{a}_{ii,t}$ is the i^{th} diagonal element of matrix $\hat{\Lambda}_t$ and $a_{ii,t}$ is the i^{th} element of matrix Λ_t for $i = 1, \dots, p$ and $t = 1, \dots, T$.

The forecast errors per stock shown in Figure 9.3 for Portfolio 1 are generally smallest using the Modified O-GARCH volatility model, except for the ABSA, Group Five and SA Breweries stocks, that had a larger forecast error for the O-GARCH models. It seems from this illustration that both multivariate O-GARCH models predicted the variance of the Group Five stock relatively poorly. The multivariate Total FE for both portfolios, displayed in Table 9.4, are much lower than for the best univariate equivalent. Specifically, for Portfolio 1, the Modified O-GARCH has a slightly smaller forecast error than the Alexander O-GARCH method.

Figure 9.4 shows the forecasting errors per stock in Portfolio 2 for the O-GARCH and Modified O-GARCH models. The forecasting error per stock for the univariate GARCH model for Portfolio 2 was not included in Figure 9.4 due to scaling differences for the three plots. Table 9.4 shows that the total sum of square error due to the GARCH model, the best univariate model for Portfolio 2, was 0.0963 which is 47 times larger than the equivalent forecasting error for the Alexander

	EGARCH (Portfolio 1) GARCH (Portfolio 2)	Alexander O-GARCH	Modified O-GARCH
Total FE for Portfolio 1	0.0157	0.0107	0.0095
Total FE for Portfolio 2	0.0963	0.0020	0.0011

Table 9.4: Sum of forecast square errors for different volatility models

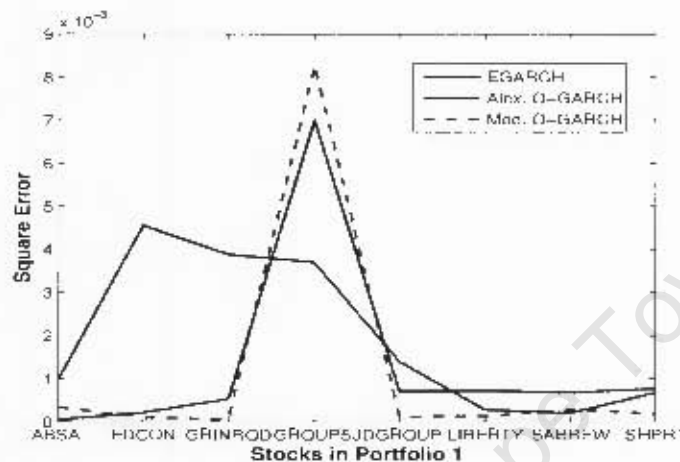


Figure 9.3: Forecast error per stock using Alexander O-GARCH, Modified O-GARCH and EGARCH volatility models: Portfolio 1

O-GARCH model. The Modified O-GARCH model has a sum of square error of 0.0011, almost half that of the Alexander O-GARCH. The multivariate O-GARCH decomposition models clearly outperform the univariate models in both Portfolio 1 and 2, a characteristic particularly marked in the high-correlation Portfolio 2. Furthermore, the Modified O-GARCH model better predicted variance than the Alexander O-GARCH model.

9.3.3 The Frobenius norm error

In order to compare an actual and a forecast covariance matrix, a summary scalar measure of the elements of a matrix, using any of a number of matrix norms available [33], is required. The Frobenius norm, defined for any $m \times n$ matrix A as

$$\|A\|_F = \left(\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2 \right)^{\frac{1}{2}} \quad (9.4)$$

where a_{ij} are elements of matrix A , was used in this thesis with

$$E_t = A_t - \hat{A}_t$$

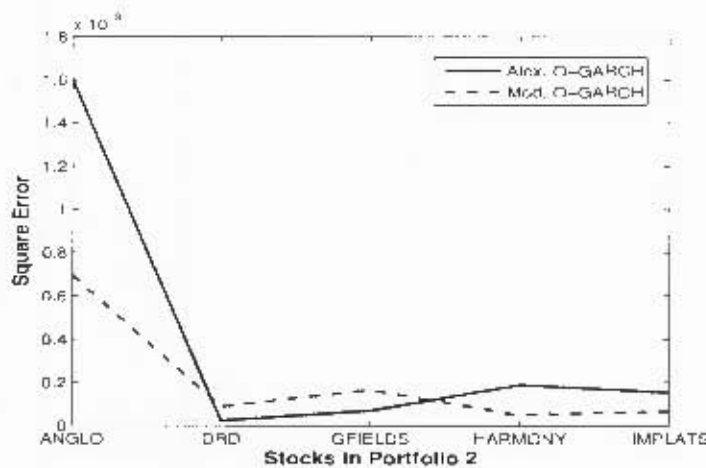


Figure 9.4: Forecast error per stock using Alexander O-GARCH and Modified O-GARCH volatility models: Portfolio 2

for $t = 1, \dots, T$ and where Λ_t is the actual covariance matrix of the innovations \mathbf{a}_t at time t and $\hat{\Sigma}_t$ is the forecast covariance matrix. The Frobenius norm of the matrix E_t , for a single value of t , is defined as

$$\|E_t\|_F = \left(\sum_{i=1}^p \sum_{j=1}^p (a_{ij,t} - \hat{a}_{ij,t})^2 \right)^{\frac{1}{2}} \quad (9.5)$$

where a_{ij} and \hat{a}_{ij} for $i, j = 1, \dots, p$ are elements of the covariance matrices Λ_t and $\hat{\Lambda}_t$ respectively. This matrix norm represents the sum of squared error between the actual and forecast elements of the covariance matrix, and will therefore be used to draw comparisons between the covariance forecasting ability of the different multivariate models.

The forecasting error over M one-period-ahead forecasts was found using a multivariate time series of size $T^* \times i$ for $i = 1, \dots, p$. A volatility model was fitted using $(T^* - (M - 1)) \times i$ observations from the data set and the one-step-ahead forecast for the innovation series of the portfolio was used to generate the covariance matrix $\hat{\Lambda}_{T^*-M}(1)$ for time $T^* - M$. The actual observed covariance matrix for a portfolio Λ_{T^*-M} at time $T^* - M$ was extracted from the existing data and used to generate the Frobenius norm $\|E_t\|_F$ specified in Equation (9.5). This process was repeated for $T^* - k$ with $k = M, \dots, 1$ and a sum of the square forecasting error for a portfolio of stocks computed as

$$\sum_{t=T^*-M+1}^{T^*} \|E_t\|_F^2 \quad (9.6)$$

	Alexander O-GARCH	Modified O-GARCH
Total norm error for Portfolio 1	12.607	12.924
Total norm error for Portfolio 2	5.123	5.942

Table 9.5: Sum of Frobenius norms for the difference between actual and forecast covariance matrices

This measure is called the total norm of forecast error, hereafter referred to as the total norm error. In this thesis M was set as 100, so each volatility model was fitted and forecast a hundred times, for each of Portfolios 1 and 2, to compute the total norm errors displayed in Table 9.5.

In both portfolios, the Alexander O-GARCH model better forecast covariance than the Modified O-GARCH. The Frobenius norms do not, however, vary by large amounts, suggesting that although the Modified O-GARCH model outperformed the Alexander O-GARCH for predicting variances, it may be weaker at predicting covariances. The total norm error in Portfolio 1 was more than twice that in Portfolio 2.

9.4 Modified Cholesky decomposition volatility model

The ordering of stocks in the covariance matrix creates an algorithmic problem of factorial complexity for the Cholesky matrix decomposition method. For a p -stock portfolio the potential number of orderings is $p!$, and iterating through all possible orderings becomes computationally intractable with increasing p . Some type of computational shortcut is therefore required to find an order approximating the optimal for forecasting performance for the covariance matrix.

9.4.1 Volatility model specification for the Modified Cholesky decomposition

Practically, the Modified Cholesky decomposition was implemented based on the estimated average covariance matrix $\hat{\Lambda}$ defined as

$$\hat{\Lambda} = (L^\circ)(G^\circ)(L^\circ)',$$

where the matrices L° and G° are specified in Section 7.4.3. The matrices L° and G° differed per ordering, but the fitting of the volatility model will be discussed referring only to the matrices L° and G° to avoid repetition. A time-varying vector \mathbf{b}_t was calculated using the L° matrix as $\mathbf{b}_t = (L^\circ)^{-1}\mathbf{a}_t$. A GARCH(1,1)

model was fitted to each of the b_{it} series for $i = 1, \dots, p$ over time $t = 1, \dots, T^*$ as

$$g_{ii,t} = \alpha_{0i} + \alpha_{1i}b_{i,t-1}^2 + \phi_{1i}g_{ii,t-1}$$

where $\theta_i = \{\alpha_{0i}, \alpha_{1i}, \phi_{1i}\}$ is a set of positive GARCH parameters for $i = 1, \dots, p$, and $g_{ii,t}$ is the i^{th} diagonal element of the matrix G_t° for $t = 1, \dots, T^*$. From the parameter estimates time-varying matrices \widehat{G}_t were constructed recursively for $t = 2, \dots, T^*$ and $i = 1, \dots, p$. The conditional covariance matrix $\widehat{\Sigma}_t$ of \mathbf{a}_t was estimated by $(L^\circ)\widehat{G}_t(L^\circ)'$ where \widehat{G}_t represents a diagonal matrix of the conditional variances of the GARCH(1,1) processes for $t = 1, \dots, T^*$.

An estimate of the forecast for the covariance matrix was therefore $\widehat{\Lambda}_{T^*+1} = \widehat{\Lambda}_{T^*}(1) = L'(\widehat{G}_{T^*}(1))L$, where $\widehat{G}_{T^*}(1)$ is the forecast of Equation (7.14) using GARCH(1,1) forecasting methodology from Section 7.8 for $i = 1, \dots, p$.

9.4.2 Effect of various stock orderings on Portfolio 2

The smaller of the two stock portfolios considered, namely Portfolio 2, consists of five stocks which means that there are $5!$ or 120 potential orderings for the covariance matrix, as opposed to 40 320 different orderings for the eight stocks in Portfolio 1. 120 reorderings of Portfolio 2 stocks with concomitant reordering of the rows and columns of the covariance matrix Λ were indexed by an order vector.

The Total FE and Total (Frobenius) norm error measures for multivariate volatility models introduced in Section 9.3.2 were calculated for all 120 different covariance matrix orderings. The best Modified Cholesky decomposition volatility model with regard to Total FE was given by the stock ordering: Anglo Gold; Goldfields; Harmony; Impala; DRD. This is the order of market impact, or market capitalization, suggesting that perhaps order determination for the Cholesky class of models may require ordering the stocks in a portfolio based on market impact using analyst experience and market information.

On the other hand, with regards to the Frobenius norm measure of covariance error, the best Modified Cholesky decomposition volatility model was that with the stocks ordered as: DRD; Harmony; Goldfields; Impala; Anglo Gold. If the sample variances of the stocks were computed over the entire sample period and the stocks were sorted in descending order of sample variance the order would be: DRD; Harmony; Goldfields; Anglo Gold; Impala, closely resembling the order based on Frobenius norm performance.

The best Cholesky decomposition volatility model produced a Total FE of 0.0054, the worst Cholesky decomposition volatility model 0.0068, compared to a Total

FE of 0.0011 for the Modified O-GARCH and 0.096 for the GARCH model. The worst-performing ordering of the Modified Cholesky decomposition volatility model still outperforms the univariate GARCH volatility model in Portfolio 2 with respect to forecasting variance, but the best ordering of the Modified Cholesky decomposition volatility model is outperformed by the Modified O-GARCH model in Portfolio 2.

The best Cholesky decomposition volatility model produced a Total norm error of 4.713 and the worst Cholesky decomposition volatility model 4.855, compared to 5.123 for the Modified O-GARCH model. There is a relatively small difference between the best and worst input orders of the Modified Cholesky decomposition volatility model with regard to covariance estimation as measured by the Frobenius norm. Even the worst of the Modified Cholesky decomposition volatility models predicted covariance better than the Alexander O-GARCH model. This suggests that in Portfolio 2, the Cholesky model is relatively efficient at forecasting covariance matrices, independent of order.

9.4.3 Order determination for the covariance matrix

Stock inputs into the Cholesky model could be sorted due to market influence, market capitalization or analyst opinion. In this thesis, a more robust order selection technique was explored.

A covariance matrix of dimension 2×2 was decomposed as $\Sigma_t = L_t G_t L_t'$, so the second diagonal element of G_t could be written as

$$g_{22,t} = \sigma_{22,t} - \frac{\sigma_{21,t}^2}{\sigma_{11,t}}, \quad (9.7)$$

which is in the form of the least squares solution for the slope of the simple linear regression and is expressed in terms of innovations as $a_{2t} = \beta a_{1t} + b_{2t}$.

If an innovation series a_{2t} has a larger market impact than a second stock with innovation series a_{1t} , then a_{2t} must be regressed against a_{1t} , since the former can provide information about the movement of the latter, but the converse is not necessarily true. The stocks should therefore be ordered increasingly according to market influence, with the most influential stock providing the first innovation series.

A potentially effective ordering strategy was to sort the stocks in descending order based on respective sums of coefficients of determination (coefficient of determination ordering), calculated as the sum of the squares of the sample correlation between each of stocks i and j , where $i, j = 1, \dots, p$. If a stock has a

strong linear relationship with all other stocks, then the sum of its individual coefficients of determination will be greater than that of the other stocks.

Another ordering technique, named the maximum variance ordering, sorted stocks in descending order of sample variance. A stock with high average variance implied higher-than-average risk and larger price fluctuation, hence greater volatility.

The Modified Cholesky decomposition model was fitted by using two ordering techniques discussed above. The first ordering technique is the coefficient of determination order and the second is called the maximum variance order. Tables 9.6 and 9.7 show the parameter estimates for the Modified Cholesky decomposition volatility model fitted to the stocks of Portfolio 1 and 2 for the coefficient of determination ordering and the maximum variance ordering respectively. The one-step ahead forecast of the diagonal elements of matrix Λ_t , namely $diag(\Lambda_T(1))$, is presented in Table 9.8 for the stocks in Portfolio 1 and 2. This is a forecast of the unique risk only and is done for ease of comparison with the univariate models.

A full analysis of all potential orderings of Portfolio 2 found the coefficient of determination ordering to be the 61st best model based on the Total FE measure and the 51st best with regard to total norm error. The maximum variance ordering was the second-best in terms of total norm error prediction for Portfolio 2, but 115th for predicting variance using the Total FE measure.

9.4.4 Model selection for the Modified Cholesky decomposition volatility model

Figures 9.5 and 9.6 show the Total FE per stock for Portfolios 1 and 2 respectively. Figure 9.6 indicates that, for Portfolio 2, the Modified O-GARCH Total FE was lower than both the Cholesky decomposition methods. The Group Five stock was weakly forecast by both the Cholesky and O-GARCH volatility models, as shown in Figure 9.5. Table 9.9 shows that the sum of forecast square error of variance for Portfolio 2 is better forecast by the Modified O-GARCH than by the Cholesky decomposition model orderings. The coefficient of determination ordering however outperforms the maximum variance ordering for Portfolio 2.

The results for Portfolio 1, presented in Table 9.9, show that both Cholesky decomposition model orderings outperform the Modified O-GARCH model with respect to variance forecasting. The principal component-based O-GARCH model performed less well on Portfolio 1, since it is not as highly-correlated as Portfolio 2. The coefficient of determination ordering and the maximum variance ordering perform comparably, with the maximum variance ordering being slightly more

	Coefficient of Determination Order (Coeff. of Det.)		
	α_{0i}	α_{1i}	β_{1i}
ABSA	0.044 (0.031)	0.071 (0.032)	0.887 (0.053)
EDCON	0.054 (0.017)	0.116 (0.035)	0.838 (0.039)
GRINROD	0.007 (0.005)	0.031 (0.007)	0.96 (0.01)
GROUP5	0.005 (0.007)	0.024 (0.009)	0.97 (0.013)
JDGROUP	0.101 (0.066)	0.095 (0.056)	0.803 (0.114)
LIBERTY	0.037 (0.027)	0.077 (0.029)	0.886 (0.049)
SABREW	0.005 (0.008)	0.068 (0.019)	0.929 (0.017)
SHPRT	0.006 (0.009)	0.027 (0.012)	0.966 (0.014)
	Maximum Variance Order (Max. Var.)		
ABSA	0.006 (0.03)	0.07 (0.031)	0.894 (0.049)
EDCON	0.137 (0.017)	0.106 (0.032)	0.846 (0.038)
GRINROD	0 (0.005)	0.033 (0.007)	0.958 (0.011)
GROUP5	0.054 (0.001)	0.033 (0.009)	0.965 (0.008)
JDGROUP	0.038 (0.086)	0.109 (0.065)	0.754 (0.14)
LIBERTY	0.004 (0.032)	0.09 (0.033)	0.861 (0.057)
SABREW	0 (0.008)	0.067 (0.02)	0.931 (0.018)
SHPRT	0.049 (0.009)	0.026 (0.011)	0.967 (0.014)

Table 9.6: Modified Cholesky decomposition parameter set for coefficient of determination and maximum variance orderings: Portfolio 1 (standard errors in brackets)

	Coefficient of Determination Order (Coeff. of Det.)		
	α_{0i}	α_{1i}	β_{1i}
ANGLO	0.026 (0.024)	0.059 (0.027)	0.915 (0.042)
DRD	0.323 (0.105)	0.154 (0.035)	0.529 (0.113)
GFIELDS	0 (0)	0.01 (0.008)	0.988 (0.006)
HARMONY	0.07 (0.064)	0.04 (0.035)	0.89 (0.094)
IMPLATS	0.065 (0.041)	0.065 (0.032)	0.871 (0.06)
	Maximum Variance Order (Max. Var.)		
ANGLO	0.023 (0.023)	0.038 (0.022)	0.939 (0.038)
DRD	0.658 (0.455)	0.074 (0.05)	0.268 (0.459)
GFIELDS	0.007 (0.013)	0.038 (0.017)	0.953 (0.025)
HARMONY	0.762 (0.246)	0.172 (0.059)	0.077 (0.264)
IMPLATS	0.065 (0.041)	0.065 (0.032)	0.871 (0.06)

Table 9.7: Modified Cholesky decomposition parameter set for coefficient of determination and maximum variance orderings: Portfolio 2 (standard errors in brackets)

Portfolio 1	Max. Var.	Coeff. of Det.
ABSA	0.064	0.047
EDCON	0.050	0.053
GRINROD	0.053	0.041
GROUP5	0.056	0.031
JDGROUP	0.044	0.052
LIBERTY	0.041	0.023
SABREW	0.032	0.057
SHPR	0.022	0.064
Portfolio 2		
ANGLO	0.094	0.069
DRD	0.073	0.052
GFIELDS	0.055	0.056
HARMONY	0.055	0.088
IMPLATS	0.051	0.051

Table 9.8: Modified Cholesky decomposition standard deviation forecast of unique risk for different input orderings

accurate.

As noted in Table 9.9, the Cholesky decomposition volatility models have the best prediction of covariance for Portfolio 2. Table 9.9 reiterates this in the sum of Frobenius norm covariance error for the coefficient of determination and maximum variance orderings. The total norm errors in Table 9.9 were very similar over the stocks in Portfolio 1, with the Alexander O-GARCH model slightly outperforming both Modified Cholesky decomposition orderings.

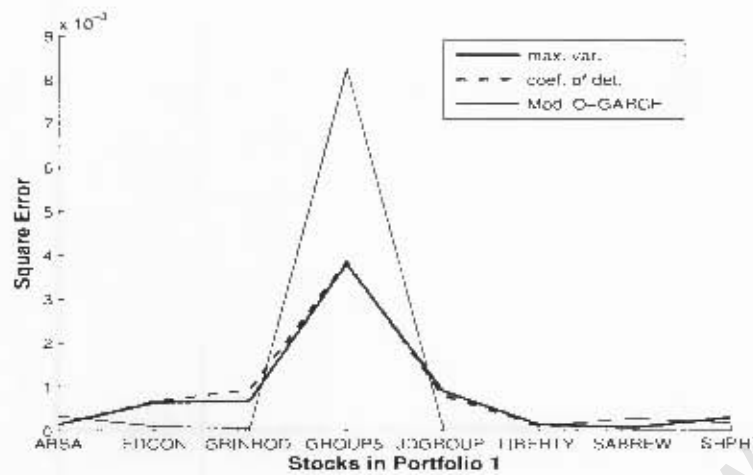


Figure 9.5: Forecast error per stock using different volatility models: Portfolio 1

	Max. Var.	Coeff. of Det.	Mod. O-GARCH
Total FE for Portfolio 1	0.0068	0.0069	0.0095
Total FE for Portfolio 2	0.0068	0.0063	0.0011
	Max. Var.	Coeff. of Det.	Alex. O-GARCH
Total norm error for Portfolio 1	12.734	12.725	12.607
Total norm error for Portfolio 2	4.715	4.743	5.123

Table 9.9: Sum of forecast square error of variance for different volatility models

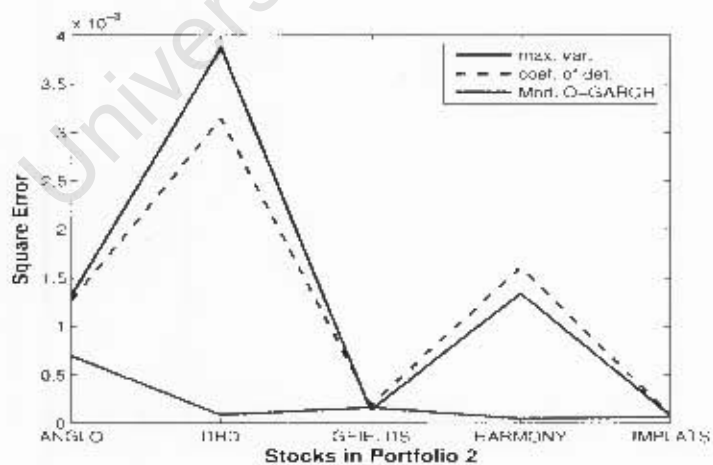


Figure 9.6: Forecast error per stock using different volatility models: Portfolio 2

Chapter 10

Value at Risk

Financial market risk includes credit, liquidity, and market risk. Market risk, defined as a single estimate of the amount by which a position of an institution could decline due to general market movements during a given holding period [3], is rare but dangerous due to unpredictability and lack of ability to control or prohibit it. The only real measure to counter market risk is planning, and institutions must realize that monetary loss could occur with any extreme market shift.

Value at Risk (VaR) is described as the maximal loss in a financial position such that losses greater than that loss occur with a given probability [3]. This involves defining $\Delta V(l)$ as the change in asset value from time t to $t+l$, with an associated unknown cumulative distribution function (CDF), $F_l(x)$. The $\text{VaR}=\nu_p$ over time horizon l with probability P is then defined as

$$P = Pr[\Delta V(l) \geq VaR] = 1 - Pr[\Delta V(l) \leq VaR] = 1 - F_l(VaR). \quad (10.1)$$

Hence VaR_P is the P^{th} quantile of the distribution F_l . Equation (10.1) states that the probability of losing more money than the VaR amount over some time horizon l is P [3]. Extreme losses are associated with the left tail of the distribution of $\Delta V(l)$, while extreme gains are concerned with the right-hand tail.

In this thesis, the CDF for the VaR was estimated by multiple econometric models built to study volatility. The normality assumption which holds for all the log-returns in this thesis enabled VaR for Portfolios 1 and 2 to be computed using a normal approximation to the CDF $F_l(x)$. The 5% and 1% quantiles for the standard normal distribution are 1.645 and 2.33 respectively. Hence $\mathbf{r}_{T+1} \sim N(\mathbf{r}_{T+1}, \Sigma_{T+1})$ where \mathbf{r}_{T+1} is a vector of forecast returns and Σ_{T+1} is the forecast covariance matrix.

Now the VaR of stock $i = 1, \dots, p$ for a one-period-ahead forecast is

$$r_{i,T+1} - 1.645\sigma_{ii,T+1}$$

$$r_{i,T+1} - 2.33\sigma_{ii,T+1} \quad (10.2)$$

at the 5% and 1% levels of probability respectively, where $r_{i,T+1}$ is an element of vector \mathbf{r}_{T+1} and $\sigma_{ii,T+1}$ are diagonal elements of matrix Σ_{T+1} for $i = 1, \dots, p$.

All calculations performed in this thesis thus far used log-return series r_{it} for $i = 1, \dots, p$ and $t = 1, \dots, T$. For a meaningful VaR figure, the log-return calculated in Equation (10.2) was converted to a return using the expression $R_{it} = \exp(r_{it}) - 1$ for $i = 1, \dots, p$ and $t = 1, \dots, T$. VaR is reported for an arbitrary figure of one million rands, and is therefore defined for the purpose of this thesis as

$$VaR_P = R_{i,T+1} \times 1000000$$

where P refers to the probability associated with the computed VaR quantile.

Tables 10.1 and 10.2 show the VaR on a million Rand for a one-week holding period using probabilities of 5% and 1% respectively for Portfolio 1, that is the amount of money that could be lost per million Rand invested over a week for each quantile. Quantiles are related to an expected frequency of this loss by the relationship $\frac{1}{P}$. At a 5% quantile, the VaR losses in Table 10.1 are expected to occur once per twenty investment periods, or, in this case, once every twenty weeks. The 1% quantile in Table 10.2 describes the loss expected to occur once every hundred periods. Tables 10.3 and 10.4 show the VaR results for Portfolio 2.

	EGARCH	O-GARCH		Modified Cholesky	
		Alexander	Modified	Max. VaR.	Coeff. of Det.
ABSA	105030	73442	61172	76560	83110
EDCON	100800	59020	56749	83750	87540
GRINROD	80880	60725	38114	106940	102680
GROUP5	77720	59512	44700	77510	98830
JDGROUP	63850	76863	62175	88720	82320
LIBERTY	56840	69365	58282	65400	50590
SABREW	42030	65470	49153	44250	52930
SHOPRITE	47700	72034	52832	75430	64700

Table 10.1: Value at Risk estimated as 5% quantile for one-week loss in Rands: Portfolio 1

10.1 Backtesting

According to the previous section, a VaR estimated by the P th quantile should not be exceeded more than $\frac{1}{P}$ times over the holding period considered, meaning

	EGARCH	O-GARCH		Modified Cholesky	
		Alexander	Modified	Max. VaR.	Coeff. of Det.
ABSA	145600	102580	85699	106830	115820
EDCON	140720	83630	80497	116800	121830
GRINROD	112880	85200	53855	148180	142350
GROUP5	108320	83280	62765	109030	136120
JDGROUP	89360	107240	87055	123350	115610
LIBERTY	79000	96270	80991	90380	70950
SABREW	57990	90470	67899	62180	73580
SHOPRITE	66970	100560	74085	104610	90680

Table 10.2: Value at Risk estimated as 1% quantile for one-week loss in Rands: Portfolio 1

	GARCH	O-GARCH		Modified Cholesky	
		Alexander	Modified	Max. VaR.	Coeff. of Det.
ANGLO	70678	69267	63739	91753	82508
DRD	144550	73413	54536	139030	139030
GFIELDS	82560	68014	26938	80121	81770
HARMONY	110500	76824	64306	113930	103260
IMPLATS	68225	65917	27011	74953	74640

Table 10.3: Value at Risk estimated as 5% quantile for one-week loss in Rands: Portfolio 2

	GARCH	O-GARCH		Modified Cholesky	
		Alexander	Modified	Max. VaR.	Coeff. of Det.
ANGLO	100780	98846	91255	126040	114930
DRD	198500	102480	76474	193010	189770
GFIELDS	117880	98001	41180	114550	116800
HARMONY	151470	105620	88397	157560	145110
IMPLATS	99102	95939	42144	108300	107870

Table 10.4: Value as Risk estimated as 1% quantile for one-week loss in Rands: Portfolio 2

that if the 5% VaR is estimated by a certain model, only 5% of values would be expected to exceed the VaR estimate. In a time series of length T , the P th quantile should only be exceeded by $T \times P$ observations.

This concept leads to the VaR model testing technique of backtesting. The VaR estimates in Tables 10.1, 10.2, 10.3 and 10.4 were divided by one million to give the return that should only be exceeded $P \times 100\%$ of the time, which was compared with the observed data set of returns and the number of times these were exceeded.

As shown in Tables 10.5 and 10.6 for Portfolio 1, the expected number of exceedences for the 5% quantile for Portfolio 1, consisting of eight time series of length 360, is $360 \times 0.05 = 18$. For Portfolio 2, with five time series of length 358, this number is also approximately eighteen exceedences at the 5% level and approximately four for the 1% level. The number of exceedences for Portfolio 2 are shown in Tables 10.7 and 10.8.

Tables 10.5 and 10.6 show that actual exceedences usually far outnumber those expected, suggesting that the data contain far more extreme values than predicted. This observation could be due to the fact that the given models do not possess fat enough tails to adequately model the extreme events of financial markets. A possible remedy for this problem is to use a non-normal distribution such as the Student-t or Negative Inverse Gaussian to model share volatility, but this is beyond the scope of this thesis.

In order to compare the ability of different volatility models to correctly

	EGARCH	O-GARCH		Modified Cholesky	
		Alexander	Modified	Max. VaR.	Coeff. of Det.
ABSA	13	31	44	26	22
EDCON	15	44	46	22	19
GRINROD	37	58	83	24	27
GROUP5	41	58	77	41	28
JDGROUP	48	38	48	26	34
LIBERTY	33	19	33	23	41
SABREW	54	25	39	52	37
SHOPRITE	49	22	46	21	33

Table 10.5: Number of log-return exceedences per stock from the $VaR_{0.05}$ level: Portfolio 1

estimate VaR, a sum of square statistic was computed by taking the sum of squares of observed less expected exceedences divided by expected exceedences. This resembles the χ^2 goodness of fit test statistic. Table 10.9 shows that the

	EGARCH	O-GARCH		Modified Cholesky	
		Alexander	Modified	Max. VaR.	Coeff. of Det.
ABSA	6	13	22	13	12
EDCON	9	22	23	12	12
GRINROD	20	36	64	12	14
GROUP5	23	37	58	22	14
JDGROUP	25	17	29	10	11
LIBERTY	14	9	14	11	17
SABREW	34	9	22	31	19
SHOPRITE	29	4	21	3	8

Table 10.6: Number of log-return exceedences per stock from the $VaR_{0.01}$ level: Portfolio 1

	GARCH	O-GARCH		Modified Cholesky	
		Alexander	Modified	Max. VaR.	Coeff. of Det.
ANGLO	41	43	52	22	30
DRD	13	67	93	17	17
GFIELDS	38	52	118	41	39
HARMONY	16	51	67	14	26
IMPLATS	38	39	94	29	30

Table 10.7: Number of log-return exceedences from the $VaR_{0.05}$ level per stock: Portfolio 2

	GARCH	O-GARCH		Modified Cholesky	
		Alexander	Modified	Max. VaR.	Coeff. of Det.
ANGLO	11	12	23	5	6
DRD	4	40	67	4	4
GFIELDS	12	24	94	14	14
HARMONY	7	21	39	7	7
IMPLATS	19	19	70	15	15

Table 10.8: Number of log-return exceedences from the $VaR_{0.01}$ level per stock: Portfolio 2

	EGARCH	O-GARCH		Modified Cholesky	
		Alexander	Modified	Max. VaR.	Coeff. of Det.
VaR _{0.05} Portfolio 1	239.22	250.61	639.78	105.50	87.17
VaR _{0.01} Portfolio 1	678.00	684.30	2134.70	337.00	196.80
VaR _{0.05} Portfolio 2	75.4	317.3	1386.6	37.9	44.1
VaR _{0.01} Portfolio 2	86.7	568.5	4502.7	57.8	58.5

Table 10.9: Sum of squares of (observed exceedences-expected exceedences)/expected exceedences for different VaR levels

Modified O-GARCH model performed worst with regards to VaR prediction in both portfolios, while the Modified Cholesky decomposition volatility model best predicted VaR with both orderings.

Although backtesting is one of the only techniques available to test VaR predictive ability of models, it is flawed. VaR and backtesting involve testing whether a given CDF $F_i(x)$ fits well in the case of extreme events; therefore only testing whether an approach models the tails of the distribution of VaR well, regardless of how well the given distribution models the non-extreme returns.

Chapter 11

Conclusions

This thesis set out to evaluate the one-step-ahead forecasting performance of novel matrix decomposition multivariate volatility models, using forecasts based on the more established univariate volatility models as benchmarks for comparison.

As described in Chapter 5, the GARCH, IGARCH and EGARCH univariate models and mean equations were fitted to each stock within the two selected portfolios using seven years of weekly data, found by way of simulation to minimize parameter estimation error. The EGARCH and GARCH models performed best for Portfolios 1 and 2 respectively, based on forecasting performance as quantified by Total Forecasting Error (Total FE). However, the slight performance advantage of the EGARCH model was outweighed by its inability to adequately fit many of the stock data sets many of the stocks due to insignificant leverage effects in the parameter estimates, rendering the GARCH model, which provided a good fit for all the stocks, more useful in general.

Chapter 9 used the Total norm error measure, based on the Frobenius norms, to test the covariance matrix forecasts of multivariate matrix decomposition volatility models, namely the spectral decomposition Alexander O-GARCH and Modified O-GARCH models, and Cholesky decomposition models, focusing on the Modified Cholesky decomposition volatility model. Upon application of an algorithm to implement the Total FE and Total norm error measures to each of 120 possible stock input orderings for the Modified Cholesky decomposition volatility model on Portfolio 2, the most efficient order for the covariance matrix was in descending order of market impact. Only the order specifications based on the sum of coefficients of determination and the maximum variance of stocks were applied to Portfolio 1 for the Modified Cholesky decomposition volatility model due to limitations on computational capacity.

All multivariate models fitted in this thesis greatly outperformed the univariate

models with respect to one-step-ahead forecasting ability of unique risks based on Total FE. Although this cannot be extrapolated to all situations and stock portfolios, the possibility that multivariate decomposition models have an advantage over univariate models when forecasting innovation series one time-step into the future bears mention and further research is recommended. The variance forecasting of the O-GARCH models outperformed the other formulations in the highly-correlated Portfolio 2 and in the CCC(1,1) simulations. The Modified Cholesky decomposition volatility model most accurately forecast variance, for orderings based on both coefficient of determination and maximum variance in the larger, less-correlated Portfolio 1. The Total norm error of the Modified Cholesky decomposition volatility model was comparable to those of the O-GARCH volatility models in forecasting between-stock covariance in Portfolio 1, and better than the O-GARCH models for Portfolio 2. As stock input ordering was not built into the CCC(1,1) simulated data sets in Chapter 8 no difference in forecasting ability was found for any specified ordering, and Modified Cholesky decomposition was inferior to the other volatility models when applied to the CCC(1,1) simulated data sets.

When comparing the VaR estimation ability of the best univariate volatility models and all the multivariate matrix decomposition methods using the backtesting methodology, it was found that the O-GARCH volatility models estimated stock VaR poorly compared to the best univariate models. Furthermore, the O-GARCH volatility models tested weakly predicted extreme values, although this feature is compensated for by more accurate prediction of ordinary events. The Modified Cholesky decomposition volatility model, however, provided the best backtesting VaR estimates for both the coefficient of determination and maximum variance-based orderings.

Models that estimate volatility close to the historical average can yield reasonable volatility forecasts, despite an inability to predict large deviations or extreme values. This quality is inappropriate to real-world forecasting, since it under-predicts volatility and by implication, the true risk of holding an asset over time. The Modified Cholesky decomposition model appears to be the most efficient of the volatility models examined in this thesis, since it provides the best compromise between accurate forecasting of both volatility and VaR for the data sets in this thesis displaying inherent orderings of stocks.

More in-depth further study of the Modified Cholesky decomposition model is therefore recommended, particularly with regards to the problem that altering the ordering of stocks to create the covariance matrix produces very different covariance forecasts. Research into application on larger stock portfolios would ascertain the practicality of implementing the Modified Cholesky decomposition model instead of the well-established univariate volatility models.

Principal component analysis is already an established multivariate statistical technique, which may be extended to better estimate the O-GARCH family of models. One possible avenue of study involves using factor analysis to provide a multitude of orthogonal rotations in p -space in order to generate more appropriate eigenvector estimates for the data set analyzed.

University of Cape Town

Bibliography

- [1] Engle R. Autoregressive Conditional Heteroscedasticity with estimates of the variance of UK inflation. *Econometrica*, 50:987–1008, 1982.
- [2] Andersen TG and Bollerslev T. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39(4):885–905, 1998.
- [3] Tsay RS. *Analysis of Financial Time Series*. John Wiley and Sons, 2nd edition, 2005.
- [4] Engle R. Dynamic Conditional Correlation: A simple class of multivariate GARCH models. *Journal of Business and Economic Statistics*, 20:339–350, 2002.
- [5] Bauwens L; Laurent S and Rombouts JVK;. Multivariate GARCH models: A survey. *Journal of Applied Econometrics*, 21:79–109, 2001.
- [6] Hansen PR and Lunde A. A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Economics*, 20:873–889, 2005.
- [7] Markowitz H. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.
- [8] Winston WL. *Operations Research: Applications and Algorithms*. Thompson, 3rd edition, 1994.
- [9] Steward J. *Calculus: Concepts and Contexts*. Brooks/Cole, 1st edition, 1998.
- [10] Ruppert D. *Statistics and Finance: An Introduction*. Springer, 2005.
- [11] Hull JC. *Options, Futures and Other Derivatives*. Prentice Hall, 5th edition, 2003.
- [12] Bodie Z; Kane A and Marcus AJ. *Investments*. Irwin, 3rd edition, 1996.
- [13] Sharpe WF. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442, 1964.
- [14] Ross SA. Return, risk and arbitrage. *Risk and Return in Finance*, 1st, 1976.
- [15] Allen L; Boudoukh J and Saunders A. *Understanding Market, Credit, and Operational Risk: The Value At Risk approach*. Blackwell, 1st edition, 2004.
- [16] Elton E; Gruber M; Brown S and Goetzmann W. *Modern Portfolio Theory and Investment Analysis*. John Wiley and Sons, 6th edition, 2003.
- [17] Alexander C. *Market Models: A Guide to Financial Data Analysis*. John Wiley and Sons, 2001.
- [18] Bollerslev T. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.
- [19] Shumway RH and Stoffer DS. *Time Series Analysis and its Applications*. Springer, 2nd edition, 2000.
- [20] Tsay RS. *Analysis of Financial Time Series*. John Wiley and Sons, 1st edition, 2002.

- [21] Nelson DB. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59:349–370, 1991.
- [22] Bollerslev T; Engle RF and Wooldridge JM. A capital asset pricing model with time varying covariances. *Journal of Political Economy*, 96:116–131, 1988.
- [23] SAS, www.sas.com. The Varimax Procedure.
- [24] Engle RF and Kroner KF. Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11:122–150, 1995.
- [25] Engle R. New frontiers for ARCH models. *Applied Econometrics*, 17:425–446, 2002.
- [26] Bollerslev T. Modeling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *Review of Economics and Statistics*, 72:498–505, 1990.
- [27] JP Morgan and Reuters, <http://www.jpmorgan.com/RiskManagement/RiskMetrics/RiskMetrics.html>. RiskMetrics Technical Document, December 1996.
- [28] Engle R and Shepard K. Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Discussion paper: University of California at San Diego, 2001.
- [29] Manly B. *Multivariate Statistical Methods*. Chapman and Hall, 2nd edition, 1986.
- [30] Alexander CO and Chibumba AM. Multivariate orthogonal factor GARCH. Mimeo, University of Sussex, 1997.
- [31] Alexander C. A primer on the orthogonal GARCH model. ISMA Centre, 2:1–38, 2001.
- [32] Vrontos ID; Dellaportas P and Politis N. A full-factor multivariate GARCH model. *Econometrics Journal*, 6:312–334, 2003.
- [33] Gerald CF and Wheatley PO. *Applied Numerical Analysis*. Addison-Wesley, 6th edition, 1999.