

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.

Volatility Forecasting using Double-Markov Switching GARCH Models under Skewed Student- t Distribution

Author: Batsirai Winmore Mazviona

Supervisor: Mr Allan Clark

Department of Statistical Sciences

A mini-thesis submitted at University of Cape Town in partial fulfillment
of the requirements for the degree of Masters of Philosophy in Mathematical Finance

10 February 2012

Abstract

This thesis focuses on forecasting the volatility of daily returns using a double Markov switching GARCH model with a skewed Student- t error distribution. The model was applied to individual shares obtained from the Johannesburg Stock Exchange (JSE). The Bayesian approach which uses Markov Chain Monte Carlo was used to estimate the unknown parameters in the model. The double Markov switching GARCH model was compared to a GARCH(1,1) model. Value at risk thresholds and violations ratios were computed leading to the ranking of the GARCH and double Markov switching GARCH models. The results showed that double Markov switching GARCH model performs similarly to the GARCH model based on the ranking technique employed in this thesis.

Acknowledgement

Sir Isaac Newton, a great scientist once said ‘if I have been able to see further than others it is because I was able to stand on the shoulders of giants.’ There are many ‘giants’ whose efforts and contribution deserve to be mentioned in this work. First and foremost, I would like to thank the Lord for His guidance in my life and in the preparation of this work, I know very well that it would have been impossible without His guidance. Special thanks goes to Mr Allan Clark for his invaluable comments, guidance and supervision during completion of this thesis. Lastly I thank my family in particular Dr E. Moyo and Mrs P. Moyo and my friends Stewart, Kakanyo and Trevor for the support and encouragement throughout my studies at the University of Cape Town.

Plagiarism declaration

1. I know that plagiarism is wrong. Plagiarism is using other people works and to pretend that it is ones own.
2. I have used the Harvard system as the convention for citation and referencing. Each significant contribution to, and quotation in this thesis from the work, or works of other people has been attributed and has cited and referenced.
3. This thesis is my own work with supervisory assistance from my supervisor.
4. 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.
5. I acknowledge that copying someone else's thesis or part of it, is wrong, and declare that this is my own work.

Signature:..... Date:.....

Contents

1	Introduction	2
1.1	Background of study	2
1.2	Aims and objectives	3
1.3	Scope of study	3
1.4	Limitation of the study	3
1.5	Significance of study	4
1.6	Layout	4
2	Preliminaries	5
2.1	ARMA	5
2.2	GARCH	5
2.3	MS-ARCH	7
2.4	MS-GARCH	8

2.5	DMS-GARCH Model	8
2.6	Bayesian inference	9
2.7	Estimation of GARCH Models	10
2.7.1	QMLE	11
2.7.2	MCMC	11
2.8	Likelihood function for the DMS-GARCH model	13
2.8.1	Prior Distributional Assumptions	14
2.8.2	Conditional Posterior Distributions	15
2.9	An application of the DMS-GARCH Model	17
2.10	Assessing VaR models	17
3	Literature Review	19
4	Results	22
4.1	Data	22
4.2	Model description	23
4.3	Model estimation and convergence diagnostics	24
4.4	Forecasting performance	26
5	Conclusions	34

<i>CONTENTS</i>	vi
A Appendix	35
A.1 Probability Densities	35
A.1.1 Normal	35
A.1.2 Student- t	35
A.1.3 Skewed Student- t	36
Bibliography	38

University of Cape Town

List of Figures

4.1	Simulations from posterior distribution of ν (last 15 000 simulations) for AMS	25
4.2	Example illustrating posterior draws of some of the DMS-GARCH(1,1) parameters for AMS	27
4.3	One day ahead volatility forecasts for AMS using a GARCH(1,1) model . .	28
4.4	One day ahead volatility forecasts for AMS using a DMS-GARCH(1,1) model	29

List of Tables

4.1	Summary of statistics of daily returns	23
4.2	Posterior mean of the unknown parameters for the GARCH SS- t model . .	26
4.3	Posterior mean of the unknown parameters for the DMS-GARCH model .	30
4.4	Posterior mean of the unknown parameters for the DMS-GARCH model .	31
4.5	Ratio $(\frac{VR}{\alpha})$	32
4.6	Ranking models	33

1. Introduction

1.1 Background of study

Volatility is a measure of the variability of financial time series data. Volatility forecasting is a time series concept whose objective is to use financial data obtained from financial markets to make rational decisions about the future. The concept of volatility dates back to the work of Markowitz (1952). The standard deviation of returns was used as a measure of the downside risk. However, this is only valid if asset returns are jointly normally distributed. Empirical evidence in financial literature suggest that financial returns data tend to exhibit some stylised facts. These facts about financial returns data were derived from Mandelbrot (1963) and documented. These include the following:

1. Volatility clustering present in the evolution of asset returns. This entails that asset returns tend to fluctuate when a either a cluster of high or low volatility occurs.
2. Volatility of asset returns is heteroscedastic implying that it changes over time. This is due to the stochastic nature of asset prices triggered by varying economic conditions.
3. The asset returns exhibit skewness and excess kurtosis (heavy tails) indicating the insufficient of normal distribution in modelling volatility.
4. Volatility of asset returns in particular share returns is inversely related to the performance of financial markets. For instance, it tends to be high when the market is on a downward trend. This asymmetric effect to positive and negative shocks has been documented in Glosten et al (1993).

Various models have been developed to forecast volatility. Straumann (2005) argues that every model for analysing financial time series data is likely to be incorrect. However, there are good and bad models. The clear distinction on the goodness of a model lies on how well it captures the observed data and reality. The main categories of volatility models are the GARCH-type models and the stochastic volatility (SV) models. The difference between the GARCH and SV models is that the latter models volatility as a random and

unobservable process whereas the former describes volatility as a deterministic function. According to Rachev et al (2008), GARCH-type models have become popular in forecasting volatility due to their analytical tractability. The following paper uses a double Markov switching GARCH model with a skewed Student- t error distribution to forecast share return volatility and the estimate Value at Risk of a financial time series.

1.2 Aims and objectives

The research specifically addressed the following objectives:

1. To forecast volatility of selected shares on the Johannesburg Stock Exchange (JSE) using the DMS-GARCH model presented in chapter 2.
2. To forecast Value at Risk (VaR) of selected shares on the JSE.
3. To compare the GARCH(1,1) and DMS-GARCH(1,1) models and decide which performs better at forecasting the value at risk.

1.3 Scope of study

This paper focus on the South African financial market. In particular we extract data for the top ten stocks from the JSE over the period 14 January 2002 to 13 January 2012.

1.4 Limitation of the study

The results are only valid for the particular data set and time period that was chosen and further studies should be undertaken to generalise the results.

1.5 Significance of study

Forecasting VaR may assist the South African Reserve Bank to determine the appropriate amount of capital which banks should allocate to protect against market risk and stock price volatility. Portfolio managers are mainly involved in portfolio rebalancing of their client's portfolios. This study is an addition to the portfolio management toolbox and therefore of use in managing risk at a portfolio level. Portfolio managers can decide which stocks to select based on the risk profile determined through volatility forecasting. Option traders are interested in volatility forecasting so that they can price derivatives fairly.

1.6 Layout

The paper is organised as follows; Chapter 2 presents a description of volatility models, estimation methods and the model formulation, Chapter 3 provides a small literature review, Chapter 4 discussed results of an empirical study and finally Chapter 5 provides the conclusion and discusses future areas of research.

2. Preliminaries

This chapter provides a background on volatility models which include the generalized autoregressive conditional heteroscedasticity (GARCH) models found in time series literature. It also provide estimation approaches for determining unknown parameters of volatility models.

2.1 ARMA

The autoregressive moving average (ARMA) model is one of the foundation of traditional time series analysis. ARMA model is a hybrid of the autoregressive (AR) and the moving average (MA) models. The ARMA models are important since they are the building block used to define GARCH models. A stationary process u_t is said to be an ARMA(k, g) process if:

$$u_t = \alpha_1 u_{t-1} + \dots + \alpha_k u_{t-k} + \beta_1 \epsilon_{t-1} + \dots + \beta_g \epsilon_{t-g} + \epsilon_t \quad (2.1)$$

where k, g are positive integers and (k, g) identifies the order of the ARMA model. ϵ_t is an independent and identical (iid) sequence having a distribution $f(\epsilon_t)$ with a mean of zero and variance σ^2 . ARMA models are favoured since they adapt well when modelling stationary processes. However, Fan and Yao (2003) highlights that ARMA models are not desirable in a nonlinear setting. According to Fan and Yao (2003), the term “nonlinear” encompasses features such as non-Gaussian, asymmetric effect, nonlinear relationship of lagged variables and forecasting over specified periods. The mentioned features motivates the use of GARCH-type models which is discussed in the following sections. Francq and Zakoian (2010) praised nonlinear models for providing better forecasts given the stylised facts of financial time series data (see section 1.1).

2.2 GARCH

The GARCH model developed by Bollerslev (1986) originated from the study by Engle (1982) on autoregressive conditional heteroscedasticity (ARCH). Let ϵ_t be an independent

and identical (iid) sequence having a distribution $f(\epsilon_t)$ with a mean of zero and variance of one. A process u_t is said to be an ARCH(k, g) process with respect to ϵ_t if:

$$\begin{aligned} u_t &= \sqrt{h_t} \epsilon_t \\ h_t &= \sigma_t^2 = \text{Var}(u_t | u_s, s < t) = \alpha_0 + \sum_{i=1}^g \alpha_i u_{t-i}^2 \end{aligned} \quad (2.2)$$

where h_t is the conditional variance of a series and the constants α_i is strictly positive and defined for $i = 0, \dots, g$. The order of the ARCH model is g . The volatility described by equation 2.2 is heteroscedastic implying that the conditional variance, h_t is not constant. In the ARCH model, volatility is defined by weights of squared past observations of the financial time series data. These weights that is, the α_i 's need to be estimated by methods to be discussed later on. A key note for an ARCH case is that larger values of the observations (financial data) would result in larger volatility (more "jumpiness"). Bollerslev (1986) found that it was difficult to determine the appropriate number of ARCH lags while ensuring that the α_i parameters remain positive. This then motivated the introduction of the GARCH model which allowed more parsimonious modelling of volatility. The definition is as follows:

Definition 2.2.1 *Let ϵ_t be an independent and identical (iid) sequence having a distribution $f(\epsilon_t)$ with zero mean and variance of one. A process u_t is said to be a strong GARCH(k, g) process with respect to ϵ_t if:*

$$\begin{aligned} u_t &= \sqrt{h_t} \epsilon_t \\ h_t &= \alpha_0 + \sum_{i=1}^g \alpha_i u_{t-i}^2 + \sum_{j=1}^k \beta_j h_{t-j} \end{aligned} \quad (2.3)$$

where the constants α_i and β_j defined for $i = 0, \dots, g$ and $j = 1, \dots, k$ are strictly positive.

The necessary and sufficient condition for (2.3) defining a unique strictly stationary process u_t with $E(u_t^2) < \infty$ is:

$$\sum_{i=1}^g \alpha_i + \sum_{j=1}^k \beta_j < 1$$

Furthermore $E(u_t) = 0$, $\text{Cov}(u_t, u_s) = 0$ for $s < t$ and $\text{Var}(u_t) = \frac{\alpha_0}{1 - \sum_{i=1}^g \alpha_i - \sum_{j=1}^k \beta_j}$.

In addition $E(u_t^4) < \infty$, provided $\max(1, E(\epsilon_t^4)^{0.5}) \cdot \frac{\sum_{i=1}^g \alpha_i}{1 - \sum_{j=1}^k \beta_j} < 1$.

The GARCH model in equation 2.3 differs from the ARCH model because it adds extras term known as the GARCH terms, $\sum_{j=1}^k \beta_j h_{t-j}$, which is the weighted sum of past conditional variances. The extras term ensures a parsimonious model and often it is found that only one GARCH term is appropriate. The value of α_0 is the average conditional variance in the long run. The GARCH model introduced can improve the modelling of the evolution of the financial time series data. For example, in stock markets, when the markets are falling investors tend to panic which translate to higher volatility and the inverse is true when markets are rising. This phenomenon can be explained by refining the GARCH model so that it captures the regime effect. This ensures that volatility is not modelled using a single structure over a specified period. For this reason, the models to follow provides a discussion of the innovations that take into account the regime effects.

2.3 MS-ARCH

The Markov switching autoregressive conditional heteroscedasticity also known as MS-ARCH model was introduced by Cai (1994). The model was a synthesis of Engle (1982) ARCH model and Hamilton (1988) MS model. The MS endeavoured to capture the effects of economic shocks on financial time series data. A process u_t is said to follow an MS-ARCH(g) process if:

$$\begin{aligned} u_t &= \sqrt{h_t} \epsilon_t \\ h_t &= \kappa(s_t) + \sum_{i=1}^g \alpha_i u_{t-i}^2 \end{aligned} \quad (2.4)$$

where ϵ_t is an iid sequence having a mean of zero and variance of one, $\kappa(s_t) = c_0 + c_1 s_t$, $c_0 > 0$ and $c_1 > 0$. $s_t = 0$ or 1 denotes the unobserved states. The latent variable s_t is assumed to be stationary and of first-order Markov process with transition probabilities:

$$f(s_t = j \mid s_{t-1} = i) = \lambda_{ij} \quad (2.5)$$

where $\lambda_{11} + \lambda_{12} = 1$, $\lambda_{21} + \lambda_{22} = 1$ and $0 < \lambda_{ij} < 1$.

The key feature in the MS-ARCH(g) model is the term $\kappa(s_t)$ which enables the volatility to switch regimes or states. The motivation of this term can be explained by the fact that high and low volatility are associated with negative and positive economic shocks

respectively (see Cai (1994)). The MS-ARCH model was tuned to form the MS-GARCH model and is described in the next section.

2.4 MS-GARCH

Gray (1996) extended the MS-ARCH model to define a Markov switching generalized autoregressive conditional heteroscedasticity known as the MS-GARCH model. The motivation for the formulation in modelling volatility was to capture volatility clustering present in financial data. Gray (1996) also argues that the MS-GARCH is able to model nonlinearity in financial time series data through the inclusion of regimes. A process u_t is said to follow an MS-GARCH(k, g) process with respect to ϵ_t , an iid sequence if:

$$\begin{aligned} u_t &= \sqrt{h_t^{(j)}} \epsilon_t \\ h_t^{(j)} &= \alpha_0^{(j)} + \sum_{i=1}^g \alpha_i^{(j)} u_{t-i}^2 + \sum_{i=1}^k \beta_i^{(j)} h_{t-i} \end{aligned} \quad (2.6)$$

where h_{t-i} is past conditional variance, j is an element of natural numbers. The superscript j denotes the current state of the process. The determination of the states are governed by similar probabilistic formulation in equation 2.5. Marcucci (2005) spelt out that the MS-GARCH model by Gray (1996) does not provide simple expressions because of the tendency to have multi-step ahead volatility forecasts which are difficult to calculate recursively as provided in basic GARCH models. For this reason another variation of the model is discussed on the next section.

2.5 DMS-GARCH Model

Chen et al (2009) assumes that the error process follows a t distribution. However, this section provides an outline of the model developed in Chen et al (2009) and thereafter we extend the model to include a Skewed Student- t distribution as the residual distribution process in order to model non-normality of financial time series data (see section 1.1 item 3). The model has different parameters for different regimes or states. The number of states used is the same as in Chen et al (2009). This sufficiency of two states in carrying

out volatility forecasts is supported by Klaassen (2002).

Let r_t be the return data, h_t the conditional volatility of $r_t | r_1, \dots, r_{t-1}$ and x_t be an exogenous variable observed at time t . The latent state variables are denoted by the sequence $\{s_t\}$ and is assumed to be a stationary, irreducible Markov process with discrete state space $\{1, 2\}$ and transition probabilities:

$$f(s_t = j | s_{t-1} = i) = \lambda_{ij} \quad (2.7)$$

where $\lambda_{11} + \lambda_{12} = 1$, $\lambda_{21} + \lambda_{22} = 1$ and $0 < \lambda_{ij} < 1$.

Define the DMS-GARCH model as follows:

$$r_t = \begin{cases} \phi_0^{(1)} + \sum_{i=1}^p \phi_i^{(1)} r_{t-i} + \sum_{i=1}^q \varphi_i^{(1)} x_{t-i} + u_t & \text{if } s_t = 1, \\ \phi_0^{(2)} + \sum_{i=1}^p \phi_i^{(2)} r_{t-i} + \sum_{i=1}^q \varphi_i^{(2)} x_{t-i} + u_t & \text{if } s_t = 2, \end{cases} \quad (2.8)$$

$$u_t = \sqrt{h_t} \epsilon_t \quad \text{and} \quad \epsilon_t \sim D(0, 1) \quad (2.9)$$

$$h_t = \begin{cases} \alpha_0^{(1)} + \sum_{i=1}^g \alpha_i^{(1)} u_{t-i}^2 + \sum_{i=1}^k \beta_i^{(1)} h_{t-i} & \text{if } s_t = 1, \\ \alpha_0^{(2)} + \sum_{i=1}^g \alpha_i^{(2)} u_{t-i}^2 + \sum_{i=1}^k \beta_i^{(2)} h_{t-i} & \text{if } s_t = 2, \end{cases} \quad (2.10)$$

where ϵ_t are independent and identically distributed error terms from a distribution with mean 0 and variance 1. The variance parameters are restricted to ensure stationarity and positivity as follows:

$\alpha_0^{(j)} > 0$, $\alpha_i^{(j)}, \beta_i^{(j)} \geq 0$ and $0 < \sum_{i=1}^g \alpha_i^{(j)} + \sum_{i=1}^k \beta_i^{(j)} < 1$ for $j=1,2$. Before proceeding to the formulation of the DMS-GARCH with a Skewed Student- t distribution we provide a description of Bayesian inference and estimation methods for GARCH models.

2.6 Bayesian inference

Let y_1, y_2, \dots, y_n be the data (see Rachev et al (2008)). This may be an observed share returns series. Suppose the observed data can be described by a probability density function $\pi(y_1, y_2, \dots, y_n | \boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is a parameter vector specific to the density function. $\pi(y_1, y_2, \dots, y_n | \boldsymbol{\theta})$ is treated as a function of the unknown parameter $\boldsymbol{\theta}$ and is known as the likelihood function. In Bayesian analysis, $\boldsymbol{\theta}$ is random variable described by a prior

density function denoted $\pi(\boldsymbol{\theta})$. The prior density may depend on one or more parameters termed hyperparameters. Bayes rule is applied to define the posterior density of $\boldsymbol{\theta}$ which is denoted as $\pi(\boldsymbol{\theta} \mid y_1, y_2, \dots, y_n)$ after observing the data. The evaluation of the posterior might become a problem when there are many parameters or when the posterior is not of a known distribution type. However, the posterior distribution can be evaluated using a sampling method known as Markov Chain Monte Carlo (MCMC) simulation. This is discussed in section 2.7.2.

2.7 Estimation of GARCH Models

The following section presents two methods for estimating the parameters of GARCH type models. The popular and most widely used method is the quasi maximum likelihood estimator (QMLE) while Markov Chain Monte Carlo (MCMC) simulations could be used as an alternative estimation method. Other alternative methods can be found in the literature. These include Whittle's method which is a close proxy to the Gaussian log likelihood. Ardia (2008) recommends the MCMC method as it avoids local maximum encountered when using the QMLE for estimating MS-GARCH model. This emanates from the fact that maximisation of the QMLE function must be done in a restricted and optimal way while ensuring that the parameters being estimated and the conditional variance are positive. The problem that arises is of achieving convergence when regime switching models such as the MS-GARCH and the DMS-GARCH are used. This can however be tackled by Bayesian MCMC methods as they are able to incorporate the restrictions through specifications of appropriate prior distributions. Ardia (2008) also argues that the MCMC methods are able to explore the joint posterior distribution of model parameters as compared to the QMLE because posterior draws are done for the models parameters until a stationary state is reached. In addition, the true distribution of the parameters in the regime switching model can be attained through MCMC simulations from the joint posterior distributions.

2.7.1 QMLE

Lets consider the GARCH(k, g) process given in Definition 2.2.1 with n observations. The order (k, g) and the realisation on u_t are assumed to be known (see Francq and Zakoian (2010)). In addition, let the parameter vector be $\mathbf{Z} = (\alpha_0, \dots, \alpha_g, \beta_1, \dots, \beta_k)'$ where \mathbf{Z} is an element of V which is a set $V \subset (0, \infty) \times [0, \infty)^{k+g}$. The true parameter vector is unknown and is defined as $\bar{\mathbf{Z}} = (\bar{\alpha}_0, \dots, \bar{\alpha}_g, \bar{\beta}_1, \dots, \bar{\beta}_k)'$. For illustrative purposes, suppose $\epsilon_t \sim N(0, 1)$ and the initial values $u_0, \dots, u_{1-g}, h_0, \dots, h_{1-g}$ are specified. The quasi-likelihood is given by

$$L_n(\mathbf{Z}) = L_n(\mathbf{Z}; u_1, \dots, u_n) = \prod_{t=1}^n \frac{1}{\sqrt{2\pi h_t}} \exp\left(\frac{-u_t^2}{2h_t}\right) \quad (2.11)$$

where h_t is defined for $t \geq 1$ by

$$h_t = h_t(\mathbf{Z}) = \alpha_0 + \sum_{i=1}^g \alpha_i u_{t-i}^2 + \sum_{j=1}^k \beta_j h_{t-j} \quad (2.12)$$

The QMLE of \mathbf{Z} is a solution $\hat{\mathbf{Z}}_n$ such that

$$\hat{\mathbf{Z}}_n = \arg \max_{\mathbf{Z} \in V} L_n(\mathbf{Z}) \quad (2.13)$$

Equivalently, by taking logarithms both sides in equation 2.13 and defining the log-likelihood as $l_n(\mathbf{Z})$, the same solution for the QMLE of \mathbf{Z} is given by

$$\hat{\mathbf{Z}}_n = \arg \min_{\mathbf{Z} \in V} l_n(\mathbf{Z}) \quad (2.14)$$

2.7.2 MCMC

The two commonly used Markov Chain Monte Carlo (MCMC) methods are the Metropolis-Hastings and the Gibbs Sampler. The MCMC involves generating samples of defined parameter space from a complex posterior probability distribution. The previous values generated are then used to generate the next values. The generated samples enable Bayesian inference to be carried out from the set of posterior distributions.

2.7.2.1 M-H Algorithm

The Metropolis-Hasting (M-H) algorithm emanates from the work pioneered by Metropolis et al (1953) and later re-engineered by Hastings (1970). Consider a non-standardised posterior density function $\pi(\mathbf{Z} \mid \mathbf{y})$ which is intractable for sampling procedures. Let \mathbf{Z} be a Q -dimensional parameter vector, $\mathbf{Z} = (Z_1, Z_2, \dots, Z_q)'$ and \mathbf{y} is the data series under consideration. Let a proposed density be denoted by $q(\mathbf{Z} \mid \mathbf{Z}^{(t-1)})$. The steps involved in M-H algorithm are as follows:

1. Initialise a value $\mathbf{Z}^{(0)}$ from the parameter space \mathbf{Z} ;
2. Draw a value \mathbf{Z}^* , from $q(\mathbf{Z} \mid \mathbf{Z}^{(t-1)})$ at t ;
3. Accept \mathbf{Z}^* with probability $U\left(\mathbf{Z}^*, \mathbf{Z}^{(t-1)}\right) = \min\left(1, \frac{\pi(\mathbf{Z}^* \mid \mathbf{y})/q(\mathbf{Z}^* \mid \mathbf{Z}^{(t-1)}, \mathbf{y})}{\pi(\mathbf{Z}^{(t-1)} \mid \mathbf{y})/q(\mathbf{Z}^{(t-1)} \mid \mathbf{Z}^*, \mathbf{y})}\right)$
4. Draw a from a uniform distribution between 0 and 1.
If $a \leq U\left(\mathbf{Z}^*, \mathbf{Z}^{(t-1)}\right)$ then $\mathbf{Z}^{(t)} = \mathbf{Z}^*$ otherwise $\mathbf{Z}^{(t)} = \mathbf{Z}^{(t-1)}$;
5. Return to step 2.

The above steps are repeated a finite large number of times until the Markov Chain converges. Yu and Mykland (1994) used a cumsum convergence monitoring to diagnose convergence. It involves making a visual inspection of the plot of the standardised posterior means of the parameters being estimated against iterations after the transient or burn-in period. The burn-in period being defined as the phase in which estimation of parameters is discarded. The statistic is as follows

$$CS_{i,m} = \frac{1}{m} \frac{\sum_{j=1}^m (\theta_i^{(j)} - \hat{\theta}_i)}{\hat{\sigma}_i}$$

for $m = 1, \dots, M$ where M is the number of simulations which are not discarded, $\hat{\theta}_i$ and $\hat{\sigma}_i$ are the posterior mean and standard deviations respectively of the unknown parameters being estimated in a model. When the statistic is close to 0, it suggests that the Markov Chain have converged.

2.7.2.2 Gibbs Sampler

The Gibbs sampler is a special case of the M-H algorithm deriving its christened name from the work of Geman and Geman (1984). However, the distinction is that there is a **full** conditional posterior distribution of which sampling is possible. Denote the Q -dimensional parameter vector, partitioned into c components as $\mathbf{Z} = (Z_1, Z_2, \dots, Z_c)'$ and the posterior distribution of $Z_i, i = 1, \dots, c$ as $\pi(Z_i | Z_1, \dots, Z_{i-1}, Z_{i+1}, \dots, Z_c, \mathbf{y}) = \pi(Z_i | \mathbf{Z}_{(-i)}, \mathbf{y})$.

The steps involved in the Gibbs Sampler are as follows:

1. Initialise values $Z_i^{(0)}$ for components $i = 1, \dots, c$;
2. Draw $\mathbf{Z} = (Z_1, Z_2, \dots, Z_c)'$ and adjust recursively the components by: Drawing an observation $Z_1^{(t)}$ from $\pi(Z_1 | \mathbf{Z}_{(-1)}, \mathbf{y})$ and similarly for the remaining components Z_2, \dots, Z_c ;
3. Return to step 2 and repeat the process several times until the Markov Chains have all converged.

2.8 Likelihood function for the DMS-GARCH model

The framework used is similar to Chen et al (2009). Denote the returns data as $\mathbf{r}^{(l+1,n)} = (r_{l+1}, \dots, r_n)'$, the state vector $\mathbf{S}^{(l+1,n)} = (s_{l+1}, \dots, s_n)'$. Let $l = \max\{p, q, g, k\}$ and denote the mean equation parameter vector as

$\boldsymbol{\phi}_j = (\phi_0^{(j)}, \dots, \phi_p^{(j)}, \varphi_1^{(j)}, \dots, \varphi_q^{(j)})'$ and the variance equation parameter vector as

$\boldsymbol{\alpha}_j = (\alpha_0^{(j)}, \dots, \alpha_g^{(j)}, \beta_1^{(j)}, \dots, \beta_k^{(j)})'$. ν and η are defined as the degrees of freedom and asymmetry parameters respectively. Let $\boldsymbol{\Xi}$ be the unknown parameter vector defined as $\boldsymbol{\Xi} = (\boldsymbol{\phi}'_1, \boldsymbol{\phi}'_2, \boldsymbol{\alpha}'_1, \boldsymbol{\alpha}'_2, \lambda_{11}, \lambda_{22}, \nu, \eta)'$.

The likelihood function for the DMS-GARCH model is:

$$f\left(\mathbf{r}^{(l+1,n)} | \mathbf{S}^{(l+1,n)}, \boldsymbol{\Xi}\right) = \prod_{t=l+1}^n \left[\frac{1}{\sqrt{h_t}} f_\epsilon\left(\frac{r_t - \omega_t}{\sqrt{h_t}}\right) \right] \quad (2.15)$$

where

$$\omega_t = \phi_0^{(j)} + \sum_{i=1}^p \phi_i^{(j)} r_{t-i} + \sum_{i=1}^q \varphi_i^{(j)} x_{t-i}$$

for $j = 1, 2$ and h_t is defined in (4.4).

In the following paper we assume that ϵ_t follows a Skewed Student- t distribution (SS- t) as in Hansen (1994). The probability density function is defined as:

$$f_\epsilon(\epsilon_t | \nu, \eta) = \begin{cases} bc \left[1 + \frac{1}{\nu-2} \left(\frac{b\epsilon_t+a}{1-\eta} \right)^2 \right]^{-\frac{(\nu+1)}{2}} & \text{if } \epsilon_t < -\frac{a}{b}, \\ bc \left[1 + \frac{1}{\nu-2} \left(\frac{b\epsilon_t+a}{1+\eta} \right)^2 \right]^{-\frac{(\nu+1)}{2}} & \text{if } \epsilon_t \geq -\frac{a}{b}, \end{cases} \quad (2.16)$$

We require $2 < \nu < \infty$ and $-1 < \eta < 1$. The constants a , b and c are defined as

$$\begin{aligned} a &= 4\eta c \left(\frac{\nu-2}{\nu-1} \right) \\ b^2 &= 1 + 3\eta^2 - a^2 \\ c &= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu-2)\Gamma\left(\frac{\nu}{2}\right)}} \end{aligned}$$

It is important to note that when $\eta = 0$, the SS- t distribution reduces to a Student- t distribution and as ν becomes large the SS- t gravitates towards a Gaussian distribution.

2.8.1 Prior Distributional Assumptions

The prior distributions for the unknown parameters are as follows:

1. $\phi_j \sim N(\boldsymbol{\mu}_{\phi_j}, \boldsymbol{\Sigma}_{\phi_j})$ where $\boldsymbol{\mu}_{\phi_j}$ is the mean vector and $\boldsymbol{\Sigma}_{\phi_j}$ is the variance-covariance matrix. $\boldsymbol{\mu}_{\phi_j}$ was chosen to be a vector of zero and $\boldsymbol{\Sigma}_{\phi_j}$ as the variance covariance matrix of the observed data.
2. The conditional variance parameters are assumed to have uninformative prior distributions:

$$f(\boldsymbol{\alpha}_j) \propto I\left(0 < \alpha_0^{(j)} < z\right) I(M_j)$$

where $z > 0$ is preset by the user. The value of $\alpha_0^{(j)}$ should be positive and was chosen from a truncated normal distribution with zero mean and the sample variance of the data. To ensure $0 < \sum_{i=1}^g \alpha_i^{(j)} + \sum_{i=1}^k \beta_i^{(j)} < 1$ that is to achieve stationarity. The value of z was specified to be one.

3. We assume that $\lambda_{ii} \sim \text{Beta}(d_{i1}, d_{i2})$ where d_{i1} and d_{i2} are both chosen to be greater than 0. The values of d_{i1} and d_{i2} were chosen to be equal to ten.
4. To ensure the attainability of the first four moments of the SS- t distribution, a re-parameterisation of degrees of freedom is undertaken as in Chen et al (2009). Define $\Delta = \nu^{-1}$ and the prior distribution of Δ as $U(0, \frac{1}{4})$.
5. The prior distribution of η is assumed to be $U(-1, 1)$.

2.8.2 Conditional Posterior Distributions

The posterior distributions are defined as follows:

(i) Define $\Xi_{(-\phi_j)}$, the parameter vector Ξ without ϕ_j . The conditional posterior distribution of ϕ_j is given by:

$$f(\phi_j | \mathbf{r}^{(l+1,n)}, \mathbf{S}^{(l+1,n)}, \Xi_{(-\phi_j)}) \propto f(\mathbf{r}^{(l+1,n)} | \mathbf{S}^{(l+1,n)}, \Xi) f(\phi_j) \quad (2.17)$$

(ii) There is no closed solution for the conditional posterior distribution of the variance parameters α_j which is given by:

$$f(\alpha_j | \mathbf{r}^{(l+1,n)}, \mathbf{S}^{(l+1,n)}, \Xi_{(-\alpha_j)}) \propto f(\mathbf{r}^{(l+1,n)} | \mathbf{S}^{(l+1,n)}, \Xi) f(\alpha_j) \quad (2.18)$$

(iii) The conditional posterior distribution of transition probability λ_{ii} is given by:

$$f(\lambda_{ii} | \mathbf{r}^{(l+1,n)}, \mathbf{S}^{(l+1,n)}, \Xi_{(-\lambda_{ii})}) \propto f(\mathbf{r}^{(l+1,n)} | \mathbf{S}^{(l+1,n)}, \Xi) f(\lambda_{ii}) \quad (2.19)$$

which results in a posterior, $\lambda_{ii} \sim \text{Beta}(d_{i1} + n_{ii}, d_{i2} + n_i - n_{ii})$. for $i = 1, 2$, n_{ii} is the number of times the Markov process remains in state i at time t after having been in the

same state at time $t - 1$ and n_i is the number of times observed whilst $s_t = i$.

(iv) The conditional posterior of state s_t is

$$f\left(s_t \mid \mathbf{r}^{(l+1,n)}, \mathbf{S}^{(l+1,n)}, \Xi\right) \propto \prod_{j=t}^n f\left(r_j \mid \mathbf{r}^{(l+1,j-1)}, \mathbf{S}^{(l+1,j)}, \Xi\right) \\ \times f\left(s_t \mid s_{t-1}, \lambda_{11}, \lambda_{22}\right) f\left(s_{t+1} \mid s_t, \lambda_{11}, \lambda_{22}\right) \quad (2.20)$$

For $t = l + 1, \dots, n - 1$ and when $t = n$

$$f\left(s_n \mid \mathbf{r}^{(l+1,n)}, \mathbf{S}^{(l+1,n)}, \Xi\right) \propto f\left(r_n \mid \mathbf{r}^{(l+1,n-1)}, \mathbf{S}^{(l+1,n)}, \Xi\right) \times f\left(s_n \mid s_{n-1}, \lambda_{11}, \lambda_{22}\right) \quad (2.21)$$

The conditional posterior probability of $s_t = i$ is defined as:

$$f\left(s_t = i \mid \mathbf{r}^{(l+1,n)}, \mathbf{S}_{(-t)}^{(l+1,n)}, \Xi\right) = \frac{l(s_t = i) f(s_t = i \mid \mathbf{S}_{(-t)}^{(l+1,n)}, \Xi)}{\sum_{j=1}^2 l(s_t = j) f(s_t = j \mid \mathbf{S}_{(-t)}^{(l+1,n)}, \Xi)} \quad (2.22)$$

for $i = 1, 2$ and $t = l + 1, \dots, n$ where $l(s_t = i)$ is log-likelihood function defined under the SS- t distribution as:

$$l(s_t = i \mid \mathbf{r}^{(l+1,n)}, \mathbf{S}^{(l+1,n)}) = \begin{cases} \sum_{j=t}^n \left(-\frac{1}{2} \ln(h_j) - \frac{\nu+1}{2} \ln \left(1 + \frac{1}{\nu-2} \left(\frac{b\epsilon_j + a}{1-\eta} \right)^2 \right) \right) & \text{if } \epsilon_t < -\frac{a}{b}, \\ \sum_{j=t}^n \left(-\frac{1}{2} \ln(h_j) - \frac{\nu+1}{2} \ln \left(1 + \frac{1}{\nu-2} \left(\frac{b\epsilon_j + a}{1+\eta} \right)^2 \right) \right) & \text{if } \epsilon_t \geq -\frac{a}{b}, \end{cases} \quad (2.23)$$

where $\epsilon_j = \frac{r_j - \omega_j}{\sqrt{h_j}}$.

This is in contrast with what was done in Chen et al (2009) where a t distribution was used.

(v) We define the conditional posterior of Δ as

$$f\left(\Delta \mid \mathbf{r}^{(l+1,n)}, \mathbf{S}^{(l+1,n)}, \Xi_{-\Delta}\right) \propto f\left(\mathbf{r}^{(l+1,n)} \mid \mathbf{S}^{(l+1,n)}, \Xi\right) f(\Delta) \quad (2.24)$$

(vi) The conditional posterior of η is

$$f\left(\eta \mid \mathbf{r}^{(l+1,n)}, \mathbf{S}^{(l+1,n)}, \Xi_{-\eta}\right) \propto f\left(\mathbf{r}^{(l+1,n)} \mid \mathbf{S}^{(l+1,n)}, \Xi\right) \quad (2.25)$$

$f(\eta)$ is not given in 2.25 because it is a constant in the prior distribution. A combination of the random walk M-H algorithm and the independence chain M-H algorithm is used for the simulation of parameters in (i), (ii), (v) and (vi). The details of the procedure are provided in Chen and So (2006).

2.9 An application of the DMS-GARCH Model

In time series literature, GARCH models have been associated with forecasting volatility and determining value at risk (VaR) thresholds. In this thesis, the DMS-GARCH model is used to forecast returns, volatility and VaR. Following the Basel amendment in 1996, the VaR was adopted as a standard technical tool by risk professionals to assess the level of economic capital needed to guard against market risk. VaR is defined as the loss that could arise over a given investment period with a specified level of probability. In general, the one-period VaR at α level is given as:

$$\alpha = Pr(r_{t+1} < -VaR)$$

where r_{t+1} is the forecasted return for the period $t+1$ at time point t . In order to forecast a one-step ahead VaR at α -level quantile, the predictive distribution $f(r_{n+1} | \mathbf{r}, \Xi)$ is first estimated using MCMC simulation. The MCMC generates samples denoted $\Xi^{(i)}$ for iterations $i = M+1, \dots, N$ where M is the burn-in period length and N is the length of the sampled chains. The resulting quantile VaR under a DMS-GARCH formulation is:

$$VaR_{n+1}^{(i)} = -\left(\omega_{n+1}^{(i)} + D_{\alpha}^{-1}(\Xi^{(i)})\sqrt{h_{n+1}^{(i)}}\right) \quad (2.26)$$

where $\omega_{n+1}^{(i)}$ and $h_{n+1}^{(i)}$ are conditional mean and variance of the prediction distribution $f(r_{n+1} | \mathbf{r}, \Xi^{(j)})$ determined using the DMS-GARCH model. D_{α}^{-1} is the inverse cumulative distribution (CDF) for SS- t distribution. The forecasted VaR at time $n+1$ is then calculated as a posterior mean quantity:

$$VaR_{n+1} = \frac{1}{N-M} \sum_{i=M+1}^N VaR_{n+1}^{(i)} \quad (2.27)$$

2.10 Assessing VaR models

In order to assess VaR model, a similar approach from Chen et al (2009) is used. A standard test for comparing the forecasting performance of VaR models is the use of

violation rates which can be denoted (VR). VR looks at the number of violations that have occurred in m forecasting days. VR is defined in Chen et al (2009) as:

$$VR = \frac{1}{N - M} \sum_{t=n+1}^{n+m} I(r_t < -\text{VaR}_t) \quad (2.28)$$

where n is the number of observation for the in-sample period. To obtain VaR for the complete m days, a moving window forecast approach is adopted as in Chen et al (2009). According to Chen et al (2009), the first n observations will forecast the $n + 1$ observation. The $n + 2$ is forecasted using the sample ranging from the second to the $n + 1$ observation. The moving window approach is repeated until the end of the sample. A ranking criteria is used to choose the best model that forecast VaR fairly well. The details of the ranking procedures are found in Wong and So (2003) and So and Yu (2006). The ranking methodology presented in So and Yu (2006) involves choosing the α values. The VR are computed for each data series. A good VaR method is expected to have a VR which is close to α . The smaller the absolute deviation between VR and α the better the VaR estimation method and this will entail a higher ranking.

3. Literature Review

Volatility forecasting plays a key role in the analysis of financial markets as alluded in section 1.5. For the purpose of this study, a description of volatility models that take into account the regime effects is explored. The use of Markov switching (MS) models have gained momentum in forecasting volatility data as they allow parameters to change depending on a state or regime variable. Hamilton (1988,1989) pioneered the use of MS models with the emphasis on modelling the evolution of interest rates and business cycles under Autoregressive (AR) and Autoregressive Integrated Moving Average (ARIMA) representations. For a discussion of AR and ARIMA models see Fan and Yao (2003). The parameters driving the processes were estimated using the maximum likelihood method. The above papers assumed that the residual processes could be modelled by a normal distribution. Hamilton (1988,1989) noted inconsistencies in the linear model, where parameters were not allowed to switch regimes. The results from these papers provided evidence that regime switching models offered a better approach in explaining the generating processes in economic variables. However, Hamilton's papers does not consider using GARCH models or different residual processes besides.

The question that has come under the spotlight in financial literature is that of forecasting the volatility of financial time series data. Nelson (1991) attempted to address the question by using a regime switching volatility model, based on the assumption that the regime is driven by an observable variable. Nelson models asset returns in relation to conditional variance. Nelson (1991) found that the new ARCH model introduced allowed simplicity and flexibility in representing conditional variances.

Another category of regime switching volatility models assumes that the state or regimes are influenced by an unobserved variable. Kim (1993) extended the MS model of Hamilton (1988,1989) by incorporating a ARCH component in the model. He modelled inflation and its uncertainty.

Cai (1994) analysed the persistence of volatility in financial time series data using a MS-ARCH model. Cai's paper highlighted that regime changes have significant impact on financial time series data. Hamilton and Sumel (1994) also explored persistence of volatility and concluded that economic recessions were associated with high volatility regimes

and that using a Student- t error distribution offered a better explanation of volatility changes as compared to the normal distribution assumption.

A generalised MS-GARCH model was developed by Gray (1996). Their results indicated that the MS-GARCH model outperformed single-regime models in the out-of sample forecasting. Klaassen (2002) expanded the model by providing a convenient way for multi-period volatility forecasting.

Dueker (1997) examined different specification for MS-GARCH models and found that the model with Student- t residuals predicted options implied volatilities better than other specifications. Li and Li (1996) extended the threshold models of Tong (1978), Tong and Lim (1980) and Tong (1983,1990). They called their model the double threshold ARCH (DTARCH) model since it handled both the conditional mean and conditional variance thresholds. Li and Li's results unveiled the possibility of asymmetry behaviour in volatility which has an influence when modelling the evolution of financial time series data. An innovation to the DTARCH model was developed by Brooks (2001) which allows the mean and variance to be drawn from two regimes under a GARCH representation. The threshold used in the DTARCH and Brooks (2001) models were not stochastic and there is no clear approach to determine the thresholds.

So et al (1998) proposed a Markov switching stochastic volatility (MSSV) model under an ARCH representation. Estimation of parameters were done using a Bayesian framework as discussed in section 2.6. Their findings were similar to Cai (1994) and Hamilton and Susmel (1994). A simulation study by Carvalho and Lopes (2006) was carried out to determine the performance of So et al (1998) MSSV model. Their result showed that the MSSV model was quite robust in forecasting as compared to a simple stochastic model.

Yoo (2004) dealt with the implementation problem encountered when an QMLE is used for estimation of model parameters. The QMLE method encounters problems when the conditional variance depends on history of states. Yoo demonstrated numerically that a Bayesian inference using a Markov Chain Monte Carlo (MCMC) method provided a simple way to estimate MS-GARCH models. Shibata and Watanabe (2005) applied a Bayesian MCMC in a MSSV model. In contrast, Marcucci (2005) and Piplack (2007)

implemented an MS-GARCH model but instead used the QMLE approach despite the difficulties highlighted by Yoo (2004). The advantages of the Bayesian MCMC over QMLE has been spelt out in section 2.7.

A key feature in the threshold heteroscedastic model of Chen and So (2006) is that the threshold variable is described as a weighted average of auxiliary variables. The model also incorporates important exogenous variables which affect for instance the dynamics of local market returns. Chen et al (2009) took a unique approach and introduced a double Markov switching GARCH (DMS-GARCH) model where the mean and volatility are modelled simultaneously. Another distinctive element is that the regimes are assumed to be unobserved and follow a first-order Markov process. Chen et al (2009) used a Bayesian formulation and the estimation of parameters were done using MCMC. In their DMS-GARCH model, they used a t error distribution. Haas (2010) introduced a skew-normal (SN) mixture density function to model a MS-GARCH model and the results indicated that a SN mixture outperformed a stand alone Gaussian model. Although the results are crucial, they do not demonstrate whether the SN mixture performs the same under a DMS-GARCH model of Chen et al (2009).

The following paper is different from the above cited papers in that it utilises a Skewed- t distribution under a DMS-GARCH model specification in order to forecast the volatility of JSE returns.

4. Results

4.1 Data

The empirical study carried out in this thesis utilises the daily closing prices of ANGLO American Plc (AGL), ANGLO American Platinum Ltd (AMS), ANGLO Ashanti Ltd (ANG), BHP Billiton Plc (BHP), Compagnie Fin Richemont (CFR), Firstrand (FSR), MTN Group Ltd (MTN), Sabmiller (SAB), Sasol (SOL), Standard Bank (SBK) and the JSE Top 40 index (TOP40), making a total of 11 data series. These shares were selected since they are the top 10 shares of the JSE Top 40 index and constitutes about 70% of the index. The data was obtained from the McGregor BFA database. The data was extracted for the period 14 January 2002 to 13 January 2012, generating 2501 observations. Table 4.1 provides the summary statistics for daily returns of the data series. The log returns were computed using the following formula:

$$r_t = (\log P_t - \log P_{t-1}) \times 100$$

where P_t and P_{t-1} are daily closing prices at times t and $t - 1$ respectively. All shares with the exception of AGL, AMS, ANG and CFR outperformed the TOP40 based on the average returns earned in the period investigated. The share returns of the data range from a minimum of about -47.6% to a maximum of close to 18% . Skewness measures the asymmetry of the distribution of returns and it is evident from the summary table that all the data series are positively skewed except for AGL, AMS, CFR, FSR and TOP40. The skewness coefficients are significant at 5% level of significance for most of the data. The standard errors for the skewness under the null hypothesis of normality is 0.0490. The kurtosis is greater than 3 implying that the share returns distribution are leptokurtic. The standard error when testing for kurtosis is $\sqrt{\frac{24}{2501}} = 0.0980$. At 5% significance level, the kurtosis coefficients are significant. The Jarque-Bera (JB) tests whether the series follow a normal distribution. The JB p -values for the data series are all less than 0.05 and therefore the null hypothesis of assuming a normal distribution is rejected for all series.

Table 4.1: Summary of statistics of daily returns

	Mean	Std	Skewness	Kurtosis	JB p -value	Min	Max
AGL	0.0173	2.5796	-0.0988	6.9107	0.0000	-17.2995	13.8586
AMS	0.0089	2.7361	-0.3180	5.5688	0.0000	-17.5891	11.9529
ANG	0.0167	2.5352	0.1752	5.8042	0.0000	-12.3268	17.5643
BHP	0.0578	2.4149	0.2254	6.3997	0.0000	-11.4209	17.9971
CFR	0.0242	2.2656	-3.7262	81.4242	0.0000	-47.5863	9.9621
FSR	0.0410	2.0756	-0.1288	4.9253	0.0000	-12.5769	10.7615
MTN	0.0852	2.3913	0.2689	5.7098	0.0000	-12.3690	16.2519
SAB	0.0514	1.6348	0.0959	5.1304	0.0000	-7.7560	9.2363
SOL	0.0543	2.2593	0.0784	5.2376	0.0000	-10.6444	11.4342
SBK	0.0460	1.9808	0.1099	4.9015	0.0000	-10.4722	10.4167
TOP40	0.0408	1.4559	-0.0725	5.8669	0.0000	-7.9594	7.7069

4.2 Model description

According to Bollerslev et al (1992), the GARCH(1,1) model is adequate to explain the conditional variance of share returns. As such, a comparison of the two models for this study is carried out, namely GARCH(1,1) and the DMS-GARCH(1,1) with SS- t error innovations. The choice of the order (1,1) for the GARCH and DMS-GARCH models follow the same pattern from the paper presented by Chen et al (2009). From Table 4.1 it is clear that all of the series exhibits large kurtosis and that a heavy tailed distribution is required to model the return series. The GARCH(1,1) model is specified as follows:

$$r_t = u_t \quad (4.1)$$

$$u_t = \sqrt{h_t} \epsilon_t \quad \text{and} \quad \epsilon_t \sim \text{SS-}t(\nu, \eta) \quad (4.2)$$

$$h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} \quad (4.3)$$

The TOP40 return series will be used as the exogenous variable in the formulation of the DMS-GARCH(1,1) model. The TOP40 assists in explaining the dynamics of individual share returns on the JSE given the prevailing market conditions. The DMS-GARCH(1,1)

model is defined as:

$$r_t = \begin{cases} \phi_0^{(1)} + \phi_1^{(1)}r_{t-1} + \varphi_1^{(1)}x_{t-1} + u_t & \text{if } s_t = 1, \\ \phi_0^{(2)} + \phi_1^{(2)}r_{t-1} + \varphi_1^{(2)}x_{t-1} + u_t & \text{if } s_t = 2, \end{cases} \quad (4.4)$$

$$u_t = \sqrt{h_t}\epsilon_t \quad \text{and} \quad \epsilon_t \sim \text{SS-}t(\nu, \eta) \quad (4.5)$$

$$h_t = \begin{cases} \alpha_0^{(1)} + \alpha_1^{(1)}u_{t-1}^2 + \beta_1^{(1)}h_{t-1} & \text{if } s_t = 1, \\ \alpha_0^{(2)} + \alpha_1^{(2)}u_{t-1}^2 + \beta_1^{(2)}h_{t-1} & \text{if } s_t = 2, \end{cases} \quad (4.6)$$

In the above formulation we assume that $s_t = 1$ relates to a low volatility period and $s_t = 2$ is a high volatility period.

4.3 Model estimation and convergence diagnostics

All computations were done using MATLAB 2010 version. Posterior samples of unknown parameters were obtained using a combination of Metropolis Hastings and Gibbs sampling techniques. In order to initialise the MCMC process, it was necessary to propose a density function that explore the unknown parameters' range. The proposal density for ϕ_j and α_j was a normal and truncated normal distribution respectively whereas for ν and η , uniform distribution were used. We ran 30 000 simulations and the last 15 000 were considered for posterior inference. Before proceeding with the posterior analysis, we computed the convergence statistic given in section 2.7.2.1. The statistics ranged between 0.01 and 0.07 suggesting convergence of the MCMC chains (Yu and Mykland (1994)). An examination of the trace plots (of the parameters) also provide further evidence that the MCMC chains has converged. The trace plots appears to wander about a constant mean. An examination of the autocorrelation functions indicates that the samples are weakly correlated (not displayed here). Fig 4.1 provides an illustration of the simulation of ν using the last 15 000 runs. The posterior mean and standard deviation (shown in brackets) of the unknown parameters are displayed in Tables 4.2-4.4 (after thinning the simulations using an appropriate thinning parameter).

Table 4.2 provides a summary of the Bayesian estimates of the GARCH(1,1) model defined in equation 4.3. The volatility persistence measure ($\alpha_1 + \beta_1$) for all shares lies in

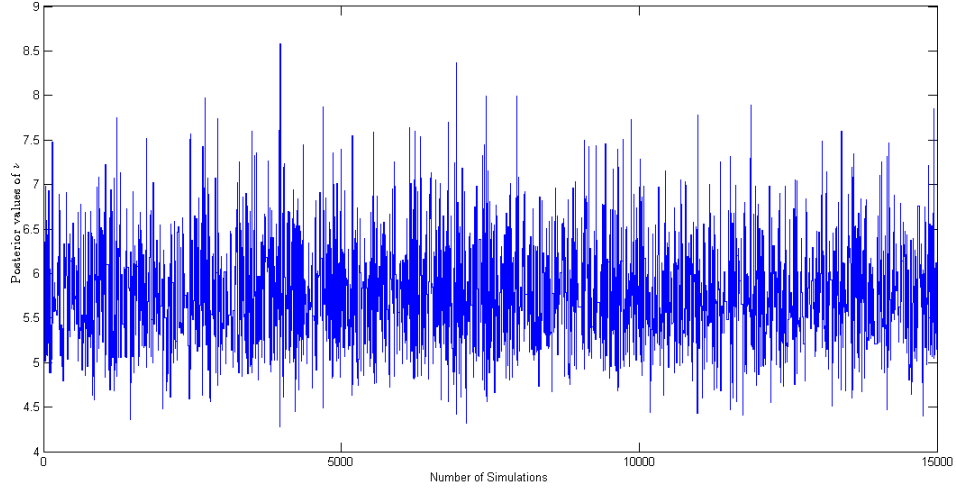


Figure 4.1: Simulations from posterior distribution of ν (last 15 000 simulations) for AMS

the interval 0.91 to 0.96. The number of degrees of freedom parameter ν is greater than 4 implying the need for a heavy tailed distribution. However, the value of the asymmetry parameter η is close to zero which suggests that the distribution of the share returns might well be approximated by a Student- t distribution.

In comparison, Table 4.3-4.4 summarises the estimates of DMS-GARCH(1,1) model. $\phi_1^{(1)}$ is close to zero with BHP, CFR and SAB having positive values. The values of $\phi_1^{(2)}$ are generally positive but insignificantly different from zero. The results suggests that in general $\phi_0^{(1)} = \phi_0^{(2)} = 0$, $\phi_1^{(1)} = \phi_1^{(2)} = 0$ and $\varphi_1^{(1)} = \varphi_1^{(2)} = 0$ which indicates that the conditional mean of the return series does not depend on the state variable s_t . The lagged and the exogenous variables are not significant in explaining the dynamics of the return series data. The values of $\alpha_0^{(1)}$ and $\alpha_0^{(2)}$ lie between 0.4 and 0.63. The volatility persistence measure values reveals that $\alpha_1^{(2)} + \beta_1^{(2)}$ is greater than $\alpha_1^{(1)} + \beta_1^{(1)}$. The values of $\alpha_1^{(i)} + \beta_1^{(i)}$ for $i = 1, 2$ lie between 0.71 and 0.96 for all of the shares. To characterise volatility, a computation of unconditional variance is done as in Chen et al (2009). The unconditional variance in state i , denoted σ_i^2 is given by:

$$\sigma_i^2 = \frac{\alpha_0^{(i)}}{1 - \alpha_1^{(i)} - \beta_1^{(i)}}$$

The values of σ_2^2 is greater than σ_1^2 and a distinction can therefore be made which labels

regime 1 and 2 as low and high volatility states respectively.

Table 4.2: Posterior mean of the unknown parameters for the GARCH SS- t model

	α_0	α_1	β_1	ν	η
AGL	0.6064 (0.2665)	0.1062 (0.1027)	0.8502 (0.1121)	5.2091 (0.4041)	0.0067 (0.0269)
AMS	0.6113 (0.2659)	0.1155 (0.1088)	0.8435 (0.1174)	5.7761 (0.5446)	-0.0387 (0.0252)
ANG	0.6209 (0.2620)	0.1060 (0.0991)	0.8535 (0.1066)	6.1935 (0.6061)	0.0082 (0.0287)
BHP	0.6160 (0.2585)	0.0971 (0.0970)	0.8583 (0.1062)	5.9650 (0.5321)	-0.0248 (0.0258)
CFR	0.6372 (0.2510)	0.0946 (0.0954)	0.8602 (0.1048)	4.4470 (0.2837)	-0.0101 (0.0275)
FSR	0.6146 (0.2603)	0.1420 (0.1376)	0.7946 (0.1509)	7.8212 (1.0061)	0.0066 (0.0276)
MTN	0.6094 (0.2647)	0.1073 (0.1059)	0.8467 (0.1155)	6.3399 (0.6413)	0.0225 (0.0256)
SAB	0.6331 (0.2502)	0.1601 (0.1581)	0.7515 (0.1817)	7.2759 (0.8634)	-0.0147 (0.0263)
SOL	0.6200 (0.2627)	0.1194 (0.1166)	0.8293 (0.1281)	6.6053 (0.6869)	-0.0181 (0.0261)
SBK	0.6258 (0.2572)	0.1324 (0.1287)	0.8021 (0.1432)	6.6499 (0.7476)	0.0212 (0.0275)

4.4 Forecasting performance

We used the MCMC approach with 30 000 simulations of which the last half were used to estimate the unknown parameters of the GARCH(1,1) and DMS-GARCH(1,1) models. The last 500 data observations were used to assess the forecasting performance of the GARCH(1,1) and DMS-GARCH(1,1) models. A rolling window approach was used to produce 500 volatility forecasts points. We used the first 2001 observations to predict the volatility value at time 2002 (the first forecasted point). The process is continued until

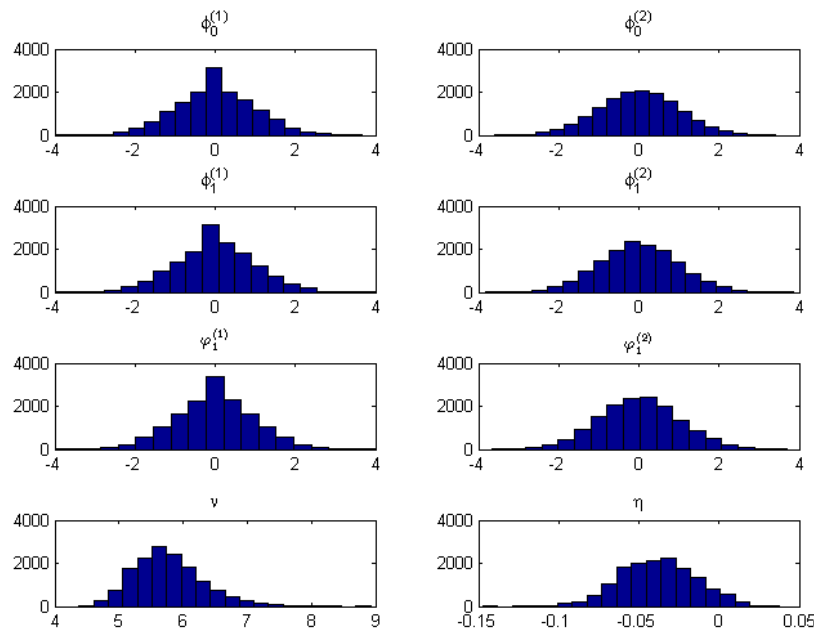


Figure 4.2: Example illustrating posterior draws of some of the DMS-GARCH(1,1) parameters for AMS

500 forecasts are produced. Fig 4.3 and 4.4 displays the one day ahead volatility forecast for AMS using the GARCH(1,1) and DMS-GARCH(1,1) model with SS- t innovations. The illustration clearly shows that there is more ‘jumpiness’ in Fig 4.4 as compared to Fig 4.3, this is demonstrated by the many spikes in Fig 4.4 throughout the forecasting period which could be attributed to volatility shifts in the DMS-GARCH(1,1) model.

Model performance is undertaken by the computation of violation ratios ($\frac{VR}{\alpha}$) displayed in Table 4.5 and the application of the ranking method by So and Yu (2006). So and Yu’s method uses the absolute value between the violation ratio and 1 in order to rank model performance. Models with violation ratio closer to 1 are ranked higher than ones with violation ratio further away from 1. We used the two α values namely 1% and 5%. These should be interpreted as the 1% and 5% cut off values. In Table 4.6, the model with a violation ratio close to 1 is ranked 1 and the next closest to 1 is ranked 2. It is observed that DMS-GARCH model has better rankings at 1% level while both models perform equally well at the 5% level. This suggests that there is no clear distinction in terms of performance for the GARCH and DMS-GARCH models. The GARCH model might well

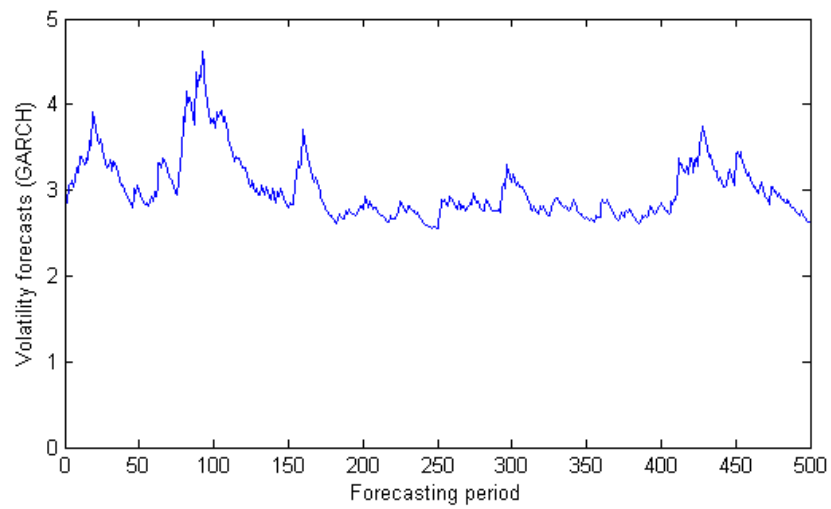


Figure 4.3: One day ahead volatility forecasts for AMS using a GARCH(1,1) model

be chosen since it is more parsimonious as compared to the DMS-GARCH model.

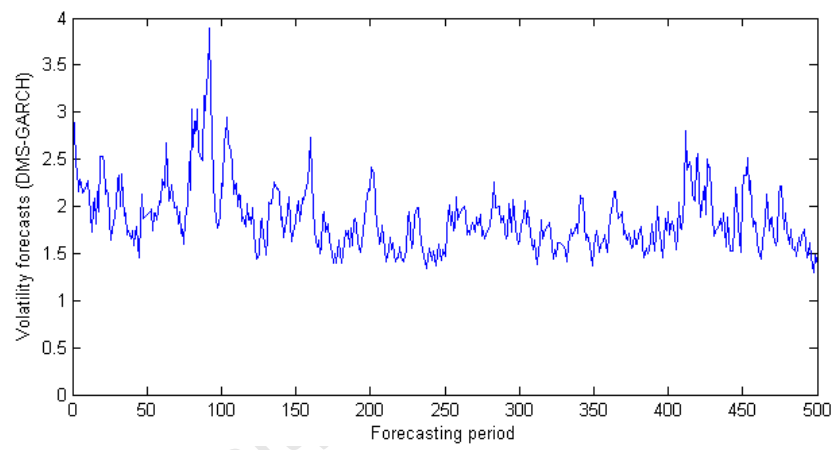


Figure 4.4: One day ahead volatility forecasts for AMS using a DMS-GARCH(1,1) model

Table 4.3: Posterior mean of the unknown parameters for the DMS-GARCH model

	AGL	AMS	ANG	BHP	CFR
$\phi_0^{(1)}$	0.0065 (0.9590)	0.0030 (0.9809)	-0.0223 (0.9641)	0.0126 (0.9729)	-0.0041 (0.9671)
$\phi_1^{(1)}$	-0.0004 (0.9592)	-0.0022 (0.9763)	-0.0003 (0.9756)	0.0042 (0.9734)	0.0044 (0.9630)
$\varphi_1^{(1)}$	0.0051 (0.9718)	0.0048 (0.9799)	0.0086 (0.9826)	-0.0083 (0.9793)	0.0115 (0.9684)
$\phi_0^{(2)}$	0.0026 (0.9997)	-0.0010 (1.0082)	0.0008 (0.9947)	0.0067 (1.0021)	-0.0075 (1.0006)
$\phi_1^{(2)}$	-0.0043 (1.0149)	0.0006 (0.9884)	0.0137 (0.9910)	0.0094 (0.9967)	-0.0046 (0.9968)
$\varphi_1^{(2)}$	-0.0059 (1.0073)	-0.0021 (0.9872)	-0.0029 (1.0053)	0.0257 (1.0008)	-0.0042 (0.9920)
$\alpha_0^{(1)}$	0.4659 (0.3448)	0.4623 (0.3371)	0.4719 (0.3475)	0.4896 (0.3408)	0.4788 (0.3497)
$\alpha_1^{(1)}$	0.0807 (0.0997)	0.0865 (0.1034)	0.0797 (0.0976)	0.0772 (0.0943)	0.0675 (0.0876)
$\beta_1^{(1)}$	0.6569 (0.3732)	0.6644 (0.3650)	0.6562 (0.3735)	0.6827 (0.3611)	0.6490 (0.3840)
$\alpha_0^{(2)}$	0.6092 (0.2645)	0.5988 (0.2653)	0.6173 (0.2609)	0.6111 (0.2598)	0.6251 (0.2484)
$\alpha_1^{(2)}$	0.1055 (0.1032)	0.1182 (0.1074)	0.1013 (0.0967)	0.1017 (0.1006)	0.0895 (0.0893)
$\beta_1^{(2)}$	0.8527 (0.1135)	0.8404 (0.1172)	0.8550 (0.1113)	0.8539 (0.1165)	0.8649 (0.1067)
p_{11}	0.4080 (0.2925)	0.4123 (0.2908)	0.4106 (0.2921)	0.4164 (0.2929)	0.4066 (0.2915)
p_{22}	0.5908 (0.2926)	0.5871 (0.2906)	0.5887 (0.2919)	0.5827 (0.2927)	0.5925 (0.2917)
ν	5.2003 (0.4050)	5.7885 (0.5328)	6.1794 (0.5999)	5.9516 (0.5350)	4.4380 (0.2756)
η	0.0101 (0.0299)	-0.0382 (0.0256)	0.0115 (0.0290)	-0.0147 (0.0275)	-0.0040 (0.0266)
σ_1^2	1.7761	1.8566	1.7868	2.0389	1.6885
σ_2^2	14.5425	14.4422	14.1176	13.7435	13.7046

Table 4.4: Posterior mean of the unknown parameters for the DMS-GARCH model

	FSR	MTN	SAB	SOL	SBK
$\phi_0^{(1)}$	0.0021 (0.9656)	0.0063 (0.9851)	0.0030 (0.9775)	0.0081 (0.9552)	-0.0166 (0.9640)
$\phi_1^{(1)}$	-0.0195 (0.9672)	-0.0021 (0.9763)	0.0071 (0.9712)	-0.0073 (0.9693)	-0.0030 (0.9710)
$\varphi_1^{(1)}$	-0.0013 (0.9691)	-0.0123 (0.9663)	-0.0094 (0.9670)	-0.0039 (0.9680)	0.0154 (0.9743)
$\phi_0^{(2)}$	0.0167 (0.9930)	0.0013 (1.0072)	-0.0071 (0.9959)	-0.0028 (1.0149)	0.0212 (0.9948)
$\phi_1^{(2)}$	-0.0022 (1.0061)	0.0116 (0.9975)	-0.0010 (0.9994)	0.0002 (0.9971)	0.0128 (1.0063)
$\varphi_1^{(2)}$	0.0012 (1.0048)	0.0068 (0.9863)	-0.0140 (1.0008)	0.0099 (1.0005)	-0.0052 (0.9983)
$\alpha_0^{(1)}$	0.5146 (0.3351)	0.4818 (0.3416)	0.5451 (0.3239)	0.4856 (0.3419)	0.5186 (0.3315)
$\alpha_1^{(1)}$	0.1168 (0.1356)	0.0824 (0.1014)	0.1422 (0.1622)	0.0890 (0.1063)	0.1073 (0.1262)
$\beta_1^{(1)}$	0.6574 (0.3310)	0.6619 (0.3655)	0.6350 (0.3153)	0.6550 (0.3602)	0.6642 (0.3318)
$\alpha_0^{(2)}$	0.6093 (0.2657)	0.6199 (0.2619)	0.6287 (0.2549)	0.6220 (0.2622)	0.6154 (0.2560)
$\alpha_1^{(2)}$	0.1390 (0.1338)	0.1103 (0.1031)	0.1701 (0.1636)	0.1199 (0.1177)	0.1271 (0.1232)
$\beta_1^{(2)}$	0.7941 (0.1551)	0.8446 (0.1136)	0.7394 (0.1874)	0.8295 (0.1292)	0.8087 (0.1410)
p_{11}	0.4118 (0.2905)	0.4152 (0.2909)	0.4128 (0.2909)	0.4105 (0.2922)	0.4143 (0.2938)
p_{22}	0.5870 (0.2903)	0.5836 (0.2911)	0.5862 (0.2908)	0.5885 (0.2924)	0.5848 (0.2937)
ν	7.8304 (1.0295)	6.3758 (0.6687)	7.2810 (0.8628)	6.6100 (0.7035)	6.6626 (0.7729)
η	0.0137 (0.0254)	0.0347 (0.0281)	-0.0052 (0.0290)	-0.0128 (0.0277)	0.0300 (0.0275)
σ_1^2	2.2794	1.8845	2.4465	1.8966	2.2692
σ_2^2	9.1095	13.7468	6.9418	12.3052	9.5867

Table 4.5: Ratio ($\frac{VaR}{\alpha}$)

VaR	Share	GARCH	DMS-GARCH
$\alpha = 1\%$	AGL	1.34	1.19
	AMS	0.64	1.07
	ANG	0.85	1.01
	BHP	1.24	1.06
	CFR	1.31	0.68
	FSR	1.12	0.97
	MTN	1.22	1.14
	SAB	0.83	0.78
	SOL	0.61	0.84
	SBK	1.18	1.28
$\alpha = 5\%$	AGL	0.91	0.77
	AMS	0.73	1.09
	ANG	1.38	0.67
	BHP	0.80	1.39
	CFR	1.32	0.66
	FSR	1.29	0.61
	MTN	0.73	0.96
	SAB	1.12	1.04
	SOL	0.83	1.13
	SBK	0.92	1.39

Table 4.6: Ranking models

α	Share	GARCH	DMS-GARCH
$\alpha = 1\%$	AGL	2	1
	AMS	2	1
	ANG	2	1
	BHP	2	1
	CFR	1	2
	FSR	2	1
	MTN	2	1
	SAB	1	2
	SOL	2	1
	SBK	1	2
$\alpha = 5\%$	AGL	1	2
	AMS	2	1
	ANG	2	1
	BHP	1	2
	CFR	1	2
	FSR	1	2
	MTN	2	1
	SAB	2	1
	SOL	2	1
	SBK	1	2

5. Conclusions

The DMS-GARCH model with $SS-t$ distribution error innovation was motivated by the Chen et al (2009) paper. The distinct features of this thesis are; a different error distribution was used and the DMS-GARCH model was applied to the JSE individual share data in order to forecast volatility. The $SS-t$ distribution is favoured since it captures the heavy-tails of financial data. This is motivated by the significance of the skewness and kurtosis coefficients. The results showed that the lagged and the exogenous variables are insignificant in explaining the dynamics of individual share returns series on the JSE. The regimes do not play a significance role in forecasting volatility of JSE. The paper also selected a model using violation ratios and a ranking technique. The results showed that DMS-GARCH model with $SS-t$ distribution performs similarly to the GARCH $SS-t$ model. Future research in this area could entail using the DMS-GARCH model with $SS-t$ innovations to forecast volatility in order to calculate Conditional Value at Risk (CVaR) levels for individual shares and portfolio. This paper can also be extended by incorporating other performance evaluation methods and using data for other stock markets to forecast volatility.

A. Appendix

A.1 Probability Densities

Peters (2001) argued that in financial time series data, the error or residual terms in GARCH models exhibit excess kurtosis and skewness. This section provides commonly used distributions.

A.1.1 Normal

Normal distribution also known as the Gaussian distribution, is a symmetric distribution and has quite tractable features. However, it is not a suitable distribution when modelling volatility as it does not capture the observed asymmetry in security returns. The standard normal has a probability density function (pdf) given by:

$$f(\epsilon_t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{s_t^2}{2}} \quad (\text{A.1})$$

and its corresponding log-likelihood is:

$$l_N = -\frac{1}{2} \sum_{t=1}^N (\ln 2\pi + s_t^2) \quad (\text{A.2})$$

where $s_t = \frac{\epsilon_t - \mu_t}{\sqrt{h_t}}$ is the standardising parameter that ensures a mean of 0 and variance of 1. N is the number of observations of the series data under investigation.

A.1.2 Student- t

The standardised Student- t has a pdf given by

$$f(\epsilon_t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right) \left(1 + \frac{s_t^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}}{\Gamma\left(\frac{\nu}{2}\right) \sqrt{\pi(\nu-2)}} \quad (\text{A.3})$$

the associated likelihood is

$$l_N = \ln \left(\Gamma \left(\frac{\nu + 1}{2} \right) \right) - \ln \left(\Gamma \left(\frac{\nu}{2} \right) \right) - \frac{1}{2} \ln (\pi(\nu - 2)) - \frac{1}{2} \sum_{t=1}^N \left((\nu + 1) \ln \left(1 + \frac{s_t^2}{\nu - 2} \right) \right) \quad (\text{A.4})$$

where $\nu > 2$ and $\Gamma(\cdot)$ is a gamma function.

The parameter ν is the degrees of freedom parameter that measures the tail thickness. Note that the Student- t distribution approximates to a normal distribution as $\nu \rightarrow \infty$.

A.1.3 Skewed Student- t

Peters (2001) considers skewness to be of paramount importance in explaining the “jumpiness” of financial asset returns, particularly in asset and derivative pricing. Hence this thesis incorporated the Skewed Student- t distribution because it has two important features the tail and the asymmetry parameters. The probability density of a Skewed Student- t as defined in Hansen (1994) is:

$$f_\epsilon(\epsilon_t | \nu, \eta) = \begin{cases} bc \left[1 + \frac{1}{\nu-2} \left(\frac{b\epsilon_t + a}{1-\eta} \right)^2 \right]^{-\frac{(\nu+1)}{2}} & \text{if } \epsilon_t < -\frac{a}{b}, \\ bc \left[1 + \frac{1}{\nu-2} \left(\frac{b\epsilon_t + a}{1+\eta} \right)^2 \right]^{-\frac{(\nu+1)}{2}} & \text{if } \epsilon_t \geq -\frac{a}{b}, \end{cases}$$

where ν and η are degrees of freedom and asymmetry parameters respectively. We require $2 < \nu < \infty$ and $-1 < \eta < 1$. The constants a , b and c are defined as

$$\begin{aligned} a &= 4\eta c \left(\frac{\nu - 2}{\nu - 1} \right) \\ b^2 &= 1 + 3\eta^2 - a^2 \\ c &= \frac{\Gamma \left(\frac{\nu+1}{2} \right)}{\sqrt{\pi(\nu - 2)\Gamma \left(\frac{\nu}{2} \right)}} \end{aligned}$$

It is important to note that when $\eta = 0$, the SS- t distribution reduces to a Student- t distribution and as ν becomes large the SS- t gravitates towards a Gaussian distribution.

The likelihood of Skewed Student- t is:

$$l_N = \begin{cases} \sum_{j=t}^n \left(\ln(bc) - \frac{\nu+1}{2} \ln \left(1 + \frac{1}{\nu-2} \left(\frac{b\epsilon_j+a}{1-\eta} \right)^2 \right) \right) & \text{if } \epsilon_t < -\frac{a}{b}, \\ \sum_{j=t}^n \left(\ln(bc) - \frac{\nu+1}{2} \ln \left(1 + \frac{1}{\nu-2} \left(\frac{b\epsilon_j+a}{1+\eta} \right)^2 \right) \right) & \text{if } \epsilon_t \geq -\frac{a}{b}. \end{cases}$$

University of Cape Town

Bibliography

Ardia D. (2008) *Financial Risk Management with Bayesian Estimation of GARCH Models: Theory and Applications*. Springer: Berlin.

Black F. (1976) Studies of Stock Market Volatility Changes. *1976 Proceedings of the American Statistical Association, Business and Economic Statistics Section* 177-181.

Bollerslev T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31(3): 307-327.

Bollerslev T, Chou RY, Kroner KF. (1992) ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics* 52(1): 5-59.

Brooks C. (2001) A double-threshold GARCH model for the French Franc/Deutschmark exchange rate. *Journal of Forecasting* 20(2): 135-143.

Cai J. (1994) A Markov model of switching-regime ARCH. *Journal of Business and Economic Statistics* 12(3): 309-316.

Carvalho CM, Lopes HF. (2006) Simulation-based sequential analysis of Markov switching stochastic volatility models. *Computational Statistics and Data Analysis* 51(9): 4526-4542.

Chen CWS, So MKP. (2006) On a threshold heteroscedastic model. *International Journal of Forecasting* 22(1): 73-89.

- Chen CWS, So MKP, Lin EMH. (2009) Volatility forecasting with double Markov switching GARCH models. *Journal of Forecasting* 28(8): 681-697.
- Ding Z, Granger C, Engle RF. (1993) A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1(1): 83-106.
- Drost FC, Nijman TE. (1993) Temporal aggregation of GARCH processes. *Econometrica* 61(4): 909-927.
- Dueker MJ. (1997) Markov switching in GARCH processes and mean-reverting stock market volatility. *Journal of Business and Economic Statistics* 15(1): 26-34.
- Engle RF. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50(4): 987-1007.
- Engle RF, Bollerslev T. (1986) Modelling the persistence of conditional variances. *Econometric Reviews* 5(1): 1-50.
- Fan J, Yao Q. (2003) *Nonlinear Time Series Modelling and Prediction*. Springer: New York.
- Francq C, Zakoian JM. (2010) *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley: West Sussex.
- Geman S, Geman D (1984). Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6(6): 721-741.
- Glosten LR, Jagannathan R, Runkle DE. (1993) On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48(5): 1779-1801.

- Gray SF. (1996) Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* 42(1): 27-62.
- Haas M. (2010) Skew-normal mixture and Markov switching GARCH processes. *Studies in Nonlinear Dynamics and Econometrics*, The Berkeley Electronic Press, 14(4): Article 1.
- Hamilton JD. (1988) Rational expectations econometric analysis of change in regime: an investigation of the term structure of interest rates. *Journal of Economic Dynamics and Control* 12(2-3): 385-423.
- Hamilton JD. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57(2): 357-384.
- Hamilton JD, Susmel R. (1994) Autoregressive conditional heteroscedasticity and changes in regime. *Journal of Econometrics* 64(1-2): 307-333.
- Hansen B. (1994) Autoregressive conditional density estimation. *International Economic Review* 35(3): 705-730.
- Hastings WK (1970). Monte Carlo Sampling Methods Using Markov Chains and their Applications. *Biometrika* 57(1): 97-109.
- Kim CJ. (1993) Unobserved-components time series models with Markov switching heteroscedasticity: changes in regime and the link between inflation rates and inflation uncertainty. *Journal of Business and Economic Statistics* 11(3): 341-349.
- Klaassen F. (2002) Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics* 27(2): 363-394.
- Li CW, Li WK. (1996) On a double-threshold autoregressive heteroscedastic time series model. *Journal of Applied Econometrics* 11(3): 253-274.

Mandelbrot B. (1963) The variation of certain speculative prices. *Journal of Business* 36(4): 394-419.

Marcucci J.(2005) Forecasting stock market volatility with regime-switching GARCH Models, *Studies in Nonlinear Dynamics and Econometrics*, The Berkeley Electronic Press, 9(4): Article 6.

Markowitz H. (1952) Portfolio selection. *Journal of Finance* 7(1): 77-91.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953). Equations of State Calculations by Fast Computing Machines. *Journal of Chemical Physics* 21(6): 1087-1092.

Nelson DB (1991) Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59(2): 347-370.

Peters JP. (2001) Estimating and forecasting volatility of stock indices using asymmetric GARCH models and (Skewed) Student-t densities. *Ecole d'Administration des Affaires, University of Li'ege, Belgium*.

Piplack J. (2007) Estimating and forecasting asset volatility and its volatility: a Markov-switching range model. Working paper.

Rachev ST, HSU JSJ, Bagashev BS, Fabozzi FJ. (2008) *Bayesian methods in finance*. Wiley: New Jersey.

Shibata M, Watanake T. (2005) Bayesian analysis of Markov switching stochastic volatility model. *J. Japan Statist. Soc* 35(2): 205-219,

So MKP, Lam K, Li WK. (1998) A stochastic volatility model with Markov switching. *Journal of Business and Economic Statistics* 16(2): 244-253.

So MKP, Yu PLH. (2006) Empirical analysis of GARCH models in value at risk estima-

- tion. *Journal of International Financial Markets, Institutions and Money* 16(2): 180-197.
- Straumann D. (2005) *Estimation in Conditionally Heteroscedastic Time Series Models*. Springer: Berlin.
- Tong H. (1978) On a threshold model. *In Pattern Recognition and Signal Processing*, Chen CH (ed). Sijhoff and Noordhoff: Amsterdam; 101-141.
- Tong H, Lim KS. (1980) Threshold autoregression, limit cycles and cyclical data (with discussion). *Journal of the Royal Statistical Society, Series B* 42(3): 245-292.
- Tong H. (1983) *Threshold Models in Non-linear Time Series Analysis*, Vol. 21 of *Lecture Notes in Statistics*, Krickeberg K (ed.). Springer: New York.
- Tong H. (1990) *Nonlinear Time Series: A Dynamical System Approach*. Oxford University Press: Oxford.
- Watson MW, Engle RF. (1983) Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics* 23(3): 385-400.
- Wong CM, So MKP. (2003) On conditional moments of GARCH models, with applications to multiple period value at risk estimation. *Statistica Sinica* 13(4): 1015-1044.
- Yoo BH. (2004) A Bayesian analysis of Markov switching models with ARMA-GARCH error. Working paper, Rutgers University.
- Yu B, Mykland P.(1994) Looking at Markov samplers through CUMSUM path plots: a simple diagnostic idea. *Technical Report 413*. Department of Statistics, University of California at Berkeley.
- Zakoian JM. (1994) Threshold heteroscedastic models. *Journal of Economic Dynamics and Control* 18(5): 931-955.

