Bayesian Estimation of Stochastic Volatility Models with Fat Tails and Correlated Errors Applied to the South African Financial Market

A Mini-Thesis submitted to the University of Cape Town in partial fulfillment of the requirements for an M.Phil degree in Mathematics of Finance in the faculty of Commerce

by

Richard Savanhu

Supervised by

Professor. Ronald Becker

May 23, 2011

PLAGIARISM DECLARATION

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is my own.
2. I have used the vancouver referencing style as the convention for citation and referencing. Each contribution to, and quotation in this mini thesis from the work(s) of other people has been contributed, and has been cited and referenced.
3. This mini thesis is my own work.
4. I have not allowed, and will not allow, anyone to copy my work.

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

Abstract

In this study we apply Markov Chain Monte Carlo methods in the Bayesian framework to estimate Stochastic Volatility models using South African financial market data. A single move Gibbs sampler is used to sample parameters from the posterior distribution. Volatility is used as measure of an asset's risk. It is particularly important in risk management, derivatives pricing, and portfolio selection. When pricing derivatives it is important to quote the correct volatility trading in the market, hence there is need for good estimates of volatility. To capture the stylised facts about asset returns we used the model extended for fat tails and correlated errors. To support this model against the basic model of Taylor (1986), we computed Bayes Factors of Jacquier, Polson and Ross (2004). The extended model was found to be far superior to the basic model.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Definitions

Volatility is the variability of a financial time series. It is a measure of magnitude and speed of asset returns variance of yet unrealised asset return [20]. Volatility is influenced by factors such as trading volume, information arrival and transaction costs among other factors (Andersen [1], Clark [4]). These factors are not scheduled, as a result volatility is not predictable and not constant. Volatility varies with time and in such a case, returns are said to be *hetroskedastic*. If σ_t^2 is the volatility and r_t is the return at time t, then

$$\sigma_t^2 = \mathrm{Var}(r_t|\mathcal{I}_{t-1}),$$

where $\mathcal{I}_{t-1}$ is the set of information available up to and including time $t-1$.

A *latent* variable is defined to be a variable that is not directly observed but is rather inferred from other variables that are observed (Wikipedia). *Stochastic volatility* is an unobservable (latent) volatility which is driven by a stochastic process. Given all the information, $\mathcal{I}_{t-1}$, volatility can not be completely determined at time t (Clark [4], Taylor 1986).

In finance, *leverage* refers to a technique used by individuals or firms to increase (or decrease) expected returns. To achieve leverage they can borrow money, use derivatives and so on. In the context of volatility modelling, *leverage effects* refer to the riskiness of a firm in relation to stock price movements. When stock prices are falling the value of equity decreases and the debt to equity ratio increases. This will cause the firm to be more risky and induces a high future volatility. Thus leverage effects are measured by a negative correlation between returns and volatility.

1.2 Literature review

Andersen [1] argued that volatility of asset price changes is directly related to the rate of flow of information to the market. This was first observed by Taylor (1986), he then argued that there must be an unpredictable component in volatility as some news is not scheduled and this led to the popular basic model of stochastic volatility of Taylor (1986). This concept and all other concepts such as time deformation (Clark [4]), support the view that volatility must be viewed as a latent process as opposed to the GARCH[1] volatility.

Much literature has concerned stochastic volatility (SVOL) models, but until recently the GARH family volatility has been extensively applied in practice due to its tractability. As we are going to see, the likelihood function of parameters of SVOL model is not only nonlinear, it contains these stochastic volatilities which need to be integrated out. There are several methods which have been proposed in trying to estimate the SVOL model, these include the efficient methods of moments (EMM) and quasi-maximum likelihood (QML). These methods are generally known to be inefficient [20]. However, the invention of the Markov Chain Monte Carlo (MCMC) methods has proved to be efficient in estimating the SVOL models. Jacquier, Polson and Ross (JPR) proposed a MCMC simulation algorithm to conduct Bayesian estimates of SVOL models, which has an advantage of yielding exact filtering and smoothing solutions to the problem of making inferences about unobservable volatility [9].

Empirical evidence has revealed that returns are not normally distributed, in fact they have a leptokurtic distribution. Figure 1.1 shows returns distribution of a JSE stock, Naspers. It can be seen that distibution of returns has fat tails and is more highly peaked than the normal distribution. Ghysels and Jasiak [9] commenting on the work by JPR [11] on estimating SVOL models, said the emphasis should be more on changing the model to improve the characterisation of volatility. There is need for a model that takes account of these stylised facts. Such a model can be derived via the distribution of the innovations in the mean equation and (or) innovations in the volatility process of the basic volatility model when modelling fat tailed (leptokurtic) distribution, this was noted by Geweke [8] and others. The main assumption in the model by Taylor is that innovation in the mean equation and innovation in the log volatility equation are independent. The independence between innovations has motivated the extension of the basic model of Taylor (1986) to include the leverage effect, by introducing a correlation parameter between the two innovations in the model. JPR [12] extended the basic volatility model in the Bayesian framework to cater for these stylised facts of returns, and they also provided the MCMC algorithm for such a model.

Faced with two or more models, Bayes Factors provide a way of selecting a model that is best explained by the data. Carlin and Chib [2] provided a framework for Bayesian model selection and a MCMC algorithm for computing Bayes Factors. JPR [12] noted that direct evaluation of Bayes Factors can be numerically unstable for latent variables. They then provided simple functions for computing Bayes Factors using MCMC output.

[1] $\sigma^2_{t|t-1} = \alpha_0 + \alpha_1 \sigma^2_{t-1|t-2} + \beta \varepsilon^2_{t-1}$, where α_0, α_1, β are model parameters, ε_{t-1} is a white noise that carries all the new information that is available at time $t-1$. GARCH(1,1) of Bollerslev(1986).

1.3 Aims and thesis outline

In this mini-thesis we investigate if it is necessary to extend the basic SVOL model of Taylor (1986) for fat tails and correlated errors in the South African financial market.

Chapter 2 gives the basic concepts. It starts with looking at Bayesian inference in general where Bayes' theorem and Bayes factors are introduced. Monte Carlo methods and algorithms are also discussed in this chapter. The chapter closes by looking at the bivariate normal distribution which plays a pivotal role when implementing the SVOL models in WinBUGS, a software for performing MCMC simulations.

Chapter 3 gives a Bayesian analysis of SVOL models. It starts with the basic SVOL model of Taylor (1986) and discusses the ability of the model to capture the features of empirical returns and the model shortcomings. The discussion of parameter estimation of the basic model is presented. Model extensions are then discussed and the section ends by deriving a form of the model extended for fat-tails and correlated errors that enables it to be implemented in WinBUGS.

Chapter 4 presents empirical results. It starts with details of the data used in this study. Findings and analysis of results are also presented, and Chapter 5 concludes our study by giving details of our achievements and some comments.

WinBUGS codes for implementing the Basic model and the model extended for fat tails and correlated errors are given in Appendix C. Appendix D presents R-codes for implementing formulas of JPR [12] which calculate Bayes Factors.

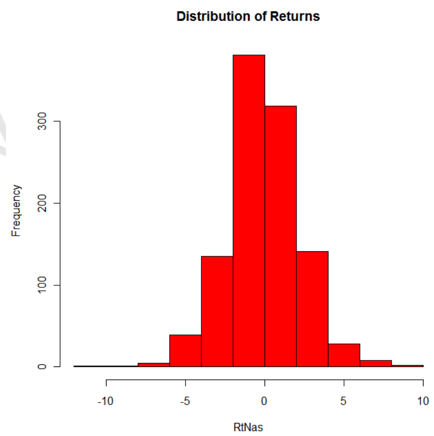

Figure 1.1: Distribution of mean corrected returns of the JSE stock Naspers

Chapter 2

Preliminaries

2.1 Bayesian inference

Bayesian inference is concerned with the posterior distribution of parameters given data and the prior distribution of parameters. We denote by $\mathbf{d}$ the data vector and θ the vector of parameters. Before data one possesses a belief about the parameters which is modeled by $p(\theta)$, which is called the prior distribution, prior to data. Given the prior distribution we have to assume some distribution of data, denoted by $p(\mathbf{d}|\theta)$. For example, if we are modelling stock returns, we assume that returns were generated from a normal distribution. After observing data, our prior beliefs are updated and they are modeled by the posterior distribution, $p(\theta|\mathbf{d})$. To move from the prior distribution, $p(\theta)$, to the posterior distribution, $p(\theta|\mathbf{d})$, we use *Bayes' Theorem*,

$$p(\theta|\mathbf{d}) = \frac{p(\mathbf{d}|\theta)p(\theta)}{p(\mathbf{d})}.$$

Since $p(\mathbf{d})$ is a constant independent of θ, the posterior distribution can be expressed as;

$$p(\theta|\mathbf{d}) \propto p(\mathbf{d}|\theta)p(\theta). \tag{2.1}$$

This forms the basis of Bayesian inference. Samples for inference can then be drawn from the posterior distribution.

Before the data, $p(\mathbf{d}|\theta)$ is defined as the joint distribution $p(d_1, \cdots, d_n|\theta)$ of d_i's given θ. When data has been realised, $p(\mathbf{d}|\theta)$ becomes a function of θ for fixed $\mathbf{d}$, and is called the *likelihood function*.

2.1.1 Prior distributions

The question which arises naturally is, What should the prior distribution, $p(\theta)$, of θ look like? Laplace (1786) observed that computational simplification arises from assuming that the prior has the same form as the likelihood function, when he was solving the Beta-Bernoulli model,

$$\theta \sim p(\theta)$$

$$d_i|\theta \sim Bernoulli(\theta), \quad i = 1, \cdots, n.$$

Thus d_i takes value 1 with probability θ and value 0 with probability $1 - \theta$. Laplace then chooses $p(\theta) = Beta(\alpha_0, \beta_0)$. To see this we consider the likelihood function, $\mathcal{L}(\theta|\mathbf{d})$.

$$\begin{aligned}
\mathcal{L}(\theta|\mathbf{d}) &= \prod_{i=1}^{n} f(d_i|\theta) \\
&= \prod_{i=1}^{n} \theta^{d_i}(1-\theta)^{1-d_i} \\
&= \theta^{\sum_{i=1}^{n} d_i}(1-\theta)^{n-\sum_{i=1}^{n} d_i} \\
&= \theta^{\alpha_0 - 1}(1-\theta)^{\beta_0 - 1},
\end{aligned}$$

where $\alpha_0 = \sum_{i=1}^{n} d_i + 1$ *and* $\beta_0 = n - \sum_{i=1}^{n} d_i + 1$. Thus we have a kernel density of a Beta distribution with parameters α_0 and β_0. If we take the prior to be the beta distribution, the resulting posterior distribution has an analytic solution. The choice of a prior distribution is governed by the need to obtain an analytically tractable and convenient posterior distribution[20].

2.1.2 Conjugate prior distribution

A tractable, analytical posterior distribution $p(\theta|\mathbf{d})$ is not always available. $p(\mathbf{d}|\theta)p(\theta)$ will not always yield a function form that is of known distribution. However, as we have noted from the problem of Beta-Bernoulli model, employing a prior that belongs to the same family of distributions as the data generating function will ensure the existence of an analytic solution for the posterior distribution. Such a prior distribution is called a *conjugate prior* distribution. This class of prior distributions is particularly important in the MCMC methods, as the closed form of the posterior distribution is available. Let us consider the following example in [20]:

Example 1 *Let* $\boldsymbol{r}$ *be a vector of returns which are assumed to have been generated from a normal distribution with a location parameter* μ *and a scale parameter* σ. *The likelihood function is then given by;*

$$\begin{aligned}
\mathcal{L}(\mu, \sigma^2|\boldsymbol{r}) &= \prod_{i=1}^{n} f(r_i|\mu, \sigma^2) \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(r_i - \mu)^2\right\}.
\end{aligned}$$

We define $\bar{r} = \frac{1}{n}\sum_{i=1}^{n} r_i$ *to be the sample mean and* $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(r_i - \bar{r})^2$ *to be the sample variance.*

We now consider

$$\begin{aligned}\sum_{i=1}^{n}(r_i-\mu)^2 &= \sum_{i=1}^{n}((r_i-\bar{r})-(\mu-\bar{r}))^2\\ &= \sum_{i=1}^{n}(r_i-\bar{r})^2-2\sum_{i=1}^{n}(r_i-\bar{r})(\mu-\bar{r})+\sum_{i=1}^{n}(\mu-\bar{r})^2.\end{aligned}$$

Note that the middle term vanishes, since $\sum_{i=1}^{n}(r_i-\bar{r})=0$*. Thus,*

$$\begin{aligned}\sum_{i=1}^{n}(r_i-\mu)^2 &= \sum_{i=1}^{n}(r_i-\bar{r})^2+n(\mu-\bar{r})^2\\ &= \nu s^2+n(\mu-\bar{r})^2, \qquad \textit{where } \nu=n-1.\end{aligned}$$

Hence,

$$\mathcal{L}(\mu,\sigma^2|\boldsymbol{r}) = \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left\{-\frac{\nu s^2}{2\sigma^2}\right\}\exp\left\{-\frac{1}{2(\sigma^2/n)}(\mu-\bar{r})^2\right\}.$$

This is a product of an inverse chi square, χ^{-2}*, distribution in* σ^2 *and a normal distribution in* μ *given* σ^2*.*

It then follows that the prior distributions of μ *and* σ^2 *respectively are*

$$p(\mu|\sigma^2)\sim N(\mu_0,\frac{\sigma^2}{n_0}),$$

and

$$p(\sigma^2)\sim\chi^{-2}(\nu_0,c_0^2),$$

where μ_0*,* n_0*,* ν_0*, and,* c_0^2 *are parameters to be determined outside the model and are called hyperparameters.*

Hyperparameters allow one to incorporate the beliefs one has on parameters before observing the data. Tsay [23], page 548, gives detailed results on conjugate prior distributions and the corresponding posterior distributions.

2.1.3 Bayes factors

In Bayesian analysis, hypothesis testing is concerned with how likely we are inclined towards a particular hypothesis in light of new information carried in the data. In this section we follow a paper by Kass and Raftery [13]. Let $\mathbf{d}$ be the vector of data, which is assumed to have arisen under one of the two hypothesis H_1 or H_2 according to a probability density $p(\mathbf{d}|H_1)$ or $p(\mathbf{d}|H_2)$. Given a prior probability $p(H_1)$ and $p(H_2)$, the data produces a posterior distribution $p(H_1|\mathbf{d})$ and $p(H_2|\mathbf{d})$. Using Bayes' theorem

$$p(H_m|\mathbf{d})=\frac{p(\mathbf{d}|H_m)p(H_m)}{p(\mathbf{d})} \quad for\ m=1,2$$

and

$$p(\mathbf{d}|H_m)=\int p(\mathbf{d}|\theta_m,H_m)p(\theta_m|H_m)d\theta_m,$$

where θ_m is the vector of parameters under H_m and $p(\theta_m|H_m)$ is its prior density.

To summarise the two hypotheses, odds ratios are used. The posterior odds (PO) of H_1 versus H_2 is given by

$$\begin{aligned} PO &= \frac{p(H_1|\mathbf{d})}{p(H_2|\mathbf{d})} \\ &= \frac{p(\mathbf{d}|H_1)p(H_1)}{p(\mathbf{d}|H_2)p(H_2)}. \end{aligned}$$

The *Bayes factor* B_{12} is defined as

$$B_{12} = \frac{p(\mathbf{d}|H_1)}{p(\mathbf{d}|H_2)}. \tag{2.2}$$

The Bayes factor is the ratio of the posterior odds regardless of the prior odds [13]. If the hypothesis H_1 and H_2 are equally likely so that $p(H_1) = p(H_2) = 0.5$, then the Bayes factor, B_{12} is equal to the posterior odds in favor of H_1.

A Bayes factor is a quantity that summarises the evidence provided by the data in favor of a particular hypothesis against another.

2.2 Markov Chain Monte Carlo methods

Let $\theta = (\theta_1, \cdots, \theta_k)$ be the vector of parameters and $\mathbf{d} = (d_1, \cdots, d_n)$ be the data vector. Suppose we are interested in the posterior mean of some measurable function f,

$$\begin{aligned} I(f) &= \mathbb{E}_{p(\theta|\mathbf{d})} f(\theta) \\ &= \int f(\theta)p(\theta|\mathbf{d})d\theta. \end{aligned}$$

If we can simulate G independent and identically distributed (iid) samples $\{\theta^{(g)}\}_{g=1,\cdots,G}$ according to the posterior distribution $p(\theta|\mathbf{d})$, compute

$$I_G(f) = \frac{1}{G}\sum_{g=1}^{G} f(\theta^{(g)}).$$

By the Law of Large Numbers, $I_G(f)$ converges to $I(f)$ almost surely as G becomes large [7]. This procedure is termed *Monte Carlo integration.* In particular, if $f(\theta) = \theta$ then we are estimating the posterior mean of the parameters θ. All properties one might be interested in such as shapes, modes, quartiles etc, of the marginal distribution of θ can easily be inferred from the Monte Carlo sample.

This method is independent of the dimension of the integrand which makes it possible to solve otherwise intractable numerical integrals [16].

The problem now is to sample the iid samples. However it is not always the case that one can sample from $p(\theta|\mathbf{d})$. In practice, the closed form of $p(\theta|\mathbf{d})$ is mostly not known and the normalising constant $p(\mathbf{d})$ is not available. Algorithms such as Metropolis-Hastings algorithm address these problems. Metropolis and Ulam [16] noted that draws from $p(\theta|\mathbf{d})$ need not to be made in an iid manner for the Monte Carlo method to be valid.

This leaves us with generally two simulation classes. *Independent simulations*, algorithms in this class include Acceptance-Rejection algorithm (AR). AR generally works by finding an envelope $h(\theta) \geq p(\theta|\mathbf{d})$ of $p(\theta|\mathbf{d})$, which is easy to simulate from. The idea is that, observations are equally likely to be drawn from $p(\theta|\mathbf{d})$, so one makes a draw θ^* (say) from $h(\theta)$ and reject it if it falls between the density h and p. Else it is accepted with probability

$$\gamma = \frac{p(\theta^*|\mathbf{d})}{h(\theta^*)}. \tag{2.3}$$

One will have to repeat this process until a draw has been accepted.

Dependent simulations will results in dependent draws. Algorithms in this class includes Metropolis-Hastings algorithm and Gibbs sampling algorithm which will generate a *Markov chain*, hence the name *Markov Chain Monte Carlo* methods.

Tierney [22] gives detailed theorems on convergence issues of Markov chains used to explore posterior distributions. Chib and Greenberg [3] give an exposition of the Metropolis-Hastings algorithm.

2.2.1 Metropolis-Hastings algorithm

We define $q(x, y)$ to be the proposal or the candidate generating distribution, which generates the value y given that the process is in state x. We also define

$$\gamma(x, y) = min\left\{\frac{p(y|\mathbf{d})}{q(y,x)} / \frac{p(x|\mathbf{d})}{q(x,y)}, 1\right\},$$

to be the probability that the move is made.

We briefly describe the algorithm in the following sequential order.

1. Initialise the chain, $\theta^{(0)}$,
2. At time t, sample θ from the proposal distribution $q(\theta^{(t)}, \theta)$,
3. Given the proposal value θ, compute $\gamma(\theta^{(t)}, \theta)$ and generate a random variable u form a uniform distribution $U(0, 1)$,
4. If $u \leq \gamma(\theta^{(t)}, \theta)$ then $\theta^{(t+1)} = \theta$ else $\theta^{(t+1)} = \theta^{(t)}$.

Repeating step 2 through to step 4 a G number of times, will result in a Markov chain $\{\theta^{(1)}, \cdots, \theta^{(g)}, \cdots, \theta^{(G)}\}$.

2.2.2 Gibbs Sampling

Let $\theta = (\theta_1, \cdots, \theta_k)$. Here one considers the distribution of one variable when all other variables are fixed. Following the notation by Draper (2000), θ_{-i} is the vector θ with component i omitted. Let $\theta_i^{(t)}$ be the current state of component i. The proposal distribution for component i,

$$h_i(\theta_i|\theta_i^{(t)}, \theta_{-i}, \mathbf{d})$$

depends on the current state of component i, $\theta_{-1}^{(t)}$ which is the current state of θ_{-i} after step $i-1$ of iteration $t+1$ and θ_i is the proposed value at iteration $t+1$.

If we choose the proposal conditional distribution to be the full conditional distribution,

$$p_i(\theta_i|\theta_{-i}, \mathbf{d})$$

then the probability of making a move γ is 1, (Geman and Geman (1984)). The algorithm is called *Gibbs sampling*, and a draw is always accepted. The algorithm iterates as follows,

1. Initialise the chain; $\theta_1^{(0)}, \theta_2^{(0)}, \cdots, \theta_k^{(0)}$,
2. sample $\theta_1^{(t+1)} \sim p(\theta_1|\theta_2^{(t)}, \theta_3^{(t)}, \theta_4^{(t)}, \cdots, \theta_k^{(t)})$,
3. sample $\theta_2^{(t+1)} \sim p(\theta_2|\theta_1^{(t+1)}, \theta_3^{(t)}, \theta_4^{(t)}, \cdots, \theta_k^{(t)})$,
4. sample $\theta_3^{(t+1)} \sim p(\theta_3|\theta_1^{(t+1)}, \theta_2^{(t+1)}, \theta_4^{(t)}, \cdots, \theta_k^{(t)})$,

   Continue in this manner until you,
5. sample $\theta_k^{(t+1)} \sim p(\theta_k|\theta_1^{(t+1)}, \theta_2^{(t+1)}, \theta_3^{(t+1)}, \cdots, \theta_{k-1}^{(t)})$.

We call this a single sweep. To make another sweep, replace t with $t+1$. Making G sweeps, generates a Gibbs sample $\{\theta^{(g)}\}_{g=1,\cdots,G}$.

2.3 The Bivariate Normal distribution

Definition 1 *Let X and Y be two random variable, that are distributed according to a bivariate normal distribution, then the joint probability density function of (X, Y) is,*

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{Q(x,y)}{2(1-\rho^2)}\right\}, \tag{2.4}$$

where

$$Q(x,y) = \left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2, \tag{2.5}$$

μ_X and μ_Y, σ_X and σ_Y are means and standard deviations of X and Y respectively, and ρ is the correlation coefficient of X and Y.

In [10] it is shown that the probability density function $f_{XY}(x,y)$ factorises as;

$$f_{XY}(x,y) = g(y)h(x,y), \tag{2.6}$$

where

$$g(y) = \frac{1}{2\pi\sigma_Y} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma}\right)^2\right\},$$

and

$$h(x,y) = \frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma_X} \exp\left\{-\frac{1}{2\sigma_X^2(1-\rho^2)}\left(x-\mu_X-\frac{\rho\sigma_X}{\sigma_Y}(y-\mu_Y)\right)^2\right\}.$$

For a given value of y, $h(x,y)$ is a probability density function of normal distribution of X with mean

$$\mu_X + \frac{\rho\sigma_X}{\sigma_Y}(y-\mu_Y) \tag{2.7}$$

and variance

$$\sigma_X^2(1-\rho^2). \tag{2.8}$$

And $g(y)$ is just nothing but a probability density function of a normal distribution with mean μ_Y and variance σ_Y^2.

Property 1 *The marginal distributions are normal. That is $f_Y(y) = g(y)$, see [10] for the proof of this property.*

Property 2 *The conditional distributions are normal.*

By definition,

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{XY}(x,y)}{f_Y(y)} \\ &= \frac{g(y)h(x,y)}{g(y)} \\ &= h(x,y). \end{aligned}$$

Thus $f_{X|Y}(x|y)$ is normally distributed with mean and variance given by (2.7) and (2.8) respectively, and this proves Property (2) [10].

Chapter 3

Stochastic Volatility Models

3.1 The basic SVOL model

The basic SVOL model and its implications in relation to empirical evidence of asset returns is discussed in this section. The basic SVOL model of Taylor (1986) is given by:

Let $y_t = r_t - \mu$ be returns in excess of the mean μ at time t, then

$$y_t = \sqrt{h_t}\varepsilon_t \qquad t = 1, \cdots, T, \tag{3.1}$$

$$\log h_t = \alpha + \delta(\log h_{t-1} - \alpha) + \tau\eta_t \qquad t = 2, \cdots, T, \tag{3.2}$$

and

$$\log h_1 \sim N\left(\alpha, \frac{\tau^2}{1-\delta^2}\right) \tag{3.3}$$

where,

$$\begin{pmatrix}\varepsilon_t \\ \eta_t\end{pmatrix} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right)$$

and where $\mu,\ \alpha,\ \delta,\ \tau > 0$ are constants.

Equation (3.1) is the mean equation of returns and Equation (3.2) is an AR(1) process that governs the dynamics of log volatility.

Here the beginning log volatility, $\log h_1$, follows the unconditional distribution. To derive the unconditional distribution, consider Equation (3.2). Using the method of lag operators, Equation (3.2) can be written as

$$(1 - \delta L)\log h_t = \alpha(1-\delta) + \tau\eta_t,$$

where L is the lag operator[1].

[1] $L^i y_t = y_{t-i}$.

Assuming that $|\delta| < 1$, the above equation reduces to

$$\log h_t = \alpha + \tau \sum_{i=0}^{\infty} \delta^i \eta_{t-i}.$$

It then follows that;

$$\begin{aligned} \mathbb{E}(\log h_t) &= \mathbb{E}(\alpha) + \tau \sum_{i=0}^{\infty} \delta^i \mathbb{E}(\eta_{t-i}) \\ &= \alpha, \end{aligned}$$

and

$$\begin{aligned} \mathrm{Var}(\log h_t) &= 0 + \tau^2 \sum_{i=0}^{\infty} \delta^{2i} \mathrm{Var}(\eta_{t-i}) \\ &= \tau^2 \sum_{i=0}^{\infty} \delta^{2i} \\ &= \tau^2 (1/(1-\delta^2)). \end{aligned}$$

Hence,

$$\log h_t \sim N\left(\alpha, \frac{\tau^2}{1-\delta^2}\right). \tag{3.4}$$

δ measures volatility persistence[2], we will see why δ measures volatility persistence shortly. If $\delta = 1$, then the log volatility process is a random walk process and its unconditional variance is undefined. Hence the process is not stationary. To impose stationarity on Equation (3.2) we require $|\delta| < 1$ [6]. τ is the scale parameter of the log volatility disturbance. τ^2 measures the variability of log volatility. α is the long run mean, the unconditional mean, of log volatility. Under the stationarity condition, volatility reverts back to is long run mean [6].

To motivate why δ is the volatility persistence parameter, we consider the autocorrelation function (ACF) of volatility. [21] page 282, shows that the theoretical ACF, $\rho_{s,h}$, of volatility is given by;

$$\rho_{s,h} = \frac{\exp(4\sigma^2\delta^s) - 1}{\exp(4\sigma^2) - 1}, \quad s = 0, 1, 2, \cdots \tag{3.5}$$

where

$$\sigma^2 = \frac{\tau^2}{1-\delta^2}.$$

Under the stationarity condition $\exp(4\sigma^2\delta^s)$ approaches one since δ^s approaches zero as the lag length, s, increases. Hence the ACF given by Equation (3.5) decays to zero at an exponential rate. Thus current volatility has an effect on future volatilities. The extent to which current volatility stretches into future

[2]Mandelbrot (1963) noted that large changes of asset prices tend to be followed by large changes of asset prices. Thus periods of high volatility see large magnitudes of asset returns(both positive and negative) while in periods of low volatility returns do not fluctuate much [20].

volatilities is clearly determined by the size of δ. For this reason δ is the volatility persistence parameter.

In what follows the ability of the basic SVOL model to account for the empirical evidence of asset returns and its shortcomings is briefly discussed. To consider the leptokurtic distribution of returns implied by the basic SVOL model we consider the the kurtosis of returns. Define by k_r the kurtosis of returns, it then follows that

$$k_r = k_\varepsilon \frac{\mathbb{E}[h_t^2]}{(\mathbb{E}[h_t])^2}$$

where k_ε is the kurtosis of ε_t.

Using relation (3.4) and a result in Statistics[3] it follows that

$$\mathbb{E}[h_t] = \exp(\alpha + \frac{1}{2}\sigma^2)$$

and

$$\mathbb{E}[h_t^2] = \exp(2\alpha + 2\sigma^2)$$

where σ^2 is defined in Equation (3.5). k_r then reduces to;

$$k_r = k_\varepsilon \exp(\sigma^2). \tag{3.6}$$

Under the normality assumption Equation (3.6) becomes,

$$k_r = 3\exp(\sigma^2) > 3. \tag{3.7}$$

Hence the model is consistent with the empirical results discussed in paragraph 3 of Section (1.2). However the kurtosis implied by the model is not enough to capture the empirical kurtosis of asset returns. The model does not allow volatility to react in an asymmetric fashion to return shocks since ε_t and η_t are assumed to be independent [20]. Thus the model fails to capture the leverage effects mentioned in Section 1.1. JPR [12] showed that the basic model is not robust in the face of outliers. Thus there is need for extending the model to include these stylised facts.

3.1.1 Parameter estimation

In this section we present the sampling of the basic SVOL model parameters. We give full conditional distributions from which the Gibbs sampling algorithm discussed in Section 2.3.2 is applied.

Let $\theta = (\alpha, \delta, \tau)$ be a vector of parameters, $\mathbf{h} = (h_1, \cdots, h_T)$ be a vector of volatilities, and $\mathbf{y} = (y_1, \cdots, y_T)$ be vector of data. The volatility model can be expressed as a probability model as follows;

$$p(h_1|\theta)$$

[3]If $\ln Y \sim N(\psi, \Sigma)$ then for $k > 0$ $\mathbb{E}(Y^k) = \exp(k\psi + \frac{1}{2}k^2\Sigma)$.

$$p(h_t|h_{t-1}, \theta), \quad t = 2, \cdots, T$$
$$p(y_t|h_t), \quad t = 1, \cdots, T.$$

Thus the volatility model views the data, $\mathbf{y}$, as a vector generated from the probability model $p(\mathbf{y}|\mathbf{h})$, and volatilities $\mathbf{h}$ are assumed to be generated by the probability model $p(\mathbf{h}|\theta)$ [11].

To make inferences about the parameters and volatilities, we are interested in the posterior distribution

$$p(\mathbf{h}, \theta|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{h})p(\mathbf{h}|\theta)p(\theta).$$

However noting that the likelihood function,

$$p(\mathbf{y}|\theta) = \int p(\mathbf{y}|\mathbf{h}, \theta)p(\mathrm{h}|\theta)d\mathbf{h},$$

is a T-dimensional integration with respect to the unknown and unobservable volatilities, making it impossible to have a closed form solution of the posterior distribution. One way of attacking this curse is to use the MCMC methods discussed in Section 2.2. JPR [12] then proposed a single move Gibbs sampler which samples a single parameter at a time.

In Section 2.1 we discussed how to determine conjugate prior distributions. To sample τ^2, the inverse gamma distribution denoted by $\mathcal{IG}(\cdot, \cdot)$ is used for the prior distribution. Thus,

$$p(\tau^2) \sim \mathcal{IG}(\tau_0/2, s_\tau/2),$$

where τ_0 *and* s_τ *are hyperparameters, see Example* (1) *of Section* 2.1.

τ^2 is then sampled from ;

$$\tau^2|\mathbf{y}, \mathbf{h}, \alpha, \delta \sim \mathcal{IG}(a/2, b/2), \tag{3.8}$$

where

$$a = \tau_0 + T$$

and

$$b = s_\tau + (H_1 - \alpha)^2(1 - \delta^2) + \sum_{t=2}^{T}(H_t - \alpha - \delta(H_{t-1} - \alpha))^2 \text{ see [14].}$$

For α, the normal distribution denoted by $N(\cdot, \cdot)$ is used for the prior distribution and it follows that

$$p(\alpha) \sim N(\alpha_0, s_\alpha^2),$$

where α_0 *and* s_α^2 *are hyperparameters.*

It then follows that α is sampled from;

$$\alpha|\mathbf{y}, \mathbf{h}, \delta, \tau^2 \sim N(a/b, 1/a), \tag{3.9}$$

where

$$a = \left((1-\delta^2) + (T-1)(1-\delta)^2\right)/\tau^2 + 1/s_\alpha^2,$$

and

$$b = \left((1-\delta^2)H_1 + (1-\delta)\sum_{t=2}^{T}(H_t - \delta H_{t-1})\right)/\tau^2 + \alpha_0/s_\alpha^2.$$

To sample δ, JPR [12] used the normal distribution $N(0,10)$ for the prior distribution. To impose stationarity, they truncated the posterior distribution [12]. Another alternative suggested by KSC [14] is to use the Beta distribution denoted by $Be(\cdot,\cdot)$ which has a support on the interval $(0,1)$. Letting

$$\delta = 2\delta^* - 1$$

where

$$\delta^* \sim Be(\delta_1, \delta_2),$$

and where $\delta_1, \delta_2 > 0.5$ *are hyperparameters.*

It then follows that $\delta \in (-1,1)$. KSC [14] used $\delta_1 = 20$ and $\delta_2 = 1.5$ which implies a prior mean of 0.86 for δ. The posterior density of δ is determined by;

$$p(\delta|\mathbf{y},\mathbf{h},\alpha,\tau^2) \propto p(\mathbf{h}|\alpha,\delta,\tau^2)p(\delta).$$

KSC noted that $\log p(\mathbf{h}|\alpha,\delta,\tau^2)$ is a concave function in δ, and hence the acceptance algorithm mentioned in Section 2.3 can be applied, see [14].

For the derivation of the full conditional distributions of parameters τ^2 and α see the appendix.

Sampling volatilities

We first remind ourselves of what we wish to achieve. We seek to sample from the joint posterior distribution $p(\mathbf{h},\theta|\mathbf{y})$. JPR [12] break the joint distribution into two Gibbs blocks $p(\theta|\mathbf{h})$ and $p(\mathbf{h}|\theta,\mathbf{y})$. The full conditional distributions of $p(\theta|\mathbf{h})$ had been dealt with. The task now remains is to sample $\mathbf{h}$ from $p(\mathbf{h}|\theta,\mathbf{y})$, JPR [12] break $p(\mathbf{h}|\theta,\mathbf{y})$ into T univariate conditional distributions and augmented the volatility space with h_0 and h_{T+1}. As mentioned in [20], the Markov property will imply that h_t has its full conditional distribution defined as;

$$p(h_t|\mathbf{h}_{-t},\theta,y_t) \propto p(h_t|h_{t-1})p(h_{t+1}|h_t)p(y_t|h_t).$$

Noting that there is a slight difference in the specification of the log volatilities with that used by JPR, we derive the kernel of the posterior distribution of volatilities otherwise everything will remain unchanged. In this study we have considered a centralised log volatility process so as to improve convergence, see for example [5].

It can be shown easly that

$$p(h_t|h_{t-1}) \propto \frac{1}{h_t}\exp\left\{-\frac{(H_t - (\alpha + \delta(H_{t-1}-\alpha)))^2}{2\tau^2}\right\},$$

$$p(h_{t+1}|h_t) \propto \exp\left\{-\frac{(H_{t+1} - (\alpha + \delta(H_t - \alpha)))^2}{2\tau^2}\right\},$$

and

$$p(y_t|h_t) \propto \frac{1}{\sqrt{h_t}} \exp\left\{-\frac{y_t^2}{2h_t}\right\},$$

where $H_t = logh_t$.

By combining the above distributions the conditional distribution reduces to;

$$p(h_t|h_{t-1}, h_{t+1}, \theta, y_t) \propto \frac{1}{\sqrt{h_t}} \exp\left\{-\frac{y_t^2}{2h_t}\right\} \times \frac{1}{h_t} \exp\left\{-\frac{(H_t - \mu_t)^2}{2\sigma^2}\right\},$$

where

$$\mu_t = [\alpha(1-\delta)^2 + \delta(H_{t+1} - H_{t-1})]/(1 + \delta^2),$$

and

$$\sigma = \frac{\tau^2}{1 + \delta^2}.$$

To construct the proposal distribution, $q(\cdot, \cdot)$, JPR [12] noted that the kernel of the conditional distribution is the product of two kernels, the inverse gamma kernel and the lognormal kernel [20]. JPR [12] proposed to approximate the lognormal kernel with the inverse gamma kernel by equating their means and variances [20]. For the derivation of the proposal distribution see [20] or [12]. The proposal distribution is defined by;

$$q(h_t|\cdot) \propto h_t^{-(\psi+1)} \exp(-\varphi_t/h_t), \tag{3.10}$$

where

$$\psi = \frac{\exp(\sigma^2) + 1}{2(\exp(\sigma^2) - 1)},$$

and

$$\varphi_t = \frac{3\exp(\sigma^2) - 1}{2(\exp(\sigma^2) - 1)} \exp(\mu_t + 0.5\sigma^2) + \frac{y_t^2}{2}.$$

To this end we can now appeal to the Gibbs sampler, to sample from $p(\mathbf{h}, \theta|\mathbf{y})$. However JPR [12] combined the rejection sampling and the Metropolis-Hastings algorithm by first applying the rejection method and if the draw is accepted it then enters the Metropolis-Hastings when sampling volatilities, see [12].

3.2 Model extensions

In the previous section shortcomings of the basic SVOL model were discussed, in this section possible solutions to these shortcomings are presented. For full conditional distributions for sampling of parameters under the extended models see JPR [12].

3.2.1 Extending for fat tails

To model fat tails, the t-distribution is a natural choice since it has fatter tails than the normal distribution.

Definition 2 *Let $\boldsymbol{X}$ be a vector of continuous random variables with location parameter $\bar{\mu}$ and scale parameter Σ. If the pdf of $\boldsymbol{X}$ can be expressed into the following mixture representation*

$$f(x|\bar{\mu}, \Sigma) = \int_0^\infty N(x|\bar{\mu}, \kappa(\lambda)\Sigma)p(\lambda)d\lambda,$$

where $N(x|\cdot,\cdot)$ is a multivariate normal pdf, $\kappa(\lambda)$ is a positive function of λ and $p(\cdot)$ is a pdf defined on $\mathbb{R}^+$, then the pdf of $\boldsymbol{X}$ has a scale mixture of normals representation.

λ is referred to as the mixing parameter and $p(\cdot)$ the mixing density of the scale mixture of normals [24].

The distribution of ε_t can be modeled as a scale mixture of normals. JPR [12] considered the mixing parameter to be a latent variable at time t. Another alternative is to model both ε_t and η_t with the t-distribution (Choy et al 2008) [24]. Under the specification of JPR, y_t can now be expressed as

$$y_t = \sqrt{h_t}\sqrt{\lambda_t}z_t,$$

where z_t is a white noise. The t-distribution representation of scale mixture of normals implies that

$$\sqrt{\lambda_t}z_t \sim t_\nu,$$

where $\lambda_t|\nu \sim IG(\nu/2, \nu/2)$ for a given value of ν.

JPR [12] used a uniform [3, 40] discrete prior of ν. Noting that the variance of t_ν, is $\nu/(\nu-2)$, this prior specification guarantees existence of a finite variance. Another possible prior on ν is a χ^2_4 which is relatively flat over the posterior range and has a prior mean of 4.

It is worth noting that JPR [12] used a block sampling for the additional parameters, ν and $\lambda|\nu$. This was to avoid the problem of ν getting absorbed into lower bound. The problem was noted by Eraker et al (1998) [12].

3.2.2 Extending for correlated errors

To model the leverage effects, a correlation coefficient, ρ, between the mean innovation and the innovation in the log volatility equation is introduced. Thus,

$$\begin{pmatrix}\varepsilon_t\\ \eta_t\end{pmatrix} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right)$$

The prior used on $\rho = \text{corr}(\varepsilon_t, \eta_t)$ is the uniform distribution denoted by $U(\cdot,\cdot)$. The natural choice for this specification is

$$\rho \sim U(-1,\ 1).$$

This way the empirically observed asymmetry is implied by negative correlation between ε_t and η_t.

3.2.3 Full model

We complete this section by looking at the model extended for both the fat tails and the correlated errors (full model). When modelling fat tails and correlated errors, one considers a model extended for correlated errors and replace y_t with $y_t/\sqrt{\lambda_t}$, see [12]. This will result in the model extended for both fat-tails and correlated errors. The model is given by;

$$y_t = \sqrt{h_t \lambda_t} z_t, \quad t = 1, \cdots, T, \tag{3.11}$$

$$\log h_t = \alpha + \delta(\log h_{t-1} - \alpha) + \tau \eta_t, \quad t = 2, \cdots, T, \tag{3.12}$$

$$\lambda_t \sim IG(\nu/2, \nu/2), \quad t = 1, \cdots, T, \tag{3.13}$$

where

$$\begin{pmatrix} z_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

and

$$\log h_1 \sim N(\alpha, \frac{\tau^2}{1-\delta^2}).$$

The above model only models the fat tails in the mean equation. If we choose to model for fat tails in both innovations, and the correlation between innovations, then we use the bivariate t-distribution. The t-distribution was then expressed as a scale mixture of normals by Choy et al [24]. In this case the model is given by;

$$y_t = \sqrt{h_t} \varepsilon_t, \quad t = 1, \cdots, T, \tag{3.14}$$

$$\log h_t = \alpha + \delta(\log h_{t-1} - \alpha) + \tau \eta_t, \quad t = 2, \cdots, T, \tag{3.15}$$

$$\lambda_t \sim IG(\nu/2, \nu/2), \quad t = 1, \cdots, T, \tag{3.16}$$

where

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \lambda_t \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),$$

and

$$\log h_1 \sim N(\alpha, \frac{\lambda_1 \tau^2}{1-\delta^2}).$$

In this mini-thesis we are going to rely on the user friendly WinBUGS software which requires the specification of the model and the prior distribution of parameters. For the discussion of algorithm on sampling from the full conditional distributions under the extended model see [12].

3.3 Volatility estimation using WinBUGS

3.3.1 WinBUGS

Bayesian inference **U**sing **G**ibbs **S**ampling (BUGS) is a program for Bayesian modelling, designed to handle complex models for which there is no analytic solution [15]. The current version of BUGS is WinBUGS 1.4 and is available on the website *http: www.mrc-bsu.cam.ac.uk/bugs/*.

There are two ways of representing a model in WinBUGS, which are graphical representation and the text-based BUGS language. In this mini-thesis we used the text-based language. For the codes see the appendix.

WinBUGS will determine the sampling method for the estimation of the target distribution, see [15], pg 46. If a node's[4] full conditional distribution is available in closed form, the software can identify the closed form, and if the nodes full conditional distribution is not available in closed form, the software examines the circumstances and chooses an approximate sampling method [15].

To make inferences using WinBUGS one is only required to know the prior distribution of variables. The program only requires the specification of the model and the prior distributions, data and initial values of the model parameters.

3.3.2 Model implementation

In this mini-thesis we consider a full model with fat tails in both innovations, given by Equations (3.14) – (3.16). To implement this model using WinBUGS, we first appeal to the results in Section 2.4 on bivariate normal distribution. We first consider the expectations and variances of returns and log volatilities under our selected model. Thus,

$$\mathbb{E}(\log h_t) = \alpha + \delta(\log h_{t-1} - \alpha),$$

$$\mathrm{Var}(\log h_t) = \lambda_t \tau^2,$$

$$\mathbb{E}(y_t) = 0,$$

$$\mathrm{Var}(y_t) = h_t \lambda_t.$$

Using Property 2 of Section 2.4,

$$\begin{aligned}
\mathbb{E}(y_t|\log h_t) &= 0 + \rho \frac{\sqrt{h_t \lambda_t}}{\sqrt{\lambda_t \tau^2}} (\log h_t - \alpha - \delta(\log h_{t-1} - \alpha)) \\
&= \rho \frac{\sqrt{h_t}}{\tau} (\log h_t - \alpha - \delta(\log h_{t-1} - \alpha)),
\end{aligned}$$

and

$$\mathrm{Var}(y_t|\log h_t) = h_t \lambda_t (1 - \rho^2).$$

[4]In WinBUGS, all variables are called nodes.

Since conditional and marginal distributions are normally distributed, by Property (2) of Section 2.4, the full model is then given by

$$\log h_1 \sim N\left(\alpha, \frac{\lambda_t \tau^2}{1-\delta^2}\right), \tag{3.17}$$

$$\log h_t \sim N\left(\alpha + \delta(\log h_{t-1} - \alpha), \lambda_t \tau^2\right), \quad t = 2, \cdots, T, \tag{3.18}$$

$$y_1 | h_1 \sim N\left(\rho \frac{\sqrt{h_1}}{\tau} \sqrt{1-\delta^2}(\log h_1 - \alpha), h_1 \lambda_1 (1-\rho^2)\right), \tag{3.19}$$

$$y_t | h_t, h_{t-1} \sim N\left(\rho \frac{\sqrt{h_t}}{\tau} (\log h_t - \alpha - \delta(\log h_{t-1} - \alpha)), h_t \lambda_t (1-\rho^2)\right), \quad t = 2, \cdots, T, \tag{3.20}$$

and

$$\lambda_t | \nu \sim IG(\nu/2,\ \nu/2), \quad t = 1, \cdots, T. \tag{3.21}$$

Prior distributions

We adopted the prior distributions used by JPR [12] and KSC [14].

1. $\alpha \sim N(0, 10)$,
2. $\delta = 2\delta^\star - 1$, where $\delta^\star \sim Be(20, 1.5)$,
3. $\tau^2 \sim \mathcal{IG}(2.5, 0.025)$,
4. $\rho \sim U(-1, 1)$ and
5. $\nu \sim \chi^2_4$.

The model given by Equations (3.17) – (3.21) together with the above prior distributions can now be implemented in WinBUGS, see [17].

Chapter 4

Empirical Results

4.1 Data

In this study we employed the Johannesburg Stock Exchange (JSE) market data. The data was downloaded from Datastream. We used both the daily and weekly series. The daily series comprises two indices, All Share Index (ALSI) and the Top 40 Index (TOPI40), two stocks Anglo Platinum and Naspers, and two exchange rates EUR/ZAR and USD/ZAR where EUR is the European currency, USD is the United States of American currency and ZAR is the South African currency. These series were from 12/8/2006-12/31/2010. The weekly series comprises two indices Oilgas index and the Telecom index, and three portfolios, first, fifth and tenth deciles of the top 165 shares. The period of the data for the Oilgas index is 7/3/1995-12/27/2010 and for the Telkom index is 1/15/1995-12/27/2010. Table 4.1 gives details of the series used in this study. Table 4.2 shows the composition of each portfolio formed by equal weights.

4.2 Methodology

To implement the model, we adopted a model form given by Equations (3.17) – (3.21). We convert this form into a BUGS language, for codes see Appendix A. We used mean corrected returns , y_t, calculated as a percentage using the formula;

$$y_t = 100 \left\{ \ln \left(\frac{PI_t}{PI_{t-1}} \right) - \frac{1}{T} \sum_{t=1}^{T} \ln \left(\frac{PI_t}{PI_{t-1}} \right) \right\}$$

where PI_t is the price index at time t.

4.3 Findings

4.3.1 Convergence diagnostics

To make inference we need to check that the chain has converged to the target density. To check for convergence we used the chain history which is a time series plot of parameter. A chain that has converged will have its history plot exhibiting randomness as the chain iterates. This shows that the chain has found the region of high likelihood and it is iterating over the posterior distribution of interest [5]. Figure B.1 shows the chain history for the full model of the series OILGAS. For parameters α, δ, and ρ graphs are exhibiting randomness. The chain of parameter τ seems to get stuck in some parts of the parameter space, this can be seen from a clear pattern exhibited by the history plot of the parameter, see figure B.1. However it must be noted that the pattern is not that strong and the chain still retains high degree of randomness.

We also used the Gelman-Rubin statistic, R, which is easily accessible in WinBUGS. To calculate R, two or more chains need to be initiated at different starting points. R compares the ratio of pooled variance (variance between samples and variance within sample) to variance within sample [18]. Once convergence has been reached, R is 1. WinBUGS calculates the R statistic and plots it against iteration number. Convergence can now be easily assessed from these plots. Figure B.2 shows the plots of the Gelman-Rubin statistic for series OILGAS. The R statistic is given by the red line. From the graphs given by Figure B.2, the chains of all parameters seem to have converged by the 15 000^{th} iteration.

In this mini-thesis we used a burning-in period of 50 000 with a follow up of 250 000 for all the series. Following results on convergence diagnostics, we are certain that the samples were generated from the target density. Figure B.3 shows the kernel density of parameters of the full model for the index OILGAS.

4.3.2 Posterior analysis of weekly series

Table 4.3 summaries results for weekly series. Volatility persistence measured by δ is significantly high, with mean δ close to one. The mean of δ for indices TELCOM and OILGAS is greater than 0.98 in both series. Among the firm classes, medium firms indexed by P5 exhibit the highest value of mean δ of about 0.95 and small firms indexed P10 have the smallest mean δ of about 0.94. This indicates that market participants trades more in large and medium stocks.

We now turn to leverage effects mentioned in Section 1.1, measured by ρ. The mean ρ for large firms indexed by P1, and small firms is below -0.42. Leverage effects seem to be increasing with firms size, that is large firms are highly leveraged than small firms. OILGAS and TELCOM indices exhibits very low leverage effects and the 95% confident interval of these indices contains zero which is an indication that leverage effects may not be significant. The parameter ν is the number of degrees of freedom in the Student-t distribution which models fat tails. However as noted by JPR(2004), we cannot assess the extent to which the value of ν supports the fat tail distribution. Large firms have mean ν of 13 and small firms having the smallest value of about 6.

4.3.3 Posterior analysis of daily series

Table 4.4 summaries results for daily series. Volatility persistence is quite high in the daily series when compared with the weekly series. This is consistent with the temporal aggregation of weekly returns [12]. The variability of volatility is quite low compared to the weekly series. ALSI and TOPI40 have almost equal mean of ρ. These two indices exhibits very high leverage effects. Thus, stock price falls are associated with very sharp increase in volatility. The exchange rates show no evidence of leverage effects at all with mean of ρ greater than 0.6 The lower quartile of mean ρ is greater than 0.3 and 0.4 for EUR/ZAR and USD/ZAR respectively.

4.3.4 Posterior odds analysis

Table 4.5 gives the results of the estimated Bayes factors. ALSI, P10 and Anglo Platinum do not favor the fat tailed model when compared with the basic model. When comparing the fat tailed model and the full model, all the series favor the full model although it is not as strong as some series when comparing the basic model and the fat tailed model. To get the Bayes Factors for the basic model against the full model we multiply the Bayes Factors of the basic model versus the fat tail model and the fat tail model versus the full model [12]. All the series except ALSI and P10 showed overwhelmingly evidence in support of the full model. Even though some series favors the basic model when compared with the fat tailed model, when combined with leverage effect the data had weaker evidence to support the basic model.

4.3.5 Unit root test

In this section we test for the presents of unit roots in series being studied just to confirm that the series are indeed not unit root processes.

The log volatility process is of the form,

$$x_t = a_0 + \delta x_{t-1} + \eta'_t,$$

where $\eta'_t \sim N(0, \tau^2)$ *and* $x_t = \log h_t$.

Subtracting x_{t-1} form all sides of the above equation yields the following difference equation,

$$\Delta x_t = a_0 + \gamma x_{t-1} + \eta'_t.$$

The null hypothesis that the series has a unit root is equivalent to saying $\gamma = 0$, see [6]. Thus,

$$H_0 : \ \gamma = 0 \text{ (Unit root)}$$

$$H_1 : \ \gamma \neq 0$$

In this test the Augmented Dickey-Fuller (ADF) test is used, see [6] page 181. If the ADF statistic is less than the critical value the null hypothesis is rejected else we fail to reject the null hypothesis. Table 4.6 shows results of unit root test. In all the series the null hypothesis was rejected at 1% significant level. A Durbin-Watson statistic is close to two, which is an indication of small residual autocorrelations. Thus we conclude that the series are not unit root processes.

Table 4.1: Series used in this study

Indices	Portfolios	Shares	Exchange rates
ALSI(JSEOVER)	P1	ANGLO PLUTINUM(930523)	USD/ZAR(USSARCM)
TOPI40(JSECA40)	P5	NASPERS(152214)	EUR/ZAR(SAEURSP)
OILGAS(OILGSSA)	P10	-	-
TELKOM(TELCMSA)	-	-	-

Note: *The code in the brackets is the FTSE/JSE code.*

Table 4.2: Portfolio constituents

JSE SHARES								
P1			P5			P10		
MARKET CAP RANK	JSE CODE	SHARE	MARKET CAP RANK	JSE CODE	SHARE	MARKET CAP RANK	JSE CODE	SHARE
1	BIL	BHP BILLITON	64	AFE	AECI	150	SFN	SASFIN
2	AGL	ANGLO AMERICAN	66	MDC	MEDI-CLINICRP	151	SIM	SIMMER AND JACK MINES
3	SAB	SABMILLER	68	PAP	PANGBOURNE PROP	153	BDM	BUILDMAX
4	MTN	MTN GROUP	70	JDG	JD GROUP	155	CMH	COMBINED MOTOR HOLDINGS
5	SOL	SASOL	71	FPT	FOUNATIN HEAD PROPERTY TRUST	156	COM	COMAIR
6	SBK	STANDARD BANK	72	HYP	HYPROP INVERSTIMENTS	157	DTC	DATACENTRIX
8	NPN	NASPERS	73	NHM	NORTHAM PLATINUM	158	DGC	DIGCORE
9	ANG	ANGLOGOLD ASHANTI	74	SAC	SA CORPORATE REAL ESTATE	159	ESR	ESORFRANKI
10	IMP	IMPALA PLATINUM	75	DTC	DATATEC	160	GIJ	GIJIMA AST GROUP
11	OML	OLD MUTUAL	78	CPL	CAPITAL PROPERTY FUND	161	KAP	KAP INTERNATIONAL
12	FSR	FIRST RAND LIMITED	79	ILV	ILLOVO SUGAR	163	PET	PETMIN
13	GFI	GOLD FIELDS	80	GNP	GRINDROD	165	MML	METMAR
14	AMS	ANGLO PLATINUM	-	-	-	-	-	-
15	SLM	SANLAM	-	-	-	-	-	-
16	SHP	SHOPRITE	-	-	-	-	-	-

Note: *These shares are extracted from the JSE top 165 shares ranked by market capitalisation as at 07 Sep 2010. Since one decile of 165 is 16 (rounded off to the nearest whole number), portforlio P5 and P10 have some of the shares droped becauses they where only included at the JSE a few years ago and they do not have enough information.*

Table 4.3: Posterior analysis for weekly series

	TELCMSA	OILGASSA	P1	P5	P10
α	3.062	2.418	2.13	1.192	2.267
	0.456	0.418	0.175	0.251	0.195
	(2.245, 3.875)	(1.554, 3.044)	(1.808, 2.494)	(0.729, 1.641)	(1.911, 2.631)
δ	0.978	0.977	0.95	0.966	0.939
	0.012	0.01	0.018	0.017	0.033
	(0.949, 0.996)	(0.956, 0.996)	(0.908, 0.98)	(0.925, 0.991)	0.862, 0.986)
τ	0.149	0.127	0.193	0.1275	0.162
	0.036	0.024	0.037	0.0307	0.057
	(0.092, 0.24)	(0.086, 0.178)	(0.127, 0.274)	(0.076, 0.197)	(0.081, 0.3)
ρ	-0.036	-0.197	-0.682	-0.443	-0.427
	0.158	0.152	0.095	0.136	0.128
	(-0.341, 0.276)	(-0.474, 0.111)	(-0.837, -0.47)	(-0.661, -0.13)	(-0.654, -0.154)
ν	8.301	9.255	13.22	8.918	6.08
	2.387	2.498	3.607	2.579	1.662
	(4.955, 13.78)	(5.572, 15.4)	(7.74, 21.69)	(5.185, 15.2)	(3.875, 10.32)

Note: *For each parameter the first number is the posterior mean, below the first number is the standard deviation and the two numbers in brackets is the* 95% *confident interval. P1, P5, P10 are large, medium and small firms respectively.*

Table 4.4: Posterior analysis for daily series

	JSEOVER	JSECA40	AngloPlatinum	Naspers	EUR/ZAR	USD/ZAR
α	0.936	1.051	1.761	1.382	-0.313	0.145
	0.268	0.314	0.533	0.246	0.319	0.252
	(0.457, 1.494)	(0.511, 1.731)	(0.735, 2.68)	(0.952, 1.862)	(-0.856, 0.312)	(-0.302, 0.65)
δ	0.989	0.989	0.991	0.981	0.987	0.982
	0.003	0.003	0.005	0.009	0.006	0.007
	(0.983, 0.995)	(0.982, 0.995)	(0.98, 0.998)	(0.962, 0.996)	(0.973, 0.997)	(0.964, 0.994)
τ	0.13	0.141	0.108	0.115	0.099	0.131
	0.015	0.017	0.0174	0.022	0.018	0.023
	(0.103, 0.16)	(0.098, 0.1671)	(0.078, 0.145)	(0.071, 0.163)	(0.071, 0.141)	(0.093, 0.184)
ρ	-0.897	-0.8856	-0.394	-0.564	0.611	0.646
	0.049	0.0693	0.151	0.142	0.121	0.099
	(-0.973, -0.783)	(-0.959, -0.761)	(-0.669, -0.086)	(-0.838, -0.263)	(0.343, 0.812)	(0.426, 0.808)
ν	15.06	15.64	14	11.85	10.45	13.42
	3.322	3.888	3.637	3.051	2.405	2.423
	(9.24, 23.65)	(9.62, 24.65)	(8.5, 22.65)	(7,35 19.17)	(6.74, 16.11)	(8.193, 21.48)

Note: *EUR/ZAR and USD/ZAR are the exchange rates of Euro and United States of America Dollar against the South African Rand respectively.*

Table 4.5: Bayes Factors

	$BF_{Basic/Fattail}$	$BF_{Fattail/Full}$	$BF_{Basic/Full}$
ALSI	2.577e+06	0.06	1.541e+05
TOPI40	0.981	0.987	8.196e−05
TELCOM	4.925e−04	0.35	1.724e−04
OILGAS	2.79e−07	0.373	1.041e−07
P1	0.521	0.187	0.974
P5	1.03e−10	0.187	1.926e−11
P10	9.839e+10	0.009	8.855e+08
AngloPlatinum	2.793	0.228	0.637
Naspers	7.462e−14	0.287	2.142e−14
USD/ZAR	0.109	0.066	7.159e−03
EUR/ZAR	2.994e−06	0.289	8.66e−07

Note: *$BF_{Basic/Fattail}$ is the Bayes Factor of the basic model against the fat tailed model. If the data favors the fat tail, we expect the Bayes factor to be less than zero.*

Table 4.6: ADF Test

	ADF stat	Prob.	Critical value 1% sig. level	DW stat
ALSI	-8.033	0.0000	-3.436	2.002
TOPI40	-7.668	0.0000	-3.436	1.998
TELCOM	-5.994	0.0000	-3.439	2.003
OILGAS	-6.382	0.0000	-3.438	1.984
P1	-6.768	0.0000	-3.441	1.999
P5	-7.396	0.0000	-3.441	2.001
P10	-5.847	0.0000	-3.441	1.999
AngloPlatinum	-6.177	0.0000	-3.436	2.002
Naspers	-7.371	0.0000	-3.436	1.999
USD/ZAR	-7.988	0.0000	-3.436	1.999
EUR/ZAR	-8.297	0.0000	-3.436	2.003

Note: *The ADF stat is the Augmented Dickey Fuller test statistic. And DW stat is the Durbin-Watson statistic.*

Chapter 5

Conclusion

The following main objectives have been achieved;

1. Review of the basic stochastic volatility model of Taylor(1986) under Bayesian approach and its extensions by JPR (2004), and
2. Found evidence to support a model extended for fat-tails and correlated errors against a basic model of Taylor (1986).

It has been found that the South African market is characterised by periods when volatility is not stationary and the stock market is characterised by high leverage effects.

Running MCMC simulations in WinBUGS is fairly simple. It only requires one's patience as running long chains requires constant monitoring. The main disadvantage with WinBUGS is the single move updating, which is very slow in converging. Volatilities will be updated one at a time. To improve convergence, KSC (1995) proposed a Monte Carlo procedure that allows for sampling of all log volatilities at once, block sampling. Recently Omori et al (2007) developed the idea of KSC (1995) and proposed an algorithm based on the multivariate normal approximation of the conditional posterior density [19]. Their algorithm has an advantage of sampling parameters in blocks, hence improved convergence. However WinBUGS still retains high degree of flexibility when considering exploiting different models. For example, the two codes provided in appendix are almost identical. It is just a matter of manipulating a few lines to obtain the desired model.

To this end we conclude that the model extended for fat tails and correlated errors is more practical. Not only that the model is superior to the basic model, it provides real economic state of the market. Thus the model extended for fat tails and correlated errors is not only important in quantitative finance, it is also useful to policy makers. It provides more information on the direction of the market hence good decisions on implementation of macroeconomic policies, say, are made. Thus evidence provided by this study supports the SVOL model extended for fat tails and correlated errors in estimating stochastic volatility in the South African market.

Appendix A

Full conditional distributions

The derivation of full conditional distribuions of parameters τ^2 and α is presented.

Posterior distribution of τ^2

$$p(\tau^2) \propto (\tau^2)^{-(\tau_0^2/2+1)} \exp\left(-\frac{s_\tau}{2\tau^2}\right).$$

Let $H_t = \log h_t$, the posterior distribution is defined by;

$$\begin{aligned} p(\tau^2|\mathbf{y},\mathbf{h},\alpha,\delta) &\propto p(\mathbf{h}|\alpha,\delta,\tau^2)p(\tau^2) \\ &\propto p(h_1|\alpha,\delta,\tau^2)\prod_{t=2}^{T} p(h_t|h_{t-1},\alpha,\delta,\tau^2)p(\tau^2). \end{aligned}$$

The density distribution of h_1 and h_t can easily be shown to be

$$f(h_1|\alpha,\delta,\tau^2) = \frac{1}{h_1\sqrt{2\pi s_H^2}} \exp\left\{-\frac{(H_1-\alpha)^2(1-\delta^2)}{2\tau^2}\right\},$$

and

$$f(h_t|\alpha,\delta,\tau^2) = \frac{1}{h_t\sqrt{2\pi\tau^2}} \exp\left\{-\frac{(H_t-\alpha-\delta(H_{t-1}-\alpha))^2}{2\tau^2}\right\},$$

where $s_H^2 = \dfrac{\tau^2}{1-\delta^2}$.

Thus the posterior distribution of τ^2 becomes

$$\begin{aligned} p(\tau^2|\mathbf{y},\mathbf{h},\alpha,\delta) &\propto \frac{1}{(\tau^2)^{T/2}} \exp\left\{-\frac{(H_1-\alpha)^2(1-\delta^2)}{2\tau^2}\right\} \prod_{t=2}^{T} \exp\left\{-\frac{(H_t-\alpha-\delta(H_{t-1}-\alpha))^2}{2\tau^2}\right\} \\ &\times \exp(-\frac{s_\tau}{2\tau^2})(\tau^2)^{-(\tau_0^2/2+1)} \\ &= \exp\left\{-\frac{s_\tau + (H_1-\alpha)^2(1-\delta^2 + \sum_{t=2}^{T}(H_t-\alpha-\delta(H_{t-1}-\alpha))^2}{2\tau^2}\right\} \\ &\times (\tau^2)^{-(\tau_0+T)/2-1}. \end{aligned}$$

Hence

$$p(\tau^2|\mathbf{y},\mathbf{h},\alpha,\delta) \propto \mathcal{IG}(a/2, b/2), \tag{A.1}$$

where

$$a = \tau_0 + T$$

and

$$b = s_\tau + (H_1 - \alpha)^2(1-\delta^2) + \sum_{t=2}^{T}(H_t - \alpha - \delta(H_{t-1} - \alpha))^2.$$

Posterior distribution of α

$$p(\alpha) \propto \exp\left(-\frac{(\alpha - \alpha_0)^2}{2s_\alpha^2}\right).$$

The posterior distribution is defined by,

$$\begin{aligned}
p(\alpha|\mathbf{y},\mathbf{h},\delta,\tau^2) &\propto p(\mathbf{h}|\alpha,\delta,\tau^2)p(\alpha) \\
&\propto p(h_1|\alpha,\delta,\tau^2)\prod_{t=2}^{T} p(h_t|h_{t-1},\alpha,\delta,\tau^2)p(\alpha) \\
&\propto \exp\left\{-\frac{(H_1-\alpha)^2(1-\delta^2)}{2\tau^2} - \frac{\sum_{t=2}^{T}(H_t-\alpha-\delta(H_{t-1}-\alpha))^2}{2\tau^2} - \frac{(\alpha-\alpha_0)^2}{2s_\alpha^2}\right\} \\
&= \exp\left\{-\frac{(\alpha^2 - 2\alpha H_1 + H_1^2)(1-\delta^2)}{2\tau^2}\right\} \\
&\times \exp\left\{-\frac{\sum_{t=2}^{T}[(H_t - \delta H_{t-1})^2 - 2\alpha(1-\delta)(H_t - \delta H_{t-1}) + \alpha^2(1-\delta)^2]}{2\tau^2}\right\} \\
&\times \exp\left\{-\frac{(\alpha^2 - 2\alpha\alpha_0 + \alpha_0^2)}{2s_\alpha^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\frac{(\alpha^2 - 2H_1\alpha)(1-\delta^2)}{\tau^2} + \frac{\alpha^2(T-1)(1-\delta)^2}{\tau^2} - \frac{2\alpha(1-\delta)}{\tau^2}\right]\right\} \\
&\times \exp\left\{-\frac{1}{2}\left[\sum_{t=2}^{T}(H_t - \delta H_{t-1}) + \frac{\alpha - 2\alpha\alpha_0}{s_\alpha^2}\right]\right\}.
\end{aligned}$$

We define

$$a = \left((1-\delta^2) + (T-1)(1-\delta)^2\right)/\tau^2 + 1/s_\alpha^2,$$

and

$$b = \left((1-\delta^2)H_1 + (1-\delta)\sum_{t=2}^{T}(H_t - \delta H_{t-1})\right)/\tau^2 + \alpha_0/s_\alpha^2.$$

The posterior distribution becomes,

$$\begin{aligned}
p(\alpha|\mathbf{y},\mathbf{h},\delta,\tau^2) &\propto \exp\left\{-\frac{1}{2}(a\alpha^2 - 2b\alpha)\right\} \\
&= \exp\left\{-\frac{1}{2}a(\alpha^2 - \frac{2b}{a}\alpha)\right\} \\
&\propto \exp\left\{-\frac{1}{2}a(\alpha - \frac{b}{a})^2\right\}.
\end{aligned}$$

Hence

$$p(\alpha|\mathbf{y}, \mathbf{h}, \delta, \tau^2) \propto N(\hat{\alpha}, \sigma^2_\alpha), \tag{A.2}$$

where

$$\hat{\alpha} = \frac{b}{a}$$

and

$$\sigma^2_\alpha = a^{-1}.$$

Relations $(A.1)$ and $(A.2)$ defines the full conditional distributions of τ^2 and α respectively.

Appendix B

Figures

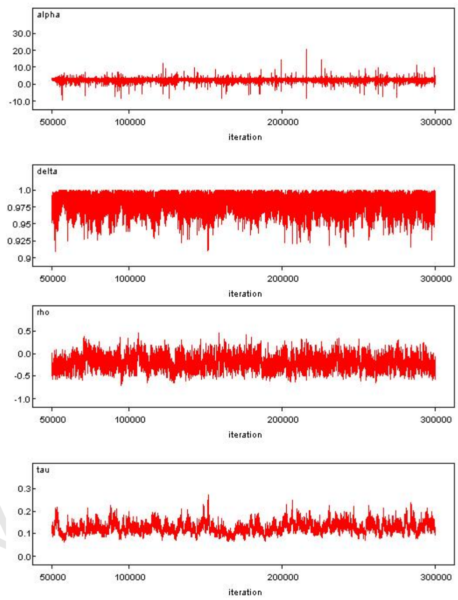

Figure B.1: Trace of parameter estimates for the series OILGAS

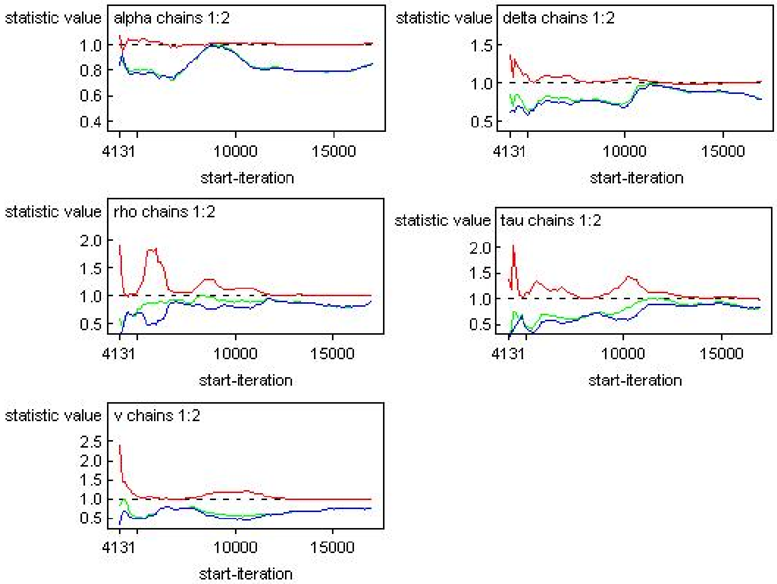

Figure B.2: The Gelman-Rubin statistic for the series OILGAS

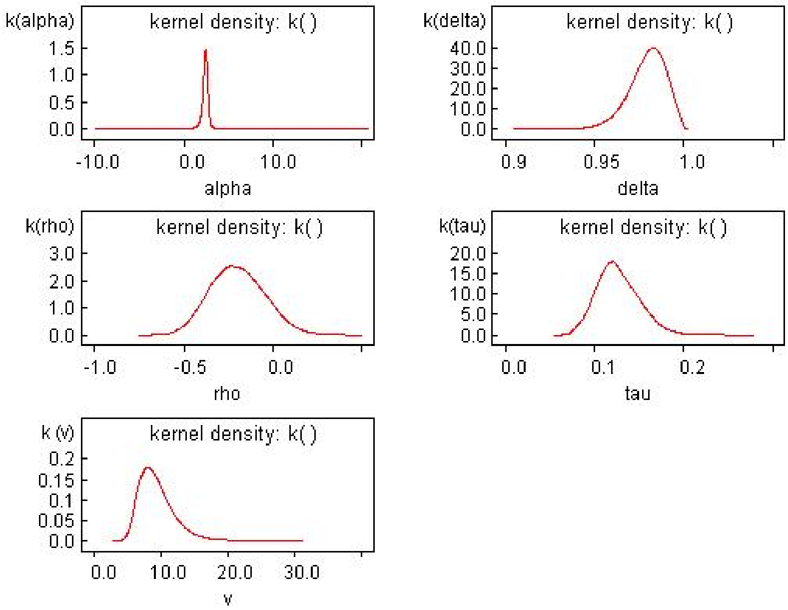

Figure B.3: Kernel densities of parameter for the series OILGAS

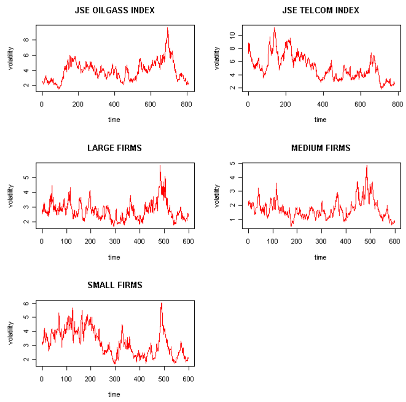

Figure B.4: Smoothed Volatility for weekly series

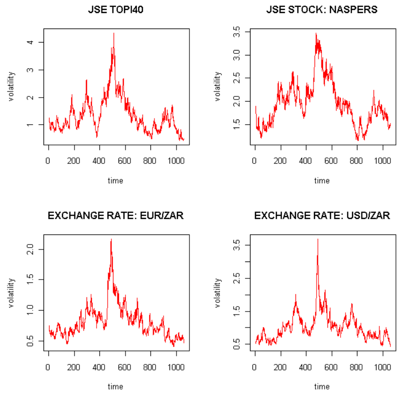

Figure B.5: Smoothed Volatility for daily series

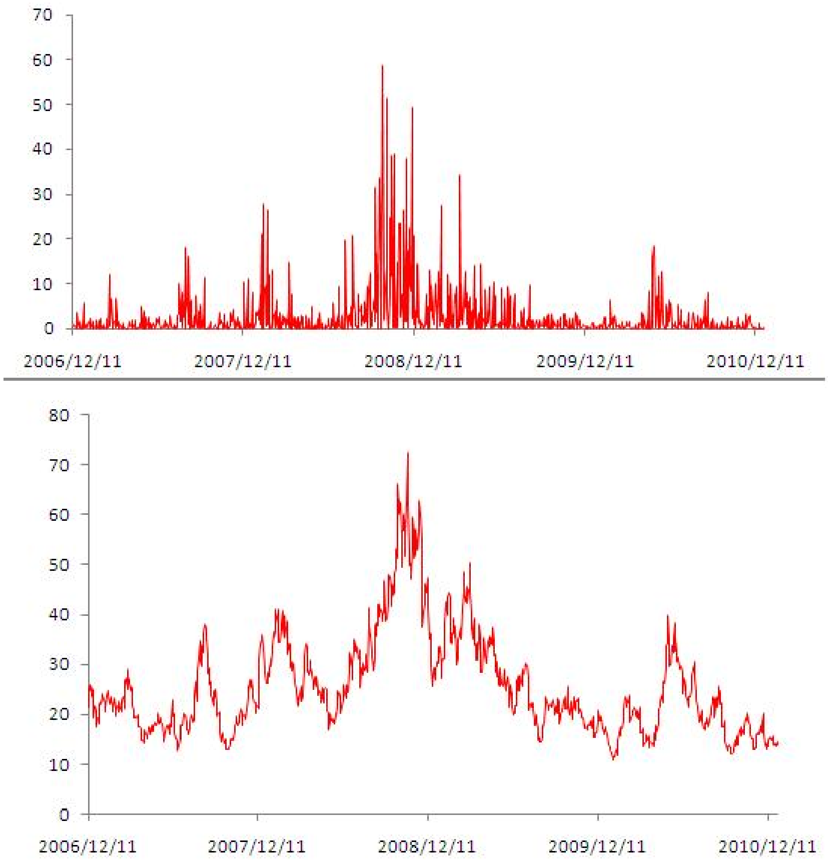

Figure B.6: ALSI squared returns and anualised smoothed volatility

Appendix C

BUGS codes

C.1 Basic model

```
 model
{
  #model
  for(i in 1:n)
  {
     rsig2[i]<- 1/ h[i]
     r[i] ~ dnorm(0,rsig2[i])
  }
   isig2<- itau2 * (1-pow(delta,2))
   logh[1] ~ dnorm(alpha, itau2)
   h[1] <- exp(logh[1])
   for(i in 2:n)
   {
     meanlogh[i] < −alpha + delta*(logh[i-1]-alpha)
     logh[i] ~ dnorm(meanlogh[i], itau2)
     h[i] <- exp(logh[i])
   }
   #prior distributions
   alpha ~ dnorm(0,0.1)
   deltastar ~ dbeta(20,1.5)
   itau2 ~ dgamma(2.5,0.025)
   tau<- sqrt(1/itau2)
   delta<- 2*deltastar-1
}
```

C.2 Full model

```
 model
{
  #model
   isig2[1]<- itau2 * ilambda[1] *(1-pow(delta,2))
   logh[1] ~ dnorm(alpha, isig2[1])
   h[1]<- exp(logh[1])
   ilambda[1] ~ dgamma(v0, v0)
   for(i in 2:n)
   {
     isig2[i]<- itau2 * ilambda[i]
     meanlogh[i]<- alpha + delta * (logh[i-1] - alpha)
     logh[i] ~ dnorm(meanlogh[i], isig2[i])
     h[i]<- exp(logh[i])
     ilambda[i] ~ dgamma(v0, v0)
   }

   risig2[1]<-ilambda[1]/(h[1]*(1-pow(rho,2)))
   rmean[1]<-rho/ tau*sqrt(h[1])*(1-pow(delta,2)) * (logh[1]-alpha)
   r[1] ~ dnorm(rmean[1], risig2[1])

   for(i in 2:n)
   {
     risig2[i]<-ilambda[i]/(h[i]*(1- pow(rho, 2)))
     rmean[i]<- rho/ tau * sqrt(h[i])*(logh[i+1] - alpha- delta*(logh[i] - alpha))
     r[i] ~ dnorm(rmean[i], risig2[i])
   }

  #prior distributions
   alpha ~ dnorm(0,0.01)
   deltastar ~ dbeta(20,1.5)
   delta<- 2 * deltastar - 1
   itau2 ~ dgamma(2.5,0.025)
   rho ~ dunif(-1, 1)
   v ~ dchisqr(4)
   v0<- 0.5 * v
}
```

Appendix D

Bayes Factors R-code

The following piece of R-code implement the Bayes Factors of JPR(2004). To implement the algorithms, we require an MCMC output. We made use of the package R2WinBUGS which has a function bugs(). The function provokes WinBUGS which in turn runs a model passed to it as an argument of the function bugs(). When iterations are completed, parameters of interest are extracted and stored in R for the calculations of Bayes Factors.

D.1 Basic model vs Fat-tail model

```
 # ---------------------------Normal pdf--------------------------
normalpdf = function(x,mu,sig2){
  pdfV = 1/sqrt(2*pi*sig2)*exp(-0.5*(x-mu) ^ 2/sig2)
  return(pdf)
}
# ---------------Bayes Factors:  Basic vs Fat-tail------------------------
N.iter = n.iter-n.burnin
BF1 = rep(0,N.iter)
p = rep(0,n)
pB = rep(0,n*N.iter)
pB=matrix(pB,N.iter,n)
pF = rep(0,n*N.iter)
pF=matrix(pF,N.iter,n)
for (i in 1:N.iter){
   for (t in 1:n){
       pB[i,t] = normalpdf(r[t],0,h[i,t])
       pF[i,t] = normalpdf(r[t],0,h1[i,t]*lambda[i,t])
       p[t] = pB[i,t]/pF[i,t]
   }
  BF1[i]= prod(p)
}
BF = sum(BF1)/N.iter
BF
```

D.2 Fat-tail model vs Full model

```
 # ---------------Bayes Factors:  Fat-tail vs Full model------------------------
N.iter = n.iter-n.burnin
z = rep(0,n)
u = rep(0,n)
uz = rep(0,n)
a11 = rep(0,N.iter)
a12 = rep(0,N.iter)
a22 = rep(0,N.iter)
a22.1 = rep(0,N.iter)
psitilde = rep(0,N.iter)
BF2 = rep(0,N.iter)

for (i in 1:N.iter){
   z[1] = r[1]/sqrt(lambda[i,1]*h2[i,1])
   u[1] = log(h2[i,1])-alpha[i]
   uz[1]= u[1]*z[1]
   for (t in 2:n){
       z[t] = r[t]/sqrt(lambda[i,t]*h2[i,t])
       u[t] = log(h2[i,t])-alpha[i]-delta[i]*(log(h2[i,t-1])-alpha[i])
       uz[t]= u[1]*z[t]
  }
  a11[i] = sum(z ^ 2)
  a12[i] = sum(uz)
  a22[i] = sum(u ^ 2)
  a22.1[i]= a22[i]-a12[i] ^ 2/a11[i]
  psitilde[i]= a12[i]/(a11[i]+2)
  BF2[i]= sqrt((1 + 0.5*a11[i])/(1+ a22.1[i]/0.005))*
          (1+psitilde[i] ^ 2/(0.5*0.005)) ^ (-0.5*(1+n))
}
BF = ratiogamma(1,n)*sum(BF2)/N.iter
BF
```

Bibliography

[1] Torben G. Andersen. Return Volatility and Trading Volume: An information flow interpretation of stochastic volatility. *The Journal of Finance*, 51(1), 1996.

[2] Bradley P. Carlin and Siddhartha Chib. Bayesian Model Choice via Markov Chain Monte Carlo Methods. *Jounal of the Royal Statistical Society. Series B (Methodological)*, 57(3):473–484, 1995.

[3] Siddhartha Chib and Edward Greenberg. Understanding the Metropolis-Hasting Algorithm. *Jounal of the American Statistics Association*, 49(4):327–335, 1995.

[4] Peter K. Clark. A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. *Econometrica*, 41(1):135–155, 1973.

[5] Annette J. Dobson and Adrian G. Barnett. *An Introduction to Generalised Linear Models*. Taylor and Francis Group, LLC., 2008.

[6] Walter Enders. *Applied Econometric Time Series*. John Wiley and Sons, Inc., 2004.

[7] George S. Fishman. *Monte Carlo: Concepts, Algorithms and Applications*. Springer-Verlag New York, Inc., 1996.

[8] John Geweke. Priors for Macroeconomic Time Series and Their Application. *Econometric Theory*, 10:609–632, 1994.

[9] Eric Ghysels and Joanna Jasiak. Bayesian Analysis of Stochastic Volatility Models. *Business and Economic Statistics*, 12(4):399–401, 1994.

[10] Freedom Gumedze, Theodor Stewart, and Christien Thiart. *Probability Theory and Statistical Inference(STA2004F)*. Lecture notes: Department of Statistical Sciences, University of Cape Town, 2009.

[11] Eric Jacquier, Nicholas G. Polson, and Peter E. Ross. Bayesian Analysis of Stochastic Volatility Models. *Journal of Business American Statistical Association*, 12(4):185–212, 1994.

[12] Eric Jacquier, Nicholas G. Polson, and Peter E. Ross. Bayesian analysis of stochastic models with fat-tails and correlated errors. *Econometrics*, pages 185–212, 2004.

[13] Robert E. Kass and Arian E. Raftery. Bayes Factor. *Jounal of the American Statistics Association*, 90(430):773–795, 1995.

[14] Sangjoon Kim, Niel Shephard, and Siddhartha Chib. Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *The Review of Economics Studies*, 65(3):361–393, 1995.

[15] Andrew B. Lawson, William J. Browne, and Carmen L. Vidal Rodeiro. *Disease Mapping with WinBUGS and MLwiN*. John Wiley and Sons, Ltd, West Sussex., 2003.

[16] Nicholas Metropolis and S. Ulam. The Monte Carlo Method. *Jounal of the American Association*, 44(247):335–341, 1949.

[17] Renate Meyer and Jun Yu. BUGS for Bayesian Analysis of Stochastic Volatility Models. *Econmetrics*, 3:198–215, 2000.

[18] Ioanis Ntzoufras. *Bayesian Modeling Using WinBUGS*. John Wiley and Sons, Inc., Hoboken, New Jersey., 2009.

[19] Yasuhiro Omori and Tashiaki Watanabe. Block Sampler and Posterior Mode Estimation for Asymmetric Stochastic Volatility Models. *Computational Statistics and Data Analysis*, 56, 2008.

[20] Svetlozar T. Rachev, John S. J. Hsu, Biliana S Bagasheva, and Frank J. Fabozzi. *Bayesian Methods in Finance*. John Wiley and Sons, Inc, New Jersey., 2008.

[21] S J. Taylor. *Asset Price Dynamics, Volatility, and Predictability*. Princeton University Press, Princeton., 2005.

[22] Luke Tierney. Markov Chains for Exploring posterior distributions. *The Annals of Statistics*, 22(4):1701–1728, 1994.

[23] Ruey S. Tsay. *Analysis of Financial Time series*. John Wiley and Sons, Inc, Hoboken, New Jersey., 2005.

[24] Joanna J. J. Wang, Jennifer S. K. Chan, and S. T. Choy. Stochastic Volatility Models with Leverage and Heavy-tailed Distributions: A Bayesian approach using scale mixtures. *Computational Statistics and Data Analysis*, 55:852–862, 2010.