Estimating Stochastic Volatility Models with Student-t Distributed Errors

Vishal Rama

A dissertation submitted to the Faculty of Science, University of Cape Town, in partial fulfilment of the requirements for the degree of Master of Science.
December 19, 2019

MSc in Decision Sciences and Analytics,
University of Cape Town.

Declaration

I declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Science in the University of Cape Town. It has not been submitted before for any degree or examination at any other University.

Signed by candidate

December 19, 2019

Abstract

This dissertation aims to extend on the idea of Bollerslev (1987), estimating GARCH models with Student-t distributed errors, to estimating *Stochastic Volatility* (SV) models with Student-t distributed errors. It is unclear whether Gaussian distributed errors sufficiently account for the observed leptokurtosis in financial time series and hence the extension to examine Student-t distributed errors for these models. The quasi-maximum likelihood estimation approach introduced by Harvey (1989) and the conventional Kalman filter technique are described so that the SV model with Gaussian distributed errors and SV model with Student-t distributed errors can be estimated. Estimation of GARCH(1,1) models is also described using the method maximum likelihood. The empirical study estimated four models using data on four different share return series and one index return, namely : Anglo American, BHP, FirstRand, Standard Bank Group and JSE Top 40 index. The GARCH and SV model with Student-t distributed errors both perform best on the series examined in this dissertation. The metric used to determine the best performing model was the Akaike information criterion (AIC).

Acknowledgements

I would like to thank my supervisors, Doctor Maria Kulikova and Melusi Mavuso, for their guidance and patience with me throughout my dissertation. To my parents, I am forever indebted to you both for your support and for giving me all the opportunities that both of you never had.

I am truly grateful.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Volatility is a central concept in finance, as it is regarded as a measurement of risk. It plays an important role in many financial applications such as derivative pricing and hedging, risk management and portfolio management. In derivative pricing, the canonical equity option pricing model of Black and Scholes (Black and Scholes (1973)) assumes constant stock volatility throughout the life of the option. Numerous studies, however, have indicated that many financial time series exhibit changes in volatility over time. Stylized facts, as defined by Cont (2001), are a set of properties common across many instruments, markets and time periods, which has been observed by independent studies. They suggest that volatility changes with time. See, Cont (2001), for more details about stylized facts.

Some of these stylized facts include; volatility clustering, the leverage effect, absence of autocorrelation, heavy-tails, aggregational Gaussianity, intermittency, conditional heavy-tails, slow decay of autocorrelation in absolute returns, volume or volatility correlation. Volatility clustering is when large changes (in absolute value) in the price of an asset are followed by other large changes and small changes are followed by other small changes. The leverage effect is the negative correlation of the volatility of an asset with the asset returns. Various models have been developed to capture these effects.

The approach based on autoregressive conditional heteroskedasticity (ARCH), introduced by Engle (1982) and later generalised to generalised autoregressive conditional heteroskedasticity (GARCH) by Bollerslev (1986), was the first attempt to account for stochastic volatility. These models capture such effects, by allowing the conditional variance to be a function of the squares of previous observations and past variances.

An alternative approach is to use Stochastic Volatility (SV) models. The shift from the GARCH model was due to these SV models generalising to multivariate series more naturally and with SV models having desirable statistical attributes. In an SV model, the logarithm of volatility is modelled as an autoregressive process of the first order, separate from the original time series.

Variants of the method of moments and estimation techniques based on auxiliary models (Broto and Ruiz (2004)) have been used in past literature to estimate SV models. Maximum likelihood (ML) estimators have also been used in the estimation of SV models. A prominent ML estimator, which is often used, is the Expectation-Maximization (EM) algorithm. See, for example, Elliott and Hyndman (2007), Shumway and Stoffer (1982),Koopman (1993) and many others for application to econometric models. The EM estimator is different compared to some other ML estimators since it does not compute the ML parameter estimates by direct maximization of the likelihood function (LF). It is an iterative algorithm for maximizing the likelihood, with each iteration consisting of two steps: Expectation (E-step) and Maximization (M-step). For the models under examination in this dissertation, EM estimation strategies have been well designed and can be found in Némesin and Derrode (2012). An attractive feature of the EM algorithm is its derivative-free nature. What this means is that no LF gradient(score) or curvature matrices need to be evaluated Dempster et al. (1977). The disadvantage of the EM algorithm, however, is its slow convergence rate, its sensitivity to initial values Mader et al. (2014) and difficulties in partial learning, where partial learning is the strategy of improving parameter estimation efficiency through reducing the dimension of the search space Némesin and Derrode (2014). There are alternative gradient-based approaches which will give higher convergence rates, however, these require the score evaluation Gupta and Mehra (1974). The log LF gradient (score) computation for the examined state-space models can be carried out by using either only the filtering equations for the LF and derivative computation [see, for example, Zadrozny (1989) and Klein and Neudecker (2000)], or by implementing a combined filter and smooth approach [see, for example, Segal and Weinstein (1988),Koopman and Shephard (1992) and Wills and Ninness (2008)].

Following Harvey et al. (1994), we utilize a Quasi-maximum likelihood (QML) approach for estimating SV models, by using the Kalman filter equation. In their paper, Harvey et al. (1994) proposed a QML estimation approach based on the Kalman filter to estimate SV models, which assumes that the models are Gaussian. The technique initially had an appeal due to its simplicity. The use of the technique, however, diminished when it was realised that SV models are not Gaussian. The problem found is that using Gaussian QML estimation produces highly inefficient estimators. Harvey et al. (1994) indicates that SV models can be generalised to have errors which are Student-t distributed. Harvey et al. (1994) stress that this is important because the kurtosis exhibited in many financial series is higher than the kurtosis accounted for by a Gaussian process.

Bollerslev (1987) modelled returns using a GARCH model with conditional Student-t distributed errors. The reason for doing this was because for the ARCH/GARCH model, the conditional error distribution, where conditioning is on the past data, is normal but the unconditional error distribution of the ARCH model is leptokurtic. It was unclear whether the Gaussian GARCH model sufficiently captured the observed leptokurtic error distribution in financial time series. It was suggested that a heavy-tailed distribution might be better at accounting for the observed leptokurtosis and as a result, a model with conditional Student-t distributed errors was investigated. What was found is that the GARCH model with conditionally Student-t distributed errors was superior to the Gaussian GARCH model. It is noted, however:

> ”It remains an open question whether other conditional error distributions provide an even better description.” (Bollerslev, 1987, p.546)

This statement forms the basis of this dissertation. This dissertation looks to extend on the idea of Bollerslev (1987) using the GARCH model with conditional Student-t distributed errors, to that of exploring the SV model with conditionally Student-t distributed errors. The reason for that is because South African stock price data has been empirically shown to exhibit high kurtosis and because SV models are a more general class of models compared to the GARCH model.

The purpose of the study is to test whether Student-t distributed errors fit better than Gaussian distributed errors. This will be done by comparing the Gaussian GARCH model against the GARCH model with Student-t distributed errors and the SV model with Gaussian errors against the SV model with Student-t distributed errors. The best performing between the GARCH models and the SV models will then be compared against each other. These models will be applied to data from the South African financial market, specifically, the JSE Top 40 index and four different share return series.

The dissertation is organized as follows. Chapter 2 describes the Gaussian GARCH model and the GARCH model with Student-t distributed errors, along with how these models are estimated. Chapter 3 presents the conventional Kalman Filter and the QML method used to estimate SV models. Chapter 4 presents and analyses an empirical application using the JSE Top 40 index, along with four different share return series from the index. Finally, Chapter 5 concludes the dissertation.

Chapter 2

Modelling Stochastic Volatility as a GARCH Process

Two prominent classes of volatility models have been developed. These are GARCH and SV models. In this chapter, we discuss GARCH models. GARCH models have been used extensively in econometrics and finance to model volatility of time series data. Many extensions of GARCH models have been proposed in the literature such as; the exponential GARCH (EGARCH), the power GARCH (PGARCH), the threshold GARCH (TGARCH) to name a few. We will, however, only be looking at Gaussian GARCH and GARCH with Student-t distributed errors in this dissertation. We begin by defining a Gaussian GARCH(1,1) model.

2.1 Gaussian GARCH(1,1) Model

The GARCH model which was developed by Bollerslev (1986) is a generalisation of the ARCH model introduced by Engle (1982). The GARCH(1,1) model with normally distributed errors is given by

$$y_t = \mu + \epsilon_t, \epsilon_t = \sigma_t z_t, \tag{2.1}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{2.2}$$

$$\epsilon_t | \psi_{t-1} \sim \mathcal{N}(0, \sigma_t^2) \tag{2.3}$$

where y_t is the returns series at time t (i.e. $y_t = \ln(\frac{S_t}{S_{t-1}})$), S_t is the closing price on day t, z_t is a discrete white noise process and σ_t^2 the conditional variance of ϵ_t at time t given ψ_{t-1}, the information set at time $t-1$ (i.e. $\psi_{t-1} = \{\epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_0\}$). Stationarity is ensured for the GARCH(1,1) process in equation 2.2, if, α_0 is strictly positive, $\alpha_1 \geq 0$, $\beta_1 \geq 0$ and $\alpha_1 + \beta_1 < 1$ Bollerslev (1986).

2.2 Properties of GARCH Models

By simply looking at the GARCH(1,1) model, the advantages and disadvantages of GARCH models can be seen. Looking at equation 2.2 above, a large ϵ_{t-1}^2 or σ_{t-1}^2, results in σ_t^2 being relatively large. What this means is that a large value of ϵ_{t-1}^2 tends to be followed by another large value of ϵ_t^2. This captures volatility clustering in financial time series.

It can be shown that if $1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 > 0$ (Bollerslev (1986)), then

$$\frac{E(\epsilon_t^4)}{[E(\epsilon_t^2)]^2} = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2} > 3. \tag{2.4}$$

What this indicates is that ϵ_t has positive excess kurtosis; therefore, the GARCH(1,1) process has a heavier tail distribution than that of a normal distribution. This agrees with the stylized fact that “outliers” are more often observed in asset returns than that which are produced from a normal random process.

Some of the weaknesses of GARCH models are as follows (Tsay (2005))

- GARCH models do not capture the leverage effect. They assume that both positive and negative shocks have the same effect on volatility. The reason for this is that the volatility in the model depends on the square of the previous shocks.

- A GARCH model is an uninformative model as it does not provide any insight into what the source of the variation of financial time series might be. It simply provides a mechanical way to describe the behaviour of the conditional variance, as it does not give an indication as to what caused such behaviour to occur.

- GARCH models respond slowly to large isolated shocks and as a result, tend to overpredict volatility.

We note that there are GARCH models that are designed to address some of these weaknesses mentioned above, such as the exponential GARCH (EGARCH) model. The EGARCH model aims to address the leverage effect. This, however, will not be investigated in this dissertation and for more information on the EGARCH model, refer to Nelson (1991).

2.3 GARCH(1,1) Model with Student-t Distributed Errors

The GARCH(1,1) model with Student-t distributed errors described by Bollerslev (1987) is given by

$$y_t = \mu + \epsilon_t, \epsilon_t = \sigma_t z_t, \tag{2.5}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{2.6}$$

$$z_t|\psi_{t-1} \sim \boldsymbol{std.} f_\nu(z_t|\psi_{t-1}), \nu > 2 \tag{2.7}$$

where the parameters are similarly defined for the Gaussian GARCH(1,1) model, except for $z_t|\psi_{t-1}$, which is now Student-t distributed with ν degrees of freedom. The term ν is known as the normality parameter. The reason for this is that as $\nu \to \infty$, the Student-t distribution converges to a normal distribution.

2.4 Properties of the Student-t Distribution

Let t_ν denote a Student-t distribution with ν degrees of freedom. The r-th moment of a Student-t distribution exists if and only if $r < \nu$;

$$E(X^r) = \begin{cases} \Gamma(\frac{r+1}{2})\Gamma(\frac{\nu-r}{2})\nu^{\frac{r}{2}}/\sqrt{\pi}\Gamma(\frac{\nu}{2}), & \text{if } r \text{ is even.} \\ 0, & \text{if } r \text{ is odd.} \end{cases} \tag{2.8}$$

The variance of a t_ν is finite and equals

$$Var(t_\nu) = \frac{\nu}{(\nu - 2)} \tag{2.9}$$

if $\nu > 2$. If $0 < \nu \leq 1$, then the expected value of the Student-t distribution does not exist and as a result, the variance is not defined. If $1 < \nu \leq 2$, then the expected value is 0 and the variance is infinite. The skewness of a Student-t distribution is 0 if $\nu > 3$. When $\nu \leq 3$, the skewness is undefined. The kurtosis of a Student-t distribution is finite and is given by

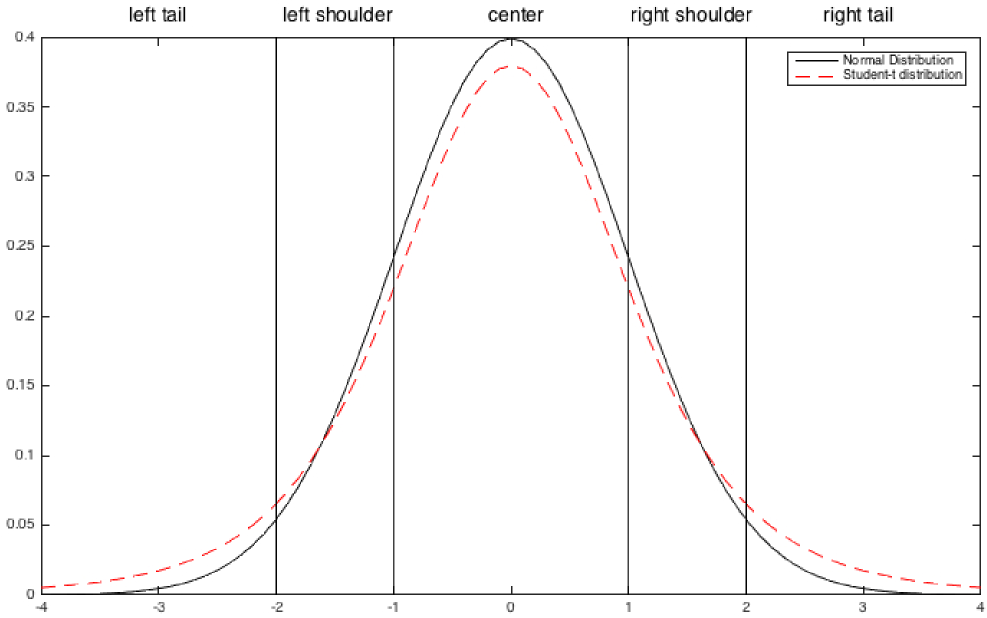

Figure 2.1: *Comparison of a Standard Normal distribution and a Student-t distribution with 5 degrees of freedom. The plot also shows the center, shoulders and tail regions.*

$$Kur(t_\nu) = 3 + \frac{6}{(\nu - 4)} \tag{2.10}$$

if $\nu > 4$. If $\nu \leq 4$, the kurtosis of Student-t distribution is undefined.

From Figure 2.1 it can be seen that a Student-t distribution with 5 degrees of freedom has heavier tails than a normal distribution. What this means is that the probability of obtaining values very far from the mean is larger than in the normal distribution. Figure 2.2, graphically shows how an increase

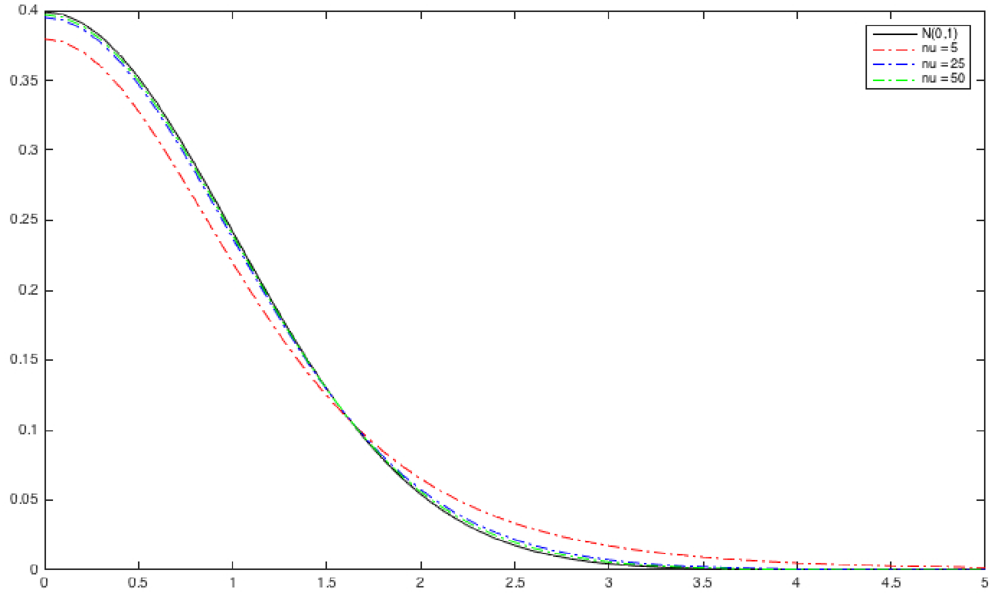

Figure 2.2: *Comparison of the tails of a Standard Normal distribution and Student-distribution with 5,25 and 50 degrees of freedom, respectively.*

in the degrees of freedom, leads the Student-t distribution to converge to a normal distribution. For further properties, see Nilmini Kumari (2013).

2.5 Estimation of the GARCH Model

Usually, the parameters of GARCH models are estimated by using the method of maximum likelihood. If the density function of z_t, where z_t is discrete white noise as defined previously, is given by $f(z_t)$ and the sample data given by $y_T, y_{T-1}, \ldots, y_1$, then the sample log LF in general form is given by

$$\mathcal{L}(\theta; y_T, \ldots, y_1) = \sum_{t=1}^{T} [\ln(f(\epsilon_t \sigma_t^{-1})) - \ln(\sigma_t)] \tag{2.11}$$

where θ is the unknown parameter vector to be estimated and $\epsilon_t = \sigma_t z_t$.

As an example, if ϵ_t is zero-mean Gaussian distributed conditional on past history ψ_{t-1} , that is

$$p(\epsilon_t|\psi_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{\epsilon_t^2}{2\sigma_t^2}} \tag{2.12}$$

the log LF of the GARCH model is

$$\mathcal{L}(\theta; y_T, \ldots, y_1) = \sum_{t=1}^{T} l_t = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln(\sigma_t^2) - \frac{1}{2}\sum_{t=1}^{T}\frac{\epsilon_t^2}{\sigma_t^2}. \tag{2.13}$$

For a GARCH(1,1) model, substituting $\sigma_t^2 = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2$ into equation 2.13, the log LF becomes only a function of ϵ_t and the parameters needing to be estimated.

Chapter 3

Modelling Stochastic Volatility as an Autoregressive Process

As mentioned previously, SV models are an alternative approach to model volatility compared to the GARCH model. For the GARCH model, the conditional volatility is a deterministic function of past quantities, whereas with SV models the conditional volatility is itself a random process. It is difficult to obtain the exact LF for SV models since we can not observe the conditional variance. As a consequence, QML estimation can be used to obtain approximate results. We begin by defining the general case of SV models.

3.1 The SV Model

A simple stationary SV model is given by Ruiz (1994) as

$$y_t = \varepsilon_t \exp(h_t/2), \tag{3.1}$$

$$h_t = \gamma + \phi h_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \sigma_\eta^2) \tag{3.2}$$

where $\{y_t\}_{t=1}^T$ are the returns, $h_t = \ln(\sigma_t^2)$, σ_t^2 is the variance of y_t and ε_t is a white noise process with unit variance, generated independently of η_t. Logarithms are used to ensure that σ_t^2 is always positive. Andersen et al. (2001, 2003) show the assumption of Gaussianity of η_t is well justified since the log-volatility process can be well approximated by a normal distribution. The uncertainty attached to future volatility is measured by the variance of the log-volatility process, σ_η^2. As σ_η^2 and ϕ get close to zero and one, respectively, we see that the evolution of volatility is smooth over time. If we, however, look at the case when σ_η^2 and ϕ are zero and one, respectively in the limit, volatility is constant over time. As a result, the returns are homoscedastic. The parameter, ϕ, can be seen as a measure of the persistence of shocks to the volatility. If $|\phi| < 1$, this ensures h_t is strictly stationary. If ε_t and η_t are allowed to be correlated with each other, the model can pick up the kind of asymmetric behaviour which is often found in stock prices Ghysels et al. (1996). When there is negative correlation between ε_t and η_t this induces a leverage effect.

In the literature the errors ε_t are generally assumed to have a Gaussian distribution, however, several authors have considered heavy-tailed distributions. See, for example, Chib et al. (2002) and Jacquier et al. (2004) among many others. We note that even when ε_t is assumed to be Gaussian, y_t conditional on past observations up to time $t-1$, is not Gaussian distributed.

The simple SV model can be linearised by squaring the observations in (3.1) and taking the logarithms which gives

$$\ln(y_t^2) = \mathbf{E}\{\ln(\varepsilon_t^2)\} + h_t + \zeta_t \ , \tag{3.3}$$

$$h_t = \gamma + \phi h_{t-1} + \eta_{t-1}, \quad \eta_t \sim \mathcal{N}(0, \sigma_\eta^2) \tag{3.4}$$

where ζ_t is non-Gaussian, zero mean white noise, and its statistical properties depend on the distribution of ε_t. The non-Gaussian linear state space model is shown in equations (3.3) and (3.4), where (3.3) is the measurement equation and (3.4) is the transition equation. We will discuss this further later in the dissertation.

3.2 Estimation Methods of SV Models

Estimation methods of SV models can be classified into two general groups. The first of these two estimation methods, being a method which is based directly on the statistical properties on the returns, y_t. Within this group, there are three main classes: (i) estimation based on the method of moments (MM) ; (ii) estimation based on the Maximum Likelihood (ML) principle and (iii) estimation based on an auxiliary model. The second of these groups is an estimation based on the linear model in (3.3) and (3.4). We give a review on some of the models from the first group and then go to the state-space approach and QML technique based on the Kalman filter.

3.2.1 Methods Based on Statistical Properties of y_t

Method of Moments

Estimators based on the Method of Moments (MM) or the Generalised Method of Moments (GMM) have been used since the early stages of SV literature. These procedures are popular, for at least two reasons. Firstly, the moments of financial time series are important due to their agreements with stylized facts of asset returns. Secondly, due to their simplicity, as some may be reluctant to make distributional assumptions and due to the exact LF being difficult to evaluate. The empirical implementation is very simple and as a result, this technique has been used extensively to estimate the parameters of SV models. See Ghysels et al. (1994) , Vetzal (1997), Andersen et al. (1999) for examples of implementation of the procedure.

Within the group of methods based on the statistical properties of y_t, MM is the simplest estimator. Wiggins (1987) and Scott (1987) used a MM approach to estimate the parameters of a stochastic process and found that the parameter estimates were sensitive to the moments which they fitted. They were, however, unsure if this was as a result of sampling error. GMM, using an approach along the lines of Hansen (1982), was then proposed by Melino and Turnbull (1990) to be used to estimate the parameters of SV models. See Hamilton (1994) for a discussion on the GMM properties. Alternatively, a Simulated Method of Moments (SMM) as introduced by Lee and Ingram (1991) and Duffie and Singleton (1990) was proposed, which replaces the analytic moments by moments produced by a simulated process.

It is well known that these procedures are inefficient relative to ML methods and have poor finite sample properties, both in terms of bias and root mean squared error (RMSE) of the estimated parameters. This becomes particularly problematic for estimating stochastic volatility models since the score function, which is used to indicate which moments to be used for MM estimation, cannot be calculated. Jacquier et al. (1994) show that the performance of the techniques worsen when ϕ is close to 1 and for large values of the coefficient of variation, $V(\sigma_t^2)/(E(\sigma_t^2))^2$. Given that financial time series returns tend to show highly persistent volatilities, this tends to suggest that MM methods are not a suitable method to estimate the parameters of SV models. Andersen and Sørensen (1996) reports a large amount of non-converging estimation, particularly for small sample sizes. This is due to the GMM criterion surface being highly irregular, leading to the optimization failing to converge. Andersen and Sørensen (1996) found that the frequent failure of GMM to converge was due to the imprecise estimation of the long-run covariance and consequently imprecise estimation of the GMM-weighted matrix. This issue, however, can be solved by obtaining an accurate approximation of the GMM-weighted matrix.

A drawback of the GMM method is that GMM estimators do not generate estimates of the latent volatility process and hence other techniques such as the Kalman filter, need to be used to get them. See, Ghysels et al. (1994) for example.

Maximum Likelihood Estimation

ML estimators of the parameters of SV models have made strides forward with the development of techniques such as importance sampling and Markov Chain Monte Carlo (MCMC). These numerical methods are needed to evaluate the likelihood of the form of 2.11.Geweke et al. (1994) applied the first importance sampling algorithm to SV models and introduced the idea of applying MCMC to SV models. Danielsson (1994), Shephard and Pitt (1997), Sandmann and Koopman (1998) and Jacquier et al. (1994), to name a few, all followed suit applying MCMC in subsequent research. The main attraction of importance sampling over MCMC algorithm is that it is less computationally demanding and avoids convergence problems Broto and Ruiz (2004). Importance sampling, however, becomes less useful as T, the sample size, becomes large or $dim(h_t)$ increases beyond 1 Shephard (1996). To try to address these issues MCMC algorithms could be used as they are more flexible, allowing large dimensional problems to be split into smaller dimensional tasks.

Jacquier et al. (1994) and Shephard (1993) both, independently, introduced the MCMC algorithm approach to estimating the parameters of SV models. Jacquier et al. (1994) adopts a Bayesian approach in which the specification of the model has a hierarchical structure. On the other hand, Shephard (1993) put their multi-move or block sampler within a simulated EM algorithm. MCMC has subsequently been used extensively. See for example, Steel (1998),So et al. (1998) and Mahieu and Bauer (1998), among many others for use of MCMC. See (Broto and Ruiz, 2004, p. 8-9) for an explanation of the steps of MCMC procedure.

Danielsson and Richard (1993) proposed the simulated maximum likelihood (SML) estimator, as a general method to estimate dynamic latent models. In a subsequent paper, Danielsson (1994) used SML to estimate parameters of SV models. SML works by evaluating the LF with simulation and then the estimates are obtained using a derivative-free optimiser. See Liesenfeld and Jung (2000) and Liesenfeld (1998), for examples of an empirical application of SML to financial returns series.

There are however issues with SML. The likelihood integral is not directly evaluated, and hence the accuracy of the approximation is difficult to measure. This is also limited to only obtaining an exact evaluation of the likelihood, with $\theta = 0$, otherwise, it is not available for the rest of the parameters. Jacquier et al. (1994) proposed a method to try to solve this issue.

Fridman and Harris (1998) propose an alternative maximum likelihood approach for estimating the parameters of SV models. This approach uses a recursive numerical integration procedure that directly calculates the marginal likelihood, which was suggested by Kitagawa (1987) for non-Gaussian state-space model filtering and smoothing procedure. This method can be considered an extended Kalman filter. What makes this approach flexible and simple, is that only conventional integration techniques are used. Fridman and Harris (1998) consider both Gaussian and Student-t errors and through a small simulation experiment, show their approach performs better than GMM and QML, and similar to SML, the MCMC by Jacquier et al. (1994) and the simulated expectation maximization (SEM) by Kim et al. (1998)

An alternative method based on an extended Kalman filter is by Watanabe (1999) who proposes a non-linear-filtering maximum likelihood, which is seen as an extension of QML. The procedure yields the exact likelihood, which is obtained by solving a series of integrals using piecewise linear approximation, which is then maximised to obtain the parameter estimates of the SV model. A smoothing algorithm for volatility is also proposed. Watanabe (1999) conduct a comparative study and find that the efficiency of the method is in line with the MCMC approach of Jacquier et al. (1994) and SML. The smoothing algorithm is also shown to be superior to the standard smoothing solution used in QML.

The main disadvantage of both extended Kalman filter approaches proposed by Fridman and Harris (1998) and Watanabe (1999) is that they have slow computational convergence. According to Sandmann and Koopman (1998) the procedures require a priori a fixed grid over which the volatility process will be integrated. In some instances, the optimal grid may not exist and hence there is a trade-off between numerical accuracy and computational efficiency with these approaches.

Estimation Based on an Auxiliary Model

To implement MM, GMM and QML are fairly simple, however, estimation based on an auxiliary model can be far more complex to implement. Estimation based on an auxiliary model are more elaborate and computationally intensive procedures, however, may be justified as they show efficiency gains. We propose methods which have auxiliary models that are easy to estimate. We do this because to simulate SV models are fairly simple, however, to estimate them can be complex. There are two main methods which are proposed within this group: Indirect Inference and Efficient Methods of Moments (EMM).

The Indirect Inference estimator proposed by Gourieroux et al. (1993) is given as follows

$$\tilde{\theta}_T^H = \arg min_{\theta\in\Theta}(\hat{\beta}_T - \tilde{\beta}_{TH}(\theta))'\hat{\Omega}_T(\hat{\beta}_T - \tilde{\beta}_{TH}(\theta)) \tag{3.5}$$

where $\hat{\beta}_T$ is obtained by maximizing an auxiliary criterion from the auxiliary model $Q_T(\beta, y_t)$, $\tilde{\beta}_{TH}(\theta)$ is an estimate of the binding function obtained by maximizing $Q_{TH}(\beta, y_{TH}(\theta))$ and $y_{TH}(\theta) = (y_1, \ldots, y_{tH})$ is a vector of simulated observations from the SV model. There are several models which have been proposed as auxiliary models. For example as auxiliary models, Gourieroux et al. (1993) proposed the use of the quasi-likelihood function of Harvey et al. (1994), Engle and Lee (1996) and Lombardi and Calzolari (2009) use GARCH models, *EmmPack 1.01: C/C++ code for use with O* (n.d.) use EGARCH models, Fiorentini et al. (2002) use a non-linear asymmetric GARCH model and Monfardini (1998) uses AR(p) and ARMA(1,1) models.

The EMM approach developed by Gallant and Tauchen (1996) and Bansal et al. (1994, 1995), and which was then expanded on by Gallant et al. (1997), Gallant and Tauchen (1999) and Andersen and Lund (1997), is developed based off GMM. It looks to address some of the problems of GMM. EMM, is similar to Indirect Inference, but fits by using scores obtained from a criterion function and improves upon the Indirect Inference approach in computational terms. Gallant et al. (1997) and Jiang and van der Sluis (1998, 1999), among others, implement the EMM procedure to estimate discrete time univariate and multivariate SV models.

Gourieroux et al. (1993) show that the Indirect Inference and EMM estimators are asymptotically equivalent. The Indirect Inference estimator is, however, more computationally intensive since the QML estimates of the score generator needs to be computed continuously. The EMM estimator, however, only need to be computed once. Both estimators are consistent and asymptotically normal.

3.2.2 State-Space Representation and the Kalman Filter

We now describe the state-space model, to which the Kalman Filter will be applied to obtain optimal estimates. The state-space form allows for the possibility of formulating models that are much wider and richer than which are normally considered. These models are also structured in a way that leads to a more natural interpretation and provides more useful information on the nature of the underlying economic processes Harvey (1987).

Consider the state-space model representation given by the discrete-time ($t = 1, \ldots, T$) linear stochastic system of equations; see the problem statement in Grewal and Andrews (2015)

$$x_t = Fx_{t-1} + Bu_{t-1} + Gw_{t-1} \qquad w_t \sim \mathcal{N}(0, Q), \tag{3.6}$$

$$y_t = Hx_t + v_t \qquad v_t \sim \mathcal{N}(0, R) \tag{3.7}$$

where x_t, the hidden random sequence in time-varying variance models (i.e. in GARCH and SV models), is the variance of a series of returns on an asset denoted by y_t. In our application, y_t will be the log returns. The vector u_t refers to the explanatory variables, which is also interpreted as a control input in engineering applications Wilcox and Hamano (2017). For instance, the classical GARCH(p,q) description in Bollerslev (1986) incorporates a term $u_t^\intercal \beta$ with the vector β standing for the regression coefficients involved. Usually, these are unknown and need to be estimated in parallel with x_t, where in the case of SV models given by 3.1, 3.2, the unknown dynamic state to be estimated is the log of the variance, i.e. $x_t := h_t = \ln(\sigma_t^2)$. See (Koopman and Hol Uspensky, 2002, p. 3) where the econometric application, as above, is also presented.

Equations (3.6) and (3.7) are called the *state* and *observation* equations, respectively. The measurement noise $\{v_t\}$ and the state process noise $\{w_t\}$ are independent zero-mean white-noise processes with covariance matrices Q (non-negative definite) and R (positive definite), respectively. They are also uncorrelated with initial state $x_0 \sim \mathcal{N}(\bar{x}_0, \Pi_0)$, Π_0 is non-negative definite.

Given the input and output measurements

$$\mathcal{Y}_0^T := \{y_0, y_1, u_1, y_2, \ldots, y_T, u_T\}, \tag{3.8}$$

the objective is to obtain an estimate $\hat{x}_{t|t}$ of x_t. If $t = T$, this is known as a filtering problem[1], which is what will be looked at. Several estimators $\hat{x}_{k|k}$ of the state sequence $\{x_k\}$, which are all functions of $\mathcal{Y}_0^T$, can be considered and are as follows

$$\mathbf{E}(x_t|\mathcal{Y}_0^T), \qquad \text{conditional mean} \tag{3.9}$$

$$\arg min_{\hat{x}_{t|t}} \mathbf{E}(\|x_t - \hat{x}_{t|t}\|^2), \quad \text{minimum expected mean square error (MSE)} \tag{3.10}$$

$$\arg min_{\hat{x}_{t|t} \in span(\mathcal{Y}_0^T)} \mathbf{E}(\|x_t - \hat{x}_{t|t}\|^2). \qquad \text{minimum linear expected MSE} \tag{3.11}$$

Computing the conditional mean 3.9 can be difficult when using non-Gaussian distributions. It can be shown that the estimators 3.9 and 3.10 are equal, due to the conditional mean representing the minimum variance estimate. The proof is as follows

We show that given information $\mathcal{Y}_0^T$, the best mean-square estimator of a random variable x_t is the conditional expectation $\mathbf{E}(x_t|\mathcal{Y}_0^T)$. Let Y be an arbitrary random variable that depends on $\mathcal{Y}_0^T$, we have to show that

$$\mathbf{E}((x_t - \mathbf{E}(x_t|\mathcal{Y}_0^T))^2) \leq \mathbf{E}((x_t - Y)^2).$$

---

[1] If $T > t$ this is called a smoothing problem, and if $T < t$ it is called a prediction problem.

To that end, we calculate $\mathbf{E}((x_t - Y)^2)$ as

$$\mathbf{E}((x_t - Y)^2) = \mathbf{E}((x_t - \mathbf{E}(x_t|\mathcal{Y}_0^T) + \mathbf{E}(x_t|\mathcal{Y}_0^T) - Y)^2)$$

$$= \mathbf{E}((x_t - \mathbf{E}(x_t|\mathcal{Y}_0^T))^2 + \mathbf{E}((\mathbf{E}(x_t|\mathcal{Y}_0^T) - Y)^2) + 2\mathbf{E}((x_t - \mathbf{E}(x_t|\mathcal{Y}_0^T))(Y - \mathbf{E}(x_t|\mathcal{Y}_0^T)))$$

$$= \mathbf{E}((x_t - \mathbf{E}(x_t|\mathcal{Y}_0^T))^2 + \mathbf{E}((\mathbf{E}(x_t|\mathcal{Y}_0^T) - Y)^2) + 2\mathbf{E}(\mathbf{E}((x_t - \mathbf{E}(x_t|\mathcal{Y}_0^T))(Y - \mathbf{E}(x_t|\mathcal{Y}_0^T))|\mathcal{Y}_0^T))$$

$$= \mathbf{E}((x_t - \mathbf{E}(x_t|\mathcal{Y}_0^T))^2 + \mathbf{E}((\mathbf{E}(x_t|\mathcal{Y}_0^T) - Y)^2) + 2\mathbf{E}(\mathbf{E}((x_t - \mathbf{E}(x_t|\mathcal{Y}_0^T))(Y - \mathbf{E}(x_t|\mathcal{Y}_0^T))|\mathcal{Y}_0^T))$$

$$= \mathbf{E}((x_t - \mathbf{E}(x_t|\mathcal{Y}_0^T))^2 + \mathbf{E}((\mathbf{E}(x_t|\mathcal{Y}_0^T) - Y)^2) + 2\mathbf{E}(\mathbf{E}((Y - \mathbf{E}(x_t|\mathcal{Y}_0^T))|\mathcal{Y}_0^T)(x_t - \mathbf{E}(x_t|\mathcal{Y}_0^T)))$$

$$= \mathbf{E}((x_t - \mathbf{E}(x_t|\mathcal{Y}_0^T))^2 + \mathbf{E}((\mathbf{E}(x_t|\mathcal{Y}_0^T) - Y)^2) + 0 \geq \mathbf{E}((x_t - \mathbf{E}(x_t|\mathcal{Y}_0^T))^2).$$

This allows us to use the minimum expected MSE 3.10, which is relatively simpler to calculate compared to the conditional mean. We typically only have estimates of the first and second moments of the signal and the observation. This is not enough information, in general, to calculate the minimum MSE. The minimum linear expected MSE, however, can be commuted from this information. In addition, when w_k,v_k and the initial state x_0 are jointly Gaussian, all three estimators coincide (Aravkin et al. (2017)).

For linear Gaussian state-space models, the conventional Kalman filter yields the minimum mean-square estimate 3.10, when applied for estimating the hidden dynamic state process $\{x_t\}_{t=1}^T$ from the observed sequence $\{y_t\}_{t=1}^T$.

“R.E. Kalman in 1960 published his famous paper describing a recursive solution to the discrete-data linear filtering problem. Since then the Kalman filter has been the subject of extensive research and application.”(Bishop et al., 2001, p.1). Simply, the Kalman filter is an optimal recursive data processing algorithm. It incorporates all information, which can be provided to it to estimate past, present and future values of states. The Kalman Filter algorithm is as follows Javaheri et al. (2003):

Algorithm 1: The Conventional Kalman Filter

```
1 Initial Data (at $t = 0$) The initial state is $x_0 \sim \mathcal{N}(\bar{x}_0, \Pi_0)$
      Set initial values for the filter: $\hat{x}_{0|0} = \bar{x}_0$ and $P_{0|0} = \Pi_0$
2 Time Update: (at $t = 1, \ldots, T$) At time $t-1$ the filtered estimates
      $\hat{x}_{t-1|t-1}$ and $P_{t-1|t-1}$ are known.            ▷ Priori estimation
      Predict for time $t$:
      $\hat{x}_{t|t-1} = F\hat{x}_{t-1|t-1} + Bu_{t-1}$            Predicted state estimate
      $P_{t|t-1} = FP_{t-1|t-1}F^T + GQG^T$                  Predicted error covariance
3 Measurement Update (at $t = 1, \ldots, T$) At time $t$ the predicted estimates
      $\hat{x}_{t|t-1}$ and $P_{t|t-1}$ are known.                ▷ Posteriori Estimation
   Use measurement data $y_t$ at time t to compute:
      $e_t = y_t - H\hat{x}_{t|t-1}$                          Residual
      $R_{e,t} = HP_{t|t-1}H^T + R,$                       Residual error covariance
      $K_t = P_{t|t-1}H^T R_{e,t}^{-1},$                        Kalman Gain
      $\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t e_t$                    Updated state estimate
      $P_{t|t} = (I - K_t H)P_{t|t-1}.$                      Updated error covariance
```

The conventional Kalman filter algorithm 1 has three stages. The first called the Initial Data, the second being the Time Update and the third the Measurement Update. In simple words, the Algorithm can be described as follows:

- The initial values $\hat{x}_{0|0}$ and $P_{0|0}$, that are chosen for the filter do not need to be close to the unobservable true values. It is preferable to set $\hat{x}_{0|0}$ close to its real value, however, this is not possible since the state vector of the model is unobservable. Some constant for $\hat{x}_{0|0}$ could be selected, which would mean we initialize the filter with incorrect values. As a result, the Kalman filter will follow a wrong state trajectory. At every sampling time, however, true measurements are entered into the filter which improves the state estimates. This moves the state trajectory, each time step, closer to the true solution.

- The time update stage 2, is the stage of going from $\hat{x}_{0|0}$ to $\hat{x}_{1|0}$. This one-step ahead predicted state estimate, $\hat{x}_{1|0}$, is computed simultaneously with the corresponding predicted error covariance matrix $P_{1|0}$.

- Next, at time instant 1, the Kalman filter receives information about the true state from the available measurements. This stage is called the measurement update stage 3. There the filtered estimate $\hat{x}_{1|1}$ is calculated simultaneously with the corresponding error covariance matrix $P_{1|1}$. The Kalman gain in this stage, K_t, relates to the mean of the conditional distribution of $x_{t|t}$ given observation y_t or equivalently, the matrix that would minimize the mean square error of $P_{t|t}$.

- In general the time update step is the step of going from $\hat{x}_{t-1|t-1}$ to $\hat{x}_{t|t-1}$ with the corresponding $P_{t|t-1}$ error covariance matrix, and the measurement update step is where $\hat{x}_{t|t}$ and $P_{t|t} = \mathbf{E}[(x_t - \hat{x}_{t|t})(x_t - \hat{x}_{t|t})^T]$ are calculated.

Econometric state-space models are often parameterized. What this means is that some or all of the entries of the system matrices $\{F, H, R, Q, B\}$ in (3.6)-(3.7) may depend on the unknown system parameter vector $\theta \in \mathbb{R}^p$. This needs to be estimated simultaneously with the dynamic state, x_t, from the only available signal $Y_t = \{y_1, \ldots, y_t\}$. In practice the ML estimation method is often used for estimating the system parameters θ_i, $i = 1, \ldots, p$. The log LF for the state-space model (3.6)-(3.7) is given by Harvey (1989) as follows

$$\begin{aligned} \ln L\left(\theta|Y_T\right) &= \sum_{t=1}^{T} \ln p(y_t|Y_{t-1}) \\ &= -\frac{mT}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left\{\ln\left(\det R_{e,t}\right) + e_t^T R_{e,t}^{-1} e_t\right\} \end{aligned} \tag{3.12}$$

where $Y_T = \{y_1, \ldots, y_T\}$ is the T-step measurement history and $\{e_t\}$ are the discrete-time Kalman filter innovations generated by the underlying Algorithm(1).

The derivation of the log LF 3.12 is based on the assumption of optimality of the Kalman filter estimator for Gaussian state-space models with the property $e_t \sim \mathcal{N}\left(0, R_{e,t}\right)$. For more information, refer to Schweppe (1965). In the non-Gaussian setting, however, the Kalman filter performs sub-optimally with regards to MSE, yielding minimum linear mean-square estimate 3.11 rather than minimum mean square estimate 3.10, with the above property violated. This is because the Kalman filter estimate obtained is optimal for the class of linear functions of measurements 3.11. Ruiz (1994) and Harvey et al. (1994) explain the quasi-maximum likelihood estimation by formula 3.12 in a non-Gaussian setting.

3.2.3 Econometric State-Space Modelling Approach

Over the past 20-25 years, econometricians have begun using Kalman's framework to represent a wide variety of structural econometric models into state-space form. This provides many advantages to their analysis and Wilcox and Hamano (2017) have recently published a comprehensive survey on how Kalman's work has been applied. Wilcox and Hamano (2017) summarise three innovative applications of state-space models (3.6)-(3.7) in the field of econometrics which are: (i) time-varying coefficient regression models; (ii) time-varying variance models, e.g. SV models; (iii) time series clustering systems.

The QML method, computed using the Kalman filter can be used to estimate SV models. The QML estimator was proposed independently by Nelson (1988) and Harvey et al. (1994). The state-space representation of the SV model equation (3.1) needs to be linearised and to do this the logarithm is taken. Doing this the equations (3.3)-(3.4) are obtained.

It is often assumed in the literature that $\varepsilon_t \sim \mathcal{N}(0,1)$ and it is proved in Ruiz (1994) that the mean of $\ln(\varepsilon_t^2)$ is approximately -1.27. As a result, the SV model often studied in the literature is given as

$$\ln(y_t^2) = -1.27 + h_t + \zeta_t, \tag{3.13}$$

$$h_t = \gamma + \phi h_{t-1} + \eta_{t-1}, \quad \eta_t \sim \mathcal{N}(0, \sigma_\eta^2) \tag{3.14}$$

where ζ_t is non-Gaussian, zero mean white noise.

As mentioned previously, for non-Gaussian state-space models, the Kalman filter will only yield minimum mean square linear estimators, rather than minimum mean square estimators. The exact likelihood will not be obtained from the resulting prediction errors since the model is not Gaussian. According to Ruiz (1994) estimates can be obtained by making $\zeta_t \sim \mathcal{N}(0, \pi^2/2)$ and then maximising the resulting quasi-likelihood function.

The SV model with Student-t distributed errors, will be estimated using the approach proposed by Ruiz (1994) in this dissertation. See, (Ruiz, 1994, p.7), for exact approach. Ruiz (1994) estimate the SV model with Student-t distributed errors, estimating the same parameters, ϕ, σ_η^2 and σ_ζ^2, however, the degree of freedom parameter, ν, is estimated from the covariance of the measurement noise estimate.

In this dissertation, the SV models given in a general form are considered.

They are represented in the state-space form as follows

$$\ln(y_t^2) = h_t + \zeta_t, \tag{3.15}$$

$$h_t = \phi h_{t-1} + \eta_{t-1}, \tag{3.16}$$

where y_t is the mean-adjusted return series, the state process noise η_t and measurement noise ζ_t, are zero mean white noise processes with covariances $Q = \sigma_\eta^2$ and $R = \sigma_\zeta^2$, respectively, but they are not necessarily Gaussian. Two possibilities are considered (i) the Gaussian model (3.15)-(3.16) (i.e. $\zeta_t \sim \mathcal{N}(0, \sigma_\zeta^2)$ and $\eta_t \sim \mathcal{N}(0, \sigma_\eta^2)$) and (ii) Student-t distributed.

When comparing state-space model (3.6)-(3.7) with (3.15)-(3.16), it is concluded that $F := \phi$, $B := 0$, $G := 1$, $Q := \sigma_\eta^2$ and $H := 1$, $R := \sigma_\zeta^2$. The Kalman filter (Algorithm (1)) is used for computing the log LF for the QML estimation of the unknown system parameters and for extracting the unknown state process, x_t, where $x_t := h_t = \ln(\sigma_t^2)$ and the measurement vector is the $\ln(y_t^2)$ where y_t is the daily share price return at time t.

Chapter 4

Empirical Results

4.1 Data Description

The Johannesburg Stock Exchange is the oldest existing and largest stock exchange in Africa. The JSE Top 40 index is used as a proxy for the South African market and consists of 40 of the largest stocks on the exchange, based on market capitalisation. Four stocks were selected from the JSE, which are seen as representative components of the JSE Top 40 index. Table 4.1 provides a description of the stocks to be analysed. Continuously-compounded daily log returns are used. This is calculated using $y_t = \ln(\frac{S_t}{S_{t-1}})$ for $t = 1, \ldots, T$, where S_t is the daily closing price at time t and y_t is the daily log return at time t. The daily closing prices were extracted from Bloomberg and was taken over the period 29/12/2006 to 30/12/2016. Based on the period taken 2500 observations were used per .

Table 4.1: *Company Profiles of the JSE Top 40 Index*

Series	**Company**	**Sector**	**Market Cap**
AGL	Anglo American PLC	Basic Resources (Mining)	ZAR 241,33 bn
BIL	BHP	Basic Resources (Mining)	ZAR 403,49 bn
FSR	FirstRand Ltd	Banking	ZAR 233,39 bn
SBK	Standard Bank Group Ltd	Banking	ZAR 267,29 bn

4.2 Preliminary Data Analysis

Table 4.2 below gives summary statistics of the daily log return series for the JSE Top 40 index and four different shares from the index. All the return series exhibited excess kurtosis, with the two mining stocks having positive skewness, while the two banking stocks and the JSE Top 40 index having negative skewness. This indicates that they are asymmetrically distributed and have fatter tails than the normal distribution.

The Dickey-Fuller [1] unit root test was conducted on each of the return series and none had a unit root present at a 5% significance level. As a result, all the return series are stationary at a 5% significance level. To test if any of the return series exhibit conditional heteroscedasticity, the Engle's test was conducted. All the return series displayed conditional heteroscedasticity. Appendix A has figures of the sample autocorrelation plots of each of the return series. Each of the return series autocorrelation plots is similar in what is seen for lag 1, which will be discussed below. These properties of the return series conform with the stylized statistical properties of asset returns as mentioned by Cont (2001).

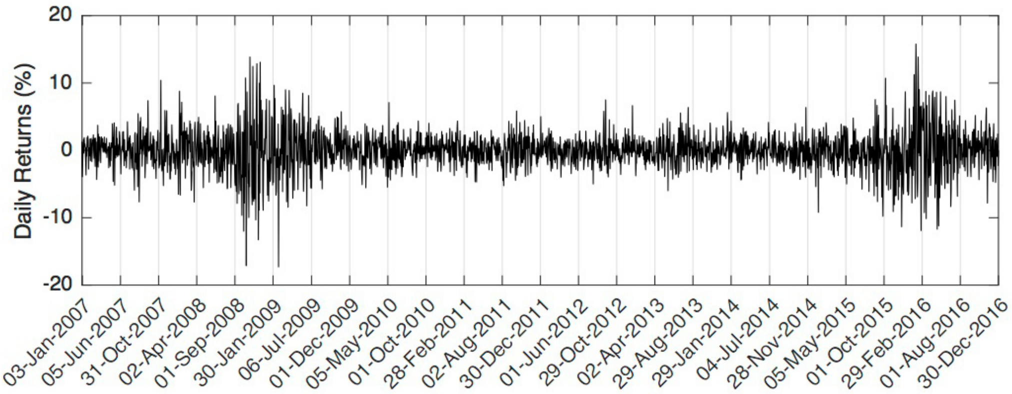

Figure 4.1: *Daily Share Price Returns (%) of Anglo American*

The daily log returns of the two mining stocks, Anglo American and BHP, are illustrated in Figure 4.1 and Figure 4.3 respectively. Both stocks were affected by the 2008 financial crisis with high levels of volatility experienced around the 2008 period. We expected this, as both stocks have businesses

---

[1] only the first lag, p = 1, is of interest because of the result of the Dickey-Fuller test.

Table 4.2: *Summary Statistics of Return Series for JSE Top 40 Index and four Selected Shares*

JSE Code	**AGL**	**BIL**	**FSR**	**SBK**	**Top 40**
No. of observations	2500	2500	2500	2500	2500
Min.	−0.173	−0.114	-0.161	-0.146	-0.08
1st Quartile	−0.015	−0.012	-0.011	-0.01	-0.007
Median	−0.0002	0.0005	0.0005	0	0.0008
3rd Quartile	0.015	0.013	0.012	0.011	0.007
Max.	0.158	0.18	0.122	0.104	0.077
Mean	-0.0002	0.0002	0.0004	0.0001	0.0003
Std. Dev.	0.029	0.024	0.02	0.019	0.014
Skewness	0.017	0.236	-0.216	-0.096	-0.083
Kurtosis	6.655	6.869	6.714	6.722	6.437

which are US Dollar-denominated and as a result would be affected directly by high levels of uncertainty in the US market. We also note that Anglo American has the highest single-day decrease of 17.3% and BHP having the highest single-day increase of 18%. Both stocks also had positive skewness, as mentioned previously, and had the highest volatility over the entire period analysed. There is also a second period around the end of 2015-2016 of high volatility for both stocks. This is more evident in the daily log returns of Anglo American than BHP from Figure 4.1 and Figure 4.3. This period was when there were high levels of uncertainty surrounding the South African market, due to the South African finance minister being dismissed. Around this period there were also calls to nationalise mines and banks in South Africa. In Appendix A, Figure A.1 and Figure A.2 illustrate the sample autocorrelation plots of Anglo American and BHP respectively. Since both are not significantly different from zero at the first lag, both stocks may be seen as efficient over the period under study.

The daily log returns of the two banking stocks, FirstRand and Standard Bank Group, are illustrated in Figure 4.2 and Figure 4.4, respectively. It can be seen that both stocks were affected by the 2008 financial crisis, from the increased volatility of the daily log returns. The second period where there was high volatility was around the end of 2015. This is where both stocks had there highest negative daily log return of 16.1% and 14.6% for FirstRand and Standard Bank Group, respectively. This exactly coincides

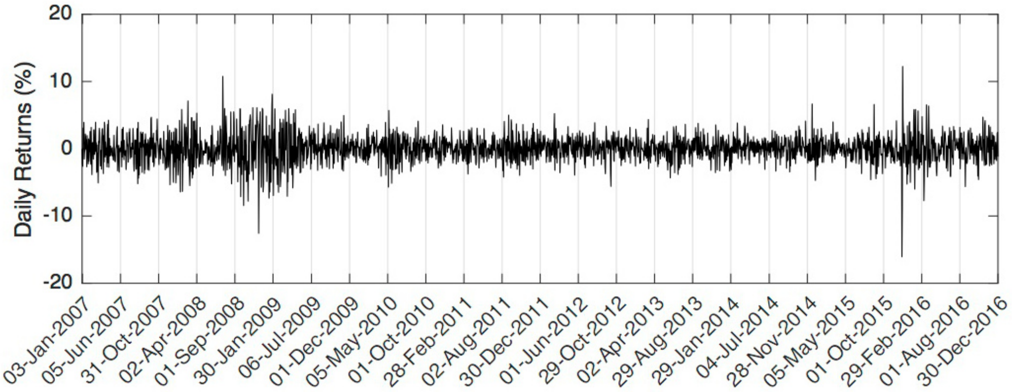

Figure 4.2: *Daily Share Price Returns (%) of FirstRand*

with the dates that the South African Minister of Finance was dismissed and it sent these share prices falling on the day. FirstRand did, however, have its highest positive daily log return a few days later of 12.2%. In Appendix A, Figure A.3 and Figure A.4 illustrate the sample autocorrelation plots of FirstRand and Standard Bank Group, respectively. Both plots again indicate that the first lag is not significantly different from zero and suggests that both FirstRand and Standard Bank Group may be efficient over the period under study. Figure 4.5 illustrates the daily log returns of the JSE Top 40 index.

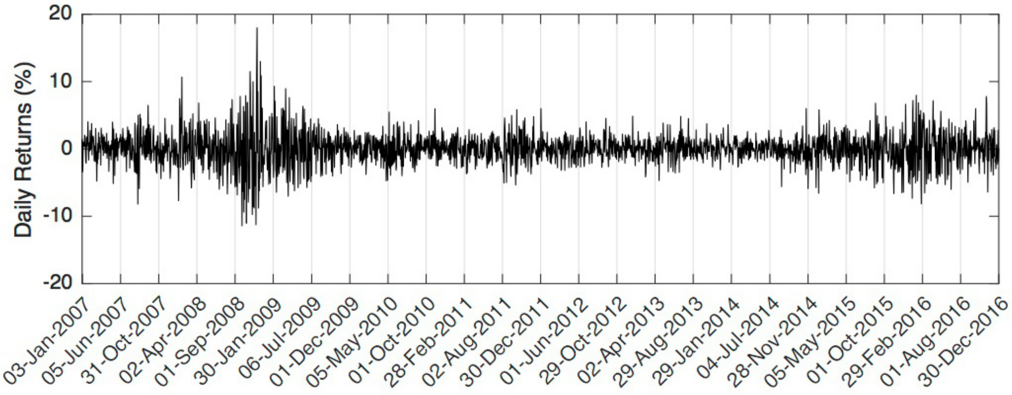

Figure 4.3: *Daily Share Price Returns (%) of BHP*

As was the case with the four components of the JSE Top 40 index, the volatility that the 2008 financial crisis caused, can be seen from Figure 4.5 around the 2008 period. The other period, which had the highest levels of volatility and common amongst the four shares discussed above, was the

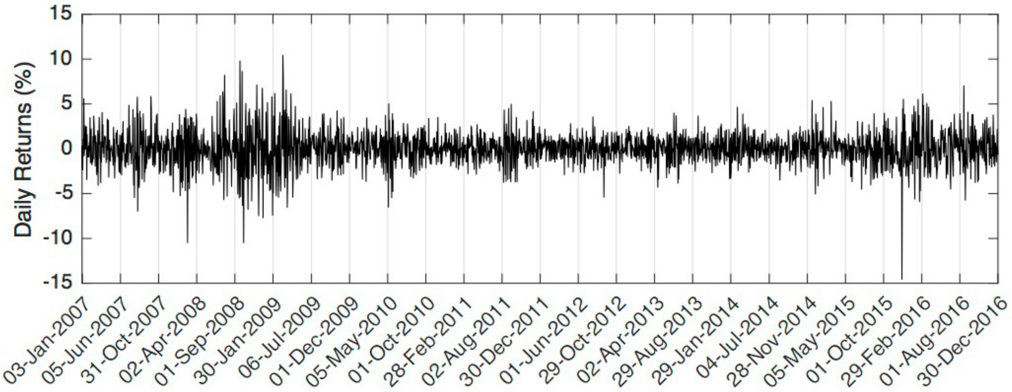

Figure 4.4: *Daily Share Price Returns (%) of Standard Bank Group*

period around the end of 2015. This period, however, is not as evident in Figure 4.5 as was with the four components of the JSE Top 40 index. There is a spike in volatility, however, it is not as pronounced. This is expected as the JSE Top 40 index is made up of 40 stocks, which collectively leads to a reduction in volatility. In Appendix A, Figure A.5 illustrates the sample autocorrelation plot of the JSE Top 40 index. As a result, we see that the autocorrelation for the first lag is not significantly different from zero. This suggests the JSE Top 40 index may be efficient over the period under study.

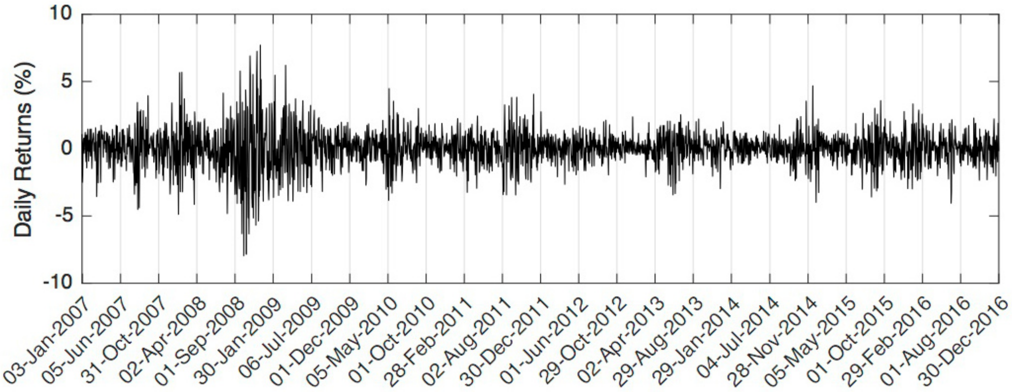

Figure 4.5: *Daily Index Returns (%) of the JSE Top 40 Index*

4.3 Estimation Results

To estimate stochastic volatility of the five series; the Gaussian GARCH(1,1) model, GARCH(1,1) model with Student-t distributed errors, Stochastic Volatility model with Gaussian errors and Stochastic Volatility model with Student-t distributed errors were used. These models were estimated using the MATLAB econometric toolbox. Table 4.4 and Table 4.5 below summarise the estimated parameters for the Gaussian GARCH(1,1) model and the GARCH(1,1) model with Student-t distributed errors, respectively. Table 4.7 and Table 4.8 below represent the resulting estimates for the Stochastic Volatility model with Gaussian errors and Stochastic Volatility model with Student-t distributed errors, respectively.

Table 4.3: *AIC Values for the Examined Models, where * and ▼ Indicates the Lowest AIC for the GARCH and SV models, respectively*

Series	**AGL**	**BIL**	**FSR**	**SBK**	**JSE Top 40**
Gaussian GARCH(1,1)	11538.2	10703.9	9829.3	10233.4	7899.9
GARCH(1,1) Student-t	11510.7*	10677.5*	9752.6*	10132.9*	7884.8*
SV model Gaussian	11270.2	11339.2	11324.9	11493.8	11116.5
SV model Student-t	7985.5▼	8158.7▼	8219.7▼	8224.9▼	7995.2 ▼

The Akaike information criterion (AIC) is used to select the best fitting model for each of the series. A lower AIC value corresponds with a better fitting model. The AIC is defined as (Akaike (1987))

$$AIC_{\pi} = -2log(L_{\pi}) + 2k, \tag{4.1}$$

where π indicates a fitted model and L_{π} is the likelihood of model π maximized over k parameters. The AIC is seen as a comprise between model fit and model complexity. The goodness of fit of a model can be improved by employing more complicated models, however, this is penalized by the second term of the equation 4.1. The AIC values for the models are summarised in Table 4.3. It can be seen that among the GARCH and SV classes, both with Student-t distributed errors perform the best for each of the four different return series and the index.

Table 4.4: *Parameter Estimation of the Gaussian GARCH(1,1) Model, where the Corresponding Standard Errors are given in Parenthesis*

Series	**Estimated Parameters**		
	Constant($\hat{\alpha}_0$)	ARCH ($\hat{\alpha}_1$)	GARCH ($\hat{\beta}_1$)
AGL	0.0543	0.0567	0.9357
	(0.0163)	(0.0074)	(0.0081)
BIL	0.0343	0.0531	0.9402
	(0.0103)	(0.0065)	(0.0073)
FSR	0.0700	0.0568	0.9256
	(0.0169)	(0.0082)	(0.0106)
SBK	0.0476	0.0721	0.9154
	(0.0122)	(0.0083)	(0.0099)
JSE Top 40	0.0237	0.0909	0.8958
	(0.0057)	(0.0113)	(0.0127)

Looking at the parameter estimates $\hat{\alpha}_1$ and $\hat{\beta}_1$, from Table 4.4 and Table 4.5, it is seen that $\hat{\alpha}_1 + \hat{\beta}_1$ yields values close to 1 for all the series. This was expected as it confirms with a stylized fact of asset returns (Cont (2001)), that the autocorrelation function remains positive and decays slowly, which is the well-known phenomenon of volatility clustering.

From Table 4.6 and Table 4.7, we see that the estimate $\hat{\phi}$ for each of the series are all close to one. The parameter $\hat{\phi}$ is an indication of volatility persistence and hence a value for it close to one was to be expected since this is a stylized fact of asset return. The values for the estimates $\hat{\sigma_\eta^2}$ are relatively small, which indicates a low influence of the process model uncertainties. On the other hand, the values for the estimates $\hat{\sigma_\zeta^2}$, are larger with values between 2.27 and 6.2. These values for $\hat{\sigma_\zeta^2}$ are to be expected, because looking at the Figures (A.1 - A.5), large deviations in the daily returns are observed and as a result this is reflected through the estimates $\hat{\sigma_\zeta^2}$.

The parameter, $\frac{1}{\hat{\nu}}$ in both Table 4.5 and 4.7 is not close to 0 which would be an indication of normality. As a result, this supports why the normal distribution is not an accurate model and why the Student-t distribution was considered.

Table 4.8 summarises the results of the likelihood ratio test for the examined GARCH models. For AGL, the maximum log LF value under the Gaussian GARCH(1,1) model is -5764.14, which is less than the GARCH(1,1)

Table 4.5: *Parameter Estimation of the GARCH(1,1) Model with Student-t Distributed Errors, where the Corresponding Standard Errors are given in Parenthesis*

Series	**Estimated Parameters**			
	Constant($\hat{\alpha}_0$)	ARCH ($\hat{\alpha}_1$)	GARCH ($\hat{\beta}_1$)	$\frac{1}{\hat{\nu}}$
AGL	0.0511	0.0565	0.9366	0.0807
	(0.0202)	(0.0092)	(0.0102)	
BIL	0.0261	0.0513	0.9442	0.0817
	(0.0118)	(0.0079)	(0.0086)	
FSR	0.0579	0.0599	0.9248	0.1108
	(0.0193)	(0.0107)	(0.0129)	
SBK	0.0518	0.0717	0.9136	0.0954
	(0.0164)	(0.0113)	(0.0131)	
JSE Top 40	0.0206	0.0902	0.8991	0.0659
	(0.0065)	(0.0127)	(0.0139)	

model with Student-t distributed errors of -5749.39. Looking at the likelihood ratio test statistics of 29.50, it is highly significant when compared with the $\chi^2_{k,0.99} = 6.63$. We can conclude that the Gaussian GARCH(1,1) model is restrictive when compared with the GARCH(1,1) model with Student-t distributed errors. The same conclusion can be drawn for each of the other series.

The noisy return data was passed into an optimization routine to search for optimal parameters. This process involved the use of the Kalman filter in conjunction with ML estimation. Given random but feasible initial parameters, the optimization iterates to obtain a parameter set that gives a minimum likelihood proxy. The optimized parameters were then used in recovering the optimized state processes (hidden volatility process) via Kalman filtering. This can be achieved by using the fmincon function in MATLAB with the objective function being the likelihood proxy, within reasonable upper and lower bounds. Furthermore, parameter recovery was very much dependent on the initial values used. For each run, initial parameters were randomly selected within a feasible range. It was found that only when the initial values were closed to the actual (assumed) values, were parameters approximately recovered. In some cases, the optimizer fails to find optimal parameters when the initial values used, are far from the local/global maximum. The quality of the results obtained can be improved by running the

Table 4.6: *Parameter Estimation of the Stochastic Volatility Model with Gaussian Distributed Errors, where the Corresponding Standard Errors are given in Parenthesis*

Series	**Estimated Parameters**		
	$\hat{\phi}$	$\hat{\sigma}_\eta^2$	$\hat{\sigma}_\zeta^2$
AGL	0.9928	0.1136	2.2512
	(0.0032)	(0.0175)	(0.0205)
BIL	0.9899	0.1157	2.2854
	(0.0040)	(0.0198)	(0.0209)
FSR	0.9880	0.1059	2.2857
	(0.0050)	(0.0223)	(0.0223)
SBK	0.9873	0.1182	2.3602
	(0.0052)	(0.0229)	(0.0216)
JSE Top 40	0.9954	0.1265	2.1728
	(0.0026)	(0.0177)	(0.0236)

optimization for a larger sample of initial values. The optimized parameters can then be cycled into the optimizer to give more accurate results. Due to the time constraints and the computing power available, however, these improvements were not possible.

The likelihood ratio test results for the examined SV models are presented in Table 4.9. For each of the series, it can be seen that the SV model with Student-t distributed errors is highly significant when compared to $\chi^2_{k,0.99}$ = 6.63. We can conclude that the Stochastic Volatility model with Gaussian distributed errors is too restrictive when compared to the Stochastic Volatility model with Student-t distributed errors.

The results obtained from the likelihood ratio test, concur with the results of the AIC, for the GARCH models and the SV models. The models with Student-t distributed errors, outperform the corresponding models with Gaussian errors.

Table 4.7: *Parameter Estimation of the Stochastic Volatility Model with Student-t Distributed Errors*

Series	**Estimated Parameters**			
	$\hat{\phi}$	$\hat{\sigma}^2_\eta$	$\hat{\sigma}^2_{\zeta_t}$	$\frac{1}{\hat{\nu}}$
AGL	0.9975	0.0061	5.2781	0.2697
BIL	0.9953	0.0075	5.5530	0.2610
FSR	0.9887	0.0115	5.5639	0.2517
SBK	0.9839	0.0160	6.1949	0.3011
JSE Top 40	0.9943	0.0129	4.8304	0.2264

Table 4.8: *Likelihood ratio (LR) test statistics, with the Gaussian GARCH model represented by GARCH-$\mathcal{N}$ and the GARCH model with Student-t distributed errors represented by GARCH-t*

Series	**Maximum Log LF**		**LR test statistics**			
	GARCH-$\mathcal{N}$	GARCH-t	LR	k	$\chi^2_{k,0.99}$	p-value
AGL	-5764.14	-5749.39	29.50	1	6.63	0.00
BIL	-5346.98	-5332.77	28.44	1	6.63	0.00
FSR	-5111.73	-5060.49	102.48	1	6.63	0.00
SBK	-4909.64	-4870.30	78.68	1	6.63	0.00
JSE Top 40	-3944.93	-3936.38	17.09	1	6.63	0.00

Table 4.9: *Likelihood ratio (LR) test statistics, with the Stochastic Volatility model with Gaussian distributed errors (abbreviated as SVM-$\mathcal{N}$) and the Stochastic Volatility model with Student-t distributed errors (abbreviated as SVM-t)*

Series	**Maximum Log LF**		**LR test statistics**			
	SVM-$\mathcal{N}$	SVM-t	LR	k	$\chi^2_{k,0.99}$	p-value
AGL	-5632.12	-3988.78	3186.7	1	6.63	0.00
BIL	-5666.58	-4075.35	3182.3	1	6.63	0.00
FSR	-5659.47	-4105.94	3107.4	1	6.63	0.00
SBK	-5743.90	-4108.45	3271.0	1	6.63	0.00
JSE Top 40	-5555.25	-3993.68	3123.2	1	6.63	0.00

Chapter 5

Conclusion

In this dissertation, GARCH(1,1) models and SV models were estimated, using financial series data from the South African market. The GARCH models and SV models were compared using Gaussian distributed errors and Student-t distributed errors. The method of maximum likelihood was used to estimate the GARCH models, while the SV models were estimated using a quasi-maximum likelihood method, computed using the Kalman filter.

In the empirical study, AIC was used as the metric to select the best performing models. We found that both the GARCH and SV model using Student-t distributed errors perform best for the series examined in this dissertation. This indicates that the leptokurtosis observed in the South African market financial series, is captured better by the Student-t distributed errors than by the Gaussian distributed errors.

As an extension, further research could be done on estimating SV models with other distributions. In particular, distributions with heavy tails and that can capture the skewness observed in the return series could perform better. This could be performed by using implied volatility of the underlying as the volatility proxy, and comparing how well these models compare out-of-sample to implied volatility. The reason why implied volatility could be used is due to the real world and risk neutral volatility being the same under the assumption of geometric Brownian motion. The quality of the results can also be improved by running the optimization for a larger sample of initial values, as well as iterating through the accurate parameter values. In addition, high-performance computing can be employed to improve optimization times.

Bibliography

Akaike, H. (1987). Factor analysis and AIC, *Selected Papers of Hirotugu Akaike*, Springer, pp. 371–386.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2001). The distribution of realized exchange rate volatility, *Journal of the American Statistical Association* **96**(453): 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003). Modeling and forecasting realized volatility, *Econometrica* **71**(2): 579–625.

Andersen, T. G., Chung, H.-J. and Sørensen, B. E. (1999). Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo Study, *Journal of Econometrics* **91**(1): 61–87.

Andersen, T. G. and Lund, J. (1997). Estimating continuous-time stochastic volatility models of the short-term interest rate, *Journal of Econometrics* **77**(2): 343–377.

Andersen, T. G. and Sørensen, B. E. (1996). GMM estimation of a stochastic volatility model: A Monte Carlo Study, *Journal of Business & Economic Statistics* **14**(3): 328–352.

Aravkin, A., Burke, J. V., Ljung, L., Lozano, A. and Pillonetto, G. (2017). Generalized Kalman smoothing: Modeling and algorithms, *Automatica* **86**: 63–86.

Bansal, R., Gallant, A. R., Hussey, R. and Tauchen, G. (1994). Computational aspects of nonparametric simulation estimation, *Computational Techniques for Econometrics and Economic Analysis*, Springer, pp. 3–22.

Bansal, R., Gallant, A. R., Hussey, R. and Tauchen, G. (1995). Nonparametric estimation of structural models for high-frequency currency market data, *Journal of Econometrics* **66**(1-2): 251–287.

Bishop, G., Welch, G. et al. (2001). An introduction to the Kalman filter, *Proc of SIGGRAPH, Course* **8**(27599-3175): 59.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities, *Journal of Political Economy* **81**(3): 637–654.

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

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return, *The Review of Economics and Statistics* pp. 542–547.

Broto, C. and Ruiz, E. (2004). Estimation methods for stochastic volatility models: A survey, *Journal of Economic Survey* **18**(5): 613–649.

Chib, S., Nardari, F. and Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models, *Journal of Econometrics* **108**(2): 281–316.

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues, *Quantitative Finance* **1**(2): 223–236.

Danielsson, J. (1994). Stochastic volatility in asset prices estimation with simulated maximum likelihood, *Journal of Econometrics* **64**(1): 375–400.

Danielsson, J. and Richard, J.-F. (1993). Accelerated Gaussian importance sampler with application to dynamic latent variable models, *Journal of Applied Econometrics* **8**(S1): S153–S173.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society: Series B (Methodological)* **39**(1): 1–22.

Duffie, D. and Singleton, K. J. (1990). Simulated moments estimation of Markov models of asset prices.

Elliott, R. J. and Hyndman, C. B. (2007). Parameter estimation in commodity markets: A filtering approach, *Journal of Economic Dynamics and Control* **31**(7): 2350–2373.

EmmPack 1.01: C/C++ code for use with O (n.d.). *Technical report*.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica: Journal of the Econometric Society* **50**: 987–1007.

Engle, R. F. and Lee, G. (1996). Estimating diffusion models of stochastic volatility. In: P.E. Rossi (ed.), *Modelling Stock Market Volatility: Bridging the Gap to Continuous Time* pp. 333 – 355.

Fiorentini, G., Leon, A. and Rubio, G. (2002). Estimation and empirical performance of Heston's stochastic volatility model: the case of a thinly traded market, *Journal of Empirical Finance* **9**(2): 225–255.

Fridman, M. and Harris, L. (1998). A maximum likelihood approach for non-Gaussian stochastic volatility models, *Journal of Business & Economic Statistics* **16**(3): 284–291.

Gallant, A. R., Hsieh, D. and Tauchen, G. (1997). Estimation of stochastic volatility models with diagnostics, *Journal of Econometrics* **81**(1): 159–192.

Gallant, A. R. and Tauchen, G. (1996). Which moments to match?, *Econometric Theory* **12**(4): 657–681.

Gallant, A. R. and Tauchen, G. (1999). The relative efficiency of method of moments estimators, *Journal of Econometrics* **92**(1): 149–172.

Geweke, J. et al. (1994). Bayesian comparison of econometric models, *Technical report*, Working Paper, Federal Reserve Bank of Minneapolis, Minnesota.

Ghysels, E., Harvey, A. C. and Renault, E. (1996). "Stochastic Volatility" in G.S. Maddala (ed.), *Handbook of Statistics* **14**: 119–191.

Ghysels, E., Jasiak, J. et al. (1994). *Stochastic Volatility and Time Deformation: An Application of Trading Volume and Leverage Effects*, Université de Montréal, Centre de recherche et développement en économique.

Gourieroux, C., Monfort, A. and Renault, E. (1993). Indirect inference, *Journal of Applied Econometrics* **8**(S1): S85–S118.

Grewal, M. S. and Andrews, A. P. (2015). *Kalman Filtering the Practice Using MATLAB*, New Jersey: Prentice Hall.

Gupta, N. and Mehra, R. (1974). Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations, *IEEE Transactions on Automatic Control* **19**(6): 774–783.

Hamilton, J. D. (1994). Time Series Econometrics, *Princeton U. Press, Princeton* .

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators, *Econometrica: Journal of the Econometric Society* **50**: 1029–1054.

Harvey, A. (1987). Applications of the Kalman Filter in Econometrics, Chapter 8, *Cambridge University Press* **77**: 285.

Harvey, A. C. (1989). *Forecasting Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.

Harvey, A., Ruiz, E. and Shephard, N. (1994). Multivariate stochastic variance models, *The Review of Economic Studies* **61**(2): 247–264.

Jacquier, E., Polson, N. G. and Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models, *Journal of Business & Economic Statistics* **12**(4): 69–87.

Jacquier, E., Polson, N. G. and Rossi, P. E. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors, *Journal of Econometrics* **122**(1): 185–212.

Javaheri, A., Lautier, D. and Galli, A. (2003). Filtering in finance, *Wilmott* **3**: 67–83.

Jiang, G. J. and van der Sluis, P. J. (1998). Forecasting volatility under multivariate stochastic volatility via reprojection.

Jiang, G. J. and van der Sluis, P. J. (1999). Index option pricing models with stochastic volatility and stochastic interest rates, *Review of Finance* **3**(3): 273–310.

Kim, S., Shephard, N. and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models, *The Review of Economic Studies* **65**(3): 361–393.

Kitagawa, G. (1987). Non-Gaussian statespace modeling of nonstationary time series, *Journal of the American Statistical Association* **82**(400): 1032–1041.

Klein, A. and Neudecker, H. (2000). A direct derivation of the exact fisher information matrix of Gaussian vector state space models, *Linear Algebra and its Applications* **321**(1-3): 233–238.

Koopman, S. J. (1993). Disturbance smoother for state space models, *Biometrika* **80**(1): 117–126.

Koopman, S. J. and Hol Uspensky, E. (2002). The stochastic volatility in mean model: empirical evidence from international stock markets, *Journal of Applied Econometrics* **17**(6): 667–689.

Koopman, S. J. and Shephard, N. (1992). Exact score for time series models in state space form, *BIOMETRIKA-CAMBRIDGE-* **79**: 823–823.

Lee, B.-S. and Ingram, B. F. (1991). Simulation estimation of time-series models, *Journal of Econometrics* **47**(2-3): 197–205.

Liesenfeld, R. (1998). Dynamic bivarsate mixture models: Modeling the behavior of prices and trading volume, *Journal of Business & Economic Statistics* **16**(1): 101–109.

Liesenfeld, R. and Jung, R. C. (2000). Stochastic volatility models: conditional normality versus heavy-tailed distributions, *Journal of Applied Econometrics* **15**(2): 137–160.

Lombardi, M. J. and Calzolari, G. (2009). Indirect estimation of α-stable stochastic volatility models, *Computational Statistics & Data Analysis* **53**(6): 2298–2308.

Mader, W., Linke, Y., Mader, M., Sommerlade, L., Timmer, J. and Schelter, B. (2014). A numerically efficient implementation of the expectation maximization algorithm for state space models, *Applied Mathematics and Computation* **241**: 222–232.

Mahieu, R. and Bauer, R. (1998). A Bayesian analysis of stock return volatility and trading volume, *Applied Financial Economics* **8**(6): 671–687.

Melino, A. and Turnbull, S. M. (1990). Pricing foreign currency options with stochastic volatility, *Journal of Econometrics* **45**(1-2): 239–265.

Monfardini, C. (1998). Estimating stochastic volatility models through indirect inference, *The Econometrics Journal* **1**(1): 113–128.

Nelson, D. B. (1988). *The time series behavior of stock market volatility and returns*, PhD thesis, Massachusetts Institute of Technology.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach, *Econometrica: Journal of the Econometric Society* pp. 347–370.

Némesin, V. and Derrode, S. (2012). Robust blind pairwise Kalman algorithms using QR decompositions, *IEEE Transactions on Signal Processing* **61**(1): 5–9.

Némesin, V. and Derrode, S. (2014). Robust partial-learning in linear Gaussian systems, *IEEE Transactions on Automatic Control* **60**(9): 2518–2523.

Nilmini Kumari, S. (2013). Characterization of students t- distribution with some application to finance, *Mathematical Theory and Modeling* **3**: 1–9.

Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models, *Journal of Econometrics* **63**(1): 289–306.

Sandmann, G. and Koopman, S. J. (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood, *Journal of Econometrics* **87**(2): 271–301.

Schweppe, F. (1965). Evaluation of likelihood functions for Gaussian signals, *IEEE Transactions on Information Theory* **11**(1): 61–70.

Scott, L. O. (1987). Option pricing when the variance changes randomly: Theory, estimation, and an application, *Journal of Financial and Quantitative analysis* **22**(4): 419–438.

Segal, M. and Weinstein, E. (1988). A new method for evaluating the log-likelihood gradient (score) of linear dynamic systems, *IEEE Transactions on Automatic Control* **33**(8): 763–766.

Shephard, N. (1993). Fitting nonlinear time-series models with applications to stochastic variance models, *Journal of Applied Econometrics* **8**(S1): S135–S152.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility, *Monographs on Statistics and Applied Probability* **65**: 1–68.

Shephard, N. and Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series, *Biometrika* **84**(3): 653–667.

Shumway, R. H. and Stoffer, D. S. (1982). An approach to time series smoothing and forecasting using the EM algorithm, *Journal of Time Series Analysis* **3**(4): 253–264.

So, M. E. P., Lam, K. and Li, W. K. (1998). A stochastic volatility model with Markov switching, *Journal of Business & Economic Statistics* **16**(2): 244–253.

Steel, M. F. (1998). Bayesian analysis of stochastic volatility models with flexible tails, *Econometric Reviews* **17**(2): 109–143.

Tsay, R. S. (2005). *Analysis of Financial Time Series*, Vol. 543, John Wiley & Sons.

Vetzal, K. R. (1997). Stochastic volatility, movements in short term interest rates, and bond option values, *Journal of Banking & Finance* **21**(2): 169–196.

Watanabe, T. (1999). A non-linear filtering approach to stochastic volatility models with an application to daily stock returns, *Journal of Applied Econometrics* **14**(2): 101–121.

Wiggins, J. B. (1987). Option values under stochastic volatility: Theory and empirical estimates, *Journal of Financial Economics* **19**(2): 351–372.

Wilcox, B. A. and Hamano, F. (2017). Kalmans expanding influence in the econometrics discipline, *IFAC-PapersOnLine* **50**(1): 637–644.

Wills, A. and Ninness, B. (2008). On gradient-based search for multivariable system estimates, *IEEE Transactions on Automatic Control* **53**(1): 298–306.

Zadrozny, P. A. (1989). Analytic derivatives for estimation of linear dynamic models, *Computers & Mathematics with Applications* **18**(6-7): 539–553.

Appendix A

Autocorrelation Plots

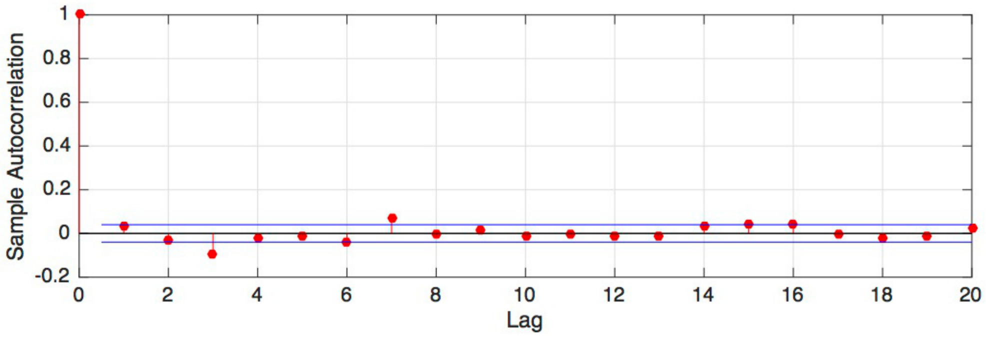

Figure A.1: *Plot of Sample Autocorrelation of Anglo American*

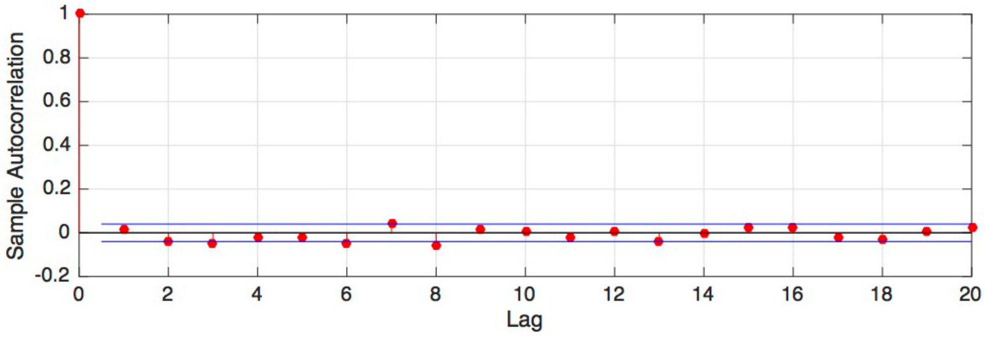

Figure A.2: *Plot of Sample Autocorrelation of BHP Billiton*

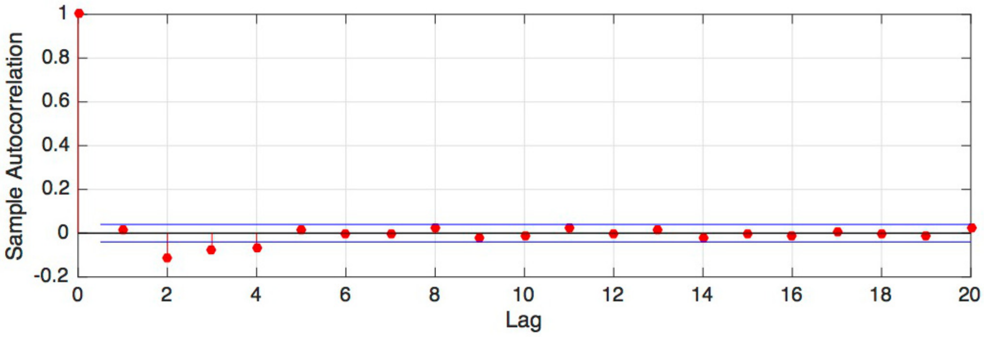

Figure A.3: *Plot of Sample Autocorrelation of FirstRand*

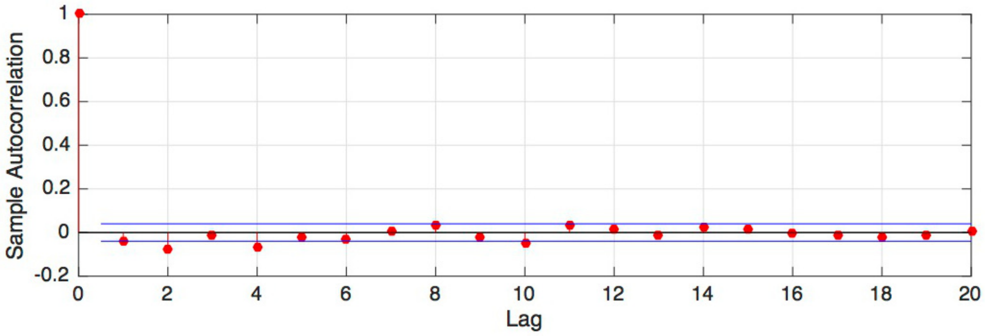

Figure A.4: *Plot of Sample Autocorrelation of Standard Bank Group*

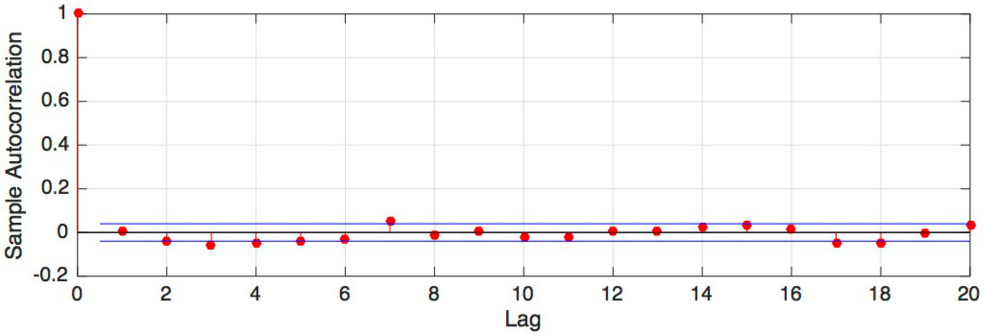

Figure A.5: *Plot of Sample Autocorrelation of the JSE TOP40 Index*

Appendix B

MATLAB Code for Kalman Filter

```
function [neg_LLF,hatX,hatDP] = KF(matrices,initials_filter,measurements)
       [F,G,Q,H,R] = deal(matrices{:});         % get system matrices
             [X,P] = deal(initials_filter{:});   % initials for the filter

            [m,n]  = size(H);                  % dimensions
           N_total = size(measurements,2);     % number of measurements
             hatX  = zeros(n,N_total+1);       % prelocate for efficiency
             hatDP = zeros(n,N_total+1);       % prelocate for efficiency
           neg_LLF = 1/2*m*log(2*pi)*N_total; % set initial value for the neg Log LF
  hatX(:,1)  = X; hatDP(:,1) = diag(P);     % save initials at the first entry
  for k = 1:N_total
             [X,P]        = kf_predict(X,P,F,G,Q);
             [X,P,ek,Rek] = kf_update(X,P,measurements(:,k),H,R);

             neg_LLF = neg_LLF+1/2*log(det(Rek))+1/2*ek'/Rek*ek;
             hatX(:,k+1)  = X; hatDP(:,k+1) = diag(P);
  end;
end
```

Time update: a priori estimates

```
function [X,P] = kf_predict(X,P,F,G,Q)
    X = F*X;                   % Predicted state estimate
    P = F*P*F' + G*Q*G';       % Predicted error covariance
end
```

Measurement update: a posteriori estimates

```
function [X,P,residual,cov_residual] = kf_update(X,P,z,H,R)
    residual     = z - H*X;            % residual
    cov_residual = R + H*P*H';         % residual covariance matrix
    Kalman_gain  = P*H'/cov_residual;  % Gain matrix

    X = X + Kalman_gain*residual;      % Filtered state estimate
    P = P-Kalman_gain*H*P;             % Filtered error covariance
end
```