THE ESTIMATION OF DAILY VOLATILITY USING HIGH FREQUENCY DATA IN THE SOUTH AFRICAN EQUITY MARKET

I.M. PAGEL

Dissertation submitted in partial fulfilment of the requirements for the degree Magister Scientiae in Risk Analysis at the North-West University

Supervisor: Prof P.J. de Jongh

2005

Potchefstroom

PREFACE AND ACKNOWLEDGEMENTS

We would like to thank:

- Prof Riaan de Jongh for his guidance
- Ettiene van Wyk for his assistance with the Matlab code used for the data preparation as well as the calculation of our various volatility estimators
- Willemien Malan and HSBC for providing us with high frequency equity data for the study
- Dr Suria Ellis for verifying our ABDE and SSR variance estimator calculations
- Chris Atkinson for assistance with the pricing of the options
- Dr Eben Mare for his suggestions
- My family for their love and support

All remaining mistakes and omissions remain our own.

ABSTRACT

Financial market volatility is central to the theory and practice of asset pricing, option pricing, asset allocation, portfolio selection, portfolio rebalancing and hedging strategies as well as various risk management applications. Most textbooks assume volatility to be constant; however in practice this is a very dangerous assumption to make and has lead to a research program regarding the distributional and dynamic properties of financial markets. Given that financial markets display high speeds of adjustment, studies based upon daily observations may fail to capture information contained in intraday or high frequency market movements and until relatively recently the use of daily or equally spaced data was considered the highest meaningful sampling frequency for financial market data.

Recently, the volatility modelling literature took a significant step forward. Andersen et al. (2001) proposed a new approach called 'realized' volatility that exploits the information in high frequency returns. Basically, the approach is to estimate daily volatility by taking the square root of the sum of the squared intraday returns which are sampled at very short intervals. We discuss several theoretical measures for volatility of which quadratic variation (QV), integrated variance (IV) and conditional variance (CV) are the most popular. Realized variance is a consistent estimator for QV and can approximate IV and CV under various conditions. GARCH models are only concerned with estimating CV. Furthermore we will discuss the NIG-GARCH model proposed by Venter and de Jongh (2002, 2004) and show how these models may be fitted to daily data to provide estimates and forecasts for daily volatility. We will also investigate three realized volatility methods which use intraday data to estimate dialy volatility namely; $\sqrt{SSR}$, $\sqrt{ABDE}$ and $\sqrt{VARHAC}$. Here SSR is the sum of the squared intraday returns, ABDE is the realized variance estimator proposed by Andersen, Bollerslev, Diebold and Ebens (2001) and VARHAC is the Vector Autoregressive Heteroscedastic and Autocorrelation Consistent estimator introduced by Den Haan and Levin (1996).

Our study will be conducted along the lines of the study by Andersen et al. (2001), utilising South African equity data. However, as far as the scaled returns are concerned we will investigate alternative volatility estimators for daily return volatilities. In particular, the VARHAC estimator (see e.g., Bollen and Inder (2002)) will be compared with the ABDE estimator (see e.g., Andersen et al. (2001)) as well as with daily volatility estimators based on the NIG-GARCH model proposed by Venter and de Jongh (2002, 2004). To the best of our knowledge a similar study has not yet been performed in the South African equity market context.

Consistent with the Andersen et al. (2001) study we found the unconditional distributions of the realized variances to be positively skewed (right skewed), while the realized logarithmic standard deviations are approximately Gaussian, as are the distributions of the returns scaled by realized standard deviations or realized volatilities. This conclusion holds for all the realized variance estimators (ABDE, SSR and VARHAC). The distribution of the returns scaled by the GARCH and NIG-GARCH volatility estimates are also close to normality. However, the distributions of the daily variance estimates and the log of the volatility (standard deviations) estimates based on the GARCH and NIG-GARCH models are clearly not normally distributed. We therefore conclude that our findings in the South African market context do agree with the claims made by Andersen et al. (2001) for the realized volatility estimation techniques and that the realized volatility techniques are a much better match to the criteria as set out by Andersen et al. (2001) than the GARCH and NIG-GARCH volatility estimates.

SINOPSES

Volatiliteit in die finansiële mark is sentraal tot die teorie en die praktyk van bateprysing, opsieprysing, bate-allokasie, portefeulje seleksie, portefeulje herbalansering en verskansingstrategieë sowel as verskeie riskobeheer toepassings. Meeste handleidings veronderstel dat volatiliteit konstant is; maar in die praktyk is dit nie 'n geldige aanname nie en het dit ondermeer gelei tot 'n nuwe navorsingsveld wat die verdeling en dinamiese eienskappe van finansiële markte ondersoek. Gegewe dat die finansiële mark vinnig aanpas by veranderende omstandighede, sal studies wat gebaseer word op daaglikse waarnemings misleidend wees, aangesien hoë frekwensie mark bewegings nie weerspieël word nie. Daaglikse data (einde van die dag sluitingspryse) was tot onlangs die mees betekenisvolle steekproef frekwensie data vir die finansiële mark.

Betekenisvolle vooruitgang in die veld van volatiliteit modellering is onlangs gemaak. Navorsing deur Andersen et al. (2001) beveel 'n nuwe benadering aan; naamlik "gerealiseerde volatiliteit" wat al die inligting in hoë frekwensie opbrengste (soos gebaseer op intradag prysdata) gebruik om daaglikse volatiliteit te beraam. Die gerealiseerde volatiliteitsbenadering skat daaglikse volatiliteit deur die vierkanstwortel van die som van die gekwadreerde intra-daaglikse obrengste oor kort intervalle te bereken. In hierdie studie word getoets of die bevindinge van Andersen et al. (2001) ook van toepassing gemaak kan word op Suid-Afrikaanse aandeel data. Andersen et al. (2001) bestudeer die SSR en ABDE beramers, wat beide op intradag data gebaseer is. SSR word gedefineer as die som van die gekwadreerde intra-daaglikse obrengste en is 'n beramer van daaglikse variansie, ABDE is die gerealiseerde variansie beramer soos bekend gestel deur Andersen, Bollerslev, Diebold and Ebens (2001). Anders as Andersen et al. (2001) sal ons ook ander beramers vir daaglikse variansie bestudeer, naamlik die VARHAC beramer van Bollen en Inder (2002), wat op intra-dag data gebaseer is, en die GARCH en NIG-GARCH model van Venter en de Jongh (2002, 2004), wat op daaglikse data gebaseer is VARHAC is die Vector Autoregressive Heteroscedastic en Autocorrelation Consistent beramer soos bekend gestel deur Den Haan and Levin (1996). Tot die beste van ons wete is 'n soortgelyke studie nog nie op die Suid–Afrikaanse effektebeurs aangepak nie. Verskeie teoretiese maatstawwe vir daaglikse variansie word in die tesis bespreek, soos kwadratiese variansie (QV), geïntegreerde variansie (IV) en voorwaardelike variansie (CV). Gerealiseerde variansie (SSR) is 'n konsekwente beramer vir QV en kan IV en CV benader onder sekere omstandighede. Daarenteen word GARCH modelle gebruik om voorwaardelike variansie (CV) te beraam.

In lyn met Andersen et al. (2001) se bevindinge het ons die volgende gevind. Die onvoorwaardelike verdelings van gerealiseerde volatiliteit soos beraam deur die vierkantswortels van SSR, ABDE en VARHAC is positief skeef. Die gerealiseerde logaritmiese standaardafwykings en die verdeling van die opbrengste wat gestandaardiseer is deur die gerealiseerde standaard afwykings is min of meer normaal verdeel. Hierdie gevolgtrekking geld vir al die gerealiseerde variansieberamers (ABDE, SSR en VARHAC). Ook het ons bevind dat die verdeling van die opbrengste gestandaardiseer deur die GARCH en NIG-GARCH variansieberamers naby normal verdeel is. Die verdelings van die daaglikse variansie is nie normaal verdeel vir enige beramer nie, maar die verdeling van die logaritme van die volatiliteite (standaard afwykings) soos beraam deur die gerealiseerde variansie beramers is wel normal verdeel, terwyl dit nie die geval is vir die GARCH en NIG-GARCH beramers nie. Ons maak dus die gevolgtrekking dat ons resultate en bevindings soos gebaseer op Suid–Afrikaanse effektebeurs data, met Andersen et al. (2001) se afleidings saamstem.

TABLE OF FIGURES

TABLE OF FIGURES (continued)

TABLE OF TABLES

TABLE OF TABLES (continued)

TABLE OF TABLES (continued)

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Financial market volatility is central to the theory and practice of asset pricing, option pricing, asset allocation, portfolio selection, portfolio rebalancing and hedging strategies as well as various risk management applications. It is widely recognised that, even although most textbooks assume volatilities to be constant, they vary over time. For example, in financial return series, the occurrence of volatility clustering and the non-normal distributional characteristics are well-known stylized facts. This recognition has spurred research into the distributional and dynamic properties of stock market volatility. Most of what we have learned from this literature is based on the estimation of parametric Generalised Autoregressive Conditional Heteroskedastic (GARCH) models, stochastic volatility models for the underlying returns, realized volatility models or the analysis of implied volatilities from options or other derivative prices. However, the validity of such volatility measures generally depends upon specific distributional assumptions, and in the case of implied volatilities, further assumptions concerning the market price of volatility. In contrast the realized volatility approach is non-parametric in nature.

Financial return volatility data is influenced by time dependent information flows which result in pronounced temporal volatility clustering. These time series can be parameterised using GARCH models. The GARCH model generalizes the autoregressive ARCH model to an autoregressive moving average model (see e.g., Engle (2003)) and it has been found by numerous authors that GARCH models can provide good in-sample parameter estimates and, when the appropriate volatility measure is used, reliable out-of-sample volatility forecasts (see e.g., Andersen and Bollersev (1998), Barndorff-Nielsen and Shephard (2001)). Realized volatility estimators have been introduced recently as an alternative method for estimating daily volatility. Realized volatility is a methodology that exploits all available information in high frequency returns (see e.g., Andersen et al. (2001), Ghysels et al. (2003), Giot and Laurent (2004) and Oomen (2001)). Basically, the approach is to estimate daily volatility by taking the square root of the sum of the squared intraday returns which are sampled at very short intervals. In ideal circumstances increasing the sampling frequency yields arbitrarily precise estimates of volatility on any given day. Therefore, daily volatility becomes almost observable via realized volatility. Up and until the introduction of realized volatility the two main approaches towards modelling volatility in financial time series were stochastic volatility models and GARCH models of which the latter is the most popular.

As the value of the numerous volatility estimation techniques available to the researcher to estimate and forecast daily volatility becomes generally recognized, research attention has shifted towards the potential gains that may be obtained from using intraday or high frequency data as an information source for the estimation and forecasting of daily volatility. Given that financial markets display high speeds of adjustment, studies based upon daily observations may fail to capture information contained in intraday or high frequency market movements and until relatively recently the use of daily or equally spaced data was considered the highest meaningful sampling frequency for financial market data. There are many reasons for the choice of a lower or daily (end of day close to close) sampling frequency: to collect, collate, store, retrieve, and manipulate high frequency data is still rather costly and time consuming and the observations are subject to a wide range of factors such as intra-day seasonal effects, measurement errors, data gaps etc. A lower frequency sampling pattern was perhaps dictated by the general view that, whatever drove share prices and returns, probably did not vary significantly over short time intervals and therefore it was deemed unnecessary to investigate shorter time intervals.

Hence, it is natural to ask what the impact of a higher sampling frequency such as 5 minute, 10 minute etc. would be and if such a finer sampling frequency would contain more information about share prices and volatility. A further contentious issue with regard to high frequency data is the selection of the high frequency sampling interval; namely 1 minute, 5 minute etc. When a very high sampling frequency (e.g., 1 minute or more frequent intraday returns) is chosen it may lead to the introduction of a bias in the realized volatility estimate of daily volatility. This is usually the result of market microstructure effects such as bid-ask bounces, price discreteness or non-synchronous trading (see e.g., Zhou (1996), Andreo and Ghysels (2002), and Oomen (2001)). These factors can distort the quality of the constructed realized volatility as a proxy for true daily volatility. Hence there is a trade-off between bias and variance when choosing the sampling frequency and this is the reason that returns are typically sampled at moderate frequency, such as 5 minute sampling intervals.

An alternative way to handle the bias problem is to use bias correction techniques. For example a moving average filter was used by Andersen et al. (2001) and Bandi and Russell (2003 a&b) and an autoregressive filter by Bollen and Inder (2002). Research is being conducted to determine the 'optimal' sampling frequency to use. At the moment, sampling frequencies that range from 5 minute to 30 minute intraday returns are popular for liquid shares. Andersen et al. (2001) found the 5 minute frequency to be the highest at which the properties of the return series are not seriously distorted by irregular quoting and the discreteness of prices. Oomen (2001) found the optimal sampling frequency for the data set to be 25 minute returns. Giot and Laurent (2004) found a sampling frequency of approximately 15 minutes to be optimal for the CAC 40 and S&P 500 stock indices and a sampling frequency of 1 hour for their study utilising the YEN-USD and DEM-USD exchange rates).

Not only are the actual price quotations important in the understanding of the structure of financial markets, but there is also additional information in the duration or time interval between quotations. A fundamental property of high frequency data is that observations can occur at non equidistant time intervals. Share trades are not equally spaced throughout the day, resulting in intraday 'seasonals' in the volume of trade (see e.g., Engle (2000)), the volatility of prices and the behaviour of spreads. During some of the time intervals, no transactions may occur, dictating that even measuring returns may be problematic. These difficulties are less pronounced when fixed daily data is used but becomes more important when high frequency data is analyzed.

In Chapter 2 of our study we will investigate and discuss various methodologies to estimate daily volatility. We will focus our attention on GARCH models and describe in detail how these models may be used to estimate volatility. Specifically the focus will be on the AR(1)-GARCH(1,1) model assuming Normal Inverse Gaussian (NIG) innovations (see e.g., Barndorff-Nielsen and Prause (2001), Lillestol (2000) or Venter and de Jongh (2002, 2004)). We continue by discussing the concept of realized volatility and the asymptotic distribution theory of the realized volatility estimator.

Practical problems with regards to the implementation of the realized volatility estimator and in particular the bias problem induced by market microstructure noise will be discussed. In order to correct for the bias problem several procedures have been suggested in the literature. We will focus our attention on two of these approaches, viz the estimators proposed by Andersen et al. (2001) and by Bollen and Inder (2002). The realized volatility estimator of Andersen et al. (2001) will be referred to as the ABDE (after Andersen, Bollerslev, Diebold and Ebens) estimator and the realized volatility estimator proposed by Bollen and Inder (2002) will be referred to as the VARHAC estimator (Vector Autoregressive Heteroskedastic and Autocorrelation Consistent estimator).

Our study will be similar to the Andersen et al. (2001) study utilising JSE Securities Exchange SA share data. However, as far as the scaled returns are concerned we will investigate alternative volatility estimators for daily return volatilities. Andersen et al. (2001)

examined realised daily equity return volatilities obtained from high-frequency intraday transaction prices on individual stocks in the Dow Jones Industrial Average. They found that the unconditional distributions of the realised variances are highly right skewed, while the realised logarithmic standard deviations are approximately Gaussian, as are the distributions of the returns scaled by realised standard deviations. While a number of studies have examined the characteristics of intraday return volatility in the International markets, to the best of our knowledge no previous research has been conducted in the South African share market with regards to high frequency return volatility of individual shares.

We proceed with the Bollen and Inder (2002) study that proposed a new approach to the estimation of realized volatility in financial markets. The authors' estimated daily volatility by utilizing all available transactions (high frequency) data on a specific trading day by paying close attention to the resulting market microstructure effects (VARHAC estimator). As previously discussed, although the intraday seasonality in high frequency data is of concern this was not addressed in the construction of the Andersen et al. (2001) and Bollen and Inder (2002) variance estimates respectively and therefore we did not address intraday seasonality in this thesis. In particular, the VARHAC estimator (see e.g., Bollen and Inder (2002)) will be compared with the ABDE and SSR estimator as well as with daily volatility estimators based on the NIG-GARCH model proposed by Venter and de Jongh (2002, 2004). We conclude Chapter 2 with a discussion of the other estimators Bollen and Inder (2002) studied, viz. the simple volatility estimator, the PARK volatility estimator and the GK volatility estimator.

Our ability to analyse the working of financial markets in a high frequency environment is limited by the availability of high frequency data of a high quality and obtaining such high frequency data in the South African share market context was an initial hurdle to our study.
In Chapter 3 we discuss the many issues pertaining to the filtering of high frequency data (erroneous data points, corporate actions etc.), the shares selected for our study and the construction of the necessary time series in more detail. In order to filter the data of erroneous data points we first had to identify these errors as well as their causes (e.g., human capturing errors and the nature of trading on the exchange). Furthermore it was necessary to identify market characteristics that influence volatility such as corporate events, the arrival of news, introduction of a new share settlement system and non-economic world events such as the 11 September 2001 terrorist attacks. The ultimate aim of this chapter was to identify outliers or suspicious data points and then to use our knowledge of the market events that influence volatility to classify the suspicious data points as erroneous or not.

In Chapter 4 we will conduct an empirical study into the behaviour of the estimators of daily volatility. Before we can start with the empirical study of our various volatility estimators it is important to ensure that the data has been filtered of all erroneous data points and that the estimates of volatility are true reflections of actual market movements. We investigate and discuss the data outliers for the ABDE, SSR and VARHAC variance estimates. We describe the methodology we followed and we continue with a general discussion on additional market events in the South African equity market context that had an impact on the variance estimates.

In analogy with the Andersen et al. (2001) study we found the unconditional distributions of the realized variances to be positively skewed (right skewed), while the realized logarithmic standard deviations are approximately Gaussian, as are the distributions of the returns scaled by realized standard deviations or realized volatilities. This conclusion holds for all the realized variance estimators studied namely ABDE, SSR and VARHAC. The distribution of the returns scaled by the GARCH and NIG-GARCH volatility estimates are also close to normality. However, the distributions of the daily variance estimates and the log of the volatility (standard deviations) estimates based on the GARCH and NIG-GARCH models are not normally distributed. We therefore conclude that our findings in the South African market context do agree with the claims made by Andersen et al. (2001) for the realized volatility estimation techniques and that the realized volatility techniques are a much better match to the criteria as set out by Andersen et al. (2001) than the GARCH and NIG-GARCH volatility estimates.

In Chapter 5 we summarize the findings of our study and we conclude the thesis with suggestions for further research.

CHAPTER 2

OVERVIEW OF THE ESTIMATION OF DAILY AND REALIZED VOLATILITY

2.1 INTRODUCTION

As stated in Chapter 1 the focus of this thesis is on estimating daily volatility with particular emphasis on using realized volatility-based estimators. Realized volatility estimators have been introduced recently as an alternative method for estimating daily volatility. Before the introduction of realized volatility estimators the two main approaches towards modelling volatility in financial time series were stochastic volatility models and GARCH models of which the latter is the most popular.

In Section 2.2, we will focus our attention on GARCH models and describe in detail how these models may be used to estimate and forecast volatility. Specifically the focus will be on the AR(1)-GARCH(1,1) model assuming Normal Inverse Gaussian (NIG) innovations, see e.g., Venter and de Jongh (2004). The concept of realized volatility, practical problems with regards to the implementation thereof, the bias problem induced by market microstructure noise and the asymptotic distribution theory of the realized volatility estimator will be discussed in Section 2.3. We will focus our attention on two of these approaches namely, the ABDE (after Andersen, Bollerslev, Diebold and Ebens) realized variance estimator (of Andersen et al. (2001)) and the VARHAC (Vector Autoregressive Heteroscedastic Autocorrelation Consistent) variance estimator (Bollen and Inder (2002)). In Section 2.4 we will discuss some of the other volatility estimators proposed by Bollen and Inder (2002), namely the simple volatility estimator, the PARK variance estimator and the GK variance estimator. We conclude the chapter in Section 2.5.

2.2 GARCH MODELS

2.2.1 Introduction and motivation

Traditionally daily volatility is estimated by utilizing end of day closing prices in order to compute the daily return time series. In this setting, the intraday or high frequency price movements are not used in the construction of a daily volatility estimator. One well-known example of this methodology is the ARCH model of Engle (1982) and subsequent ARCH type models such as the GARCH (Generalised Autoregressive Conditional Heteroscedastic) models (see e.g., Bollerslev 1986). The ability of ARCH models to provide reliable estimates of equity daily return volatility is well documented (see e.g., Blair et al. (2001), Nelson (1992), Ederington and Lee (2001) and Nelson and Foster (1995)). More recently, research has focused on inter temporal dependence models to explain the empirical observation of volatility clustering. The latter characteristic is a natural application for the ARCH model and Bollerslev's generalized ARCH model (GARCH). In their unpublished manuscript Rhaman et al. (2000) show that GARCH models have been applied to a wide variety of financial market instruments such as stock indices and individual equities. These research papers found that the GARCH processes showed a better statistical fit to the time series than traditional ARCH models. GARCH imposes an autoregressive structure on the conditional variance, allowing volatility shocks to persist over time. This persistence captures the propensity of returns of like magnitude to cluster in time and can explain the well-documented non-normality and non-stability of empirical asset return distributions. Other variants of the original GARCH model have been developed, with many of these models showing statistical advantages over the original ARCH model when applied to a particular time series. A vast assortment of different ARCH and GARCH models are available to analyze statistical data and when choosing one of the many GARCH models available today it is important to select a model that is most applicable and statistically reliable when modelling and forecasting daily volatility. Bollerslev and Wright (2001) state that the estimation of daily GARCH models is one of the most popular approaches to volatility forecasting available to the researcher. Mandlebroth (1963), Fama (1965) as well as Bollerslev et al. (1992) indicate that the rates of return implicit in the time series of share prices are time dependent. The evidence shows that leptokurtosis, skewness and volatility clustering characterise the distribution of daily share returns. Several recent studies provide evidence that the GARCH methodology is capable of capturing these characteristics.

Stochastic volatility models present another approach to the modelling of volatility. These use continuous time diffusion type processes specified via stochastic differential equations. In practice, however, financial series are mostly observed in discrete time. Ignoring this difference between the time specifications of model and observations when fitting diffusion stochastic volatility models can result in inconsistent estimators. Model fitting when following the stochastic volatility approach is a matter of ongoing research, which does not seem to have yet culminated in routinely available and recommendable methodology. A good recent entry into this literature is provided by Ait-Sahalia (2002). On the other hand, the GARCH approach proceeds from the discrete time nature of observed time series but makes the simplifying assumption that current volatility is at most a function of past data. This makes it relatively easy to determine the likelihood function when fitting GARCH models to observed time series so that the standard inference tools of maximum likelihood estimation become available. However, the underlying innovation density must be specified to make this possible. Since there is no universally accepted choice of this density, the GARCH approach also has open issues of its own to which much research has been devoted. Progress on resolving these issues is urgently needed since routine application of GARCH models in the financial industry is becoming increasingly important, stimulated by findings such as those of Berkowitz and O'Brien (2002), namely that in banking, profit and loss modelling and VaR estimation, using GARCH models permits comparable risk coverage while requiring less regulatory capital than when using structural bank VaR models. In this thesis we will estimate GARCH models assuming a Normal Inverse Gaussian (NIG) innovation density. This approach, described in detail by Venter and de Jongh (2004) will be discussed next.

2.2.2 Modelling volatility using GARCH models

Formally, let $Y_1, Y_2, \ldots, Y_T$ denote a time series and we describe it in terms of a GARCH model of the form

$$Y_t = \mu_t + \sqrt{h_t} Z_t \quad \text{for } t = 1, 2, \ldots, T, \tag{2.1}$$

where μ_t represents an expected (or structural) component, $\sqrt{h_t}$ is the volatility and Z_t the innovation at time t. In our context $Y_1, Y_2, \ldots, Y_T$ will be a series of log returns, i.e. $Y_t = \ln(\frac{P_t}{P_{t-1}}) = \ln(P_t) - \ln(P_{t-1})$ where P_t denotes the price of the share at time t. It is assumed that μ_t and h_t are at most functions of the past observations $Y_1, Y_2, \ldots, Y_{t-1}$ and these functions may involve a number of parameters. (By fitting the GARCH model to a financial series of daily returns an estimate for h_t, the daily conditional variance, may be obtained. This will be illustrated in detail in the next section.) Further, $Z_1, Z_2, \ldots, Z_T$ are assumed independent random variables with common density function (the "innovation density") g which is a unit density in the sense that its expectation is zero and its variance is one. This restriction is necessary to ensure identifiability of the parameters in the model. For example, note that, conditional on $Y_1, Y_2, \ldots, Y_{t-1}$ we have that $Var(\sqrt{h_t} Z_t) = h_t$. The innovation density may possibly depend on a number of further parameters also and we represent all the parameters needed in the model by a vector $\boldsymbol{\theta}$. The log of the likelihood function may then be written as

$$l_g(\boldsymbol{\theta}) = \sum_{t=1}^{T} [\log\{g((Y_t - \mu_t)/\sqrt{h_t})\} - \tfrac{1}{2}\log(h_t)], \tag{2.2}$$

and the maximum likelihood estimate of $\boldsymbol{\theta}$ is found by maximizing (2.2) over $\boldsymbol{\theta}$. To do this we must make an explicit choice for the innovation density g. In much of the GARCH literature the unit normal density $g = \varphi$ is chosen, but there is now abundant empirical evidence that this normality assumption is often violated in practical financial contexts especially when dealing with high frequency (e.g., daily) series. Figure 2.1 provides an illustrative example. The tick by tick log return series for Anglo American PLC (AGL) on 11 Sep 2001, the day of the terrorist attack on the World Trade Centre, is depicted in the graph. Inspection of the graph reveals normal-like activity until news of the attack became known. The returns start to vary significantly, giving rise to, especially, large negative returns. If we would want to model this process with a GARCH model, a heavy tailed, negatively skewed innovation density model is appropriate. Unless we are sure that the true innovation density is normal, estimating the parameters of a GARCH model by maximizing $l_{\varphi}(\boldsymbol{\theta})$ cannot be described as true maximum likelihood and is referred to as "pseudo" maximum likelihood estimation (PMLE). We will refer to this as normal MLE.

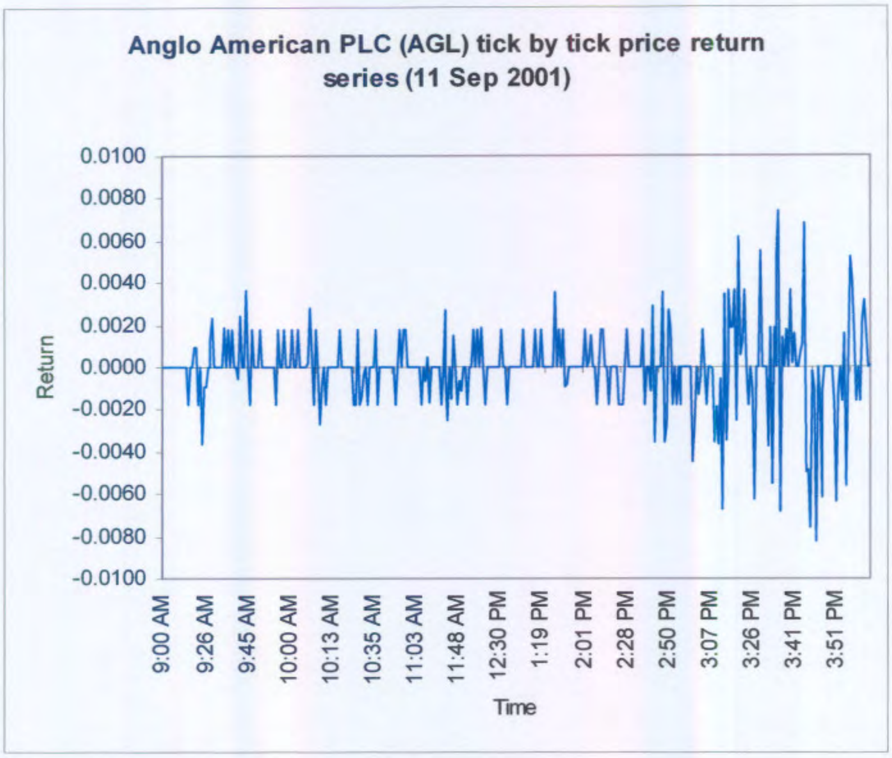


Figure 2.1: Anglo American PLC (AGL) tick by tick price return series on 11 Sep 2001.

Engle and Gonzalez-Rivera (1991) studied the loss of estimation efficiency inherent in normal MLE and found that it may be severe if the true innovation density is heavy tailed and especially so if it is skewed. Bollerslev and Wooldridge (1992) and Gourieroux (1997) provide conditions under which normal MLE yields consistent and asymptotically normally distributed estimates; they also give expressions for asymptotic variances and covariances of these estimates, valid even if the true innovation density is not normal. These results make the use of normal MLE practical but the issue of efficiency of the estimators remained open. The Normal Inverse Gaussian (NIG) distribution has already been used in financial contexts with considerable success by a number of authors (see e.g., Barndorff-Nielsen and Prause (2001), Lillestol (2000) and Venter and de Jongh (2002)). De Jongh and Venter (2004) compare the NIG distribution with a variety of other distributions (e.g., the t, skewed t, and normal distributions). Their main finding is that the NIG based approach is competitive with the other methods and they propose the NIG as the choice for the innovation density g . The NIG distribution is discussed in the next section.

2.2.3 The NIG-distribution

The NIG-distribution can be parametrized in many ways. The most common specification is the $(\alpha, \beta, \mu, \delta)$ parameters. Its parameter space is

$$0 < |\beta| < \alpha \ , \ -\infty < \mu < \infty \ , \ \delta > 0 \tag{2.3}$$

and in terms of $\zeta = \delta\sqrt{\alpha^2 - \beta^2}$ and the function $q(x) = \sqrt{1+x^2}$ the NIG density is

$$g_{NIG(\alpha,\beta,\mu,\delta)}(z) = \frac{\alpha \exp(\varsigma + \beta(z-\mu))K_1\left(\alpha\delta q\left((z-\mu)/\delta\right)\right)}{\pi q\left((z-\mu)/\delta\right)}, \tag{2.4}$$

where K_1 is the modified Bessel function of the third kind and index 1. Here μ and δ are location and scale parameters respectively while α and β are parameters specifying the shape of the distribution. In particular $\beta = 0$ corresponds to a symmetric distribution (see Lillestol (2000) or the cited references for more details). Barndorff-Nielsen et al. (1985) also introduced the tail-heaviness (or peakedness) and asymmetry parameters ξ and χ given by

$$\xi = \left(1 + \delta\sqrt{\alpha^2 - \beta^2}\right)^{-1/2} = \left(1+\zeta\right)^{-1/2} \text{ and } \chi = \beta\xi/\alpha = \rho\xi \ . \tag{2.5}$$

The parameters ξ and χ are scale and translation invariant and they may be used as shape parameters instead of α and β. Their domain is the so-called NIG shape triangle

$$0 \le |\chi| < \xi < 1 \ . \tag{2.6}$$

For $\chi < 0$ we get negatively skewed distributions, for $\chi = 0$ symmetric and for $\chi > 0$ positively skewed distributions. The parameter ξ controls the tail thickness of the distributions, with ξ close to 0 yielding normal-like tails and ξ close to 1 yielding heavier tails depending on what we do with the other parameters. To serve as the innovation distribution in GARCH models, we want the mean to be 0 and the variance to be 1. In general the mean and variance of the NIG-distribution is given by

$$\kappa_1 = \mu + \delta\rho\big/\sqrt{1-\rho^2} \text{ and } \kappa_2 = \delta^2\big/\varsigma(1-\rho^2) \text{ with } \rho = \beta/\alpha \ . \tag{2.7}$$

These equations together with (2.6) may be solved to express $(\alpha, \beta, \mu, \delta)$ in terms of $(\xi, \chi, \kappa_1, \kappa_2)$ so that we may parameterize the NIG-distribution in terms of $(\xi, \chi, \kappa_1, \kappa_2)$ rather than the original $(\alpha, \beta, \mu, \delta)$. In this new parameter set $(\xi, \chi, \kappa_1, \kappa_2)$ we now take

$\kappa_1 = 0$ and $\kappa_2 = 1$ to get the unit form of the NIG-distribution which depends only on the shape parameters (ξ, χ) and is skew when $\chi \neq 0$. We will denote it by $SUNIG(\xi, \chi)$.

2.2.4 Estimating volatility using an AR(1)-GARCH(1,1) model

In this thesis we consider only the popular AR(1)-GARCH(1,1) model (see e.g., Hansen and Lunde (2004a), Olsen (1999)) which has the form of (2.1) with

$$\mu_t = \nu + \phi Y_{t-1} \text{ and } h_t = \alpha_0 + (\alpha_1 Z_{t-1}^2 + \beta)h_{t-1}. \tag{2.8}$$

Here ν is the process mean, ϕ the AR parameter, α_0, α_1 the two ARCH parameters and β the GARCH parameter.

Under this model the log likelihood is:

$$\begin{aligned} l_g(\boldsymbol{\theta}) &= \sum_{t=1}^{T} [\log\{g((Y_t - \mu_t)/\sqrt{h_t})\} - \tfrac{1}{2}\log(h_t)] = \\ l_{SUNIG(\xi,\chi)}(\nu, \phi, \alpha_0, \alpha_1, \beta) &= \sum_{t=1}^{T} [\log\{g((Y_t - \nu - \phi Y_{t-1})/\sqrt{h_t})\} - \tfrac{1}{2}\log(h_t)] \end{aligned} \tag{2.9}$$

where $h_t = \alpha_0 + (\alpha_1 Z_{t-1}^2 + \beta)h_{t-1}$ and $g = SUNIG(\xi, \chi)$

The MLE estimates $\left(\hat{\nu}, \hat{\phi}, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}, \hat{\xi}, \hat{\chi}\right)$ are then obtained by maximising (2.9). In this way estimates for the structural part $\hat{\mu}_t = \hat{\nu} + \hat{\phi} Y_{t-1}$ and variance $\hat{h}_t = \hat{\alpha}_0 + (\hat{\alpha}_1 \hat{Z}_{t-1}^2 + \hat{\beta})\hat{h}_{t-1}$ (where $\hat{Z}_t$ denotes the estimated residual $(Y_t - \hat{\nu} - \hat{\phi} Y_{t-1})/\sqrt{\hat{h}_t}$ at time t) may be obtained. In order to estimate conditional daily volatility of the returns of a particular share, the closing prices on a particular day are used to construct a log return series on which the particular GARCH model is then fitted.

In Figure 2.2 below we give the price and corresponding log return series of Anglo American PLC (AGL) (one of the shares we selected for our study). The estimated variances resulting from an AR(1)-GARCH(1,1) fit assuming NIG innovations, are depicted graphically.

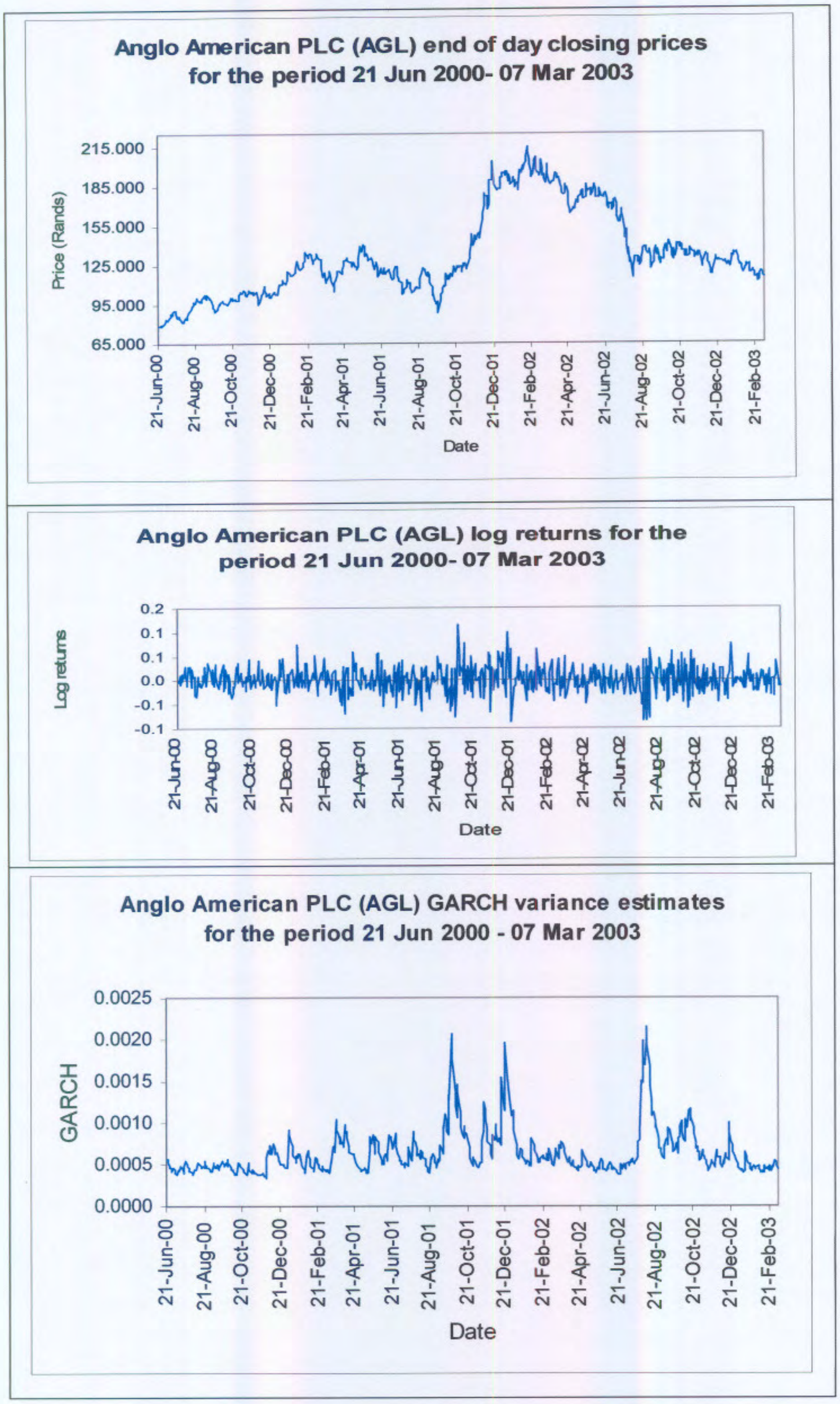


Figure 2.2: Anglo American PLC (AGL) end of day closing prices, log return and GARCH variance estimator graphs for the period 21 Jun 2000 – 07 Mar 2003.

In the NIG-GARCH variance estimate graph in Figure 2.2 there are three distinct periods of volatility clustering. The first being around the 11 Sep 2001 terrorist attacks in the USA, followed by the futures close out in Dec 2001 in the South African market and lastly the well documented leaked draft of the proposed Mining Charter in Aug 2002. When comparing the graph containing the series of variance estimates with the corresponding price and log return series, the reason for the distinct volatility clusters becomes clear. In all three cases considerable movement (volatility) in the prices and returns are clear.

2.2.5 Forecasting volatility

Assuming that the GARCH model postulated in the previous section holds, the one period ahead forecast for volatility is $\sqrt{\hat{h}_{T+1}}$ where $\hat{h}_{T+1} = \hat{\alpha}_0 + (\hat{\alpha}_1 \hat{Z}_T^2 + \hat{\beta})\hat{h}_T$, and $\hat{Z}_t$ denotes the estimated residual $(Y_t - \hat{\nu} - \hat{\phi} Y_{t-1}) \big/ \sqrt{\hat{h}_t}$ at time t). Once a forecast for the volatility exists, it is easy to provide forecasts for other risk measures. For example, consider VaR (value at risk) and ESF (expected shortfall) as risk measures. If the true innovation density is g with distribution function G, then the true innovation VaR at probability level τ is given by $VaR_\tau = G^{-1}(\tau)$ and the corresponding expected shortfall by $Esf_\tau = \tau^{-1} \int_{-\infty}^{VaR_\tau} xg(x)dx$. The corresponding estimates are given by the same expressions with g and G replaced by their estimates $\hat{g}$ and $\hat{G}$.

Therefore, assuming NIG innovations, the one period ahead forecast for VaR due to innovation and volatility is $\sqrt{\hat{h}_{T+1}} G^{-1}_{\hat{\xi},\hat{\chi}}(\tau)$ and the one period ahead VaR forecast for the return generating process is $\hat{\nu} + \hat{\phi} Y_T + \sqrt{\hat{h}_{T+1}} G^{-1}_{\hat{\xi},\hat{\chi}}(\tau)$. The forecasts for ESF may be obtained in a similar fashion.

2.2.6 Estimation and forecasting issues

As seen in the previous section, when GARCH estimation techniques are implemented, "volatility forecasts" can be obtained as an extension of the current model. There are however practical issues when fitting GARCH models to financial time series. For instance one has to specify a parametric model and make certain distributional assumptions, which need not be correct. Therefore one has to carry out a proper "goodness of fit" test to assess the adequacy of the fit. GARCH models should also be fit to time series containing "enough" data points. If the time series does not have enough data points the Maximum Likelihood Estimates (MLE) obtained might be flawed. In practice it is assumed that at least 250 data points are required to ensure an adequate fit.

GARCH volatility estimates change every time a new fit is made. For example, suppose that the same GARCH model is fitted to exactly the same time series, with the only difference being that the one time series is lagged one period. One would want the historical volatility estimates to be the same at the same time point. However, this is not the case. The introduction of the new data point causes the historic volatility estimates to change. This is not a very desirable property of GARCH models, because a newly observed return should not influence past volatility.

2.3 REALIZED VOLATILITY MODELS

2.3.1 Introduction and motivation

In the previous section we described how GARCH models might be used to estimate and forecast conditional daily volatility. An estimate for the volatility of a particular share on a particular day may be obtained by fitting a GARCH model to a series of daily log returns obtained from closing price data. We also mentioned that until recently most of what was learnt regarding financial market volatility was based on parametric GARCH and stochastic volatility models for the underlying returns, or on the analysis of implied volatilities from options or other derivative instruments. However, the validity of such volatility measures generally depends upon specific distributional assumptions (e.g., assuming Normal Inverse Gaussian distributed innovations) and, in the case of implied volatilities, further assumptions concerning the market price of volatility risk. As such, the existence of multiple competing models immediately calls into question the robustness of previous findings. An alternative approach, based on intraday squared returns over the relevant return horizon, provides model-free unbiased estimates of the ex-post realized volatility. Recently, Andersen et al. (2001) investigated the realized volatility approach. Basically, the approach is to estimate daily volatility by taking the square root of the sum of the squared intraday returns, which are sampled at very short intervals. In ideal circumstances increasing the sampling frequency yields arbitrarily precise estimates of volatility on any given day. Therefore, daily volatility becomes almost observable via realized volatility. The above-mentioned authors also found that realized volatility behaves like a long memory process and that this feature may be used to construct potentially more accurate forecasts of daily volatility. Using currency data, Pong et al. (2004) compared the volatility forecasting ability of four methods, viz. GARCH forecasts obtained from daily returns, short memory ARMA and long memory ARFIMA forecasts from high frequency returns, and implied volatilities obtained from option prices. They found that the methods based on realized volatility provided the most accurate forecasts in general with the long memory ARFIMA forecasts performing best.

2.3.2 Realized volatility modelling

Daily realized volatility can be defined as the square root of the sum of the high frequency or intra day squared returns. Andersen et al. (2001) motivate the realized volatility approach by the limitations of traditional approaches which were mentioned previously. Naturally they argue that having more relevant data available should improve the efficiency and accuracy with which daily volatility may be estimated. The authors' approach to construct a realized volatility estimator follows similar lines than that of earlier work by French et al. (1987), Schwert (1990 a&b) and Schwert and Sequin (1991). These authors rely primarily on daily return observations which they use for the construction of monthly realized share return volatility. However, these earlier studies did not provide any formal justification for such measures, which Andersen et al. (2001) provides. This justification, summarised by Hansen and Lunde (2004b), is presented below.

Let $\{p(t)\}_{t\in I}$ be a logarithmic price process over a time interval I, and let $[a,b]\subset I$ be a compact interval that is partitioned into m intervals of equal length $\Delta_m \equiv (b-a)/m$. The interval $[a,b]$, will typically span a trading day, so we refer to $Y_{t,m} \equiv p(a+i\Delta_m)-p(a+i\Delta_m-\Delta_m),\ i=1,\dots,m$ as intraday returns. The realized variance at frequency m is defined as

$$RV_{[a,b]}^{(m)} \equiv \sum_{i=1}^{m} Y_{i,m}^2.$$

When $p(t)$ is a semi-martingale the RV is by definition a consistent estimator of the quadratic variation (QV), of $\{p(t)\}_{t\in[a,b]}$ (see e.g., Andersen and Bollerslev (1998) and Barndorff-Nielsen and Shephard (2002)). The stochastic volatility models define a particular class of semi-martingales. These satisfy a stochastic differential equation of the form

$$dp(t) = \mu(t)dt + \sigma(t)d\omega(t),$$

where $\mu(t)$ and $\sigma(t)$ are time varying random functions and $\omega(t)$ is standard Brownian motion. The integrated variance for such processes is defined by

$$IV_{[a,b]} \equiv \int_a^b \sigma^2(t)dt,$$

and equals QV for this class of semi-martingales. The IV is fundamental for pricing derivative securities, see e.g., Hull and White (1987), which makes it a natural population measure of volatility.

Another popular measure of volatility is the conditional variance (CV) that plays a pivotal role in ARCH-type models. The CV is defined by

$$CV_{[a,b]} \equiv \mathrm{var}(r_{[a,b]} | F_a),$$

where $r_{[a,b]} = p(b) - p(a)$ and F_a denotes the information set at time a. So $CV_{[a,b]}$ is the variance of the innovation in $p(t)$ over the interval $[a,b]$, conditional on the information at time a. The relations between the various quantities are the following: RV is generally consistent for QV, which (sometimes) equals the IV. Further, $E(IV|F_a) \approx CV$ with equality if $\{\mu(t)\}_{t=a}^{b}$ is F_a measurable. Thus the RV can be used to approximate these population measures under various assumptions. The empirical properties of the RV have been studied in various settings by Andersen et al. (2001) and Andersen et al. (2003). A theoretical comparison between the IV and RV (in relation to the IV) is established in Barndorff-Nielsen and Shephard (2002).

In a multivariate setting and using similar, but slightly different notation, Andersen et al. (2001) provide the following results. Assume that a logarithmic $N \times 1$ vector price process $\underline{p}_t$, follows a multivariate continuous time stochastic volatility diffusion,

$$d\underline{p}_t = \underline{\mu}_t dt + \mathbf{\Omega}_t^{1/2} d\underline{W}_t, \tag{2.10}$$

where $\underline{W}_t$ denotes a standard N dimensional Brownian motion, the process for the N x N positive definite diffusion matrices, $\mathbf{\Omega}_t$ is strictly stationary, and the time unit interval, or $h = 1$, is normalized to represent one trading day.

Conditional on the sample path realization of $\mathbf{\Omega}_t$ and $\underline{\mu}_t$, the distribution of the continuously compounded h–period returns, $\underline{r}_{t+h,h} \equiv \underline{p}_{t+h} - \underline{p}_t$, is then

$$\underline{r}_{t+h,h} | \sigma\{\underline{\mu}_{t+\tau}, \mathbf{\Omega}_{t+\tau}\}_{\tau=0}^{h} \sim N\left(\int_0^h \underline{\mu}_{t+\tau} d\tau, \int_0^h \mathbf{\Omega}_{t+\tau} d\tau \right), \tag{2.11}$$

where $\sigma\{\underline{\mu}_{t+\tau}, \mathbf{\Omega}_{t+\tau}\}_{\tau=0}^{h}$ denotes the σ-field generated by the sample paths of $\underline{\mu}_{t+\tau}$ and $\mathbf{\Omega}_{t+\tau}$ for $0 \le \tau \le h$.

The integrated diffusion matrix thus provides a natural measure of the true latent h-period volatility. This notion of integrated volatility plays a central role in the stochastic volatility option pricing literature, where the price of an option typically depends on the distribution of the integrated volatility process for the underlying asset over the life of the option.

By the theory of quadratic variation and under weak regularity conditions, Andersen et al. (2001) show that

$$\sum_{j=1,\dots,[h/\Delta]} \underline{r}_{t+j\Delta,\Delta}\,\underline{r}'_{t+j\Delta,\Delta} - \int_0^h \boldsymbol{\Omega}_{t+\tau} d\tau \to 0\,, \tag{2.12}$$

almost surely for all t as the sampling frequency of the returns increases, or $\Delta \to 0$.
Thus by summing sufficiently finely sampled high frequency returns, it is possible to construct ex-post realized volatility measures for the integrated latent volatilities. This contrasts sharply with the common use of the cross product of the h period returns, $\underline{r}_{t+h,h}\underline{r}'_{t+h,h}$, as a simple ex post volatility measure.

Although the squared return over the forecast horizon provides an unbiased estimate for the realized integrated volatility, it is an extremely noisy estimator, and predictable variation in the true latent volatility process is typically dwarfed by measurement error. For longer horizons any conditional mean dependence will tend to contaminate the variance measure. In contrast, as the length of the return horizon decreases the impact of the drift term vanishes, so that the mean is effectively annihilated.

2.3.3 Problems with the practical implementation of the realized volatility approach

The theoretical results in the previous section effectively state that realized volatility yields a perfect estimate of volatility in the hypothetical situation where prices are observed in continuous time and without measurement error. This result suggests that the realized variance (RV), which is the sum of squared error returns (SSR), should be based on returns that are sampled at the highest possible frequency (tick-by-tick data). However, in practice this leads to the well-known bias problem due to market microstructure noise, see e.g., Zhou (1996), Andreo and Ghysels (2002), and Oomen (2001). The bias is particularly evident from volatility signature plots that were introduced by Andersen et al. (2000). Hence, there is a trade-off between bias and variance when choosing the sampling frequency and this is the reason that returns are typically sampled at moderate frequency, such as 5 minute sampling. An alternative way to handle the bias problem is to use bias correction techniques. For example a moving average filter was used by Andersen et al. (2001) and Bandi and Russell (2003 a&b) and an autoregressive filter by Bollen and Inder (2002). These estimators referred to as the ABDE (after Andersen, Bollerslev, Diebold and Ebens) estimator and the VARHAC (Vector Autoregressive Heteroscedastic Autocorrelation Consistent) estimator will be the focus of this thesis and will be defined and discussed in detail in the following two sections. Other bias correction techniques were recently introduced by Zhang et al. (2003) and Hansen and Lunde (2004b). The first approach considers time independent noise and uses a sub-sampling approach, while the latter allows for time dependence in the noise process and uses a kernel-based approach. We will not investigate these techniques in this thesis, because we only recently became aware of their existence.

2.3.4. The ABDE realized variance estimator

We now continue by discussing the construction of the ABDE realized variance estimator. Andersen et al. (2001) construct a 5 minute returns series from the logarithmic difference between the prices recorded at or immediately before the corresponding 5 minute mark. Although the limiting result in Equation (2.12) is independent of the value of the drift parameter μ_t, the use of a fixed discrete time interval could allow dependence in the mean to systematically bias the volatility measures. In order to purge the high frequency data of the negative serial correlation induced by uneven spacing of the observed prices and the inherent bid ask spread, an MA(1) model for each of the 5 minute return series is estimated.

In Section 2.3.2 we used the notation $RV_{[a,b]}^{(m)}$ to denote the realized variance at frequency m over the interval $[a,b]$. To simplify notation we use the subscript t to refer to day t and write $RV_t^{(m)}$ in place of $RV_{[a,b]}^{(m)}$ where $[a,b]$ represents the hours of day t that the market is open. In the case of 5 minute intraday returns, $m = \frac{b-a}{5} = 96$ assuming the market opens at $a = 9\ am$ and closes at $b = 5\ pm$. Suppose we denote the resulting 5 minute intraday return series by $\{Y_i; i = 1,2,\ldots,Tm\}$ where $T = 690$ is the number of trading days and $m = 96$, the number of 5 minute intraday returns. Prior to the introduction of the new trading methodology of the JSE Securities Exchange SA in May 2002, the market opened at $a = 9\ am$ and closed at $b = 4\ pm$ where $m = 84$. As previously we have $Y_t = \ln(\frac{P_t}{P_{t-1}})$, with P_t the price of the trade at time t. Then the realized variance RV_t on day t is equal to the sum of the squared intraday returns, SSR_t, on day t. Therefore we have that the realized volatility on day t is $\sqrt{SSR_t}$. In our notation $RV_t = SSR_t = \sum_{i=(t-1)m+1}^{tm} Y_i^2$, with $m = 96$ because we are concerned with 5 minute intraday returns. Suppose we fit an MA(1) model to the return series $\{Y_i; i = 1,2,\ldots,Tm\}$ as Andersen et al. (2001) suggested in order to purge the high frequency data series of negative serial correlations. The residuals resulting from the fit are then used to calculate the realized volatilities. Although standard packages like SAS and MATLAB provides the residuals from such a fit as standard output we give the mathematical details below. Assuming a MA(1) process, let $Y_i = e_i - \theta e_{i-1} + \mu$, where e_i is ith error, θ the moving average parameter and μ the mean of the process. Note that the first order autocorrelation of the MA(1) process is $\frac{-\theta}{(1+\theta^2)}$, so that a positive θ implies a negative first order autocorrelation. After fitting the MA(1) process to the time series of intraday returns, the estimated errors or residuals from the fit are obtained as

$$\begin{aligned} \hat{e}_1 &= Y_1 - \hat{\mu}; \\ \hat{e}_i &= Y_i + \hat{\theta}\hat{e}_{i-1} - \hat{\mu}; \quad \text{for} \quad i = 2,\ldots,Tm. \end{aligned} \tag{2.13}$$

The residuals will now form the new time series $\{Y_i^* = \hat{e}_i;\ i = 1,2,\ldots,Tm\}$ of intraday returns which is demeaned and the first order serial correlation removed. The ABDE realized variance estimator on day t is then obtained as $ABDE_t = \sum_{i=(t-1)m+1}^{tm} (Y_i^*)^2$; for $t = 1,2,\ldots,T$ which is the sum of squared intraday residual returns resulting from the MA(1) fit. Again the ABDE realized volatility estimator on day t is obtained as $\sqrt{ABDE_t}$.

In Figure 2.3 an example of daily realized variance estimates obtained via the ABDE variance estimator is shown. The ABDE variance estimator suggests two periods of noticeable increase in volatility namely, the terrorist attacks in the USA and the well documented leaking of the Aug 2002 Mining Charter. The two periods of high volatility corresponds to significant movements in the price and log return series. In comparing the variance estimates in Figure 2.2 with those in Figure 2.3, the GARCH and ABDE variance estimators indicate similar volatility clusters. However the volatility experienced around the Dec 2001 futures close out is not as prominent in the ABDE variance graph as in the GARCH variance graph. The opposite is true for the Dec 2000 future close out. We could provide the following explanation. Remember that the GARCH estimates are based on end of day closing prices and the ABDE estimates are based on high frequency 5 minute data. The high frequency data was more volatile during the Dec 2000 futures close out period in comparison to the end of day data, explaining the Dec 2000 volatility cluster in the case of the ABDE variance estimator. In the case of the Dec 2001 futures close out the end of day time series was more volatile (close to close moves of 9%) in comparison to the high frequency time series, contributing to the higher GARCH variance estimates.

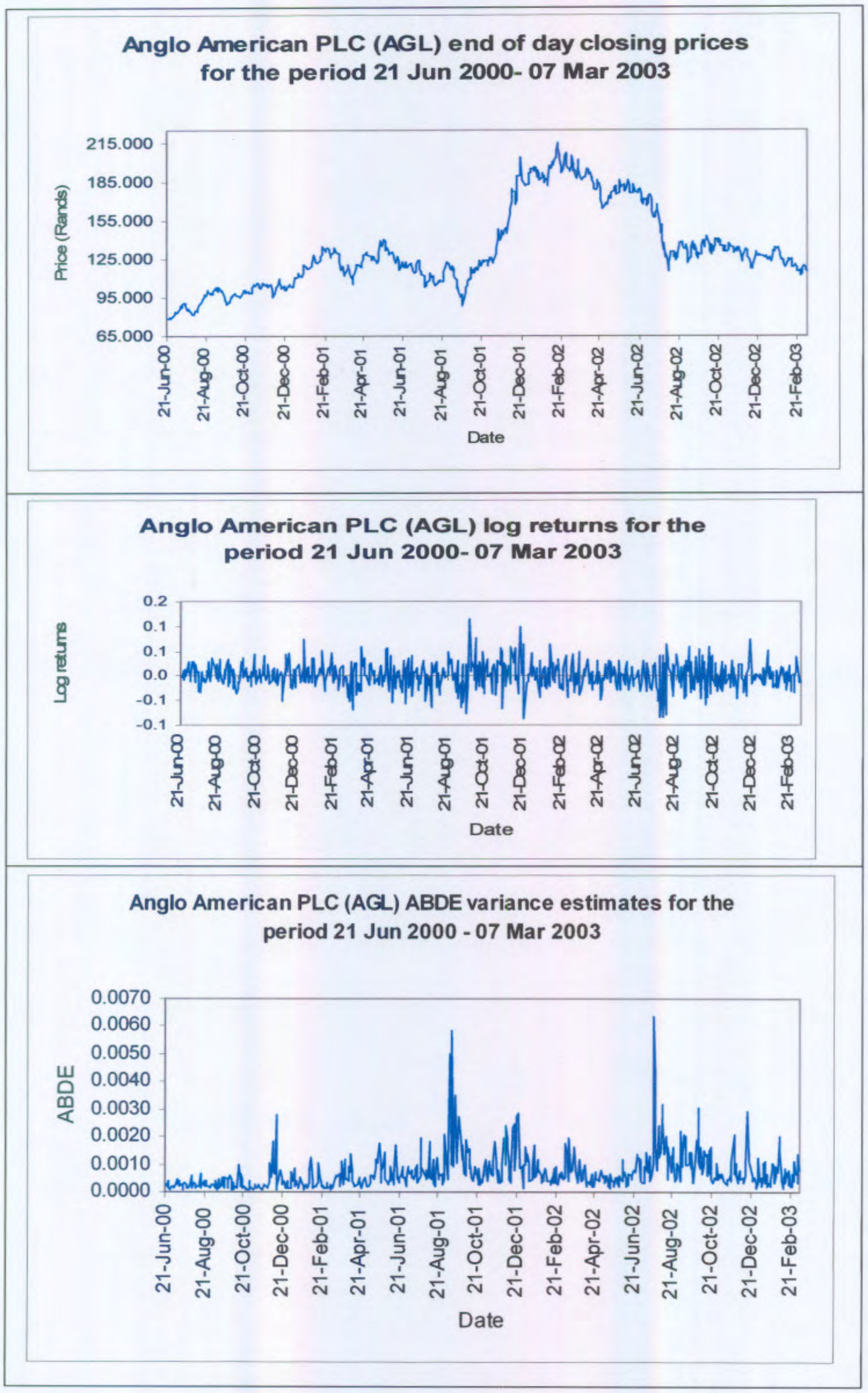


Figure 2.3: Anglo American PLC (AGL) end of day closing prices, log return and ABDE variance estimator graphs for the period 21 Jun 2000 – 07 Mar 2003.

2.3.5 The VARHAC realized variance estimator

The Vector Autoregressive Heteroscedastic and Autocorrelation Consistent (VARHAC) estimator was introduced by den Haan and Levin (1996) as a parametric spectral estimation procedure for constructing heteroscedastic and autocorrelation consistent (HAC) covariance matrices. The VARHAC estimator is based on earlier work by Parzen (see e.g., Parzen (1983)). Bollen and Inder (2002) used the VARHAC and some other estimators to estimate daily volatility. The VARHAC realized variance estimator makes use of all transaction (tick-by-tick) data on a specific trading day.

As motivation for the VARHAC estimator of daily realized variance, Bollen and Inder (2002) considered some stylized facts with regards to the properties of high frequency returns, which lead the authors to propose a set of assumptions for the data generating process for high frequency returns.

They mention that previous studies highlighted the following characteristics of high frequency day returns:

- Andersen and Bollerslev (2001) found that the number of transactions per 5 minute interval is highly seasonal and strongly related to the 5 minute volatility pattern. Admati and Pfeiderer (1998) developed a theoretical model in which concentrated trading at certain times is the result of strategic trading behaviour of both informed and liquidity traders.

- Roll (1984) showed that price fluctuation between the bid and ask spread induces first order autocorrelations into the high frequency return series.

- Hasbrouck and Ho (1987) after researching the autocorrelation structure of a number of securities, found that the pattern consists of large negative autocorrelation at first lag, followed by positive autocorrelation of decreasing magnitude that are statistically significant. The positive autocorrelation at lags greater than one may be due to traders' working orders (buying and selling on behalf of clients, executing hedges for their options etc.).

From the above, Bollen and Inder (2002) conclude that each trading day is characterized by both predictable (seasonal) and unpredictable periods of volatility in response to news arrivals, etc. Each trading day may be characterized by a different auto-correlation structure, since autocorrelation in high frequency returns appears to be closely related to periods of concentrated trading. The existence of auto-correlation at lag lengths greater than one (strong microstructure effects) would invalidate the suitability of volatility estimators based on daily or high and low data input parameters. An alternative approach would be to allow for the latter properties of high frequency data and to incorporate all available high frequency data as in the VARHAC estimator proposed by Den Haan and Levin (1996).

Bollen and Inder (2002) describe the VARHAC estimator as follows. Assume n_t returns are observed in a trading day t (where n_t maybe, for example, the number of 5 –minute returns or transaction returns on day t). Define $r^*_{i,t} = n_t^{0.5} r_{i,t}$ so that the daily return can be written as $r_t = \sum_{i=1}^{n_t} r_{i,t} = n_t^{-0.5} \sum_{i=1}^{n_t} r^*_{i,t}$. The goal is to estimate the variance of the daily returns parameter $\sigma_t^2 = E[r_t^2]$. The latter parameter can be expressed as a function of the high frequency data, i.e.

$$\sigma_t^2 = E[r_t^2] = E\left[\left(\sum_{i=1}^{n_t} r_{i,t}\right)^2\right],$$

which can be written as

$$\sigma_t^2 = \frac{1}{n_t}\sum_{i=1}^{n_t}\sum_{j=1}^{n_t} E[r^*_{i,t} r^*_{j,t}], \qquad (2.14)$$

where σ_t^2 represents the spectral density of $r^*_{i,t}$ at frequency zero.

One of the concerns of Den Haan and Levin (1996) was to estimate parameters underlying the data generating mechanism for a process $\{V_t\}_{t=-\infty}^{\infty}$. In particular they are interested in estimating

$$S_T = \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} E[V_s V_t'], \qquad (2.15)$$

where V_s and V_t are N dimensional vectors. Suppose , we simplify their notation and let N=1 then V_t is a scalar. If we let $V_t = r^*_{i,t}$ and $T = n_t$ (by comparing equations (2.14) and (2.15)) then S_T is equal to σ_t^2).

We can further appeal to the literature on spectral density estimation to guide us in our estimation of σ_t^2. Den Haan and Levin (1996) argue strongly for the use of the VARHAC estimator. Before the estimator can be adopted it is necessary to verify whether $r_{i,t}^{*}$ satisfies the assumptions required for valid use of the estimator.

The following conditions must be met in order to implement the VARHAC estimator (Den Haan and Levin Section 3.2 (1996)):

Conditions:

Let $\{V_t\}_{t=-\infty}^{\infty}$ be a mean zero sequence of N – vectors, which satisfy the following conditions:

A. $\sup_{t\geq 0} E[V_t V_t'] < +\infty$;

B. $\Gamma(0) = \lim_{T\to\infty} 1/T \sum_{t=1}^{T} E[V_t V_t']$ is positive definite, and ;

C. $\sum_{j=1}^{\infty} \sup_{t\geq 1} | E[V_t V_{t+j}'] |_{\infty} < +\infty$.

These conditions allow V_t to possess a degree of heterogeneity and autocorrelation, and they provide sufficient conditions for the valid estimation of S_T using the VARHAC estimator. In order to generalise the assumptions to high frequency returns the authors rewrote the conditions, with reference to $r_{i,t}$.

Modified Conditions :

Assume $r_{i,t}$ is a mean zero sequence of scalars satisfying:

A. $\sup_{i\geq 0} n_t E[r_{i,t}^2] < +\infty$;

B. $\sigma_t^2 = \lim_{n_t\to\infty} \sum_{t=1}^{n_t} E[r_{i,t}^2] > 0$, and ;

C. $\sum_{j=1}^{\infty} \sup_{i\geq 1} | E[r_{i,t} r_{i+j,t}] |_{\infty} < +\infty$..

We now discuss the above-mentioned modified conditions:

- Condition (A) requires that high frequency returns have bounded variance,

- Condition (B) rules out the possibility of too many values having zero variance.

- Condition (C) controls the degree of dependence across observations, it implies that the $r_{i,t}$ are asymptotically independent, and that the sums of the autocovariances are bounded. No further restrictions are placed and these conditions allow for a wide range of possible heteroscedasticity and autocorrelation in high frequency returns. The autocorrelations are not restricted to be constant across time – the only constraint is the absolute summable restriction in condition (C).

The key advantage of the weak set of conditions on $r_{i,t}$ is that they are consistent with the theory regarding time series behaviour of high frequency returns. Such conditions allow $r_{i,t}$ to be unequally spaced transaction returns. Unequally spaced data produces another potential source of heteroscedasticity, and also cause auto variances to vary across observations. Neither of these implications leads to a distortion of the assumptions stated in Equation (2.16). Following den Haan and Levin (1998) we can apply the VARHAC estimator to obtain a consistent estimator for σ_t^2, which incorporates all of the available high frequency information. The VARHAC estimator is constructed as follows:

Step 1: Lag order selection for each day

Consider a series of observations on transaction returns $\{r_{1,t}, r_{2,t}, \ldots, r_{n_t,t}\}$ for a specific trading day t, and furthermore assume the researcher has data available for S such trading days (i.e. $1 \le t \le S$). For each of the S days of data, the following AR(K) model is estimated by least squares:

$$r_{i,t} = \sum_{k=1}^{K} \hat{\alpha}_{k,t}(K) r_{i-k,t} + \hat{e}_{i,t}(K) \text{ for } i = K+1, \ldots, n_t, \qquad (2.16)$$

for each possible lag order $K = 1, \ldots, \bar{K}$. For $K = 0$, set $\hat{e}_{i,t}(0) = r_{i,t}$.

Now calculate the Schwarz BIC:

$$BIC(K,t) = \ln\left(\frac{\sum_{i=K+1}^{n_t} \hat{e}_{i,t}^{2}(K)}{n_t}\right) + \frac{\ln(n_t)K}{n_t}. \qquad (2.17)$$

For each day, the optimal lag order K_t is chosen as the value of K which minimizes $BIC(K,t)$.

Step 2: Calculate the estimate of σ_t^2

For each day, the selected lag order K_t and estimates of $\hat{\alpha}_{k,t}(K_t)$ and $\hat{e}_{i,t}(K_t)$ from Equation (2.16) can be used to compute the VARHAC estimate:

$$VARHAC_t = \frac{n_t \hat{\sigma}^2}{\left[1 - \sum_{k=1}^{K_t} \hat{\alpha}_{k,t}(K_t)\right]^2}, \qquad (2.18)$$

where

$$\hat{\sigma}^2 = \sum_{i=K_t+1}^{n_t} \frac{\hat{e}_{i,t}(K_t)^2}{n_t}.$$

The estimator defined in Equation (2.18) will be referred to as the VARHAC variance estimator $VARHAC_t$. We will sometimes drop the subscript t for notional simplicity.

In Figure 2.4 we graphically depict the VARHAC variance estimator for Anglo American PLC (AGL) from 22 Jun 2000 – 07 Mar 2003. Once again the two periods of noticeable volatility can be ascribed to the 11 Sep 2001 terrorist attacks in the USA and the well documented leaking of the Aug 2002 Mining Charter. Both the realized variance estimators (VARHAC and ABDE) in Figure 2.3 and 2.4 and the GARCH variance estimator in Figure 2.2 depict similar volatility clusters for Anglo American PLC (AGL). For these specific volatility clusters the ABDE variance estimates are larger than the VARHAC variance estimates which, in turn, are larger than the GARCH variance estimates. In Section 4.2 we discuss the characteristics of the realized variance estimators and the market factors contributing to the higher variance estimates in more detail. As mentioned previously, the GARCH variance estimations are based on end of day closing prices and would not reflect the "inter market" moves (between close to close) as the realized variance estimates that are based on high frequency data would.

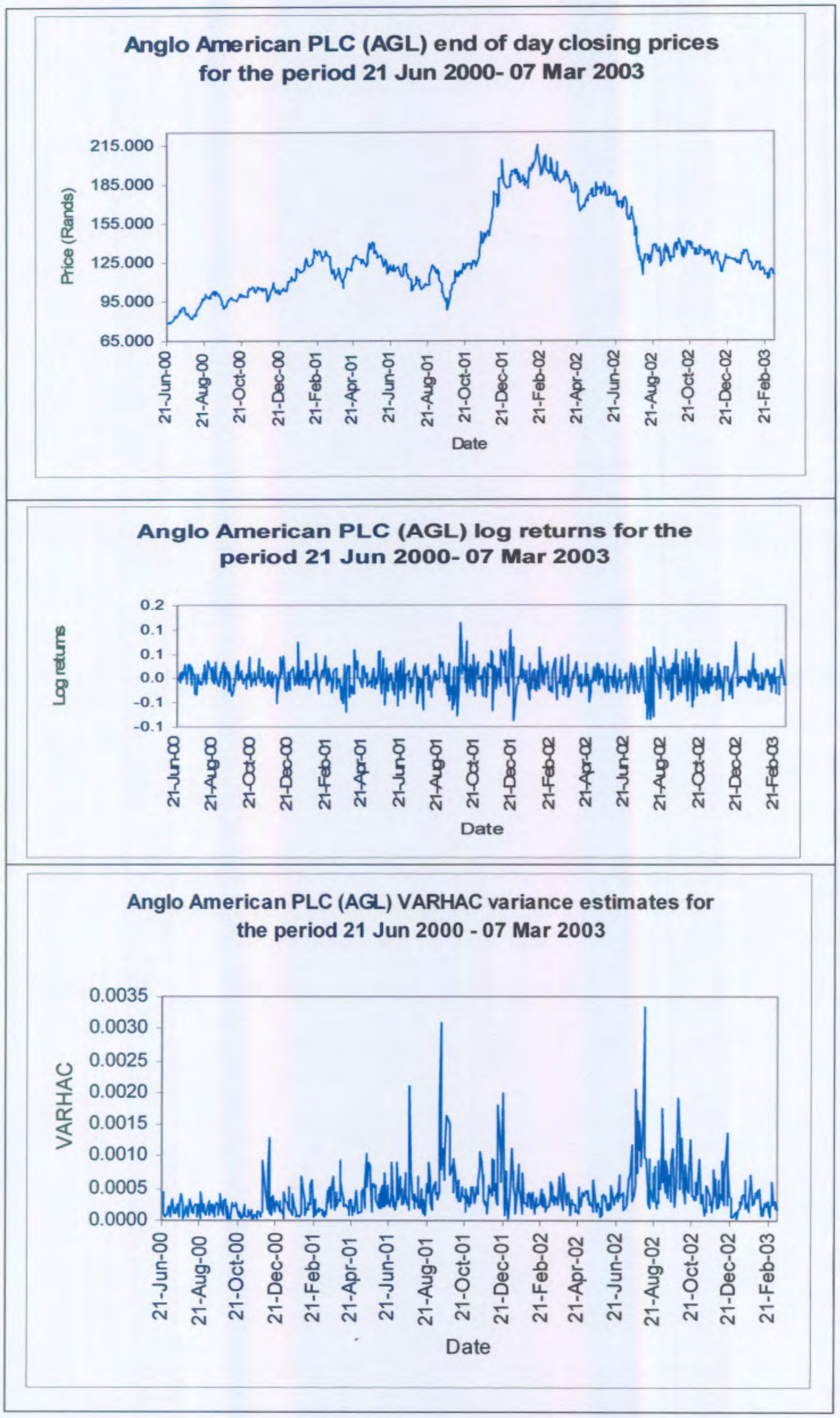


Figure 2.4: Anglo American PLC (AGL) end of day closing prices, log return and VARHAC variance estimator graph for the period 21 Jun 2000 – 07 Mar 2003.

2.4.6 Estimation and forecasting issues

When a realized volatility estimation technique is used to estimate volatility a different technique is usually used to forecast volatility. The most popular technique is currently autoregressive fractionally integrated moving average models (ARFIMA).

Oomen (2001) calculated and analysed realized volatility by employing minute by minute data on the FTSE 100 index. Oomen found that in order to implement the "sum of squared returns approach" to calculate realised volatility, a careful check on serial dependence in high frequency returns was necessary in order to avoid biases in the resulting measure for average daily return volatility. Motivated by several test statistics for the presence of long memory, Oomen modelled the realised volatility time series as an ARFIMA model and found lagged returns and trading volumes to be significant regressors. In a simulation study the forecasting performance of the ARFIMA model for daily realised volatility was assessed and the authors found that the ARFIMA model outperformed conventional GARCH models. It should be noted that the ARFIMA model requires more input data than the GARCH models and that the GARCH model's ability to account for persistence in the volatility process makes it a reasonable and easier alternative to the data intensive and complicated ARFIMA model for realised volatility.

Hol and Koopman (2002) explored the forecasting of values of high frequency time series in conjunction with a variety of volatility models (realized volatility, GARCH and stochastic volatility) on the S&P 100 index. The authors found that for forecasting horizons ranging from one day to one week the most accurate out of sample volatility forecasts were obtained with the realised volatility and extended stochastic volatility models. Although the GARCH model extended with intraday volatility appeared to perform well when its forecasts are evaluated on the basis of regression methods, other evaluation criteria indicated that it tends to overestimate volatility. As the GARCH (1,1) model also suffers from this problem the authors conclude that in the absence of intraday volatility information the stochastic volatility model is the preferred model for forecasting volatility.

2.4 SOME OTHER VOLATILITY ESTIMATORS

In the following section we will discuss some of the other volatility estimators proposed by Bollen and Inder (2002) in their article. We will not be pursuing these estimators in our empirical analysis, however in the following Sections 2.4.1-2.4.3 we will discuss the reason as to why these estimators have been eliminated from the study.

2.4.1 Simple volatility estimator

Bollen and Inder (2002) begin by assuming the daily return $Y_t = \ln(p_t) - \ln(p_{t-1})$ where p_t is the closing price on day t and Y_t is a drawing from the normal distribution with time varying volatility:

$$Y_t \sim N\left(0, \sigma_t^2\right). \tag{2.19}$$

The first and most basic estimator is based upon the observation that the expected value of the folded normal distribution is given by $E\{|Y_t|\} = \sigma_t \sqrt{2/\pi}$. Bollen and Inder (2002) defined the Simple Volatility Estimator as:

$$\hat{\sigma}_{(simple,t)} = \frac{|Y_t|}{\sqrt{2/\pi}}. \tag{2.20}$$

The Simple Volatility Estimator is based on only one observation to estimate each day's volatility, suggesting that the estimator will be highly inefficient in comparison to the other estimators selected for our study. The simple volatility estimator for the share AGL for the period 21 Jun 2000 – 7 Mar 2003 is depicted in Figure 2.5. Similarly to our more complex volatility estimators the simple volatility estimates also increased noticeably for the share Anglo American PLC (AGL) around the time of the 11 Sep 2001 terrorist attacks, during the Dec 2001 futures close out as well as at the leaking of the Mining Charter in Aug 2002.

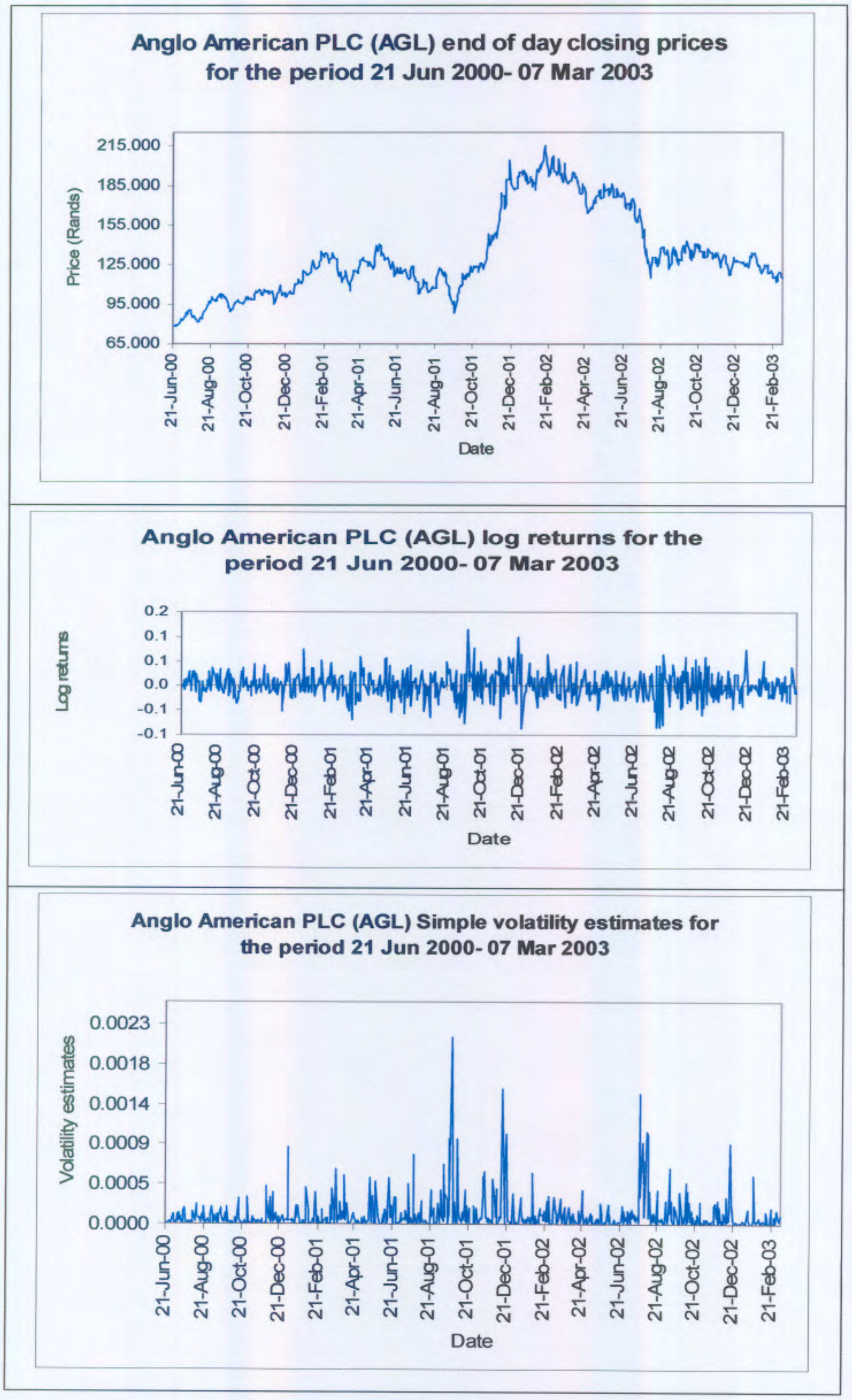


Figure 2.5: Anglo American PLC (AGL) end of day closing price, log return and simple volatility estimator graphs for the period 21 Jun 2000 – 7 Mar 2003.

2.4.2 PARK variance estimator

Extreme value estimators make some use of the information in the high frequency data such as the high, low, opening and closing prices. Define H_t as the highest price and L_t as the lowest price on day t. Parkinson (1980) developed the PARK daily variance estimator, based upon the assumption that high frequency prices follow a geometric Brownian motion process. The corresponding PARK daily variance estimator under this assumption is defined as:

$$\hat{\sigma}^2_{(PARK),t} = \frac{(\ln(H_t) - \ln(L_t))^2}{4\ln(2)}. \tag{2.21}$$

In Figure 2.6 we show the PARK variance estimator for the share AGL for the period 21 Jun 2000 – 7 Mar 2003. In contrast to the other estimation techniques discussed previously the PARK estimator is much more volatile and we therefore decided to discard the PARK variance estimator. Furthermore, we refer the reader Section 2.4.3, where after personal communication with Dr Bollen, the decision was made to discard the GK estimator, which is more efficient than the PARK estimator and therefore the PARK estimator should also be discarded.

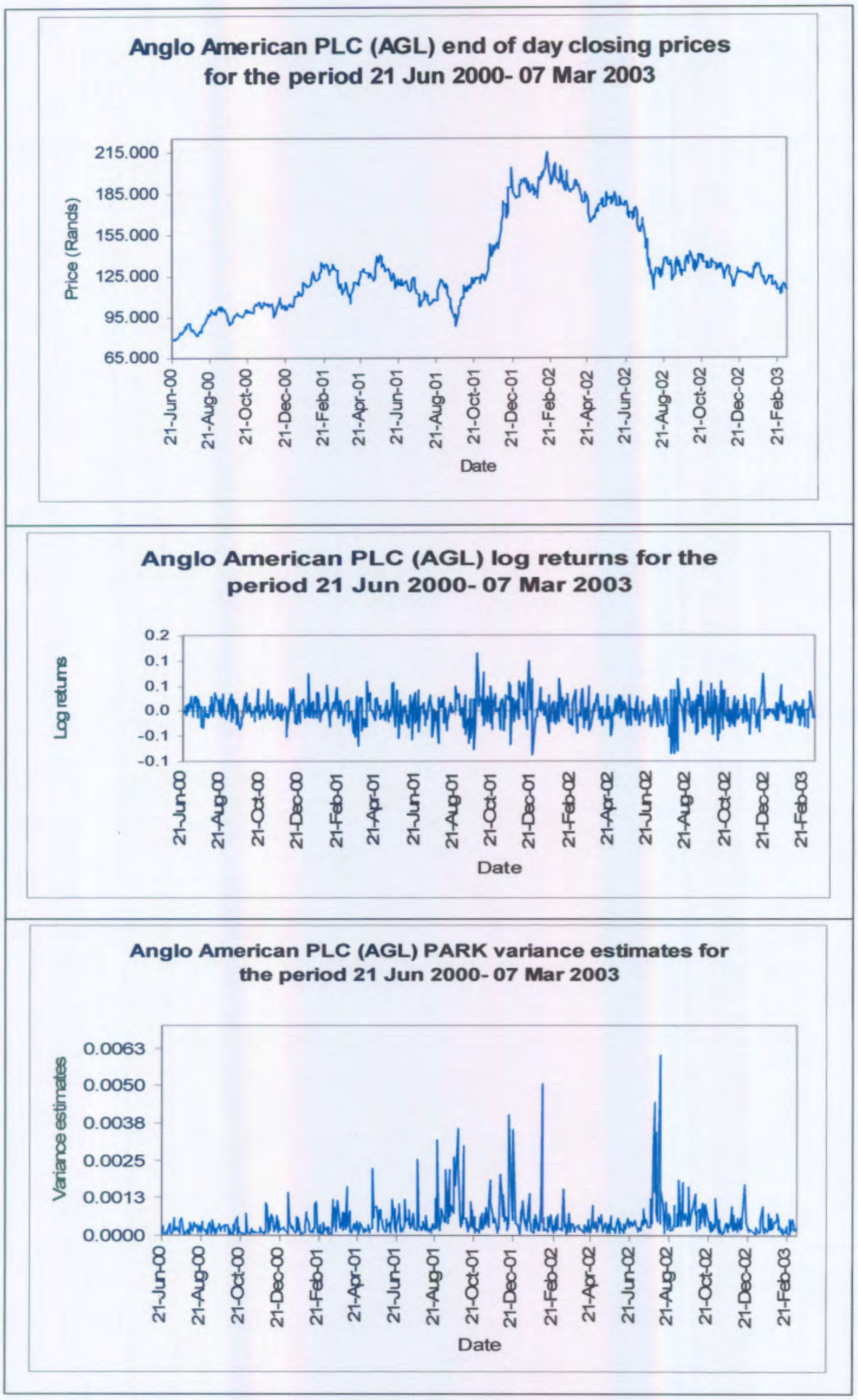


Figure 2.6: Anglo American PLC (AGL) end of day closing prices, log return and PARK variance estimator graphs for the period 21 Jun 2000 – 7 Mar 2003.

2.4.3 GK variance estimator

Garman and Klass (1980) proposed an estimator, which they claim is more efficient than PARK. The corresponding estimator assuming a geometric Brownian motion price process shall be referred to as the GK variance estimator and is defined as:

$$\hat{\sigma}^2_{(GK),t} = 0.5(\ln(H_t / L_t))^2 - 0.39(\ln(p_t / p_{t-1}))^2. \tag{2.22}$$

Similarly to the Simple and PARK variance estimator, the GK estimator is also an inefficient estimator of realized variance based on high frequency data. This is because the high and low prices of a share are not a very accurate reflection of trading and, therefore of volatility.

In our analysis (Figure 2.7) we found that the GK variance estimator returned negative volatilities on some trading days. The reason being that the difference between the previous day's close to the current days closing price was greater than the difference between the high and low prices on that particular trading day. The latter volatility characteristic is counter intuitive and it was necessary to recheck our GK volatility estimator calculations. We confirmed that we applied the GK variance estimator in Equation (2.22) as Bollen and Inder had stated. In e-mail correspondence with Dr Bollen we explained our findings and he confirmed that they had seen similar occurrences of a negative GK variance estimator in previous studies. We therefore decide to also exclude the GK variance estimator from our study.

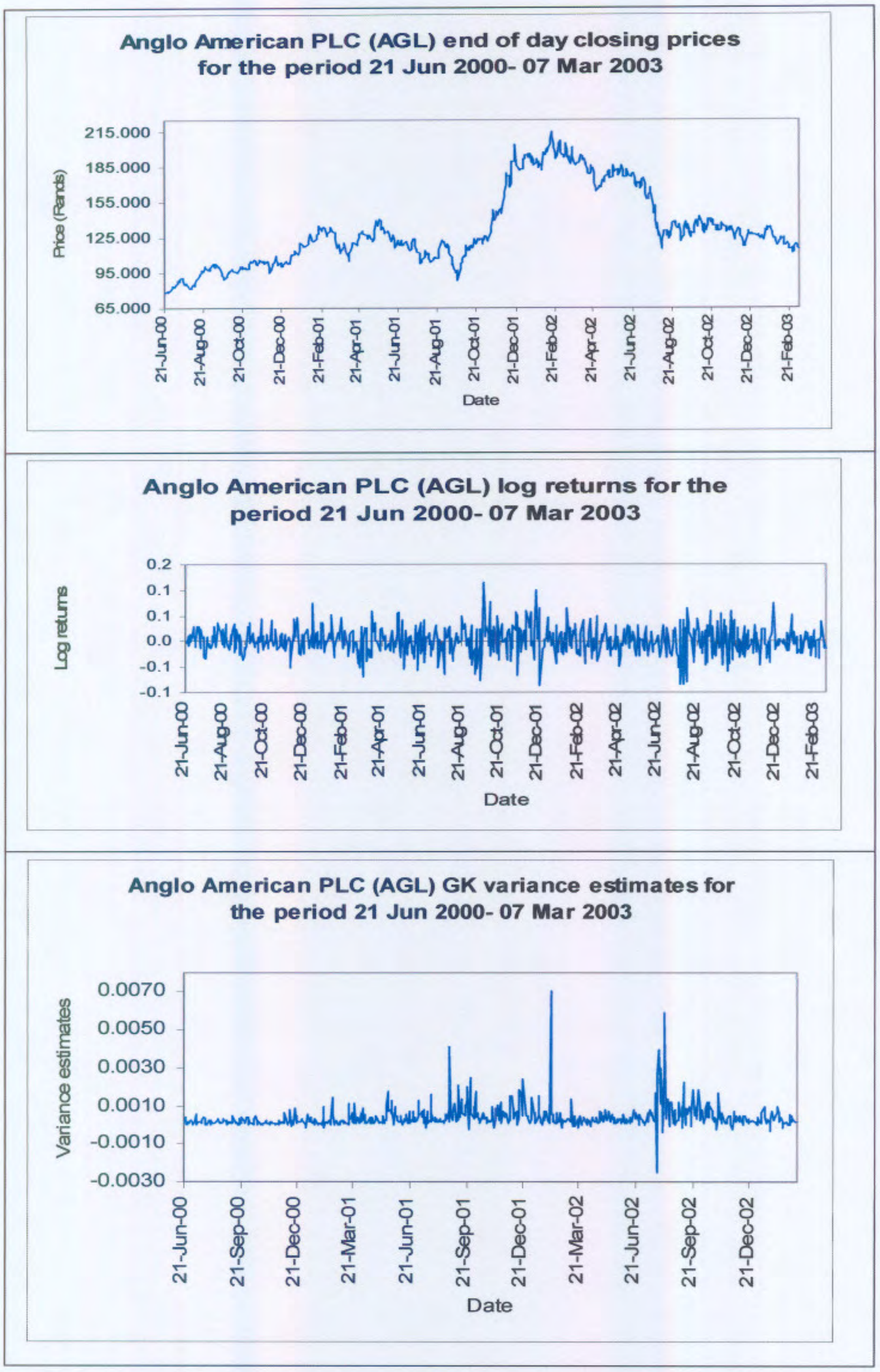


Figure 2.7: Anglo American PLC (AGL) end of day closing prices, log return and GK variance estimator graphs for the period 21 Jun 2000 – 7 Mar 2003.

2.5 CONCLUSION

We have discussed several theoretical measures for volatility of which quadratic variation (QV), integrated variance (IV) and conditional variance (CV) are the most popular. Realized variance is a consistent estimator for QV and can approximate IV and CV under various assumptions. GARCH models are typically only concerned with CV. In this chapter we have also discussed the GARCH and NIG-GARCH model proposed by Venter and de Jongh (2002) and showed how these models may be fitted to daily data to provide estimates and forecasts of daily volatility. We have also proposed three realized volatility methods which use intraday data to estimate daily volatility, namely $\sqrt{SSR_t}$, $\sqrt{ABDE_t}$ and $\sqrt{VARHAC_t}$.

CHAPTER 3

HIGH FREQUENCY DATA FILTERING

3.1 INTRODUCTION

Traditional studies of financial markets have relied on price observations drawn at fixed intervals, usually end of day (close to close). This sampling pattern was perhaps dictated by the general view that whatever drove equity prices and returns probably did not vary significantly over short time intervals and therefore it was deemed unnecessary to investigate shorter time intervals. Not only are the actual price quotations important in the understanding of the structure of financial markets, there is also additional information in the duration or time interval between quotations. A fundamental property of high frequency data is that observations can occur at non-equally spaced time intervals. Equity trades are not normally equally spaced throughout the day, resulting in intraday "seasonals" in the volume of trade, the volatility of prices, and the behaviour of spreads. During some of the time intervals, no transactions may occur, dictating that even measuring returns may be problematic. These difficulties are less pronounced when fixed daily data is used but becomes more important when high frequency data is analyzed. Given that financial markets display high speeds of price movements/adjustment, studies based upon daily observations may fail to capture information contained in intraday or high frequency data and market movements. With the development of computer technology, data availability is becoming less of a problem for researchers and the last few years has seen a rapid development of econometric and statistical techniques applicable to financial market data.

In this chapter we provide the necessary foundations we had to complete in order to construct a reliable time series to utilize in the calculation of our various volatility estimators. Before we could design a data filter to purge erroneous data points we first had to identify these errors as well as their causes (e.g., human capturing errors and the nature of trading on the exchange). Furthermore it was also necessary to identify market characteristics that influence volatility, such as corporate events, the arrival of news, introduction of a new share settlement system and non-economic world events such as the 11 Sep 2001 terrorist attacks. Once these events had been identified we could design a data filter in order to filter erroneous data points from our South African data. While a number of studies have examined the characteristics of intraday return volatility in the international markets (refer to Chapter 2), to the best of our knowledge no previous research has been conducted in the

South African equity market with regards to high frequency return volatility of individual shares.

Before a high frequency time series can be statistically analysed the following processes (depicted in Figure 3.1) should take place. We will use this process to motivate the layout of the chapter:

- Firstly, high frequency data of a reliable quality must be obtained. At the time of our study it was extremely difficult to obtain high frequency data for the South African equity market, as data wasn't easily available to the researcher (we obtained the data used in our study from a market participant). In Section 3.2 of this chapter we begin with a discussion of the underlying characteristics and trade mechanics of international financial markets and specifically South African equity markets. In Section 3.3 we describe the subset of six South African ALSI 40 shares selected for our study. We conclude the section with a discussion on the format of the data files and some descriptive statistics for the subset of six shares.

- Once the high frequency data had been obtained it was necessary to filter the data of potentially erroneous data points that could have affected the resulting volatility and the reliability of the statistical results. In Section 3.4 we discuss the various factors identified in our study that would impact volatility, namely erroneous data points in the time series, random market events (corporate actions, macro economic news announcements), once-off events (such as the change in trading methodology on the JSE Securities Exchange SA (JSE) and the introduction of STRATE) as well as extreme market events (terrorism etc.). We continue in Section 3.4 by discussing the different aspects the researcher had to keep in mind when filtering high frequency data and the data filtering methodology we applied in filtering erroneous data points for our subset of six shares in more detail.

- Once the data had been filtered it was necessary to construct the time series as specified by the model chosen to analyse the data. The construction of the time series applicable to each of the different approaches chosen to calculate daily volatility (refer to Chapter 2) will be discussed in Section 3.5.

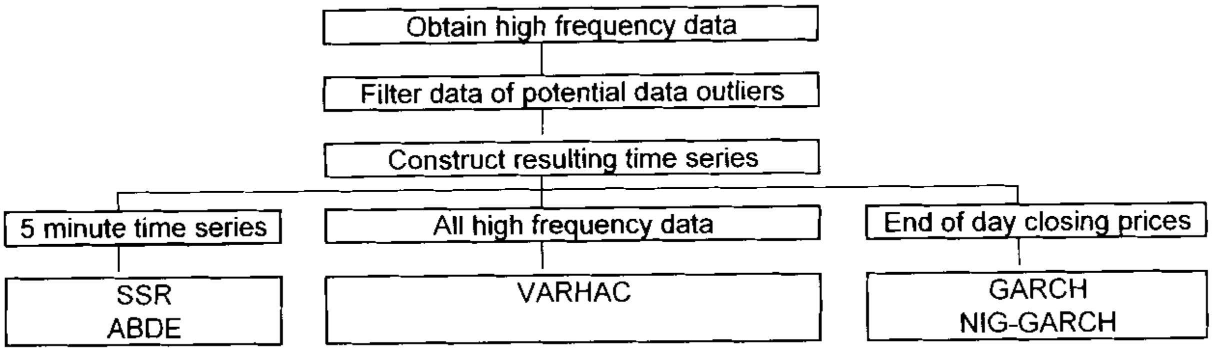


Figure 3.1: Data gathering and filtering process followed before the researcher can begin analyzing high frequency data.

3.2 INTERNATIONAL AND SOUTH AFRICAN FINANCIAL MARKETS

Markets across the world use different trading methodologies to facilitate trading, e.g., trading floors with traders shouting sell and buy orders, whereas other markets utilize automated screen trading. In this section we will discuss each of these trading methodologies in more detail as well as the potential data capturing errors that can occur within each.

3.2.1 International markets

Trading system

The following major trading methodologies can be identified:

Dealing on a trading floor market is done by means of open outcry. Traders meet each other face to face and "shout" their buying and selling prices out loud on a physical trading floor.

In a *Screen quotation telephone trading market*, buyers and sellers negotiate a financial transaction over the telephone. The screen (usually a Reuter's or Bloomberg's screen) is used to advertise potential non-binding prices. Screen quotation usually takes place in informal markets such as the bond and foreign exchange markets.

In a "*Screen trading only*" system, firm quotes are made on screen and transactions are finalized by telephone. Screen trading requires a central clearing house. Currently the South African bond market utilizes both an automated screen trading system (BATS) as well as the screen quotation telephone system (Reuter's or Bloomberg's). Although the screen market is a major improvement on screen telephone trading it is still regarded as an interim step between floor and automated trading systems.

In an *Automated trading system* environment all trading is fully automated and executed from a trading screen platform. Buyers and seller transact using the central automated trading system of the exchange. Examples of automated trading systems in a South African context are TALX or Hermes (equity platforms), ATS (derivative platform) and BATS (bonds).

The trading model employed by the exchange will have an effect on the record keeping quality of the particular exchange's daily trading data, namely whether the trading data is being captured manually (human intervention) or automatically (by sophisticated computer systems). Each of the above mentioned trading methodologies will have their own data capturing "issues" and potential data errors. In Section 3.2.1 we will discuss some of these issues briefly (the focus of the chapter being the identification of South African data errors and issues).

In open outcry markets such as the Chicago Board of Trade and the New York Stock Exchange, participants "voicing" bids and offers generate trades on physical exchanges and trained pit reporters located on the trading floor report executed trades. The various trained pit reporters will manually enter the various hand signals and verbal quotes that confirm trade bid and offers into hand held devices which are routed to servers. Due to the manual process of capturing trades into hand held devices the following error may occur; namely: Software running on the hand held devices reduce, but do not eliminate, the possibility of multiple reporters entering the same trade (e.g., trade duplication). The trade reporters may enter incorrect prices and quantities. For example, decimal errors may occur, i.e. a share may trade at $100 and be entered at $1000 or $1 in the hand held device.

In telephonic and screen-only trading markets all trades will be reported at the end of the trading day and manually captured by each market participant into their respective in-house trading systems. Due to the nature of trading, maintaining a database of all trades in the market will be a problematic task. As mentioned in the previous section, with any process where human intervention is necessary, data capturing errors could occur.

In electronic markets, such as the JSE and the London Stock Exchange, the buy and sell orders are matched electronically. All electronic exchanges keep records of transactions consummated on the exchange, the price, volume and the counterparties involved and time of the deal. Data is collected electronically in sequence and released to data vendors and the various market participants. Although the trades are entered electronically, data transmission errors such as decimal places and incorrect trading prices may still occur.

High frequency data

All electronic trading international exchanges keep records of transactions consummated on the exchange (e.g., price, volume and counterparties involved, time of the deal). The exchange will publish almost all of the trade information (sometimes at a cost). The names of the counterparties are generally regarded as private and potentially commercially sensitive and will not be included in the data sets. Our ability to analyze the working of financial markets is limited by the availability of relevant historical high frequency data. Market microstructure studies depend on access to high frequency data as well as the use of information technology to store and process the data sets. High frequency data is available for most of the international derivatives and stock exchanges, e.g., Chicago Board of Option Exchange, Chicago Mercantile Exchange. Researchers can purchase historical high frequency data on the internet from websites such as www.tickdata.com (historical intraday data of the international commodity and equity markets) and http://cbotdataexchange.if5.com (the Chicago Board of Trade is a global commodity futures exchange trading treasury bonds, corn, soybean, wheat, gold, silver and more).

There are many exchanges for which such detailed information isn't easily available to researchers. While traditional trading venues involve personal interactions between traders either on exchanges or by telephone, the advent of technology permits the development of electronic exchanges devoid of such interactions. Moreover, with only two exceptions, all new derivative exchanges established since 1996 are fully automated and, increasingly, new stock exchanges are similarly structured (see e.g., Goodhart and O'Hara (1997)).

3.2.2 The South African equity market

Before we start our research with regard to the estimation of the volatility of a small subset of shares on the JSE, it is important to gain an understanding of the mechanics of equity markets in general as well as the market events that have had an effect on volatility.

Equity market structure

Depending on whether current or new capital is being issued, South African equity markets can be conceptually divided into two components, namely primary and secondary markets.

- A *primary market* deals with the creation or termination of shares. Transactions in the primary market affect the size of the available equity market. Examples of transactions in the primary market would be new listings, rights issues, share splits, winding up of existing companies and the buying back by a company of its own shares. In our study, two of the shares selected were impacted by a share split and had to be adjusted accordingly. We will discuss the process in more detail later in the chapter.

- A *secondary market* engages in the trading of existing shares. Transactions in the secondary market affect the flow of money through the market. They also impact share prices and thus market capitalization (market value of the total shares in issue), as well as market liquidity, but do not affect the size of the share pool.

South African equity market participants

There are different market participants active in the South African equity market and their different trading styles will have an effect on the information content/profile of trades. The market participants can be classified as follows:

- *Brokers or Agents* operate solely as agents for buyers and sellers. They do not speculate on price movements in the market as they concentrate solely on their roles as agents for which they receive a commission. Clients place orders at specific levels with their brokers and the brokers will execute the deal at the same level or better depending on the market. The deals may have an effect on volatility as the trades may have been executed at a predetermined level requested by the client

- *Proprietary traders* (merchant banks, institutions, pension funds) trade for their own account. Proprietary trading includes prearranged deals (i.e. deals arranged on the telephone in order to execute the delta associated with an option trade) and these could give rise to periods of time when financial markets are particularly active and volatile. At other times the opposite may be true and the market may be extremely quiet. Proprietary trading and its effect on volatility is difficult to identify due to the anonymity of the trading parties.

Trading system

The JSE utilizes an automated trading system methodology and all trading is fully automated and executed from a trading screen platform. Buyers and sellers transact using the central automated trading system of the exchange. Large investors (who are exchange members) can therefore deal on their own behalf and aren't reliant on brokers to find them a "best price". In contrast, clients who are not members of the stock or derivative exchanges and market participants who would like to trade anonymously will still have to utilize the services of a stock broker or derivative exchange member to trade on their behalf.

High frequency data availability

Obtaining high frequency data for the South African equity market was an initial hurdle to our study. In the South African equity market, outside vendors such as Reuters, I-net bridge and Bloombergs provide access to high frequency data, but due to the large volume of transactions the information is only stored for a limited period. Data vendor high frequency data is usually stored for a period of 3 months or "set number of data entries" depending on the liquidity of the share. The other alternative is to keep a record of all transactions on an in-house server, which is a rather costly and not very feasible solution for the researcher. The JSE, in contrast to their international counterparts, does not keep a database of high frequency equity data and thus availability of high frequency data in the South African equity market was an initial hurdle to our study (we obtained the data used in our study from HSBC). The Centre for Business Mathematics and Informatics, North-West University (Potchefstroom Campus), has subsequently started maintaining a high frequency historical database of the South African equity market.

3.3 SHARES SELECTED FOR THE STUDY

The aim of the present study is the estimation of daily volatility utilizing high frequency data and in order to obtain these estimates it was important to select a subset of shares that are actively traded in the South African equity market. Due to the large amount of data involved in a high frequency data study it was necessary to select a small subset of shares for our study. We selected a subset of six shares based on the following criteria:

- Firstly, the shares had to be liquid (actively traded) with a high average daily turn-over.
- Secondly, they had to represent different market sectors such as mining, banking (refer to Table 3.1).
- Thirdly, they had to be included in the FTSE/JSE Top 40 (ALSI 40/TOPI).

Even within the same trading mechanism (exchange) there can be large differences in the trade outcomes for different shares. When estimating volatility, an area of concern is the pricing of infrequently traded shares. On the JSE the constituents of the FTSE/JSE Top 40 (ALSI 40) are traded in a different sector to the mid cap shares. Due to the infrequent trading nature of the mid cap shares compared to the ALSI 40, these were not included in the study.

All shares selected have a high market capitalisation (number of shares issued multiplied by the market price) and form part of the JSE All Share 40 Actuaries Index (ALSI 40). In practice the ALSI 40 constituents make up the majority of trade value on the Exchange on any particular day. Furthermore, ALSI 40 futures and options on futures are listed on the South African Futures Exchange (SAFEX) thus making arbitrage and hedge trading possible across the equity and derivative markets.

Table 3.1: Subset of six shares selected for the study from the JSE Securities Exchange SA All Share 40 Index (ALSI 40).

Share Code	Share name	Trading Sector
AGL	Anglo American PLC	Other Mineral Extractors and Mines
BIL	BHP Billiton PLC	Other Mineral Extractors and Mines
BVT	Bidvest Limited Ord	Business Support Services
HAR	Harmony GM Co Limited	Gold Mining
RCH	Richmond Securities AG	Household Appliances and Housewares
SBK	Standard Bank Group Limited	Banks

In Table 3.2 we list the average, median, minimum, maximum and quartile statistics of the number of trades per day applicable to the subset of six shares selected for our study. For all the shares selected for our study the median of the number of trades per day is above 99 and we can conclude that they are actively traded on a daily basis. AGL has the highest number of trades per day, followed by RCH, HAR, BIL, SBK and, lastly, BVT. In our study we have included trading days which are considerably less active than others, for example 24 Dec 2000-2003 and 31 Dec 2000-2003 which are traditionally halfday, hence have lower trading volumes than normal.

Table 3.2: Average, median, minimum, maximum and quartile statistics for the number of trades per day applicable to the subset of six shares selected for our study

	AGL	BIL	BVT	HAR	RCH	SBK
Average	323	183	110	149	227	194
Median	329	176	105	99	210	181
Q1	190	136	86	45	164	141
Q3	427	219	125	228	272	236
Min	7	10	6	4	16	4
Max	882	509	335	586	629	501

Some interesting statistics on the price return distribution for subset of six shares selected for our study are given in Table 3.3.

Table 3.3: Price return trading statistics (mean, standard deviation, minimum, maximum and kurtosis) for the subset of shares (AGL, BIL, BVT, HAR, RCH and SBK) applicable to our study for the period 21 Jun 2000-07 Mar 2003.

Share Code	Mean	Std dev	Min	Max	Kurtosis	Skewness
AGL	0.00000	0.0018	-0.1094	0.1114	202.31	0.21931
BIL	0.00000	0.0025	-0.0604	0.1621	181.52	2.48590
BVT	0.00000	0.0019	-0.0339	0.0425	30.81	0.17594
HAR	0.00001	0.0033	-0.1147	0.1420	164.24	0.77934
RCH	0.00000	0.0020	-0.0569	0.0433	37.15	-0.33980
SBK	0.00000	0.0018	-0.0984	0.0394	92.57	-1.03600

From Table 3.3 we note the following:

- The share with the greatest range between minimum and maximum is HAR and the share with the smallest range between minimum and maximum is BVT.

- Secondly all shares selected for the study display highly leptokurtic behaviour (with a kurtosis greater than 3)

- The return distributions for AGL, HAR, SBK and BIL are all skewed to the right. In contrast, RCH and BVT are the only two shares skewed to the left.

Example of tick data utilized in the study

Our primary data set consists of share codes, date and time of transaction, trade code, value, volume and price and trade number for the period 21 Jun 2000 - 7 Mar 2003. The JSE transmits the data in the following format:

Table 3.4: Example of tick data used in the study.

ShrCode	Date	Tcode	TotalValue (Rand)	TotalVolume	Price (Cents)	TradeNo
AGL	00/12/29 15:53	MB	407200	1000	40720	6935
AGL	00/12/29 15:54	MB	81440	200	40720	6960

The fields used in Table 3.4 are described as follows:

- ShrCode: - Share code (e.g., AGL, BIL, refer to Table 3.1)
- Date: - Date and time stamp
- Tcode- Transaction code (refer to Table 3.10 for a full breakdown of all transaction codes used by the JSE)
- TotalValue – Total Rand value of trade
- TotalVolume – Total volume assigned to the trade
- Price – Share price (quoted in South African cents)
- TradeNo – Unique trade code (identifier) assigned to each trade

3.4 DESIGNING A DATA FILTER

In order to design a data filter to filter out erroneous points in our time series it was necessary to develop a good understanding of the data as well as the events (normal and extreme) in a South African equity market context that would have had an effect on volatility, as well as gain an understanding of the underlying mechanics and requirements of a data filter in general. In Section 3.4.1 we discuss the characteristics of data filtering in general (i.e. not specific characteristics applicable to the South African study).

In a South African equity market context we identified three categories/classes which have an important impact on volatility. Firstly, "everyday" market events such as corporate actions, data capturing errors and news releases. Secondly, certain "once-off" market events such as the introduction of STRATE and the change in trading methodology on the JSE. Lastly, our sample of data included the 11 Sep 2001 terrorist attacks, which we classified as an extreme market event. Later in this section we will discuss each of these categories/classes with practical examples from our study in more detail. We conclude the section with a summary of the methodology applied to build a data filter for the shares selected for the South African study.

3.4.1 A general introduction to data filtering

The importance of correct data, and hence the emphasis on filtering erroneous and "suspicious" data, has risen in recent years. Advances in technology have made analysis of large sets of high frequency data more accessible to market participants. Before the analysis of the constructed time series can be undertaken all erroneous data points must be filtered out of the original time series. The latter process calls for a good understanding of the underlying market as the "wrong trading prices" may compromise the results of the statistical analysis by introducing artificial price movements (volatility).

One of the most difficult aspects of filtering high frequency data is identifying erroneous and "suspicious" data points. There may be obvious erroneous data points, such as decimal errors, and there are borderline errors such as losing a fraction of the decimal. The removal of obvious erroneous data points is an easy problem to solve, but the complexity lies in the handling of borderline errors and "suspicious" data points.

Tick data consists of all trades concluded in the market. These trades are associated with a time stamp showing exactly when the transaction occurred. The information is kept in its raw form by exchanges and data vendors and has to be processed before any reliable statistical analysis can be performed. An effective data filter should create a time series for historical research that eliminates erroneous data points in the trader's or researcher's base unit without eliminating important statistical information from the time series.

There are various approaches/strategies a researcher can follow with regards to the treatment of erroneous data points, namely:

- "Search and replace/delete", once an erroneous tick (price) is identified. The erroneous tick could be replaced with the last known value or deleted from the time series. When deleting an erroneous tick (price) the researcher has to decide if the erroneous tick's volume should be assigned to the data point before the erroneous tick or should the volume associated with the erroneous trade be eliminated all together?

- Some data contributors send test ticks (not actual trading data) to the system when the market isn't liquid. Care should be taken, as these ticks may appear "valid" to the filter (if weekends and public holidays haven't been flagged in the filter) and therefore could be incorrectly included in the study. If at the open or close of the trading day no trading has taken place the researcher needs to make the decision if the last trade price is to be used in the time series. An example applicable to our study is given in Table 3.6 below.

Furthermore the filter employed should have the following characteristics:

- It should be computationally fast. This requirement excludes algorithms starting from scratch at each new incoming tick and therefore has to be sequential and iterative. For large data sets computation time is crucial.

- There is a limit to the number of erroneous ticks in succession that can be filtered. A neighbourhood of trades or a filtering window is needed to perform a check with regard to the feasibility of a trade and how many erroneous ticks will be allowed. In other words, is the change in price due to an erroneous data point or is the change in price due to an actual change in price (e.g., looks suspicious but should not be filtered out)

- The cause of data errors is rarely known; therefore the validity of a quote is judged according to its plausibility given the statistical properties of the series and common sense.

- For a specific share several trades may take place at the same time stamp but be executed at different levels. The filter has to take this characteristic into consideration. In order to simplify the large amount of high frequency data applicable to our study, a decision was made to amalgamate trades that took place simultaneously. If more than one trade took place at the same time e.g., at 9:01, the VWAP (volume weighted average price) for all the trades was calculated. Later in Section 3.4 the latter process and examples applicable to our study will be discussed in more detail.

- One of the most important tasks when preparing data for analysis is the process of generating a time price series from the original data file/source. The definition of data outliers or erroneous data points is also dependent on the time series used; a higher frequency data error may be insignificant to a trader using hourly data. In Figure 3.2 we graphically depict a high frequency tick data price series compared to a 5 minute price series for AGL. The "5 minute" data series was compiled using the Andersen et al. (2001) nearest neighbour methodology (refer to Section 3.4). If the latter methodology is utilized the noticeably higher price between 9.15 and 9.20 in Figure 3.2 would have been excluded in the Andersen et al. (2001) 5 minute price series. The 25 minute price series sequence contains a "data outlier" when used by a tick trader, which wouldn't have had an impact on a lower frequency trader.

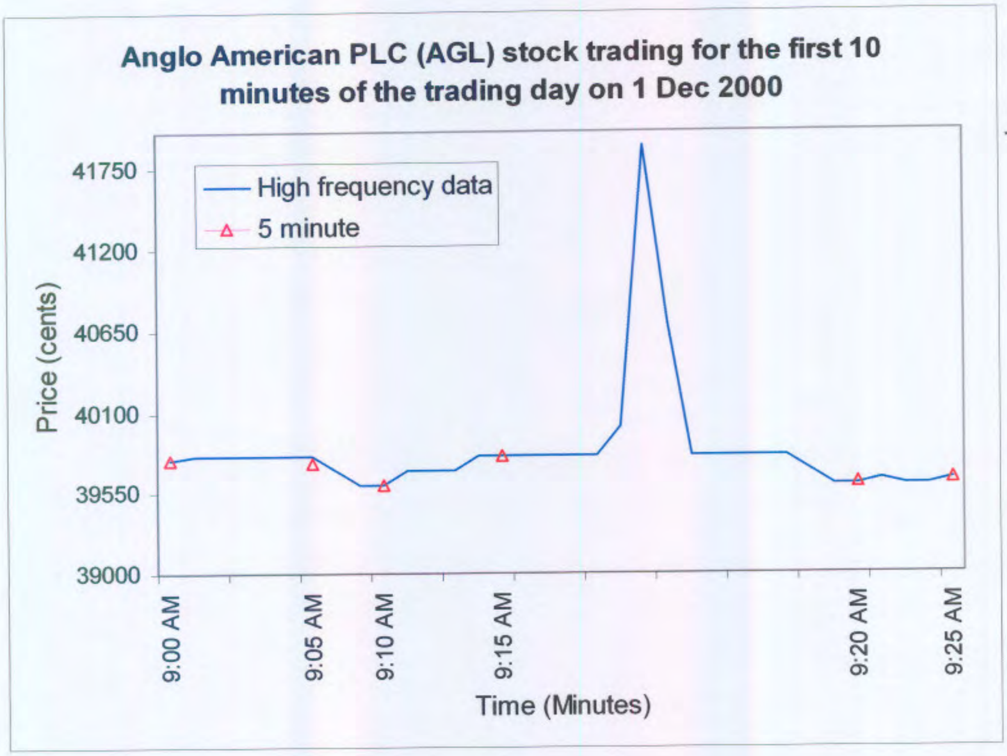


Figure 3.2: Anglo American PLC (AGL) high frequency tick data price series compared to a 5 minute price interval. The 25 minute data sequence contains a "data outlier" when used by a tick trader, which wouldn't have had an influence on a lower frequency trader.

- Borderline data filtering is a difficult and tedious task and the researcher must be able to differentiate between a market crash, market correction and a incorrectly recorded trade. There is no perfect solution to the problem. One has to rely on statistical procedures as well as on a "human factor" to flag the erroneous data points and exclude them from the data sample. The latter is easier said than practically applied, especially when large volumes of data have to be filtered (as is the case with high frequency data).

Filtering of borderline errors is a tradeoff between under-and over-scrubbing of the time series. If the time series is filtered too loosely or "under-scrubbed" the data could still contain many erroneous data points and the resulting time series will be unusable and unreliable for statistical analysis. On the other hand, if you over compensate and "over-scrub", you increase the possibility that you over compensate for potential errors, thereby taking the "reality" out of the data and changing its statistical properties. Defining borderline errors is the crux in the tradeoff between over and under scrubbing the available data. This process calls for a good understanding of the market and the underlying mechanics thereof. In developing a set of tick filters the primary objective is to maintain a balance between over-and under-scrubbing the time series in such a manner as to produce a time series that removes outliers but is still statistically reliable (see e.g., Falkenberry (2002)).

While data filtering should take place at the tick level, filter parameters should remain a function of the trader's base unit (tick, hourly, daily data) of analysis. Data may appear to be filtered of all erroneous data points at a higher frequency, but by focusing into a finer time interval the data may contain erroneous information that is unusable for reliable research. In order to have reliable data it is necessary to filter data at its finest granularity. Depending on the volume of data being filtered, the researcher may utilize a computer algorithm or decide to filter the time series manually. When using a computer algorithm to filter out erroneous data points the following factors must be taken into consideration:

- Time efficiency – especially if large volumes of data must be filtered

- The accuracy of the resulting time series is debatable. Computer algorithms utilize programming rules to "scrub" data. As an example - if the "data outlier" is more than "X percent" away from a previous data point the data point will be removed. If the algorithm doesn't take the preceding data into consideration the latter rule may lead to over or under scrubbing the resulting data series.

3.4.2 Designing a data filter for the South African study

In order to keep the human touch and intuition with regards to the identification of "suspicious data" points and the filtering of erroneous data points in the data series for the shares selected for the study, a decision was made that the data would be filtered manually. We begin by summarizing the market characteristics/events that have an important impact on volatility, namely every day, once off and extreme market events (which don't occur on a daily basis).

Every day market events

- Data capturing errors such as decimal place errors (trades captured in Rand and not cents will dramatically alter the volatility estimates and characteristics of the time series).

- Macro economic news releases - news is a very broad concept and therefore difficult to quantify. News can range from a phone call to a client, economic forecasts of the research department of a bank, general economic and political news as well as major economic news announcements (the transparency of the news announcement will have an effect on the share price).

- Many markets impose limits (volatility bands) on the amount (percentage) assets prices can change within a trading day to prevent the market from overreacting and, hence, to dampen volatility in the short term.

- Corporate events (e.g., rights issues, splits and consolidations) have the effect of dramatically altering the price at which a share trades. In the case of a share split it was necessary to back-adjust the time series to incorporate the share split otherwise incorrect volatilities will be created.

- The relationship between volume and volatility is important due to the information content associated with volume. Also, volume and the underlying associated price movement thereof is thought to be a good estimator of volatility.

Once-off market events

- The JSE introduced a new electronic settlement system and custody system known as Share Transactions Totally Electronic (STRATE) in 2001. The move to the STRATE system involved many fundamental changes from the paper based settlement system (where settlement took place once every seven days), to a rolling 5-day electronic settlement system.

- Introduction of a new trading methodology on the JSE. In May 2002 the JSE introduced a new trading methodology, as well as a new trading platform (e.g., TALX) to replace the previous trading system (JET). The trading model provides for continuous automated trading, open and optional intraday and closing auctions, static and dynamic price monitoring as well as an auction or VWAP (volume weighted average price) closing price calculation.

Extreme market events

- Non-economic world events such as terrorism and war also had an impact on local and international financial markets. Financial markets worldwide displayed considerable turmoil in the weeks following the terror attacks of 11 Sep 2001.

The statistical distributions and characteristics of the various volatility estimators as noted in Chapter 2 will be discussed in more detail in Chapter 4. In Sections 3.4.3 to 3.4.5 we will be discussing the above mentioned characteristics and their influence on volatility in more detail.

3.4.3 Every day market events

3.4.3.1 Data errors in the time series

In the next section we note and discuss examples of data errors we found in the South African study:

Data errors due to the trading methodology followed by the Exchange

In electronic markets (JSE, London Stock Exchange) the buy and sell orders are matched electronically. All electronic exchanges keep records of transactions consummated on the exchange, the price, volume and the counterparties involved and estimates of the time of the transaction. Data is collected electronically in sequence and released to data vendors and the markets. In a South African context all the orders are entered electronically and data transmission errors may still occur as traders manually enter prices into the trading system and as a result decimal place errors occur.

Data transmission errors

High frequency data strings are commercially transmitted by the exchange or data vendors as a parcel of real time information to data users. The data users are mostly professionals who know the context as well as if and when erroneous data is transmitted (trade out of sequence or at an incorrect price); the professional users will immediately understand and implicitly filter the data by using the information they have available.

The situation changes if the users are researchers or when computer algorithms extract information from high frequency quotes. If erroneous trades are used the results of the study may be tainted and not valid. Most researchers have found that high frequency data contain erroneous data and data filtering is necessary, in order to ensure reliable results.

The source of errant data points is difficult to assess. The root of the problem can be traced to the speed and volume at which the data is generated as well as human intervention between the point of trade and data transmission. Erroneous data strings originate from the process inherent to trading. Various scenarios can arise in which trades are reported out of sequence, cancelled, replaced or reported in error. To identify the latter may not always be a simple task and the data preceding the erroneous data point must always be taken into consideration.

There are many causes for data errors. Human errors include isolated erroneous ticks, multiple erroneous ticks in succession, decimal errors, transposition errors, typing errors and the loss of the decimal portion of a number. Table 3.5 gives examples of decimal place errors from the unfiltered HAR time series used in the study. If the incorrect data points aren't filtered with care the resulting volatility analysis will be compromised and unreliable.

Table 3.5: The table contains examples of two decimal place errors for Harmony GM Co Limited (HAR) on different trading days (the data points are consecutive trading points).

Date	Data point 1	Data point 2	Data point 3
27 Dec 2002	156.70	1.56	156.00
		Decimal place error	
7 Feb 2003	120.60	12.54	121.30
		Decimal place error	

Figure 3.3 depicts examples of data outliers, which may be due to human errors or decimal place errors. There were numerous data outliers or data transmission errors that had to be investigated and filtered. The extreme data outlier on 29 Jan 2003 was caused by a decimal place error; the trader erroneously typed in the price as R1300 where it should have been entered as R130. The same explanation holds for the lower outliers; instead of typing a price of R130 on 30 Jan 2003, the trader entered a price of R0.13. The time series still includes the 4:1 split on 8 May 2001, which will be discussed in the Section 3.4.3.2.

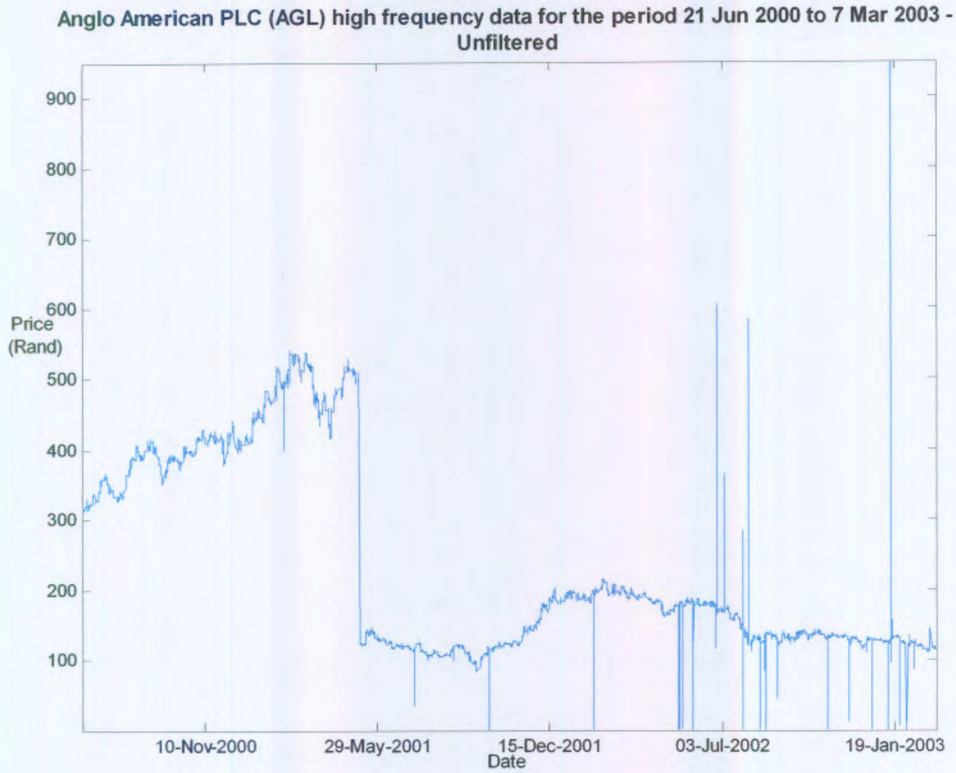


Figure 3.3: Anglo American PLC (AGL) high frequency equity data graph for the period 21 Jun 2000 – 07 Mar 2003.

Computer system failures cause system errors. Some data contributors send test ticks to the system in order to test if their software is operating optimally. These quotes can be identified by their time stamps as the vendors usually perform the tests after hours when the market is not open. In the AGL data the following example of a test tick was found:

Table 3.6: The table contains an example of an incorrect data point for Anglo American PLC (AGL). The data point was captured on 2 Feb 2003 a Saturday, which is not a valid trading day.

Date	Data point 1
2 Feb 2003	60

There are two aspects of the above time series which made the data point easy to identify. Firstly, 2 Feb 2003 was a Saturday (not a valid trading day) and secondly, there was only one data entry on the specific day. The data point shown in Table 3.6 was the only example of a weekend data point found. Furthermore, we eliminated all non trading days such as Saturdays, Sundays and the various South African public holidays.

3.4.3.2 Equity corporate actions/events

Corporate events (e.g., rights issues, share splits and consolidations excluding ordinary dividends) have the effect of dramatically altering the price at which a share trades. Consolidations cause the price to rise as there are now fewer shares in issue, while splits cause a drop in the price as there are now more shares in issue. An example of a share split found in our data is shown in Figure 3.4.

Figure 3.4: Anglo American PLC (AGL) high frequency data graph for the period 21 Jun 2000-7 Mar 2003. The data has been filtered for potential erroneous data points as discussed in Section 3.2.1 (decimal place errors, trade types etc.). Note the dramatic change in the share price (8 May 2001) due to the 4:1 share split.

Table 3.7 lists corporate events applicable to the shares selected in our study.

Table 3.7: Table of corporate events applicable to the shares selected for the study.

Code	Date of corporate event	Corporate event
AGL	8 May 2001	4:1 split
BIL	None	
BVT	None	
HAR	None	
RCH	16 Nov 2001	10:1 split
SBK	None	

The price movements due to the corporate action can be dramatic and the time series must be adjusted to incorporate the price adjustment that occurs between trading days to avoid the introduction of artificial volatility movements. In Table 3.8 we show an example of AGL price data before the 4:1 split and after the data has been adjusted (divided by four) for the share split.

Table 3.8: Example of the Anglo American PLC (AGL) price data on 21 Jun 2000 before the share split and after the share split.

Date	Price (Before split)	Price (After split)
6/21/00 9:27	317.71	79.43
6/21/00 9:28	317.80	79.45
6/21/00 9:33	317.82	79.46
6/21/00 9:40	317.60	79.40
6/21/00 9:49	317.20	79.30

In Figure 3.5 the AGL graph in Figure 3.4 has been adjusted to incorporate the share split on 8 May 2001.

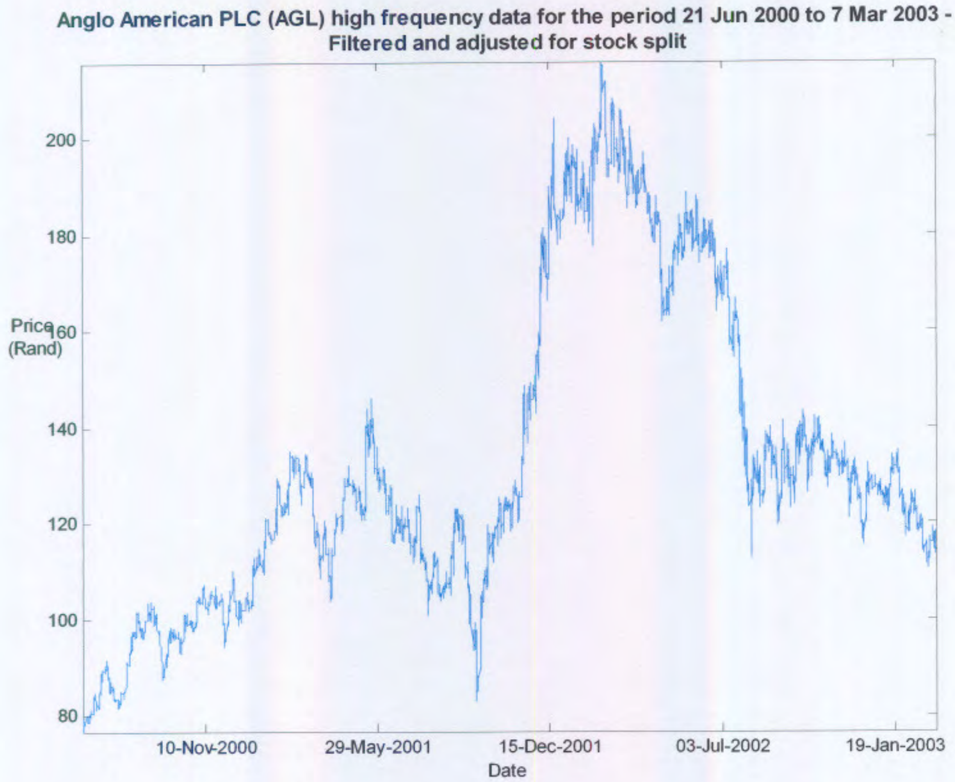


Figure 3.5: Anglo American PLC (AGL) high frequency data graph for the period 21 Jun 2000-7 Mar 2003. The data has been adjusted to incorporate the 4:1 share split.

3.4.3.3 Every day market events that influence volatility

Some factors that influence volatility will now be discussed. These factors may give rise to periods of high and low volatility and even outliers. Because these observations are inherently part of the process generating volatility they should not be filtered out of the time series.

Macro economic news releases

The recent literature on microstructure of financial markets has produced models that treat the arrival of public information or "news" as an important factor, with models differing in institutional trading characteristics, such as the presence of informed and uninformed traders (see e.g., Goodhart and O'Hara (1997)). News is a very broad concept and therefore difficult to quantify. Examples of news events are: a phone call to a client, economic forecasts of the research department of a bank, specific news from a company (e.g., profit warnings and financial statements) as well as general economic and political news announcements.

Although it is evident from both theory and stylized facts that the intensity of information flow impacts the level of market volatility, it is in general difficult to explicitly identify the information arrival process underlying systematic intraday patterns. The latter has been well researched and numerous studies have been performed regarding the effects of public news releases. Ederington and Lee (2001) found that regularly scheduled US macroeconomic news releases lead to significant time-of-day and day-of-week patterns in the volatility of US Treasury and Foreign Exchange futures.

In contrast, Cutler et al. (1989) find little association between the largest daily price changes in the US stock markets and readily identifiable economic news. Mitchell and Mulherin (1994) show that the number of Dow Jones announcements is only weakly related to the US stock market volume and volatility, while Berry and Howe (1994) report no significant relationship between US equity volatility and the total number of Reuters news releases.

Ederington and Lee (2001) compared scheduled and unscheduled news announcements in the interest and foreign exchange markets. Unscheduled news announcements showed higher volatility persistence than scheduled news announcements. Foreign exchange markets, in contrast to the bond markets, display pronounced intraday volatility patterns, but have smaller, yet still significant volatility responses to macroeconomic announcements (see e.g., Andersen et al. (2000)). The findings of a U-shaped volatility pattern and a rather low sensitivity to macroeconomic releases seemed ubiquitous across US data studies in the equity markets.

Andersen and Bollersev (1998), in their study of the foreign exchange market, take the first step towards characterizing the explanatory power of intraday patterns, macroeconomic news releases and long term volatility factors. Tests for significant volatility responses to specific events are reliable only if the researcher can control appropriately for the intraday pattern. The intraday pattern may only be extracted with precision if one accounts for scheduled macroeconomic announcements. A large quantity of information regarding the current state of volatility is contained in high frequency data, but the relationship with the intraday pattern as well as macroeconomic news releases should be taken into account.

Volatility clustering is related to both the arrival and transmission of news. It has been argued that items of news that have a large impact on prices have a tendency to be clustered together, as do items that have a small impact on prices (see e.g., Antoniou et al. (1998)). The role of transmission of news in explaining volatility clustering relates to market dynamics. Differences in investor's expectations may take some time to be eradicated and some traders may have access to insider information which may filter into the market at different stages.

The timing of news announcements may also have an impact on volatility. Some news announcements are specifically released outside of normal trading hours and this tendency reflects the market expectation that particular releases are likely to have a disruptive impact on the market. On the other hand, the market may be aware of news releases beforehand and price the effect of the news in the current market price, thus incorporating the potential price expectations. Some examples are: news announcements with predetermined fixed dates e.g., CPI, interest rate adjustments and the release of financial statements.

In our study we found it difficult to specifically attribute higher and lower volatility estimates to macro economic news events. It is not humanly possible to "back date" potential price movements and link the price movement to a "phone call" or research forecast of a bank if the researcher doesn't have access to the information. However, knowledge of events like these may aid significantly towards identifying outlying observations as errors or not.

Effects of price limits

Many markets impose limits (barriers) on the amount (percentage) by which assets prices are allowed to change within a trading day to prevent the market from overreacting and, hence, to dampen volatility. Examples of financial markets that impose price limits are Austria, Belgium, France, Italy, Japan, Korea, Malaysia, Mexico, the Netherlands, South Africa, Spain, Switzerland, Taiwan and Thailand (see e.g., Cho et al. (2003)).

Share price limits have two attributes that decrease the price volatility:

- The limits literally set a ceiling and floor within which the price can move in a trading day.

- Price limits provide a cooling off period (limits market participant over-reaction, but does not interfere with trading activity).

Fama (1989) argues that price limits can have adverse effects on the equity market:

- The *Volatility Spillover Hypotheses* states that price limits will increase the volatility on subsequent trading days because the limits prevent one-day price changes and immediate corrections.

- A second potential problem with price limits is a delay in price discovery. In the *Delayed Price Discovery Hypotheses* Fama (1989) argues that trading price constraints prevent trading prices from reacting to new information arrivals as well as reaching new levels of equilibrium.

- Thirdly, if the share price hits a limit, then the share becomes less liquid and the trading will be heavier on the following days. Lee et al. (1994) found that trading halts increase both the volatility and volume for the New York Stock Exchange when trading resumes.

- The *Magnet Effect* (Fama, 1989) suggests that prices accelerate towards the limit, as it gets closer to the limits. There are two reasons proposed for the magnet effect: illiquidity and behavioral investors.

- The *Cool-off Effect* (Fama, 1989) is the opposite of the Magnet Effect. Price limits may cause a halt in trading and the market will have time to reassess the fundamental value to counter the overreaction, if there is any in the market.

In the South African equity market as well as the agricultural derivative markets the price limit principle has been applied:

  - The JSE implemented equity volatility trading bands in May 2002. The volatility auction period can be described as follows:

"An auction call period which occurs during automated trading when an order is entered that would execute at a price that is a percentage, as specified by the JSE from time to time, or more away from the reference price". If and when the share volatility band is triggered an automatic auction period will occur and trading will be halted for a period of 10 minutes only (see JSE Equities markets and trading functionality overview).

- The Agricultural Products Division of the JSE applies the following methodology to daily price limits (see JSE Notice to members, A043, 2003):

  "If one or more White and Yellow maize contracts trade at the predetermined price limits (up or down) for two consecutive days, the daily price limit will be extended from R45 per ton to R65 per ton and will remain at extended limits until the daily price movements on all like contracts is equal to or below R45 per ton. "

  An interesting possibility for future research would be to investigate if the imposed price limits do have the desired effects of prevention (market overreaction) and dampening of volatility. The latter effects may not always be in the market's best interest as artificial volatility levels are introduced and market participants may be forced to execute deals at unrealistic market levels.

The interaction of volume and volatility

The traded volume of an instrument is not uniformly distributed over the trading day as certain periods of the trading day will be more liquid (actively traded) than others. These variations in volume are due to the release of economic figures (macro economic news releases), market participant activity (orders in the market by institutions etc.) and unexpected clusters of information.

Volume of a trade is, loosely speaking, inversely related to the time between trades and the underlying share price movement on a specific day is dependent on whether the volume is high or low. In other words, if a market player has a large size order to execute and there is a lack of liquidity in the market it would possibly create a problem of price slippage due to the lack of liquidity. For a trader the distribution of volume through the trading day is important. Firstly because of the information content associated with volume and, secondly because, volume and the underlying associated price movement is thought to be a good estimator of volatility. The indication of market depth (volume) might assist the trader in executing orders at better levels (see e.g., Lequex (1997)).

Volatility may be significantly higher at certain periods in the trading day, e.g., at the opening in the morning session of trading and at the close in the afternoon leading to a U shaped volatility pattern. On the New York Stock Exchange the volume of deals, the volatility of equity prices and the spread between bid and ask quotes all broadly follow a U shaped pattern. All three variables are at the highest point at the opening, fall to lower levels during the mid-day and rise again towards the close. Andersen et al. (2000) show that the Nikkei 225-index volatility is significantly higher at the opening in the mornings and the close of the afternoon sessions than during the morning and mid afternoon sessions (double U shape).

The temporal intra-day pattern is not easily explained theoretically (using a basic model that splits agents in the market into informed, uninformed and market makers). One expects uninformed liquidity traders, with discretion over the timing of their trades, to trade in time periods when trading costs are low. Given the resulting market depth and subsequent liquidity, privately informed traders would also want to trade in such intervals in order to remain anonymous. Nevertheless more information is subsequently revealed and the prices become more volatile (see e.g., Goodhart and O'Hara (1997)). In the financial markets greater volatility is associated with two factors, namely the "revelation" of information, as well as market uncertainty.

Similarly Piccinato et al. (1998) found that intraday tick activity in Euro futures displays a U shape as well as evidence of a day of week effect (similar to that documented for stock indices). Another interesting finding was that Euro futures trading activity displayed the lowest trading volumes on Mondays and the highest trading volumes on the last two working days of the week.

Prior to the implementation of STRATE the JSE settled all equity trades on a fixed day of the week (Tuesdays) for the previous weeks trades (Monday to Friday). An interesting aspect for future research would be to analyze if a day of week pattern or a "U shape pattern" would be applicable to South African equities. Two different timelines would be applicable to this study:

- The trading period prior to the implementation of STRATE. If a day of week effect is to be studied in the South African equity market care should be taken when analyzing data before STRATE was introduced.

- The trading period after the implementation of STRATE.

3.4.4 Once-off market events

In this section we will discuss two events that have had an important impact on the volatility calculations in our study. Due to the infrequent nature of these events they can be classified as once off occurrences, namely the introduction of STRATE and the change of trading methodology on the JSE.

3.4.4.1 The introduction of STRATE

The JSE introduced the new electronic settlement and custody system known as Share Transactions Totally Electronic (STRATE) in 2001. The move to the STRATE system involved many fundamental changes from the paper based settlement system (where settlement took place once every seven days). STRATE achieves secure, electronic settlement of share transactions on the JSE on a rolling T+5 (trade day + 5 business days) basis. All paper shares were dematerialized (conversion from paper share certificates to an electronic record keeping system) in different stages (starting in 1999 and completed by the beginning of 2002) and settlement now takes place on a contractual rolling settlement (T+5) basis. The dematerialization dates applicable to the shares selected for our study are shown in Table 3.9.

Table 3.9: An example of the dematerialization schedule for the subset of shares selected for the study.

JSE Code	Share	Dematerialization date
AGL	Anglo American PLC	22 Oct 2001
BIL	BHP Billiton PLC	21 Jan 2002
BVT	Bidvest Limited Ord	3 Sep 2001
HAR	Harmony GM Co Limited	25 Oct 1999
RCH	Richmond Securities AG	15 Oct 2001
SBK	Standard Bank Group Limited	24 Dec 2001

Prior to the introduction of STRATE, the physical settlement of the net trades (buys and sells of each stock) concluded from a Monday to a Friday of each week took place on the subsequent Tuesday. If a market participant did not wish to take physical delivery of stocks traded they would ensure that they buy and sell the same quantity of each stock during the week. Due to the increased trading activity by market participants managing this process on Fridays, trading activity would increase therefore we would expect higher volatilities on these days.

3.4.4.2 Change in trading methodology JSE Securities Exchange SA

In May 2002 the JSE introduced a new trading methodology, as well as trading platform to replace the previous trading system (JET). The new trading model is very similar to the European Alliance Market model (1999) and is the preferred model for the operation of an order book to support trading in liquid securities. The trading model provides for continuous automated trading, open and optional intraday and closing auctions, static and dynamic price monitoring as well as auction or VWAP (volume weighted average price) closing prices (see e.g., JSE equities market and trading functionality overview (2003)).

The new trading methodology included the introduction of a volatility auction period as well as a variety of new trade types that replaced the previous trade types. There are certain times when unusual events occur in auction call periods and to minimize their impact and to ensure optimal price formation and execution the events need to be brought to the attention of the market. This will be achieved by the JSE through implementation of auction call extension periods. There are two types of auction extensions:

- If the execution price lies outside a predetermined tolerance level a price monitoring extension that lasts for a period of 5 minutes will be implemented in order to increase the likelihood that the price movement may be reduced to more acceptable prior levels.

- If market orders within the order book are not executable or only partially executable at the end of an auction call period, the call period could be extended for an additional two minutes.

The effect of the volatility auctions must be taken into account when studying volatility of the individual shares and it may be necessary to divide the study into two categories e.g., before and after the implementation of the new trading methodology. The official trading hours (after May 2002) of the JSE are 09h00 to 17h00 and prior to May 2002 the market opened at 09h00 and closed at 16h00. The trading day is divided into phases to facilitate liquidity, price formation and market integrity through volatility interruptions. The JSE trading system (new trading methodology) operates on every business day according to the following standard periods and times:

- Market opening period (08h30 to 08h35)

- Opening auction call period (08h35 to 09h00)

- Automated trading period (09h00 to 16h50)

- Intraday auction call period for selected securities (12h00 to 12h15)

- Closing auction call period (16h50 to 17h00)

- Runoff period (17h00 to 18h00)

- System close (18h00)

The JSE uses various trade types to define different trading activities in their database, e.g., trade type "OX" describes options exercises and "MB" is the main board indicator. For a full breakdown of all different trade types used please refer to Table 3.10. For the purpose of our study the decision was taken to only include MB (main board) trade types prior to May 2002 and AT (Automatic trades) after May 2002. The reasons being as follows:

- Firstly, computational time and data management had to be taken into consideration.

- Secondly, if all trade types were included the results may have been unreliable. Not all non main board and automatic trades are reported in sequence. As an example, if a share trades at an incorrect market level or price (usually decimal place errors) the surveillance department of the JSE will request the market participants involved in the incorrect trade to reverse the incorrect trade (trade type = CT for a correction trade) and to re-book the trade at the correct price (trade type = AT). Care must be taken when analyzing the resulting time series as the incorrect trade as well as the correction trade would have to be filtered out. The data filter implemented has to be sensitive with regards to the different trade types and the effect they may have on the data. In Figure 3.6 there are three distinct suspicious trades or outliers. On further investigation we found the three trades to be of trade type ST. ST or special trades are equity trades that are traded cum dividend (including the dividend) and only reported when the share has gone ex dividend (after the dividend has been paid or received by the holder of the share).

Figure 3.6: Anglo American PLC (AGL) trading for the last 10 minutes on 27 June 2000, the three outliers in the graph are “special trades” (ST).

Table 3.10: JSE Securities Exchange SA (JSE) equity trade types used for trade reporting.

Code	Description	JSE definition
AS	Asset swap	A transaction, which complies with all the asset swap requirements of the SARB.
AT	Automatic trade	Automatic trade.
BB	Broker to broker trade	Broker to broker trade.
BT	Block trade	A transaction where a broking member trades as an agent in a single security where the transaction has a minimum value of R5 Million; and compromises at least 100% of the average daily value.
CF	Corporate finance trade	A transaction which must be entered into in writing; requires public notification in the press; and complies with the requirements of transaction categories of the listing requirements of the JSE.
CT	Correction trade	If the price entered on an order which results in an error is 10% or more away from the previous reference price or if there were no trade on that day, the previous business day's closing price.
LC	Late cancellation trade	If conducted after hours must be reported to the JSE trading system within 15 minutes of the market-opening period the next day.
LP	Late purchase	If conducted after hours must be reported to the JSE trading system within 15 minutes of the market-opening period the next day.
LS	Late sale	If conducted after hours must be reported to the JSE trading system within 15 minutes of the market-opening period the next day.
LT	Late trade	A transaction where a broking member trades after hours with a professional market participant, as agent or principal in fulfillment of: 1) an order already entered into the JSE trading system which reflects a reasonable price at which the client wishes to trade. 2) an order received prior to the end of closing auction call period or an order received after hours.

Code	Description	JSE definition
MB	Main board	Main board.
OC	Overseas counterparty	Overseas counterparty.
OL	Odd lot	Share quantities less than 100 shares.
OP	Off order book principal trade	A transaction where a broking member trades as a principal in a single security where the transaction has a minimum value of R500 000; and compromises at least six times the normal market size.
OX	Option exercise	Exercising an option.
PF	Portfolio trade	A transaction where a broking member trades as an agent in a portfolio of shares.
ST	Special terms trade	Special terms trade.
UT	Uncrossing / auction trade	A transaction matched automatically in the JSE trading system during auction matching.
NX	Namibian stock exchange	Trading in Shares listed on the Namibian Stock Exchange.

Nature of trading and different trading classifications

The nature of trading can cause various scenarios wherein trades are reported out of sequence. For example, equity OTC options are deals that are concluded on a principal to principal or agency to agency basis between the option writer and the buyer. The options have tailor made specifications and are mostly written on the more liquid shares. Depending on the contract specification the options may be exercised in cash or equities.

Option exercises may be misleading as they could look like possible data "outliers" as the strike price of the option may be far from the market trading levels at exercise. The trades can be identified from the database by the trade type "OX" (Option exercise) or "OD" (Option delta) and can therefore be eliminated for the purpose of the study.

The data filter implemented has to be sensitive with regards to the different trading classifications (option exercises, contra trades, late trades, portfolio trades – please refer to Table 3.10 for more examples) and the effect they may have on the data. Due to the contractual nature of some of the trade types they may distort the data and have a negative impact on the meaningfulness of the study.

The following table illustrates an AGL time series example of a potential outlier due to a different trade type.

Table 3.11 Extract of a high frequency time series data for Anglo American PLC (AGL) on 4 Feb 2002. Note the data outlier (data point 2) – could be an example of an option exercise (different trade type).

Date	Data point 1	Data point 2	Data point 3
4 Feb 2002	198.40	130.00	197.86

The JSE uses various trade types to define different trading activities in their database, e.g., trade type "OX" describes options exercises and "MB" is the main board indicator. For the purpose of our study the decision was taken to only include MB (main board) trade types prior to May 2002 and AT (Automatic trades) after May 2002.

3.4.5 Extreme market events that influence volatility

Non economic world events such as terrorism and war also have an important impact on local and international financial markets. Financial markets worldwide displayed considerable turmoil in the weeks following the terror attacks of 11 Sep 2001 and the pattern of price changes were typical of crises in many ways: sharp changes in asset prices, increases in volatility, and changes in correlation patterns (see e.g., Malz and Mina (2001-2002)). Most global equity markets were already experiencing losses at the time of the attack and dropped sharply right after 11 Sep 2001 (refer to Figure 3.7).

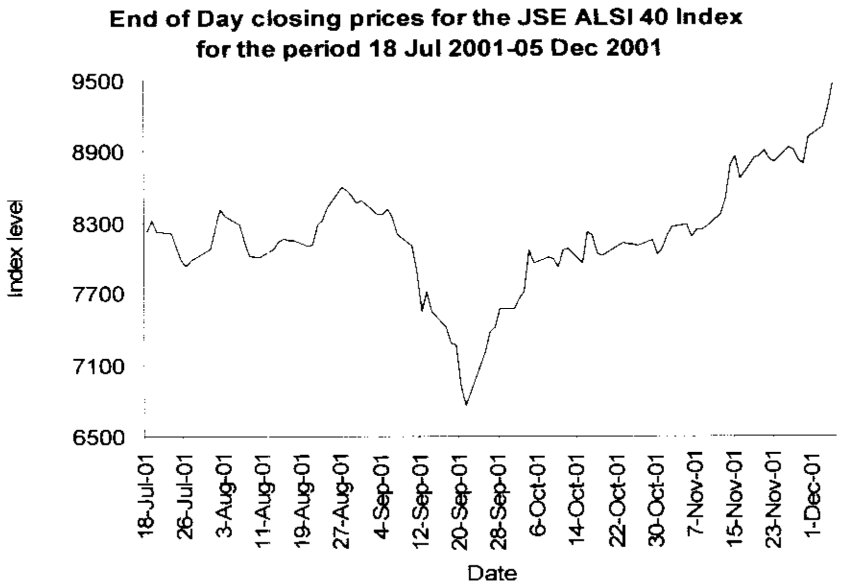


Figure 3.7: JSE ALSI 40 Index end of day closing prices graph for the period 18 Jul 2001 to 05 Dec 2001. The Index level reached the lowest point on 21 Sep 2001 (index level of 6753.065) and only showed signs of recovery at the beginning of Oct 2001.

The market moves during this period in 2001, by comparison with October 1987, were not that extreme. In Table 3.12 we list the percentage move in the international as well as the South African equity markets in the week prior to the terrorism attacks as well as two and three weeks after the attacks. In the beginning of September (one week before the attacks) International markets were already in decline in comparison to the market levels in August of that same year and the terror attacks accelerated market trends that were in place prior to the attacks. The declines in equity prices were particularly sharp in developed countries and the equity markets rallied (upward trend) across the globe during the attacks (see e.g., Malz and Mina (2001-2002)). The latter authors also state that equity volatility reached a high during the last week of September, however, with the exception of Germany; volatilities were lower than those at the peak of the Russian crisis. During the first two weeks after the terrorism attacks the markets declined to even lower levels in comparison to the opening levels of September. Only three weeks after the attacks did the international and South African equity markets begin to recover to the beginning of September market levels.

Table 3.12: Comparison of market conditions prior to and after the 11 Sep 2001 terrorist attacks.

11 September 2001 - Terror attacks			
Index	**Market levels in the week preceding**	**First two weeks after the attacks**	**Three weeks after the attack in comparison to pre attack levels**
S&P 500	↓ 4 %	↓ another 12 %	↓ 5 % from pre attack levels
CAC 40	↓ 5 %		
DAX	↓ 7 %	↓ over 20 %	↓ 9 % from pre attack levels
MIB 30		↓ over 20 %	↓ 7 % from pre attack levels
FTSE 100	↓ 5 %		
ALSI 40	↓ 6 %	↓ 6.6 %	↓ 8.2 % from pre attack levels

3.4.6 Summary of the South African data filtering methodology

Using the above factors that influence volatility as a foundation, we set a benchmark for our six shares in order to determine if a "jump" in price in comparison to the preceding data (price) point was due to a erroneous data point or to a market event (e.g., corporate actions, news etc.). Setting up benchmarks for the different shares was no easy task and after consulting various equity traders the decision was made to use a hundred percentage price move (price is quoted in cents and not Rands) as a benchmark. We began filtering the data by calculating a hundred percent price move from the previously traded price for each data point in the time series. Every hundred percent or greater data point was investigated and included in or eliminated from the study based on the following methodology:

- The percentage move error could be caused by the trader entering the price in Rand and not cents (as required by the JSE). We found the majority of the hundred percentage or greater moves to be due to decimal place errors as discussed previously. If the erroneous data point was due to a decimal place error e.g., price entered in Rand and not cents as prescribed by the JSE the erroneous data point was deleted from the time series.

- Equity corporate actions also had to be taken into account (large price movements from Friday to Monday when the trading price would include the corporate action adjustment) and the time series adjusted accordingly. In our data we had two shares to which a share split was applicable, namely AGL (4:1) and RCH (10:1). For both these shares the share prices preceding the share split had to be adjusted accordingly.

- We experienced database problems on 13 Mar 2001 and 23 Oct 2002 where almost no trades were recorded for the subset of six shares. One possible explanation is that the problem could have been caused by malfunctioning of the in-house data capturing server. Due to the unreliable nature of the data on these two trading days the decision was made to exclude the data from our study.

We refer the reader to Chapter 4, Section 4.2 where we completed one last check to ascertain if the data had been filtered of all erroneous data points and that any remaining data outliers were due to market events and not erroneous data points.

3.5 TIME SERIES CONSTRUCTION

In this section we will discuss the construction of the various time series utilized by the different variance estimators discussed in Chapter 2.

General data preparation issues

We made the following assumption in order to simplify data processing. If more than one trade occurred at the exact matching time stamp e.g., 9.00.01 the VWAP (volume weighted average price) for all the matching entries was calculated using the following formula:

$$VWAP = ((P_1 * V_1) + (P_2 * V_2) + \dots + (P_n * V_n))/V_T, \qquad (3.1)$$

where

P_t = the price of trade t,

V_t = the volume of trade t and $V_T = \sum_{t=1}^{n} V_t$ the total volume of all trades

Table 3.13 Examples of Anglo American PLC (AGL) tick by tick data used in a VWAP calculation applicable to our study.

Share code	Time stamp	Volume	Price
AGL	6/21/00 9:27 AM	100	79.00
AGL	6/21/00 9:27 AM	200	79.45
AGL	6/21/00 9:27 AM	1500	79.45
AGL	6/21/00 9:27 AM	300	79.45

From Table 3.13 and Equation (3.1) an example of a VWAP calculation applicable to our study would be as follows:

$$VWAP = ((79*100) + (79.45*2000))/2100$$
$$= 79.42857$$

For each entry in our time series where simultaneous trades took place these were replaced by the single VWAP entry.

Table 3.14 highlights the different input data series required by the various volatility estimators discussed in Chapter 2.

Table 3.14: Input data required by each model in order to construct the realized volatility estimates.

Methodology	**Input data intervals**
SSR estimator	5 minute data intervals constructed on a nearest neighbour principal.
ABDE estimator	5 minute data intervals constructed on a nearest neighbour principal and a MA(1) smoother.
VARHAC estimator	All available high frequency data (tick by tick data)
NIG-GARCH estimator	End of day closing price data
GARCH estimator	End of day closing price data

<u>Data preparation for the ABDE volatility estimator</u>

We begin by discussing the data preparation steps followed in order to prepare the 5 minute "intra" data time series Andersen et al. (2001) used for the calculation of their realized variance estimator: Firstly, a 5 minute time series had to be constructed for each share. In practise high frequency trading data may occur at irregularly spaced time intervals. In order to perform classical time series analysis it may be necessary for the researcher to convert the irregularly spaced data to a fixed time interval of their choice (5 minute, 10 minute trade intervals etc.).

Engle and Sun (2004) note the following three drawbacks when converting irregularly spaced data to fixed time intervals. Firstly, only a specific subset of data can be used if the fixed time interval is broad (e.g., hourly spaced data etc.). Important trading information may be lost if the data falls outside of the chosen time interval. In their study Andersen et al. (2001) constructed a synthetic 5 minute return series from the 5 minute logarithmic differences between prices recorded at or immediately before the corresponding 5 minute marks. By doing so important information contained between the corresponding 5 minute marks wouldn't have been included in the synthetic time series. In contrast to Andersen et al. (2001) Bollen and Inder (2002) utilized all of the available high frequency data points for the construction of the time series necessary to compute the VARHAC volatility estimator. The latter characteristic makes the VARHAC estimator an attractive alternative as it imposes few assumptions on the nature of the data generating process for high frequency returns. Lastly, when converting "irregularly spaced" data to "regularly spaced" data the researcher will lose some of the original characteristics of the data. As an example, the duration of time between trades to some degree can be seen as a reflection of how quickly the market responds to shocks and news. When choosing trading intervals (e.g., 10 minute or hourly spaced data) that fit the data too loosely the latter characteristics may be lost in the conversion. Trading duration bears important information in predicting volatility of returns as it takes a certain period of time for the trading price to incorporate all the new information of the market. Existing market microstructure theories model the latter characteristic differently and how trade duration affects volatility of returns remains an empirical question (see e.g., Engle and Sun (2004)).

To create the time series the nearest neighbour interpolation method in MATLAB was used. The method interpolates to the nearest left hand side data point of the corresponding 5 minute mark. In Table 3.15 and 3.16 a practical example of the construction of the initial 5 minute time series for the ABDE methodology is shown.

Table 3.15: Raw input data used to construct an artificial 5 minute data series for Anglo American PLC (AGL).

Time Stamp	Price (cents)
6/21/00 9:27 AM	7942.857143
6/21/00 9:28 AM	7945
6/21/00 9:33 AM	7945.588235
6/21/00 9:40 AM	7940
6/21/00 9:44 AM	7940
6/21/00 9:45 AM	7935
6/21/00 9:49 AM	7930
6/21/00 9:50 AM	7930
6/21/00 9:53 AM	7930
6/21/00 9:54 AM	7930
6/21/00 9:56 AM	7930
6/21/00 9:57 AM	7930

Table 3.16: Resulting 5 minute time series for Anglo American PLC (AGL) using the nearest neighbour method in MATLAB.

5 minute series	Price (cents)
6/21/00 9:30 AM	7945
6/22/00 9:35 AM	7945.588235
6/23/00 9:40 AM	7940
6/24/00 9:45 AM	7935
6/25/00 9:50 AM	7930
6/26/00 9:55 AM	7930
6/27/00 10:00 AM	7930

Secondly, we followed the Andersen et al. (2001) methodology to purge the high frequency returns of negative serial correlation induced by the uneven spacing of the data points, by implementing a MA(1) model for each of the 5 minute return series.

Andersen et al. (2001) found in their study that when spurious dependence introduced by non-synchronous trading was present, all the estimated moving average coefficients would be negative. In our study we found that, with the exception of BVT, the latter characteristic did not apply to our shares. The estimated parameters of each of the MA(1) models fitted to the shares selected for our study are shown in Table 3.17 (for a related discussion we refer the reader to Section 2.3.4 and Section 4.3.1)

Table 3.17: List of MA(1) equations used to calculate the 5 minute return series for ABDE.

Share Code	θ	μ
AGL	0.130000	0.000010
BIL	0.106320	0.000010
BVT	-0.019600	-0.000002
HAR	0.101345	0.000020
RCH	0.093060	-0.000006
SBK	0.053970	0.000001

The ABDE volatility estimator was calculated from the residuals of the MA(1) fit.

Data preparation for the VARHAC volatility estimator

For the VARHAC volatility estimator all available high frequency information was utilized in the construction of the input data series. No further data manipulation was necessary. The latter characteristic makes the VARHAC estimator an attractive alternative as it imposes few assumptions on the nature of the data generating process for high frequency returns.

Data preparation for the GARCH and NIG GARCH volatility estimators

In order to prepare the necessary input data for the NIG GARCH model the closing prices for the subset of six shares selected for our study was downloaded from Reuters for the period of 21 Jun 2000 – 7 Mar 2003. The time series was not constructed from the raw data because the JSE follows a complicated end of day procedure to derive the closing prices on a specific trading day. Therefore we felt there was no need to recreate the already existing Reuters data.

3.6 CONCLUSION

In this chapter we discussed the necessary ground work we had to complete in order to construct a reliable time series to utilise in the calculation of our various volatility estimators. The process of constructing a reliable intraday return series is complex. We gave the reader an overview of the factors that drive the observed price process. In order to filter the data of errors we need to recognise the factors that can generate the erroneous data points as well as market events contributing to volatility and suspicious data points. The latter process of filtering the data of potential erroneous data points was a time consuming and labour intensive task which had to be performed in order to obtain good quality high frequency data. Once the data had been filtered we could proceed with the preparation of the different time series formats (e.g., 5 minute etc.) required by the different volatility estimators discussed in Chapter 2.

We proceed in the next chapter with a discussion of the statistical and distributional characteristics of the different volatility estimators. Our main objective is to determine whether the Andersen et al. (2001) findings also apply to JSE shares. In particular, the VARHAC estimator will be compared with the ABDE estimator as well as with daily volatility estimators based on the NIG-GARCH model proposed by Venter and de Jongh (2002, 2004).

CHAPTER 4

EMPIRICAL STUDY CONCERNING ESTIMATORS OF DAILY VOLATILITY

4.1 INTRODUCTION

Andersen et al. (2001) examined realized daily equity return volatilities obtained from high-frequency intraday transaction prices on individual stocks in the Dow Jones Industrial Average. They found that the unconditional distributions of the realized variances are highly skewed to the right, while the realized logarithmic standard deviations are approximately Gaussian, as are the distributions of the returns scaled by realized standard deviations. In this chapter we investigate the above-mentioned findings by using high frequency intraday transaction prices on six individual shares listed on the JSE Securities Exchange SA (JSE). Our main objective is to determine whether their findings also apply to a South African equity environment. Our study will be similar to theirs, but will use our JSE data. However, as far as scaled returns are concerned we will investigate alternative realized volatility estimators. In particular, the VARHAC estimator, advocated by Bollen and Inder (2002), will be compared with the ABDE and SSR estimator as well as with daily volatility estimators based on the GARCH and NIG-GARCH model proposed by Venter and de Jongh (2004). The theory underlying these methods was described in detail in Chapter 2. In the subsequent analysis we will use the filtered high frequency intraday transaction price series constructed in Chapter 3.

Before we can start with the statistical analysis of our various volatility estimators it is important to ensure that the data has been filtered of all erroneous data. In Section 4.2 we investigate and discuss the data outliers for the ABDE, SSR and VARHAC variance estimates. We describe the methodology we followed and we continue with a general discussion on additional market events in the South African equity market that had an impact on the variance estimates. We conclude the section with a summary of erroneous data points and illiquid trading days we identified that had to be removed from the study.

In Section 4.3 we investigate if the Andersen et al. (2001) claims hold true in a South African equity market context: Firstly, we investigate if a MA(1) model did indeed purge the data of negative serial correlation. Secondly, we research the first claim by Andersen et al. (2001) i.e. that the distributions of the returns scaled by realized standard deviations are Gaussian, followed by their second claim, i.e. that the unconditional distributions of the realized variances are highly right skewed, while the realized logarithmic standard deviations are approximately Gaussian. The characteristics of the return series as well as the scaled returns for various volatility estimators for the six shares selected will be discussed in more detail in this section.

In Section 4.4 we discuss the role of volatility and the impact it can have for a derivative trader managing options. We illustrate the sensitivity of an option to it's volatility with a practical example of a call option on one of the shares we selected for the study and the effect that changes in volatility can have on the profit and loss and risk management of a derivative portfolio. We conclude the chapter in Section 4.5 with a summary of our findings in comparison to Andersen et al. (2001).

4.2 TESTING THE ACCURACY OF THE DATA FILTER

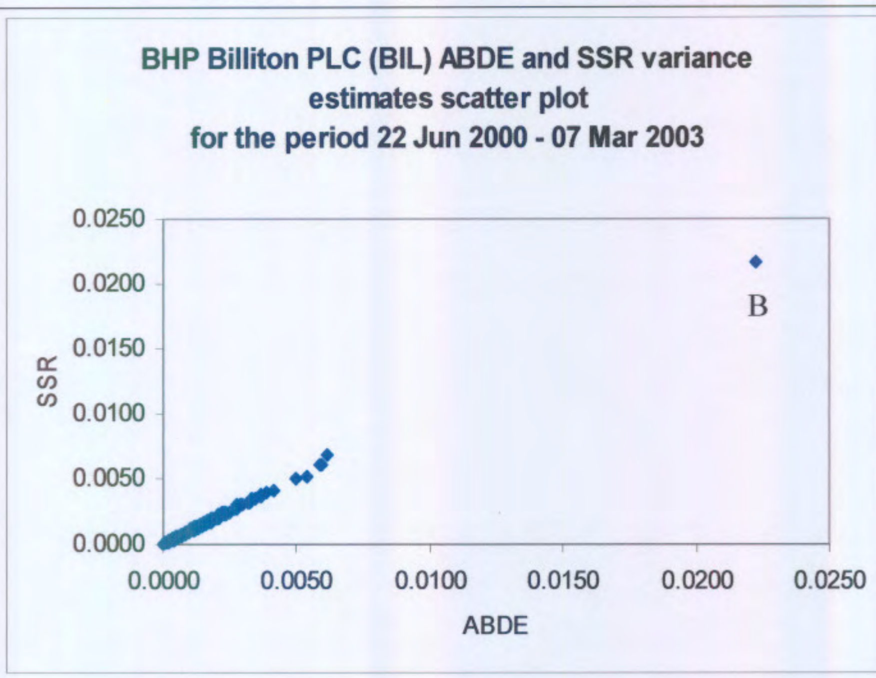
In Chapter 3 we discussed the underlying characteristics of equity markets as well as the various market events that have an impact on volatility. We also discussed the important task of identifying "suspicious" data points and filtering these out. Because we chose to filter the data manually (i.e. by hand) it was important to perform one last check to ensure that all erroneous data points had been removed. In this section we will discuss the process followed to ensure that any remaining variance outliers (higher and lower estimates of variance) were indeed true reflections of actual market events. We proceed as follows: Firstly, we will discuss the scatter plot methodology applied in order to establish if the data had been filtered efficiently. Secondly, we discuss the erroneous data points we identified that hadn't been removed by the filtering methodology employed in Chapter 3. We conclude the section with some examples of variance estimates that seemed "suspicious" at a first glance, but on further investigation turned out to be valid variance estimates due to actual price fluctuations.

Scatter plot analysis

We calculated the SSR, ABDE and VARHAC realized variance estimators for the six shares selected for our study for the period 22 Jun 2000 – 07 Mar 2003. We then compared scatter plots of the ABDE and SSR variance estimates and the ABDE and VARHAC variance estimates for each share selected in our study (note that due to the fact that the ABDE and SSR variance estimates are so similar it wasn't necessary to compare the VARHAC to both the ABDE and SSR variance estimates). For each of the scatter plots we identified suspicious or outlying volatilities and then investigated their "correctness" in detail, by scrutinising the underlying intraday return series. In Tables 4.1 - 4.4 we present the ABDE and SSR scatter plots for the six shares selected.

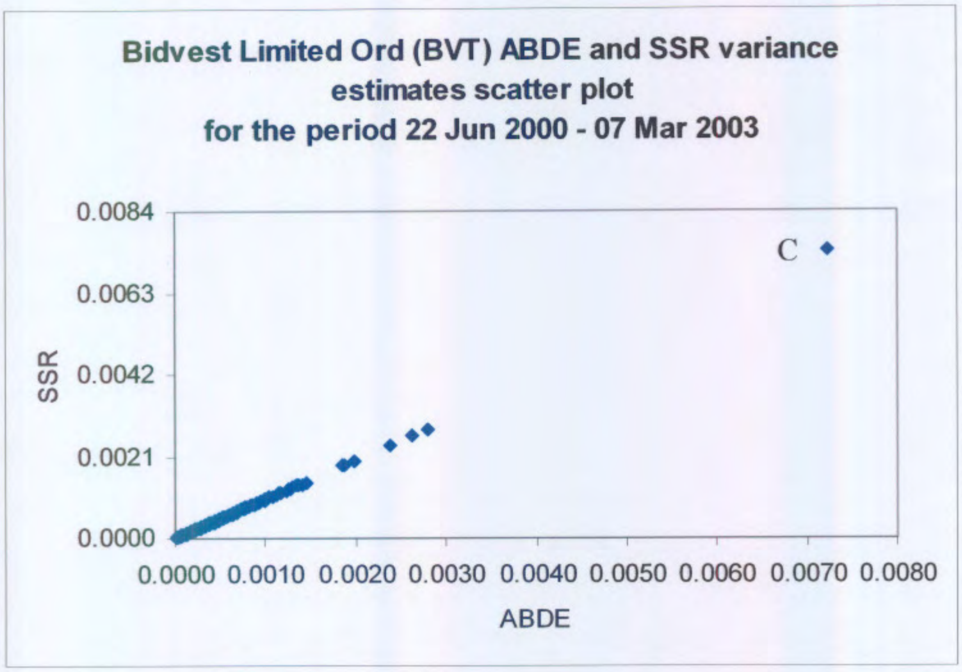
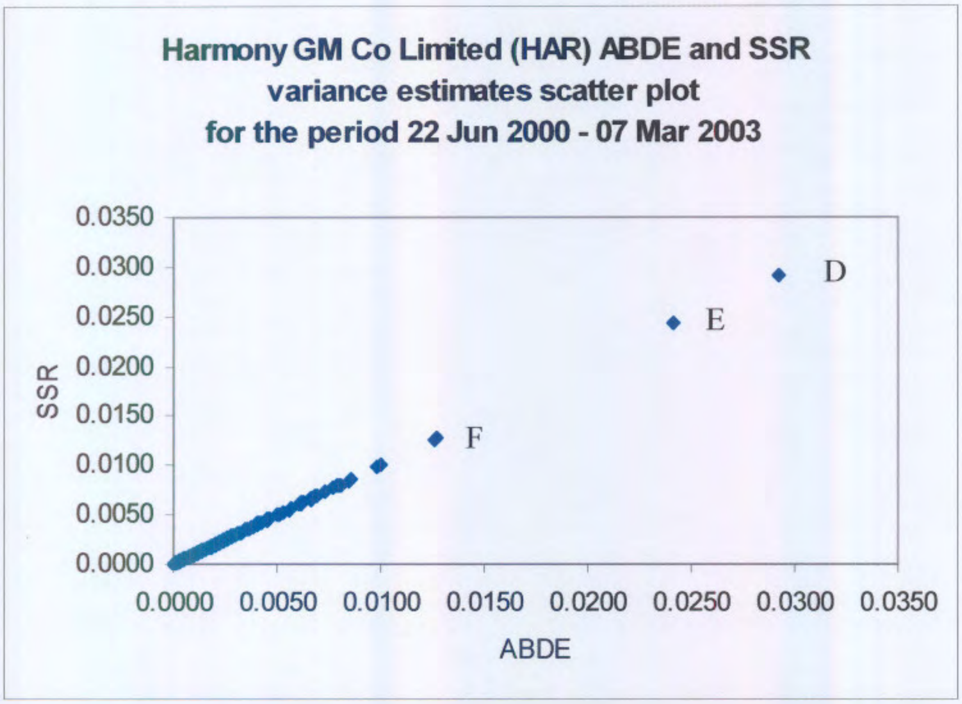
Table 4.1: ABDE and SSR scatter plot analysis for Anglo American PLC (AGL) and BHP Billiton PLC (BIL) for the period 22 Jun 2000 – 07 Mar 2003.

<table>
<tr><th>Scatter plot analysis</th><th>Comments</th></tr>
<tr><td>


</td><td>Marker A:
Erroneous data outlier included on 4 Feb 2002. The data on 4 Feb 2002 was deleted from the study.</td></tr>
<tr><td>


</td><td>Marker B:
The extreme variance estimate can be attributed to the large difference in opening price in comparison to the rest of the trading day (close on 18 Mar 2001 was at R32.20, the first trade at R32.20 and next trade at R37.86 and similar levels were sustained for rest of trading day).</td></tr>
</table>

Table 4.2: ABDE and SSR scatter plot analysis for Bidvest limited ORD (BVT) and Harmony GM Co Limited (HAR) for the period 22 Jun 2000 - 07 Mar 2003.

<table>
<tr><th>Graph</th><th>Comments</th></tr>
<tr><td>
</td><td><u>Marker C:</u>
Higher variance estimate due to market uneasiness after the 11 Sep 2001 Twin Towers terrorist attacks.</td></tr>
<tr><td>
</td><td><u>Marker D</u>
Higher variance estimates on 12 Jun 02 due to market volatility.

<u>Marker E</u>
The market closed at R38 on 26 Feb 01 and opened at R43.80 on 27 Feb 01, a volatile trading day in general.

<u>Marker F</u>
Uneasiness due to the Mining Charter.</td></tr>
</table>

Table 4.3: ABDE and SSR scatter plot analysis for Richmond Securities AG (RCH) for the period 22 Jun 2000 - 07 Mar 2003.

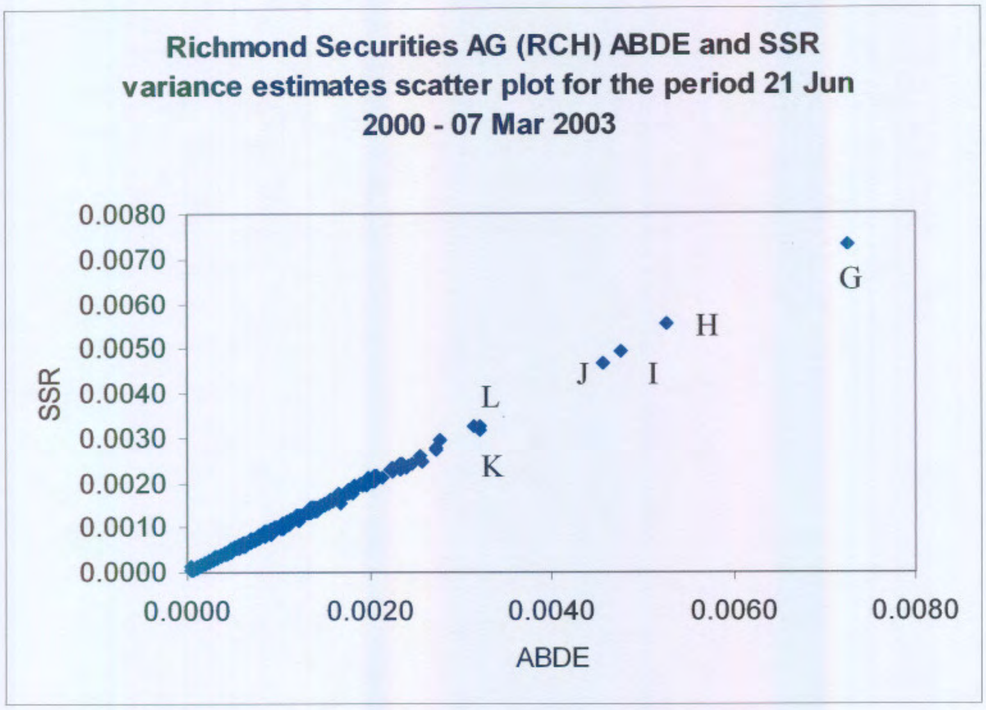
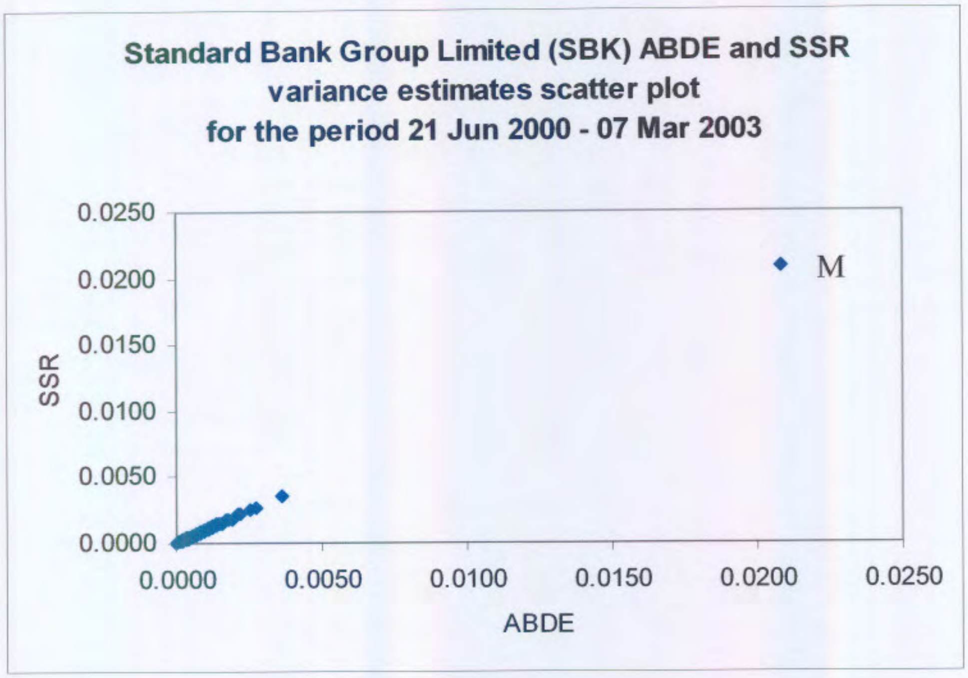
Graph	Comments
	Marker G and L: 11-13 Sep 2001 terrorist attacks. Marker H: 21 Dec 2001, volatile trading day in general. Marker I: 26 Oct 2001, volatile trading day in general. Marker J: 25 Sep 2001 a volatile trading day in general. Marker K: 27 Dec 2002, volatile trading day.

Table 4.4: ABDE and SSR scatter plot analysis for Standard Bank Group Limited (SBK) for the period 22 Jun 2000 - 07 Mar 2003.

Graph	Comments
	Marker M: 11 Sep 2001 terrorist attacks.

From the scatter plots it is clear that the SSR and ABDE estimators provide very similar estimates of daily volatility. Before we proceed with the ABDE and VARHAC scatter plot analysis there is one more interesting aspect of the SSR and ABDE scatter plots that we would like to discuss. Engle (2000) found that one of the salient features of high frequency data was that the data was irregularly spaced and that when data was aggregated to fixed intervals of time a loss of information occurred. This loss occurs partly because the large number of zero log-returns makes econometric analysis very complex if the intervals are small. In two thirds of the calculated variance estimates the ABDE variance estimates were slightly higher than the SSR variance estimates. We ascribe the latter characteristic of our time series due to the fact that a large numbers of zero log returns have been included in the calculation of our SSR and ABDE variance estimates and furthermore due to the MA(1) process being fitted for the ABDE estimators the zero log-returns (as in the case of the SSR estimator) are being replaced by "non- zero values".

On a typical trading day there were extended periods where no price change took place in our 5 minute samples used to calculate the SSR and ABDE variance estimates, i.e. the share traded at the same price and this holds true even for the more liquid shares in our sample like AGL. In Table 4.5 we list the median, average, minimum and maximum statistics for the number of zero log-returns used in the calculation of the SSR variance estimates.

Table 4.5: Summary of the median, average, minimum and maximum statistics of zero log returns per day for the period 21 Jun 2000 – 07 Mar 2003 used in the calculation of the SSR variance estimates for the six shares selected for our study.

Share	Median	Average	Min	Max
AGL	41%	43%	6%	95%
BIL	57%	56%	14%	97%
BVT	72%	72%	52%	90%
HAR	73%	66%	10%	98%
RCH	51%	51%	16%	97%
SBK	61%	60%	20%	98%

From the above table it is evident that on a typical trading day there were extended periods where no change in share price took place (resulting in a zero log return for those periods). For the most actively traded share, AGL, the average percentage of zero log returns per day was 43 %, in comparison to BVT with and average percentage of zero log returns per day of 72%. Unfortunately we cannot compare the latter result against the Andersen et al. (2001) study as the authors do not discuss the number of zero log returns obtained in their study.

We now continue with a scatter plot analysis of the ABDE versus the VARHAC estimates. Clearly, it is not necessary to also consider the SSR versus the VARHAC estimators because the SSR and ABDE variance estimates were seen above to be very similar. In Tables 4.6 – 4.9 a brief description of the outliers is given for each graph and a detailed discussion of these outliers and their respective causes will follow after all the scatter plots.

Table 4.6: ABDE and VARHAC scatter plot analysis for Anglo American PLC and BHP Billiton PLC for the period 22 Jun 2000 – 07 Mar 2003.

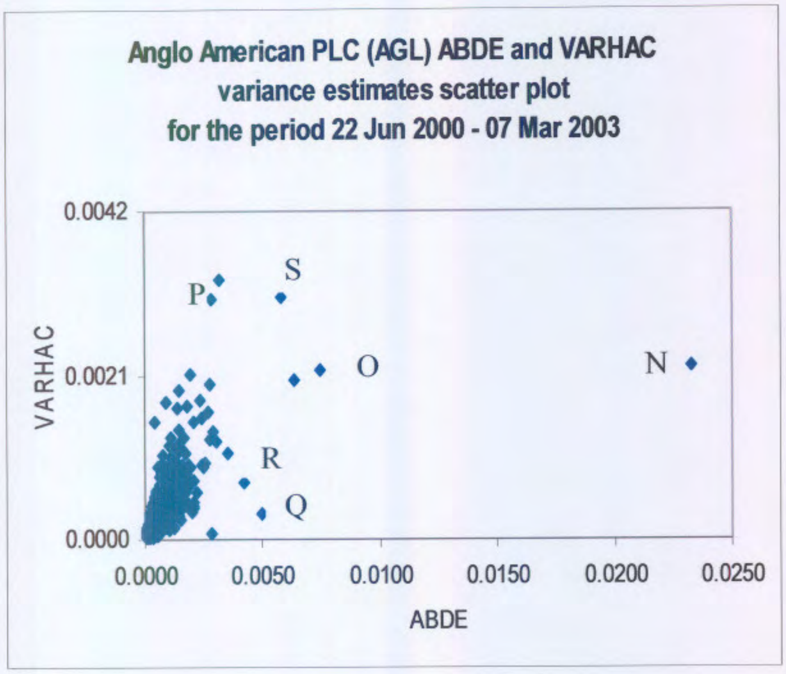
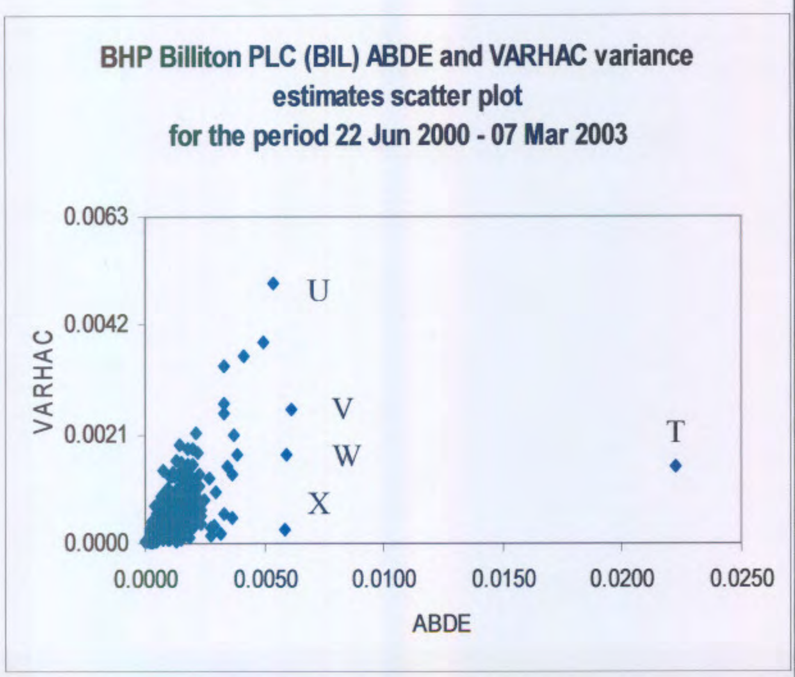
Scatter plot analysis	Comments
	Marker N: 4 Feb 2002, similar to the ABDE scatter plot this day has been deleted from the study (refer to marker A). Marker O : Brooklyn Bridge bomb scare. Marker P: Leaking of the mining Charter. Marker Q: Large price move on open of trade and a volatile trading day in general. Marker R: Festive period (previous day's close at R186, market opened at R 180). Marker S: 11 Sep 2001 terrorist attacks.
	Marker T: Large price move on open of trade and a volatile trading day in general (refer to marker B). Marker U: Volatile trading day in general. Marker V: Includes 4 trades where the price fell from R30 to R28. Marker W: Brooklyn Bridge bomb scare. Marker X: Large price move on open of trade.

Table 4.7: ABDE and VARHAC scatter plot analysis for Bidvest Limited ORD (BVT) and Harmony GM Co Limited (HAR) for the period 22 Jun 2000 - 07 Mar 2003.

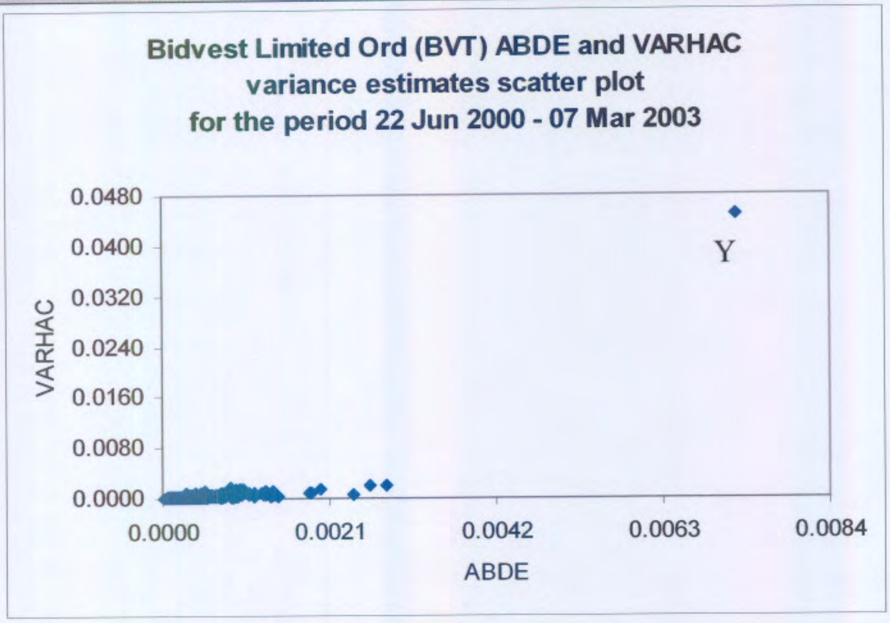
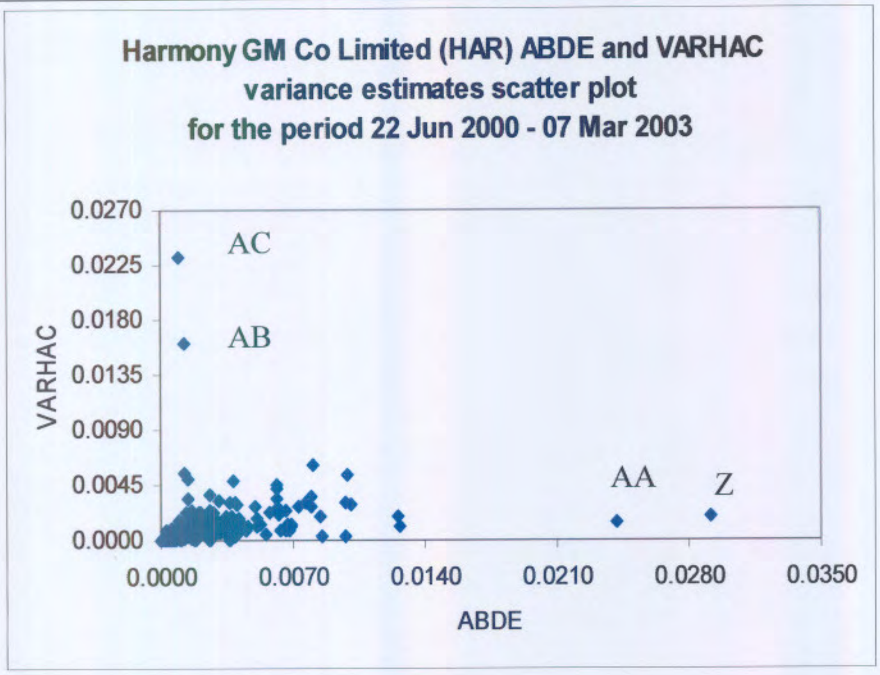
Graph	Comments
	Marker Y: 11 Sep 2001 terrorist attacks (refer to marker C).
	Marker Z: Large move in price on open of market. Marker AA: Large move in price on open of market (refer to marker E). Marker AB: VARHAC time series more volatile than the ABDE time series. Marker AC: Small data set (20 trade entries).

Table 4.8: ABDE and VARHAC scatter plot analysis for Richmond Securities AG (RCH) for the period 22 Jun 2000 - 07 Mar 2003.

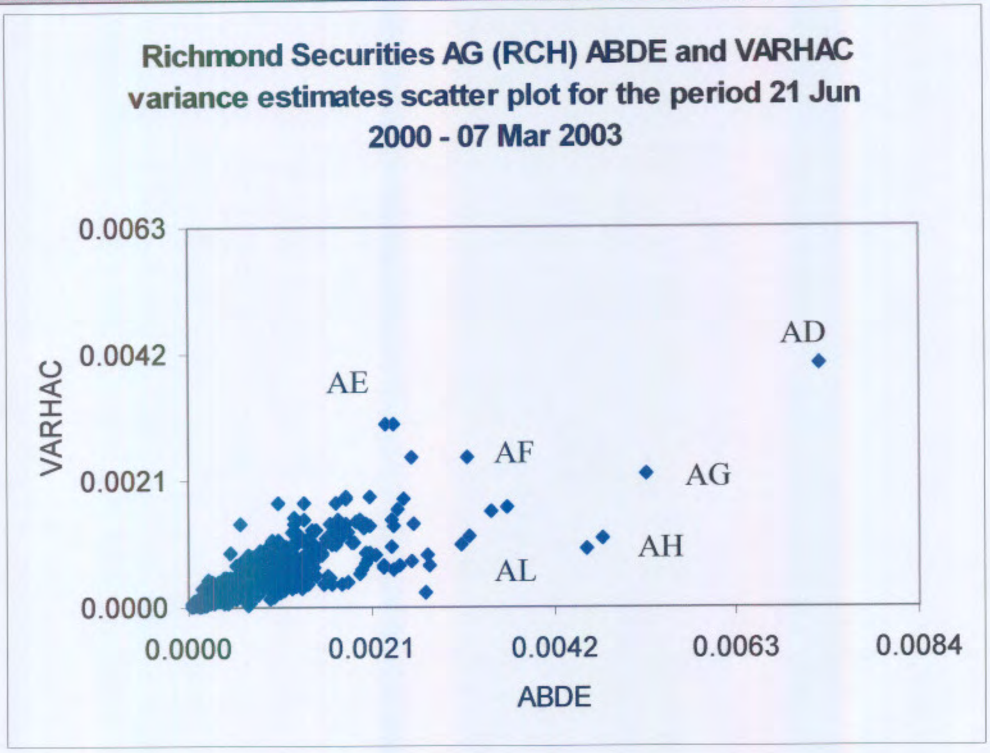
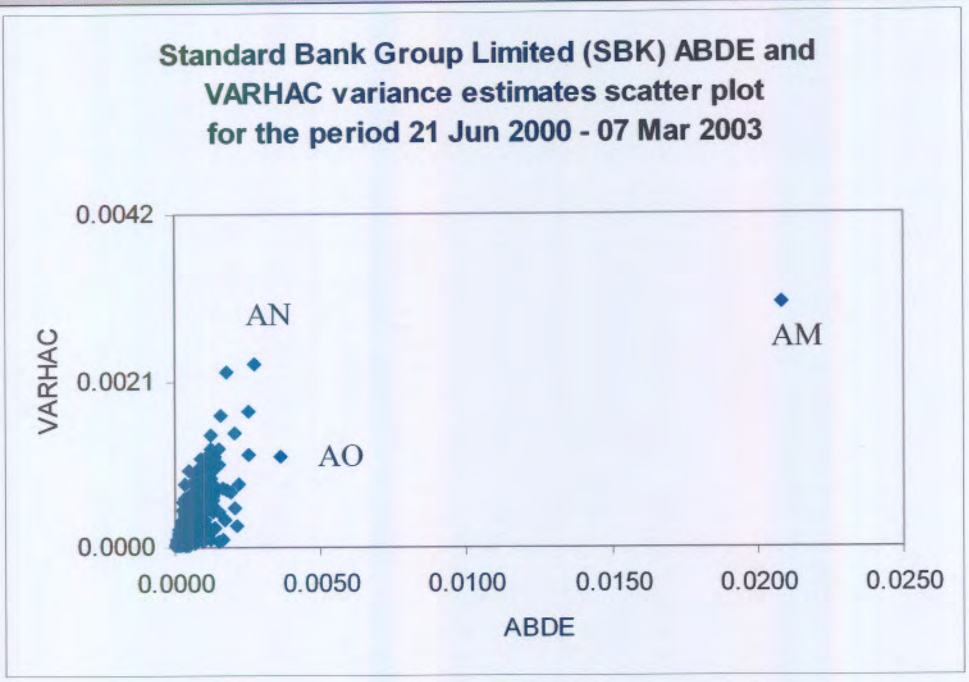
Graph	Comments
	Marker AD: 11 Sep 2001 terrorist attacks (refer to marker G). Marker AE: Market volatile with price spike at market close. Marker AF: Large move in price on open of market. Marker AG: Large move in price on open of market (refer to marker H). Marker AH: Large move in price on open of market (refer to marker I). Marker AL: Brooklyn Bridge bomb scare (refer to marker J).

Table 4.9: ABDE and VARHAC scatter plot analysis for Standard Bank Group Limited (SBK) for the period 22 Jun 2000 - 07 Mar 2003.

Graph	Comments
	Marker AM: 11 Sep 2001 terrorist attacks (refer to marker M). Marker AN: Large move in price on open of market. Marker AO: Brooklyn Bridge bomb scare.

From the ABDE and VARHAC scatter plots we identified three scenarios regarding the VARHAC and ABDE variance estimates that had to be investigated further:

- Firstly, when in our study would the VARHAC variance estimates be significantly higher than the ABDE variance estimates?
- Secondly when would the variance estimates be similar?
- Lastly when would the ABDE variance estimates be significantly higher than the VARHAC variance estimates?

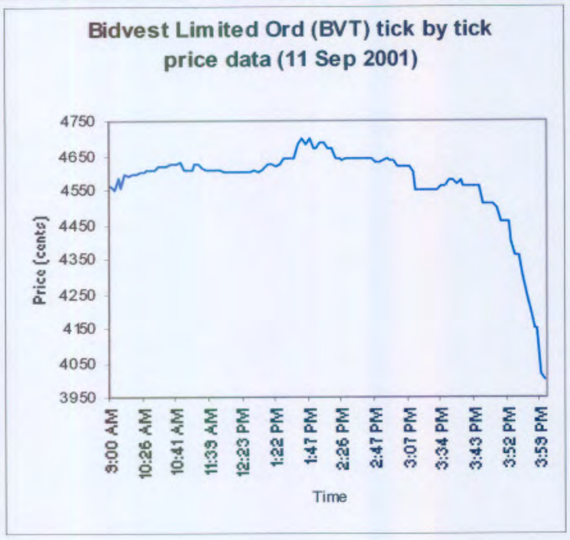
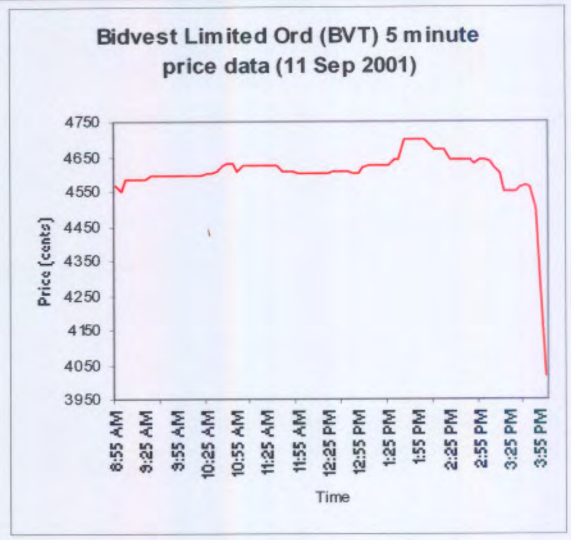
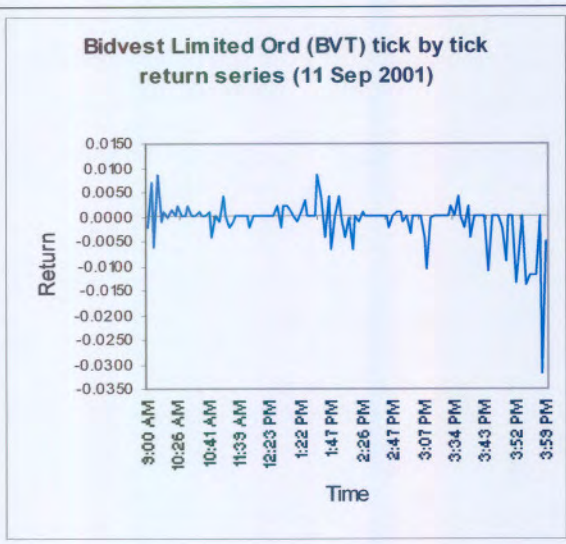
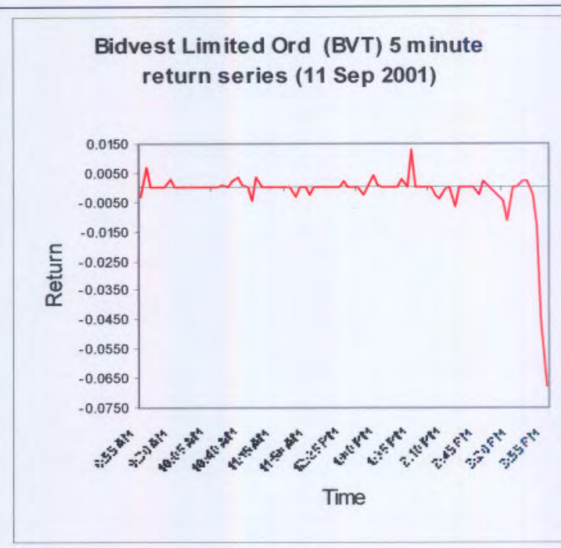
Before continuing note the following. In our study the ABDE variance estimates were based on a synthetically created time series, namely by the nearest neighbour method where the last price closest to the left of the time interval was included. The previous day's close was incorporated in the calculation of the first log return of the 5 minute time series and therefore the ABDE variance estimates would include the effect of overnight news events etc. In contrast, the VARHAC variance estimates were calculated utilizing all of the high frequency data and therefore included "inter" 5 minute price variations. Furthermore the VARHAC variance estimates were calculated from the first trade of the day to market close, not incorporating over night news events etc.

<u>VARHAC variance estimates higher than ABDE variance estimates</u>

As stated before we know that there is information contained in the high frequency data used in the calculation of the VARHAC variance estimates that hasn't been included in the 5 minute time series used in the calculation of the ABDE variance estimates. Therefore, if significant price changes occur in the "inter" 5 minute intervals (which are not included in the 5 minute time estimates) one would expect a higher VARHAC variance estimate. From the scatter plot analysis we identified three markers, namely Y (BVT - 11 Sep 2001), AB (HAR – 4 Apr 2001) and AC (HAR – 15 Feb 2001) where the VARHAC variance estimates were significantly higher than the ABDE variance estimates. In the following section we will present graphs to support this hypothesis.

In Table 4.10 we graphically depict the tick by tick, 5 minute price and corresponding return series for marker Y, BVT.

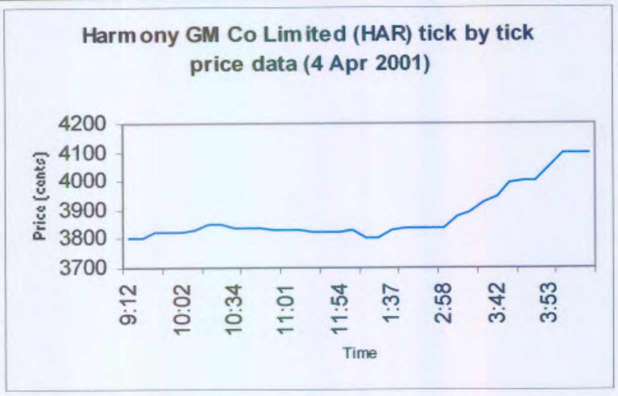
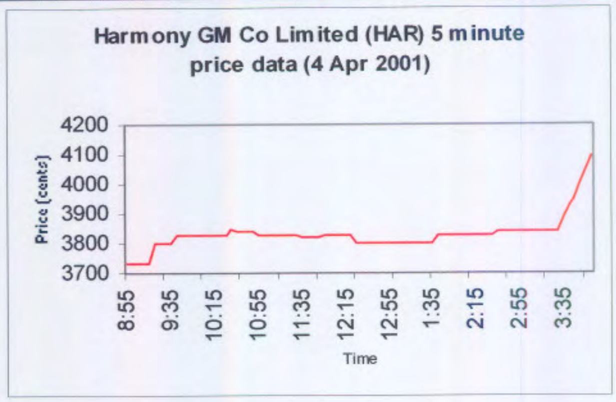
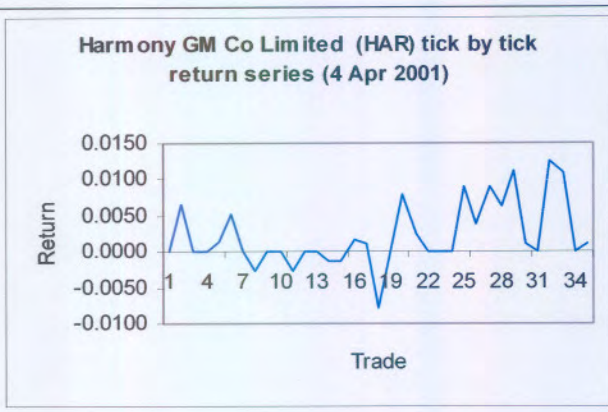
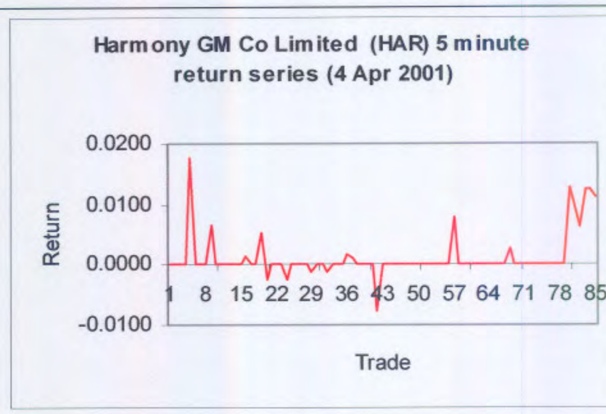
Table 4.10: A comparison of the tick by tick, 5 minute price and corresponding return series used in the Bidvest Limited ORD (BVT) VARHAC and ABDE variance estimates.

VARHAC variance estimate = 0.044839	ABDE variance estimate = 0.007220
	
	

From the graphs in Table 4.10 it is clear that the tick by tick returns were more volatile than the corresponding 5 minute returns and this seems to be the reason why the VARHAC variance estimate was higher than the ABDE variance estimate. Note that the sensitivity of the VARHAC variance estimator to large market moves could be investigated in a further study.

Following on from the previous discussion in Table 4.11 we graphically depict the tick by tick, 5 minute price and corresponding return series for the next marker AB, HAR.

Table 4.11: A comparison of the tick by tick, 5 minute price and return series used in the Harmony GM Co Limited (HAR) VARHAC and ABDE variance estimates.

VARHAC variance estimate = 0.016042	ABDE variance estimate = 0.001276
	
	

The HAR tick by tick return series in Table 4.11 seems to be more volatile in comparison to the 5 minute return series.

On some of the trading days included in our study we found that very few trades were recorded on a specific trading day. If a small quantity of trades were executed on a specific trading day the decision was made to exclude these days from the study and some examples applicable to our study can be found in Table 4.16. On 15 Apr 2001 (marker AC) only 20 trades were executed in the market for HAR resulting in a high VARHAC variance estimation. The VARHAC variance estimator is reliant on sufficient data in order to calculate the BIC criteria and when too little input data is provided inaccurate estimates may result.

VARHAC variance estimates similar to the ABDE variance estimates

There were many examples in our study where the VARHAC and ABDE variance estimates were similar to one another. We have selected the following three markers from our scatter plot analysis to illustrate the latter characteristic (Table 4.12), namely marker P, American AGL, marker AD, RCH and marker AN, SBK.

Table 4.12: VARHAC and ABDE variance estimates.

Marker	Share	VARHAC	ABDE	Comment
P	AGL	0.003328	0.003149	Early leak of the Mining Charter
AD	RCH	0.001031	0.001687	Twin Towers terrorist attacks
AN	SBK	0.00271	0.00230	15 Jan 2002 a volatile trading day

Both the VARHAC and ABDE variance estimation techniques produced similar variance estimates for the 11 Sep 2001 terrorism attacks (RCH), the well documented leak of the Mining Charter (AGL) as well as for 15 Jan 2002 (SBK) a volatile trading day (where the SBK share price spiked up sharply on open). Later in this section (p98) we will be discussing the variance estimates obtained for the 11 Sep 2001 terrorist attacks and the Mining Charter in more detail.

<u>ABDE variance estimates higher than VARHAC variance estimates</u>

As stated before, the previous days close was incorporated in the calculation of the first log return of the 5 minute return series. Therefore the ABDE variance estimates would include the effect of over night news events etc. In comparison the VARHAC variance estimates were calculated from the first trade of the day to market close, not incorporating the effect of over night news events etc. We believe that this is the reason for the higher ABDE variance estimates in comparison to the VARHAC variance estimates. In the following section we will investigate the latter hypothesis.

From the scatter plot analysis we identified the following markers, namely: T, Z, AA and AM where the ABDE variance estimates were significantly higher than the VARHAC variance estimates. To illustrate we refer the reader to marker AA - on 26 Feb 2001 (refer to Table 4.13). HAR closed at R38 and the first trade of the day on 27 Feb 2001 was at R43.80, a 15% move in price, which is not unusual for HAR due to the sensitivity of the share to movements in the Rand and gold price. Due to the fact that the VARHAC variance estimator wouldn't have included the extreme market moves between close of trade on the previous business day and open of trade on the following day we would therefore expect the ABDE variance estimates to be higher. On all of the days in question in Table 4.13 on opening of trade a large move between the previous days closing price (used in the first return) and the following days market opening price was found.

Table 4.13: Comparison of the ABDE and VARHAC variance estimates.

Marker	Share	VARHAC	ABDE	Previous day's closing price (first 5 minute price observation)	Market open (second 5 minute price observation)	Percentage price move
T	BIL	0.001480	0.022250	R32.20	R37.10	-15%
Z	HAR	0.001974	0.029189	R125	R142	-14%
AA	HAR	0.001500	0.024130	R38	R43.80	-15%
AM	SBK	0.003073	0.020842	R32	R28	13%

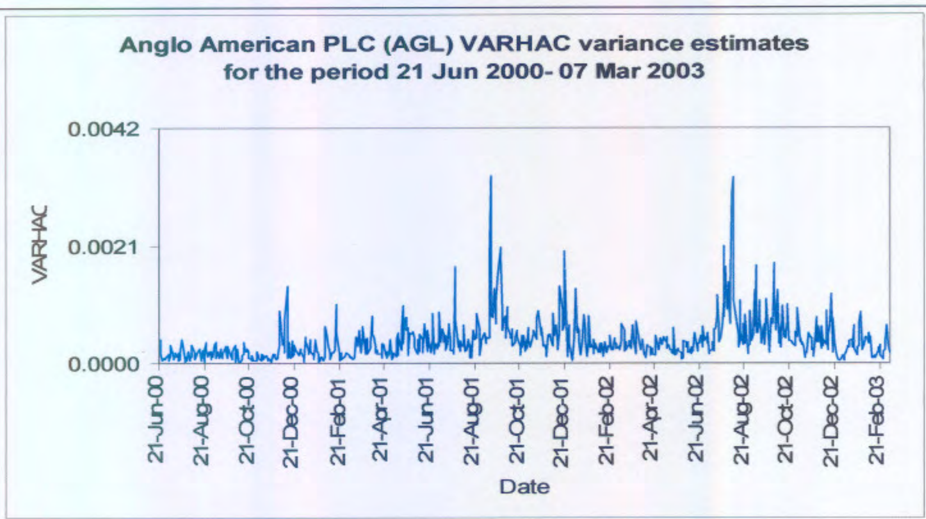
In the following section we summarize the market events and characteristics that we identified from the scatter plot analysis of the ABDE, SSR and VARHAC variance estimates.

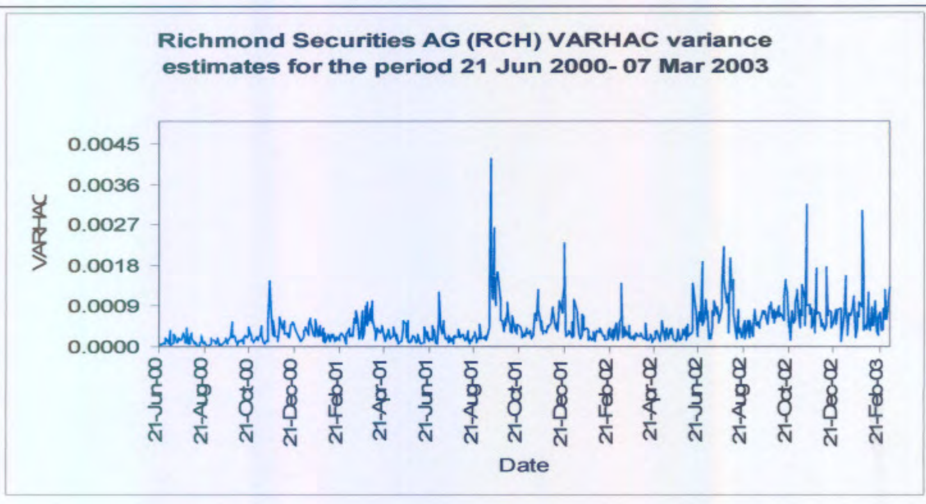
Examples of South African market events that impacted the variance estimates

Firstly, there were noticeable fluctuations in variance during festive periods and care should be taken when analysing variance estimators during Christmas and New Year.

Secondly, the terrorist attacks of Sep 2001 had a significant impact on the variance and almost all of the shares selected for our study displayed high variances during that period. To illustrate the latter point we graphically depict the VARHAC variance estimates for two of the shares in Table 4.14.

Table 4.14: VARHAC variance estimates for Anglo American PLC (AGL) and Richmond Securities AG (RCH) for the period 21 Jun 2000- 07 Mar 2003.





Related to the terrorist attacks, there was a noticeable increase in variance in five of the shares (excluding HAR) on 25 Sep 2001 (two weeks after the initial attacks when the market was already uneasy after the initial terrorist attacks) due to the Brooklyn Bridge bomb scare. On opening of trade the already unsettled markets reacted nervously with some of the shares in our sample moving as much as 5 % (RCH and SBK) to 7 % (BIL) from the previous day's close of trade and first trade of the day on 25 Sep 2001. As mentioned previously, HAR was the only share that didn't show a noticeable increase during this period. HAR is a more volatile share in general in comparison to the other shares selected for the study and the latter characteristic can be ascribed to the company's Rand hedge exposure and subsequent sensitivity to movements in the gold price (quoted in dollars).

Thirdly, futures closeouts take place four times a year namely, on the third Thursday of March, June, September and December. Volumes of equities traded during these periods will increase depending on the viewpoint of arbitrage traders (arbitrage traders will either be long shares and short futures or vice versa). The arbitrage traders may be closing down existing basket trades, opening new baskets or rolling baskets (buying/selling the future expiring and selling/buying the next futures contract). In Figure 4.1 we graphically depict the move in open interest of the Jun 2001 ALSI futures contract to the Sep 2001 ALSI futures contract the week before close out (11 Jun 2001 to 21 Jun 2001). The changeover date from the expiring futures contract to the next futures contract, namely September 2001, is an indication of the arbitrage players in the market rolling their baskets as well as the derivative traders hedging their positions with the next available contract.

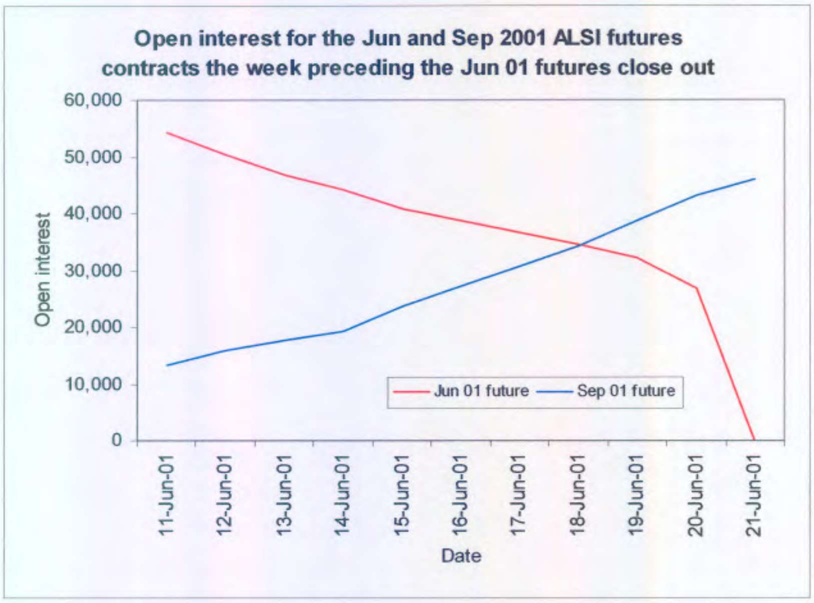


Figure 4.1: SAFEX open interest for the Jun 2001 and Sep 2001 ALSI futures contracts.

Lastly, similar to the terrorist attacks there was a noticeable increase in variance in the month of Aug 2002 for the mining sector shares (AGL, BIL and HAR) in our study. The latter increase in variance during that period can be ascribed to market concerns arising from an early (leaked) draft of the Mining Charter (http://www.mineweb.net). In Figure 4.2 we depict the closing prices of Anglo American PLC for the period Jan 2002 – Dec 2002 (note the decline in price during the Aug 2002).

Figure 4.2: Closing prices of Anglo American PLC (AGL) for the period 2 Jan 2002-31 Dec 2002.

The proposed Mining Charter caused market uneasiness due to the following reasons: Firstly, there were market rumours doing the rounds suggesting the possibility of the introduction of an 8 percent royalty tax and a 5 percent export tax on mining companies. Secondly, not all market participants had equal access to the information and some market participants were better informed than others and could act thereupon where the other participants couldn't because they weren't aware of the (leaked) information. Thirdly, the charter proposed the review of current mining rights. Lastly, the charter proposed black empowerment of South African mining companies. However, market nerves were settled by statements from Government sources indicating that the rumour wasn't true (http://www.mineweb.net). Trading in AGL during that period was a reflection on how sensitive the market was to speculation, following the controversial leaking of the Mining Charter information.

Higher variance estimates due to actual erroneous data points and illiquidity

In the following section we will note those data points we identified in the scatter plot analyses that could be ascribed as erroneous. Due to the fact that we chose to clean the data by hand it is expected that a few errors may have been included due to the large volume of data filtered manually. In Table 4.15 we list the erroneous data points that had to be excluded from the variance analysis.

Table 4.15: List of the erroneous data points that were identified and subsequently deleted from the study.

Share	Date of erroneous data point	Comment
AGL	4 Feb 2002	Refer to Table 4.6 marker A and N
BIL	19 Mar 2001 and 30 Nov 2001	Deleted prior to scatter plot analysis
RCH	26 Oct 2001 and 19 Dec 2002	Deleted prior to scatter plot analysis
SBK	17 Sep 2001 and 11 Jul 2002	Deleted prior to scatter plot analysis

In Table 4.16 a list of illiquid trading days is shown. This result is close to zero estimates of volatility which in turn results in very high values of standardized returns.

Table 4.16: List of the illiquid trading days that we identified and subsequently deleted from the study.

Share	Date	Comment
RCH	27 Dec 2002	The trading day was illiquid, with only 40 trades, compared to the other trading days included in the study (median of the number of trades for RCH equals 226)
AGL,BIL,BVT,HAR, RCH and SBK	5 Dec 2000,8 Jan 2001, 15 Feb 2001,13 Mar 01 and 6 Nov 2001	The trading day was illiquid, with very few trades (less than 12 on 13 Mar 2001) or no trades for all the shares selected for the study (could be due to database problems)/

In the following section we will discuss a few examples of the latter characteristics from our study:

Firstly, for BIL we found that the high variance estimate on 5 Dec 2002 (refer to Marker X) was an accurate reflection of market conditions on that day. Although the first trading point of the day in Figure 4.3 may look like a potential data error, the opening level for BIL (R47) on 5 Dec 2002 was lower than its closing level on 4 Dec 2002 (R49.84). The ABDE variance estimate (0.005) was larger than the VARHAC estimate (0.0002) (refer to Table 4.13 for other examples of higher ABDE than VARHAC variance estimates).

Similarly for RCH the higher variance estimate on 17 Sep 2001 (refer to marker AF) was an accurate reflection of the market conditions on that specific trading day. Although the first trading point of the day in Figure 4.4 may look like a potential data error the opening level for RCH (R16.22) on 17 Sep 2001 was lower than the closing level on 14 Sep 2001 (R16.52). During the first hour of trade the prices continued dropping to lower levels and the rest of the trading day was characterized by volatile price fluctuations. On this particular day the ABDE variance estimate (0.003) was higher in comparison to the VARHAC estimate (0.002).

Because the variance on these two trading day were due to market activity they should be included in our study.

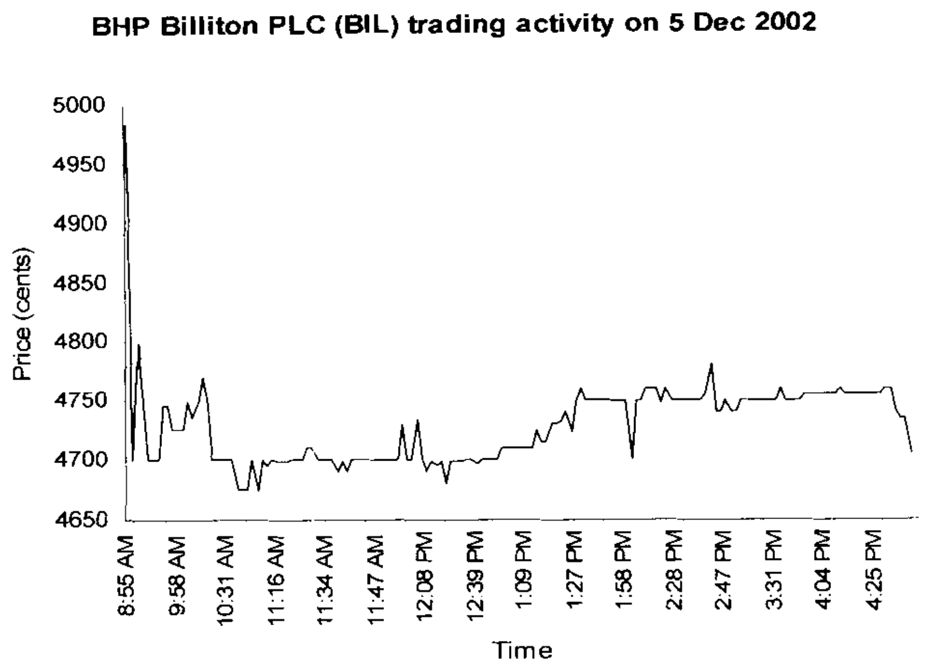


Figure 4.3: BHP Billiton PLC (BIL) trading activity on 5 Dec 2002.

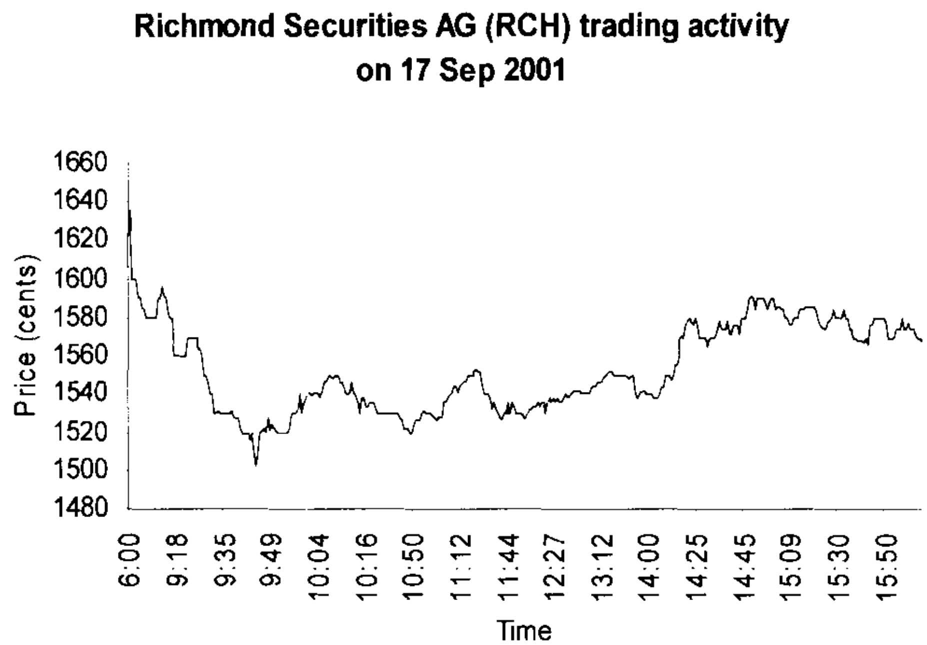


Figure 4.4: Richmond Securities AG (RCH) trading activity on 17 Sep 2001.

Finally, we can conclude that our data filtering methodology discussed in Section 3.4 was effective and that the variance outliers still present can be ascribed to valid market events namely, futures closeouts, festive periods, terrorist attacks etc.

4.3 SOUTH AFRICAN EQUITY MARKET ANALYSIS

In this section we will discuss the claims made by Andersen et al. (2001) and ascertain whether these claims are applicable in a South African market context. In the first section we investigate if the MA (1) model applied to the 5 minute log return time as proposed by Andersen et al. (2001) purges the South African data of first order negative serial correlation. In Section 4.3.2 we discuss the distributional characteristics of the six South African shares and we investigate if the distributional claims made by Andersen et al. (2001) in their article hold true in the South African market context.

4.3.1 Removal of negative first order serial correlation

Andersen et al. (2001) found, that in spite of restricting their analysis to actively traded shares, that market microstructure frictions, including price discreteness, infrequent trading and bid ask bounce effects, were still operative. In order to mitigate these effects, they used a 5 minute time horizon as the effective continuous time record. Due to the organizational structure of the market, the available quotes and transaction prices were subject to discrete clustering of bid-ask bounce effects. These microstructure frictions aren't important when longer time horizons such as days, weeks or months are selected but they can distort the distributional characteristics of high frequency data. In our study the 5 minute return series was constructed from the logarithmic difference between the prices recorded at or immediately before the corresponding 5 minute marks. In order to purge the high frequency returns of the negative first order serial correlation induced by the uneven spacing of the observed prices and the inherent bid ask spread, Andersen et al. (2001) estimated an MA(1) model for each 5 minute return series using the full sample of data.

In our study we followed the Andersen et al. (2001) methodology described above. In order to test the ability of the MA(1) model to adequately purge the 5 minute returns of negative serial correlation, we constructed two 5 minute return series. Firstly, we used the original return series and secondly we used the residuals of the return series resulting from a MA(1) fit to the original series. The lag 1 to 5 autocorrelations for the original return series are given in Table 4.17. From the table it is evident that negative first order correlation is present, especially for AGL, RCH and SBK. Note that the number of points in the return series is 66240 (960*96).

Table 4.17: Lag 1 to 5 autocorrelation values for the six shares selected for our study before a MA(1) model was applied to the 5 minute time series.

Lag	AGL	BIL	BVT	HAR	SBK	RCH
1	-0.1200	-0.0001	0.0199	-0.0131	-0.0504	-0.0851
2	-0.0283	0.0000	0.0078	-0.0132	-0.0289	-0.0371
3	-0.0126	-0.0001	-0.0065	0.0013	-0.0349	-0.0095
4	-0.0028	0.0000	-0.0027	0.0007	-0.0128	-0.0254
5	0.0018	-0.0003	-0.0098	0.0093	-0.0062	0.0014

Next we will investigate if the MA(1) model does remove the first order negative serial correlation in the MA(1) filtered series. In Table 4.18 we summarize the results of the lag 1 to 5 autocorrelations performed on the residuals of the return series resulting of the MA(1) fit.

Table 4.18: Lag 1 to 5 autocorrelation values for the six shares selected for the MA(1) filtered series.

Lag	AGL	BIL	BVT	HAR	SBK	RCH
1	0.0042	0.0050	0.0002	0.0002	0.0018	0.0037
2	-0.0299	-0.0438	0.0080	-0.0132	-0.0309	-0.0382
3	-0.0170	-0.0282	-0.0066	0.0012	-0.0374	-0.0155
4	-0.0050	-0.0143	-0.0024	0.0008	-0.0152	-0.0269
5	0.0002	-0.0038	-0.0097	0.0094	-0.0077	-0.0011

Comparing Table 4.17 (original return series) to Table 4.18 (filtered MA(1) series) it is evident that negative serial first order correlation (lag 1) induced by the uneven spacing of the data points has been successfully removed by the MA(1) model (Note that the results in table 4.17 and 4.18 are practically significant). As a concluding remark we refer the reader back to Table 3.1. It is important to keep in mind that a positive θ results in a negative first order correlation. Further a MA(1) process with a positive θ should remove the negative first order correlation which seems indeed to be the case.

4.3.2 Statistical analysis

In Chapter 2 we defined realized variance estimators that could be utilized to estimate daily variance. Andersen et al. (2001) examined realized daily equity return variances obtained from high-frequency intraday transaction prices on individual shares in the Dow Jones Industrial Average. They found that the unconditional distributions of the realized variances are highly right (positive) skewed, while the realized logarithmic standard deviations are approximately Gaussian, as are the distributions of the returns scaled by realized standard deviations.

In this section we will research the first claim by Andersen et al. (2001) i.e. that the distributions of the daily returns scaled by daily realized standard deviations are Gaussian. We will focus on their second claim, i.e. that the unconditional distributions of the realized variances are highly right (positive) skewed, while the realized logarithmic standard deviations are approximately Gaussian. The characteristics of the return series as well as those of the scaled returns for various variance estimators will be researched. Probability distributions are characterised by their moments and the comparison of the daily return distributions of the shares included in our study will be in terms of their moments. We will limit ourselves to the first four moments; i.e. the mean, standard deviation, skewness and kurtosis.

Our main objective is to determine whether Andersen's findings also apply to South African shares. Our study will be similar to theirs; however, as far as the scaled returns are concerned we will investigate alternative realized daily volatility estimators. In particular, the VARHAC estimator will be compared with the ABDE and SSR estimator as well as with daily variance estimators produced by the GARCH and NIG-GARCH model proposed by Venter and de Jongh (2002,2004).

<u>**Distributional characteristics of the return series**</u>

In this section we will be discussing the distributional characteristics of the daily return series of the six South African shares.

Table 4.19: The table summarizes statistics of the daily return distributions for the six South African shares. The sample covers the period 22 Jun 2000 – 07 Mar 2003.

Share	Mean	Std dev	Skewness	Kurtosis
AGL	0.000573	0.026115	0.277421	5.501888
BIL	0.000642	0.026422	0.258687	5.258855
BVT	-0.000145	0.019792	-0.540328	7.519316
HAR	0.001709	0.040662	0.023996	4.548265
RCH	-0.000614	0.023818	0.211115	4.439421
SBK	0.000182	0.020684	0.291057	4.108891
Mean	**0.000391**	**0.026249**	**0.086991**	**5.229439**
Std dev	**0.000797**	**0.007568**	**0.322580**	**1.237932**

In the analysis that follows we will study the distributional characteristics of the daily return distributions in terms of the estimated moments of the daily return distribution. The skewness and kurtosis of a normal distribution is 0 and 3 respectively. We will estimate the skewness and kurtosis for the daily return distribution and then assess how close the estimates are to the theoretical values assuming a normal distribution. Statistical hypotheses tests to test the significance of the deviation of skewness and kurtosis from normality do exist. However, due to the large sample sizes (for AGL 690 trading days and 66240 data points) the significance tests will reject the null hypothesis (assuming the normal distribution holds) almost always. In the light of this and the paper by Anderson et al. we have decided to not include statistical significance tests and rather judge the closeness of the estimated skewness and kurtosis to the skewness and kurtosis of the normal distribution in a rather qualitative and ad hoc way.

In Table 4.19 we summarize the moments of the daily return distributions and show that the daily returns for the six South African shares have thicker tails than the normal distribution and, for the majority of shares (excluding BVT), are also positively skewed. This can be seen from the mean skewness (0.08) and the mean kurtosis which is 5.22.The kurtosis in the BVT return series is considerably higher than the measured kurtosis for the other shares selected for our study. The BVT return series has a smaller standard deviation than the other return series. The negative skewness and higher kurtosis in the BVT time series can be attributed to market moves pertaining to 11 and 12 Sep 2001 (terrorist attacks). For illustration purposes we removed the 11 and 12 Sep 2001 data points and BVT was subsequently positively skewed (0.08) and the kurtosis reduced to 3.32. Although the effects of the terrorist attacks were extreme the decision was made not to exclude the events from the data. Furthermore the large number of zero log returns cause the return distribution to become more peaked causing large kurtosis estimates.

Distributional characteristics of the standardized daily return series

In this section we discuss the distributional characteristics of the standardized return series for the realized daily volatility estimators and the GARCH volatility estimators. In Table 4.20 we summarize the daily standardized return distributions where the returns were standardized by the realized volatility estimators ABDE, SSR and VARHAC. The return sample covers the period 22 Jun 2000 – 07 Mar 2003 for all shares excluding BVT (which has a smaller sample which covers the period 22 Jun 2000 – 12 Apr 2002). The daily volatility estimates were calculated as described in Chapter 2.

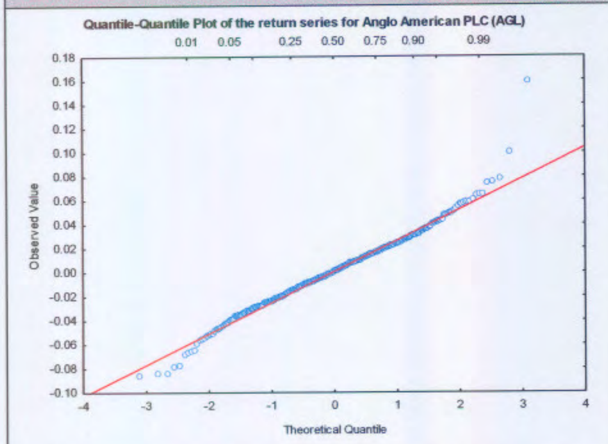
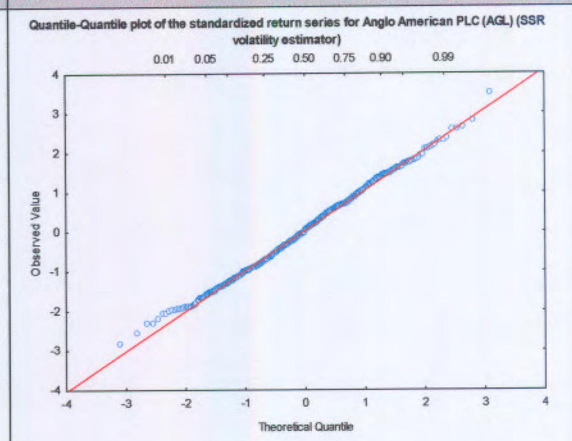
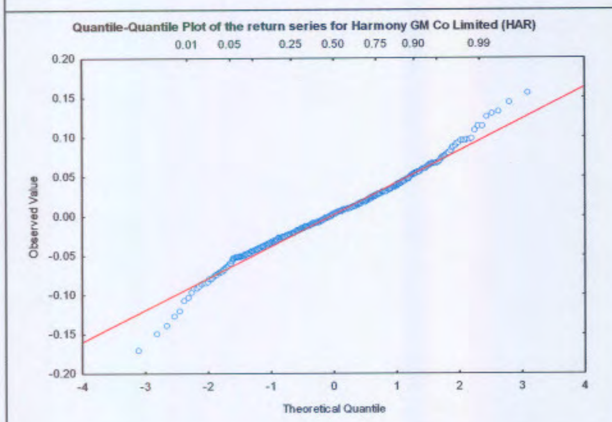
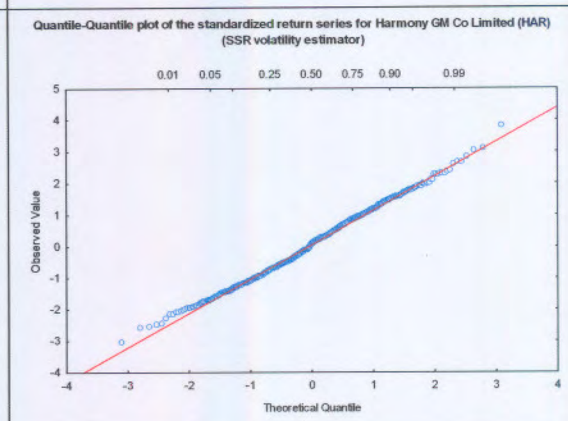
Table 4.20: The table summarizes the distributional characteristics of the daily returns scaled by the daily volatility estimates for the ABDE, SSR and VARHAC for the six South African shares. The sample covers the period 22 Jun 2000 – 07 Mar 2003.

ABDE				
Share	**Mean**	**Std dev**	**Skewness**	**Kurtosis**
AGL	0.028767	1.010508	0.090606	2.627027
BIL	0.032171	0.941347	0.661129	6.964460
BVT	-0.012994	0.965753	0.121167	2.865445
HAR	0.052888	1.085885	0.098065	2.662213
RCH	-0.001984	0.859970	0.121794	2.794276
SBK	0.010169	0.884413	0.128964	2.635494
Mean	**0.018169**	**0.957979**	**0.203621**	**3.424819**
Std dev	**0.024311**	**0.083035**	**0.224631**	**1.736699**
SSR				
Share	**Mean**	**Std dev**	**Skewness**	**Kurtosis**
AGL	0.030227	1.017135	0.095497	2.667703
BIL	0.032492	0.940867	0.636317	6.723254
BVT	-0.013237	0.965537	0.118112	2.857476
HAR	0.051676	1.087496	0.107317	2.661602
RCH	-0.008032	0.867901	0.128634	2.760298
SBK	0.010507	0.883945	0.133051	2.647123
Mean	**0.017272**	**0.960480**	**0.203154**	**3.386243**
Std dev	**0.025298**	**0.082682**	**0.212652**	**1.636750**
VARHAC				
Share	**Mean**	**Std dev**	**Skewness**	**Kurtosis**
AGL	0.024307	1.286061	0.096761	2.493821
BIL	0.050416	1.351754	0.739590	7.556016
BVT	-0.030536	1.180870	0.023761	2.915030
HAR	0.036456	1.682711	-0.091410	4.020711
RCH	-0.024430	1.136916	0.075806	2.774167
SBK	0.026146	1.135653	0.155444	2.730878
Mean	**0.013727**	**1.295661**	**0.166659**	**3.748437**
Std dev	**0.033299**	**0.208339**	**0.292757**	**1.940363**

Table 4.20 summarizes the distributional characteristics of the daily returns standardized by the ABDE, SSR and VARHAC daily volatility estimates. In Table 4.20 BVT is positively skewed for the various estimators in comparison to the negative skewness as stated in Table 4.19. On both 11 and 12 Sep 2001 when higher volatility estimates in comparison to the other days were obtained, standardizing the daily returns by the daily volatility estimates resulted in a shift in skewness. HAR, in contrast, is negatively skewed for the VARHAC volatility estimates in comparison to the raw return series (as stated in Table 4.19). The latter characteristic can be ascribed to the fact that on 3 Nov 2000 a small VARHAC volatility estimate was obtained and when the estimate is standardized a large value is obtained which contributes to the negative skewness (as a test we removed the data point and the skewness was subsequently positive). Note that a negative return divided by a very small volatility estimate will result in a large negative standardized return.

Consistent with the findings of Andersen et al. (2001) we found the mean value (over the six shares) of the kurtosis to be closer to normality (a mean skewness equal to 0 and a mean kurtosis close to 3). This is true in all three cases (ABDE, SSR and VARHAC). For the returns standardized by the ABDE the mean kurtosis dropped from 5.22 (for the unstandardized returns) to 3.42. The mean kurtosis for the SSR volatility estimates reduced to 3.38 and the mean kurtosis for the VARHAC volatility estimates to 3.74. For almost all of the shares selected the mean kurtosis appeared to be considerably lower for the standardised return series than for the raw return series discussed in the previous section and closer to 3 and therefore normality (refer to Table 4.21).

Table 4.21: Quantile-Quantile plot of the daily unstandardized returns vs. the returns standardized by the SSR daily volatility estimates.

Unstandardized returns	Standardized daily returns
 	
 	

From the QQ plots in Table 4.21 it is clear that the unstandardized return series exhibits heavy tailed behaviour and that the standardized return series is much closer to normality than the unstandardized return series.

We continue with a discussion of the return distributions standardized by the GARCH and NIG-GARCH volatility estimators. The return sample covers the period 22 Jun 2000 – 07 Mar 2003 for all shares excluding BVT which has a smaller sample covering the period 22 Jun 2000 – 12 Apr 2002.

Table 4.22: The table summarizes distributional characteristics of the daily returns standardized by the daily volatility estimates for the GARCH and NIG-GARCH) for the six South African shares. The sample covers the period 22 Jun 2000 – 07 Mar 2003.

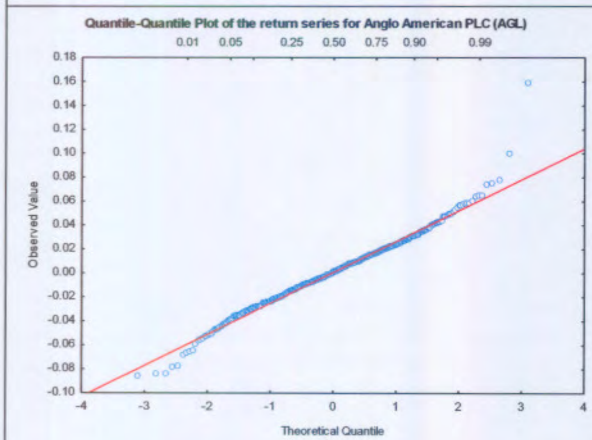
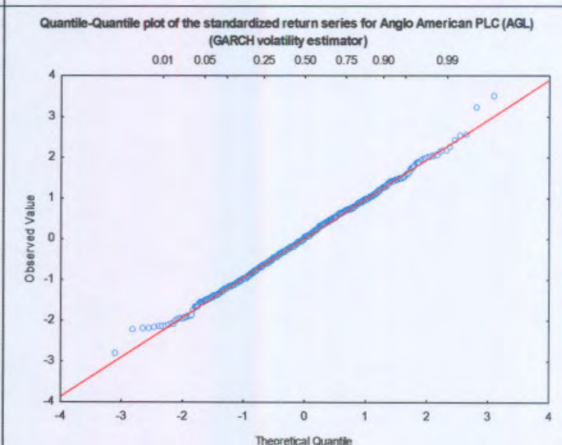
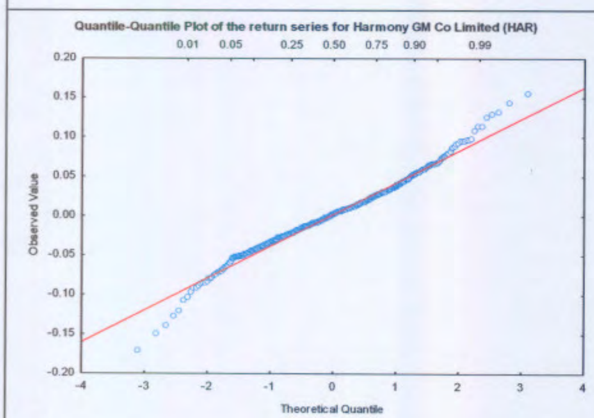
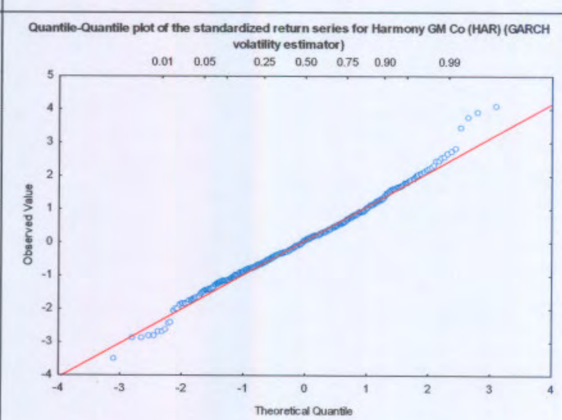
GARCH				
Share	**Mean**	**Std dev**	**Skewness**	**Kurtosis**
AGL	0.015805	0.969527	0.078328	2.947205
BIL	0.021762	1.068888	0.224996	4.319410
BVT	-0.023423	1.050263	-0.874099	9.606624
HAR	0.055624	1.032197	0.226128	3.948542
RCH	-0.028553	1.030799	0.128768	3.703410
SBK	-0.003220	1.027690	0.063266	3.257886
Mean	**0.006333**	**1.029894**	**-0.025436**	**4.630513**
Std dev	**0.031468**	**0.033438**	**0.421579**	**2.485989**
NIG-GARCH				
Share	**Mean**	**Std dev**	**Skewness**	**Kurtosis**
AGL	0.015415	0.968753	0.075269	2.945523
BIL	0.022245	1.069159	0.214590	4.348190
BVT	-0.030755	1.061373	-1.055654	11.042063
HAR	0.057180	1.035717	0.252989	4.030744
RCH	-0.027829	1.028027	0.156790	3.767402
SBK	-0.003070	1.024448	0.062191	3.299443
Mean	**0.005531**	**1.031246**	**-0.048971**	**4.905561**
Std dev	**0.033311**	**0.035572**	**0.498836**	**3.047930**

Almost all of the GARCH and NIG-GARCH standardized returns are skewed to the right (positive), excluding BVT (negative skewness due to 11 and 12 Sep 2001 as well as a small GARCH and NIG-GARCH volatility estimates). Similar to our finding for the realized volatility estimators there has been an observable decrease in the mean kurtosis for both the GARCH and NIG-GARCH standardized returns in Table 4.22, once again excluding BVT which has in fact increased in kurtosis from the previously calculated return series in Table 4.16. The GARCH mean kurtosis decreased to 4.6 and the NIG-GARCH mean kurtosis to 4.9 (in comparison to the mean kurtosis of 5.22 as stated in Table 4.19).

When comparing Table 4.22 with the results in Table 4.20 it seems that the returns scaled by the realized volatility estimators are closer to normality than those scaled by the GARCH volatility estimators (the mean kurtosis in the GARCH (4.6) and NIG-GARCH (4.9) scaled returns are further removed from normality in comparison to the mean kurtosis of the ABDE (3.42), SSR (3.38) and VARHAC (3.74) scaled returns).

In Table 4.23 we depict the QQ plots of the daily unstandardized return series vs. the returns standardized by the GARCH volatility estimates for Anglo American PLC (AGL) as well as Harmony GM Co Limited (HAR).

Table 4.23: Quantile-Quantile plot of the daily unstandardized returns vs. the returns standardized by the GARCH daily volatility estimates.

Unstandardized returns	Standardized daily returns
	
	

From the above QQ plots in Table 4.23 it is evident that the standardized return series is much closer to normality. It is not clear from the QQ plots analyzed if the distribution of the return series standardized by the realized variance estimators or the return series standardized by GARCH and NIG-GARCH estimators is closer to normality.

Distributional characteristics of the daily realized variances

The first four columns in Table 4.24 and 4.27 provide the same set of statistics for the daily realized variances of the various variance estimators as discussed in the previous section. Again the sample covers the period 22 Jun 2000 – 07 Mar 2003 for all shares except BVT which has a smaller sample covering the period 22 Jun 2000 – 12 Apr 2002.

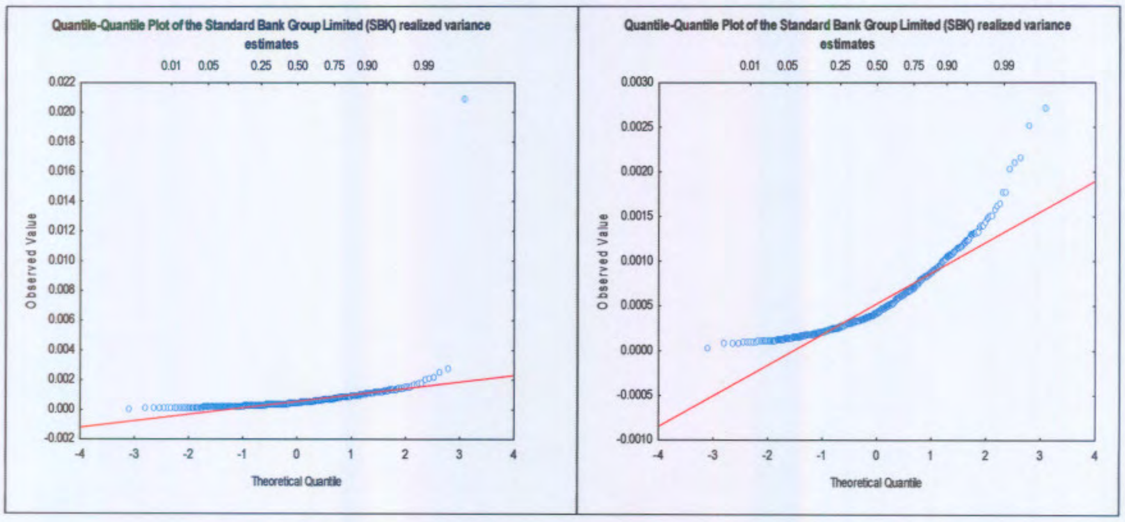
Table 4.24: The table summarizes the distributional characteristics of the ABDE, SRR and VARHAC realized daily variances for the six South African shares. The sample covers the period 22 Jun 2000 – 07 Mar 2003.

ABDE				
Statistic	**Mean**	**Std dev**	**Skewness**	**Kurtosis**
AGL	0.000691	0.000654	3.372832	21.974654
BIL	0.000857	0.000754	3.103987	17.279632
BVT	0.000427	0.000487	7.560407	95.766109
HAR	0.001523	0.002030	5.685877	61.400320
RCH	0.000745	0.000671	3.153102	21.848096
SBK	0.000559	0.000886	19.065054	436.706342
Mean	**0.000800**	**0.000914**	**6.990210**	**109.162526**
Std dev	**0.000384**	**0.000562**	**6.176404**	**163.348489**
SSR				
Statistic	**Mean**	**Std dev**	**Skewness**	**Kurtosis**
AGL	0.000699	0.000673	3.416848	22.631433
BIL	0.000867	0.000771	3.192675	18.458024
BVT	0.000427	0.000492	7.766571	100.135716
HAR	0.001522	0.002031	5.685007	61.343991
RCH	0.000751	0.000682	3.191452	22.035414
SBK	0.000560	0.000887	19.037280	435.818192
Mean	**0.000804**	**0.000923**	**7.048305**	**110.070462**
Std dev	**0.000384**	**0.000558**	**6.148800**	**162.710260**
VARHAC				
Statistic	**Mean**	**Std dev**	**Skewness**	**Kurtosis**
AGL	0.000399	0.000366	3.280452	20.242392
BIL	0.000451	0.000462	4.143349	29.857703
BVT	0.000411	0.002195	19.925505	407.026122
HAR	0.000796	0.001329	10.415815	156.591512
RCH	0.000459	0.000424	2.759907	16.077539
SBK	0.000349	0.000295	2.939857	19.922111
Mean	**0.000478**	**0.000845**	**7.244148**	**108.286230**
Std dev	**0.000161**	**0.000763**	**6.853283**	**156.071993**

It is evident from the third and fourth columns of Table 4.24 that the distributions of the realized variances are positively skewed and leptokurtic for all the shares and for all the realized variance estimators.

For the ABDE and SSR variance estimates the skewness and kurtosis in SBK is considerably more than the skewness and kurtosis in the other shares selected for the study. The extreme data point contributing to the considerably higher kurtosis is the terrorist attacks of 11 Sep 2001. If the latter point is removed from the SBK data series, the skewness and kurtosis for the ABDE variance estimate reduces to 1.67 and 4.34 respectively.

Table 4.25: Quantile-Quantile Plot of the Standard Bank Group Limited (SBK) ABDE realized variance estimates including and excluding the data point 12 Sep 2001.



From the QQ plots in Table 4.25 it is clear that the distribution of the realized variances is far removed from normality.

In Table 4.26 below we also see that the GARCH and NIG-GARCH variance estimates are positively skewed with thick tails. This result agrees with the findings of the Andersen et al. (2001) study; namely that the distributions of the variances are right skewed and leptokurtic.

Table 4.26: The table summarizes the distributional characteristics of the GARCH and NIG-GARCH realized daily variances for the six South African shares. The sample covers the period 22 Jun 2000 – 07 Mar 2003.

GARCH				
Statistic	**Mean**	**Std dev**	**Skewness**	**Kurtosis**
AGL	0.000633	0.000272	2.450342	10.594605
BIL	0.000642	0.000318	5.715313	53.544178
BVT	0.000364	0.000182	6.075821	55.943438
HAR	0.001606	0.001082	2.638014	11.749660
RCH	0.000535	0.000243	1.473137	5.412650
SBK	0.000404	0.000151	1.503253	6.076192
Mean	**0.000697**	**0.000374**	**3.309313**	**23.886787**
Std dev	**0.000460**	**0.000352**	**2.062245**	**24.040352**
NIG-GARCH				
Statistic	**Mean**	**Std dev**	**Skewness**	**Kurtosis**
AGL	0.000632	0.000268	2.500292	10.941599
BIL	0.000637	0.000306	6.094646	60.114754
BVT	0.000378	0.000252	5.403744	45.533518
HAR	0.001644	0.001203	2.703530	12.273562
RCH	0.000539	0.000248	1.652614	6.436458
SBK	0.000405	0.000142	1.292099	5.146295
Mean	**0.000706**	**0.000403**	**3.274487**	**23.407697**
Std dev	**0.000472**	**0.000396**	**1.998636**	**23.400147**

Lastly, in order to compare the variances of the six shares we annualized the average daily variances for the six shares. The results are shown in Table 4.27. We describe the process with an example

AGL annualized GARCH value = $\left(\sqrt{(0.000272 * 252)}\right) * 100 = 26.16\%$

Table 4.27: Annualized average daily variance estimates for the different variance estimation techniques (ABDE, GARCH, NIG-GARCH and VARHAC).

Shares	**ABDE**	**SSR**	**VARHAC**	**GARCH**	**NIG-GARCH**
AGL	40.60	41.18	30.35	26.16	25.97
BIL	43.58	44.08	34.12	28.32	27.76
BVT	35.04	35.23	74.38	21.40	25.18
HAR	71.53	71.54	57.88	52.21	55.07
RCH	41.12	41.44	32.69	24.75	24.97
SBK	47.26	47.28	27.28	19.49	18.94
Average	**46.52**	**46.79**	**42.78**	**28.72**	**29.65**

For the ABDE variance estimator an annualized variance of around 46.5 % was found for the sample of six shares in our study. There is considerable variation in the average variance of the six shares, ranging from a high of 71.5% for HAR (the most volatile share in our subset) to a low of 35 % for BVT (similar results were obtained for SSR). Although the average for the VARHAC annualized variance estimates is close to the other two realized variance estimators, the results for BVT stand in sharp contrast. As expected, if we exclude BVT, the VARHAC annualized variance estimates for the remaining five shares are lower than the ABDE and SSR variance estimates and the GARCH and NIG-GARCH methodologies produced the lowest (due to the smoothing) variance estimations in our study.

Distributional characteristics of the realized logarithmic standard deviations

In the next section we discuss the distributional characteristics of the realized logarithmic standard deviations of the six shares. As mentioned previously the sample covers the period 22 Jun 2000 – 07 Mar 2003 for all shares excluding BVT which has a smaller sample which covers the period 22 Jun 2000 – 12 Apr 2002.

Table 4.28: The table summarizes the distributional characteristics of the ABDE, SSR and VARHAC logarithmic standard deviations for the six South African shares. The sample covers the period 22 Jun 2000 – 07 Mar 2003.

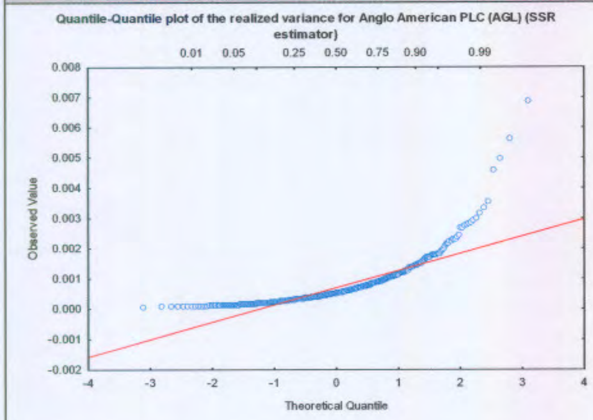
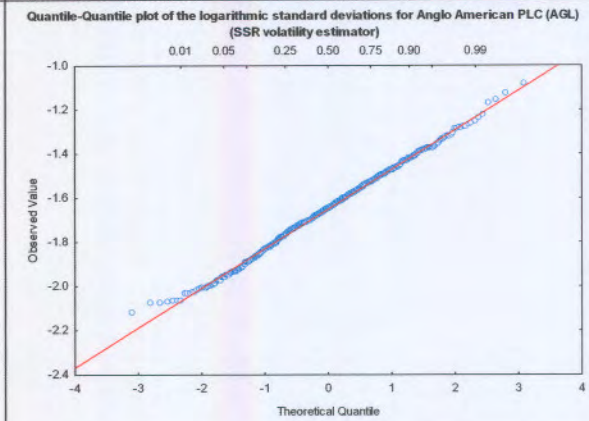
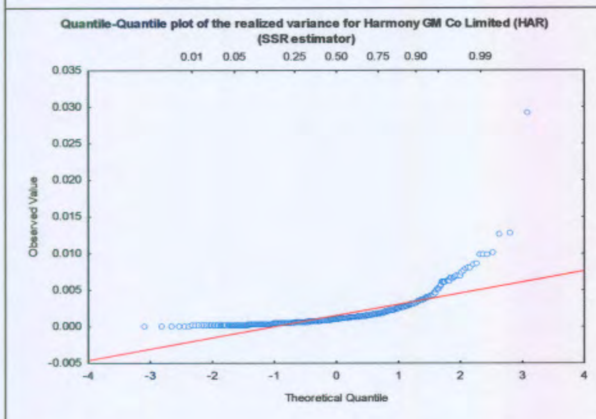
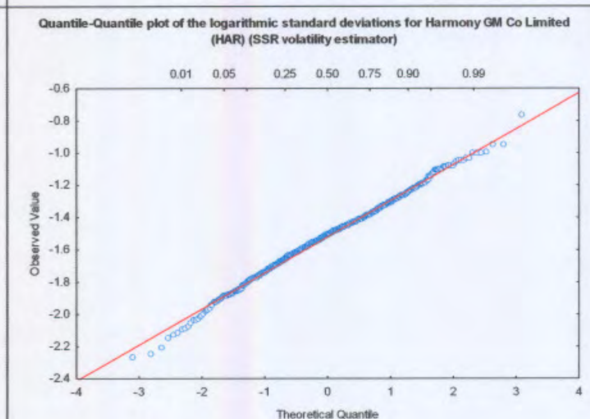
ABDE				
Share	**Mean**	**Std dev**	**Skewness**	**Kurtosis**
AGL	-1.650968	0.176151	-0.020977	2.840975
BIL	-1.592752	0.159455	-0.057112	3.978968
BVT	-1.753479	0.167962	0.072016	3.551931
HAR	-1.520065	0.223683	-0.177631	3.371189
RCH	-1.631959	0.175106	-0.129911	3.043227
SBK	-1.686739	0.152297	0.134534	3.967720
Mean	**-1.639327**	**0.175776**	**-0.029847**	**3.459002**
Std dev	**0.079826**	**0.025197**	**0.118382**	**0.469112**
SSR				
Share	**Mean**	**Std dev**	**Skewness**	**Kurtosis**
AGL	-1.650604	0.179097	-0.025440	2.818908
BIL	-1.591047	0.160384	-0.048575	3.964222
BVT	-1.753638	0.168112	0.074295	3.563741
HAR	-1.520097	0.223695	-0.177911	3.373174
RCH	-1.630923	0.175655	-0.117770	3.021237
SBK	-1.685875	0.151951	0.141775	3.979569
Mean	**-1.638697**	**0.176482**	**-0.025604**	**3.453475**
Std dev	**0.079970**	**0.025169**	**0.118503**	**0.478696**
VARHAC				
Share	**Mean**	**Std dev**	**Skewness**	**Kurtosis**
AGL	-1.769065	0.180534	-0.388038	3.636257
BIL	-1.745055	0.177315	-0.111000	3.592781
BVT	-1.838260	0.204963	0.168771	5.285080
HAR	-1.675604	0.248849	-0.513481	3.513945
RCH	-1.746208	0.190307	-0.300758	3.212720
SBK	-1.794173	0.177002	-0.421919	3.522805
Mean	**-1.761394**	**0.196495**	**-0.261071**	**3.793931**
Std dev	**0.054588**	**0.027762**	**0.250824**	**0.745460**

Table 4.28 gives summary statistics of the logarithmic realized standard deviations. Our findings are consistent with Andersen et al. (2001) in that the mean skewness in Table 4.28 for the three estimators has been reduced in comparison to the mean skewness in Table 4.24. The ABDE mean skewness for the realized variance was 6.99 in comparison to -0.02 for the mean logarithmic realized standard deviations (similar results were obtained for the SSR and VARHAC) estimators.

The mean kurtosis of the realized logarithmic standard deviations (over all the shares) for the ABDE and SSR realized volatility estimators are reasonably close to the normal value of three (3.45) while the mean kurtosis of the realized variance for the same estimators is far removed. The mean skewness of the realized logarithmic standard deviations (over all the shares) for the ABDE and SSR realized volatility estimators are reasonably close to the normal value of zero (-0.02) while the mean skewness of the realized variance for the same estimators is far removed (7.0).Therefore, one can conclude that the distribution of the realized logarithmic standard deviations for the ABDE and SSR estimators are much closer to the normal distribution than the corresponding distributions for the realized variances.

In Table 4.29 we graphically depict QQ plots for the realized variances and realized logarithmic standardized deviations for Anglo American PLC (AGL) and Harmony GM Co Limited (HAR) for the SSR volatility estimations. It is evident from the QQ plots that the assumption of normality holds much better for the realized logarithmic standard deviations.

Table 4.29: QQ plots of the realized variances vs. the realized logarithmic standard deviations for Anglo American PLC (AGL) and Harmony GM Co Limited (HAR) (SSR volatility estimations).

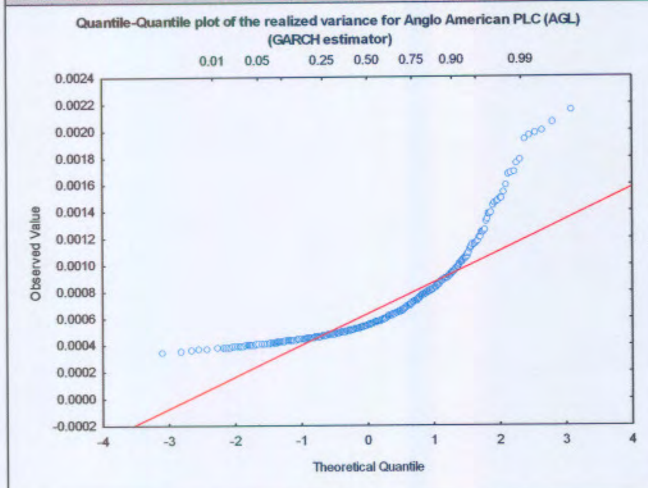
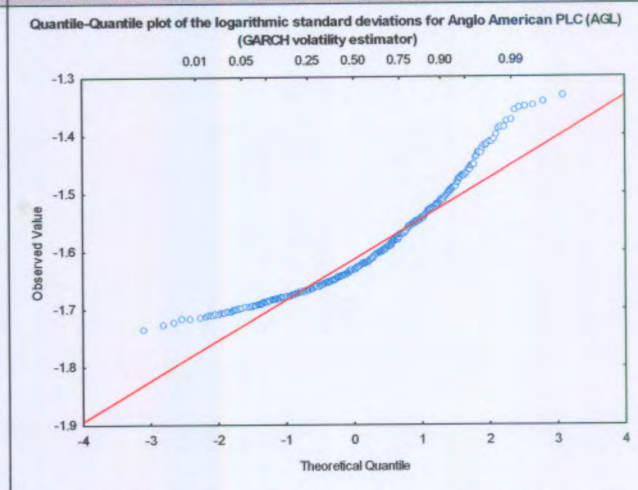
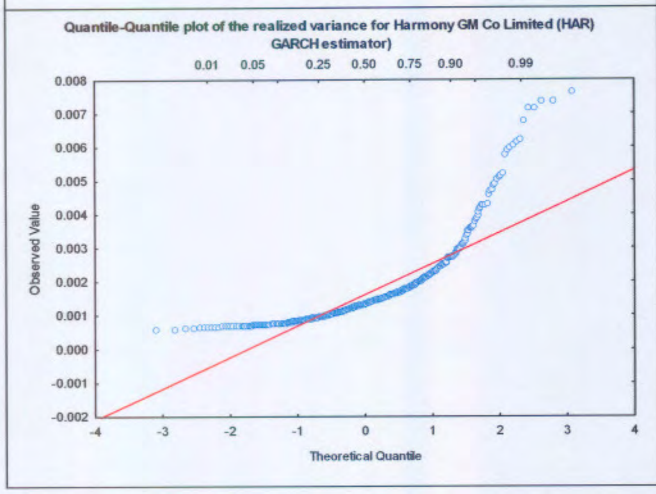
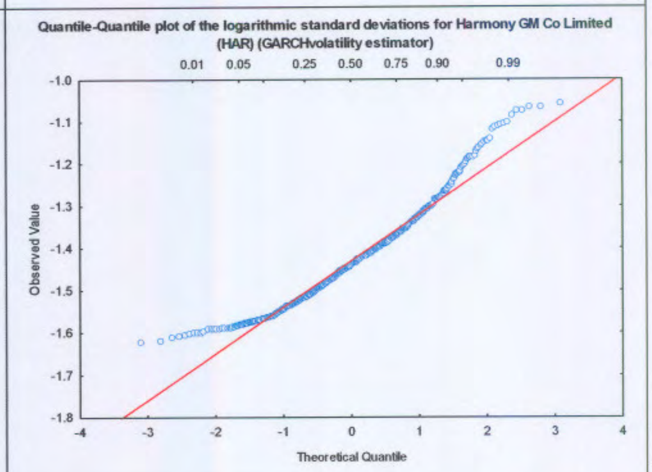
<table>
<tr><th>Realized variances</th><th>Realized logarithmic standard deviations</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

Similar to the results for the realized variance estimators, the mean skewness of the GARCH variance estimates reduced if Table 4.26 is compared with Table 4.30. The GARCH variance mean skewness was 3.30 in comparison to the logarithmic standard deviation mean skewness of 1.17. In contrast to the finding in the previous section the assumption of normality doesn't hold for all the logarithmic standard deviations. The QQ plots for the GARCH variances and GARCH logarithmic standard deviations for Anglo American Limited (AGL) and Harmony GM Co Limited (HAR) are shown in Table 4.31.

Table 4.30: The table summarizes the GARCH and NIG-GARCH logarithmic standard deviations for the six South African shares. The sample covers the period 22 Jun 2000 – 07 Mar 2003.

GARCH				
Share	**Mean**	**Std dev**	**Skewness**	**Kurtosis**
AGL	-1.613473	0.073383	1.211883	4.446977
BIL	-1.610120	0.068272	2.053883	9.402860
BVT	-1.733357	0.068460	1.880974	9.540902
HAR	-1.430080	0.111995	0.787551	3.530616
RCH	-1.654540	0.087733	0.535117	2.597627
SBK	-1.709946	0.072748	0.550963	2.868701
Mean	**-1.625253**	**0.080432**	**1.170062**	**5.397947**
Std dev	**0.107868**	**0.017026**	**0.666489**	**3.219738**
NIG-GARCH				
Share	**Mean**	**Std dev**	**Skewness**	**Kurtosis**
AGL	-1.613325	0.072112	1.253110	4.589548
BIL	-1.610535	0.065082	2.218241	10.496016
BVT	-1.735524	0.092879	1.237659	5.853962
HAR	-1.430740	0.121612	0.704018	3.368371
RCH	-1.653082	0.086975	0.614045	2.770681
SBK	-1.708204	0.070161	0.453014	2.668654
Mean	**-1.625235**	**0.084804**	**1.080014**	**4.957872**
Std dev	**0.107712**	**0.020924**	**0.648383**	**2.971885**

Table 4.31: QQ plots of realized variances vs. the realized logarithmic standard deviations for Anglo American PLC (AGL) and Harmony GM Co Limited (HAR) (GARCH volatility estimators).

Realized variances	Realized logarithmic standard deviations
	
	

It is evident from the QQ plots that the distributions of the logarithmic standard deviations are closer to normality than the distributions of the realized variances, but that significant deviations from normality are still present.

We can therefore conclude that our findings in the South African market context do agree with the claims made by Andersen et al. (2001) for the realized variance estimators.

4.4 THE IMPACT OF VOLATILITY ON THE RISK MANAGEMENT OF OPTIONS

In the following section we introduce the reader to the importance of using the correct volatility when pricing, valuing and managing a portfolio of options. There we use a practical example to illustrate the sensitivity of an option to changes in the volatility. The aim of the section is to illustrate the impact that a change in volatility has on the valuation and risk management of options.

In theory, good estimators have the following properties:
Firstly, they are consistent i.e. they converge towards the "true" or population parameter value as the sample used in the calculation increases in size. Secondly, they have a small standard error; of two unbiased estimators, the more desirable is the one with the smaller variance (noise). High kurtosis is generally indicative of a heavy tailed distribution and that the data from such a distribution show a greater propensity for extreme behaviour. The above is from a statistical perspective. However, what are the requirements of "good" volatility estimation in a derivatives trading environment, and are these different from the enumerated attributes? The reality of a trading environment can be summarised as follows:

Derivatives replication (synthesising) is a standard practice, and is also the main purpose of most derivative pricing models (providing hedging ratios). A financial institution that trades derivative products in the over the counter market (OTC) as well as listed instruments (exchange traded) has to manage the resulting risks. One of the easiest ways of managing the risk is by trading the equal and opposite option, an alternative which may not always be a feasible solution especially in the OTC market where options are tailored to a client's specific need. Therefore it may be necessary for the trader to create the options synthetically in order to hedge the position with the underlying share. This strategy involves taking a position in the underlying asset (or future on the underlying asset) so that the *delta* (defined as the rate of change of the option price with respect to the price of the underlying asset) of the underlying position is maintained equal to the delta of the required option. To create a put option synthetically, the trader should ensure that at any given time a certain proportion of share delta in the original portfolio should have been sold and the proceeds invested in a risk-less asset. As the value of the original portfolio declines, the delta of the put becomes more negative and the proportion of shares sold must be increased. As the value of the original portfolio increases, the delta of the put becomes less negative and it will be necessary to buy shares back (see e.g., Hull, 1989).

Most derivative traders use sophisticated hedging schemes and these involve calculating measures such as the delta, theta, gamma and vega (also referred to as the "Greeks" of an option) of the option portfolio they are managing.

The option Greeks can be defined as follows (see e.g., Hull, 1989, p 310-329)

- *Delta* - defined as the rate of change of the option price with respect to the price of the underlying asset,

- *Theta* - rate of change of the option price with respect to the passage of time when all else stays the same,

- *Gamma* - defined as the rate of change of the portfolio's delta with respect to the price of the underlying asset. If the gamma is small the delta changes slowly and adjustments to keep the portfolio delta neutral need only be made infrequently. However if the gamma is large the delta will be very sensitive to the underlying asset and it is risky to leave a delta neutral portfolio unchanged for any length of time.

- *Vega* – rate of change of the value of the portfolio with respect to the volatility of the underlying asset. Vega is an indication to the sensitivity of the portfolio to small changes in volatility

In Table 4.32 we illustrate the sensitivity of the hedge model to changes in the underlying volatility (using a one year, R1 000 000 nominal, at the money (R55.20) strike Harmony call option with a Black and Scholes pricing model). Suppose the trader bought the option but has erroneously marked the option at a level of 25 % instead of 45 %.

Table 4.32: Harmony GM Co Limited (HAR) Delta and Vega change implications as the underlying volatility is changed.

Annualized volatility	Price	Delta	Gamma	Vega	Theta
25%	6.88560	12573	-532	-3693	223
45%	10.75449	12429	-300	-3717	311

The impact on the option would be as follows:

- The option is worth less than when you bought it (lower price)

- In order to keep the portfolio delta neutral it is necessary to buy more of the underlying share. Normally this may not be an issue except if the share is illiquid or the trader may be close to the limits implied by the risk department. Furthermore the trader may not be able to trade the delta due to the fact that the execution of the hedge may push the trader over his allocated limits.

- The gamma exposure increases to 532 and if a change of R10 in the price of the underlying asset occurs, there would be a decrease in the value of the portfolio of approximately R26 800.

- If the option was marked at 25 % the theta earned on the option per day would be less (R223 per R1 million)

In an ideal world traders would rebalance their portfolios frequently to maintain zero delta, zero gamma etc. When managing a large and diverse portfolio the latter may not be possible and the traders may only be able to delta hedge once a day. It is therefore very important to have accurate estimates of volatility as incorrect estimates can lead to over or under trading of the Greeks which in turn have an impact on the profitability of a portfolio.

Good replication hedging results in a low variance of daily profit and loss and more specifically in no extreme profit and losses, i.e. the distribution of the profit and loss for a well-replicated derivatives portfolio should be concentrated around zero, with very little propensity for outliers. The general practice of stop-loss limits makes for a high priority as concerns the frequency of outlying profit and loss. Hedge models perform well when they accurately capture the dynamics of the underlying asset's behaviour over time. This implies that assumptions made about the distribution of returns must be honoured; the parameters supplied to the model also have to conform (be as close as possible) to actual (empirical) parameter values. The volatility estimators studied represent an attempt to a closer tracking of one of the inputs required by most derivatives models, namely volatility. Now, it is generally preferable to keep hedge errors as small as possible; this is achieved by having a hedge model that incorporates the empirical behaviour of the underlying asset and volatility estimation. A model that estimates volatility accurately is important for various reasons. Firstly, as illustrated in the above example, the hedge is dependent on the volatility used (the required hedge for the option will change as the option volatility is changed). Incorrect volatility estimates will result in an over or under hedged position with cost and risk implications. Secondly, if incorrect volatilities are used the profit and loss and risks associated with the option may not be an accurate reflection of the truth. In the past inaccurate models and procedures have caused the financial ruin of big financial institutions (such as Barings or Long Term Capital).

4.5 CONCLUSION

In this chapter we examined the realized daily equity return volatilities obtained from high-frequency intraday transaction prices on six South African shares. Similar to the Andersen et al. (2001) study we found the unconditional distributions of the realized variances to be positively skewed (right skewed), while the realized logarithmic standard deviations are approximately Gaussian, as are the distributions of the returns scaled by realized standard deviations or realized volatilities. This conclusion holds for all the realized variance estimators (ABDE, SSR and VARHAC). The distribution of the returns scaled by the GARCH and NIG-GARCH volatility estimates are also close to normality. However, the distributions of the daily variance estimates and the log of the volatility (standard deviations) estimates based on the GARCH and NIG-GARCH models are clearly not normally distributed. We therefore conclude that our findings in the South African market context do agree with the claims made by Andersen et al. (2001) for the realized volatility estimators and that the realized volatility techniques are a much better match to the criteria as set out by Andersen et al. (2001) than the GARCH and NIG-GARCH volatility estimates.

CHAPTER 5

CONCLUSION AND SUGGESTIONS FOR FURTHER RESEARCH

5.1 SUMMARY AND CONCLUSIONS

Accurate measures and good forecasts of volatility are critical for asset pricing, option pricing, asset allocation, portfolio selection, portfolio rebalancing and hedging strategies as well as various risk management applications. Most textbooks assume volatility to be constant; however in practice this is a very dangerous assumption to make and has lead to a research program regarding the distributional and dynamic properties of financial markets. Given that financial markets display high speeds of adjustment, studies based upon historical close to close observations (end of day) may fail to capture information contained in intraday or high frequency market movements and until relatively recently the use of daily or equally spaced data was considered the highest meaningful sampling frequency for financial market data. In the financial markets, market participants determine their trading decisions by observing high frequency or tick-by-tick data, and until recently most of the studies published in financial literature deal with low frequency (usually end of day close to close), regularly spaced data.

In the thesis we discussed the process of constructing a reliable high frequency return series and we gave the reader an overview of the factors that drove the price process that was eventually observed. In order to filter the data of errors we had to understand the factors generating the erroneous data points as well as market characteristics that contributed to the underlying price series. The latter process of filtering the data of potential erroneous data points was a time consuming and labour intensive task which had to be performed in order to obtain good quality high frequency data. Once the data had been filtered we could proceed with the preparation of the different time series formats (e.g., 5 minute etc.) as required by the different volatility estimators we were investigating.

Our study was conducted along the lines of the study by Andersen et al. (2001), utilizing South African equity data, however, as far as the scaled returns were concerned we investigated alternative volatility estimators for daily return volatilities. In particular, we compared the VARHAC estimator (see e.g., Bollen and Inder (2002)) to the ABDE estimator (see e.g., Andersen et al. (2001)) as well as with daily volatility estimators based on the NIG-GARCH model proposed by Venter and de Jongh (2004).

Our findings for all the realized variance estimators studied namely, ABDE, SSR and VARHAC were consistent with the Andersen et al. (2001) study i.e. the right skewed distributions of the variances, the normal distributions of the logarithmic standard deviations and the normal distributions of daily returns standardized by realized standard deviations. The distribution of the returns scaled by the GARCH and NIG-GARCH volatility estimates were close to normality. However, the distributions of the daily variance estimates and the log of the volatility (standard deviations) estimates based on the GARCH and NIG-GARCH models are clearly not normally distributed.

We can therefore conclude that our findings in the South African market context do agree with the claims made by Andersen et al. (2001) for the realized volatility estimation techniques and that the realized volatility techniques are a much better match to the criteria as set out by Andersen et al. (2001) than the GARCH and NIG-GARCH volatility estimates.

5.2 SUGGESTIONS FOR FURTHER RESEARCH

We conclude the thesis with suggestions for further research:

Optimal sampling frequency

A study could be undertaken to ascertain the "optimal" choice in sampling frequency, e.g., 5 minute, 10 minute etc. in a South African market context with regards to the construction of the realized variation estimation techniques discussed in this thesis and the effect that the choice of sampling frequency would have on the distributional and dynamic properties of these estimators. For previous research conducted with regards to high frequency realized variance estimators see e.g., Andersen et al. (2001), Ghysels et al. (2003), Giot and Laurent (2004) and Oomen (2001).

Sensitivity of the VARHAC estimator

On some of the trading days included in the study the tick by tick data (as utilized by the VARHAC estimator) was more volatile than the corresponding 5 minute data (as utilized by the ABDE estimators) and this latter characteristic seems to have contributed to the VARHAC variance estimates being higher than the ABDE variance estimates. A further study could investigate the sensitivity of the VARHAC variance estimator to large market moves and extreme data behaviour.

Market limits

Many markets impose limits (volatility bands) on the amount (percentage) asset prices can change within a trading day to prevent the market from overreacting and, hence, to dampen volatility see e.g., Cho et al. (2003). Fama (1989) argued that price limits could have an adverse effect on the equity market. In the South African equity market (May 2002) as well as the agricultural derivative markets the price limit principle has been applied. An interesting possibility for future research would be to investigate if the imposed price limits do have the desired effects of prevention (market overreaction) and dampening of volatility. The latter effects may not always be in the markets best interest as artificial volatility levels are introduced and market participants may be forced to execute deals at unrealistic market levels.

Economic value of realized volatility in investment

Recently Fleming et al. (2002) attempted to measure the economic value of the realized volatility and correlation approach in the context of investment decisions. They evaluated the empirical performance of conditionally mean-variance efficient portfolios that are rebalanced daily based on estimates of the conditional covariance matrix. They used daily as well as intraday returns to estimate the covariance matrix and assumed the target expected return to be constant. The investors that they consider are assumed to follow a volatility-timing strategy where the weights vary only with changes in the covariance matrix. The basic assumption is that more precise estimates of the conditional covariance matrix will improve the performance of the volatility-timing strategies. *They found that an investor, implementing a volatility-timing strategy, would be willing to pay on the order of 50 to 100 basis points per year to switch from a daily returns based estimator of the daily conditional covariance matrix to an estimator based on realized volatility.* Moreover, they claim that these benefits are not restricted to short-horizon investors. In fact they claim that cumulative gains to volatility timing at the daily level are substantial for performance measurement horizons as long as one year. Moreover these gains are robust to transaction costs and estimation risk regarding expected returns. This should be investigated in a South African context.

Forecasting and Value at Risk

The assumption that financial market volatility is predictable has important implications for asset pricing and portfolio management. Investors seeking to avoid risk may choose to adjust their portfolios by reducing their commitments to assets whose volatilities are predicted to increase or by hedging against the possibilities of such an occurrence. Due to the increasingly important role played by Value at Risk (VAR) in risk assessment, it is becoming more important to have a good measure and forecast short-term volatility. Because volatility is a key input to VAR models, the characterization of asset volatility is of paramount importance when implementing and testing VAR models. The literature on VAR models has grown over the last decade mostly due to the popular RiskMetrics VAR specification of JP Morgan and the risk adjusted measures of capital adequacy enforced by the Basel committee. VAR has become widely used by corporate treasurers, fund managers and financial institutions. The South African Reserve Bank also requires local banks and financial institutions to report VAR figures as a measure to ensure that adequate capital is being held. A study investigating the feasibility of high frequency volatility estimates in a VAR environment could be undertaken in a South African context.

Accurate measures and good forecasts of volatility are critical for the implementation and evaluation of derivative pricing, trading, hedging strategies and risk management. In the past, conventional ARCH models were used but recently the focus has shifted to incorporate tick by tick (high frequency) data when forecasting volatility (see e.g., Engle and Sun (2004), Hol and Koopman (2002)). In this thesis the focus was on the estimation of daily volatility. A follow up study could also focus on the forecasting performance of the realized volatility estimators (see e.g., Areal and Taylor (2002), Hol and Koopman (2002), Oomen (2001), Thomakos and Wang (2003)) in a high frequency environment.

High frequency GARCH

In this thesis we concentrated on the use of high frequency data in a realized variance framework. GARCH models may also incorporate high frequency data (see e.g., Venter et al. (2005)). These methods of estimating and forecasting volatilities should be compared with realized volatility techniques.

Financial markets

Share correlation is an important aspect of equity trading and further attention should be given to this topic in the South African market (in this study we did not investigate the distributional and dynamic properties of correlations between different shares (see e.g., Andersen et al. (2001), Djupsjöbacka (2002) and Giot and Laurent (2004)).

The results and conclusions of this study do not have to be restricted to a South African equity market context but the application thereof could be investigated for a wide range of financial instruments traded in the South African financial markets. Previous research (over a diverse range of financial instruments) in the international markets include:

- Indices - see e.g., Ebens (1999), Speight et al. (2000)
- Futures - see e.g., Areal and Taylor (2002), Fleming et al (2002), Piccinato et al. (1998) and Thomakos and Wang (2003)
- Currencies - see e.g., Andersen et al. (1999) and Pong et al. (2004)
- Debt Instruments – see e.g., Speight and McMillan (2000)

Furthermore a study could be undertaken investigating the following:

- The impact of futures trading on the equity market (see e.g., Antoniou et al. (1998))
- A high frequency futures study on the South African agricultural market could be undertaken.
- Combining high frequency data on derivatives trading and high frequency data on spot trading, would allow the researcher to research volatility in several dimensions.

REFERENCES

Y. Aït-Sahalia, 2002. Maximum Likelihood Estimation of Discretely Sampled Diffusions: a Closed-Form Approximation Approach. *Econometrica*, 70(1):223-262.

A. Admati and P. Pfeiderer, 1998. A Theory of Intraday Patterns: Volume and Price Variability. *The Review Of Financial Studies*, 1:3-40.

T.G. Andersen and T. Bollerslev, 1998. Answering the Skeptics: Yes Standard Volatility Models do Provide Accurate Forecasts. *International Economic Review*, 39(4):885-904.

T.G. Andersen, T. Bollerslev, F.X. Diebold and P. Labys, 1999. The Distribution of Exchange Rate Volatility. NBER Working Paper, 6961.

T.G. Andersen, T. Bollerslev, F.X. Diebold and P. Labys, 2000. Modeling and Forecasting Realized Volatility. NBER Working Paper, 8160.

T.G. Andersen, T. Bollerslev, F.X. Diebold and H. Ebens, 2001. The Distribution of Realized Stock Return Volatility. *Journal of Financial Economics*, 61:43-76.

T.G. Andersen, T. Bollerslev and F.X. Diebold, 2003. Some Like It Smooth, and Some Like it Rough: Untangling Continuous Jump Components in Measuring, Modeling, and Forecasting Asset Return Volatility. Available online from http://www.eco.fundp.ac.be/cerefim/realvol/home.html (Accessed: 24 Sep 2005).

E. Andreou and E. Ghysels, 2002. Rolling sample Volatility estimators: Some New Theoretical, Simulation and Empirical Results. *Journal of Business & Economic Statistics*, 20(3): 363-376.

A. Antoniou, P. Holmes and R. Priestly, 1998. The Effects of Stock Index Futures Trading on Stock Index Volatility: An Analysis of the Asymmetric Response of Volatility to News. *The Journal of Futures Markets*, 18:151-166.

N.M.P.C. Areal and S.J. Taylor, 2002. The Realized Volatility of FTSE 100 Futures Prices. *The Journal of Futures Markets*, 22:636-648.

O.E. Barndorff-Nielsen, P. Blaesid, J.L. Jensen and M. Sorensen, 1985. The Fascination of Sand, in A.C. Atkinson and S.E. Feinberg, eds. *A Celebration of Statistics*, Springer-Verlag, New York, 57-87.

O.E. Barndorff-Nielsen and K. Prause, 2001. Apparent Scaling. *Finance and Stochastics*, 5: 103-113.

O.E. Barndorff-Nielsen and N. Shephard, 2001. Normal Modified Stable Processes. Unpublished manuscript available from the Centre for Mathematical Physics and Stochastics, University of Aarhus, Denmark.

O.E. Barndorff-Nielsen and N. Shephard, 2002. Econometric Analysis of Realized Volatility and its use in Estimating Stochastic Volatility Models. *Journal of the Royal Statistical Society*, B64: 253-280.

F.M. Bandi and J.R. Russell, 2003a. Microstructure Noise, Realized Volatility and Optimal Sampling. Paper written for the CIRANO conference "Realized Volatility", Montreal, November 7-8, 2003. Available online from http://www.eco.fundp.ac.be/cerefim/realvol/home.html (Accessed: 24 Sep 2005).

F.M. Bandi and J.R. Russell, 2003b. Volatility or Microstructure Noise? Available online from http://www.eco.fundp.ac.be/cerefim/realvol/home.html (Accessed: 24 Sep 2005).

J. Berkowitz and J. O'Brien, 2002. How Accurate are Value-at-Risk Models at Commercial Banks? *The Journal of Finance*, LVII, 1093-1111.

T.D. Berry and K.M. Howe, 1994. Public Information Arrival. *Journal of Finance*, 49: 1331-1346

B.J. Blair, S. Poon and S.J. Taylor, 2001. Forecasting S&P 100 Volatility: the Incremental Information Content of Implied Volatilities and High Frequency Index Returns. *Journal of Econometrics*, 105:5-26.

B. Bollen and B. Inder, 2002. Estimating Daily Volatility in Financial Markets Utilizing Intraday Data. *Journal of Empirical Finance*, 9:551-562.

T. Bollerslev, 1986. Generalised Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31:307-27.

T. Bollerslev and J.H. Wright, 2001. High Frequency Data, Frequency Domain Inference, and Volatility Forecasting. *The Review of Economics and Statistics*, 83(4):596-602.

T. Bollerslev and J.M. Wooldridge, 1992. Quasi-maximum Likelihood Estimation of Dynamic Models with Time Varying Covariances. *Econometric Reviews*, 11: 143-172.

T. Bollerslev, R.Y. Chou, and K.F. Kroner, 1992. ARCH Modeling in Finance. *Journal of Econometrics*, 52:5-59.

D.D. Cho, J. Russell, G.C. Tiao and R. Tsay, 2003. The Magnet Effect of Price Limits: Evidence from High Frequency Data on Taiwan Stock Exchange. *Journal of Empirical Finance*, 10:133-168.

D.M. Cutler, J.M. Poterba and L.H. Summers, 1989. What Moves Stock Prices ? *Journal. Of Portfolio Management*, 15: 4-12.

W.J. Den Haan and A. Lavin, 1996. Inferences from Parametric and Non-Parametric Covariance Matrix rstimation Procedures. Working Paper, NBER, 195.

D. Djupsjöbacka, 2002. Properties of Realized Volatility and Correlation. Swedish School of Economics and Business Administration. Available online from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=393469 (Accessed: 24 Sep 2005).

D. Easley, N.M. Kiefer and M. O'Hara, 1995. The Information Content of the Trading Process. *Journal of Empirical Finance*, 4 159-186.

H. Ebens, 1999. Realized Stock Index Volatility. Unpublished working paper #420. Department of Economics, John Hopkins University. Available online from www.econ.jhu.edu/pdf/papers/wp430ebens.pdf (Accessed: 24 Sep 2005).

L. Ederington and Jah Ha Lee, 2001. Intraday Volatility in Interest Rate and Foreign Exchange Markets: ARCH, Announcement and Seasonality Effects. *The Journal of Futures Markets*, 21(6):517-522.

R.F. Engle, 1982. Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50, 987-1008.

R.F. Engle and G. Gonzalez-Rivera, 1991. Semi-parametric ARCH Models, *Journal of Business and Economic Statistics*, 9:345-359.

R.F. Engle, 2000. The Econometrics of Ultra-High-Frequency Data. *Econometrica*, 68(1):1-22.

R.F. Engle, 2003. Risk and Volatility: Econometric Models and Financial Practise. Nobel lecture, December 8, 2003.

R.F. Engle and Z. Sun, 2004. Forecasting Volatility Using Tick by Tick Data. Available online from http://www.eco.fundp.ac.be/cerefim/realvol/home.html (Accessed: 24 Sep 2005).

T.N. Falkenberry, 2002. High Frequency Data Filtering: A Review of the Issues Associated with Maintaining and Cleaning a High Frequency Financial Database. Available online from www.tickdata.com (Accessed: 24 Sep 2005).

E. Fama, 1965. The Behavior of Stock Market Prices. *Journal of Business*, 38:34-105.

E.F. Fama, 1989. Perspectives on October 1987, or What did we learn from the crash in *Black Monday and the Future of Financial Markets*, edited by R.W. Kamphuis, Jr., R.C. Kormendi, and J.W.H. Watson (Homewood: Dow-Jones-Irwin, Inc).

J. Fleming, C. Kirby and B. Ostdiek, 2002. The Economic Value of Volatility Timing Using Realized Volatility. *Journal of Financial Economics*, article in press – received 20 November 2002, accepted 1 February 2002.

K.R. French, G.W. Schwert and R.F. Stambaugh, 1987. Expected Stock Returns and Volatility. *Journal of Financial Economics*, 19(1): 3-29.

M. Garman and M. Klass, 1980. On the Estimation of Security price Volatilities from Historical Data. *Jourbal of Business*, 53: 67-78.

E. Ghysels, P. Santa-Clare and R. Valkanov, 2003. Predicting Volatility: Getting the Most out of Return Data Sampled at Different Frequencies. Available online from http://www.eco.fundp.ac.be/cerefim/realvol/home.html (Accessed: 24 Sep 2005).

P. Giot and S. Laurent, 2004. Modelling Daily Value at Risk Using Realized Volatility and ARCH Type Models. *Journal of Empirical Finance*, 11 (2004):379-398.

C.A.E. Goodhart and M. O'Hara, 1997. High Frequency Data in Financial Markets: Issues and Applications. *Journal of Empirical Finance*, 4:73-114.

C. Gourieroux, 1997. *ARCH Models and Financial Applications*. Springer, New York.

P.R. Hansen and P.R. Lunde, 2004a. A Forecast Comparison of Volatility Models : Does Anything Beat a GARCH(1,1)?. Brown University working paper. Available online from http://www.stanford.edu/~prhansen/papers.html (Accessed: 24 Sep 2005).

P.R. Hansen and P.R. Lunde, 2004b. An Unbiased Realized Variance. Brown University working paper. Available online from http://www.stanford.edu/~prhansen/papers.html (Accessed: 24 Sep 2005).

J. Hasbrouck and T. Ho, 1987. Order Arrival, Quote Behaviour and the Return-Generating Process. *Journal of Finance*. 42: 1035-1048.

E. Hol and S.J. Koopman, 2002. Stock Index Volatility Forecasting with High Frequency Data. 2002. Available online from http://ideas.repec.org/e/pko46.html (Accessed: 24 Sep 2005).

J. Hull and A. White, 1987. *The Pricing of Options on Assets with Stochastic Volatility*, Journal of Finance, 42: 281-300.

J.C. Hull, 1989. *Options, Futures & Other Derivatives (fourth edition)*. Prentice- Hall, Inc.

C.M. Jones, G. Kaul and M.L. Lipson, 1994. Transactions, Volume and Volatility. *Review of Financial Studies*, 7: 631-651.

JSE Securities Exchange SA (JSE) Equities Markets and Trading Functionality Overview. Available online from www.jse.co.za (Accessed: 24 Sep 2005).

C.M.C. Lee, M.J. Ready and P.J. Sequin, 1994. Volume, Volatility and New York Stock exchange Trading Halts. Journal of Finance, 49: 183-214.

P. Lequeux, 1997. On the Impact of High Frequency Data in the Financial Markets. Quantitative Research and Trading Desk, Banque National de Paris P.L.C.

J. Lillestol, 2000. Risk Analysis and the NIG-distribution. *The Journal of Risk*, 2(4): 41-56.

A. Malz and J. Mina, 2001-2002. Financial Markets in the Aftermath of the Terrorist Attack. *RiskMetrics Journal* , 2 (2):49-64.

B.B. Mandelbrot, 1963. The Variation of Certain Speculative Prices. *Journal of Business,* 36:394-419.

Mineweb : Available online from http://www.mineweb.net (Accessed: 24 Sep 2005).

M.L. Mitchell and J.H. Mulherin, 1994. The Impact of Public Information on the Stock Market. *Journal of Finance*, 49: 923-950.

D.B. Nelson, 1992. Filtering and Forecasting with Misspecified ARCH Models 1: Getting the Right Variance with the Wrong Model. *Journal of Econometrics*, 52:61-90.

D.B. Nelson and D.P. Foster, 1995. Filtering and Forecasting with Misspecified ARCH Models 2: Getting the Right Variance with the Wrong Model. *Journal of Econometrics*, 67:303-335.

R. Olsen, 1999. Real-Time Trading Models and Statistical Properties of Foreign Exchange Rates. Available online from http://atlas-conferences.com/c/a/c/q/03.htm (Accessed: 24 Sep 2005).

R.C.A. Oomen, 2001. Using High Frequency Stock Market Index Data to Calculate, Model and Forecast Realized Volatility. European University Institute, Department of Economics Via dei Roccetinni. Available online from www.panagora.com/research/2001crowell/2001cp_30.pdf (Accessed: 24 Sep 2005).

I.M. Pagel, (marizap@nedbank.co.za) 2004. Estimating Daily Volatility in Financial Markets Utilizing Intraday Data. E-mail correspondence to: Dr B. Bollen (Bernie.Bollen@BusEco.monash.edu.au) on 29 June 2004.

M. Parkinson, 1980. The Extreme Value Method for Estimating the Variance of the Rate of Return. *Journal of Business*, 53: 61-65.

E. Parzen, 1983. Autoregressive Spectral Estimation, *Handbook of Statistics* 3, D.R. Brillinger and P.R. Krisnaiah, eds., Elsevier Press, 221-247.

B. Piccinato, G. Ballocchi and M.M. Dacorogna, 1998. A Closer Look at the Eurofutures Market: Intraday Statistical Analysis. O&A Research Group Preprint. Available online from www.olsen.ch/research/workingpapers/ (Accessed: 24 Sep 2005).

S.E. Pong, M.B. Shackleton, S.J. Taylor and X. Xu, 2004. Forecasting Currency Volatility: a Comparison of Implied Volatilities and AR(FI)MA Models. Lancaster University Working Paper, EFA 2002 Berlin Meeting presented paper. Available online from http://papers.ssrn.com (Accessed: 24 Sep 2005).

S. Rahman, and Kian Ping Ang, 2000. Intraday Return Volatility Process: Evidence from NASDAQ Stocks. Available online from www.ntu.edu.sg/nbs/crefs/working_papers/wp2000-11.pdf (Accessed: 24 Sep 2005).

R. Roll, 1984. A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market. *Journal of Finance*, 39: 1127-1139

G.W. Schwert, 1990a. Stock Market Volatility. *Financial Analysts Journal*, 46: 23-34.

G.W. Schwert, 1990b. Stock Volatility and the crash of '87. *Review of Financial Studies*, 3: 77-102.

G.W. Schwert and P.J. Sequin, 1991. Heteroskedasticity in Stock Returns. *Journal of Finance*, 45: 1129-1155.

A.E.H. Speight and D.G. McMillan, 2000. Nonlinear Dynamics in High Frequency Intraday Financial Data: Evidence from the UK Long Gilt Futures Market. *The Journal of Futures Markets*, 22(11):1037-1058.

A.E.H. Speight, D.G. McMillan and O.A. Gwilym, 2000. Intraday Volatility Components in FTSE 100 Stock Index Futures. *The Journal of Futures Markets*, 20(5):425-444.

D.D. Thomakos and T. Wang, 2003. Realized Volatility in the Futures Markets. *Journal of Empirical Finance*, 10(3):321-353.

J.H. Venter and P.J. de Jongh, 2002. Risk Estimation using the Normal Inverse Gaussian Distribution. *The Journal of Risk*, 4(2):1-23.

J.H. Venter and P.J. de Jongh, 2004. Selecting an Innovation Distribution for GARCH Models to Improve Efficiency of Risk and Volatility Estimation. *The Journal of Risk*, 6(3):27-52.

J.H. Venter, P.J. de Jongh and G.Griebenow, 2005. NIG-GARCH Models Based on Open, Close, High and Low Prices. *South African Statistical Journal*, 39(2):79-102.

L. Zhang, P.A. Mykland and Y. Aït-Sahalia, 2003. A Tale of Two Scales: Determining Integrated Volatility with Noisy High Frequency Data. Available online from http://www.eco.fundp.ac.be/cerefim/realvol/home.html (Accessed: 24 Sep 2005)

B. Zhou, 1996. High Frequency Data and Volatility in Foreign Exchange Rates. *Journal of Business & Economic Statistics*, 14(1): 45-52.