Analysis of Financial Time Series via Estimating Functions

BY

Melody Ghahramani

A Thesis
Submitted to the Faculty of Graduate Studies
In Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Faculty of Graduate Studies

University of Manitoba

Winnipeg, Manitoba

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION

Analysis of Financial Time Series via Estimating Functions

BY

Melody Ghahramani

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirement of the degree

DOCTOR OF PHILOSOPHY

Dedication

To my parents and my brother

Abstract

This thesis studies moment properties, applications of combining estimating functions, improved estimation in the presence of structural change, and forecasting, as well as hypothesis testing of a new class of Generalized Autoregressive Conditional Heteroscedastic (GARCH) and Random Coefficient Autoregressive (RCA) models for financial time series data. The autocorrelation structure of the squared process is derived.

The kurtosis for various classes of GARCH models is also derived. Using the estimating function method, the asymptotic correlation between the sample mean and sample median is used to identify the marginal distribution of the error term in the class of GARCH models. The methodology has also been illustrated in simulation studies and with real data examples.

Keywords: Estimating functions, Financial data, Forecasting, Hypothesis testing, Kurtosis, Model identification, Structural change, Volatility,

Acknowledgements

I am deeply indebted to my Supervisor, Dr. A. Thavaneswaran for his patience, guidance and financial support for living and conference funding during my doctoral studies. I would also like to thank Dr. Thavaneswaran for being so generous with his time, for his infectious enthusiasm for statistical research and for introducing me to an interesting area in statistics. I feel very fortunate to have had Dr. Thavaneswaran as my Ph.D. Supervisor.

I would also like to thank my Advisory Committee members, Dr. Ken Mount and Dr. M. Samanta for their helpful suggestions and their useful discussions on all matters statistical during my studies at the University of Manitoba. My thanks go also to the external examiners, Dr. Ruppa Thulasiram (internal external examiner) of the Computer Science department and my External Examiner, Dr. Bovas Abraham of the University of Waterloo for careful reading of the dissertation.

Many thanks go to the Department of Statistics for awarding me a Faculty of Science Studentship for the period of 2003-2005 and for giving me the opportunity to teach an introductory course. Thanks also go to the departmental support staff, Ms. Margaret Smith and Ms. Stana Drobko.

Thanks also to Dr. S.M. Moghadas of the Institute of Biodiagnostics at the National Research Council for providing me with financial support and statistical experience during the Summer of 2006.

Coffee and enjoyable conversation with Heather Ashton of Winnipeg provided a necessary respite from academic life. Thank you Heather for being a good friend.

Last but not least, I would like to thank my parents and brother for their constant encouragement and unending support. This thesis is dedicated to my father who taught me to see the joy in discovering patterns at an early age, my mother who never stopped believing in me and my brother whose focus and ambition inspired me.

Papers

The following papers have been published from this thesis.

1. Ghahramani M., Thavaneswaran A. (2007). Identification of ARMA Models with GARCH Errors. The Mathematical Scientist, 32(1), 1-10.

2. Ghahramani M., Thavaneswaran A. (2006). Financial Applications of ARMA Models with GARCH Errors. Journal of Risk Finance 7(5), 525-543.

3. Ghahramani M., Thavaneswaran A. (2006). Improved Estimation of Volatility with some Applications, Journal of Statistical Theory and Applications 5(3), 260-270.

4. Ghahramani M., Thavaneswaran A. (2005). Moment properties of Hidden Semi-martingale(HSM) models with GARCH errors. InterStat, August 2005.

5. Thavaneswaran A., Ghahramani M. (2005). Forecasting with structural change. InterStat, March 2005.

6. Appadoo S.S., Ghahramani M., Thavaneswaran A. (2005). Moment properties of some time series models. The Mathematical Scientist 30, 5663.

7. Thavaneswaran A., Ghahramani M. (2004). Applications of combining estimating functions. In Proceedings of the International Sri Lankan Conference:

Visions of Futuristic Methodologies, University of Peradeniya and Royal Melbourne Institute of Technology (RMIT), pp. 515-532.

List of Tables

List of Figures

Contents

Chapter 1

Introduction

1.1 Motivation

A lot of research in finance is concerned with measuring and managing financial risk. Portfolio optimization requires maximizing rewards and minimizing risks, the latter of which are not rewarded equally. The topic of risk management has a long history in economics. Markowitz (1952) and Tobin (1958) associated risk with the variance in the value of a portfolio. Sharpe (1964) developed the theory of Capital Asset Pricing Model (CAPM) which shows that there is a relationship between expected returns and variance where returns are computed as the logarithm of the price today divided by the price yesterday. These contributions were recognized by Nobel Prize committees in 1981 and 1990, respectively.

Black and Scholes (1972) and Merton (1973) developed a model to evaluate the pricing of options. Call options give the owner the right but not the obligation to purchase an asset at a particular price at a future date. By purchasing such call options, the risk of a portfolio can be minimized as the call option offers protection against loss. The cost of this insurance depends upon the risks and these risks are quantified by the variance of the asset returns.

1.1.1 Volatility Modeling

Estimates of variance are required to implement the option pricing formulas. Typically, the square root of the variance, known as *volatility* is reported. An estimate of volatility widely in use today is *historical volatility.* This estimate is calculated as the sample standard deviation of the returns over a range of time periods. There is no guidance however, on how to choose the length of time period optimally. Furthermore, it is not reasonable to assume the volatility calculated based on two different yearly intervals should be equal. From an economics point of view, the time-varying nature of volatility arises out of changes in economic conditions, reaction to unexpected news announcements from government agencies, as well as reaction to unexpected world events.

The class of Autoregressive Conditional Heteroscedastic (ARCH) models proposed by Engle (1982) provides a class of models for which the time-varying nature of volatility can be adequately modeled. The conditional heteroscedasticity of the variance of return series can be seen in the sample autocorrelation function (ACF) plot of the square of a process. While the return series is an uncorrelated process, its square is a correlated sequence.

Analysts had also observed that the sample kurtosis of return series is far greater than the kurtosis implied by a normal distribution. The class of ARCH models is able to capture the leptokurtic nature of return series.

Heuristically, ARCH models are a generalization of the sample variance in that they take weighted averages of past squared observations. Bollerslev (1986) proposed a moving average type generalization of ARCH models to include a weighted average of past (unobserved) volatilities.

Further generalizations of ARCH models have been proposed by many researchers. These generalizations were surveyed in Bollerslev, Chou and Kroner (1992), Bollerslev (1994), Engle (2002) and Engle & Ishida (2002). These extensions recognized that there may be non-linearity, asymmetry and long-memory present in return series and that returns can be conditionally non-normal with numerous parametric and non-parametric distributions. A non-exhaustive list of these extensions includes: AARCH (Bera and Lee (1990)), APARCH (Ding et al.(1993)), FIGARCH (Bollerslev and Mikkelsen (1996)), FIEGARCH (Baillie et al.(1996)), STARCH (Lee and Degennaro (2000)), SWARCH (Fornari and Mele (1997)), GJR-GARCH (Glosten et al.(1993)), MARCH (Bollerslev (1987)), QGARCH (Sentana, 1995), NARCH (Higgins and Bera, (1992)), Component ARCH (Ding and Granger (1996)), Asymmetric Component ARCH (Ding and Granger (1996)), Taylor-Schwert (Schwert (1990)), TGARCH (Zakoian (1994)), Student-t-GARCH (Bollerslev (1986)), Generalized Exponential Distribution GARCH (Nelson (1992)).

1.1.2 An Example

In this Section, an illustration of some statistical properties of GARCH models are given through analysis of Standard and Poor's 500 Composite Index of daily price levels from January 2, 1963 through November 23, 2005. The index represents the bulk of the U.S. equity market and provides a proxy for the U.S. financial history for this period. All statistics and plots have been generated using the *finmetrics* module of S-Plus© 6.2.

In examining the price period before 1987 and from 1987 onwards in Figure 1.1

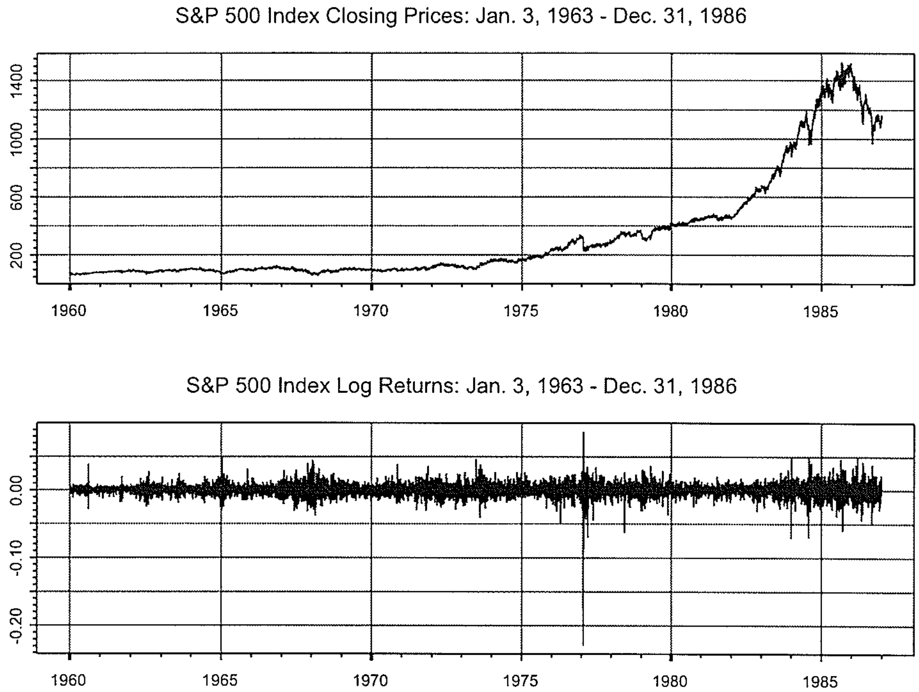

Figure 1.1: S&P 500 Index of Price and Returns from 1963 - 1986.

and Figure 1.2, the great growth of equity prices over the period and the subsequent decline after January 2000 is observed. Economists focus attention on returns as it is the relative price from the purchase point to the sale point that matters. The return series is centered around zero throughout the sample period even though prices are sometimes increasing and sometimes decreasing. The largest negative return corresponds to the crash of October 1987 (see Figure 1.2). The amplitude of the returns is changing. The magnitude of the changes is sometimes large and sometimes small. This is the effect that economists call *volatility clustering.* There is however another interesting feature in this graph. It is clear that the volatility is higher when prices are falling. This is the asymmetric volatility effect that Nelson (1992) described

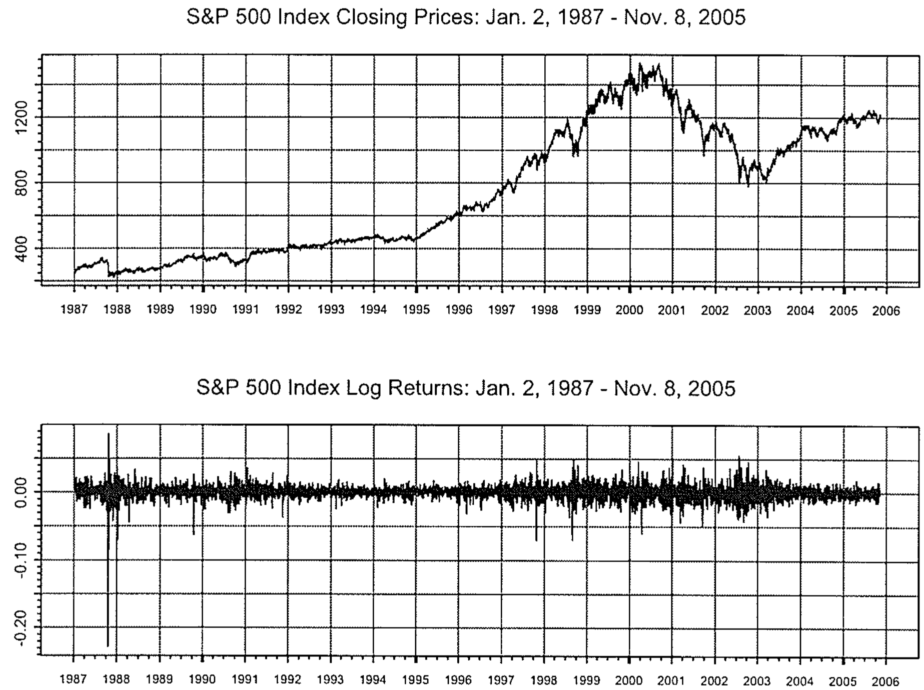

Figure 1.2: S&P 500 Index of Price and Returns from 1987 - 2005.

with his Exponential GARCH (EGARCH) model.

Looking at the next sub-period after the 1987 crash in Figure 1.2, the record low volatility period of the middle 1990's is observed. This was accompanied by a slow and steady growth of equity prices. The volatility began to rise reaching very high levels from 1998 onwards. Looking at the last period since 1998 in Figure 1.2, the high volatility continues as the market turned down. Only at the end of the sample, since the official conclusion of the Iraq war does volatility decline.

Sample moments of this data are presented in Table 1.1. The mean is close to zero relative to the standard deviation for both periods. It is .03% per trading day or about 7.8% per year. The standard deviation is higher in the 1990's.

Statistic	Full	Since 1990
Mean	0.0003	0.0003
Standard Deviation	0.0094	0.107
Skewness	-1.39	-0.101
Kurtosis	40.012	6.75

Table 1.1: Moments of S&P 500 Returns.

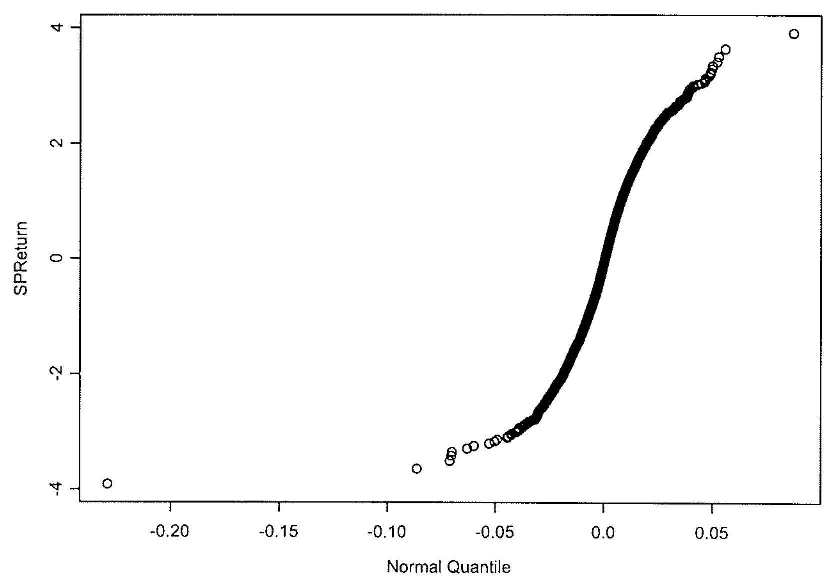

Figure 1.3: Quantile plot of S&P 500 Returns from 1990 - 2005.

The skewness is small throughout. The kurtosis; a measure of the magnitude of the extreme observations is substantial at 6.8 during the nineties, while for the full sample it is 40. This is strong evidence that extremes are much larger than would be expected from a normal random variable. Similar evidence is seen in Figure 1.3, which is a quantile-quantile plot for the post 1990 data. The autocorrelations of the return series are non-significant as shown in Figure 1.4 while squared returns(and absolute returns; not shown here) are significant.

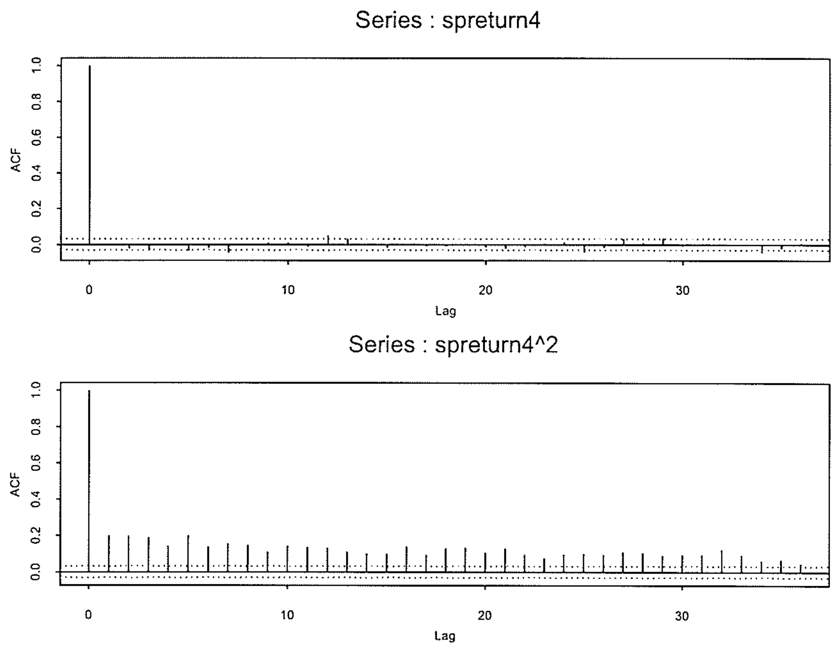

Figure 1.4: Sample ACF of S&P 500 Returns and the Squared Process Returns from 1990 - 2005.

Parameter	Estimate	Std.Error	t value	Pr(t > $\|t\|$)
ω	5.3×10^{-7}	0.0015	4.627	< 0.0001
α_1	0.0057	0.0045	12.678	< 0.0001
β_1	0.9381	0.0049	192.064	< 0.0001

Table 1.2: Parameter Estimates of GARCH(1, 1) model fitted to S&P 500 Returns (from 1990 - 2005)

The GARCH (1, 1) was fitted to the S&P 500 data using maximum likelihood estimation. The model and estimation will be more thoroughly discussed in the following chapter. The parameter estimates are reported in Table 1.2. The bulk of the information comes from the previous day forecast (see the estimate of β_1). The new information (which corresponds to the estimate of α_1) changes this a little and the long run average variance has a very small effect (see the estimate of ω). The

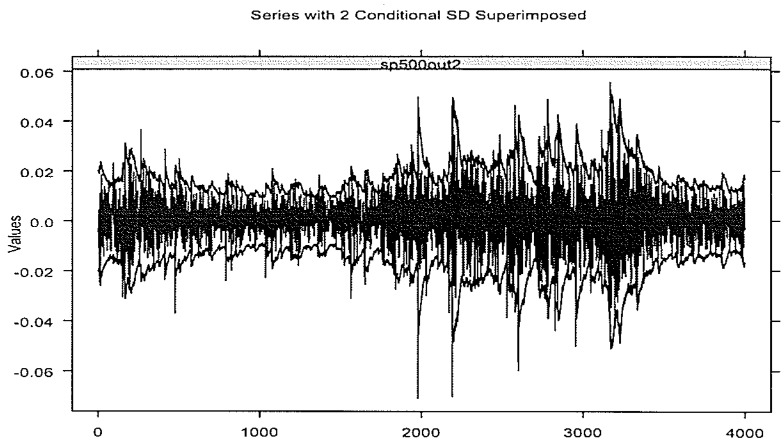

Figure 1.5: S&P 500 Index of returns with estimates of 2 standard deviations fitted using AR(1) model with GARCH(1, 1) errors.

long run variance is naturally small because of the use of daily data.

To address the question of volatility forecasting for this series, suppose an analyst proceeds naively by assuming an autoregressive model of order one (AR(1)) for the return series. That is, let $\{y_t\}$ denote the observed return series and assume

$$y_t = \phi y_{t-1} + e_t, \quad e_t \sim NID(0, \sigma_e^2).$$

Under this model, conditional on past values, we usually assume that the data are normally distributed with conditional variance $Var(y_t|y_{t-1}) = \sigma_e^2$, where σ_e^2 is the variance of the error terms e_t. Under this model, the estimate of unconditional variance is found by substituting estimated model parameters into $Var(y_t) = \dfrac{\sigma_e^2}{1-\phi^2}$. The 'prediction' interval in this case has constant width.

However, an AR(1) model with GARCH (1, 1) errors results in a prediction interval that is changing in a more believable fashion as seen in Figure 1.5. The difference between these two prediction intervals has important applications in finance.

Consider the following simple example due to Gouriéroux (1997) that illustrates the importance of moding time-varying volatilit for portfolio theory. Suppose an investor has a portfolio consisting of exactly two assets: a risk-free asset with constant return r_f and a risky return whose future return is predicted via a prediction interval, $(r_{l,t}, r_{u,t})$, where t denotes time. Assume that at the initial date, the whole portfolio is invested in the risk-free asset. The investor wishes to reallocate his portfolio in the following manner. Because of transaction costs, he will only change his portfolio when $r_{l,t} > r_f$. Since this means that the risky asset is more profitable, he will change his entire portfolio to be reinvested in the risky asset. The portfolio will remain unchanged until $r_{u,t} < r_f$, at which time, the entire portfolio will be reinvested in the risk-free asset, and so on. Clearly, the frequency of the reallocations depends on the evolution of the lower and upper bounds of the prediction interval. But the predictions from the model with GARCH errors are more variable. In practice, fitting an AR(1) model with GARCH errors to the returns would imply more frequent tradings in order to make profits by exploiting the volatility variations. Another widely used application of volatility modeling is illustrated in the next Subsection.

1.1.3 Options Pricing

One of the important applications of volatility modeling in finance is in the area of pricing call options. A call option gives one the right, but not the obligation to buy a specified number of shares of a certain asset such as a stock at the *exercise* or *strike* price. Option contracts are generally for 100 shares; see Jarrow and Turnbull(2000). An option has an *exercise date*, which is also called the *strike date*, *maturity*, or *expiration date*. American options may be exercised at any time up to their exercise date, while European options can only be exercised at their exercise date. In this thesis, an application of volatility modeling for the pricing of European call options is considered. From hereon, a call option refers to a European call option.

To clarify the idea behind call options, consider the following example found on page 258 of Ruppert(2004). Suppose that an investor has purchased a call option of 100 shares of Stock A with an exercise price of \$70. At the expiration date, suppose that stock A is selling at \$73. The option allows the investor to purchase 100 shares for \$70 and to immediately sell them for \$73, which leaves a net gain of \$300 on the 100 shares. However, the net profit is not exactly \$300 since the investor had to pay a premium for the option. For example, if the option cost \$2 per share, then the premium cost \$200. Moreover, the investor had to pay \$200 upfront only to receive \$300 at the expiration date. Suppose that the expiration date was 3 months after the purchase date and that the continuously compounded risk-free rate is 6% per annum or 1.5% for 3 months. Then the dollar value of the investor's net profit at the time of purchase is

$$\exp(-0.015)300 - 200 = 95.53$$

and is

$$300 - \exp(-0.015)200 = 96.98$$

at the exercise date.

A call is seldom exercised if the exercise price is greater than the price of the stock, since exercising the option corresponds to buying the stock for more than it would cost on the market. If a call is not exercised, then the investor only looses the cost of the premium.

The investor could lose money on an option even if it is exercised, because the amount gained by exercising the option might be less than the premium. In the example above, if stock A were selling for \$71 at the exercise date, then the net gain at the exercise date would be \$100; which is less than the \$200 paid for the option. Even though exercising the option results in a loss in this case, the loss is less than it would be had the option not been exercised. An option should always be exercised if the stock's price on the exercise date exceeds the strike(exercise) price.

The Black-Scholes formula for pricing European call options requires an estimate of volatility which it assumes is a constant. The classic Black-Scholes formula for a European call option gives the cost of an option based on the Geometric Brownian motion model for the log return of stock with price $S(t)$ at time t as

$$c(S,t) = S\Phi(d) - Ke^{-r(T-t)}\Phi(d - \sigma\sqrt{T-t}),$$

where

$$d = \frac{\log[S/K] + r(T-t)}{\sigma\sqrt{T-t}} + \frac{\sigma\sqrt{T-t}}{2},$$

$S(t) = S$ is the price of the underlying security at time t usually $t = 0$, K is the strike price, T is the maturity date, r is the spot rate, σ is the volatility (i.e. the instantaneous standard deviation of the rate of return of the underlying security), and $\Phi(x)$ is the distribution function of the standard normal variate. The option price depends on the initial stock price S, strike price K, maturity date T, spot rate r and the the 'unknown' volatility parameter σ and hence the better estimate/forecast of the volatility will improve the performance of the option pricing formula. The Black-Scholes formula assumes that both the spot rate r and the volatility σ are constant. Historical volatility has been used to estimate the volatility. In the next Section, we provide an outline of the Thesis.

1.2 Thesis Organization

In Chapter 2, we begin by revisiting the class of ARCH models first introduced by Engle (1982). Extensions of ARCH models that allow for a time-varying conditional mean to be specified are proposed and their moment properties are studied. The class of Random Coefficient Autoregressive (RCA) processes due to Nicholls and Quinn (1982) parallels the time-varying conditional mean and variance structure of ARMA models with GARCH errors. We introduce generalizations of the class of RCA models and study their moment properties. Heinen (2003) introduced the class of Autoregressive Conditional Poisson (ACP) models that account for discreteness, autocorrelation and overdispersion of count data such as the daily number of price change durations of \$0.75 on IBM stock. A \$0.75 price- change duration is defined as the time it takes the stock price to move by at least \$0.75. The variable of interest is the daily number of such durations, which is a measure of intradaily volatility. The class of ACP models is shown to possess properties that parallel those of the class of GARCH models. We extend the moment property results of the ACP(1,1) model due to Heinen(2003) to the class of ACP(p,q) models, where $p > 1$ and $q > 1$. An extension of Hamilton's (1989) Markov-switching model is proposed and its properties are studied. In Chapter 2, we also study a new class of nonlinear GARCH models developed for the purpose of modeling conditional leverage and conditional skewness. The Heston and Nandi (2000) GARCH models model the presence of conditional leverage; whereas the Inverse Gaussian GARCH model of Christoffersen, Heston and Jacobs (2006) models conditional skewness in addition to the presence of conditional leverage. Model identification for nonlinear GARCH models remains an open problem. We develop data-driven procedures for identifying this recently developed class of nonlinear GARCH models. Simulation studies are extensively used in Chapter 2 in order to illustrate to show that moment properties are useful for the puropose of volatility model identification.

In Chapter 3, we revisit the theorem on optimal estimating functions for stochastic processes due to Godambe (1985). We also revisit the theorem on optimal estimating functions due to Thompson and Thavaneswaran (1999). We show that both theorems lead to the same optimal estimating function. Godambe's theorem is used to combine estimating functions for volatility. The combined estimating functions are shown to have more information than each of the component estimating functions. The combination theorem due to Thompson and Thavaneswaran (1999) is used to demonstrate the usefulness of the correlation between the least squares estimating function and least absolute deviation estimating function as a GARCH model identification tool. Applications of combining estimating functions in hypothesis testing problems for volatility models were also studied.

In Chapter 4, the forecasting problem for stationary processes (such as ARMA models) with GARCH errors is studied. The forecasting problem for stationary series in the presence of structural change such as change in the mean or change in the volatility is also studied. The proposed forcasts which we term *improved estimates*; were motivated from the biased minimum mean squared error estimates proposed by Shalabh (2001). In practice, the biased minimum mean squared error forecasts are estimable if all nuisance parameters are known. We derive recursive improved estimates in the presence of structural change that only require starting values for nuisance parameters.

In Chapter 5, some of the financial applications of the theoretical results obtained in the previous chapters. The correlation between the LS and LAD estimating functions as a GARCH model identification tool is extensively studied in simulation studies and illustrated using real financial data. The superiority of the optimal estimating function approach for hypothesis testing is also demonstrated through some simulation studies. European call options prices are calculated using historical volatility and compared to to predicted prices obtained by assuming an adequate GARCH model for the volatility. Interval estimation of the marginal standard deviation of GARCH models using nonparametric and parametric bootstrapping is studied in a preliminary simulation study.

We end with possible extensions for future work. Appendix A contains algorithms for simulating random variables from an Inverse Gaussian distribution. Appendix B contains details of the proofs for combining estimating functions for volatility discussed in Chapter 3.

Chapter 2

Moment properties of some time series models

In the introductory chapter, we have introduced several GARCH models for modeling financial series. Other time series models are also able to capture salient features of financial series such as time varying conditional variance and leptokurtosis. For a simple model considered by Gouriéroux (1997) of the form

$$y_t = \varepsilon_{t-1}^2 \varepsilon_t,$$

where ε_t is a Gaussian white noise process with variance σ_ε^2, we obtain

$$\begin{aligned}
\mathrm{E}(y_t) &= 0 \\
\mathrm{E}(y_t^2) &= \mathrm{E}(\varepsilon_{t-1}^4)\mathrm{E}(\varepsilon_t^2) = 3\sigma_\varepsilon^6 \\
\mathrm{E}(y_t y_{t-h}) &= \mathrm{E}(\varepsilon_{t-1}^2 \varepsilon_t \varepsilon_{t-h-1}^2 \varepsilon_{t-h}) = \mathrm{E}(\varepsilon_t)\mathrm{E}(\varepsilon_{t-1}^2 \varepsilon_{t-h-1}^2 \varepsilon_{t-h}) = 0, \quad (h > 0)
\end{aligned}$$

with $\rho_k^y = 0$ for $k > 0$. We see that the process y_t is weakly stationary having variance $3\sigma_\varepsilon^6$, where as its conditional variance given its past, $\mathrm{V}(y_t|\mathcal{F}_{t-1}^y) = \mathrm{V}(\varepsilon_{t-1}^2\varepsilon_t|\mathcal{F}_{t-1}^y) = \sigma^2\varepsilon_{t-1}^4$, depends on the lagged residuals. Since,

$$\mathrm{E}(\varepsilon_t^{2n}) = \frac{(2n)!}{2^n(n!)}\sigma_\varepsilon^{2n} \text{ and } \mathrm{E}(y_t^4) = \mathrm{E}(\varepsilon_{t-1}^8)\mathrm{E}(\varepsilon_t^4) = 315\sigma_\varepsilon^{12},$$

the kurtosis is given by $K^{(y)} = \dfrac{E(y_t^4)}{[Var(y_t)]^2} = \dfrac{315\sigma_\varepsilon^{12}}{9\sigma_\varepsilon^{12}} = 35$. This clearly shows that even a simple volatility model $y_t = \varepsilon_{t-1}^2\varepsilon_t$ with Gaussian error term ε_t can generate very high peakedness which is very common for financial time series. For the y_t^2 process, $y_t^2 = \varepsilon_{t-1}^4\varepsilon_t^2$, the correlation is given by,

$$\rho_k^{y^2} = \begin{cases} 1 & k = 0 \\ 0.1145 & k = 1 \\ 0 & k \geq 2. \end{cases}$$

In contrast, for the process given by $Z_t = \varepsilon_{t-1}\varepsilon_t$, where ε_t is an uncorrelated Gaussian sequence with mean zero and variance σ_ε^2, we have the following:

$$\begin{aligned} \mathrm{E}(Z_t) &= 0, \\ \mathrm{E}(Z_t^2) &= \mathrm{E}(\varepsilon_{t-1}^2)\mathrm{E}(\varepsilon_t^2) = \sigma_\varepsilon^4 \\ \mathrm{E}(Z_t^4) &= \mathrm{E}(\varepsilon_{t-1}^4)\mathrm{E}(\varepsilon_t^4) = 9\sigma_\varepsilon^8, \end{aligned}$$

so that the kurtosis is given by

$$K^{(Z)} = \frac{\mathrm{E}(Z_t - \mu)^4}{(\mathrm{Var}(Z_t))^2} = \frac{9\sigma_\varepsilon^8}{\sigma_\varepsilon^8} = 9.$$

The autocorrelation Z_t^2 of the process is given by

$$\rho_k^{Z^2} = \begin{cases} 1 & k = 0 \\ 0.25 & k = 1 \\ 0 & k \geq 2. \end{cases}$$

Hence, these simple white noise driven models can exhibit leptokurtosis and time-varying conditional variance observed in financial data.

Recently, Thavaneswaran et al. (2005) have studied the moment properties of the zero mean GARCH processes. In this chapter, we extend the results of Thavaneswaran et al. (2005) by studying the moment properties such as the autocorrelation of the square of GARCH processes and the kurtosis of ARMA(p, q) models with GARCH(P, Q) errors. The models studied are useful for volatility modeling in the sense that they are conditionally heteroscedastic and the kurtosis implied by such models is larger than three. Moment properties are useful for model identification. In this chapter, the moment properties of two classes two classes of volatility models are presented. In Section 2.1, moment properties of the class of Random Coefficient Autoregressive(RCA) models due to Nicholls and Quinn (1982) are given. The moment properties of the class of zero-mean ARCH models due to Engle (1982) are studied in Section 2.2. Moment properties of RCA models extended to have GARCH errors are studied in Section 2.3.1 and Section 2.3.2, while moment properties of stationary models with GARCH errors are studied in Section 2.3.3.

Moment properties of Hidden Semimartingale models (HSMs) with GARCH errors are studied in Section 2.4. The class of HSMs with GARCH errors is an extension of Hidden Markov Models with autocorrelated errors due to Hamilton (1989). In Section 2.5, we extend the results on the moment properties of a class of volatility models for time series of counts proposed by Heinen (2003).

Identification of two classes of non-linear GARCH models are explored in Section 2.7.1 and in Section 2.7.3 by using their moment properties. These classes of non-linear GARCH models can incorporate conditional leverage. In particular, the class of Inverse Gaussian GARCH models allows for conditional skewness to be modelled, as well. Our main contribution is to use moment properties for the purpose of model identification.

In the next section, we extend the results on moment properties of RCA models.

2.1 Random coefficient autoregressive models

Random coefficient autoregressive time series were introduced by Nicholls and Quinn (1982) and their moment properties have been studied recently by Aue (2004) and Appadoo, Ghahramani and Thavaneswaran (2005). RCA models exhibiting long memory properties have been considered in Leipus and Sugailis (2003). A sequence of random variables $\{y_t\}$ is an RCA(1) time series if it satisfies the equations

$$y_t = (\phi + b_t)y_{t-1} + e_t \qquad t \in Z,$$

where $\mathbf{Z}$ denotes the set of integers, and

$$\begin{pmatrix} b_t \\ e_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_b^2 & 0 \\ 0 & \sigma_e^2 \end{pmatrix}\right), \tag{2.1.1}$$

$$\phi^2 + \sigma_b^2 < 1. \tag{2.1.2}$$

The sequences $\{b_t\}$ and $\{e_t\}$ respectively, are the errors in the model. According to Nicholls and Quinn (1982), (2.1.2) is a necessary and sufficient condition for the second order stationarity of $\{y_t\}$. (2.1.1) also ensures strict stationarity. Moreover, Feigin and Tweedie (1985) showed that $Ey_t^{2k} < \infty$ for some $k \geq 1$ if the moments of the noise sequences satisfy $Ee_t^{2k} < \infty$ and $E(\phi + b_t)^{2k} < 1$, for the same k.

Let $\{y_t\}$ be a stationary Gaussian linear process with mean zero and variance σ_y^2. Then it can easily be shown that the joint moment generating function of the pair (y_t, y_{t-k}) is given by

$$m(u,v) = E(e^{uy_t + vy_{t-k}}) = exp\left(\frac{1}{2}\sigma_y^2\left(u^2 + v^2 + 2\rho_k^y uv\right)\right),$$

where $\sigma_y^2 = Var(y_t)$ and ρ_k^y is the autocorrelation function of y_t. Since, $E\left[y_t^2 y_{t-k}^2\right] = \sigma_y^4(1+2(\rho_k^y)^2)$ and $Var(y_t^2) = 2\sigma_y^4$ we have

$$\rho_k^{y^2} = \frac{E[y_t^2 y_{t-k}^2] - \sigma_y^4}{2\sigma_y^4} = (\rho_k^y)^2.$$

That is, for any stationary Gaussian process $\{y_t\}$, the autocorrelation of the squared process $\{y_t^2\}$ is the square of the autocorrelation of $\{y_t\}$ and hence the autocorrelation of any stationary Gaussian process $\{y_t\}$ is larger than the autocorrelation of $\{y_t^2\}$(i.e $|\rho_k^y| \geq \rho_k^{y^2}$). The squared process plays an important role in model identification (See Thavaneswaran et al. (2005)) and volatility forecasting (See Thavaneswaran, Appadoo and Peiris (2005)). In the following set of theorems, we study the first four moments, the autocorrelation structure of the process and its square process.

Part (i) of the following theorem have been derived by Aue (2004). Parts (ii) and (iii) have been derived by Appadoo, Ghahramani and Thanveswaran (2005).

Theorem 2.1.1 *Let $\{y_t\}$ be a RCA(1) time series satisfying conditions (2.1.1) and (2.1.2), and let γ_y be its covariance function.*

(i) We have

$$\begin{aligned} E(y_t) &= 0, \\ E(y_t^2) &= \frac{\sigma_e^2}{1-\phi^2-\sigma_b^2}, \end{aligned}$$

and the k-th lag autocovariance for y_t is given by

$$\gamma_k^y = \frac{\phi^k \sigma_e^2}{1-\phi^2-\sigma_b^2},$$

while the autocorrelation for y_t is $\rho_k^y = \phi^k$ for all $k \in Z$. That is, the usual AR(1) process has same autocorrelation as the RCA(1).

(ii) If $\{b_t\}$ and $\{e_t\}$ are normally distributed random variables then the kurtosis $K^{(y)}$ of the RCA process $\{y_t\}$ is given by

$$K^{(y)} = \frac{3(1-(\sigma_b^2+\phi^2)^2)}{1-(\phi^4+6\phi^2\sigma_b^2+3\sigma_b^4)},$$

and for an AR(1) process $K^{(y)}$ reduces to 3.

(iii) The autocorrelation of y_t^2 assuming b_t and e_t are normally distributed is given by

$$\rho_k^{y^2} = (\phi^2+\sigma_b^2)^k$$

and for a Gaussian AR(1) process it turns out to be

$$\rho_k^{y^2} = \phi^{2k}.$$

Proof. By conditioning on y_{t-1}, we obtain $E(y_t) = EE(y_t|y_{t-1}) = 0$, and

$$\begin{aligned}
\mathrm{Var}(y_t) &= \mathrm{E}(\mathrm{Var}(y_t|y_{t-1})) + \mathrm{Var}(\mathrm{E}(y_t|y_{t-1})) \\
&= \mathrm{E}(\sigma_b^2 y_{t-1}^2 + \sigma_e^2) + \mathrm{Var}(\phi y_{t-1}) \\
&= \sigma_e^2 + \sigma_b^2 \mathrm{E}(y_{t-1}^2) + \phi^2 \mathrm{Var}(y_{t-1}),
\end{aligned}$$

so that

$$\mathrm{Var}(y_t) = \gamma_0^y = \frac{\sigma_e^2}{(1-\sigma_b^2-\phi^2)}.$$

The l-lag autocovariance is

$$\begin{aligned}
\mathrm{E}(y_t y_{t-1}) &= \mathrm{E}(\phi y_{t-1} y_{t-1}) + \mathrm{E}(b_t y_{t-1} y_{t-1}) + \mathrm{E}(e_t y_{t-1}) \\
\gamma_1^y &= \phi \gamma_0^y.
\end{aligned}$$

Similarly, the k-th lag autocovariance is,

$$\gamma_k^y = \phi^k \gamma_0^y.$$

For the proof of (iii), we have

$$\begin{aligned} y_t^4 &= [(\phi + b_t)y_{t-1} + e_t]^4 \\ \mathrm{E}(y_t^4) &= \mathrm{E}((\phi + b_t)^4 y_{t-1}^4) + 6\mathrm{E}((\phi + b_t)^2 y_{t-1}^2 e_t^2) + 3\sigma_e^4 \qquad (2.1.3) \\ \mathrm{E}((\phi + b_t)^4 y_{t-1}^4) &= (\phi^4 + 6\phi^2\sigma_b^2 + 3\sigma_b^4)E(y_{t-1}^4), \end{aligned}$$

and

$$\begin{aligned} \mathrm{E}((\phi + b_t)^2 y_{t-1}^2 e_t^2] &= E[(\phi^2 e_t^2 + 2\phi b_t e_t^2 + b_t^2 e_t^2) y_{t-1}^2) \\ &= \mathrm{EE}((\phi^2 e_t^2 + 2\phi b_t e_t^2 + b_t^2 e_t^2) y_{t-1}^2 | y_{t-1}) = \mathrm{E}((\phi^2\sigma_e^2 + \sigma_b^2\sigma_e^2) y_{t-1}^2) \\ &= \sigma_e^2(\phi^2 + \sigma_b^2)\frac{\sigma_e^2}{(1 - \sigma_b^2 - \phi^2)} \\ &= \frac{\sigma_e^4(\phi^2 + \sigma_b^2)}{1 - \sigma_b^2 - \phi^2}, \end{aligned}$$

By substitution, we obtain the following results:

$$\begin{aligned} \mathrm{E}(y_t^4) &= (\phi^4 + 6\phi^2\sigma_b^2 + 3\sigma_b^4)\mathrm{E}(y_{t-1}^4) + 6\frac{\sigma_e^4(\phi^2 + \sigma_b^2)}{(1 - \sigma_b^2 - \phi^2)} + 3\sigma_e^4 \\ &= \frac{3\sigma_e^4(1 + \phi^2 + \sigma_b^2)}{(1 - \sigma_b^2 - \phi^2)(1 - (\phi^4 + 6\phi^2\sigma_b^2 + 3\sigma_b^4))} \end{aligned}$$

Hence, the kurtosis is given by

$$K^{(y)} = \frac{3(1 - (\sigma_b^2 + \phi^2)^2)}{1 - (\phi^4 + 6\phi^2\sigma_b^2 + 3\sigma_b^4)}.$$

We now derive the autocorrelation of y_t^2 of the RCA(1) process. Since

$$\begin{aligned} E(y_t^2 y_{t-k}^2) &= E(\phi_1^2 y_{t-1}^2 y_{t-k}^2) + E(b_t^2 y_{t-1}^2 y_{t-k}^2) + E(e_t^2 y_{t-k}^2) \\ &= (\phi_1^2 + \sigma_b^2)E(y_{t-1}^2 y_{t-k}^2) + \sigma_e^2 E(y_{t-k}^2) \end{aligned}$$

we obtain

$$\rho_1^{y^2} = \frac{E[y_t^2 y_{t-1}^2] - \sigma_y^4}{E(y_t^4) - \sigma_y^4} = \frac{\dfrac{E[y_t^2 y_{t-1}^2]}{\sigma_y^4} - 1}{K^{(y)} - 1} = (\phi^2 + \sigma_b^2).$$

Similarly, the autocorrelation of y_t^2 at lag k is given by $\rho_k^{y_t^2} = (\phi^2 + \sigma_b^2)^k$.

Note that when σ_b=0, the kurtosis of y_t in part (iii) reduces to that of a standard AR(1) process, which is equal to 3. Easy computation shows that $K_{\text{AR}}^{(y)} \leq K_{\text{RCA}}^{(y)}$. Consider a more general form of the RCA(1) model of the form

$$y_t = \theta_t y_{t-1} + e_t. \tag{2.1.4}$$

Then the following corollary is true.

Corollary 2.1.1 *Let $\{y_t\}$ be a RCA(1) time series of the form 2.1.4 satisfying the stationarity conditions (2.1.1)-(2.1.2), and let ρ_k^y denote its autocorrelation function. Then the following hold.*

(i) When $\theta_t = \phi + b_t$, $E(y_t) = 0$, $E(y_t^2) = \sigma_e^2/(1 - \phi^2 - \sigma_b^2)$, the k-th lag autocorrelation for y_t is given by $\rho_k^y = (E(\phi + b_t))^k = \phi^k$.

(ii) When $\theta_t = sgn(b_t)$ where $b_t \sim (0, \sigma_b^2)$ and $sgn(b_t) = I_{(b_t \geq 0)} - I_{(b_t < 0)}$, ($I_{(.)}$ is the indicator function), then $\rho_k^y = (1 - 2F(0))^k$, where F is the cumulative distribution function of b_t. Thus, when the coefficient θ_t is driven by a binary random variable $\{b_t\}$ taking values -1 and +1, the autocorrelation of the process depends on the marginal distribution of $\{b_t\}$.

(iii) When $\theta_t = (\phi + |b_t|^\alpha)$ where $b_t \sim N(\mu, \sigma_b^2)$ then, the autocorrelation is

$$\rho_k^y = \left(\phi + \frac{(2\sigma_b^2)^{\frac{\alpha}{2}}}{\sqrt{\pi}} \Gamma\left(\frac{\alpha + 1}{2}\right)\right)^k,$$

where $\Gamma(.)$ *is the Gamma function.*

Proof: The proof of (i) follows by observing, that

$$\mathrm{EE}(y_t y_{t-k}|\theta_t) = \mathrm{EE}(\theta_t y_{t-1} y_{t-k} + e_t y_{t-k}|\theta_t) = \mathrm{E}(\theta_t \mathrm{E}(y_{t-1} y_{t-k})) = \mathrm{E}(\theta_t)\gamma_{k-1}^y$$

so that

$$\rho_k^y = \frac{\mathrm{E}(\theta_t)\gamma_{k-1}^y}{\gamma_0^y} = \mathrm{E}(\theta_t)\frac{\gamma_{k-1}^y}{\gamma_0^y} = \mathrm{E}(\theta_t)\rho_{k-1}^y = (E(\theta_t))^k = (E(\phi + b_t))^k = \phi^k$$

For the proof of (ii), if we let $y_t = \theta_t y_{t-1} + e_t$ and $\theta_t = \mathrm{sgn}(b_t) = I_{(b_t \geq 0)} - I_{(b_t < 0)}$, then

$$\mathrm{E}(\theta_t) = \mathrm{E}(\mathrm{sgn}(b_t)) = P(b_t \geq 0) - P(b_t < 0) = 1 - 2F(0),$$

so that $\rho_k^y = (1 - 2F(0))^k$. Part (iii) follows from the fact that for a normal random variable X, having mean 0 and variance σ^2 the α^{th} absolute moment of X is

$$E|X|^\alpha = \frac{(2\sigma^2)^{\frac{\alpha}{2}}}{\sqrt{\pi}}\Gamma\left(\frac{\alpha+1}{2}\right).$$

2.1.1 Random coefficient autoregressive with moving average errors

The RCA models with moving average innovation terms are an extension of the classical RCA model of Nicholls and Quinn (1982). This new formulation was proposed by Appadoo, Ghahramani and Thavaneswaran (2005). The proof of the moments and the kurtosis is similar to that of the RCA model discussed in the previous section. A sequence of random variables $\{y_t\}$ is called an RCA(1)–MA(1) time series if it satisfies the equations.

$$y_t = (\phi + b_t)y_{t-1} + e_t + \theta e_{t-1} \qquad t \in Z, \tag{2.1.5}$$

such that (2.1.1)-(2.1.2) hold.

Lemma 2.1.1 *Let $\{y_t\}$ be the RCA(1)–MA(1) time series model described by equation (2.1.5) and let γ_k^y denotes its autocovariance function. Then the following hold.*

$$
\begin{aligned}
E(y_t) &= 0, \\
\gamma_0^y &= \frac{\sigma_e^2(1+\theta^2)}{1-\sigma_b^2-\phi^2}, \\
\gamma_k^y &= \phi\gamma_{k-1}^y, \quad |k|>1, \\
\gamma_1^y &= \phi\gamma_0^y + \theta\sigma_e^2
\end{aligned}
$$

and the autocorrelation and kurtosis of the process are given by

$$
\rho_k^y = \begin{cases} 1 & k=0 \\ \phi + \dfrac{\theta}{1+\theta^2}(1-\sigma_b^2-\phi^2) & |k|=1 \\ \phi\ \rho_{k-1}^y & |k|>1 \end{cases}
$$

and

$$
K^{(y)} = \frac{3(1+\phi^2+\sigma_b^2)(1-\sigma_b^2-\phi^2)}{1-(\phi^4+6\phi^2\sigma_b^2+3\sigma_b^4)}.
$$

Proof: The proof is somewhat similar to the proof of Theorem 2.1.1. Some of the details are as follows.

$$
\begin{aligned}
y_t^2 &= (\phi+b_t)^2y_{t-1}^2 + 2(\phi+b_t)y_{t-1}(e_t+\theta e_{t-1}) + (e_t+\theta e_{t-1})^2, \\
E(y_t^2) &= (\phi^2+\sigma_b^2)E(y_{t-1}^2) + (1+\theta^2)\sigma_e^2,
\end{aligned}
$$

$$
\begin{aligned}
y_t^4 &= (\phi+b_t)^4y_{t-1}^4 + 4(\phi+b_t)^3y_{t-1}^3(e_t+\theta e_{t-1}) \\
&+ 6(\phi+b_t)^2y_{t-1}^2(e_t+\theta e_{t-1})^2 + (\phi+b_t)y_{t-1}(e_t+\theta e_{t-1})^3 \\
&+ (e_t+\theta e_{t-1})^4.
\end{aligned}
$$

And,

$$
E(y_t^4) = \frac{3\sigma_e^4(1+\theta^2)^2[1+(\phi^2+\sigma_b^2)]}{[1-(\phi^2+\sigma_b^2)][1-(\phi^4+6\phi^2\sigma_b^2+3\sigma_b^4)]}
$$

It is of interest to note that an AR(1) process, an RCA(1) process and an RCA(1)-MA(1) process have the same autocorrelation structure.

Lemma 2.1.2 *For a simple time series model of the form $y_t = \phi y_{t-2}\varepsilon_{t-1} + \varepsilon_t$, where ε_t is a Gaussian white noise with variance σ_ε^2, the variance of the process is given by*

$$Var(y_t) = \frac{\sigma_\varepsilon^2}{1-\phi^2\sigma_\varepsilon^2},$$

and the kurtosis is given by

$$K^{(y)} = \frac{3[1-\phi^4\sigma_\varepsilon^4]}{(1-3\phi^4\sigma_\varepsilon^4)}.$$

Proof. We obtain

$$\begin{aligned}
E[y_t] &= 0 \text{ by conditioning,} \\
E[y_t^2] &= \phi^2 E[y_{t-2}^2]\sigma_\varepsilon^2 + \sigma_\varepsilon^2 = \frac{\sigma_\varepsilon^2}{1-\phi^2\sigma_\varepsilon^2}, \text{ assuming stationarity of } y_t \\
E[y_t^4] &= E[\phi^4 y_{t-2}^4\varepsilon_{t-1}^4 + 4\phi^3 y_{t-2}^3\varepsilon_{t-1}^3\varepsilon_t + 6\phi^2 y_{t-2}^2\varepsilon_{t-1}^2\varepsilon_t^2 + 4\phi y_{t-2}\varepsilon_{t-1}\varepsilon_t^3 + \varepsilon_t^4] \\
&= \phi^4 E[y_{t-2}^4]3\sigma_\varepsilon^4 + 6\phi^2 E[y_{t-2}^2]\sigma_\varepsilon^2\sigma_\varepsilon^2 + 3\sigma_\varepsilon^4.
\end{aligned}$$

Now

$$\begin{aligned}
E[y_t^4] - \phi^4 E[y_{t-2}^4]3\sigma_\varepsilon^4 &= 6\phi^2 E[y_{t-2}^2]\sigma_\varepsilon^4 + 3\sigma_\varepsilon^4 \\
&= 6\phi^2\left[\frac{\sigma_\varepsilon^2}{1-\phi^2\sigma_\varepsilon^2}\right]\sigma_\varepsilon^4 + 3\sigma_\varepsilon^4, \\
E[y_t^4] &= \frac{6\phi^2\sigma_\varepsilon^6 + 3\sigma_\varepsilon^4(1-\phi^2\sigma_\varepsilon^2)}{(1-\phi^2\sigma_\varepsilon^2)[1-3\phi^4\sigma_\varepsilon^4]} \\
&= \frac{3\sigma_\varepsilon^4[1+\phi^2\sigma_\varepsilon^2]}{(1-\phi^2\sigma_\varepsilon^2)[1-3\phi^4\sigma_\varepsilon^4]}
\end{aligned}$$

Hence, the kurtosis follows.

Example 2.1.1 *Let us consider,*

$$y_t = \theta_1 \varepsilon_{t-1}^2 \varepsilon_t + \theta_2 \varepsilon_{t-2}^2 \varepsilon_t$$

where ε_t is a Gaussian white noise with variance σ_ε^2. We have

$$\begin{aligned} E(y_t) &= 0 \qquad (2.1.6) \\ E(y_t^2) &= \sigma_\varepsilon^6(3\theta_1^2 + 2\theta_1\theta_2 + 3\theta_2^2) \\ E(y_t^4) &= [315(\theta_1^4 + \theta_2^4) + 180(\theta_1^3\theta_2 + \theta_1\theta_2^3) + 162\theta_1^2\theta_2^2\ ]\sigma_\varepsilon^{12} \end{aligned}$$

so that,

$$\begin{aligned} K^{(y)} &= \frac{E(y-\mu)^4}{[Var(y)]^2} = \frac{[315(\theta_1^4 + \theta_2^4) + 180(\theta_1^3\theta_2 + \theta_1\theta_2^3) + 162\theta_1^2\theta_2^2\ ]\sigma_\varepsilon^{12}}{\sigma_\varepsilon^{12}(3\theta_1^2 + 2\theta_1\theta_2 + 3\theta_2^2)^2} \\ &= \left[\frac{[315(\theta_1^4 + \theta_2^4) + 180(\theta_1^3\theta_2 + \theta_1\theta_2^3) + 162\theta_1^2\theta_2^2\ ]}{(3\theta_1^2 + 2\theta_1\theta_2 + 3\theta_2^2)^2}\right]. \end{aligned}$$

The derivation of the autocorrelation of y_t is similar to that for y_t in the motivating examples in the beginning of this chapter due to Gouriéroux (1997) and hence is omitted. Moreover, for any conditionally Gaussian process of the form,

$$y_t = f(\varepsilon_{t-1}, \ldots \varepsilon_{t-k})\varepsilon_t$$

where f is a measurable function of $\varepsilon_{t-1}, \ldots, \varepsilon_{t-k}$ and ε_t is a zero mean Gaussian process, then

$$E(y_t^4) = EE(y_t^4|y_{t-1}) = E3[E(y_t^2|y_{t-1})]^2 \geq 3(EE(y_t^2|y_{t-1}))^2 = 3(E(y_t^2))^2,$$

and

$$K^{(y)} = \frac{E(y_t^4)}{E(y_t^2)^2} \geq 3.$$

2.2 GARCH Models

Consider the general class of GARCH (P, Q) model for the time series y_t, where

$$\begin{aligned} y_t &= \sqrt{h_t} Z_t, && (2.2.1) \\ h_t &= \omega + \sum_{i=1}^{P} \alpha_i y_{t-i}^2 + \sum_{j=1}^{Q} \beta_j h_{t-j} && (2.2.2) \end{aligned}$$

and Z_t is a sequence of independent, identically distributed random variables with zero mean, unit variance. Let $u_t = y_t^2 - h_t$ be the martingale difference and let σ_u^2 be the variance of u_t. Rearranging terms and by noting that $h_{t-j} = y_{t-j}^2 - u_{t-j}$, then (2.2.1) and (2.2.2) could be written as:

$$\begin{aligned} y_t^2 - u_t &= \omega + \sum_{i=1}^{P} \alpha_i y_{t-i}^2 + \sum_{j=1}^{Q} \beta_j h_{t-j}, && (2.2.3) \\ \left[1 - \sum_{i=1}^{P} \alpha_i B^i - \sum_{j=1}^{Q} \beta_j B^j\right] y_t^2 &= \omega - \sum_{j=1}^{Q} \beta_j B^j u_t + u_t, && (2.2.4) \\ \Phi(B) y_t^2 &= \omega + \beta(B) u_t. && (2.2.5) \end{aligned}$$

where, $\Phi(B) = 1 - \sum_{i=1}^{R} \Phi_i B^i$, $\Phi_i = (\alpha_i + \beta_i)$, $\beta(B) = 1 - \sum_{j=1}^{Q} \beta_j B^j$ and $R = \max(P, Q)$. We shall make the following stationarity assumptions for y_t^2 which has an ARMA(R, Q) representation. We can view the square of the process as a measure of variability of the return process. The ARMA representation in terms of y_t^2 allows us to forecast volatility, for example.

Assumption 2.1 All zeroes of the polynomial $\Phi(B)$ lie outside of the unit circle.

Assumption 2.2 We have $\sum_{i=0}^{\infty} \Psi_i^2 < \infty$ where the $\Psi_i' s$ are obtained from the relation $\Psi(B)\ \Phi(B) = \beta(B)$ with $\Psi(B) = 1 + \sum_{i=1}^{\infty} \Psi_i B^i$.

Assumptions 2.1 and 2.2 ensure that the $\{u_t\}$ are uncorrelated with zero mean and finite variance and that the y_t^2 process is weakly stationary. In this case, the autocorrelation function of y_t^2 will be exactly the same as that for a stationary ARMA(R, Q) model. For any random variable y with finite fourth moment, the kurtosis is defined by $\frac{E(y-\mu)^4}{[Var(y)]^2}$ and is denoted by $K^{(y)}$. If $\{Z_t\}$ is a normal process, then the process $\{y_t\}$ defined by equations (2.2.1) and (2.2.2) is called a normal GARCH (P, Q) process. In order to calculate the GARCH kurtosis in terms of the Ψ-weights and the autocorrelation function (ACF) of the squared process, the following theorem given in Thavaneswaran et al. (2005) will be useful.

Theorem 2.2.1 *(i) For the GARCH(P, Q) process specified by (2.2.1) and (2.2.2), under the stationarity assumptions (2.1)-(2.2) and having finite fourth moment, the kurtosis $K^{(y)}$ of the process is given by*

$$K^{(y)} = \frac{E(Z_t^4)}{E(Z_t^4) - [E(Z_t^4) - 1] \sum_{j=0}^{\infty} \Psi_j^2},$$

(ii) The variance of the y_t^2 process is $\gamma_0^{y^2} = \sum_{j=0}^{\infty} \Psi_j^2 \sigma_u^2$, $\sigma_u^2 = \frac{\mu^2(K^{(y)} - 1)}{\sum_{j=0}^{\infty} \Psi_j^2}$ and $\mu =$

$$E(y_t^2) = \frac{\omega}{1 - \Phi_1 - \ldots - \Phi_R},$$

(iii) The kth lag autocovariance of the y_t^2 process is

$$\gamma_k^{y^2} = \sigma_u^2 \sum_{j=0}^{\infty} \Psi_{k+j} \Psi_j,$$

(iv) The kth-lag autocorrelation is given by

$$\rho_k^{y^2} = \frac{\gamma_k^{y^2}}{\gamma_0^{y^2}} = \frac{\sum_{j=0}^{\infty} \Psi_{k+j} \Psi_j}{\sum_{j=0}^{\infty} \Psi_j^2},$$

(v) For a normal GARCH(P, Q) process

$$K^{(y)} = \frac{3}{1 - 2\sum_{j=1}^{\infty} \Psi_j^2}.$$

It is of interest to note that Bai, Russell and Tiao (2003) have derived also studied the moment properties of GARCH processes. In particular, they have studied the relationship between the kurtosis of a GARCH process, its excess kurtosis, the kurtosis of the error term Z_t and the ψ-weights.

2.2.1 Special Cases

In the following examples we show that the results for normal GARCH(1, 1), t-distribution GARCH (1, 1), IGARCH, double-exponential GARCH, GED-GARCH and Power GARCH are special cases of (2.2.1), (2.2.2) and (2.2.5).

Example 2.2.1 *(normal GARCH(1, 1)).*

In this example, we show that the results for a normal GARCH(1,1) model and an ARCH (1) model are special cases of Theorem 2.2.1.

For the GARCH (1,1) model, (2.2.4) and (2.2.5) are given by the following:

$$\begin{aligned} y_t &= \sqrt{h_t} Z_t, \\ h_t &= \omega + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1}, \end{aligned}$$

Let $u_t = y_t^2 - h_t$. Then,

$$y_t^2 - u_t = \omega + \alpha_1 y_{t-1}^2 + \beta_1 (y_{t-1}^2 - u_{t-1}).$$

This shows that for a GARCH(1, 1) process y_t, y_t^2 has ARMA(1, 1) representation.

$$y_t^2 - \alpha_1 y_{t-1}^2 - \beta_1 y_{t-1}^2 = \omega + u_t - \beta_1 u_{t-1},$$

And, $(1-\Phi_1 B)y_t^2 = \omega + (1-\beta_1 B)u_t$, where $\Psi_1 = \alpha_1$, $\Psi_2 = \alpha_1(\alpha_1+\beta_1)$, $\Psi_3 = \alpha_1(\alpha_1+\beta_1)^2, \ldots$ $\Psi_j = \alpha_1(\alpha_1+\beta_1)^{(j-1)}$, $j \geq 1$ and

$$\sum_{j=1}^{\infty} \Psi_j^2 = \alpha_1^2 + \alpha_1^2(\alpha_1+\beta_1)^2 + \ldots = \frac{\alpha_1^2}{1-(\alpha_1+\beta_1)^2}.$$

By Theorem 2.2.1(iv),

$$\begin{aligned} K^{(y)} &= \frac{3}{1-2\sum\limits_{j=1}^{\infty}\Psi_j^2} \\ &= \frac{3(1-(\alpha_1+\beta_1)^2)}{1-(\alpha_1+\beta_1)^2-2\alpha_1^2}, \end{aligned}$$

and this turns out to be the same as that given in Bollerslev(1986).

For the ARCH(1) model of the form

$$\begin{aligned} y_t &= \sqrt{h_t}Z_t, \\ h_t &= \omega + \alpha_1 y_{t-1}^2, \\ u_t &= y_t^2 - h_t, \end{aligned}$$

if we set $\beta_1 = 0$ and use the method given earlier in this example, then we obtain

$$K^{(y)} = \frac{3(1-\alpha_1^2)}{1-3\alpha_1^2}.$$

Example 2.2.2 *The t_ν distribution GARCH model.*

For any t-GARCH (P, Q) process driven by a t-distribution with d.f. $\nu > 4$, the r^{th} central moment (when r is even) of such a heavy tailed distribution is given by $\mu_r(t_\nu) = \nu^{\frac{1}{2}r}\dfrac{1.3.\ldots.(r-1)}{(\nu-r)(\nu-r+2)\ldots.(\nu-2)}$, $\mathrm{var}[t_\nu] = \dfrac{\nu}{\nu-2}$ $(\nu \geq 3)$ and the kurtosis,

$K = \dfrac{6}{\nu - 4} + 3$ $(\nu \geq 5)$. The r^{th} central moment is zero when r is odd.

If we assume a t−distribution with ν d.f. $(\nu \geq 5)$ for Z_t,then $E[Z_t^4] = \dfrac{6}{\nu - 4} + 3$ and using Theorem 2.2.1 we can obtain the kurtosis of the t-GARCH (P, Q) process by working with a t-distribution standardized to have zero mean and unit variance in the following manner.

Suppose $\tilde{Z}_t \sim t_\nu$. Then $E(\tilde{Z}_t^2) = \dfrac{\nu}{\nu - 2}$ whenever $\nu > 2$ and $E(\tilde{Z}_t^4) = \dfrac{3\nu^2}{(\nu-4)(\nu-2)}$ whenever $\nu > 4$. Consider the standardized random variable $Z_t = \sqrt{\dfrac{\nu-2}{\nu}}\tilde{Z}_t$.

Then

$$E(Z_t^4) = \frac{(\nu-2)^2}{\nu^2}E(\tilde{Z}_t^4) = \frac{3(\nu-2)}{\nu-4}.$$

For a GARCH(1, 1) process, $\displaystyle\sum_{j=0}^{\infty}\Psi_j^2 = \frac{1 - \Phi_1^2 + \alpha_1^2}{1 - \phi_1^2}$, where $\Phi_1 = \alpha_1 + \beta_1$.

Using Theorem 2.2.1, the kurtosis of the t_ν-distribution GARCH(1, 1) process is given by the following:

$$\begin{aligned} K^{(y)} &= \frac{E(Z_t^4)}{E(Z_t^4) - [E(Z_t^4) - 1]\sum_{j=0}^{\infty}\Psi_j^2} \\ &= \frac{3(\nu-2)/(\nu-4)}{3(\nu-2)/(\nu-4) - [3(\nu-2)/(\nu-4) - 1](1 - \phi_1^2 + \alpha_1^2)/(1-\phi_1^2)} \\ &= \frac{3(\nu-2)[1 - (\alpha_1+\beta_1)^2]}{(\nu-4)(1 - 2\alpha_1\beta_1 - \beta_1^2) - 3(\nu-2)\alpha_1^2}. \end{aligned}$$

Example 2.2.3 *The double exponential GARCH model.*

For any GARCH(1, 1) process y_t where Z_t is an uncorrelated sequence following a standard double exponential density

$$f(z) = \frac{1}{\sqrt{2}}e^{\sqrt{2}|z|}, \quad -\infty < z < \infty$$

using Theorem 2.2.1 and the fact that $E(Z_t^4) = 6$, and $\sum_{j=1}^{\infty} \Psi_j^2 = \frac{1-\Phi_1^2+\alpha_1^2}{1-\Phi_1^2}$, ($\Phi_1 = \alpha_1 + \beta_1$) the kurtosis of the process is

$$K^{(y)} = \frac{6[1-(\alpha_1+\beta_1)^2]}{1-(\alpha_1+\beta_1)^2-5\alpha_1^2}. \tag{2.2.6}$$

Example 2.2.4 *The Generalized Error Distribution (GED) GARCH model.*

Consider a GARCH(1, 1) process that conditionally follows a GED distribution given by

$$\begin{aligned} y_t &= \sqrt{h_t}\, Z_t \\ h_t &= \omega + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1}, \end{aligned}$$

where $Z_t \sim (0,1)$ has the following density:

$$f(z) = \frac{\nu}{\lambda\, 2^{(1+1/\nu)}\Gamma(1/\nu)} \exp\left(-\frac{1}{2}|z/\lambda|^{\nu}\right), \quad \lambda = \left(\frac{2^{-2/\nu}\Gamma(1/\nu)}{\Gamma(3/\nu)}\right)^{1/2}, \quad \nu > 0.$$

It is easy to show that $E(Z_t^2) = 1$ and $E(Z_t^4) = \frac{\Gamma(1/\nu)\Gamma(5/\nu)}{(\Gamma(3/\nu))^2}$. Then,

$$K^{(y)} = \frac{\Gamma(1/\nu)\Gamma(5/\nu)[1-(\alpha_1+\beta_1)^2]}{(\Gamma(3/\nu))^2[1-(\alpha_1+\beta_1)^2] - [\Gamma(1/\nu)\Gamma(5/\nu) - (\Gamma(3/\nu))^2]\alpha_1^2}.$$

It is of interest to note that for $\nu = 1$, the kurtosis reduces to the kurtosis of the double-exponential GARCH(1, 1) model and that for $\nu = 2$, the kurtosis reduces to the kurtosis of the normal-GARCH(1, 1) model.

Example 2.2.5 *(Integrated GARCH(IGARCH))*

For the IGARCH model considered in He and Terasvirta (1999),

$y_t = \sqrt{h_t} Z_t, \qquad h_t = \alpha y_{t-1}^2 + (1-\alpha)h_{t-1}, \qquad$ where Z_t is a sequence of independent, identically distributed random variables with zero mean, unit variance and α is a constant $0 \leq \alpha \leq 1$. Assuming that $h_0 = 1$, the kurtosis is given by, $K^{(y)} = \dfrac{E[y_t^4]}{[E(y_t^2)]^2} = [\alpha^2(K^{(Z)}-1)+1]^t K^{(Z)}$, where $K^{(Z)}$ is the kurtosis of the (Z_t) process. For more details on the proof of the kurtosis of the IGARCH process, see Thavaneswaran, Appadoo and Samanta (2005).

Example 2.2.6 *(power GARCH(1, 1))*

Consider the power GARCH(1, 1) model given by

$$\begin{aligned} y_t &= \sqrt{h_t} Z_t, \\ h_t^\delta &= \omega + \alpha_1 |y_{t-1}|^\delta + \beta_1 h_{t-1}^\delta, \end{aligned} \tag{2.2.7}$$

where Z_t is an uncorrelated mean zero process with unit variance. Let $u_t = |y_t|^\delta - h_t^\delta$ be the martingale difference. Then, $\Phi(B)|y_t|^\delta = \omega + \beta(B)u_t$, where $\Phi(B) = 1 - \Phi_1 B$, $\Phi_1 = \alpha_1 + \beta_1$, and $\beta(B) = 1 - \beta_1 B$. This shows that the power GARCH(1, 1) model could be represented as an ARMA(1, 1) for $|y_t|^\delta$.

In the following example, some of the moment properties of a normal-GARCH(1, 1) process are illustrated.

Example 2.2.7 *In this example, we simulate $n = 8000$ observations from a normal-GARCH(1, 1) process with parameters $\omega = 0.01$, $\alpha_1 = 0.2$ and $\beta_1 = 0.75$.*

The sample ACF of the process (shown in Figure 2.1) and its square (shown in Figure 2.2), illustrate the absence of autocorrelation in the process and the presence of autocorrelation that does not dampen in the square of the process. Table 2.1 gives the parameter estimates obtained via maximum likelihood estimation using *proc autoreg* and the method of moments using *proc arima* in SAS©. As the simulated process is

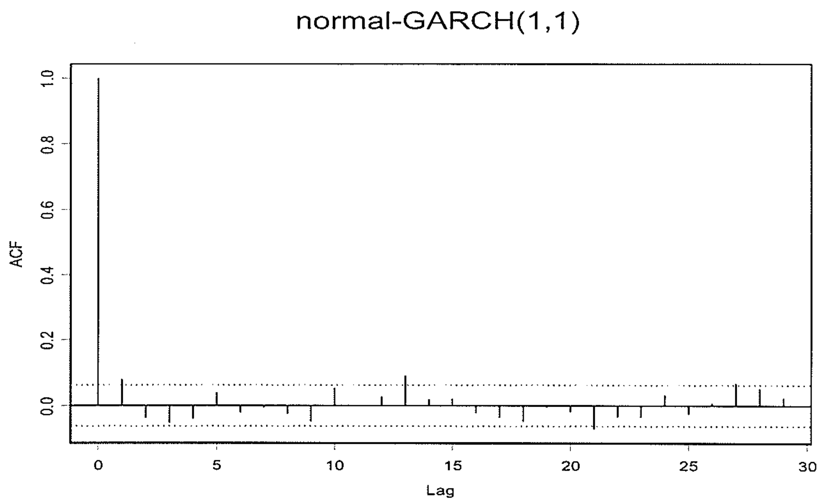

Figure 2.1: Sample ACF of a simulated normal-GARCH(1, 1) process.

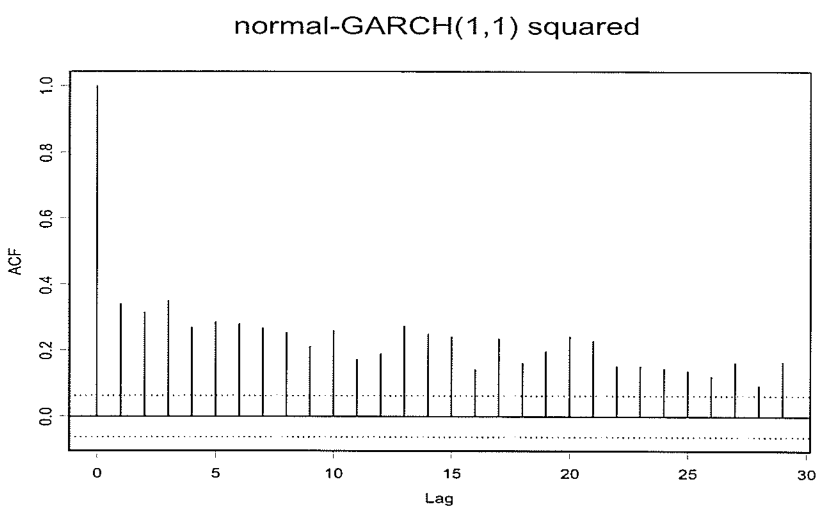

Figure 2.2: Sample ACF of the square of a simulated normal-GARCH(1, 1) process.

a normal-GARCH(1, 1) process, the maximum likelihood estimates(MLEs) will be efficient. However, the method of moments estimates are appealing as no distributional assumption is made about the Z_t process. The method of moments estimators of α_1 and β_1 are quite close to the true parameter values; whereas ω is overestimated by the method of moments estimator. Consequently, the unconditional variance will be overestimated if we were to use the moments based parameter estimates.

The empirical kurtosis is 4.93 indicating that the class of ARMA models would not be appropriate for fitting this data. The estimated kurtosis that based on the normal-GARCH(1, 1) model fit turns out to be 7.82, while the estimated kurtosis that is found by substituting the method of moments based parameter estimates is 3.20. The former estimate explains all of the sample kurtosis whereas the latter estimate does not fully explain the sample kurtosis. Therefore, identification of the appropriate error distribution is crucial. In Chapter 3, identification of the error distribution for a GARCH model will be studied.

Parameter	proc arima	proc autoreg
$\omega = 0.01$	0.0165	0.0106
$\alpha_1 = 0.2$	0.1717	0.1790
$\beta_1 = 0.75$	0.7406	0.7676

Table 2.1: Parameter estimates of a simulated normal-GARCH(1, 1) process using maximum likelihood and method of moments estimation.

In the next example, an application of ARMA models with GARCH errors is studied.

Example 2.2.8 *(S&P 500 Index of Composite Excess Returns)*

The GARCH(1, 1) model is the most widely used GARCH model for fitting financial return series (Hansen and Lunde (2004)). Tsay (2001) analysed the S&P 500 Composite Index of excess returns data by initially fitting an AR(3) model with GARCH(1, 1) errors. The motivation for fitting an autoregressive component is due to the presence of the first and third significant lags in the sample partial autocorrelation function

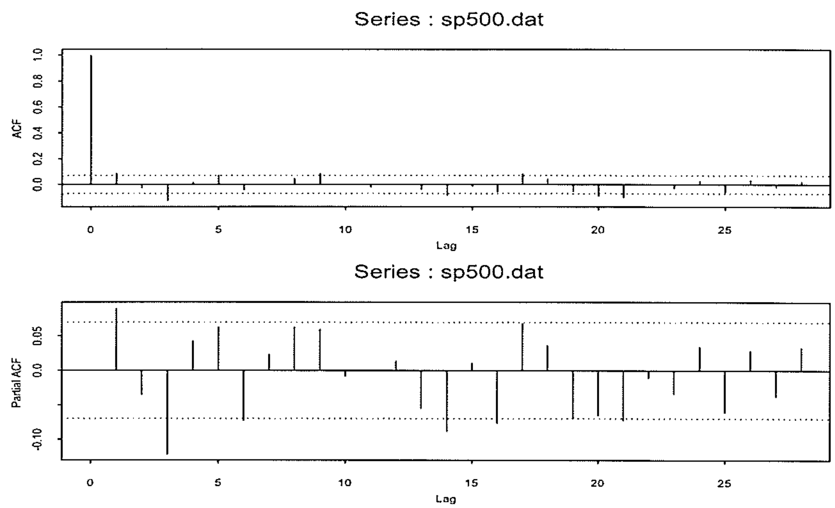

Figure 2.3: Sample ACF and PACF of S&P 500 data.

(PACF) plot as shown in Figure 2.3. Let $\{y_t\}$ denote the return process and let a_t denote the white noise process in an AR(3) model. The estimated AR(3) model is given by

$$y_t - 0.0062 = 0.089(y_{t-1} - 0.0062) - 0.024(y_{t-2} - 0.0062) - 0.13(y_{t-3} - 0.0062) + a_t.$$

The sample ACF of residuals from this fit and their squared process are examined in Figure 2.4, respectively.

However, upon fitting an AR(3) model with GARCH(1,1) errors, the autoregressive parameters failed to be significant and Tsay selects a GARCH(1, 1) for this data. The sample ACF of the standardized residuals and squared standardized residuals in Figure 2.5 reveal that the ARCH effects have been accounted for by the GARCH(1, 1) fit. Tables 2.2 and 2.3 provide the parameter estimates for an AR(3) model fit and an AR(3) model with GARCH(1, 1) errors. Table 2.4 gives the parameter estimates for a mean model with GARCH(1,1) errors. The estimated model is given by

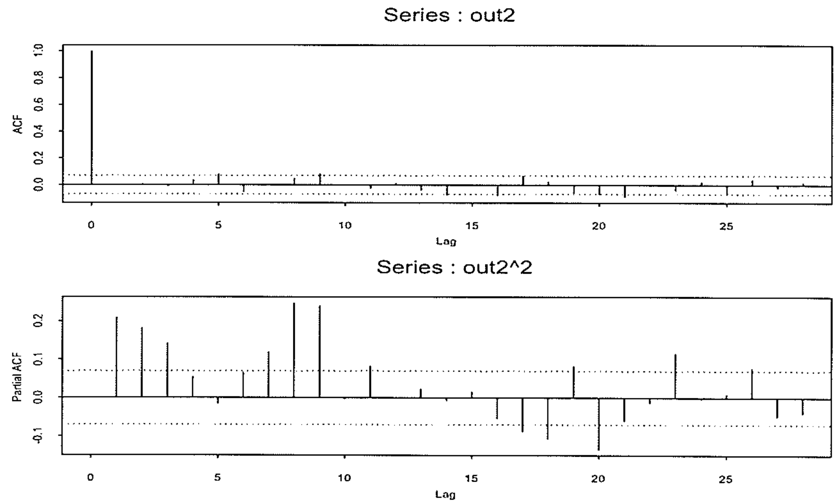

Figure 2.4: Sample ACF of residuals and squared residuals of AR(3) fit to S&P 500 data.

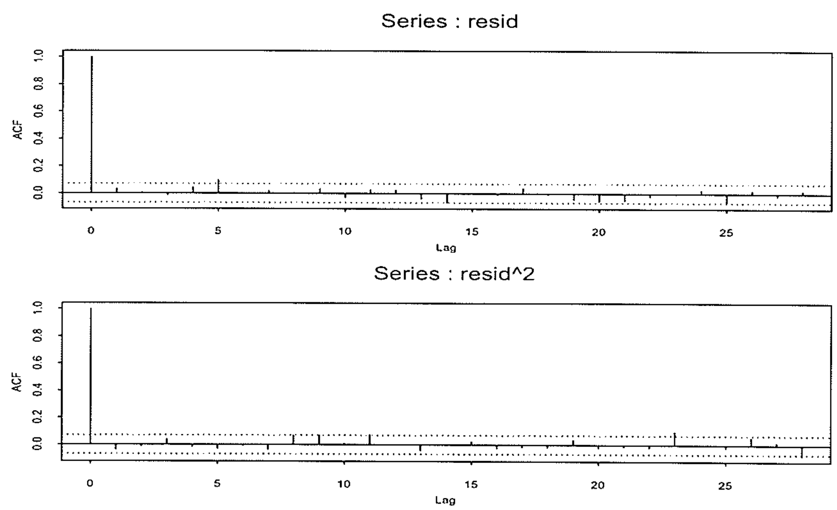

Figure 2.5: Sample ACF of residuals and squared residuals of GARCH(1, 1) fit to S&P 500 data.

$$\begin{aligned} y_t &= 0.0075 + \varepsilon_t, \\ \varepsilon_t &= \sqrt{h_t} Z_t, \\ h_t &= 0.000082 + 0.12\varepsilon_{t-1}^2 + 0.85 h_{t-1}. \end{aligned}$$

All parameter estimates are significant in this case. The estimated kurtosis assuming a normal-GARCH(1,1) model is given by $\hat{K}^{(y)} = 11.15$ while the sample kurtosis is 12.36 indicating that the normal-GARCH(1,1) explains most of the leptokurtosis. This example has motivated us to study how the ψ-weights of autoregressive models with GARCH errors depend on the autoregressive parameters.

Parameter	DF	Estimate	S.E.	t-Value	p-value
μ	1	0.006159	0.001945	3.17	0.0016
ϕ_1	1	0.08909	0.03544	2.51	0.0121
ϕ_2	1	-0.02376	0.03559	-0.67	0.5046
ϕ_3	1	-0.12297	0.03545	-3.47	0.0006

Table 2.2: Parameter Estimates of AR(3) model fit to S&P 500 index of returns.

Parameter	DF	Estimate	S.E.	t-Value	p-value
μ	1	0.007474	0.001580	4.73	<.0001
ϕ_1	1	-0.0337	0.0385	-0.87	0.3819
ϕ_2	1	0.0312	0.0383	0.82	0.4146
ϕ_3	1	0.0101	0.0356	0.28	0.7759
ω	1	0.0000805	0.0000240	3.35	0.0008
α_1	1	0.1200	0.0202	5.94	<.0001
β_1	1	0.8552	0.0196	43.54	<.0001

Table 2.3: Parameter Estimates of AR(3)/GARCH(1,1) model fit to S&P 500 index of returns.

Parameter	DF	Estimate	S.E.	t-Value	p-value
μ	1	0.007453	0.001547	4.82	<.0001
ω	1	0.0000818	0.0000238	3.44	0.0006
α_1	1	0.1203	0.0197	6.12	<.0001
β_1	1	0.8545	0.0189	45.15	<.0001

Table 2.4: Parameter Estimates of GARCH(1,1) model fit to S&P 500 index of returns.

In the next section, we study a new class of models that are extensions of both the class of GARCH models and the class of RCA models. All theorems and lemmas in the following sections have been studied by Appadoo, Ghahramani and Thavaneswaran (2005).

2.2.2 Random Coefficient ARCH (1) model

By analogy with the RCA models we introduce a class of models which can be viewed as RCA versions of GARCH models. Consider the ARCH (1) model for the time series y_t, where

$$
\begin{aligned}
y_t &= \sqrt{h_t} Z_t && (2.2.8)\\
h_t &= \omega + (\alpha_1 + b_{t-1}) y_{t-1}^2, && (2.2.9)
\end{aligned}
$$

Z_t is a sequence of independently, identically distributed random variables with zero mean and variance σ_Z^2 and b_t is an uncorrelated sequence with zero mean and variance σ_b^2.

Lemma 2.2.1 *For the model*

$y_t = \sqrt{h_t}\ Z_t,\quad h_t = \omega_0 + (\alpha_1 + b_{t-1}) y_{t-1}^2,$ *where* $Z_t \sim N(0, \sigma_Z^2)$ *and is uncorrelated with* $b_t \sim NID(0, \sigma_b^2)$*, the kurtosis, is given by*
$K^{(y)} = \dfrac{3[1 - \alpha_1^2 \sigma_Z^4]}{[1 - 3\sigma_Z^2(\alpha_1^2 + \sigma_b^2)]}.$

Proof: We have $y_t^2 = h_t Z_t^2$ and $E[y_t^2] = E[h_t]\sigma_Z^2$.

We now use this relationship to find the expected value of $E(h_t)$.

$$E[h_t] = \omega_0 + \alpha_1 E[y_{t-1}^2].$$

Assuming y_t is stationary, $E(y_t^2) = \dfrac{\omega_0 \sigma_Z^2}{[1 - \alpha_1 \sigma_Z^2]}$

$$E[h_t^2] = \frac{\omega_0^2 + 2\omega_0 \alpha_1 E(y_{t-1}^2)}{1 - 3\sigma_Z^4(\alpha_1^2 + \sigma_b^2)}.$$

$$E[y_t^4] = \frac{3\sigma_Z^4(\omega_0^2 + 2\alpha_1 \omega_0 E(y_{t-1}^2))}{1 - 3\sigma_Z^4(\alpha_1^2 + \sigma_b^2)}.$$

Hence,

$$K^{(y)} = \frac{3(1 - \alpha_1^2 \sigma_Z^4)}{1 - 3\sigma_Z^4(\alpha_1^2 + \sigma_b^2)}.$$

2.2.3 Two Component Heteroscedastic Normal Mixture

Consider an RCA(1) process with two-component normal mixture innovation terms given by $y_t = (\phi + b_t)y_{t-1} + e_t$, where $\{b_t\}$ and $\{e_t\}$ are independent two-component heteroscedastic normal mixtures. That is,

$f(b_t) = \pi_1\, \phi_1(b_t; 0, \sigma_1^2) + \pi_2\, \phi_2(b_t; 0, \sigma_2^2)$,

where $\pi_1 + \pi_2 = 1$ and

$$f(e_t) = \pi_1^*\, \phi_3(e_t; 0, \sigma_3^2) + \pi_2^*\, \phi_4(e_t; 0, \sigma_4^2),$$

where $\pi_1^* + \pi_2^* = 1$ and $\phi_i, i = 1, 2, 3, 4$ are normal densities. Then,

(i) $E(e_t) = E(b_t) = E(e_t^3) = E(b_t^3) = 0$,

(ii) $Var(e_t) = E(e_t^2) = \pi_1 \sigma_1^2 + \pi_2 \sigma_2^2$ and $Var(b_t) = E(b_t^2) = \pi_1^* \sigma_3^2 + \pi_2^* \sigma_4^2$,

(iii) $E(b_t^4) = 3(\pi_1 \sigma_1^4 + \pi_2 \sigma_2^4)$ and $E(e_t^4) = 3(\pi_1^* \sigma_3^4 + \pi_2^* \sigma_4^4)$.

Lemma 2.2.2 *Let $\{y_t\}$ be a two-component heteroscedastic normal mixture model,* *then the moments of the mixture driven RCA model, are given by*

$E(y_t) = 0,$

$$\begin{aligned} Var(y_t) &= \sigma_y^2 = \frac{\pi_1^*\sigma_3^2 + \pi_2^*\sigma_4^2}{1 - \phi^2 - \pi_1\sigma_1^2 - \pi_2\sigma_2^2}, && (2.2.10) \\ E(y_t^4) &= E[(\phi + b_t)^4 y_{t-1}^4 + 4(\phi + b_t)^3 y_{t-1}^3 e_t + 6(\phi + b_t)^2 y_{t-1}^2 e_t^2 + 4(\phi + b_t) y_{t-1} e_t^3 + e_t^4], && (2.2.11) \end{aligned}$$

and the kurtosis is given by

$$K^{(y)} = \frac{E(y_t)^4}{(Var(y_t))^2}.$$

Proof: $E(y_t) = EE(y_t|y_{t-1}) = 0$. Now, the variance of the process is given by

$$\begin{aligned} Var(y_t) &= VE(y_t|y_{t-1}) + EV(y_t|y_{t-1}) \\ &= \phi^2 V(y_{t-1}) + V(e_t) + V(y_{t-1})(\pi_1\sigma_1^2 + \pi_2\sigma_2^2) \\ &= \frac{\pi_1^*\sigma_3^2 + \pi_2^*\sigma_4^2}{1 - \phi^2 - \pi_1\sigma_1^2 - \pi_2\sigma_2^2}, \end{aligned}$$

since $E(y_t) = 0$ and $Var(y_t) = E(y_{t-1}^2)$. We now outline the proof for the kurtosis $K^{(y)}$of the process y_t. Note that $E(y_t) = 0$. We obtain the following result:

$$\begin{aligned} y_t^4 &= [(\phi + b_t)^4 y_{t-1}^4 + 4(\phi + b_t)^3 y_{t-1}^3 e_t + 6(\phi + b_t)^2 y_{t-1}^2 e_t^2 + 4(\phi + b_t) y_{t-1} e_t^3 + e_t^4] \\ E(e_t^4) &= 3(\pi_1^*\sigma_3^4 + \pi_2^*\sigma_4^4). \end{aligned}$$

We know that $E(4(\phi + b_t)y_{t-1}e_t^3) = 0$ since (b_t) and (e_t) are independent and are assumed to be Gaussian. Hence,

$$\begin{aligned} E\left[6(\phi + b_t)^2 y_{t-1}^2 e_t^2\right] &= 6\phi^2 V(y_{t-1})(\pi_1^*\sigma_3^2 + \pi_2^*\sigma_4^2) + 6V(y_{t-1})(\pi_1\sigma_1^2 + \pi_2\sigma_2^2)(\pi_1^*\sigma_3^2 + \pi_2^*\sigma_4^2) \\ &= 6V(y_{t-1})(\pi_1^*\sigma_3^2 + \pi_2^*\sigma_4^2)\left[\phi^2 + \pi_1\sigma_1^2 + \pi_2\sigma_2^2\right]. \end{aligned}$$

We know that $E\left[4(\phi+b_t)^3 y_{t-1}^3 e_t\right] = 0$ since e_t and b_t are independent, zero-mean Gaussian processes. Now

$$\begin{aligned}
E\left[(\phi+b_t)^4 y_{t-1}^4\right] &= EE\left[(\phi+b_t)^4 y_{t-1}^4 | y_{t-1}\right] \\
&= E(y_{t-1}^4) \times E(\phi+b_t)^4 \\
&= E(y_{t-1}^4)\left[\phi^4 + 6\phi^2(\pi_1\sigma_1^2 + \pi_2\sigma_2^2) + 3(\pi_1\sigma_1^4 + \pi_2\sigma_2^4)\right].
\end{aligned}$$

Assume $E(y_t^4) = E(y_{t-1}^4)$ and that (y_t) is stationary. Then we derive

$$\begin{aligned}
E(y_t^4) &= \frac{1}{1-\phi^4 - 6\phi^2(\pi_1\sigma_1^2+\pi_2\sigma_2^2) - 3(\pi_1\sigma_1^4+\pi_2\sigma_2^4)} \\
&\times \ [3(\pi_1^*\sigma_3^4 + \pi_2^*\sigma_4^4) + 6V(y_{t-1})(\pi_1^*\sigma_3^2 + \pi_2^*\sigma_4^2)(\phi^2 + \pi_1\sigma_1^2 + \pi_2\sigma_2^2)].
\end{aligned}$$

and finally the kurtosis is given by using (2.2.10) and (2.2.11).

2.3 Stationary processes with GARCH errors

This section is comprised of two subsections. The class of GARCH models discussed previously has been extended to a class of volatility models where the conditional mean, as well as the conditional variance is changing over time. The class of RCA models have also been extended to include models with time-varying errors such as the GARCH model.

2.3.1 RCA models with GARCH errors

Theorem 2.3.1 *Let $\{y_t\}$ be an RCA(1) time series satisfying conditions (2.1.1) and (2.1.2). The RCA(1) model is given by*

$$y_t = (\phi + b_t)\, y_{t-1} + \varepsilon_{t-1}^2 \varepsilon_t \tag{2.3.1}$$

$e_t \sim N(0, \sigma_e^2)$, $b_t \sim N(0, \sigma_b^2)$. Then we have the following:

(a) $E(y_t) = 0, \qquad E\left(y_t^2\right) = \left[\dfrac{3\sigma_\varepsilon^6}{(1 \ - (\phi^2 + \sigma_b^2))}\right],$

(b) $K^{(y)} = \left(\dfrac{(35 - 29\,(\sigma_b^2 + \phi^2))\,(1 \ - (\phi^2 + \sigma_b^2))}{(1 - 3\sigma_b^2(2\phi^2 + \sigma_b^2) - \phi^4)}\right).$

Proof:

$$\begin{aligned} E\left(y_t^2\right) &= E\left(y_{t-1}^2\phi^2\right) + E\left(y_{t-1}^2 b_t^2\right) + E\left(\varepsilon_{t-1}^4\right) E\left(\varepsilon_t^2\right) \\ &= \phi^2 E\left(y_{t-1}^2\right) + \sigma_b^2 E\left(y_{t-1}^2\right) + 3\sigma_\varepsilon^6 \end{aligned}$$

Thus, assuming y_t is stationary we have

$$E\left(y_t^2\right) = \left[\frac{3\sigma_\varepsilon^6}{(1 \ - (\phi^2 + \sigma_b^2))}\right] \tag{2.3.2}$$

$$\begin{aligned} E(y_t^4) &= 6\phi^2\sigma_b^2 E\left(y_{t-1}^4\right) + 315\sigma_\varepsilon^{12} + 3\sigma_b^4 E\left(y_{t-1}^4\right) + \phi^4 E\left(y_{t-1}^4\right) \\ &\quad + 18\sigma_b^2\sigma_\varepsilon^6 E\left(y_{t-1}^2\right) + 18\sigma_\varepsilon^6\phi^2 E\left(y_{t-1}^2\right) \end{aligned}$$

and we have

$$E(y_t^4) = \left(\frac{9\sigma_\varepsilon^{12}}{(1 - 6\phi^2\sigma_b^2 - 3\sigma_b^4 - \phi^4)}\right)\left(\frac{35 - 29\,(\sigma_b^2 + \phi^2)}{1 - (\sigma_b^2 + \phi^2)}\right)$$

$$\begin{aligned} K^{(y)} &= \frac{E\left[y_t^4\right]}{E\left[y_t^2\right]^2} = \left(\frac{9\sigma_\varepsilon^{12}}{(1 - 6\phi^2\sigma_b^2 - 3\sigma_b^4 - \phi^4)}\right)\left(\frac{35 - 29\,(\sigma_b^2 + \phi^2)}{1 - (\sigma_b^2 + \phi^2)}\right)\left(\frac{(1 \ - (\phi^2 + \sigma_b^2))^2}{9\sigma_\varepsilon^{12}}\right) \\ &= \left(\frac{(35 - 29\,(\sigma_b^2 + \phi^2))\,(1 \ - (\phi^2 + \sigma_b^2))}{(1 - 3\sigma_b^2(2\phi^2 + \sigma_b^2) - \phi^4)}\right) \end{aligned} \tag{2.3.3}$$

When $\sigma_b^2 = 0$, the kurtosis of the process $y_t$converge to the one reported by Ghahramani and Thavaneswaran(2007), $K^{(y)} = \dfrac{35 - 29\phi^2}{(1 + \phi^2)}$ and when $\sigma_b^2 = 0$, and $\phi = 0$ the kurtosis of the process $y_t$converge to 35, this result was reported in Appadoo, Ghahramani and Thavenswaran (2005). We have considered a time series model driven by nonlinear functions of random processes. Such time series are said to have nonlinear structure. Many financial log return time series possess the property where the squared log return process has autocorrelations that vanish only very slowly, or do not even vanish at all, in the limit. This simple model has the ability to model high volatility in financial time series model.

Theorem 2.3.2 *Suppose $\{y_t\}$ is an RCA model with GARCH(1, 1) errors of the form*

$$\begin{aligned} y_t &= (\phi + b_t)\, y_{t-1} + \varepsilon_t \\ \varepsilon_t &= \sqrt{h_t} Z_t \qquad (2.3.4) \\ h_t &= \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \qquad (2.3.5) \end{aligned}$$

where $b_t \sim NID(0, \sigma_b^2)$(uncorrelated). Then, we have the following:

(i) $E(y_t) = 0$,

(ii) $E\left(y_t^2\right) = \dfrac{\sigma_\varepsilon^2}{(1 - \phi^2 - \sigma_b^2)}$

(iii) $E\left(y_t^4\right) = \dfrac{6\sigma_\varepsilon^4(\phi^2 + \sigma_b^2) + K^{(\varepsilon)}\sigma_\varepsilon^4(1 - (\phi^2 + \sigma_b^2))}{[1 - (\phi^2 + \sigma_b^2)][1 - (\phi^4 + 6\phi^2\sigma_b^2 + 3\sigma_b^2)]}$,

(iv) When $Z_t \sim N(0,1)$ (uncorrelated), then

$$K^{(y)} = \left[\frac{3(1 - \phi^2 - \sigma_b^2)[2\,(\sigma_b^2 + \phi^2)\,(1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2) + (1 - \phi^2 - \sigma_b^2)(1 - (\alpha_1 + \beta_1)^2)]}{(1 - 6\phi^2\sigma_b^2 - \phi^4 - 3\sigma_b^4)(1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2)}\right].$$

Proof:

Part (ii) follows from observing that $E(y_t) = (\phi^2 + \sigma_b^2)E(y_{t-1}^2) + \sigma_\varepsilon^2$, and by assuming that y_t is stationary.

Part (iii) follows from observing that

$$E(y_t^4) = E((\phi + b_t)^4 y_{t-1}^4) + E(6(\phi + b_t)^2 y_{t-1}^2 \varepsilon_t^2) + \varepsilon_t^4),$$

and from the fact that $E(\varepsilon_t^4) = K^{(\varepsilon)}\sigma_\varepsilon^4$.

In particular, where $Z_t \sim N(0,1)$, then $K^{(\varepsilon)} = \dfrac{3[1-(\alpha_1+\beta_1)^2]}{1-(\alpha_1+\beta_1)^2-2\alpha_1^2}$ and part (iii) follows.

When $\phi = 0$ and $\sigma_b^2 = 0$ in Part (iii),

$$K^{(y)} = \frac{3[1-(\alpha_1+\beta_1)^2]}{(1-(\alpha_1+\beta_1)^2)-2\alpha_1^2} \tag{2.3.6}$$

Note: The result given by (2.3.6) has been observed in Thavaneswaran et al.(2005).

When $\sigma_b^2 = 0$ in $K^{(y)}$ of part (iii),

$$K^{(y)} = \left[\frac{6\phi^2}{(1+\phi^2)}\right] + 3\left(\frac{1-\phi^2}{1+\phi^2}\right)\frac{[1-(\alpha_1+\beta_1)^2]}{(1-(\alpha_1+\beta_1)^2)-2\alpha_1^2} \tag{2.3.7}$$

The result given by (2.3.7) has been observed in Ghahramani and Thavaneswaran(2007).

Next, special cases of Theorem 2.3.2 are discussed.

2.3.2 Special Cases of Theorem 2.3.2:

Example 2.3.1 *Let $\{y_t\}$ be a Sign RCA-GARCH(1, 1) time series satisfying conditions (2.1.1) and (2.1.2) given by*

$$y_t = (\phi + b_t + \Phi s_t)\, y_{t-1} + \varepsilon_t \tag{2.3.8}$$

where

$$\varepsilon_t = \sqrt{h_t} Z_t, \tag{2.3.9}$$

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \tag{2.3.10}$$

where Z_t and b_t are sequences of independently, identically Gaussian distributed random variables with zero mean, variance given by σ_Z^2 and σ_b^2 respectively,

$$s_t = \begin{cases} +1 & \text{if } y_t > 0 \\ 0 & \text{if } y_t = 0 \\ -1 & \text{if } y_t < 0 \end{cases}$$

ω, α_1, β_1 and Φ are real parameters, satisfying the following conditions, $\omega > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$. $|\Phi| \leq \omega$. Note: $E(s_t^2) = 1$, and in order to calculate the kurtosis, we observe that $E(s_t^4) = 1$. Then, we have the following moment properties

(a) $E(y_t) = 0$,

(b) $E\left(y_t^2\right) = \dfrac{\omega\sigma_Z^2}{\left[1 - (\phi^2 + \sigma_b^2 + \Phi^2)\right]\left[1 - (\alpha_1 + \beta_1)\right]}$,

(c) If $\{b_t\}$ and $\{\varepsilon_t\}$ are normally distributed random variables then the kurtosis $K^{(y)}$ of the process $\{y_t\}$ is given by,

$$\begin{aligned} K^{(y)} &= \frac{3(1-(\alpha_1\sigma_Z^2+\beta_1)^2)[1-(\phi^2+\sigma_b^2+\Phi^2)]^2}{(1-3\alpha_1^2\sigma_Z^4-2\alpha_1\beta_1\sigma_Z^2-\beta_1^2)(1-6(\phi^2\Phi^2+\Phi^2\sigma_b^2+\phi^2\sigma_b^2)-\Phi^4-\phi^4-3\sigma_b^4)} \\ &\quad + \frac{6\left(\Phi^2+\phi^2+\sigma_b^2\right)\left[1-(\phi^2+\sigma_b^2+\Phi^2)\right]}{(1-6(\phi^2\Phi^2+\Phi^2\sigma_b^2+\phi^2\sigma_b^2)-\Phi^4-\phi^4-3\sigma_b^4)}. \end{aligned}$$

Some details are as follows. $\sigma_\varepsilon^2 = E(\varepsilon_t^2) = \dfrac{\omega\sigma_Z^2}{1-\alpha_1-\beta_1}$.

$$E(y_t^2) = E((\phi^2 + b_t^2 + \Phi^2 s_t^2 + 2\phi b_t + 2\phi\Phi s_t + 2\Phi b_t s_t)y_{t-1}^2) + \frac{\omega\sigma_z^2}{1-\alpha_1-\beta_1}.$$

$$\begin{aligned} E(y_t^4) &= E\left\{(\phi + b_t + \Phi s_t)^4 y_{t-1}^4 + 6(\phi + b_t + \Phi s_t)^2 y_{t-1}^2 \varepsilon_t^2 + \varepsilon_t^4\right\} \\ &= \frac{6(\phi^2+\sigma_b^2+\Phi^2)\sigma_\varepsilon^2 E(y_t^2) + K^{(\varepsilon)}\sigma_\varepsilon^4}{1-(\phi^4+6\phi^2\sigma_b^2+3\sigma_b^4+6\Phi^2(\phi^2+\sigma_b^2)+\Phi^4)}. \end{aligned}$$

And,

$$K^{(y)} = \frac{3[(1-(\alpha_1\sigma_Z^2+\beta_1)^2]}{1-(\sigma_Z^2\alpha_1+\beta_1)^2-\alpha_1\sigma_Z^2}.$$

Next, a class of RCA Sign-GARCH Volatility Models is studied.

Example 2.3.2 *Consider the GARCH (1, 1) process for the time series y_t, where*

$$\begin{aligned} y_t &= (\phi + b_t)y_{t-1} + \varepsilon_t \\ \varepsilon_t &= \sqrt{h_t}Z_t \\ h_t &= \omega + (\alpha_1 + a_{t-1} + \Phi s_{t-1})\,\varepsilon_{t-1}^2 + \beta_1 h_{t-1} \end{aligned}$$

where $Z_t \sim N(0, \sigma_Z^2)$, $b_t \sim N(0, \sigma_b^2)$ *and* $a_t \sim N(0, \sigma_a^2)$.

$$s_t = \begin{cases} +1 & if\ \ y_t > 0 \\ 0 & if\ \ y_t = 0 \\ -1 & if\ \ y_t < 0 \end{cases}$$

$$E\left(h_t\right) = \frac{\omega}{[1-(\alpha_1\sigma_Z^2+\beta_1)]} \quad since \quad E(s_t)=0 \tag{2.3.11}$$

$$E\left(h_t^2\right) = \frac{\omega^2\left(\sigma_Z^2\alpha_1+\beta_1+1\right)}{\left[1-\beta_1\left(2\sigma_Z^2\alpha_1+\beta_1\right)-3\sigma_Z^4(\Phi^2+\alpha_1^2+\sigma_a^2)\right]\left[1-\sigma_Z^2\alpha_1-\beta_1\right]} \tag{2.3.12}$$

$$Var(y_t) = \frac{\omega\sigma_Z^2}{[1\ -(\phi^2+\sigma_b^2)][1-(\alpha_1\sigma_Z^2+\beta_1)]} \tag{2.3.13}$$

$$E\left[y_t^4\right] = \frac{3\sigma_Z^4}{(1-3\sigma_b^4-\phi^4-6\phi^2\sigma_b^2)}E[h_t^2] + \frac{6\sigma_Z^4\left(\sigma_b^2+\phi^2\right)}{\left(1-3\sigma_b^4-\phi^4-6\phi^2\sigma_b^2\right)\left[1\ -(\phi^2+\sigma_b^2)\right]}\left(E\left[h_t\right]\right)^2 \tag{2.3.14}$$

The kurtosis of the process is given by

$$K^{(y)} = \frac{3[1\ -(\phi^2+\sigma_b^2)]^2[1-(\alpha_1\sigma_Z^2+\beta_1)^2]}{\left[1-\beta_1\left(2\sigma_Z^2\alpha_1+\beta_1\right)-3\sigma_Z^4(\Phi^2+\alpha_1^2+\sigma_a^2)\right]\left(1-3\sigma_b^4-\phi^4-6\phi^2\sigma_b^2\right)} + \frac{6\left(\sigma_b^2+\phi^2\right)\left[1\ -(\phi^2+\sigma_b^2)\right]}{\left(1-3\sigma_b^4-\phi^4-6\phi^2\sigma_b^2\right)} \tag{2.3.15}$$

The proof of (2.3.15) parallel the proof of Theorem 2.3.2.

Note that when $\phi = 0$, $\sigma_b = 0$, $\sigma_a = 0$, and $\sigma_Z = 1$, the kurtosis of the process converge to the one reported by Thavaneswaran et al.(2005) as follows.

$$K^{(y)} = \frac{3[1-(\alpha_1+\beta_1)^2]}{(1-(\alpha_1+\beta_1)^2)-2\alpha_1^2} > 3. \tag{2.3.16}$$

2.3.3 ARMA models with GARCH errors

For certain financial return series, the sample ACF of both the series and its square exhibit significant autocorrelation. In such cases, the time-varying nature of the conditional mean as well as time-varying conditional variance suggests that fitting ARMA models with GARCH errors, for example, would be more appropriate than fitting zero-mean GARCH models. In this section, we study identification methods for stationary processes with GARCH errors.

Lemma 2.3.1 *For a volatility process of the form*

$$y_t - \mu \quad = \quad \phi\left(y_{t-1} - \mu\right) + \varepsilon_{t-1}^2 \varepsilon_t \tag{2.3.17}$$

under the stationarity assumptions that $|\phi| < 1$, *and the* ε_t *symmetric i.i.d with mean* 0, *variance* σ_ε^2 *and finite eighth moments, then*

(a) $E\left(y_t - \mu\right)^2 = \dfrac{E\left(\varepsilon_{t-1}^4\right) E(\varepsilon_t^2)}{(1-\phi^2)}$,

(b) $K^{(y)} = \dfrac{E\left[(y_t-\mu)^4\right]}{Var(y_t)^2} = \left[\dfrac{6\phi^2\left(E\left(\varepsilon_{t-1}^4\right)E(\varepsilon_t^2)\right)^2 + E(\varepsilon_{t-1}^8)E(\varepsilon_t^4)(1-\phi^2)}{(1+\phi^2)\left(E\left(\varepsilon_{t-1}^4\right)E(\varepsilon_t^2)\right)^2}\right]$,

(c) If we assume that $\varepsilon_t \sim i.i.d.N(0, \sigma_\varepsilon^2)$, *then* $E[\varepsilon_t^{2n}] = \dfrac{(2n)!}{2^n(n!)}\sigma_\varepsilon^{2n}$ *and hence* $K^{(y)} = \dfrac{35 - 29\phi^2}{1+\phi^2}$.

Proof: $y_t - \mu = \sum_{j=0}^{\infty}\phi^j a_{t-j}$ where $a_t = \varepsilon_{t-1}^2\varepsilon_t$. Then, $E\left(y_t-\mu\right)^2 = \sigma_a^2 \sum_{j=0}^{\infty}\phi^{2j} = \dfrac{\sigma_a^2}{1-\phi^2}$, where $\sigma_a^2 = E(\varepsilon_{t-1}^4)E(\varepsilon_t^2)$.

$$E\left(y_t-\mu\right)^4 \quad = \quad \phi^4 E\left(y_t-\mu\right)^4 + 6\phi^2 E\left(y_t-\mu\right)^2 E(\varepsilon_{t-1}^4)E(\varepsilon_t^2) + E(\varepsilon_{t-1}^8)E(\varepsilon_t^4)$$

Then,

$$E\left(y_t-\mu\right)^4 \quad = \quad \frac{6\phi^2\left(E\left(\varepsilon_{t-1}^4\right)E(\varepsilon_t^2)\right)^2 + E(\varepsilon_{t-1}^8)E(\varepsilon_t^4)(1-\phi^2)}{(1-\phi^2)(1-\phi^4)}$$

$$K^{(y)} \quad = \quad \frac{E\left[(y_t-\mu)^4\right]}{Var(y_t)^2} = \left[\frac{6\phi^2\left(E\left(\varepsilon_{t-1}^4\right)E(\varepsilon_t^2)\right)^2 + E(\varepsilon_{t-1}^8)E(\varepsilon_t^4)(1-\phi^2)}{(1+\phi^2)\left(E\left(\varepsilon_{t-1}^4\right)E(\varepsilon_t^2)\right)^2}\right].$$

Part (c) follows from the fact that when, $Z_t \sim N(0,1)$, then $E[\varepsilon_t^{2n}] = \frac{(2n)!}{2^n(n!)}\sigma_\varepsilon^{2n}$. So that, $E(\varepsilon_t^4) = 3\sigma_\varepsilon^4$ and $E(\varepsilon_t^8) = 105\sigma_\varepsilon^8$.

Note that when $\phi = 0$ in (2.3.17), the kurtosis of the process y_t turns out to be 35 as observed in Appadoo, Ghahramani and Thavaneswaran (2005).

The following lemma proved Ghahramani and Thavaneswaran (2007), provides an identification method for the kurtosis of a weakly stationary process through the kurtosis of the error term and the ψ-weight representation of the $\{y_t\}$ process. Part (i) of the following lemma is a well-known result for the ψ-weight representation of stationary series.

Lemma 2.3.2 *Suppose $\{y_t\}$ is a second order linear stationary process having a MA(moving average) representation of the form*

$$y_t - \mu = \sum_{j=0}^{\infty} \psi_j a_{t-j}, \tag{2.3.18}$$

where a_t is an uncorrelated noise process with zero mean and with variance σ_a^2.

Then the variance and the kurtosis of y_t are

(i) $Var(y_t) = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j^2$ and

(ii) $$K^{(y)} = \frac{K^{(a)}\left[\sum_{j=0}^{\infty} \psi_j^4\right] + 6\sum_{i<j}^{\infty} \psi_i^2\psi_j^2}{\left(\sum_{j=0}^{\infty} \psi_j^2\right)^2}$$

where $K^{(a)}$ is the kurtosis of a_t and $\sum_{j=0}^{\infty} \psi_j^4 < \infty$.

Proof: Part(i) follows from the fact that the a_t are uncorrelated.

Part(ii) follows from the fact that

$$E(y_t-\mu)^4 = E\left(\sum_{j=0}^{\infty}\psi_j a_{t-j}\right)^4 = E(a_t^4)\sum_{j=0}^{\infty}\psi_j^4+6\sigma_a^4\sum_{i<j}^{\infty}\psi_i^2\psi_j^2+4E\left[\sum_{j=0}^{\infty}\psi_j^3 a_{t-j}+\sum_{j=0}^{\infty}\psi_j a_{t-j}^3\right].$$

$$K^{(y)} = \frac{E(y_t-\mu)^4}{\{E(y_t-\mu)^2\}^2} = \frac{E(a_t^4)\sum_{j=0}^{\infty}\psi_j^4+6\sigma_a^4\sum_{i<j}^{\infty}\psi_i^2\psi_j^2}{\left(\sigma_a^2\sum_{j=0}^{\infty}\psi_j^2\right)^2} \tag{2.3.19}$$

The following theorem due to Ghahramani and Thavaneswaran (2007), extends the result of Thavaneswaran et al. (2005) for a mean zero GARCH process to a stationary ARMA(p, q) process with GARCH(P, Q) errors.

Theorem 2.3.3 *For any second order stationary process $\{y_t\}$ with GARCH (P, Q) errors given by*

$$y_t - \mu = \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j} \tag{2.3.20}$$

$$\varepsilon_t = \sqrt{h_t}Z_t \tag{2.3.21}$$

$$h_t = \omega+\sum_{i=1}^{P}\alpha_i\varepsilon_{t-i}^2+\sum_{j=1}^{Q}\beta_j h_{t-j} \tag{2.3.22}$$

Under suitable stationarity conditions, kurtosis is given by

$$K^{(y)} = \frac{K^{(\varepsilon)}\left[\sum_{j=0}^{\infty}\psi_j^4\right]+6\sum_{i<j}^{\infty}\psi_i^2\psi_j^2}{\left(\sum_{j=0}^{\infty}\psi_j^2\right)^2} \tag{2.3.23}$$

where $K^{(\varepsilon)}$ is given by $K^{(\varepsilon)} = \dfrac{E(Z_t^4)}{E(Z_t^4)-[E(Z_t^4)-1]\sum_{j=0}^{\infty}\Psi_j^2}$.

Proof: The Proof follows from Lemma 2.3.2 and the fact that a GARCH process is an uncorrelated process.

Note: For an ARMA(p, q) process, when $p = 0$, $q = 0$, the theorem provides the kurtosis of a GARCH(P, Q) process. Moreover, when we set $P = 0$, and $Q = 0$, the theorem provides the kurtosis of an ARMA(p, q) process.

2.3.4 Special cases of stationary processes with GARCH errors

In this subsection, we provide the moment properties of special cases of Theorem 2.3.3 that have been observed by Ghahramani and Thavaneswaran (2007).

Case 1: Consider an autoregressive model of the form:

$$y_t - \mu = \phi(y_{t-1} - \mu) + \varepsilon_t, \tag{2.3.24}$$

where $|\phi| < 1$, ε_t is a zero mean GARCH(P, Q) process given by expression (2.3.21), the conditional variance h_t is given by (2.3.22) and Z_t is an uncorrelated zero mean sequence with unit variance and finite fourth moment. Let σ_ε^2 be the variance of $\{\varepsilon_t\}$. Then $\psi_j = \phi^j$, $\sum_{j=0}^{\infty} \psi_j^2 = \frac{1}{1-\phi^2}$, $\sum_{j=0}^{\infty} \psi_j^4 = \frac{1}{1-\phi^4}$ and $\sum_{i<j}^{\infty} \psi_i^2 \psi_j^2 = \frac{\phi^2}{(1-\phi^2)(1-\phi^4)}$.
Hence, $\mathrm{Var}(y_t) = \frac{\sigma_\varepsilon^2}{1-\phi^2}$, and $K^{(y)} = \frac{6\phi^2 + (1-\phi^2)K^{(\varepsilon)}}{1+\phi^2}$, where $K^{(\varepsilon)} = \frac{E(Z_t^4)}{E(Z_t^4) - [E(Z_t^4) - 1]\sum_{j=0}^{\infty} \Psi_j^2}$.
If we assume that $Z_t \sim N(0, 1)$, and $P = 1$ and $Q = 1$ then the kurtosis of the process is given by

$$K^{(y)} = \frac{1}{1+\phi^2}\left[6\phi^2 + (1-\phi^2)\left(\frac{3(1-(\alpha_1+\beta_1)^2)}{(1-(\alpha_1+\beta_1)^2 - 2\alpha_1^2)}\right)\right]. \tag{2.3.25}$$

If we assume that Z_t has a t-distribution with v degrees of freedom, then

$$K^{(y)} = \frac{1}{(1+\phi^2)}\left[6\phi^2 + (1-\phi^2)\left(\frac{3(v-2)\left(1-(\alpha_1+\beta_1)^2\right)}{\left(1-2\alpha_1\beta_1-\beta_1^2\right)(v-4) + 3\alpha_1^2(v-2)}\right)\right]. \tag{2.3.26}$$

Case 2: Consider an ARMA(1, 1) model given by

$$y_t - \mu \;=\; \phi(y_{t-1} - \mu) + \varepsilon_t - \theta\varepsilon_{t-1}, \tag{2.3.27}$$

where $|\phi| < 1$, ε_t is given by expression (2.3.9), the conditional variance h_t is given by (2.3.10) and Z_t is an uncorrelated zero mean sequence with unit variance and finite fourth moment. Let σ_ε^2 be the variance of $\{\varepsilon_t\}$. Then, $\sum_{j=0}^{\infty} \psi_j^2 = \frac{1 - 2\phi\theta + \theta^2}{1 - \phi^2}$, $\sum_{j=0}^{\infty} \psi_j^4 = \frac{1 - 4\phi^3\theta + 6\phi^2\theta^2 - 4\phi\theta^3 + \theta^4}{1 - \phi^4}$, and $\sum_{i<j}^{\infty} \psi_i^2\psi_j^2 = \frac{(\phi - \theta)^2}{1 - \phi^2} + \frac{(\phi - \theta)^4}{1 - \phi^2}\frac{\phi^2}{1 - \phi^4}$.
Hence,

$$Var(y_t) = \frac{\omega(1 - 2\phi\theta + \theta^2)}{(1 - \sum_{i=1}^{R} \Phi_i)(1 - \phi^2)},$$

and

$$K^{(y)} = K^{(\varepsilon)} \left(\frac{(1 - 4\phi^3\theta + 6\phi^2\theta^2 - 4\phi\theta^3 + \theta^4)(1 - \phi^2)}{(1 + \phi^2)(1 - 2\phi\theta + \theta^2)^2} \right) + 6\frac{(1 - \phi^4)(\phi - \theta)^2 + (\phi - \theta)^4\phi^2}{(1 + \phi^2)(1 - 2\phi\theta + \theta^2)^2},$$

where $\Phi_i = \alpha_i + \beta_i$, $i = 1, \ldots, R$, and $R = \max(P, Q)$.

2.4 Hidden Semimartingale Models(HSM) with GARCH errors

Following Hamilton (1989) who models changes in regimes via Markov switching, we consider the situation where the volatility is also changing. Building upon the ideas developed by Abraham and Thavaneswaran (1991) on state space formulation of non-linear models to draw optimal inference about the unknown parameters (see Granger

(1998) for details), we propose the estimating function method based inference for the proposed HSM of volatilities.

2.4.1 Hidden Semimartingale Models

Consider the random coefficient autoregressive(RCA) model given in Thavaneswaran and Abraham (1988):

$$y_t - \theta_t f(t, F_{t-1}^y) \quad = \quad e_t \tag{2.4.1}$$

where $\{\theta_t\}$ is a more general stochastic sequence and $f(t, F_{t-1}^y)$ is a function of the past. When θ_t is a moving average (MA) sequence of the form

$$\theta_t \quad = \quad \theta + \varepsilon_t + \varepsilon_{t-1} \tag{2.4.2}$$

where θ_t, e_t are square integrable independent random variables and $\{\varepsilon_t\}$ consists of zero mean square integrable *Gaussian random variables* independent of $\{e_t\}$. In this case $E(y_t|F_{t-1}^y)$ depends on the posterior mean $m_t = E(\varepsilon_t|F_t^y)$ and variance $\gamma_t = E[(\varepsilon_t - m_t)^2|F_t^y)$ of ε_t. Assume $y_0 = 0$, then m_t and y_t satisfy the following Kalman-like recursive algorithms,

$$m_t = \frac{\sigma_t^2 f\left(t, F_{t-1}^y\right)\left[y_t - (\theta + m_{t-1}) f\left(t, F_{t-1}^y\right)\right]}{\sigma_e^2 + f^2\left(t, F_{t-1}^y\right)\left(\sigma_e^2 + \gamma_{t-1}\right)} \tag{2.4.3}$$

and

$$\gamma_t = \sigma_\varepsilon^2 - \frac{f^2(t, F_{t-1}^y)\sigma_\varepsilon^4}{\sigma_e^2 + f^2\left(t, F_{t-1}^y\right)\left(\sigma_e^2 + \gamma_{t-1}^2\right)} \tag{2.4.4}$$

where $\gamma_0 = \sigma_\varepsilon^2$ and $m_0 = 0$. Hence, $E(y_t|F_{t-1}^y) = (\theta + m_{t-1}) f(t, F_{t-1}^y)$ and $E(h_t^2|F_t^y) = \sigma_e^2 + f^2(t, F_{t-1}^y)(\sigma_e^2 + \gamma_{t-1})$, where $h_t = y_t - E(y_t|F_{t-1}^y)$, can be calculated recursively. Then the optimal estimating function turns out to be $g_n^* = \sum_{t=2}^{n} h_t a_{t-1}^*$ where,

$$a_{t-1}^* = \frac{E\left[(dh_t/d\theta)\,|F_{t-1}^y\right]}{E\left[h_t^2|F_{t-1}^y\right]}.$$

Thus, the optimal estimate is given by

$$\widehat{\theta}_n = \frac{\sum_{t=2}^{n} a_{t-1}^* y_t}{\sum_{t=2}^{n} a_{t-1}^* f(t, F_{t-1}^y)}$$

where

$$a_{t-1}^* = \frac{f\left(t, F_{t-1}^y\right)\left(1 + (dm_{t-1}/d\theta)\right)}{\left[\sigma_\varepsilon^2 + f^2\left(t, F_{t-1}^y\right)\left(\sigma_\varepsilon^2 + \gamma_{t-1}\right)\right]}.$$

Since γ_t is independent of θ, the relation

$$\frac{dm_t}{d\theta} = -\frac{\left[\sigma_\varepsilon^2 f^2(t, F_{t-1}^y)(1 + dm_{t-1}/d\theta)\right]}{\left[\sigma_\varepsilon^2 + f^2\left(t, F_{t-1}^y\right)\left(\sigma_\varepsilon^2 + \gamma_{t-1}\right)\right]}$$

can be used to calculate this derivative recursively.

As can be seen from (2.4.3) and (2.4.4), the optimal estimate $\widehat{\theta}_n$ adopts a weighting scheme based on σ_ε^2 and σ_e^2. The superiority of the optimal estimate over the conditional least squares estimate has been demonstrated in Thavaneswaran and Abraham (1988). This shows that optimal inference for RCA models and HSM models can be studied using the estimating function method. Later we show that a GARCH(p, q) model for y_t could be written as an ARMA(r, q) for y_t^2 and hence one could study the optimal inference by combining estimating functions as in Thompson and Thavaneswaran (1999).

Hamilton (1989) used Markov switching to detect changes between positive and negative growth periods in the economy, using a Hidden Markov Model (HMM) of the form

$$y_t = n_t + \varepsilon_t, \tag{2.4.5}$$

$t = 1, \ldots, T$, where ε_t is a zero mean ARIMA(r,1,0) process and n_t (the trend term) is a random walk with drift that switches between two values α_0 and $\alpha_0 + \alpha_1$. That is,

$$n_t = n_{t-1} + \alpha_0 + \alpha_1 S_t, \tag{2.4.6}$$

$t = 1, \ldots, T$, depending on whether the unobserved state of the system S_t is in state 1 or state 0. Furthermore, transitions between the states of S_t is assumed to be governed by the Markov process with: $p_{ij} = Pr(S_{t+1} = j | S_t = i)$. Let the steady-state probabilities be denoted by $\pi = Pr(S_t = 1)$ and $1 - \pi = Pr(S_t = 0)$ respectively. The following lemma gives the corresponding semimartingale form of the binary process S_t.

Lemma 2.4.1 *Let S_t be defined as in (2.4.6). Then S_t has the following form:*

$$S_{t+1} - \pi = d(S_t - \pi) + V_{t+1}, \tag{2.4.7}$$

where $d = Corr(S_{t+1}, S_t) = p_{11} - p_{01}$ and V_t is a semimartingale with $E[V_{t+1}|S_t = i] = 0$ and $Var[V_{t+1}|S_t = i] = p_{ii}(1 - p_{ii})$, $i = 0, 1$.

Proof: Let $\underline{\pi}' = (1 - \pi \;\; \pi)'$ and let $\mathbf{P}$ denote the transition probability matrix. The steady-state probability $\pi = E(S_t) = Pr(S_t = 1)$ found by solving the Chapman-Kolmogorov equations $\underline{\pi}'\mathbf{P} = \underline{\pi}'$ is $\pi = \dfrac{p_{01}}{p_{01} + p_{10}}$. The variance of S_t is $Var(S_t) = E(S_t^2) - \pi^2 = \pi(1 - \pi)$.

The correlation between S_t and S_{t+1} follows by observing that $Cov(S_t, S_{t+1}) = E(S_t S_{t+1}) - \pi^2 = Pr(S_t = 1, S_{t+1} = 1) - \pi^2 = Pr(S_{t+1} = 1 | S_t = 1) Pr(S_t = 1) - \pi^2 = p_{11}\pi - \pi^2$.

Let $V_{t+1} = S_{t+1} - \pi - d(S_t - \pi)$ and let $i = 0$. Then, $E(V_{t+1}|S_t = 0) = E(S_{t+1}|S_t = 0) - \pi + d\pi = p_{01} - p_{01} = 0$.

$Var(V_{t+1}|S_t = 0) = Var(S_{t+1} - dS_t|S_t = 0) = Var(S_{t+1}|S_t = 0) + d^2Var(S_t|S_t = 0) - 2dCov(S_{t+1}, S_t|S_t = 0) = E(S_{t+1}^2|S_t = 0) - E^2(S_{t+1}|S_t = 0) = p_{01}(1 - p_{01}) = (1 - p_{00})p_{00}$. Using the same reasoning $E(V_{t+1}|S_t = 1) = 0$ and $Var(V_{t+1}|S_t = 1) = (1 - p_{11})p_{11}$.

It is of interest to note that for the usual continuous space AR(1) process $y_t - \mu = \phi(y_{t-1} - \mu) + a_t$, $Var(a_t|y_{t-1}) = \sigma_a^2$, that is, a constant.

2.4.2 HSMs with GARCH Errors

Suppose the observed time series is obtained from the process $y_t = n_t + \varepsilon_t$, where ε_t is a zero mean GARCH(p, q) process as in (2.3.4), (2.3.5), and n_t (the trend term) is a semimartingale as in (2.4.6). This model is of interest when the conditional variance of the series is changing and the trend term switches between two states such as contraction and expansion in the economy.

In order to calculate the variance and kurtosis for a HSM with GARCH errors in terms of the ψ weights, we have the following theorem.

Theorem 2.4.1 *For the HSM process for trend, specified by (2.4.6), (2.4.7), (2.2.1), and (2.2.2), under the assumptions of stationarity and finite fourth moment, the kurtosis $K^{(\varepsilon)}$ of the process is given by:*

(a) $K^{(\varepsilon)} = \dfrac{E(Z_t^4)}{E(Z_t^4) - [E(Z_t^4) - 1]\sum_{j=0}^{\infty}\psi_j^2}$,

(b) The variance of the ε_t^2 process is $\gamma_0^{\varepsilon^2} = \sum_{j=0}^{\infty}\psi_j^2\sigma_u^2$, *where* $\sigma_u^2 = \dfrac{\mu_{\varepsilon^2}^2(K^{(\varepsilon)} - 1)}{\sum_{j=0}^{\infty}\psi_j^2}$ *and*

$$\mu_{\varepsilon^2} = E(\varepsilon_t^2) = \frac{\omega}{1 - \Phi_1 - \Phi_2 - \cdots - \Phi_R},\ R = \max(\mathrm{P},\ \mathrm{Q}),$$

(c) Let μ_t^n and σ_n^2 denote the mean and variance of n_t conditional on n_0, respectively.

Then $\mu_t^n = n_0 + (\alpha_0 + \pi\alpha_1)t$ *and* $\sigma_n^2 = \alpha_1^2\pi(1-\pi)[t + 2\sum_{j=1}^{t-1}(t-j)d^j]$.

(d) When Z_t *is Gaussian,* $K^{(\varepsilon)}$ *and* $K^{(y)}$ *are greater than 3.*

Proof: Proof of part (a) somewhat parallels the proof of Theorem 2.1 in Thavaneswaran et al.(2005) for kurtosis. Part (b) follows from the ARMA representation of ε_t^2.

The proof of part (c) follows by using the fact that $E(S_t) = \pi$ and $Var(S_t) = \pi(1-\pi)$ in equation (2.4.6). Then using the fact that $n_t - n_0 = \alpha_0 t + \alpha_1 \sum_{k=1}^{t} S_k$ the results for mean and variance follow. We note that since S_t was shown to have an AR(1) representation, the process has autocorrelation function $\rho_k = d^k$ where $d = p_{11} - p_{01}$. Also, we use the fact that since S_t is stationary, $Var(\sum_{k=1}^{t} S_k) = \gamma_0(t + 2(t-1)\rho_1 + 2(t-2)\rho_2 + \ldots + 2\rho_{t-1})$, where $\gamma_0 = Var(S_t)$.

Note 1: $\sum_{j=1}^{t-1} d^j = \left(\frac{1-d^t}{1-d} - 1\right)$.

Note 2: $\sum_{j=1}^{t-1} j d^j = \left(\frac{-td^{t-1}(1-d) + (1-d^t)}{(1-d)^2}\right)$.

Part (d) follows from the fact that for any conditionally Gaussian GARCH process $\varepsilon_t | F_{t-1}^{\varepsilon}$, $E(\varepsilon_t^4) = EE(\varepsilon_t^4|\varepsilon_{t-1}) = E3[E(\varepsilon_t^2|\varepsilon_{t-1})]^2 = 3E[E(\varepsilon_t^2|\varepsilon_{t-1})]^2 \geq 3(E[E(\varepsilon_t^2|\varepsilon_{t-1})])^2 = 3(E(\varepsilon_t^2)^2)$.

In analogy with the RCA example, using the combination theorem given in Thompson and Thavaneswaran (1999), we can make inference about the model parameters.

2.5 Autoregressive Conditional Poisson(ACP) models

In this section, some results in Abraham and Thavaneswaran (1991) are extended to ACP models. The ACP model was introduced by Heinen(2003) to deal with issues of discreteness, overdispersion and autocorrelation of count data such as the daily number of price change durations of 0.75$ on IBM stock. A 0.75$ price- change duration is defined as the time it takes the stock price to move by at least 0.75$. The variable of interest is the daily number of such durations, which is a measure of intradaily volatility (see Heinen(2003) for details). The result of Heinen(2003) for ACP $(1, 1)$ is extended to ACP(p, q) models, by using a martingale transformation.

2.5.1 Moment properties of ACP models

Let N_t denote the time series and let F_t^N denote the σ-field generated by observations up to and including time t. The conditional distribution of N_t is assumed to follow a Poisson distribution,

$$N_t|F_{t-1}^N \sim \text{Poisson}(\mu_t) \tag{2.5.1}$$

with an autoregressive conditional intensity as in the conditional variance in the GARCH model of Bollerslev(1986) and as in the ACD model of Engle and Russell(1998):

$$E[N_t|F_{t-1}^N] = \mu_t = \omega + \sum_{i=1}^{p} \alpha_i N_{t-i} + \sum_{j=1}^{q} \beta_j \mu_{t-j}, \tag{2.5.2}$$

for positive α_j's, β_j's and ω. This model is known as the Autoregressive Conditional Poisson (ACP) model and using a martingale formulation we study its moment properties.

Thus, we see that both models specify the conditional distribution of the observed process given past values and model the conditional mean or conditional variance or possibly both in a time-varying manner. Let $u_t = N_t - E[N_t|F_{t-1}^N] = N_t - \mu_t$ be the martingale difference and let σ_u^2 be the variance of u_t. Then (2.5.1) and (2.5.2) could be written as:

$$\begin{aligned} N_t - u_t &= \omega + \sum_{i=1}^{p} \alpha_i N_{t-i} + \sum_{j=1}^{q} \beta_j (N_{t-j} - u_{t-j}), \\ \phi(B) N_t &= \omega + \beta(B) u_t \end{aligned} \qquad (2.5.3)$$

where, $\phi(B) = 1 - \sum_{i=1}^{r} \phi_i B^i$, $\phi_i = (\alpha_i + \beta_i)$, $\beta(B) = 1 - \sum_{i=1}^{q} \beta_i B^i$ and $r = \max(p, q)$. Under stationarity assumptions similar to Assumptions 2.1 and 2.2 for GARCH models, N_t has an ARMA(r, q) representation.

The $u_t's$ are uncorrelated with mean zero and finite variance and that the N_t process is weakly stationary. In this case, the autocorrelation function of N_t will be exactly the same as that for a stationary ARMA(r, q) model. In order to calculate the moments and autocorrelation function for an ACP process in terms of the ψ-weights, the following theorem will be useful.

Theorem 2.5.1 *For the ACP (p, q) process specified by (2.5.1) and (2.5.2), under the assumption of stationarity, the moments of the process, as well as the autocorrelation function and the variance of the l-steps ahead forecast error are given by the following:*

(a) $\mu = E(N_t) = \dfrac{\omega}{1 - \sum_{j=0}^{max(p,q)} (\alpha_j + \beta_j)} = \dfrac{\omega}{1 - \phi_1 - \ldots - \phi_r}$, *where* $r = max(p, q)$.

(b) The variance of the process is $\gamma_0^N = \sigma_u^2 \sum_{j=0}^{\infty} \psi_j^2 = \mu \sum_{j=0}^{\infty} \psi_j^2$, *where* $\mu = \dfrac{\omega}{1 - \phi_1 - \ldots - \phi_r}$.

(c) The k-lag autocorrelation of N_t is $\rho_k^N = \dfrac{\sum_{j=0}^{\infty} \psi_{j+k}\psi_j}{\sum_{j=0}^{\infty} \psi_j^2}$.

(d) Let $e_n(l) = N_{n+l} - E(N_{n+l}|F_n^N)$ *be the minimum mean square error l-steps ahead forecast error. Then the variance of* $e_n(l)$ *is given by* $Var(e_n(l)) = \mu \sum_{j=0}^{l-1} \psi_j^2$.

Proof: Part (a) follows from observing

$$\mathrm{E}(N_t) = \mathrm{E}\mathrm{E}(N_t|F_{t-1}^N) = \mathrm{E}(\mu_t) = \omega + \sum_{j=1}^{p} \alpha_j \mathrm{E}(N_{t-j}) + \sum_{j=1}^{q} \beta_j \mathrm{E}(\mu_{t-j}).$$

Assuming N_t is stationary with $E(\mu_t) = E(N_t) = \mu$, and solving for μ, the result follows.

Part(b) follows from the ARMA representation of N_t and by the properties of the martingale differences $u_t's$. Namely, $u_t = N_t - \mu_t$, where $E(u_t) = 0$, $\sigma_u^2 = E(u_t^2)$, $\mu_t = E(N_t|F_{t-1}^N)$ and, moreover, $\sigma_u^2 = E(N_t^2) - E(\mu_t^2)$, $Var(N_t) = E(\mu_t^2) - \mu^2 + \mu$. Thus, $E(N_t^2) = E(\mu_t^2) + \mu$. Finally, $\sigma_u^2 = E(\mu_t^2) + \mu - E(\mu_t^2) = \mu$ and hence, $Var(N_t) = \mu \sum_{j=0}^{\infty} \psi_j^2$.

Part (c) is due to the moving average(MA) representation of N_t. Part(d) follows from the fact that for a stationary ARMA process with error variance σ_u^2, $Var(e_n(l)) = \sigma_u^2(1 + \psi_1^2 + \ldots + \psi_{l-1}^2)$ and since $\sigma_u^2 = \mu$.

It is of interest to note that part (a) of Theorem 2.5.1 was previously observed by Heinen (2003). Parts (b) - (d) are however, new results. Heinen (2003) had previously observed the special cases of parts (b) and (c) demonstrated in the following example.

Example 2.5.1 *In particular, for an ACP(1,1) model,*

(i) the unconditional mean is given by $E(N_t) = \mu = \dfrac{\omega}{1 - (\alpha_1 + \beta_1)}$,

(ii) the unconditional variance is equal to

$$Var(N_t) = \sigma^2 = \frac{\mu(1 - (\alpha_1 + \beta_1)^2 + \alpha_1^2)}{1 - (\alpha_1 + \beta_1)^2} \geq \mu,$$

(iii) the unconditional autocorrelation of the ACP(1,1) model is given by

$$\rho_k^N = (\alpha_1 + \beta_1)^{k-1} \frac{\alpha_1(1 - \beta_1(\alpha_1 + \beta_1))}{1 - (\alpha_1 + \beta_1)^2 + \alpha_1^2}, \quad k \geq 1.$$

This also holds for all models with mean equation given by (2.5.2), such that $\frac{\mu_t}{N_t} = \frac{\alpha_1^2}{1 - (\alpha_1 + \beta_1)^2 + \alpha_1^2}$.

(iv) The variance of the l-steps ahead forecast error is given by

$$Var(e_n(l)) = \mu \left(1 + \alpha_1^2 \frac{1 - (\alpha_1 + \beta_1)^{2(l-1)}}{1 - (\alpha_1 + \beta_1)^2}\right).$$

for $l \geq 1$. *The mean, variance and autocorrelation function turn out to be the same ones given by Heinen(2003).*

The martingale transformation allows us to extend the result on the autocorrelation structure without calculating covariances as in Heinen(2003).

2.5.2 Doubly Autoregressive Poisson (DACP) Models

Heinen (2003) extends the ACP model specified by (2.5.1) and (2.5.2) to allow for conditional over-dispersion using the Double Poisson (DP) distribution (See Efron (1986) for more details). That is, conditional on the past,

$$N_t | F_{t-1}^N \sim DP(\mu_t, \gamma), \tag{2.5.4}$$

and μ_t is specified as in (2.5.2). Efron (1986) shows that the mean of the Double Poisson is μ_t and that the variance is approximately equal to $\frac{\mu_t}{\gamma}$. Let $u_t = N_t -$

$E(N_t|F_{t-1}^N)$ be the martingale difference. Following assumptions similar to those made for the ACP model in Section 2.5, it is easy to show that the DACP model has an ARMA representation. Using the ARMA representation, the following theorem allows us to calculate the mean and variance of the DACP(p, q) model in terms of $\psi-$weights.

Theorem 2.5.2 *For the DACP (p, q) process specified by (2.5.4) and (2.5.2), under the assumption of stationarity, as well as the autocorrelation function and the variance of the l-steps ahead forecast error are given by the following:*

(a) $\mu = E(N_t) = \dfrac{\omega}{1 - \sum_{j=0}^{max(p,q)}(\alpha_j + \beta_j)} = \dfrac{\omega}{1 - \phi_1 - \ldots - \phi_r}$, *where* $r = max(p, q)$.

(b) The variance of the process is $\gamma_0^N = \sigma_u^2 \sum_{j=0}^{\infty} \psi_j^2 = \dfrac{\mu}{\gamma} \sum_{j=0}^{\infty} \psi_j^2$, *where* $\mu = \dfrac{\omega}{1 - \phi_1 - \ldots - \phi_r}$.

(c) The k-lag autocorrelation of N_t *is* $\rho_k^N = \dfrac{\sum_{j=0}^{\infty} \psi_{j+k}\psi_j}{\sum_{j=0}^{\infty} \psi_j^2}$.

(d) Let $e_n(l) = N_{n+l} - E(N_{n+l}|F_n^N)$ *be the minimum mean square error l-steps ahead forecast error. Then the variance of* $e_n(l)$ *is given by* $Var(e_n(l)) = \dfrac{\mu}{\gamma} \sum_{j=0}^{l-1} \psi_j^2$.

The proof is similar to the proof of Theorem 2.5.1.

In the next section, volatility model identification using the kurtosis and the autocorrelation structure of the squared process are studied in simulation studies.

2.6 Simulation Studies

In this section, the moment properties of the following volatility models are examined: normal-GARCH, t-distribution GARCH, and two simple volatility models discussed in Appadoo, Ghahramani and Thavaneswaran (2005). The comparison will be made on the basis of simulating from each process, fitting the corresponding model and comparing the estimated moments with the empirical moments.

Case 1:(GARCH(1, 1)) In this example, we consider the process

$$
\begin{aligned}
y_t &= \sqrt{h_t}\, Z_t, \\
h_t &= \omega + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1},
\end{aligned}
$$

where $Z_t \sim NID(0,1)$ or Z_t follows a standardized t-distribution with 5 degrees of freedom. The model parameters are $\omega = 5.3 \times 10^{-7}$, $\alpha_1 = 0.0057$ and $\beta_1 = 0.9381$, and we generated GARCH(1, 1) processes of size $n = 1000, 5000, 10,000$, respectively. Each simulation was replicated five hundred times. For the normal-GARCH(1, 1) model, the empirical kurtosis and the estimated model based kurtosis are reported in Table 2.5. For the t_5-GARCH(1, 1) model, the empirical kurtosis and the estimated model based kurtosis are reported in Table 2.6. For these GARCH(1, 1) models, the model based kurtosis explains the empirical kurtosis. Furthermore, the estimated kurtosis is a consistent estimator for the kurtosis parameter in both cases.

Estimator	$n = 1000$	$n = 5000$	$n = 10,000$
$\hat{K}^{(y)}$	3.0132	3.0031	3.0022
	(0.0004)	(0.00002)	(6.31×10^{-6})
Empirical kurtosis	2.9885	2.9984	3.0006
	(0.02257)	(0.0049)	(0.0027)

Table 2.5: Empirical and estimated model based kurtosis for the normal-GARCH(1, 1) model in Case 1. The MSE of the estimators appear in parentheses. $K^{(y)} = 3.0020$, $\omega = 5.3 \times 10^{-7}$, $\alpha_1 = 0.0057$, $\beta_1 = 0.9381$.

Estimator	$n = 1000$	$n = 5000$	$n = 10,000$
$\hat{K}^{(y)}$	9.09	9.03	9.02
	(0.013)	(0.0017)	(0.0008)
Empirical kurtosis	6.09	6.92	7.18
	(10.72)	(6.24)	(5.05)

Table 2.6: Empirical and estimated model based kurtosis for the t_5-distribution GARCH(1, 1) model in Case 1. The MSE of the estimators appear in parentheses. $K^{(y)} = 9.1$, $\omega = 5.3 \times 10^{-7}$, $\alpha_1 = 0.0057$, $\beta_1 = 0.9381$.

Case 2: Consider the model $y_t = \varepsilon_{t-1}^2 \varepsilon_t$, where ε_t is an uncorrelated $N(0, 1)$ sequence. The moment properties of this model were previously studied in Appadoo, Ghahramani and Thavaneswaran (2005). We generated $n = 1000, 5000, 10,000$ realizations from this model five hundred times and calculated the median sample kurtosis and median sample first lag autocorrelation of the squared process. The estimate and its corresponding median absolute deviation(MAD) are reported in Table 2.7. The median and MAD were chosen as estimators and a measure of spread as the sampling distribution of $\hat{\rho}_1^{y^2}$ and the empirical kurtosis are skewed. For large sample sizes, $\hat{\rho}_1^{y^2}$ and the empirical kurtosis are useful for identifying this model.

Estimator	$\rho_1^{y^2}$	$n = 1000$	$n = 5000$	$n = 10,000$
$\hat{\rho}_1^{y^2}$	0.1175	0.1092	0.1131	0.1195
		(0.0735)	(0.0411)	(0.0372)
Estimator	$K^{(y)}$	$n = 1000$	$n = 5000$	$n = 10,000$
Empirical kurtosis	35	24.29	29.81	31.25
		(19.01)	(10.76)	(9.63)

Table 2.7: Sample moments of $\rho_1^{y^2}$ and $K^{(y)}$ for the model in Case 2. The MAD of the estimators appears below in parentheses.

Case 3: Consider the model $y_t = \varepsilon_{t-1}\varepsilon_t$, where ε_t is an uncorrelated $N(0, 1)$ sequence. The moment properties of this model were previously studied in Appadoo, Ghahramani and Thavaneswaran (2005). We generated $n = 1000, 5000, 10,000$ realizations from this model and calculated the mean sample kurtosis and sample first lag autocorrelation of the squared process. The mean value of each estimator and its corresponding MSE are reported in Table 2.8. As the sample size increases, estimates of $\rho_1^{y^2}$ and the empirical kurtosis become useful for identifying this model.

Estimator	$\rho_1^{y^2}$	$n = 1000$	$n = 5000$	$n = 10,000$
$\hat{\rho}_1^{y^2}$	0.25	0.2400	0.2505	0.2465
		(0.0044)	(0.0013)	(0.0005)
Estimator	$K^{(y)}$	$n = 1000$	$n = 5000$	$n = 10,000$
Empirical kurtosis	9	8.46	8.78	8.92
		(2.06)	(0.8456)	(0.4145)

Table 2.8: Sample moment estimates of $\rho_1^{y^2}$ and $K^{(y)}$ for the model in Case 3. The MSE of the estimators appears below in parentheses.

In the next section, we study the moment properties of two recently proposed non-linear GARCH models. The Heston and Nandi GARCH model and Inverse-Gaussian GARCH model add flexibility to the class of GARCH models by incorporating conditional leverage in the former model and conditional leverage and conditional skewness in the latter model.

2.7 Nonlinear GARCH Models

In this section, we study the moment properties of two classes of non-linear GARCH models. The Heston and Nandi (2000) GARCH (HN-GARCH) model is similar to the GARCH-in-Mean (GARCH-M) model of Engle, Lilien and Robins (1987) in that it relates the conditional variance to the return of the process. By fitting a model where the return is regressed against the conditional variance, the researcher is able to test the hypothesis that returns are higher whenever there is increased volatility. The *leverage effect* pertains to the concept that an asset's returns may become more volatile whenever its price decreases. The Exponential GARCH(E-GARCH) model due to Nelson(1992) was designed to model the leverage effect. Moreover, it is possible to fit an E-GARCH in mean or E-GARCH-M model in order to account for the presence of both conditional leverage and to explain the correlation between returns and conditional variance. Now call prices are nothing more than discounted expected payoffs under the risk-neutral measure. Numerically, the expected payoffs obtained by using the class of linear GARCH models are non-tractible integrals that have been evaluated using either simulation or through numerical approximation; (See the NGARCH model of Duan (1995) and Duan and Simonato (2001)). Call pricing formulas that use the HN-GARCH model for volatility are however, analytically tractible. It is this tractibility that has rendered the HN-GARCH model to be widely used. To date, the Heston and Nandi (2000) paper has received over one thousand citations.

Christoffersen, Heston and Jacobs (2006) extended the HN-GARCH model by allowing conditional skewness to be modelled as well. Just as with the HN-GARCH model, the option valuation expression of Christoffersen et al.(2006) is analytically tractable.

Model identification for the HN-GARCH and IG-GARCH models remains an open problem. In this section, we provide identification methods that provide researchers insight as to when a non-linear GARCH model should be fit as opposed to the usual class of linear GARCH models.

2.7.1 HN-GARCH(1, 1) Models

Consider the Heston & Nandi (2000) GARCH(1, 1) (HN-GARCH(1, 1) model given by the following:

$$\log\left(\frac{S_{t+1}}{S_t}\right) = r + \lambda h_t + \sqrt{h_t} Z_t \tag{2.7.1}$$

$$h_t = \omega + \alpha_1 (Z_{t-1} - \gamma_1 \sqrt{h_{t-1}})^2 + \beta_1 h_{t-1}, \tag{2.7.2}$$

where $Z_t \sim NID(0,1)$, r is the risk-free rate and S_t is the price of the stock or asset at time t. Assuming stationarity, it is easy to show that $E(h_t) = \dfrac{\omega + \alpha_1}{1 - (\beta_1 + \gamma_1^2 \alpha_1)}$.

The HN-GARCH(1, 1) model does not model conditional skewness but does model conditional leverage. Let E_{t-1} denote a conditional expectation operator and let Cov_t denote a conditional covariance operator. The conditional skewness of a HN-GARCH(1,1) process is given by the following:

$$\mathrm{Skew}_t(y_t) = \frac{E_{t-1}(y_t - E_{t-1}(y_t))^3}{h_t^{3/2}} = \frac{E_{t-1}(y_t - r - \lambda h_t)^3}{h_t^{3/2}} = \frac{E_{t-1}(\varepsilon_t^3)}{h_t^{3/2}} = \frac{E_{t-1}(h_t^{3/2} Z_t^3)}{h_t^{3/2}} = 0,$$

since $Z_t \sim NID(0,1)$. However, the conditional leverage is non-zero since

$$\begin{aligned}
\mathrm{Cov}_t(y_t, h_{t+1}) &= \mathrm{Cov}_{t-1}(\lambda h_t + \sqrt{h_t} Z_t, h_{t+1}) \\
&= \lambda \mathrm{Cov}_{t-1}(h_t, h_{t+1}) + \mathrm{Cov}_{t-1}(\sqrt{h_t} Z_t, h_{t+1}) \\
&= \lambda \mathrm{Cov}_{t-1}(h_t, \alpha_1 (Z_t - \gamma_1 \sqrt{h_t})^2 + \beta_1 h_t) \\
&+ \mathrm{Cov}_{t-1}(\sqrt{h_t} Z_t,\ \alpha_1 (Z_t - \gamma_1 \sqrt{h_t})^2 + \beta_1 h_t) \\
&= \lambda[\mathrm{Cov}_{t-1}(h_t, \beta_1 h_t) + \alpha_1 \mathrm{Cov}_{t-1}(h_t,\ Z_t^2 - 2\gamma_1 Z_t \sqrt{h_t} + \gamma_1^2)] \\
&+ \mathrm{Cov}_{t-1}[\sqrt{h_t} Z_t,\ \alpha_1 (Z_t^2 - 2\gamma_1 \sqrt{h_t} Z_t + \gamma_1^2 h_t) + \beta_1 h_t] \\
&= -2\alpha_1 \gamma_1 \mathrm{Cov}_{t-1}(\sqrt{h_t} Z_t,\ \sqrt{h_t} Z_t) \\
&= -2\alpha_1 \gamma_1 h_t,
\end{aligned}$$

Model	ω	α_1	β_1	γ_1	λ
HN-GARCH(1,1)	4.939×10^{-6}	1.579×10^{-6}	9.062×10^{-6}	785.3	-0.5
GARCH(1,1)	4.960×10^{-7}	7.3×10^{-2}	0.925	0	-0.5

Table 2.9: Parameter estimates from a normal-GARCH(1,1) fit and a HN-GARCH(1,1) fit. The table is adapted from Hsieh and Ritchken (2000).

since $E(Z_t^2) = 1$.

Example 2.7.1 *Comparison of normal-GARCH(1,1) and HN-GARCH(1,1) parameter estimates*

Table 2.9 (taken from Hsieh and Ritchken (2000)) gives parameter estimates from a non-zero mean normal-GARCH(1,1) model fit and a HN-GARCH(1,1) model fit. The estimate of $\sqrt{\widehat{E(h_t)}}$ obtained from fitting a normal-GARCH(1,1) model is $\sqrt{\dfrac{4.960 \times 10^{-7}}{1 - 7.3 \times 10^{-2} - 0.925}} =$ 0.0157. The estimate of $\sqrt{\widehat{E(h_t)}}$ obtained from fitting a HN-GARCH(1,1) model is

$$\sqrt{\frac{4.939 \times 10^{-6} + 1.579 \times 10^{-6}}{1 - 9.062 \times 10^{-6} - (1.579 \times 10^{-6})(785.3^2)}} = 0.0158.$$

The estimates of $\sqrt{\widehat{E(h_t)}}$ do not differ, however, the conditional leverage would not be accounted for by the normal-GARCH(1,1) model.

It is of interest to note that while a GARCH-in-mean (GARCH-M) model has the same conditional mean structure as the conditional mean structure of the HN-GARCH model, neither conditional leverage, nor conditional skewness are modelled by the GARCH-in-mean model. To see this, consider the GARCH(1,1)-M model for the log return process $y_t = \log\left(\dfrac{S_t}{S_{t-1}}\right)$, where S_t represents the stock price at time

t, given by the following:

$$\begin{aligned} y_t &= \delta\sqrt{h_t} + \varepsilon_t, \\ \varepsilon_t &= \sqrt{h_t} Z_t, \\ h_t &= \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}, \end{aligned}$$

where $Z_t \sim NID(0,1)$. This model differs from the usual GARCH model in that the conditional mean is time-varying as $E(y_t|\mathcal{F}_{t-1}^y) = \delta\sqrt{h_t}$. However, the presence of conditional skewness and conditional leverage are not modelled by the GARCH(1,1)-M model. Let Skew_t denote conditional skewness of the GARCH(1,1)-M process. Then we have the following:

$$\text{Skew}_t = \frac{E_{t-1}(y_t - E_{t-1}(y_t))^3}{h_t^{3/2}} = \frac{E_{t-1}(y_t - \delta\sqrt{h_t})^3}{h_t^{3/2}} = \frac{E_{t-1}(\varepsilon_t^3)}{h_t^{3/2}} = \frac{h_t^{3/2} E(Z_t^3)}{h_t^{3/2}} = 0,$$

since $Z_t \sim N(0,1)$.

Empirically, the presence of conditional skewness for a volatility process would be reflected by the presence of significant lags in the sample ACF of plot of the y_t^3 process. The GARCH$(1,1)$-M process does not model a conditional leverage effect as

$$\begin{aligned} \text{Cov}_t(y_t, h_{t+1}) &= \text{Cov}(\delta\sqrt{h_t} + \sqrt{h_t} Z_t, h_{t+1}|\mathcal{F}_{t-1}^y) \\ &= \text{Cov}(\sqrt{h_t}(\delta + Z_t), \alpha_1 h_t Z_t^2|\mathcal{F}_{t-1}^y) \\ &= \alpha_1 h_t^{3/2} E(Z_t^3|\mathcal{F}_{t-1}^y) = 0, \end{aligned}$$

since $Z_t \sim NID(0,1)$. Empirically, the absence of leverage can be observed by the lack of significant cross-correlations in the Cross-Correlation Function (CCF) plot between the observed series and the square of the observed series.

2.7.2 Simulation Study

We study the consequences of fitting a normal-GARCH$(1,1)$ model to a HN-GARCH$(1,1)$ model in a simulation study. We also simulate from a normal-GARCH$(1,1)$ process for comparison purposes. Specifically, 6000 observations are simulated from a normal-GARCH$(1,1)$ model with parameters $\omega = 5 \times 10^{-6}$, $\beta_1 = 0.589$ and $\alpha_1 = 1.32 \times 10^{-6}$ and 6000 observations are simulated from a HN-GARCH$(1,1)$ model with parameters: $\omega = 5 \times 10^{-6}$, $\beta_1 = 0.589$, $\alpha_1 = 1.32 \times 10^{-6}$, $\gamma = 421.39$, $r = 0.00019$, $\lambda = 0.205$. These values are the fitted values obtained in Heston and Nandi (2000). The CCF plot between the normal-GARCH(1, 1) process and its square fails to indicate the presence of any systematic cross-correlation at the first lag as observed in Figure 2.6. Hence, there is an absence of conditional leverage in a normal-GARCH(1, 1) model.

On the other hand, just as with linear GARCH models, the sample ACF of the HN-GARCH(1, 1) process shows no autocorrelation while the sample ACF of its squared process shows the presence of autocorrelation (See Figure 2.7). The absence of conditional skewness is demonstrated by the absence of significant autocorrelations in the sample ACF of the y_t^3 process (See Figure 2.8). The CCF plot between the HN-GARCH(1, 1) series and the squared HN-GARCH(1, 1) series indicates that these series are negatively correlated and therefore conditional leverage is present (See Figure 2.9). A normal-GARCH(1, 1) model is fit and the model seems to be adequate as the Ljung-Box tests of white noise of the standardized residuals and the of squared standardized residuals have p-values 0.8384 and 0.3472, respectively. For other diagnostic plots of residuals from the normal-GARCH$(1,1)$ fit see Figures 2.10 (where the ARCH effects are absent) and Figure 2.11 (where the residuals are normal).

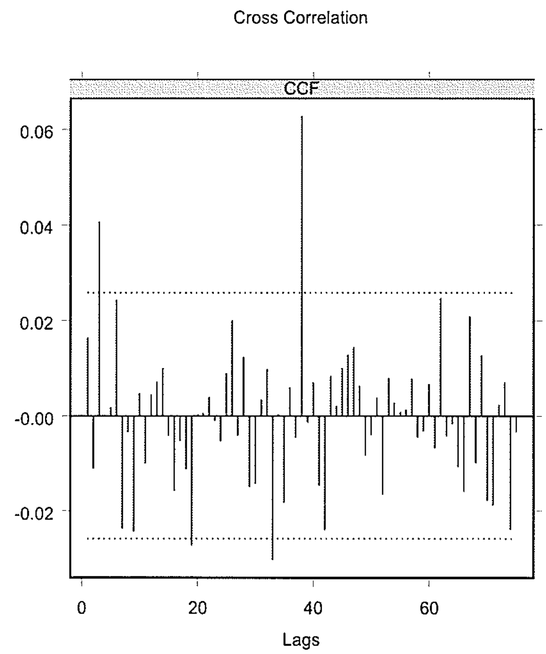

Figure 2.6: Cross-correlation function between the series and the squared series for the data simulated from a normal-GARCH(1, 1) model with parameters $\omega = 5 \times 10^{-6}$, $\beta_1 = 0.589$, $\alpha_1 = 1.32 \times 10^{-6}$.

However, the CCF plot in Figure 2.12 of the standardized residuals and the squared standardized residuals indicates that the normal-GARCH(1, 1) is unable to account for the leverage effect in the simulated HN-GARCH(1, 1) process.

In the next section, the moment properties of another nonlinear GARCH model are studied.

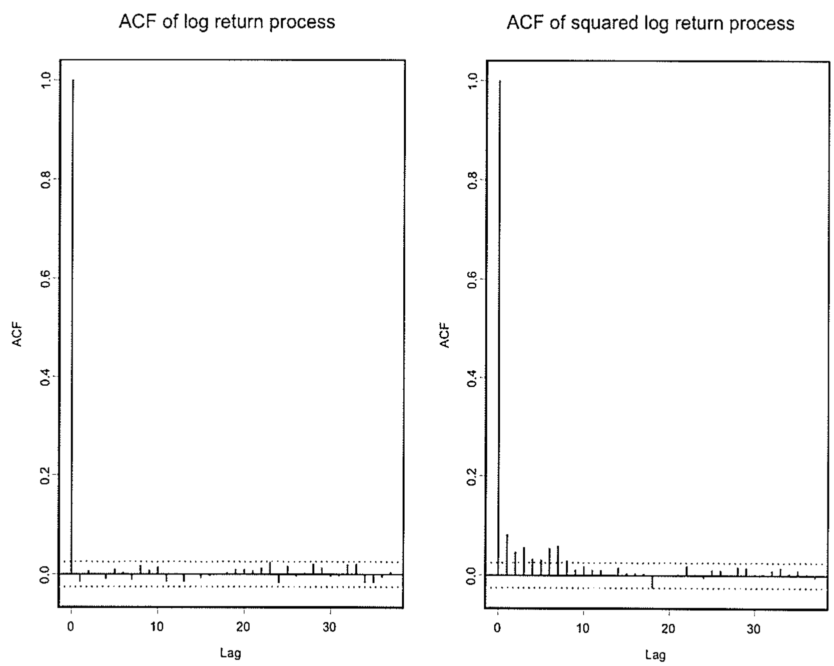

Figure 2.7: Sample ACF of log return and squared log return of a HN-GARCH(1, 1) model with parameters $\omega = 5 \times 10^{-6}$, $\beta_1 = 0.589$, $\alpha_1 = 1.32 \times 10^{-6}$, $\gamma = 421.39$, $r = 0.00019$, $\lambda = 0.205$.

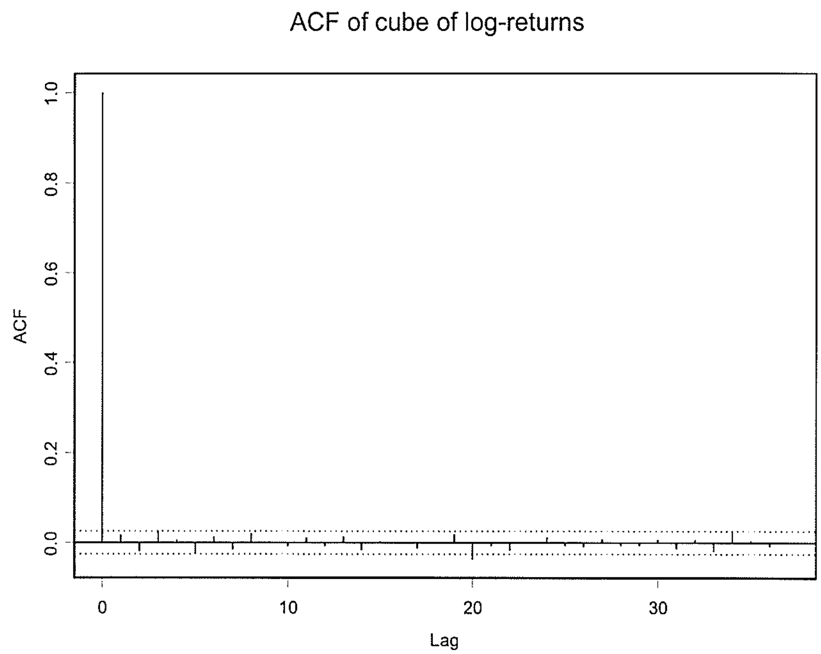

Figure 2.8: ACF of the cubed series for the data simulated from a HN-GARCH(1, 1) model with parameters $\omega = 5 \times 10^{-6}$, $\beta_1 = 0.589$, $\alpha_1 = 1.32 \times 10^{-6}$, $\gamma = 421.39$, $r = 0.00019$, $\lambda = 0.205$.

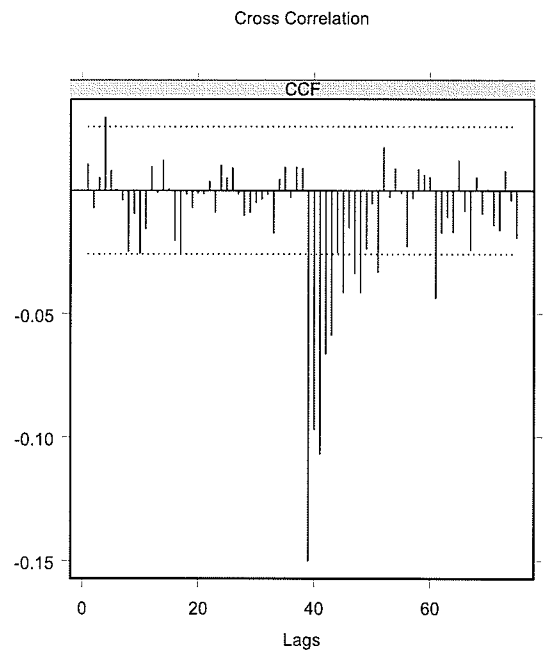

Figure 2.9: Cross-correlation function between the series and the squared series for the data simulated from a HN-GARCH(1,1) model with parameters $\omega = 5 \times 10^{-6}$, $\beta_1 = 0.589$, $\alpha_1 = 1.32 \times 10^{-6}$, $\gamma = 421.39$, $r = 0.00019$, $\lambda = 0.205$.

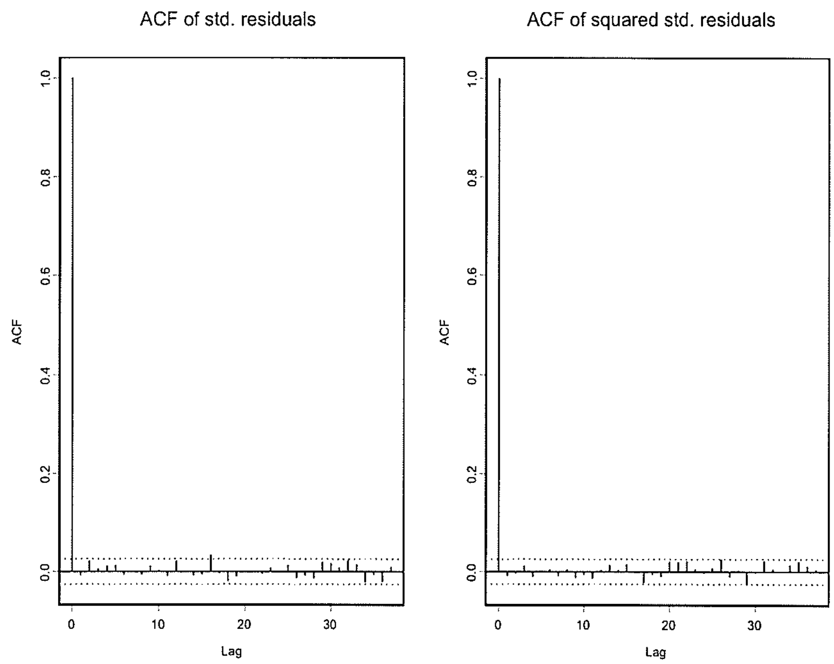

Figure 2.10: Sample ACF of the standardized residuals and the squared standardized residuals of the normal-GARCH(1, 1) model fit to the HN-GARCH(1, 1) simulated data. The data is simulated from a model with parameters $\omega = 5 \times 10^{-6}$, $\beta_1 = 0.589$, $\alpha_1 = 1.32 \times 10^{-6}$, $\gamma = 421.39$, $r = 0.00019$, $\lambda = 0.205$.

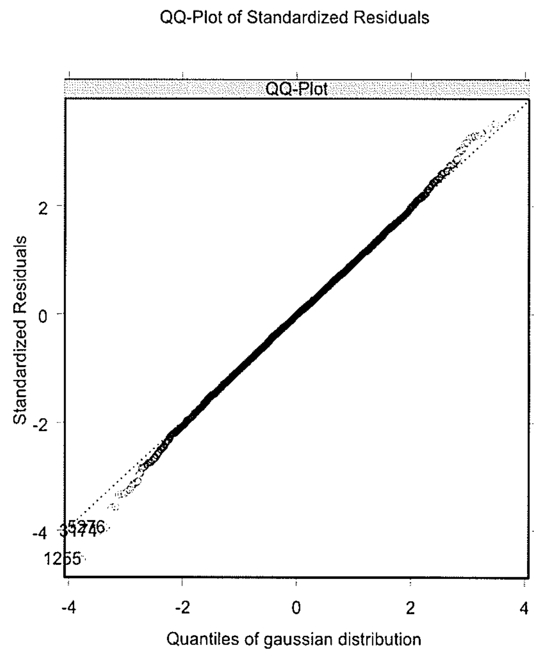

Figure 2.11: QQ-plot of the standardized residuals of normal-GARCH(1, 1) model fit to the simulated data from a HN-GARCH(1, 1) model with parameters $\omega = 5 \times 10^{-6}$, $\beta_1 = 0.589$, $\alpha_1 = 1.32 \times 10^{-6}$, $\gamma = 421.39$, $r = 0.00019$, $\lambda = 0.205$. .

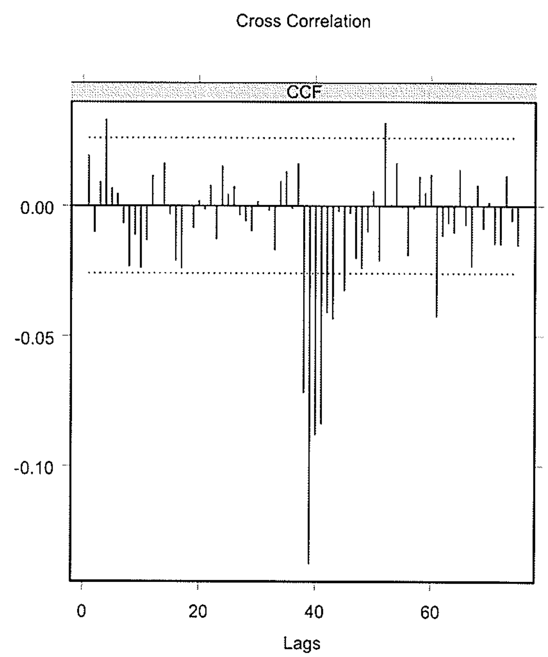

Figure 2.12: Cross-correlation plot of the standardized residuals of normal-GARCH(1,1) model fit to the simulated data from a HN-GARCH(1,1) model with parameters $\omega = 5 \times 10^{-6}$, $\beta_1 = 0.589$, $\alpha_1 = 1.32 \times 10^{-6}$, $\gamma = 421.39$, $r = 0.00019$, $\lambda = 0.205$.

2.7.3 GARCH models with conditional skewness

In this section, the Inverse Gaussian GARCH (IG-GARCH) model of Christoffersen et al.(2006) is discussed. The classical inverse Gaussian distribution denoted as $IG(\mu, \lambda)$ has the following density.

$$f(x;\mu,\lambda) = \frac{\sqrt{\lambda}}{\sqrt{2\pi x^3}} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right), \quad x > 0 \tag{2.7.3}$$

where $\mu > 0$ and $\lambda > 0$. The IG-GARCH model due to Christoffersen et al.(2006) uses a standardized one-parameter Inverse Gaussian distribution. Details on random number generation from the one-parameter Inverse Gaussian distribution are given in Appendix A.1.

The following dynamic model proposed by Christoffersen et al. (2006) specifies returns on a spot asset price at time t, S_t, and the conditional variance of return h_{t+1} :

$$\log(S_{t+1}/S_t) = r + \nu h_{t+1} + \eta y_{t+1} \tag{2.7.4}$$

$$h_{t+1} = \omega + b h_t + c y_t + a h_t^2 / y_t \tag{2.7.5}$$

where, given the available information at time t, y_{t+1} has an inverse Gaussian conditional distribution with degrees of freedom parameter $\delta_{t+1} = h_{t+1}/\eta^2$. r is the risk-free rate.

Neither the normal-GARCH$(1,1)$ model, nor the GARCH$(1,1)$-M model adequately model a process with time-varying conditional skewness and conditional leverage. We simulated 6000 observations from a IG-GARCH$(1,1)$ process and found that after fitting a normal-GARCH$(1,1)$ model to these data, the cross-correlation function between the residual process and the squared residual process contained significant autocorrelations indicating the presence of conditional leverage. Furthermore, the sample ACF of the cube of the IG-GARCH$(1, 1)$ process contains significant autocorrelations which indicates the presence of conditional skewness. The model parameters we used were the following: $r = 0.00019$, $\nu = 1625$, $\eta = 6.162 \times 10^{-4}$, $\omega = 3.768 \times 10^{-10}$, $b = -19.33$, $c = 4.142 \times 10^{-6}$, $a = 2.472 \times 10^{7}$. These model parameters are similar to the parameter estimates obtained in the paper by Christoffersen et al. (2006). We begin by examining the sample ACF of the process and the squared process in Figure 2.13. Both processes are time-varying. The CCF plot between the log return process and the squared log return process in Figure 2.14 corroborates this fact.

A normal-GARCH$(1, 1)$ model is fit to the simulated IG-GARCH$(1, 1)$ data and the adequacy of the fit is examined. This model fails to be adequate as the p-values of the Box-Ljung test of white noise for the standardized residuals and the squared standardized residuals are less than 0.001. The non-normality of the residuals is further illustrated in their QQ-plot in Figure 2.16. The presence of significant lags in the sample ACFs of the standardized residuals and the square standardized residuals in Figure 2.15 again show that the normal-GARCH$(1, 1)$ model fails to provide an adequate fit for an IG-GARCH(1, 1) process. In addition, the presence of significant cross-correlations between the standardized residuals and the squared standardized residuals in Figure 2.17 once again demonstrates that the normal-GARCH(1, 1) model fails to be adequate for this data.

The presence of significant cross-correlations between the observed log return process and the squared log return process suggests a non-linear GARCH model is more

Model	cond'l mean changes	cond'l variance changes	cond'l skewness changes	cond'l leverage changes
GARCH(1, 1)	No	Yes	No	No
GARCH(1, 1)-M	Yes	Yes	No	No
HN-GARCH(1, 1)	Yes	Yes	No	Yes
IG-GARCH(1, 1)	Yes	Yes	Yes	Yes

Table 2.10: Characteristics of some linear and nonlinear GARCH models.

appropriate than a linear GARCH model. The presence of significant autocorrelations in the sample ACF of the cube of the process, suggests conditional skewness. The IG GARCH model nests the HN-GARCH model and should be used to account for both the presence of conditional skewness and conditional leverage.

Table 2.10 summarizes the difference between linear and non-linear GARCH models.

2.8 Conclusions

The moment properties of the class of GARCH models and RCA model have been studied. Moment properties of extensions of GARCH models that allow for time-varying mean as well as time-varying variance have also been studied. In particular, for the extended class of GARCH models, the kurtosis of the observed process has been shown to be related to the kurtosis of the error process. The moment properties of a new class of models developed for autocorrelated count data have been studied. This new class of models known as ACP models has been shown to share certain properties with the class of GARCH models allowing for simplified derivations of the moment properties derived in Heinen (2003). The moment properties of a volatility model that accounts for structural change caused by a Markov process have been studied. Using

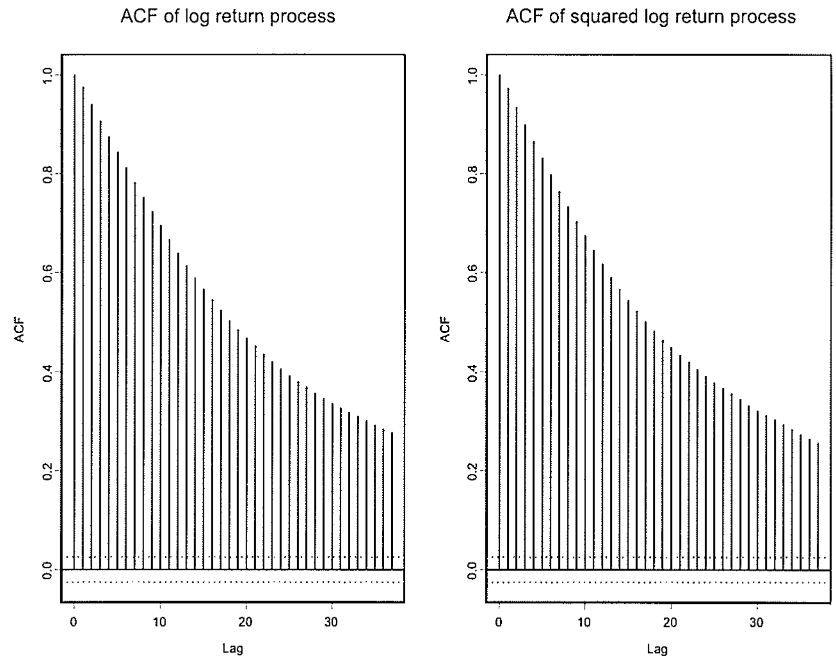

Figure 2.13: Sample ACF of log return and squared log return of a IG-GARCH(1, 1) model with parameters $r = 0.00019$, $\nu = 1625$, $\eta = 6.162 \times 10^{-4}$, $\omega = 3.768 \times 10^{-10}$, $b = -19.33$, $c = 4.142 \times 10^{-6}$, $a = 2.472 \times 10^{7}$.

simulation studies, moment properties have been shown to be useful for GARCH model identification The sample CCF plot between the lagged log-return process and the square of the process has been shown to be useful for identifying the presence of conditional leverage. The sample ACF plot of the cube of the log-return process has been shown to be useful for identifying the presence of conditional skewness. The presence of conditional skewness in addition to the presence of conditional leverage suggests that the IG-GARCH model of Christoffersen et al. (2006) would be more appropriate than the the usual GARCH model or the HN-GARCH model of Heston and Nandi (2000).

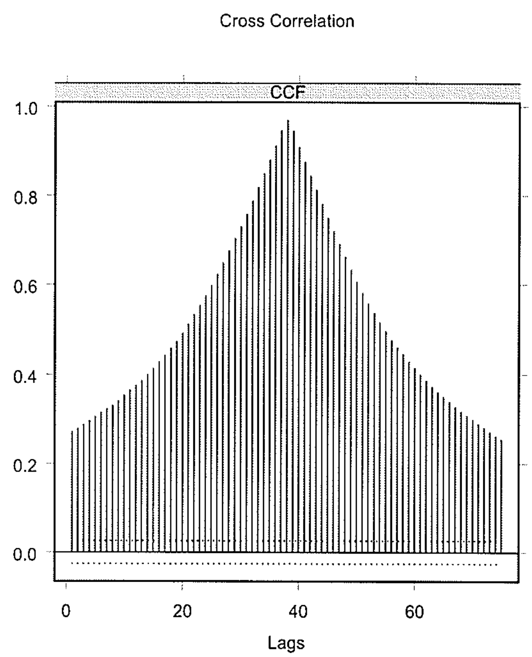

Figure 2.14: Cross-correlation function between the series and the squared series for the data simulated from a IG-GARCH(1, 1) model with parameters $r = 0.00019$, $\nu = 1625$, $\eta = 6.162 \times 10^{-4}$, $\omega = 3.768 \times 10^{-10}$, $b = -19.33$, $c = 4.142 \times 10^{-6}$, $a = 2.472 \times 10^{7}$.

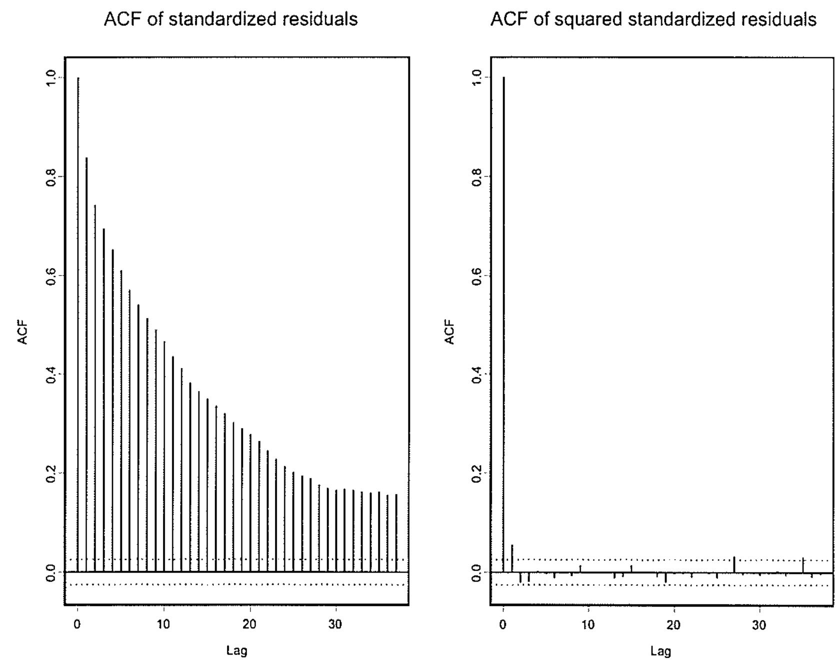

Figure 2.15: Sample ACF of the standardized residuals and the squared standardized residuals of the GARCH(1, 1) model fit to the IG-GARCH(1, 1) simulated data. The data is simulated from a model with parameters $r = 0.00019$, $\nu = 1625$, $\eta = 6.162 \times 10^{-4}$, $\omega = 3.768 \times 10^{-10}$, $b = -19.33$, $c = 4.142 \times 10^{-6}$, $a = 2.472 \times 10^{7}$.

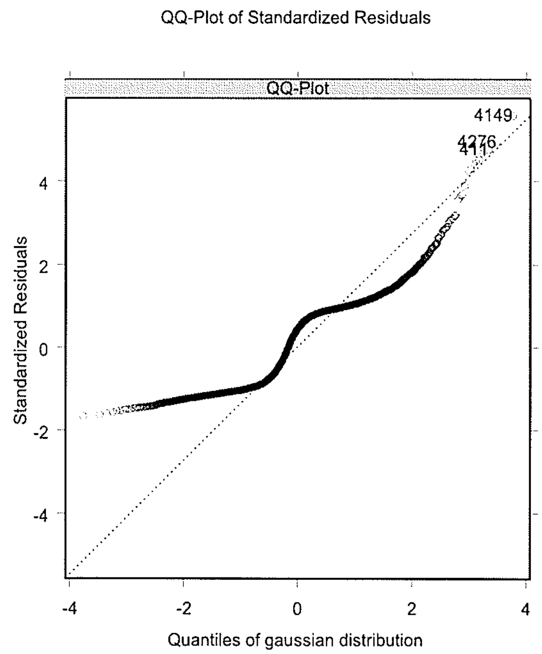

Figure 2.16: QQ-plot of the standardized residuals of the normal-GARCH(1, 1) model fit to the simulated data from a IG-GARCH(1, 1) model with parameters $r = 0.00019$, $\nu = 1625$, $\eta = 6.162 \times 10^{-4}$, $\omega = 3.768 \times 10^{-10}$, $b = -19.33$, $c = 4.142 \times 10^{-6}$, $a = 2.472 \times 10^{7}$.

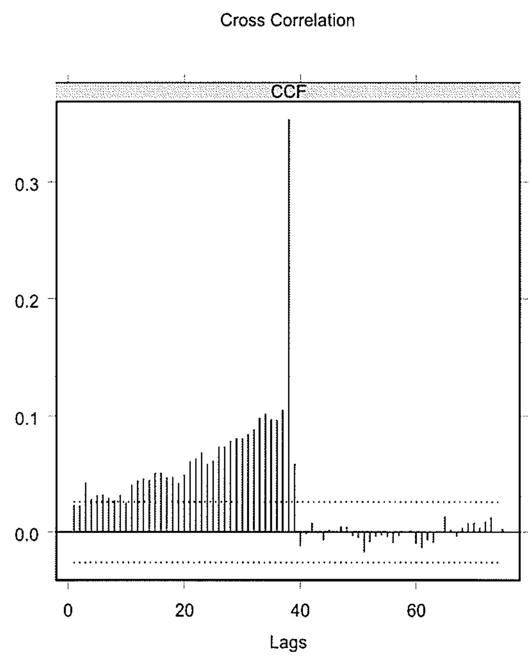

Figure 2.17: Cross-correlation plot of the standardized residuals of GARCH(1, 1) model fit to the simulated data from a IG-GARCH(1, 1) model with parameters $r = 0.00019$, $\nu = 1625$, $\eta = 6.162 \times 10^{-4}$, $\omega = 3.768 \times 10^{-10}$, $b = -19.33$, $c = 4.142 \times 10^{-6}$, $a = 2.472 \times 10^{7}$.

Chapter 3

Applications of Combining Estimating Functions

In this chapter, estimating functions are combined to obtain estimates with smaller mean square errors. In Section 3.1, we state the Combination theorem due to Thompson and Thavaneswaran (1999) for combining non-orthogonal estimating functions. In Section 3.2, we develop a GARCH model identification tool based on the theorem in Section 3.1. In Section 3.3, we study the problem of combining estimating functions for the class of AR(1) processes with GARCH errors and show how as a result, an improved point estimate of the variability of the log return series may be obtained. In Section 3.4, we study an application of combining estimating functions to the hypothesis testing problem for the class of stationary processes with GARCH errors. We end with concluding remarks in Section 3.5.

3.1 Combination Theorem

The result of Godambe (1999) uses combinations of orthogonal estimating functions. In this Section, we give some combination results when the components need not be orthogonal. Consider a probability space $(\Omega, \mathcal{F}, \mathcal{P})$, on which θ is a real valued random variable. An estimating function for θ is a function $g(\cdot, \theta)$ on $\Omega \times \mathcal{R}$, and it is unbiased if $Eg(\cdot, \theta(\cdot)) = 0$, where '$\cdot$' represents a generic point of Ω. Let $g_1(\cdot,\ \theta)$, $g_2(\cdot,\ \theta)$ be fixed unbiased estimating functions having finite and positive variances, and such that the expectations of $\partial g_1/\partial\theta$ and $\partial g_2/\partial\theta$ are finite, with $E[\partial g_1/\partial\theta] \neq 0$. For example, g_1 and g_2 could be score functions. The following theorem, given in Thompson and Thavaneswaran(1999), is used to obtain a model identification method for linear and nonlinear time series in Section 3.2.

Theorem 3.1.1 *(Thompson and Thavaneswaran(1999)): In the class of all unbiased estimating functions*

$$g = g_1 + cg_2,$$

(i) the function which minimizes Var g *is given by*

$$g^* = g_1 + C^* g_2$$

where

$$C^* = -\mathrm{Cov}(g_1, g_2)/\mathrm{Var}\ g_2$$

and

(ii) the function which minimizes Var $g/(E[\frac{\partial g}{\partial\theta}])^2$ *is given by*

$$g^0 = g_1 + C^0 g_2$$

where

$$C^0 = \frac{E\left[\frac{\partial g_2}{\partial\theta}\right]\ \mathrm{Var}\ g_1 - E\left[\frac{\partial g_1}{\partial\theta}\right]\ \mathrm{Cov}(g_1, g_2)}{E\left[\frac{\partial g_1}{\partial\theta}\right]\mathrm{Var}\ g_2 - E\left[\frac{\partial g_2}{\partial\theta}\right]\ \mathrm{Cov}(g_1, g_2)}\ .$$

We now show how to apply the results above to model identification of linear and nonlinear time series.

3.2 Model Identification

In this Section, we study GARCH model identification by applying Theorem 3.1.1. Using Theorem 3.1.1, the correlation between least squares (LS) estimating functions and least absolute deviation (LAD) estimation functions is obtained and it turns out to be the asymptotic correlation between the corresponding estimators. The following example motivates combining estimating functions.

Example 3.2.1 *Suppose we want to estimate $\theta = E(|Z|)$, where Z is standard normal. We generate n iid $N(0,1)$ variables $Z_1, \ldots, Z_n$ and compute $\hat{\theta} = \sum |Z_i|/n$. But we know that $E(Z_i^2) = 1$ and can see easily that $\hat{\theta}$ is positively correlated with $\sum Z_i^2/n$. Hence the combined estimate is $\tilde{\theta} = \hat{\theta} - c(\sum Z_i^2/n - 1)$, where the value of c is chosen to minimize the variance of $\tilde{\theta}$ and is given by $c = \dfrac{Cov(\hat{\theta}, \sum Z_i^2/n)}{Var(\sum Z_i^2/n)}$.*

In practice, the value of c can be estimated by regressing $|Z_i|$ on Z_i^2. The main message is that two different measures of variability can be combined to obtain a better estimate (see Samanta (1985) for details on combining estimates). Rather than combine estimates, estimating functions will be combined for GARCH model identification.

Following Thavaneswaran and Ghahramani (2004), we will combine LS and LAD estimating functions in order to develop an GARCH model identification tool. The least squares estimating function g_1 based on X_i from a random sample $X_1, \ldots, X_n$ from a symmetric population with location parameter θ and standard deviation σ is given by $g_1(X_i;\theta) = \dfrac{(X_i - \theta)}{\sigma}$. The LAD estimating function based on X_i is given by $g_2(X_i;\theta) = \text{sgn}(X_i - \theta)$. Then it is easy to show that the correlation, ρ, between g_1

and g_2 is

$$\rho = \mathrm{Corr}(g_1, g_2) = \frac{\mathrm{E}|X - \theta|}{\sigma}. \tag{3.2.1}$$

Hence, the estimating function that minimizes $Var(g^*)$ according to Theorem 3.1.1 part(a), is given by

$$g^* = \frac{(X_i - \theta)}{\sigma} - \rho\,\mathrm{sgn}(X_i - \theta). \tag{3.2.2}$$

Then $\mathrm{Var}(g^*) = (1 - \rho^2)$, so that $\mathrm{Var}(g^*)$ is minimized whenever ρ is maximized. If we let g_1 be the LAD estimating function and let g_2 be the LS estimating function, then again $\mathrm{Var}(g^*) = (1 - \rho^2)$ as well. Hence, the combined estimating function in (3.2.2) has smaller variance than either the LS estimating function or the LAD estimating function. This in turn implies that the asymptotic variance of the estimator corresponding to the estimating function in (3.2.2) is less than either the variance of the least squares estimator or the variance of the LAD estimator, (see Heyde (1997) for the details of the relationship between the asymptotic properties of the estimates and the finite sample properties of the estimating functions.)

It is of interest to note that, the correlation between the LS estimating function and the LAD estimating function has not been studied in the literature. Consider a random sample of size n drawn from a population having mean μ, variance σ^2, median ν and density $f(x)$. The asymptotic joint distribution of the sample mean $\bar{X}$ and the sample median $\tilde{X}$ is given by

$$\sqrt{n}\begin{pmatrix} \bar{X} \\ \tilde{X} \end{pmatrix} \longrightarrow_{\mathcal{L}} N\left(\begin{pmatrix} \mu \\ \nu \end{pmatrix}, \begin{pmatrix} \sigma^2 & E|X-\nu|/(2f(\nu)) \\ & 1/(2f(\nu))^2 \end{pmatrix}\right)$$

Distribution	E$\|X-\mu\|/\sigma$
Normal(65,25)	0.797885
t_3	0.636620
t_4	0.7071068
t_5	0.735105
t_6	0.750000
t_7	0.7592134
t_8	0.765466
Logistic(0,1)	0.764304
Double exponential(1)	0.707107
Uniform(0,1)	0.866025
Beta(2,2)	0.838525
Beta(3,3)	0.826797
Beta(4,4)	0.820313

Table 3.1: Theoretical E$|X-\mu|/\sigma$ values.

as $n \longrightarrow \infty$, (Ferguson (1999)). Hence, the asymptotic correlation between the sample mean and sample median is the same as the finite sample correlation between the LS and LAD estimating functions. The following example illustrates an application of the result.

Example 3.2.2 : *Suppose $X_1, \ldots, X_n$ is a random sample from a $N(\mu, \sigma^2)$ population. Assume σ^2 is known and interest centers on estimating μ. The correlation between the sample mean and sample median is approximately given by*

$$Corr(\bar{X}, \tilde{X}) \approx \sqrt{\frac{2}{\pi}} = 0.797. \tag{3.2.3}$$

The theoretical E$|X-\mu|/\sigma$ values have been calculated for a number of distributions (See Johnson and Kotz (1970)). Table 3.1 provides these values for some symmetric distributions.

After a time series model is fit, the moment estimator of ρ based on the residuals can be used to identify the error distribution. For example, suppose we decide to fit a GARCH model to a return series $\{y_t\}$ after inspecting the sample ACF of the series and the squared series. We can fit a normal-GARCH model as an initial model. The residuals are given by $\widehat{Z}_t = \frac{y_t}{\widehat{h}_t}$ where $\widehat{h}_t = \widehat{\omega} + \widehat{\alpha}_1 y_{t-1}^2 + \widehat{\beta}_1 h_{t-1}$ represents the estimated conditional variance at time t. The moment estimator of ρ is given by $\widehat{\rho} = \frac{\sum_{t=1}^{n} |Z_t - \bar{Z}|/n}{s_Z}$, where $\bar{Z}$ and s_Z denote the sample mean and sample standard deviation of the residuals, respectively.

$\widehat{\rho}$ as a GARCH model identification tool is studied using simulation studies in Chapter 5. The error distribution of GARCH models fit to real financial data is identified using $\widehat{\rho}$ in Chapter 5.

3.3 Combining Estimating Functions for Volatility

In this section, we study the problem of improving the estimate of volatility by combining orthogonal estimating functions. Thavaneswaran and Heyde (1999) have compared the information associated with LS and LAD estimating functions for the autoregressive parameter of a process $\{y_t\}$ and showed that the LAD estimating function is more efficient than the ordinary LS estimating function if the distribution of the error term is such that $4f^2(0) \geq 1/\sigma^2$ where σ^2 is the variance of the error term and $f(x)$ is the conditional density of $y_t|y_{t-1}$ such that $f(0) > 0$. Thavaneswaran and Heyde (1999) have also shown that for heavy tailed distributions, the efficiency of the estimator can be improved by using the LAD estimating function.

Using the result of Heyde (1997) for an autoregressive process where the conditional mean and conditional variance are a function of the autoregressive parameter, we combine LS type and LAD type estimating functions for a class of volatility models and show that the combined estimating function has more information. The information associated with the combined estimating function obtained by non-orthogonal combination turns out to be the same as the information of the orthogonal combination in Heyde (1997). We also discuss combining estimating functions for RCA models.

Following Godambe (1985), we say that any $\mathbf{R}^p$ valued function g of the variates $(y_{t_1}, \ldots y_{t_n})$ and the parameter θ, satisfying certain regularity conditions such as square integrable and differentiable in Θ with nonzero derivative, is called a regular unbiased estimating function if, $E_\theta[g(y_{t_1}, \ldots, y_{t_n})] = 0, \theta \in \Theta$. Let L be the class of estimating functions g of the form $g_n = \sum_{i=1}^{n} a_{t_{i-1}} h_{t_{i-1}}$, where the functions h_{t_i} are such that $E_\theta[h_{t_i} | F_{t_{i-1}}] = 0, \quad (i = 1, \ldots, n)$ and $a_{t_{i-1}}$ is a function of $y_{t_1}, \ldots y_{t_{i-1}}$. The following theorem due to Godambe (1985), gives the form of an optimal estimating function.

Theorem 3.3.1 *(Godambe (1985)) In the class L of unbiased estimating functions g, the optimal estimating function g^* is the one which makes the difference $Cov(b, b) - Cov(b^*, b^*)$ non-negative definite, where*

$$b = \left[E \frac{\partial g_n}{\partial \theta} \right]^+ g_n, \quad b^* = \left[E \frac{\partial g_n^*}{\partial \theta} \right]^+ g_n*$$

and $+$ denotes the pseudoinverse of a matrix. The optimal estimating function is given by

$$g_n^* = \sum_{i=1}^{n} a_{t_{i-1}}^* h_{t_i}, \quad with \ a_{t_{i-1}}^* = \left[E \frac{\partial h_{t_i}}{\partial \theta} \right] / E[h_{t_i} h_{t_i}^t | F_{t_i}].$$

The martingale information associated with the optimal estimating function is $I_{g^*} = \dfrac{E^2\left(\dfrac{\partial g_n^*}{\partial \theta}\right)}{E(g_n^{*2})}$ *where the expectation is conditional on the information set.*

In what follows, we begin by illustrating how combining estimating functions improves estimation of the parameter in situations where the variance is a function of the mean. Next we consider the same set-up for estimating the conditional mean parameter in an autoregressive model and an autoregressive model with GARCH(1, 1) errors.

Suppose

$$y_t = \theta + \varepsilon_t \tag{3.3.1}$$

where $\varepsilon_t = \varepsilon_t(\theta)$ are i.i.d with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = \sigma^2(\theta)$. Denote $\sigma^2 = \sigma^2(\theta)$. Assume $E(\varepsilon_t^4) < \infty$, $\sigma^{-3}E(\varepsilon_t^3) = \gamma$ and that $\sigma^{-4}E(\varepsilon_t^4) - 3 = \kappa$. Consider the following elementary estimating functions m_t, k_t and l_t for θ given by $m_t = y_t - \theta$ and $k_t = (y_t - \theta)^2 - \sigma^2$ and $l_t = -sgn(m_t)$. We wish to estimate θ on the basis of the sample $y_1, \ldots, y_n$.

It follows from Theorem 3.3.1 that the optimal estimating function based on m_t is given by $M = -\dfrac{1}{\sigma^2}\displaystyle\sum_{t=1}^{n} m_t$, and the corresponding information is $I_{LS} = \dfrac{n}{\sigma^2}$. Moreover, the optimal estimating function based on k_t is given by $K = -\dfrac{2\dot{\sigma}}{(\kappa+2)\sigma^3}\displaystyle\sum_{t=1}^{n} k_t$ and the corresponding information is $I_K = \dfrac{4n(\dot{\sigma})^2}{(\kappa+2)\sigma^2}$. In addition, the optimal estimating function based on l_t is given by $L = -2f(0)\displaystyle\sum_{t=1}^{n} l_t$ and its information is $I_{LAD} = 4nf^2(0)$. The following theorem gives the form of the combined estimating functions and the associated information.

Theorem 3.3.2 *(a) For the model described in Equation (3.3.1), the optimal combined estimating function and its information formed by combining $m_t = y_t - \theta$ and $k_t = m_t^2 - \sigma^2$ are given by the following:*

$$(i)\ g(\underline{y};\theta) = \frac{1}{\sigma^3(\kappa+2)}\left(1-\frac{\gamma^2}{\kappa+2}\right)^{-1}\sum_{t=1}^{n}\{\sigma(2\dot{\sigma}\gamma-(\kappa+2))m_t+(\gamma-2\dot{\sigma})k_t\},$$

$$(ii)\ I_g = \left(1-\frac{\gamma^2}{\kappa+2}\right)^{-1}\left\{I_{LS}+I_K-\frac{4\dot{\sigma}\gamma n}{\sigma^2(\kappa+2)}\right\}.$$

(b) Assuming ε_t is symmetric with p.d.f. $f(.)$, and letting $\rho = \frac{E|\varepsilon_t|}{\sigma}$, then the optimal estimating function for θ based on combining the LS and LAD estimating functions $m_t = y_t - \theta$ and $l_t = -sgn(m_t)$ and its information are given by the following:

$$(i)\ g(\underline{y};\theta) = -\frac{(1+2f(0)\sigma\rho)}{\sigma^2(1-\rho^2)}\sum_{t=1}^{n}\left\{m_t+\frac{(2f(0)\sigma^2+\rho\sigma)}{(2f(0)\sigma\rho+1)}l_t\right\},$$

$$(ii)\ I_{(LS,LAD)} = \frac{1}{(1-\rho^2)}(I_{LS}+I_{LAD}+\frac{4f(0)n\rho}{\sigma})$$

Proof: For part (a), we employ the Gram-Schmidt orthogonalization procedure to find two uncorrelated estimating functions. Consider the elementary estimating function ψ_t given by

$$\psi_t = k_t - \frac{\text{Cov}(m_t,k_t)}{\text{Var}(m_t)}m_t = k_t - \gamma\sigma m_t.$$

Then m_t and ψ_t are uncorrelated zero mean estimating functions with variances σ^2 and $(\kappa+2-\gamma^2)\sigma^4$, respectively. Applying Theorem 3.3.1 separately to m_t and ψ_t, the optimal estimating function becomes

$$g(\underline{y};\theta) = \sum_{t=1}^{n}\frac{\mathrm{E}\left(\frac{dm_t}{d\theta}\right)}{\mathrm{E}(m_t^2)}m_t + \sum_{t=1}^{n}\frac{\mathrm{E}\left(\frac{d\psi_t}{d\theta}\right)}{\mathrm{E}(\psi_t^2)}\psi_t.$$

Note that $\mathrm{E}\left(\frac{d\psi_t}{d\theta}\right) = -\sigma(\gamma - 2\dot{\sigma})$, where $\dot{\sigma} = \frac{d\sigma}{d\theta}$. Let $g_t = \sigma(2\dot{\sigma}\gamma - (\kappa + 2)m_t + (\gamma - 2\dot{\sigma})k_t$. It is easy to show that $E\left(\frac{\partial g_t}{\partial\theta}\right) = \sigma(4(\dot{\sigma})^2 - 4\dot{\sigma}\gamma + (\kappa + 2))$ and that $E(g_t^2) = \{\kappa + 2 - \gamma^2\}\sigma^4(4(\dot{\sigma})^2 - 4\dot{\sigma}\gamma + (\kappa + 2))$. After some algebra, the optimal combined estimating function in terms of m_t and k_t is given by

$$g(\underline{y};\theta) = \frac{1}{\sigma^3(\kappa + 2)}\left(1 - \frac{\gamma^2}{\kappa + 2}\right)^{-1}\sum_{t=1}^{n}\{\sigma(2\dot{\sigma}\gamma - (\kappa + 2))m_t + (\gamma - 2\dot{\sigma})k_t\}.$$

The martingale information follows from the definition.

For symmetric distributions (with a few exceptions), $\gamma = 0$ and hence the combination has more information. When ε_t follows a normal distribution, $\gamma = 0$ and assuming $\sigma^2(\theta) = \theta^2$, the information reduces to $I(\theta) = \frac{3n}{\theta^2}$. Hence, the combined estimating function has more information.

Similarly for part (b), by taking $\psi_t = l_t - \frac{Cov(m_t, l_t)}{V(m_t)}m_t = l_t + \frac{\rho}{\sigma}m_t$. The optimal estimating function turns out to be

$$\begin{aligned} g(\underline{y};\theta) &= \sum_{t=1}^{n}\frac{E\left(\frac{\partial m_t}{\partial\theta}\right)}{E(m_t^2)}m_t + \sum_{t=1}^{n}\frac{E\left(\frac{\partial\psi_t}{\partial\theta}\right)}{E(\psi_t^2)}\psi_t \\ &= -\frac{(1 + 2f(0)\sigma\rho)}{\sigma^2(1 - \rho^2)}\sum_{t=1}^{n}\left\{m_t + \frac{(2f(0)\sigma^2 + \rho\sigma)}{(2f(0)\sigma\rho + 1)}l_t\right\} \end{aligned} \tag{3.3.2}$$

Note: It is of interest to note that optimal estimating function using Theorem 3.1.1 part (ii), based on m_t and l_t as in (b), is given by $g^0 = \sum_{t=1}^{n}\left\{m_t + \frac{(1 + 2f(0)\sigma\rho)}{\sigma^2(1 - \rho^2)}l_t\right\}$ since $E\left(\frac{\partial m_t}{\partial\theta}\right) = -1$, $E\left(\frac{\partial l_t}{\partial\theta}\right) = -2f(0)$, $E(m_t^2) = \sigma^2$, $E(l_t^2) = 1$ and $Cov(m_t, l_t) = -\rho\sigma$. It is easy to show the the estimating function in (3.3.2) satisfies the relation $g = -\frac{(1 + 2f(0)\sigma\rho)}{\sigma^2(1 - \rho^2)}g^0$. Hence, both have the same information.

In the next section, we discuss combining estimating functions for correlated data.

3.3.1 Autoregressive processes

The following uncorrelated data example, provides the motivation for combining estimating functions for autoregressive processes.

Example 3.3.1 *Consider the Weighted Least Squares estimation(WLS) problem for estimating the variance in the linear regression model given by the following:*

$$y_t = \beta x_t + e_t,$$

where e_t is an uncorrelated sequence with zero mean and variance known up to a parameter σ^2 that depends on β. That is, $Var(e_t) = \sigma^2(\beta)\sigma_t^2$, $t = 1, \ldots, n$. An estimate of $\sigma^2(\beta)$ is given by

$$\widehat{\sigma^2(\beta)} = \frac{\sum_{t=1}^{n}(y_t - \widehat{y_t})^2}{n-2}.$$

Since $\widehat{\sigma^2(\beta)}$ depends on $\widehat{\beta}$ through the predicted value, $\widehat{y_t}$, improved estimation of β leads to improved estimation of $\sigma^2(\beta)$.

We consider the first order autoregressive process described in Heyde (1997). We let θ represent the autogressive parameter as in the example on combining estimating functions in Heyde (1997). Suppose

$$y_t = \theta y_{t-1} + \varepsilon_t, \tag{3.3.3}$$

where $\varepsilon_t = \varepsilon_t(\theta)$ are independent and identically distributed with $E(\varepsilon_t) = 0$, $E\varepsilon_t^2 = \sigma^2$, $E\varepsilon_t^4 < \infty$. Let $\sigma^2 = \sigma^2(\theta)$ and let $\rho = \frac{E|\varepsilon_t|}{\sigma}$. Furthermore, assume ε_t have density $f(.)$ where $f(0) > 0$ and $supf(x) < \infty$. Denote the skewness and excess kurtosis of ε by $\gamma = \sigma^{-3}E\varepsilon_t^3$ and $\kappa = \sigma^{-4}E\varepsilon_t^4 - 3$. We wish to estimate θ based on the sample $(y_t, 0 \leq t \leq T)$.

Consider the following set of elementary estimating functions m_t, k_t and l_t for θ given by $m_t = y_t - \theta y_{t-1}$, $k_t = (y_t - \theta y_{t-1})^2 - \sigma^2$, and $l_t = -\text{sgn}(\varepsilon_t)$.

Using Theorem 3.3.1, it is easy to show that the optimal estimating functions corresponding to m_t, k_t and l_t are $g_{LS} = -\frac{1}{\sigma^2}\sum_{t=1}^{T} y_{t-1}m_t$, $g_K = -\frac{2\dot{\sigma}}{(\kappa+2)\sigma^3}\sum_{t=1}^{T} k_t$ and $g_{LAD} = -2f(0)\sum_{t=1}^{T} l_t$, respectively. The martingale informations are $I_{LS} = \sigma^{-2}\sum_{t=1}^{T} y_{t-1}^2$, $I_K = \frac{4T(\dot{\sigma})^2}{(\kappa+2)\sigma^2}$ and $I_{LAD} = 4f^2(0)\sum_{t=1}^{T} y_{t-1}^2$, respectively (for the proofs see Appendix B.1).

The following theorem extends the results for an AR(1) process given in (3.3.3).

Theorem 3.3.3 *(a) The combined estimating function based on combining the estimating functions m_t and k_t along with its information are the following:*

(i)

$$g(\underline{y};\theta) = \frac{1}{\sigma^3(\kappa+2)}\left(1-\frac{\gamma^2}{\kappa+2}\right)^{-1}\sum_{t=1}^{T}\{\sigma(2\dot{\sigma}\gamma-(\kappa+2)y_{t-1})m_t+(y_{t-1}\gamma-2\dot{\sigma})k_t\},$$

where $\dot{\sigma} = d\sigma/d\theta$,

(ii)

$$\begin{aligned} I_g &= \frac{1}{\sigma^2(\kappa+2)}\left(1-\frac{\gamma^2}{\kappa+2}\right)^{-1} \\ &\quad \times \sum_{t=1}^{T}\{((\kappa+2)y_{t-1}-2\dot{\sigma}\gamma)y_{t-1}+2\dot{\sigma}(2\dot{\sigma}-y_{t-1}\gamma)\} \qquad (3.3.4) \\ &= \left(1-\frac{\gamma^2}{\kappa+2}\right)^{-1}\left\{I_{LS}+I_K-\frac{4\dot{\sigma}\gamma}{\sigma^2(\kappa+2)}\sum_{t=1}^{T} y_{t-1}\right\}. \qquad (3.3.5) \end{aligned}$$

(b) The combined estimating function based on combining the estimating functions m_t and l_t along with its information are the following:

(i)

$$g(\underline{y};\theta) = -\frac{1}{\sigma^2(1-\rho^2)}\sum_{t=1}^{T}\{(1+2f(0)\rho\sigma)y_{t-1}m_t + (\sigma\rho + 2f(0)\sigma^2)l_t\} \tag{3.3.6}$$

(ii)

$$I_g = \frac{1}{\sigma^2(1-\rho^2)}\sum_{t=1}^{T} y_{t-1}^2(1+4f^2(0)\sigma^2 + 4f(0)\rho\sigma). \tag{3.3.7}$$

Proof: The proof of part (a) is given in Appendix B.1 and Appendix B.2. The proof of part (b) is given in Appendix B.5.

In the case where (y_t) is stationary ($|\theta| < 1$) we have

$$I_{LS} \sim T\sigma^{-2}Ey_1^2 = T(1-\theta^2)^{-1} \text{ a.s.}$$

and

$$I_g \sim \left(1 - \frac{\gamma^2}{\kappa+2}\right)^{-1}\{I_{LS} + I_K\} \text{ a.s.}$$

as $T \to \infty$.

If on the other hand, $|\theta| \geq 1$, then $I_K/I_{LS} = \dfrac{4T(\dot{\sigma})^2}{(\kappa+2)\sum_{t=1}^{T} y_{t-1}^2} \to 0$ as $T \to \infty$, a.s. Consequently, in the case when $|\theta| \geq 1$, $I_g \sim \left(1 - \dfrac{\gamma^2}{\kappa+2}\right)^{-1} I_{LS}$ a.s. as $T \to \infty$. Thus, combining g_K with g_{LS} is advantageous in this latter instance if $\gamma \neq 0$.

We observe that

$$I_{LAD} \sim 4f^2(0)\frac{\sigma^2}{1-\theta^2}, \text{ a.s. as T} \to \infty.$$

Also, $I_{LS} \sim T(1-\theta^2)^{-1}$ as $T \to \infty$.

Hence,

$$I_g \sim \frac{1}{(1-\rho^2)}\left\{\frac{T}{1-\theta^2}+\frac{4f^2(0)T\sigma^2}{1-\theta^2}+\frac{4Tf(0)\rho\sigma}{1-\theta^2}\right\},$$

as $n \to \infty$. That is, in part (ii) of Theorem 2.2, we have

$$I_g \sim \frac{1}{(1-\rho^2)}\left\{I_{LS}+I_{LAD}+\frac{4Tf(0)\rho\sigma}{1-\theta^2}\right\},$$

as $T \to \infty$. Since $\dfrac{4Tf(0)\rho\sigma}{1-\theta^2} > 0$, and $(1-\rho^2)^{-1} > 1$, we see that combining the LS and LAD estimating functions is advantageous in this case.

In the next Section, we combine estimating functions for an autoregressive process with GARCH $(P,\ Q)$ errors.

3.3.2 Volatility Models

Consider the first order autoregressive process with GARCH$(P,\ Q)$ errors given by

$$y_t = \theta y_{t-1} + \varepsilon_t, \tag{3.3.8}$$

where $\varepsilon_t = \varepsilon_t(\theta)$ is a GARCH$(P,\ Q)$ process. That is,

$$\begin{aligned}\varepsilon_t &= \sqrt{h_t}\, Z_t\\ h_t &= \omega + \sum_{i=1}^{P}\alpha_i\varepsilon_{t-i}^2 + \sum_{j=1}^{Q}\beta_j h_{t-j},\end{aligned}$$

where Z_t are independently and identically distributed with $E(Z_t) = 0$, $E(Z_t^2) = \sigma^2 = \sigma^2(\theta)$. Thus, we have $E\varepsilon_t = 0$, $E(\varepsilon_t^2|\mathcal{F}_{t-1}^y) = \sigma^2 h_t$, $E\varepsilon_t^4 < \infty$. Denote the skewness and excess kurtosis of ε conditional on $\mathcal{F}_{t-1}^y$ by $\gamma = \sigma^{-3}h_t^{-3/2}E(\varepsilon_t^3|\mathcal{F}_{t-1}^y)$ and $\kappa = \sigma^{-4}h_t^{-2}E(\varepsilon_t^4|\mathcal{F}_{t-1}^y) - 3$. Let $f(.)$ denote the density of ε_t conditional of $\mathcal{F}_{t-1}^y$. We wish to estimate θ based on the sample $(y_t, 0 \le t \le T)$.

Let $m_t = (y_t - \theta y_{t-1})$, $k_t = \{(y_t - \theta y_{t-1})^2 - \sigma^2 h_t\}$ and $l_t = \operatorname{sgn}\left(\dfrac{\varepsilon_t}{\sqrt{h_t}}\right)$. Then the optimal estimating functions corresponding to m_t, k_t and l_t are given by , $g_{LS} =$

$-\frac{1}{\sigma^2}\sum_{t=1}^{T}\frac{y_{t-1}}{h_t}m_t$, $g_K = \frac{-2\dot{\sigma}}{(\kappa+2)\sigma^3}\sum_{t=1}^{T}\frac{k_t}{h_t}$ and $g_{LAD} = 2f(0)\sum_{t=1}^{T} l_t$, respectively. The martingale information of these are

$$I_{LS} = \sum_{t=1}^{T}\frac{1}{\sigma^2}\frac{y_{t-1}^2}{h_t}, \quad I_K = \frac{4T(\dot{\sigma})^2}{(\kappa+2)\sigma^2}, \quad \textit{and} \quad I_{LAD} = 4f^2(0)\sum_{t=1}^{T}\frac{y_{t-1}^2}{h_t},$$

respectively.

The following theorem gives the expression for the optimal combined estimating function and the corresponding martingale information.

Theorem 3.3.4 *For the autoregressive process of order with GARCH(P, Q) errors given by equation (3.3.8), we have the following:*

(a) *(i) The optimal combination based on* m_t *and* k_t *is given by*

$$\begin{aligned} g_{(LS,K)}(\underline{y};\theta) &= \frac{1}{\sigma^3(\kappa+2)}\left(1-\frac{\gamma^2}{(\kappa+2)}\right)^{-1}\sum_{t=1}^{T}\left\{\sigma(2\dot{\sigma}\gamma-(\kappa+2)\frac{y_{t-1}}{\sqrt{h_t}})m_t\right. \\ &+ \left.\left(\frac{\gamma y_{t-1}}{\sqrt{h_t}}-2\dot{\sigma}\right)k_t\right\}. \end{aligned}$$

(ii) $g_{(LS,K)}$ *has martingale information given by*

$$I_{(LS,K)} = \left(1-\frac{\gamma^2}{\kappa+2}\right)^{-1}\left\{I_{LS}+I_K-\frac{4\dot{\sigma}\gamma}{(\kappa+2)\sigma^2}\sum_{t=1}^{T}\frac{y_{t-1}}{\sqrt{h_t}}\right\}.$$

(b) *(i) Combining* m_t *and* l_t *optimally gives the following estimating function:*

$$g_{(LS,LAD)} = -\frac{1}{(1-\rho^2)}\sum_{t=1}^{T}\frac{y_{t-1}}{\sigma^2 h_t}\left\{(1+2f(0)\rho\sigma\sqrt{h_t})m_t+\sigma\sqrt{h_t}(\rho+2f(0)\sigma\sqrt{h_t})l_t\right\}$$

(ii) $g_{(LS,LAD)}$ *has martingale information given by the following expressions:*

$$\begin{aligned} I_{(LS,LAD)} &= \frac{1}{\sigma^2(1-\rho^2)}\sum_{t=1}^{T}\frac{y_{t-1}^2}{\sigma^2 h_t}(1+4f^2(0)\sigma^2 h_t+4f(0)\rho\sigma\sqrt{h_t}) \\ &= \frac{1}{(1-\rho^2)}\left\{I_{LS}+I_{LAD}+4f(0)\rho\sum_{t=1}^{T}\frac{y_{t-1}^2}{\sigma\sqrt{h_t}}\right\} \qquad (3.3.9) \end{aligned}$$

Proof: Part (a) is proved in Appendix B.3. Some of the details of the proof of part (b) are provided in Appendix B.6.

The estimate of θ can be found by solving the estimating function for θ iteratively using the Newton-Raphson procedure. For example, suppose we wished to solve the equation $g_{LS,K}(\phi) = 0$ in order to obtain an improved estimate of θ, where $\phi = (\theta, \omega, \alpha_1, \ldots, \alpha_P, \beta_1, \ldots, \beta_Q, \gamma, \kappa)'$. Starting values for each parameter are required in order to implement the Newton-Raphson procedure. The conditional least squares estimate of θ can be used as a starting value for θ. The residuals from the fitted model (where $\widehat{\theta}$ is the conditional least squares estimate) can be used to provide starting values for the remaining parameters in the following manner. Using the ARMA representation of the square of the residuals, the conditional least squares estimates of ω, $\alpha_i (i = 1, \ldots, P)$ and $\beta_j (j = 1, \ldots, Q)$ can be used as the starting values for the GARCH parameters.

Consequently, an improved estimate of $\mathrm{Var(y_t)}$ is obtained from the improved estimate of the conditional mean through the expression $\widehat{\mathrm{Var(y_t)}} = \dfrac{\widehat{\mathrm{E(h_t)}}}{1-\widehat{\theta}^2}$, where $\widehat{\mathrm{E(h_t)}} = \dfrac{\widehat{\omega}}{1 - \sigma_Z^2(\widehat{\theta})\widehat{\alpha_1} - \widehat{\beta}_1}$.

In the next Section, combining estimating functions for the class of Random Coefficient Autoregressive processes is studied.

3.3.3 Random Coefficient Autoregressive Models

For the random coefficient autoregressive model of order one for the process $\{y_t\}$.

$$y_t = (\theta + b_t)y_{t-1} + \varepsilon_t,$$

where $\{b_t\}$ and $\{\varepsilon_t\}$ are uncorrelated zero mean process with variance σ_b^2 and $\sigma^2(\theta)$, respectively. Further, denote the skewness and excess kurtosis of $\{b_t\}$ and $\{\varepsilon_t\}$ by $\sigma_b^{-3}E(b_t^3) = \gamma_b$, $\sigma^{-3}E(\varepsilon_t^3) = \gamma$, $\sigma_b^{-4}E(b_t^4) - 3 = \kappa_b$, and $\sigma^{-4}E(\varepsilon_t^4) - 3 = \kappa$, respectively.

Put

$$m_t = y_t - \theta y_{t-1} = b_t y_{t-1} + \varepsilon_t$$

and

$$k_t = m_t^2 - (\sigma_b^2 y_{t-1}^2 + \sigma^2).$$

Then using these elementary estimating function, we can show that M and K are two estimating functions for estimating θ, where

$$M = -\sum_{t=1}^{T} \frac{y_{t-1}}{\sigma_b^2 y_{t-1}^2 + \sigma^2} m_t \tag{3.3.10}$$

$$K = \sum_{t=1}^{T} \frac{-2\sigma\dot{\sigma}}{y_{t-1}^4(\kappa_b + 2)\sigma_b^4 + 4y_{t-1}^2\sigma^2\sigma_b^2 + (\kappa + 2)\sigma^4} k_t \tag{3.3.11}$$

Let $f(y_{t-1}) = y_{t-1}^4(\kappa_b + 2)\sigma_b^4 + 4y_{t-1}^2\sigma^2\sigma_b^2 + (\kappa + 2)\sigma^4$. These estimating functions have martingale information I_M and I_K given by $I_M = \sum_{t=1}^{T} \frac{y_{t-1}^2}{\sigma_b^2 y_{t-1}^2 + \sigma^2}$ and $I_K = \sum_{t=1}^{T} \frac{4\sigma^2(\dot{\sigma})^2}{f(y_{t-1})}$, respectively.

Theorem 3.3.5 *The combined estimating function is given by the following:*

$$\begin{aligned} g(\underline{y};\theta) &= \sum_{t=1}^{T} \left(1 - \frac{(y_{t-1}^3\sigma_b^3\gamma_b + \sigma^3\gamma)^2}{f(y_{t-1})(\sigma_b^2 y_{t-1}^2 + \sigma^2)}\right)^{-1} \\ &\times \left\{ \left(\frac{2\sigma\dot{\sigma}(y_{t-1}^3\sigma_b^3\gamma_b + \sigma^3\gamma)}{f(y_{t-1})(\sigma_b^2 y_{t-1}^2 + \sigma^2)} - \frac{y_{t-1}}{\sigma_b^2 y_{t-1}^2 + \sigma^2} \right) m_t + \left(\frac{y_{t-1}(y_{t-1}^3\sigma_b^3\gamma_b + \sigma^3\gamma)}{f(y_{t-1})(\sigma_b^2 y_{t-1}^2 + \sigma^2)} - \frac{2\sigma\dot{\sigma}}{f(y_{t-1})} \right) k_t \right\}, \end{aligned} \tag{3.3.12}$$

and the information associated with g is given by the following:

$$\begin{aligned} I_g &= \sum_{t=1}^{T} \left(1 - \frac{(y_{t-1}^3\sigma_b^3\gamma_b + \sigma^3\gamma)^2}{f(y_{t-1})(\sigma_b^2 y_{t-1}^2 + \sigma^2)}\right)^{-1} \\ &\times \left\{ \frac{4\sigma^2(\dot{\sigma})^2}{f(y_{t-1})} + \frac{y_{t-1}^2}{\sigma_b^2 y_{t-1}^2 + \sigma^2} - \frac{4\sigma(\dot{\sigma})y_{t-1}(y_{t-1}^3\sigma_b^3\gamma_b + \sigma^3\gamma)}{f(y_{t-1})(\sigma_b^2 y_{t-1}^2 + \sigma^2)} \right\}. \end{aligned} \tag{3.3.13}$$

Proof: Some details of the proof are given in Appendix B.4.

The optimal combined estimating functions contain more information than the component estimating functions.

Note: Setting $\sigma_b^2 = 0$ in (3.3.12), we obtain the combined estimating function in part (a) of Theorem 3.3.3.

In the next Section, hypothesis tests based on combined estimating functions are studied for the class of stationary processes with GARCH errors.

3.4 Tests for Volatility

Tests for nonlinear time series via estimating functions have been studied by Thavaneswaran (1991), Basawa (1991) and Thavaneswaran and Peiris (1998).(See also Chapter 9 in Heyde (1997) for more details). We extend the results of Thavaneswaran and Peiris (1998) to ARMA models with GARCH errors.

For a simple hypothesis which is of special interest for time-series data, say $H_0 : \theta = \theta_0$ against any possible alternative, we propose a test statistic T_g based on an unbiased estimating function g as

$$T_g = \frac{\hat{\theta}_g - \theta_0}{\sqrt{I_g^{-1}}}, \tag{3.4.1}$$

where $\hat{\theta}_g$ is the estimate found by solving $g(\underline{y}; \theta) = 0$. The following theorem establishes the superiority of Godambe's approach.

Theorem 3.4.1 *(Thavaneswaran and Peiris (1998)) The power of a test based on the optimal estimating function g_n^* against a particular alternative is bounded below by the power of any unbiased estimating function g i.e. Power of $T_{g^*} \geq$ Power of T_g.*

In the next set of examples, the martingale information of the optimal estimating function as well as the martingale information of the least squares type estimating function are derived. Tests based on the aforementioned estimating functions are derived and compared.

Example 3.4.1 *Consider the simple linear regression model through the origin,* $y_t = \beta x_t + \varepsilon_t$, $t = 1, \ldots, n$. *Assume that the errors are independent with* $E(\varepsilon_t) = 0$ *and* $Var(\varepsilon_t) = \frac{\sigma^2}{x_t^2}$. *The usual least squares estimate of* β *is* $\hat{\beta} = \frac{\sum_{t=1}^n y_t x_t}{\sum_{t=1}^n x_t^2}$. *In contrast, the weighted least squares estimate of* β *is* $\hat{\beta}_{WLS} = \frac{\sum_{t=1}^n x_t^3 y_t}{\sum_{t=1}^n x_t^4}$. *Using the Cauchy-Schwarz inequality,* $\left(\sum x_t^2.1\right)^2 \leq \sum x_t^4 \sum 1 = n \sum x_t^4$. *The least squares estimating function for* β *is* $g_{LS} = -\sum_{t=1}^n x_t m_t$, *where* $m_t = y_t - \beta x_t$. *It is easy to show that the information associated with least squares estimating function is* $I_{LS} = \frac{\sum_{t=1}^n x_t^2}{n\sigma^2}$ *while the information associated with the optimal estimating function is* $I_{opt} = \frac{\sum_{t=1}^n x_t^4}{\sigma^2}$. *Hence,* $I_{LS} \leq I_{opt}$.

Example 3.4.2 *Consider the autoregressive process* $y_t = \theta y_{t-1} + \varepsilon_t$, *where* ε_t *are i.i.d. with* $E(\varepsilon_t) = 0$ *and* $Var(\varepsilon_t) = \frac{\sigma^2}{y_{t-1}^2}$. *The conditional least squares estimating function is given by* $g_{CLS} = -\sum_{t=2}^n y_{t-1} m_t$ *where* $m_t = y_t - \theta y_{t-1}$. *The information associated with the conditional least squares estimating function is given by* $I_{CLS} = \frac{\left(\sum y_{t-1}^2\right)^2}{(n-1)\sigma^2}$. *On the other hand, the optimal estimating function based on elementary estimating functions* m_t *is given by* $g_{opt} = -\sum_{t=2}^n \frac{y_{t-1}^3}{\sigma^2} m_t$. *The martingale information associated with* g_{opt} *is* $I_{opt} = \frac{\sum_{t=2}^n y_{t-1}^4}{\sigma^2}$. *An application of the Cauchy-Schwarz inequality once again shows that* $I_{CLS} \leq I_{opt}$. *The test based on the statistic* $T_{opt} = \frac{\hat{\theta}_{opt} - \theta_0}{\sqrt{I_{opt}^{-1}}}$ *is more powerful.*

Example 3.4.3 *Suppose $\{y_t\}$ is the following autoregressive process:*

$$y_t = \theta y_{t-1} + \varepsilon_{t-1}^2 \varepsilon_t \tag{3.4.2}$$

where ε_t are uncorrelated with mean zero and variance σ_ε^2. The optimal estimating function in the Godambe(1985) sense based on elementary estimating functions $m_t = y_t - \theta y_{t-1}$ is $g = -\sum_{t=2}^{n} \frac{y_{t-1}}{3\sigma_\varepsilon^6} m_t$. To test $H_0 : \theta = \theta_0$ against $H_a : \theta \neq \theta_0$, we propose the statistic $T = \frac{\hat{\theta} - \theta_0}{\sqrt{V}}$ where $\hat{\theta}$ is the solution of the optimal estimating equation and $V = \left[\sum_{t=2}^{n} \frac{y_{t-1}^2}{3\sigma_\varepsilon^6} \right]^{-1}$. Using a weighted least squares approach, the optimal estimating function is $g_{opt} = -\sum_{t=2}^{n} \frac{y_{t-1}^3}{3\sigma_\varepsilon^6} m_t$ with martingale information $I_{opt} = \sum_{t=2}^{2} \frac{y_{t-1}^4}{3\sigma_\varepsilon^6}$. Since $\sum y_{t-1}^4 = \sum y_{t-1}^2 y_{t-1}^2 \geq \sqrt{\left(\sum y_{t-1}^2\right)^2} = \sum y_{t-1}^2$, $I_{opt} \geq I_g$ and the test based on the statistic $T = \frac{\hat{\theta}_g - \theta_0}{\sqrt{I_{opt}^{-1}}}$ is more powerful.

Example 3.4.4 *Consider an AR(1)-GARCH(1,1) model of the form:*

$$y_t = \theta y_{t-1} + \varepsilon_t \tag{3.4.3}$$

where $|\theta| < 1$ is a zero-mean GARCH(P, Q) process given by the following:

$$\begin{aligned} \varepsilon_t &= \sqrt{h_t} Z_t, \\ h_t &= \omega + \sum_{i=1}^{P} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{Q} \beta_j h_{t-j} \end{aligned}$$

and Z_t is an uncorrelated zero-mean sequence with variance $\sigma^2 = \sigma^2(\theta)$ and finite fourth moment. Let σ_ε^2 be the variance of ε_t. Using the elementary estimating functions $m_t = y_t - \theta y_{t-1}$, the estimating function based on the least squares approach is $g = -\sum_{t=2}^{n} y_{t-1} m_t$ while the optimal estimating function is $g_{opt} = -\sum_{t=2}^{n} \frac{y_{t-1}}{\sigma^2 h_t} m_t$. Let θ^* denote the solution of the equation $g_{opt} = 0$. In testing the hypothesis H_0 :

$\theta = \theta_0$ *against* $H_a : \theta \neq \theta_0$, *the test based on the statistic* $T = \dfrac{\theta^* - \theta_0}{\sqrt{V}}$ *where* $V = \left[\sum_{t=2}^{n} \dfrac{y_{t-1}^2}{\sigma^2 h_t}\right]^{-1}$ *is more powerful than the test based on the least squares estimating function.*

Example 3.4.5 *Consider the autoregressive order one process with power GARCH(1,1) innovations given by*

$$y_t = \theta y_{t-1} + \varepsilon_t \tag{3.4.4}$$

where

$$\varepsilon_t = \sqrt{h_t}\tilde{Z}_t, \tag{3.4.5}$$

$$h_t = \omega + \alpha_1|\varepsilon_{t-1}| + \beta_1 h_{t-1} \tag{3.4.6}$$

with $\tilde{Z}_t = \dfrac{|Z_t| - E|Z_t|}{\sqrt{Var(|Z_t|)}}$. *Based on the estimating function* $m_t = y_t - \theta y_{t-1}$, *the optimal estimating function is* $g_{opt} = -\sum_{t=2}^{n} \dfrac{y_{t-1}}{h_t} m_t$. *Let* θ^* *denote the solution of the equation* $g_{opt} = 0$. *In order to test the hypothesis* $H_0 : \theta = \theta_0$ *against* $H_a : \theta \neq \theta_0$, *use the test statistic* $T = \dfrac{\theta^* - \theta_0}{\sqrt{I_{opt}^{-1}}}$ *where* $I_{opt} = \left[\sum_{t=2}^{n} \dfrac{y_{t-1}^2}{h_t}\right]$. *The test based on the optimal estimating function is more powerful than the test based on the least squares estimating function* $g = \sum_{t=2}^{n} y_{t-1} m_t$.

Example 3.4.6 *Let* y_t *be a stationary RCA*(1) *process with normal GARCH(1,1) errors given by*

$$y_t = (\theta + b_t)y_{t-1} + \varepsilon_t \tag{3.4.7}$$

where

$$\varepsilon_t = \sqrt{h_t}Z_t, \tag{3.4.8}$$

$$h_t = \omega + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1}, \tag{3.4.9}$$

$Z_t \sim N(0,1)(i.i.d.)$ and uncorrelated with $b_t \sim N(0,\sigma_b^2)(i.i.d.)$. Based on the elementary estimating function $m_t = y_t - \theta y_{t-1}$, the optimal estimating function is given by $g_{opt} = -\sum_{t=2}^{n} \frac{y_{t-1}}{\sigma_b^2 y_{t-1}^2 + \sigma_\varepsilon^2} m_t$ where $\sigma_\varepsilon^2 = \frac{\omega}{1-(\alpha_1+\beta_1)}$. Let θ^* denote the solution of the optimal estimating function equation $g = 0$. The test statistic for the one-sided test of hypothesis $H_0 : \theta = \theta_0$ against $H_a : \theta \neq \theta_0$ is $T = \frac{\theta^* - \theta_0}{\sqrt{I_{opt}^{-1}}}$ where $I_{opt} = \left[\sum_{t=2}^{n} \frac{y_{t-1}^2}{\sigma_b^2 y_{t-1}^2 + \sigma_\varepsilon^2}\right]$ is the optimal estimating function information. This test is more powerful than the one based on the least squares estimating function given by $g = -\sum_{t=2}^{n} y_{t-1} m_t$.

Example 3.4.7 *Consider an AR model with Sign-switching GARCH errors given by*

$$y_t = \theta y_{t-1} + \varepsilon_t$$

where

$$\varepsilon_t = \sqrt{h_t} Z_t \qquad (3.4.10)$$

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} + \Theta_1 S_{t-1} \qquad (3.4.11)$$

$Z_t \sim N(0,1)(i.i.d.)$ and $S_t = I_{(\varepsilon_t > 0)} - I_{(\varepsilon_t < 0)}$. It is easy to show that the optimal estimating function for θ is given by $g = -\sum_{t=2}^{n} \frac{y_{t-1}}{h_t} m_t$ where $m_t = y_t - \theta y_{t-1}$. Let θ^* denote the solution of the optimal estimating function equation $g_{opt} = 0$. The test statistic for testing $H_0 : \theta = \theta_0$ against $H_a : \theta \neq \theta_0$ is $T = \frac{\theta^* - \theta_0}{\sqrt{I_{opt}^{-1}}}$ where $I_{opt} = \left[\sum_{t=2}^{n} \frac{y_{t-1}^2}{h_t}\right]$. This test is more powerful than the one based on the least squares estimating function given by $g = -\sum_{t=2}^{n} y_{t-1} m_t$.

Example 3.4.8 *Consider the autoregressive model with GARCH type innovations* *given by*

$$y_t = \theta y_{t-1} + \varepsilon_t,$$

where

$$\varepsilon_t = \sqrt{h_t} Z_t \tag{3.4.12}$$

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \gamma_1 \varepsilon_{t-1}^2 I_{(\varepsilon_{t-1}<0)} + \beta_1 h_{t-1} \tag{3.4.13}$$

where $Z_t \sim N(0,1)(i.i.d.)$ *and* I *denotes an indicator function. Based on the elementary estimating function* $m_t = y_t - \theta y_{t-1}$ *for* θ*, the optimal estimating function is* $g_{opt} = -\sum_{t=2}^{n} \frac{y_{t-1}}{h_t} m_t$ *while the least squares estimating function based on* m_t *is* $g = -\sum_{t=2}^{n} y_{t-1} m_t$*. For testing* $H_0 : \theta = \theta_0$ *against* $H_a : \theta \neq \theta_0$*, the test based on* $T = \frac{\theta^* - \theta_0}{\sqrt{V}}$ *where* θ^* *is the solution of* $g_{opt} = 0$ *and* $I_{opt} = \left[\sum_{t=2}^{n} \frac{y_{t-1}^2}{h_t}\right]$ *is more powerful than the test based on the least squares estimating function.*

Following Heyde (1997), we study tests based on combined estimating functions in Example 3.4.9.

Example 3.4.9 *The optimal estimate based on the estimating function in Theorem 3.3.4 part (a) is given by solving the equation* $g_{LS,K} = 0$ *iteratively. Denote the optimal estimate by* θ^**. Then the test statistic based on the optimal estimate for testing* $H_0 : \theta = \theta_0$ *against* $H_a : \theta \neq \theta_0$ *is* $T = \frac{\theta^* - \theta_0}{\sqrt{I_{opt}^{-1}}}$ *where*

$$I_{opt} = \left(1 - \frac{\gamma^2}{\kappa + 2}\right)^{-1} \left[I_{LS} + I_K - \frac{4\dot{\sigma}\gamma}{(\kappa+2)\sigma^2} \sum_{t=2}^{T} \frac{y_{t-1}}{\sqrt{h_t}}\right].$$

The optimal estimate based on the optimal combined estimating function in Theorem 3.3.4 part (b) is found by iteratively solving the equation $g = 0$*. Denote the root* *of the previous equation by $\tilde{\theta}$. The test statistic becomes* $T = \dfrac{\tilde{\theta} - \theta_0}{\sqrt{I_{opt}^{-1}}}$ *where*

$$I_{opt} = \frac{1}{(1-\rho^2)}\left[I_{LS} + I_{LAD} + 4f(0)\rho\sum_{t=2}^{T}\frac{y_{t-1}^2}{\sqrt{h_t}}\right].$$

The power of the test for the autoregressive parameter in an AR(1) model with GARCH(1, 1) errors using the least squares and the optimal estimating functions is studied in Chapter 5 through simulation studies.

3.5 Conclusions

Applications of combining estimating functions in volatility modeling have been studied. The correlation between the LS and LAD estimating functions turns out to be useful for identifying the error distribution in the class of GARCH models. Combined estimating functions for the class of autoregressive processes with GARCH errors and the class of RCA volatility models are shown to have more information than the component estimating functions. By combining estimating functions, an improved estimate of the conditional mean is obtained leading to an improved estimate of variance such as $E(h_t)$ or $E(h_t|\mathcal{F}_{t-1}^y)$.

Examples of model identification for GARCH models using real financial data will be given in Chapter 5. The superiority of hypothesis tests for the class of autoregressive processes with GARCH errors based on the optimal estimating function over tests based the least squares estimating function will be demonstrated in simulation studies in Chapter 5. In the next chapter, the forecasting problem for stationary processes with GARCH errors is studied.

Chapter 4

Forecasting for GARCH Models

In this chapter, derivations of forecasts and the forecast error variances for various classes of ARMA(p, q) models with GARCH errors are given. The latter part of this chapter includes theorems related to volatility estimation in the presence of structural change. We end with a Conclusions section.

4.1 Forecasting for ARMA Processes with Heteroscedastic Errors

Recall the GARCH(P, Q) process ε_t defined as

$$\varepsilon_t = \sqrt{h_t}\, Z_t \tag{4.1.1}$$

$$h_t = \omega + \sum_{i=1}^{P} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{Q} \beta_j h_{t-j}. \tag{4.1.2}$$

The following theorem enables us to calculate the variance of the l-steps ahead forecast error for a zero mean GARCH(P, Q) process.

Theorem 4.1.1 *For the GARCH(P, Q) process specified by (4.1.1) and (4.1.2), under the stationarity assumptions and finite fourth moment, we have the following:*

(a) $Var(\varepsilon_t^2) = (K^{(\varepsilon)} - 1)[Var(\varepsilon_t)]^2$. *That is, the square of the coefficient of variation for the squared process* ε_t^2 *is* $K^{(\varepsilon)} - 1$.

(b) *Let $\varepsilon_n^2(l)$ be the forecast of ε_{n+l}^2 based on n observations $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$. The l-steps ahead forecast error variance $Var(e_n(l))$ is given by*

$$Var(e_n(l)) = \frac{\left(\dfrac{\omega}{1 - \Phi_1 - \Phi_2 - \ldots - \Phi_R}\right)^2}{\sum_{j=0}^{\infty} \Psi_j^2} [K^{(\varepsilon)} - 1] \left[1 + \sum_{j=1}^{l-1} \Psi_j^2\right].$$

Proof: The proof of part (a) follows from the definition of coefficient of variation for ε_t^2. The proof of part (b) is given in Thavaneswaran, Appadoo and Peiris (2005).

As we have seen in Chapter 2, the simplest volatility process $y_t = \varepsilon_{t-1}^2 \varepsilon_t$, where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)(i.i.d.)$ has a kurtosis of 35. This process can be used as the innovation term in AR(1) and RCA(1) processes to generate volatility models with time-varying conditional mean and conditional variance. The following lemmas give the variance of the l-steps ahead forecast error for such processes.

Lemma 4.1.1 *For the simple volatility process of the form*

$$y_t - \mu = \phi\,(y_{t-1} - \mu) + \varepsilon_{t-1}^2 \varepsilon_t \tag{4.1.3}$$

under the stationarity assumptions that $|\phi| < 1$, ε_t i.i.d $N(0, \sigma_\varepsilon^2)$, then

(a) *Let $y_n(l)$ be the forecast of y_{n+l} based on n observations $y_1, y_2, \ldots, y_n$. Then $y_n(l) = \mu + \phi^l(y_n - \mu)$ and,*

(b) *the l-steps ahead forecast error variance for the observed process $\{y_t\}$ is $Var(e_n(l)) = 3\sigma_\varepsilon^6 \sum_{j=0}^{l-1} \phi^{2j} = 3\sigma_\varepsilon^6 \dfrac{1 - \phi^{2l}}{1 - \phi^2}$.*

Lemma 4.1.2 *Let $\{y_t\}$ be an RCA(1) process given by*

$$y_t = (\phi + b_t)y_{t-1} + \varepsilon_{t-1}^2 \varepsilon_t, \tag{4.1.4}$$

where $\{\varepsilon_t\}$ and $\{b_t\}$ are uncorrelated zero-mean Gaussian processes with variances σ_ε^2 and σ_b^2, respectively. Under the stationarity assumption that $\phi^2 + \sigma_b^2 < 1$, we have the following:

(i) The variance of the l-steps ahead forecast error $e_n(l)$ is given by

$$Var(e_n(l)) = E(y_{n+l} - y_n(l))^2 = \frac{3\sigma_\varepsilon^6(1-\phi^2)}{1-\phi^2-\sigma_b^2}(1 + \ldots + \phi^{2(l-1)}) = \frac{3\sigma_\varepsilon^6(1-\phi^{2l})}{1-\phi^2-\sigma_b^2}.$$

For financial time series data, appropriate models of log returns are the ARMA(p, q) processes with GARCH(P, Q) errors. In this Section, we study the forecasting problem for stationary ARMA(p, q) processes with GARCH(P, Q) errors. In the following theorem, we study the second and fourth moment of the l-steps ahead forecast error of $\{y_t\}$.

Theorem 4.1.2 *For any second order stationary process $\{y_t\}$ with GARCH (P, Q) errors given by*

$$y_t - \mu = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \tag{4.1.5}$$

$$\varepsilon_t = \sqrt{h_t}\, Z_t \tag{4.1.6}$$

$$h_t = \omega + \sum_{i=1}^{P} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{Q} \beta_j h_{t-j} \tag{4.1.7}$$

where Z_t is an uncorrelated sequence with zero mean and unit variance, we have the following:

(a) Under suitable stationarity conditions for a GARCH process (stated in Chapter 2), the kurtosis of the $\{y_t\}$ process is given by

$$K^{(y)} = \frac{K^{(\varepsilon)}\left[\sum_{j=0}^{\infty}\psi_j^4\right] + 6\sum_{i<j}^{\infty}\psi_i^2\psi_j^2}{\left(\sum_{j=0}^{\infty}\psi_j^2\right)^2} \tag{4.1.8}$$

where,

$$K^{(\varepsilon)} = \frac{E(Z_t^4)}{E(Z_t^4) - [E(Z_t^4) - 1]\sum_{j=0}^{\infty}\Psi_j^2}, \tag{4.1.9}$$

and Ψ_j's are found from the ARMA representation of $\{\varepsilon_t^2\}$.

(b) For the l-steps ahead forecast $y_n(l)$, the variance of the forecast error $e_n(l) = y_{n+l} - y_n(l)$ is $E[y_{n+l} - y_n(l)]^2 = \frac{\omega}{1 - \Phi_1 - \Phi_2 - \ldots \Phi_R}\sum_{j=0}^{l-1}\psi_j^2$, where $R = \max(P, Q)$ and $\Phi_i = \alpha_i + \beta_i$, $i = 1, \ldots, R$.

(c) $E[y_{n+l} - y_n(l)]^4 = \sigma_\varepsilon^4\left(K^{(\varepsilon)}\sum_{j=0}^{l-1}\psi_j^4 + 6\sum_{i<j}^{l-1}\psi_i^2\psi_j^2\right)$ where $\sigma_\varepsilon^4 = \frac{\omega^2}{\left(1 - \sum_{i=1}^{R}\Phi_i\right)^2}$ and where $K^{(\varepsilon)}$ is given as in part (a).

Proof: Proof of part (a) is given in Chapter 2. Part (b) follows from the fact that for a stationary ARMA process with error variance $\mathrm{Var}(e_n(l)) = \sigma_\varepsilon^2(1+\psi_1^2+\cdots+\psi_{l-1}^2)$, and the proof of part (c) is similar to that of part (a).

Note: When all the ARMA parameters are zero, then $K^{(y)}$ turns out to be the kurtosis of the zero mean GARCH process. When all Φ_i in the GARCH process are zero, then $K^{(y)}$ turns out to be the kurtosis of the ARMA process. The next set of examples illustrate Theorem 4.1.2.

Example 4.1.1 *Consider an AR-GARCH model of the form:*

$$y_t - \mu \;=\; \phi(y_{t-1} - \mu) + \varepsilon_t \tag{4.1.10}$$

where $|\phi| < 1$, $\{\varepsilon_t\}$ is a zero mean GARCH(P, Q) process given by expression (4.1.1), the conditional variance h_t is given by (4.1.2) and $\{Z_t\}$ is an uncorrelated zero mean sequence with unit variance and finite fourth moment. Let σ_ε^2 be the variance of $\{\varepsilon_t\}$. Then $\psi_j = \phi^j$, $\sum_{j=0}^{\infty} \psi_j^2 = \frac{1}{1-\phi^2}$ and $\sum_{i<j}^{\infty} \psi_i^2\psi_j^2 = \frac{\phi^2}{(1-\phi^2)(1-\phi^4)}$.
Hence, Var$(y_t) = \frac{\sigma_\varepsilon^2}{1-\phi^2}$, and $K^{(y)} = \frac{6\phi^2 + (1-\phi^2)K^{(\varepsilon)}}{1+\phi^2}$, where $K^{(\varepsilon)} = \frac{E(Z_t^4)}{E(Z_t^4) - [E(Z_t^4) - 1]\sum_{j=0}^{\infty} \Psi_j^2}$.
From Theorem 4.1.2, when $\{\varepsilon_t\}$ is a GARCH(P, Q) process,

(i) $E[y_{n+l} - y_n(l)]^2 = \frac{\omega}{1 - \Phi_1 - \ldots \Phi_R} \frac{1-\phi^{2l}}{1-\phi^2}$, *where* $\Phi_i = \alpha_i + \beta_i$, $i = 1, \ldots, R$, *and* $R = \max(P, Q)$.

(ii) $E[y_{n+l} - y_n(l)]^4 = \sigma_\varepsilon^4 \left(K^{(\varepsilon)} \sum_{j=0}^{l-1} \psi_j^4 + 6 \sum_{i<j}^{l-1} \psi_i^2\psi_j^2 \right)$, *where $K^{(\varepsilon)}$ is the kurtosis of the GARCH(P, Q) process, ε_t.*

Example 4.1.2 *Consider an ARMA-GARCH model given by*

$$(y_t - \mu) \;=\; \phi(y_{t-1} - \mu) + \varepsilon_t - \theta\varepsilon_{t-1} \tag{4.1.11}$$

where $|\phi| < 1$, $|\theta| < 1$, ε_t is a GARCH process given by (4.1.1) and (4.1.2), Z_t is an uncorrelated zero mean sequence with unit variance and finite fourth moment. Let σ_ε^2 be the variance of $\{\varepsilon_t\}$. Then, $\sum_{j=0}^{\infty} \psi_j^2 = \frac{1 - 2\phi\theta + \theta^2}{1-\phi^2}$, $\sum_{j=0}^{\infty} \psi_j^4 = \frac{1 - 4\phi^3\theta + 6\phi^2\theta^2 - 4\phi\theta^3 + \theta^4}{1-\phi^4}$, and $\sum_{i<j}^{\infty} \psi_i^2\psi_j^2 = \frac{(\phi-\theta)^2}{1-\phi^2} + \frac{(\phi-\theta)^4}{1-\phi^2}\frac{\phi^2}{1-\phi^4}$. Hence,

$$Var(y_t) = \frac{\omega(1 - 2\phi\theta + \theta^2)}{(1 - \sum_{i=1}^{R} \Phi_i)(1 - \phi^2)},$$

and

$$K^{(y)} = K^{(\varepsilon)} \left(\frac{(1 - 4\phi^3\theta + 6\phi^2\theta^2 - 4\phi\theta^3 + \theta^4)(1 - \phi^2)}{(1 + \phi^2)(1 - 2\phi\theta + \theta^2)^2} \right) + 6 \frac{(1 - \phi^4)(\phi - \theta)^2 + (\phi - \theta)^4\phi^2}{(1 + \phi^2)(1 - 2\phi\theta + \theta^2)^2}.$$

The variance of the l-steps ahead forecast error $e_n(l) = y_{n+l} - y_n(l)$ is given by $Var(e_n(l)) = \frac{\omega}{1 - \Phi_1 - \ldots - \Phi_R}[1 + (\phi - \theta)^2(1 + \phi^2 + \ldots + \phi^{2(l-2)})]$. In addition, $E[y_{n+l} - y_n(l)]^4 = \sigma_\varepsilon^4 \left(K^{(\varepsilon)} \sum_{j=0}^{l-1} \psi_j^4 + 6 \sum_{i<j}^{l-1} \psi_i^2 \psi_j^2 \right)$, where $K^{(\varepsilon)}$ is the kurtosis of the GARCH(P, Q) process, ε_t.

Random coefficient autoregressive (RCA) time series have been introduced by Nicholls and Quinn (1982) in order to model leptokurtosis and time-varying volatility. Some of their properties have been recently studied by Appadoo, Ghahramani and Thavaneswaran (2005). We next consider the forecasting problem for a stationary RCA(1) model with normal GARCH(1, 1) innovations.

Example 4.1.3 *Let $\{y_t\}$ be a stationary RCA(1) process with normal GARCH(1, 1) errors given by*

$$y_t = (\phi + b_t)y_{t-1} + \varepsilon_t, \tag{4.1.12}$$

where

$$\varepsilon_t = \sqrt{h_t} Z_t, \tag{4.1.13}$$

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \tag{4.1.14}$$

$Z_t \sim N(0,1)$ (i.i.d.) and uncorrelated with $b_t \sim N(0,\sigma_b^2)$(i.i.d.). Then the variance of the l-steps ahead forecast error is $E[y_{n+l} - y_n(l)]^2 = \dfrac{\omega(1-\phi^{2l})}{[1-(\alpha_1+\beta_1)](1-\phi^2-\sigma_b^2)}$. *We also have the following:* $E[y_{n+l} - y_n(l)]^4 = \sigma_\varepsilon^4 \left(K^{(\varepsilon)} \sum_{i=0}^{l-1} \psi_i^4 + 6 \sum_{i<j}^{l-1} \psi_i^2 \psi_j^2 \right)$ *where* $\sigma_\varepsilon^4 = \left(\dfrac{\omega(1-\phi^2)}{[1-(\alpha_1+\beta_1)](1-\phi^2-\sigma_b^2)} \right)^2$ *and* $K^{(\varepsilon)} = \dfrac{3(1-(\alpha_1+\beta_1))^2}{[1-(\alpha_1+\beta_1)^2 - 2\alpha_1^2]}$.

Example 4.1.4 *Consider the AR(1) model with Sign-switching GARCH errors (see Fornari and Mele (1997)) given by*

$$y_t - \mu = \phi(y_{t-1} - \mu) + \varepsilon_t$$

where, $\varepsilon_t = \sqrt{h_t} Z_t \quad \varepsilon_t | I_{t-1} \sim N(0, h_t)$

$h_t = \omega + \sum_{i=1}^{P} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{Q} \beta_j h_{t-j} + \sum_{x=1}^{m} \Theta_x s_{t-x}$

$$s_t = \begin{cases} +1 & \text{if } \varepsilon_t > 0 \\ 0 & \text{if } \varepsilon_t = 0 \\ -1 & \text{if } \varepsilon_t < 0 \end{cases}$$

where P, Q and $m > 0$, w, $\alpha_i, (i = 1, 2, \ldots P)$, $\beta_j, (j = 1, 2, \ldots Q)$ and $\Theta_x, (x = 1, 2, \ldots m)$ are real parameters, satisfying $w > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ and $|\sum_x \Theta_x| \leq \omega$. Let $u_t = \varepsilon_t^2 - h_t$ be the martingale difference and let σ_u^2 be the variance of u_t, (4.1.13) and (4.1.14) could be written as:

$$\Phi(B)\varepsilon_t^2 = \omega + \beta(B)u_t + \sum_{x=1}^{m} \Theta_x s_{t-x}$$

where $\Phi(B) = 1 - \sum_{i=1}^{R} \Phi_i B^i$ *and* $\Phi_i = (\alpha_i + \beta_i)$ $\beta(B) = 1 - \sum_{j=1}^{Q} \beta_j B^j$ *and* $R = max(P, Q)$ $Var(\varepsilon_t^2) = \sum_{i=0}^{\infty} \Psi_i^2 \sigma_u^2$ *and* $\sum_{i=0}^{\infty} \Psi_i^2 < \infty$ *where the Ψ_i-weights are obtained from the relation* $\Psi(B)\Phi(B) = \beta(B)$ *with* $\Psi(B) = 1 - \sum_{i=0}^{\infty} \Psi_i B^i$. *Under the stationarity assumptions similar to those given for GARCH models, for the special case of the Gaussian AR(1) model with Sign-switching GARCH(1, 1) errors, we have the following:*

(i) The variance of the l-steps ahead forecast error is $E[y_{n+l}-y_n(l)]^2 = \frac{\omega}{1-\Phi_1}\sum_{j=0}^{l-1}\phi^{2j}$, *where* $\Phi_1 = \alpha_1 + \beta_1$, *and*

(ii) $E[y_{n+l} - y_n(l)]^4 = \sigma_\varepsilon^4\left(K^{(\varepsilon)}\sum_{j=0}^{l-1}\psi_j^4 + 6\sum_{i<j}^{l-1}\psi_i^2\psi_j^2\right)$, *where*
$\sigma_\varepsilon^4 = \frac{\omega^2}{[1-(\alpha_1+\beta_1)]^2}$ *and* $K^{(\varepsilon)} = \frac{[\omega^2(1+\Phi_1)+\Theta_1^2(1-\Phi_1)](1-\Phi_1)}{\omega^2(1-3\alpha_1^2-\beta_1^2-2\alpha_1\beta_1)}$.

Note: $E(\varepsilon_t^2) = E(h_t)$ *and* $E(h_t) = \omega + \alpha_1 E(\varepsilon_{t-1}^2) + \Phi E(s_{t-1}) + \beta_1 E(h_{t-1})$. *Since* $E(s_{t-1}) = 0$, $E(\varepsilon_t^2) = \frac{\omega}{1-\Phi_1}$, *where* $\Phi_1 = \alpha_1 + \beta_1$.

$E(h_t^2) = \omega^2 + \alpha_1^2E(\varepsilon_{t-1}^4) + \beta_1^2E(h_{t-1}^2) + \Theta_1^2E(s_{t-1}^2) + 2\omega\alpha_1E(\varepsilon_{t-1}^2) + 2\omega\beta_1E(h_{t-1}) + 2\alpha_1\beta_1E(h_{t-1}^2)$.

Assuming $\{\varepsilon_t\}$ *is stationary and* $Z_t \sim N(0,1)$, $E(h_t^2) = \frac{\omega^2+\Theta_1^2+2\omega\sigma_\varepsilon^2\Phi_1}{1-3\alpha_1^2-\beta_1^2-2\alpha_1\beta_1}$ *where* $\sigma_\varepsilon^2 = E(\varepsilon_t^2) = \frac{\omega}{1-\Phi_1}$.

Example 4.1.5 *Consider the autoregressive model with GARCH type innovations that allow for the effect of a negative shock to be modelled given by*

$$y_t - \mu = \phi(y_{t-1} - \mu) + \varepsilon_t,$$

where

$$\varepsilon_t = \sqrt{h_t}Z_t, \tag{4.1.15}$$

$$h_t = \omega + \alpha_1\varepsilon_{t-1}^2 + \gamma_1\varepsilon_{t-1}^2 I(\varepsilon_{t-1} < 0) + \beta_1 h_{t-1} \tag{4.1.16}$$

in which Z_t *is a sequence of i.i.d. random variables (shocks), with zero mean and unit variance, and the indicator function* I *equals 1 when* $\varepsilon_{t-1} < 0$, *and 0 otherwise. Sufficient conditions for positivity of the conditional variance are that* $\omega > 0$, $\alpha > 0$, $\gamma > 0$ *and* $\beta > 0$. *Then it can be easily shown that* *(i) The variance of the l-steps ahead forecast error is*

$$E[y_{n+l} - y_n(l)]^2 = \frac{\omega}{1-(\alpha_1+\beta_1+\frac{\gamma_1}{2})}\sum_{j=0}^{l-1}\phi^{2j},$$

(ii) $E[y_{n+l}-y_n(l)]^4 = \sigma_\varepsilon^4\left(K^{(\varepsilon)}\sum_{j=0}^{l-1}\psi_j^4 + 6\sum_{i<j}^{l-1}\psi_i^2\psi_j^2\right)$ *where* $\sigma_\varepsilon^4 = \dfrac{\omega^2}{\left[1-(\alpha_1+\beta_1+\frac{\gamma_1}{2})\right]^2}$,

and

$$K^{(\varepsilon)} = \frac{3\left[1-\left(\alpha_1+\beta_1+\frac{\gamma_1}{2}\right)^2\right]}{\left(1-3\alpha_1^2-\frac{3}{2}\gamma_1^2-\beta_1^2-3\alpha_1\gamma_1-2\alpha_1\beta_1-\gamma_1\beta_1\right)}.$$

4.1.1 Application to GARCH Forecasts

Accurate forecasts of volatility are needed in various areas of financial time series modelling and forecasting. For example, the classic Black-Scholes formula for a European call option gives the cost of an option based on the Geometric Brownian motion model for the stock price $S(t)$ as

$$c(S,t) = S\Phi(d) - Ke^{-r(T-t)}\Phi(d-\sigma\sqrt{T-t})$$

where

$$d = \frac{\log[S/K]+r(T-t)}{\sigma\sqrt{T-t}} + \frac{\sigma\sqrt{T-t}}{2}$$

$S(t) = S$ is the price of the underlying security price at time t usually $t = 0$, K is the strike price, T is the maturity date, r is the spot rate, σ is the volatility (i.e. the instantaneous standard deviation of the rate of return of the underlying security), and $\Phi(x)$ is the distribution function of the standard normal variate. The option price depends on the initial stock price S, strike price K, maturity date T, spot rate r and the the 'unknown' volatility parameter σ and hence the better estimate/forecast of the volatility will improve the performance of the option pricing formula. The

Black-Scholes formula assumes that both the spot rate r and the volatility σ are constant. When these assumptions are false, use of the Black-Scholes formula can lead to serious errors. For this reason, the use of GARCH models for pricing options is a promising area of research. (see Ruppert (2004) for a recent survey of pricing options under generalized GARCH processes). In the following examples, we compute l-steps ahead forecast standard errors ($l = 1, 2, 3$) for log-return series of IBM and GE stock taken from *http://www.gsb.uchicago.edu/fac/ruey.tsay/teaching/fts.* All computations have been carried out using S-Plus©(Insightful Corp., Seattle, WA, USA) and SAS/ETS©(SAS Institute Inc., Cary, NC, USA) software.

In the following two examples, we show that forecast error variances from autoregressive models with GARCH errors may differ from those obtained from the usual autoregressive models.

Example 4.1.6 *(IBM)*

Based on examining the sample ACF and PACF of monthly log returns of IBM stock $\{y_t\}$, we fit an AR(1) model with normal-GARCH$(1,1)$ errors. The estimated AR(1) model with GARCH$(1,1)$ errors is given by the following:

$$y_t = 0.1429 y_{t-1} + \varepsilon_t \tag{4.1.17}$$

where,

$$\varepsilon_t = \sqrt{h_t}\, Z_t \tag{4.1.18}$$

$$h_t = 3.515 \times 10^{-4} + 0.0966\varepsilon_{t-1}^2 + 0.8224 h_{t-1}. \tag{4.1.19}$$

The estimate of σ_ε^2 based on the AR(1) model with GARCH$(1,1)$ errors is $\hat{\sigma}_\varepsilon^2 = \dfrac{\hat{\omega}}{1-\hat{\alpha}_1-\hat{\beta}_1} = \dfrac{3.515 \times 10^{-4}}{1-0.0966-0.8224} = 0.1583$. The marginal variance is estimated to be $E(y_t^2) = 0.1616$. Using Theorem 4.1.2, the l-steps ahead forecast error standard errors ($l = 1, 2, 3$) are 0.3979, 0.4019 and 0.4020, respectively.

When we fit an AR(1) model of the form $y_t = 0.0767y_{t-1} + \varepsilon_t$, assuming the conditional variance is a constant, the constant error variance estimate is $\hat{\sigma}_\varepsilon^2 = 0.0044$. The standard error of the 1,2 and 3-steps ahead forecast standard errors for an AR(1) fit with constant variance are 0.0663, 0.0665, and 0.0665, respectively.

The forecast standard errors from the AR(1) model with GARCH(1, 1) errors are larger than forecast standard errors from the Gaussian AR(1) model. Prediction intervals from the AR(1) model are narrower than the prediction intervals obtained from the AR(1)/GARCH(1, 1) model. Let us consider the portfolio allocation example discussed in Chapter 1 due to Gourieroux (1997). Suppose that an investor has a portfolio consisting of exactly two assets: a risk-free asset with constant return r_f and a risky return whose future return is predicted via a prediction interval, $(r_{l,t},\ r_{u,t})$, where t denotes time. Assume that at the initial date, the whole portfolio is invested in the risk-free asset. The investor wishes to reallocate his portfolio in the following manner. Because of transaction costs, he will only change his portfolio when $r_{l,t} > r_f$. Since this means that the risky asset is more profitable, he will change his entire portfolio to be reinvested in the risky asset. The portfolio will remain unchanged until $r_{u,t} < r_f$, at which time, the entire portfolio will be reinvested in the risk-free asset, and so on. For this particular dataset, because the prediction intervals derived from the Gaussian AR(1) model are narrower, this would mean that the investor would falsely infer that the risky asset will be more profitable than the risk-free asset as $r_{l,t}$ is more likely to exceed r_f, in which case, the investor is more likely to incur unnecessary risk, resulting in a loss.

Example 4.1.7 *(GE)*

We now consider fitting an appropriate model for daily log returns (in percentages) of GE stock. We fit a AR(1) model to $\{y_t\}$ with t-distribution GARCH$(1,1)$ errors. The estimate of the AR parameter is $\hat{\phi} = 0.0363$ and its standard error is estimated to be 0.0101. The estimates of the GARCH parameters are $\hat{\omega} = 0.02101$, $\hat{\alpha_1} = 0.0478$ and $\hat{\beta}_1 = 0.9418$, respectively. Based on this model, $\hat{\sigma}_\varepsilon^2 = \dfrac{0.02101}{1 - 0.0478 - 0.9418} = 2.0276$. Using Theorem 4.1.2, the estimated standard errors of the l-steps ahead $(l = 1, 2, 3)$ forecast errors are 1.4329, 1.4249, and 1.4249, respectively.

It is of interest to note that if we fit an AR(1) model with a constant error variance, using the estimate of the AR parameter $\hat{\phi} = 0.03584$, then $\hat{\sigma}_\varepsilon^2 = 2.0638$. The 1,2 and 3-steps ahead forecast standard errors are 1.4366, 1.4375 and 1.4375, respectively.

In the GE example, the forecast standard errors are nearly the same for both the AR(1) model and the AR(1) model with GARCH(1, 1) errors. This is mainly due to the large estimate of ω. The estimate of ω corresponds to the long-run variance of the return process. When the estimate of the variability in the return process is large, inferences based on the AR(1) model and the AR(1) model with GARCH errors are the same. In the next Section, inference for stationary processes with structural change is discussed.

4.2 Forecasting with Structural Change

Suppose we have observations from a GARCH(p,q) process and the unconditional mean of y_t^2, σ_0^2 shifts from σ_0^2 to σ_1^2 at time $t = T + 1$. The following lemmas are needed to prove a theorem on the improved estimator of some volatility models in the presence of structural change.

Lemma 4.2.1 *(i) The MMSE forecast of y_{T+1} is $c\bar{y}$ where $c = \frac{\mu^2}{E\bar{y}^2}$, and*

$$E\bar{y}^2 = \mu^2 + \frac{\gamma_0}{T^2}(T + 2(T-1)\rho_1 + 2(T-2)\rho_2 + \ldots + 2\rho_{T-1}) = \mu^2 + \Sigma_{\bar{y}}^2 \quad (4.2.1)$$

with $\Sigma_{\bar{y}}^2 = Var(\bar{y})$ given by

$$\Sigma_{\bar{y}}^2 = \frac{\gamma_0}{T^2}(T + 2(T-1)\rho_1 + 2(T-2)\rho_2 + \ldots + 2\rho_{T-1}) \quad (4.2.2)$$

(ii) The MSE is $MMSE_T = \sigma^2 + \mu^2(1 - \frac{\mu^2}{E\bar{y}^2}) = \sigma^2 + \frac{\mu^2\Sigma_{\bar{y}}^2}{\mu^2 + \Sigma_{\bar{y}}^2} \leq \sigma^2 + \Sigma_{\bar{y}}^2$.

Lemma 4.2.2 *Suppose that the mean of the stationary series shifts from μ to μ_1 at time $T+1$ and also assume that the autocorrelation structure remains the same after the shift.*

(i) The MMSE estimate of μ_1 is $c\bar{y}$ where $c = \frac{\mu\mu_1}{E\bar{y}^2}$, $E\bar{y}^2$ is as in (4.2.1) and $\Sigma_{\bar{y}}^2$ is as in (4.2.2).

(ii) The MSE is $MMSE_T = \sigma^2 + \mu_1^2(1 - \frac{\mu^2}{E\bar{y}^2}) = \sigma^2 + \frac{\mu_1^2\Sigma_{\bar{y}}^2}{\mu^2 + \Sigma_{\bar{y}}^2}$.

Theorem 4.2.1 *: For the GARCH processes described in (2.3.9) and (2.3.10), under suitable stationarity conditions,*

(a) The improved estimate of $\sigma_1^2 = E(y_{T+1}^2)$ is $c\bar{y^2}$ where $\bar{y^2} = \frac{1}{T}\sum_{k=1}^{T} y_k^2$ and $c = \left[1 + \frac{(K^{(y)} - 1)}{T}\right]^{-1}$.

(b) *The MSE of the estimate of σ_1^2, $c\bar{y^2}$, is less than the MSE of the estimate of $\bar{y}^2$, of σ_1^2.*

Proof: The proof of Part (a) follows from the fact that $Var(y_t^2) = (K^{(y)}-1)[Var(y_t)]^2$, and by the definition of the coefficient of variation. The proof of Part (b) is similar to the proof for linear time series models.

4.2.1 A Regression Model for a Single Structural Break

Consider the simple linear regression model studied in Pesaran and Timmerman (2004) of the form

$$y_{t+1} = \beta_1 x_t I_{(t \leq T_1)} + I_{(t > T_1)} \beta_2 x_t + a_{t+1} \tag{4.2.3}$$

possibly with a shift in its variance from σ_1^2 to σ_2^2, where I denotes the indicator function. We know that β has changed at T_1, our interest lies in forecasting y_{T+1} given the observations $(x_t, y_t), t = 1, 2, \ldots, T$.

The problem of the number of observations needed to estimate a model that, when used to generate forecasts, will minimize the expected mean squared forecast error has been studied in Pesaran and Timmermann (2004).

Let m denote the starting point of the most recent observations to be used in estimation for the purpose of forecasting y_{T+1}. Then the least squares estimator (LSE) can be written as

$$\hat{\beta}_T(m) = \left(\sum_{t=m}^{T} x_{t-1}^2 \right)^{-1} \left(\sum_{t=m}^{T} x_{t-1} y_t \right). \tag{4.2.4}$$

The forecast of y_{T+1} is given by $\hat{y}_{T+1} = \hat{\beta}_T(m) x_T$, with

$$\hat{\beta}_T(m) = \theta_m \beta_1 + (1 - \theta_m)\beta_2 + \nu_T(m) \tag{4.2.5}$$

where $\nu_T(m) = \left(\sum_{t=m}^{T} x_{t-1}^2\right)^{-1}(\sum_{t=m}^{T} x_{t-1}a_t)$, and $\theta_m = \left(\sum_{t=m}^{T} x_{t-1}^2\right)^{-1}(\sum_{t=m}^{T_1} x_{t-1}^2)$. The forecast error in the prediction of y_{T+1} will be a function of the data sample used to estimate β and is given by the following: $e_{T+1}(m) = a_{T+1} + \theta_m(\beta_2 - \beta_1)x_T - \nu_T(m)x_T$. And, $E[\nu_T(m)^2|X_T] = \left(\sum_{t=m}^{T} x_{t-1}^2\right)^{-1}(\sigma_1^2\theta_m + (1-\theta_m)\sigma_2^2)$. The corresponding conditional MSFE becomes $E[e_{T+1}^2(m)|x_T] = \sigma_2^2 + \theta_m^2(\beta_2 - \beta_1)^2 x_T^2 + [\left(\sum_{t=m}^{T} x_{t-1}^2\right)^{-1}(\sigma_1^2\theta_m + (1-\theta_m)\sigma_2^2)]x_T^2$. For the proposed the MMSE predictor as $y_{T+1} = c\hat{\beta}_T(m)x_T$ where $c = \dfrac{\beta_2 E\hat{\beta}_T(m)}{E\hat{\beta}_T^2(m)}$, $e_{T+1}(m) = y_{T+1} - c\hat{\beta}_T(m)x_T = (\beta_2 - c\hat{\beta}_T(m))x_T + a_{T+1}$. The corresponding conditional MSFE becomes $E[e_{T+1}^2(m)|x_T] = \sigma_2^2 + [\beta_2^2(1 - \dfrac{(E\hat{\beta}_T(m))^2}{E\hat{\beta}_T^2(m)}]x_T^2$ which is smaller than the conditional MSFE of Persaran and Timmerman (2004).

For the model considered in (4.2.3), the estimate of β_2 with the starting point m of the most recent observations can be written as

$$\hat{\beta}_T(m) = \left(\sum_{t=m}^{T} x_{t-1}^2\right)^{-1}(\sum_{t=m}^{T} x_{t-1}y_t) = k_m(\sum_{t=m}^{T} x_{t-1}y_t), \tag{4.2.6}$$

where $k_m^{-1} = \left(\sum_{t=m}^{T} x_{t-1}^2\right)$, $k_m^{-1} - k_{m+1}^{-1} = x_{m-1}^2$.

In order to determine m,

Step 1: Recursively calculate $\hat{\beta}_T(m)$ and k_m^{-1} for each m using $\hat{\beta}_T(m) = \hat{\beta}_T(m+1) + k_m x_{m-1}\left(y_m - x_{m-1}\hat{\beta}_T(m+1)\right)$.

Step 2: Calculate the improved forecast and the corresponding conditional minimum mean square error for each m.

Step 3: Select the value of m which minimizes the minimum value given in Step 2 as the optimal choice of the window.

Improved estimation requires knowledge of model parameters other than the parameter of interest. In the next Section, recursive improved estimators are proposed where only starting values for nuisance parameters are needed in order to estimate the parameter of interest.

4.2.2 Related Inference Problems

Before discussing improved recursive estimation for time series models, we derive improved recursive estimates for i.i.d. data in order to motivate recursive estimation for time series in the presence of structural change.

Example 4.2.1 *Let $Y_1, \ldots, Y_n$ be i.i.d. with $E(Y) = \mu$ and $Var(Y) = \sigma^2$ (known). We have the following:*

(i) Let $\tilde{\mu}_n = \bar{y}_n$ denote the sample mean based on n observations. The recursive estimate of the mean based on $n+1$ observations has the form

$$\tilde{\mu}_{n+1} = \tilde{\mu}_n + \frac{1}{n+1}(Y_{n+1} - \tilde{\mu}_n).$$

(ii) The improved estimate of μ given by $\hat{\mu}_n = c_n \tilde{\mu}_n$ where $\tilde{\mu}_n = \bar{y}_n$, $c_n = \dfrac{\mu^2}{E\bar{y}_n^2} = \dfrac{1}{1 + \dfrac{\nu^2}{n}}$ and $\nu = \sigma/\mu$ is the coefficient of variation. The following relations provide the recursive improved estimate of μ.

$$\begin{aligned} \hat{\mu}_{n+1} &= \hat{\mu}_n + \frac{c_{n+1}}{n+1}(Y_{n+1} - \hat{\mu}_n) \\ (n+1)c_{n+1}^{-1} &= nc_n^{-1} + 1 \end{aligned}$$

Proof: Part (ii) follows from observing the following. The improved estimate *is* $\hat{\mu}_n = c_n\tilde{\mu}_n$. *Then,*

$$\begin{aligned}
\hat{\mu}_{n+1} &= c_{n+1}\tilde{\mu}_{n+1} \\
&= \frac{c_{n+1}}{n+1}[n\tilde{\mu}_n + Y_{n+1}] \\
&= \frac{c_{n+1}}{n+1}[nc_n^{-1}\hat{\mu}_n + Y_{n+1}] \\
&= \frac{n}{n+1}\frac{c_{n+1}}{c_n}\hat{\mu}_n + \frac{c_{n+1}}{n+1}Y_{n+1}.
\end{aligned}$$

But, $(n+1)c_{n+1}^{-1} = nc_n^{-1} + 1$. *Multiplying by* c_{n+1}^{-1} *on both sides and rearranging terms, we obtain* $\frac{c_{n+1}}{c_n}\frac{n}{n+1} = 1 - \frac{c_{n+1}}{n+1}$. *Then,*

$$\hat{\mu}_{n+1} = \left(1 - \frac{c_{n+1}}{n+1}\right)\hat{\mu}_n + \frac{c_{n+1}}{n+1}Y_{n+1}.$$

Note: In both instances above, c_n is a function of unknown model parameters. If an off-line(non-recursive) estimation procedure were to be used, we would need to estimate these model parameters from the data as well, adding to the variability of the paramter estimate. However, with the recursive estimation procedure one simply needs to provide a starting value for c_n since the estimate will converge to the uniformly minimum variance unbiased estimator(UMVUE) as the sample size increases. c_n is chosen to minimize the MSE of the improved estimator. That is,

$$E(y_{n+1} - c\bar{y})^2 = \sigma^2(1 + \frac{1}{n}[1 + \frac{v^2}{n}]^{-1}) \leq \sigma^2(1 + \frac{1}{n}),$$

where $v = \frac{\sigma}{\mu}$ is the the coefficient of variation. If v is small then the improvement on the MSE is relatively small. However, if v is large then the improvement can be significant.

The following theorems give expressions for recursive improved estimates of mean and variance parameters. The derivation of the proofs is similar to Example 4.2.1

and will be omitted.

Theorem 4.2.2 *Consider the MA(1) model $y_t - \mu = a_t - \theta a_{t-1}$. Interest centers on recursively estimating the improved estimate of the mean, μ. The improved estimate is given by $\hat{\mu}_n = c_n \tilde{\mu}_n$, where $c_n = \dfrac{\mu^2}{\mu^2 + \frac{\gamma_0}{n^2}[n + 2(n-1)\rho_1]}$, ρ_1 is the first lag autocorrelation of y_t and $\tilde{\mu}_n$ is the sample mean based on n observations. The recursive relations needed are the following:*

$$
\begin{aligned}
\hat{\mu}_{n+1} &= \hat{\mu}_n + \frac{c_{n+1}}{n+1}(y_{n+1} - \hat{\mu}_n) && (4.2.7) \\
\frac{c_{n+1}^{-1} - 1}{c_n^{-1} - 1} &= \frac{n^2}{(n+1)^2}\left[\frac{n + 1 + 2n\rho_1}{n + 2(n-1)\rho_1}\right]. && (4.2.8)
\end{aligned}
$$

We can use the first lag autocorrelation of the sample autocorrelation function (ACF) as a starting value for ρ_1.

Theorem 4.2.3 *For a moving average process of order q (MA(q))*

$$y_t - \mu = a_t - \theta_1 a_{t-1} - \ldots \theta_q a_{t-q},$$

where a_t is an uncorrelated zero mean Gaussian process with variance σ_a^2. The recursive relations for improved estimates of μ in a MA(q) model are given by the following:

$$
\begin{aligned}
\hat{\mu}_{n+1} &= \hat{\mu}_n + \frac{c_{n+1}}{n+1}(y_{n+1} - \hat{\mu}_n), \\
\frac{c_{n+1}^{-1} - 1}{c_n^{-1} - 1} &= \frac{n^2}{(n+1)^2}\left[\frac{n + 1 + 2n\rho_1 + 2(n-1)\rho_2 + \ldots + 2(n+1-q)\rho_q}{n + 2(n-1)\rho_1 + 2(n-2)\rho_2 + \ldots + 2(n-q)\rho_q}\right].
\end{aligned}
$$

In order to obtain starting values for the first q autocorrelations $\rho_1, \ldots, \rho_q$, the first q sample autocorrelations may be used as estimates.

Theorem 4.2.4 *Consider the AR(1) model*

$$y_t - \mu = \phi(y_{t-1} - \mu) + \varepsilon_t, \quad \varepsilon_t \sim (0, \sigma_e^2), i.i.d., t = 1, \ldots, n.$$

The recursive improved estimates of the parameters are given by the following:

(i) *The recursive improved estimate of μ is given by $\hat{\mu}_n = c_n \tilde{\mu}_n$, where $c_n = \dfrac{\mu^2}{E\bar{y}^2} = \dfrac{\mu^2}{Var(\bar{y}) + \mu^2}$, and*

$$Var(\bar{y}) = \frac{\gamma_0}{n^2}(n + 2(n-1)\rho_1 + 2(n-2)\rho_2 + \ldots + 2\rho_{n-1}) = \mu^2 + \Sigma_{\bar{y}}^2.$$

The recursive relations are the following:

$$\hat{\mu}_{n+1} = \hat{\mu}_n + \frac{c_{n+1}}{n+1}(y_{n+1} - \hat{\mu}_n) \tag{4.2.9}$$

$$\frac{c_{n+1}^{-1} - 1}{c_n^{-1} - 1} = \frac{n^2}{(n+1)^2}\left[\frac{n + 2(n-1)\rho_1 + 2(n-2)\rho_2 + \ldots + 2\rho_{n-1}}{n + 1 + 2n\rho_1 + 2(n-1)\rho_2 + \ldots + 2\rho_n}\right]^{-1}$$

(ii) *Let $\tilde{\phi}_n$ denote the conditional least squares estimate (CLS) of ϕ based on a sample of n observations. The following equations are needed in order to calculate the recursive estimate of ϕ.*

$$\tilde{\phi}_{n+1} = \tilde{\phi}_n + k_{n+1} y_n (y_{n+1} - \tilde{\phi}_n y_n) \tag{4.2.10}$$

$$k_{n+1}^{-1} = k_n^{-1} + y_n^2. \tag{4.2.11}$$

(iii) *The improved estimate of ϕ based on minimizing the MSE conditional on F_{n-1}^y, the $\sigma-$field generated by $y_1, \ldots, y_{n-1}$ is given by $\hat{\phi}_n = c_n \tilde{\phi}_n$, where $\tilde{\phi}_n$ is the CLS estimate of ϕ based on a sample of size n and*

$$c_n = \frac{\phi^2}{\sigma_e^2\left(\sum_{t=2}^n y_{t-1}^2\right)^{-1} + \phi^2}.$$

The recursive relations for estimating the improved estimate are the following:

$$\hat{\phi}_{n+1} = c_{n+1}\tilde{\phi}_{n+1} \tag{4.2.12}$$

$$\tilde{\phi}_{n+1} = \tilde{\phi}_n + k_{n+1}y_n(y_{n+1} - \tilde{\phi}_n y_n) \tag{4.2.13}$$

$$k_{n+1}^{-1} = k_n^{-1} + y_n^2 \tag{4.2.14}$$

$$c_{n+1}^{-1} = \frac{c_n^{-1}k_n^{-1} + y_n^2}{k_n^{-1} + y_n^2}, \tag{4.2.15}$$

For an AR(1) process, we may estimate ϕ using the first lag autocorrelation of y_t. While c_n may be a function of unknown parameter values, the recursive estimation procedure requires only a starting value c_1 be provided and thus, these unknown quantities do not need to be estimated at each iteration. It is of interest to note that recursive improved estimation also leads to the Kalman filtering type equations of Thavaneswaran and Abraham (1988).

The ARMA representation of a GARCH(p,q) y_t process in terms of y_t^2 allows for recursive improved estimation in the following Lemma.

Lemma 4.2.3 *Consider an ARCH(1) model*

$$\begin{aligned} y_t &= \sqrt{h_t}Z_t, \\ h_t &= \omega + \phi y_{t-1}^2, \end{aligned}$$

where Z_t is an uncorrelated zero mean process with unit variance.

(i) The following pair of equations can be used to obtain the recursive estimate of ϕ:

$$\begin{aligned} \tilde{\phi}_{n+1} &= \tilde{\phi}_n + k_{n+1}y_n^2(y_{n+1}^2 - \tilde{\phi}_n y_n^2) \\ k_{n+1}^{-1} &= k_n^{-1} + y_n^4. \end{aligned}$$

(ii) *Suppose we wish to recursively estimate the improved estimator $c\phi$, of ϕ, by minimizing the conditional MSE. Then,* $c_n = \dfrac{\phi^2}{\phi^2 + \sigma_u^2 \left(\sum_{t=2}^{n} y_{t-1}^4\right)^{-1}}$. *Let* $\tilde{\phi}_n$ *denote the estimator of ϕ based on a sample of n observations and found by minimizing the conditional MSE. Let* $\hat{\phi}_n = c_n \tilde{\phi}_n$ *deonte the improved estimator based on a random sample of size n. Minimizing the conditional MSE, leads to the following recursive equations for the improved estimate* $\hat{\phi} = c_n \tilde{\phi}$.

$$\hat{\phi}_{n+1} = c_{n+1}\tilde{\phi}_{n+1} \tag{4.2.16}$$

$$\tilde{\phi}_{n+1} = \tilde{\phi}_n + k_{n+1} y_n^2 (y_{n+1}^2 - \tilde{\phi}_n y_n^2) \tag{4.2.17}$$

$$c_{n+1}^{-1} = \frac{c_n^{-1} k_n^{-1} + y_n^4}{k_n^{-1} + y_n^4}, \tag{4.2.18}$$

$$k_{n+1}^{-1} = k_n^{-1} + y_n^4. \tag{4.2.19}$$

Proof: Parts (i) and (ii) follow from the fact that y_t^2 has an AR(1) representation.

Suppose we have observations from a GARCH(p,q) process and the unconditional mean of y_t^2, σ_0^2 shifts from σ_0^2 to σ_1^2 at time $t = T+1$. The following theorem provides recursive improved estimates of volatility in the presence of structural change.

Theorem 4.2.5 *Consider (y_t) a GARCH(p,q) process with structural change at time $n+1$. Denote by $\tilde{\sigma}_n^2$ an estimate of volatility based on n observations calculated as* $\dfrac{1}{n}\displaystyle\sum_{k=1}^{n} y_k^2$. *Let* $\hat{\sigma}_n^2 = c_n \tilde{\sigma}_n^2$ *denote the improved estimate of volatility. The following relations, allow us to calculate $\hat{\sigma}_n^2$, recursively.*

$$\hat{\sigma}_{n+1}^2 = \hat{\sigma}_n^2 + \frac{c_{n+1}}{n+1}(y_{n+1}^2 - \hat{\sigma}_n^2), \tag{4.2.20}$$

$$(n+1)c_{n+1}^{-1} = nc_n^{-1} + 1. \tag{4.2.21}$$

Proof: From Theorem 4.2.1, $c_n = \left[1 + \frac{K^{(y)} - 1}{n}\right]^{-1}$ where $K^{(y)}$ denotes the kurtosis. For a GARCH(p,q) process, Thavaneswaran et al. (2005) and Bollerslev (1986) have shown that $K^{(y)} = \frac{3(1-(\alpha_1+\beta_1)^2)}{1-(\alpha_1+\beta_1)^2 - 2\alpha_1^2}$. It is then easy to show that $(n+1)(c_{n+1}^{-1} - 1) = n(c_n^{-1} - 1)$ and the recursive relation involving c_n follows. The recursive relation involving $\hat{\sigma}_n^2$ is due to the fact that $\hat{\sigma}_n^2 = E(\bar{y_n^2})$.

The recursive relations lead to the Kalman filter once more as was shown in Thavaneswaran and Abraham (1988). An application of Godambe's theorem allows us to calculate recursive improved estimators of the autoregressive parameter in an RCA model.

Another class of models useful for modeling the time-varying and leptokurtic nature of financial time series is that of RCA models first studied by Nicholls and Quinn (1982). RCA models are defined by allowing random additive perturbations of the autoregressive coefficients of linear AR models. For example, an RCA(1) process is given by

$$y_t = (\phi + b_t)y_{t-1} + \varepsilon_t, \tag{4.2.22}$$

where (b_t) and (ε_t) are uncorrelated zero mean white noise sequences with variances σ_b^2 and σ_ε^2 respectively.

Using the theory of optimal estimating functions for stochastic processes, (See Theorem 3.3.1), Thavaneswaran and Abraham (1988) showed that the optimal estimate of ϕ is given by

$$\tilde{\phi}_n = \frac{\sum_{t=2}^n a_{t-1}^* y_t}{\sum_{t=2}^n a_{t-1}^* y_{t-1}},$$

where $a_{t-1}^* = \frac{-y_{t-1}}{\sigma_\varepsilon^2 + y_{t-1}^2\sigma_b^2}$. Let $k_n^{-1} = \sum_{t=2}^n a_{t-1}^* y_{t-1}$. Then $\tilde{\phi}_n = k_n \sum_{t=2}^n a_{t-1}^* y_t$.

Theorem 4.2.6 provides the recursive relations for estimating improved estimate of ϕ.

Theorem 4.2.6 *Consider the RCA(1) model specified by (4.2.22).*

(i) The recursive equations for estimating ϕ are given by the following:

$$\begin{aligned}\tilde{\phi}_{n+1} &= \tilde{\phi}_n + a_n^* k_{n+1}(y_{n+1} - \tilde{\phi}_n y_n) \\ k_{n+1}^{-1} &= k_n^{-1} + a_n^* y_n.\end{aligned}$$

(ii) Consider the improved estimator of $c\phi$ defined as the minimizer of the conditional MSE. The estimate of c based on n observations is given by

$$c_n = \frac{\phi^2}{\phi^2 + \left(\sum_{t=2}^{n} \frac{y_{t-1}^2}{\sigma_b^2 y_{t-1}^2 + \sigma_\varepsilon^2}\right)^{-1}}.$$

The recursive relations for estimating $c\phi$ are the following:

$$\hat{\phi}_{n+1} = c_{n+1}\tilde{\phi}_{n+1} \tag{4.2.23}$$

$$\tilde{\phi}_{n+1} = \tilde{\phi}_n + a_n^* k_{n+1}(y_{n+1} - \tilde{\phi}_n y_n) \tag{4.2.24}$$

$$c_{n+1}^{-1} = \frac{k_n^{-1} c_n^{-1} + a_n^* y_n}{k_{n+1}^{-1}} \tag{4.2.25}$$

$$k_{n+1}^{-1} = k_n^{-1} + a_n^* y_n. \tag{4.2.26}$$

Since a_n^* is a function of σ_ε^2 and σ_b^2, these parameters can by estimated from the data using the method of moments and Theorem 2.1.1, while ϕ can be estimated using the sample first lag autocorrelation.

Standard errors for recursive improved estimates of volatility based on GARCH models are more computationally intensive to compute as they require block resampling techniques (See Liu and Singh (1992)). This is the subject of future work.

4.3. Conclusions

In this chapter, the variance as well as the fourth moment of the l-steps ahead forecast error have been derived for ARMA(p, q) models with GARCH(P, Q) errors. Our results extend the results on forecasting for the class of zero-mean GARCH models derived in Thavaneswaran, Appadoo and Peiris (2005). Moreover, we have demonstrated the superiority of the AR model with GARCH errors over the usual linear AR model whenever the estimate of the ω is small. In addition, recursive improved estimation in the presence of structural change has been discussed. In the next chapter, some of the theoretical results derived for volatility modeling and estimation are illustrated through simulation studies and real financial data.

Chapter 5

Applications in Finance

This chapter illustrates some of the financial applications of the theoretical results obtained in the previous chapters. The correlation between the LS and LAD estimating functions as a GARCH model identification tool is extensively studied in simulation studies and illustrated using real financial data in Section 5.1.1 and Section 5.1.2. In Section 5.2, the superiority of the optimal estimating function approach for hypothesis testing is also demonstrated through some simulation studies.

One of the applications of volatility modeling in finance is in the area of pricing call options for the purpose of risk management. Section 5.3 illustrates an application of volatility modeling in the area of option pricing. Section 5.4 deals with estimation of the expected value of the conditional variance, $E(h_t)$. All applications are illustrated with real financial data.

In the next section, we study $\widehat{\rho}$, the moment estimator of the correlation between the LS and LAD estimating functions, as a time series model identification tool in three different simulation studies.

5.1 Time Series Model Identification

5.1.1 Simulation Studies

In this subsection, we study $\hat{\rho}$ under three different classes of models. The first class of models is of the form $y_t = \epsilon_t$, where ϵ_t is assumed to follow a parametric distribution. The second class is that of autoregressive models of order one where the error distribution is assumed to follow a parametric distribution. Finally, the third class of models is that of GARCH models with normal, t-distribution or double exponential errors.

We used *S-Plus*© software to conduct the simulation studies. In the first simulation study, we have simulated 1000 random samples of size 1000 from a number of distributions where the theoretical ρ value is known. In Table 5.1 the average $\hat{\rho}$ statistics and the theoretical ρ values are given for some distributions. The estimated statistic appears to be close to the true parameter.

Distribution	ρ	$\hat{\rho}$	MAD($\hat{\rho}$)
normal(65,25)	0.797885	0.7932454	0.0175457
t_3	0.636620	0.6662141	0.06222691
t_4	0.7071068	0.7021046	0.04411666
t_5	0.735105	0.7400141	0.02608555
t_6	0.750000	0.7457525	0.02743061
t_7	0.7592134	0.7620597	0.02229045
t_8	0.765466	0.7695058	0.01874937
Logistic(0,1)	0.764304	0.7622757	0.02243997
Double exponential(1)	0.707107	0.7065042	0.02528356
Uniform(0,1)	0.866025	0.8600816	0.01228705
Beta(2,2)	0.838525	0.8350651	0.01326087
Beta(3,3)	0.826797	0.819865	0.01347214
Beta(4,4)	0.820313	0.8128436	0.01704931

Table 5.1: Theoretical ρ values and average estimated ρ statistics; The data are random samples from each population. MAD($\hat{\rho}$) represents the mean absolute deviation from ρ.

Error Distrib'n	ρ	$\hat{\rho}$ (AR)	MAD(AR)	$\hat{\rho}$ (GARCH)	MAD(GARCH)
$N(0,1)$	0.797885	0.7979	0.0025	0.7979	0.0025
t_3	0.636620	0.6451	0.0273	0.6436	0.0279
t_4	0.707107	0.7070	0.0109	0.7084	0.0106
t_5	0.735105	0.7356	0.0065	0.7354	0.0062
t_6	0.750000	0.7507	0.0045	0.7501	0.0048
t_7	0.759213	0.7590	0.0041	0.7591	0.0039
t_8	0.765466	0.7652	0.0037	0.7656	0.0036
d'ble exp'l	0.707107	0.7073	0.0038	0.7073	0.0038

Table 5.2: Theoretical ρ values and average estimated ρ statistics for a simulated AR(1) process and GARCH (1,1). MAD is the mean absolute deviation from the true value.

In our second simulation study, we have simulated 500 samples of size 5000 each from autoregressive processes of order 1 (AR(1)), with different innovation distributions. The autoregressive parameter is set to 0.7. It is hoped that the residuals obtained after fitting conditional least squares estimates to the model parameters follow the distribution from which the process was simulated from. The $\hat{\rho}$ statistics have been calculated by using the residuals. Table 5.2 presents an average $\hat{\rho}$ value along with the mean absolute deviation and the statistic is close to the true parameter.

In the third study, we have simulated 500 samples of size 5000 each from y_t, a zero mean model with GARCH(1, 1) errors assuming each of the following standardized distributions for Z_t in (2.3.9): Gaussian, t_3, t_4, t_5, t_6, t_7, t_8 and double exponential(1). We set the model parameters as $\omega = 5.3 \times 10^{-7}$, $\alpha_1 = 0.0057$, $\beta_1 = 0.9381$. This study was carried out using the *finmetrics* module of S-Plus© version 7.0.

For each simulated process, normal-GARCH(1, 1) model was fit and using the standardized residuals $\widehat{Z}_t = \dfrac{y_t}{\widehat{h}_t}$, ρ is estimated as $\widehat{\rho} = \dfrac{\sum_{t=1}^{n} |Z_t - \bar{Z}|/n}{s_Z}$, where $\bar{Z}$ and s_Z represent the sample mean and standard deviation of the standardized residuals, respectively. $\widehat{h}_t$ is the estimated conditional variance at time t.

We have presented the results in Table 5.2. Here, $\widehat{\rho}$ is close to ρ as well. Proper identification also requires and estimated standard error of $\widehat{\rho}$ and this is the subject of future work. In each of these studies, we have also reported the mean absolute deviations. The variablity of $\widehat{\rho}$ about ρ is largest for t-distribution GARCH models where the degrees of freedom is less than five as $\widehat{\rho}$ is sensitive to extreme observations. In the next section, we illustrate how $\widehat{\rho}$ can be used to select the appropriate error distribution of GARCH(1, 1) models fit to several real datasets.

5.1.2 Examples of GARCH model identification

We illustrate our informal test procedure using financial and biostatistical time series datasets. We provide descriptions of these datasets in Appendix B.7. The datasets are typical financial series in that they are leptokurtic (See Table 5.3), uncorrelated sequences whose square is autocorrelated (See figures in Appendix B.8). For each dataset, we fit a zero-mean normal GARCH(1, 1) model using *S-Plus*© software. Using the residuals from this fit, we calculated an estimate of $\rho = \dfrac{\mathrm{E}|X-\theta|}{\sigma}$ and used $\widehat{\rho}$ to identify the distribution of the error term in the GARCH model. The t-distribution GARCH(1, 1) model is another example of a GARCH model that is typically fit to log-return series. We also fit a t-distribution GARCH model to the same dataset where the degrees of freedom of the t-distribution is a parameter estimated from the data using *S-Plus*© software. We then rounded off the estimated degrees of freedom to the nearest integer. Datasets marked with a $\star$ can also be fitted using a normal-GARCH(1, 1) model. Table 5.4 shows that there is good agreement between the two methods. Jackknife estimates of the standard error of $\widehat{\rho}$ provide a more rigorous way to ascertain the proper error distribution. Incorporating the standard error of $\widehat{\rho}$ into the GARCH model identification procedure is ongoing and the subject of future study.

Dataset	Empirical Kurtosis	Dataset	Empirical Kurtosis
dell.s	4.73	d-ibmln.dat(IBM)	18.22
ford.s	9.15	BMW	10.17
hp.s	8.24	AT&T	27.11
ndx.dat	6.02	Boeing	10.55
siemens	10.75	Caterpillar	6.24
merck.s	8.4	Honeywell	17.17
hkja.dat(Japan)	7.49	Philip Morris	20.04
exch.perc.dat	14.49		

Table 5.3: Empirical Kurtosis of several real financial datasets.

Dataset	$\widehat{\rho}$	Dist'n (implied by $\widehat{\rho}$)	Estimated t-GARCH(1, 1)
dell.s*	0.7835	Gaussian	t_{12}
ford.s*	0.7596	t_7	t_7
hp.s*	0.7572	t_7	t_7
ndx.dat*	0.7975	Gaussian	t_{29}
siemens	0.72199	t_5	t_5
merck.s*	0.7558	t_6	t_6
hkja.dat(Japan)	0.7586	t_7	t_7
exch.perc.dat	0.7136	t_4	t_4
d-ibmln.dat(IBM)*	0.7486	t_6	t_6
BMW	0.70497	t_4	t_4
AT&T	0.7256	t_5	t_5
Boeing	0.7303	t_5	t_5
Caterpillar*	0.7448	t_6	t_6
Honeywell	0.7312	t_5	t_5
Philip Morris	0.6929	t_4	t_4

Table 5.4: Error distribution identification in financial series.

5.2 Tests for Volatility Models

Consider the problem of hypothesis testing for the autoregressive parameter in an AR(1) process with GARCH(1,1) errors. Let θ represent the autoregressive parameter. We are interested in constructing test-statistics for two-sided tests of the form: $H_0 : \theta = \theta_0$ against $H_a : \theta \neq \theta_0$. In Chapter 3, the test based on the optimal estimating function was theoretically shown to be superior over the test based on the least squares estimating function.

In each of the following tables, the the test based on the optimal estimating function method is empirically shown to be superior over the test based on the least squares estimating function. 1000 samples of size 500 are generated from an AR(1) model with, normal-GARCH(1, 1), t_4-distribution GARCH(1, 1) and t_5-distribution GARCH(1, 1) errors. The GARCH model parameters were $\omega = 0.1$, $\alpha_1 = 0.6$ and $\beta_1 = 0.1$. The autoregressive parameter of the simulated AR(1)/GARCH(1, 1) process was selected to be one of the following: $\theta_a = 0.4, 0.5, 0.6$ and the autoregressive parameter θ was tested at each of the following: $\theta_0 = 0.1, \ldots, 0.9$. The power of the least squares estimating function test ($\widehat{power}_{LS}$) and optimal estimating function test ($\widehat{power}_{opt}$) are obtained as the fraction of tests that reject H_0. In the following tables, the power of the test based on the optimal estimating function is shown to be at least as large as the power of the test based on the least squares estimating function.

In the next section, an application of GARCH modeling in finance is discussed.

θ_0	$\widehat{power}_{LS}$	$\widehat{power}_{opt}$
0.1	0.975	1.000
0.2	0.702	0.937
0.3	0.284	0.438
0.5	0.270	0.408
0.6	0.794	0.960
0.7	0.978	1.000
0.8	0.999	1.000
0.9	1.000	1.000

$\theta_a = 0.4$, $Z_t \sim N(0,1)$.

θ_0	$\widehat{power}_{LS}$	$\widehat{power}_{opt}$
0.1	0.999	1.000
0.2	0.966	0.994
0.3	0.772	0.942
0.4	0.293	0.427
0.6	0.337	0.426
0.7	0.853	0.960
0.8	0.992	1.000
0.9	1.000	1.000

$\theta_a = 0.5$, $Z_t \sim N(0,1)$.

θ_0	$\widehat{power}_{LS}$	$\widehat{power}_{opt}$
0.1	1.000	1.000
0.2	0.999	1.000
0.3	0.979	1.000
0.4	0.829	0.971
0.5	0.295	0.442
0.7	0.344	0.466
0.8	0.924	0.977
0.9	1.000	1.000

$\theta_a = 0.6$, $Z_t \sim N(0,1)$.

θ_0	$\widehat{power}_{LS}$	$\widehat{power}_{opt}$
0.1	0.910	0.999
0.2	0.643	0.892
0.3	0.266	0.349
0.5	0.287	0.344
0.6	0.686	0.909
0.7	0.963	1.000
0.8	0.996	1.000
0.9	1.000	1.000

$\theta_a = 0.4$, $Z_t \sim t_4$.

θ_0	$\widehat{power}_{LS}$	$\widehat{power}_{opt}$
0.1	0.985	1.000
0.2	0.914	0.998
0.3	0.647	0.915
0.4	0.259	0.370
0.6	0.311	0.350
0.7	0.734	0.931
0.8	0.983	1.000
0.9	1.000	1.000

$\theta_a = 0.5$, $Z_t \sim t_4$.

θ_0	$\widehat{power}_{LS}$	$\widehat{power}_{opt}$
0.1	0.997	1.000
0.2	0.998	1.000
0.3	0.950	1.000
0.4	0.745	0.921
0.5	0.297	0.400
0.7	0.334	0.367
0.8	0.866	0.950
0.9	0.998	1.000

$\theta_a = 0.6$, $Z_t \sim t_4$.

θ_0	$\widehat{power}_{LS}$	$\widehat{power}_{opt}$
0.1	0.921	0.998
0.2	0.670	0.915
0.3	0.278	0.350
0.5	0.289	0.394
0.6	0.707	0.911
0.7	0.972	1.000
0.8	1.000	1.000
0.9	1.000	1.000

$\theta_a = 0.4$, $Z_t \sim t_5$.

θ_0	$\widehat{power}_{LS}$	$\widehat{power}_{opt}$
0.1	0.990	1.000
0.2	0.946	1.000
0.3	0.705	0.942
0.4	0.275	0.380
0.6	0.309	0.376
0.7	0.773	0.943
0.8	0.990	1.000
0.9	1.000	1.000

$\theta_a = 0.5$, $Z_t \sim t_5$.

θ_0	$\widehat{power}_{LS}$	$\widehat{power}_{opt}$
0.1	0.996	1.000
0.2	0.993	1.000
0.3	0.957	0.999
0.4	0.751	0.951
0.5	0.306	0.366
0.7	0.328	0.386
0.8	0.859	0.950
0.9	0.998	1.000

$\theta_a = 0.6$, $Z_t \sim t_5$.

5.3 An Application of Volatility Modeling in Options Pricing

Call options on eBay^{TM} stock traded on September 28, 2006 were obtained from *The Wall Street Journal* on September 29, 2006. eBay^{TM} is the registered name of a popular auction website and its stock is traded on the New York Stock Exchange (NYSE). The adjusted daily closing stock prices were taken from the Yahoo! finance website. The call options listed in *The Wall Street Journal* are American options; meaning that they can be exercised at any time prior to the exercise date. However, it has been shown that European and American call options have the same call price (p. 276 in Ruppert (2004)). Hence, we can use the Black-Scholes formula for European call options to price eBay^{TM} options.

This section has two subsections. In Section 5.3.1, an adequate model is fit to eBay^{TM} daily log returns based on adjusted closing prices from September 24, 1998 until August 31, 2006.

In Section 5.3.2, two different estimates of volatility are obtained and used to calculate the call prices of eBay^{TM} stock options traded on September 28, 2006. The superiority of the GARCH modeling based approach over the use of historical volatility is demonstrated.

5.3.1 eBay^{TM} Volatility Modeling

The daily closing price process along with the daily log return process are displayed in Figure 5.1. There are 1931 observations in the log-return series. The daily log returns are leptokurtic as the sample kurtosis is 10.99. The sample ACF and of the daily log returns and the sample ACF of the squared process are displayed in Figure 5.2. The presence of significant autocorrelations in the squared log return process indicates the need for fitting a GARCH model. The absence of significant autocorrelations in the cube of the log return process in Figure 5.3 suggests the absence of conditional skewness. The lack of systematic cross-correlations in the sample CCF plot in Figure 5.4 is indicative of the absence of conditional leverage. Hence, an adequate linear GARCH model will explain the salient features of this data.

A normal-GARCH(1, 1) model is fit as a starting point. The $\widehat{\rho}$ statistic (developed in Chapter 3) was calculated based on the standardized residuals obtained from the fitted model. Since $\widehat{\rho} = 0.7280944$, the error distribution was identified to be a Student t-distribution with 5 degrees of freedom using Table 3.1 in Chapter 3. The quantile-

quantile plots of the standardized residuals from a normal-GARCH(1, 1) model and the t_5-distribution GARCH(1, 1) model in Figure 5.5 show that the t-distribution GARCH(1, 1) model provides a better fit. It is of interest to note that when we fit a t-distribution GARCH(1, 1) model by also estimating the degrees of freedom parameter ν, we obtained $\widehat{\nu} = 5.28$ with a standard error of 0.49. Using a Wald-test, we conclude that 5 is a plausible value for ν. We also fit typical GARCH(1, 1) models for this data and examined the kurtosis of the standardized residuals from each fit. Table 5.5 provides the empirical kurtosis of the standardized residuals from the following fitted models: normal-GARCH(1, 1), t_5-distribution GARCH(1, 1), GED-GARCH(1, 1) and Laplace-GARCH(1, 1). In comparing these kurtosis values to the kurtosis of the normal, t_5, GED and Laplace density kurtosis values, we see that the t-distribution with 5 degrees of freedom explains 65% of the kurtosis observed in the standardized residuals, whereas a normal-GARCH(1, 1) fit explains 28% of the kurtosis of the standardized residuals. We investigated the adequacy of the t_5-distribution GARCH(1,1) further by examining the sample ACF of the squared standardized residuals. As there are no significant autocorrelations present in the corresponding sample ACF plots in Figure 5.6, we conclude that a t_5-distribution GARCH(1, 1) model is adequate. The parameter estimates of the fitted t_5-distribution GARCH(1, 1) model are reported in Table 5.6.

From the parameter estimates in Table 5.6, the estimate of $E(h_t)$ is $\widehat{E(h_t)} = \dfrac{2.027 \times 10^{-6}}{1 - 0.03182 - 0.9666} = 0.001292164$ resulting in $\sqrt{\widehat{E(h_t)}} = 0.03594668$ as an estimate for volatility of eBay^{TM} daily log returns. In the next section, eBay call options will be priced and compared using two estimates of volatily: historical volatility and

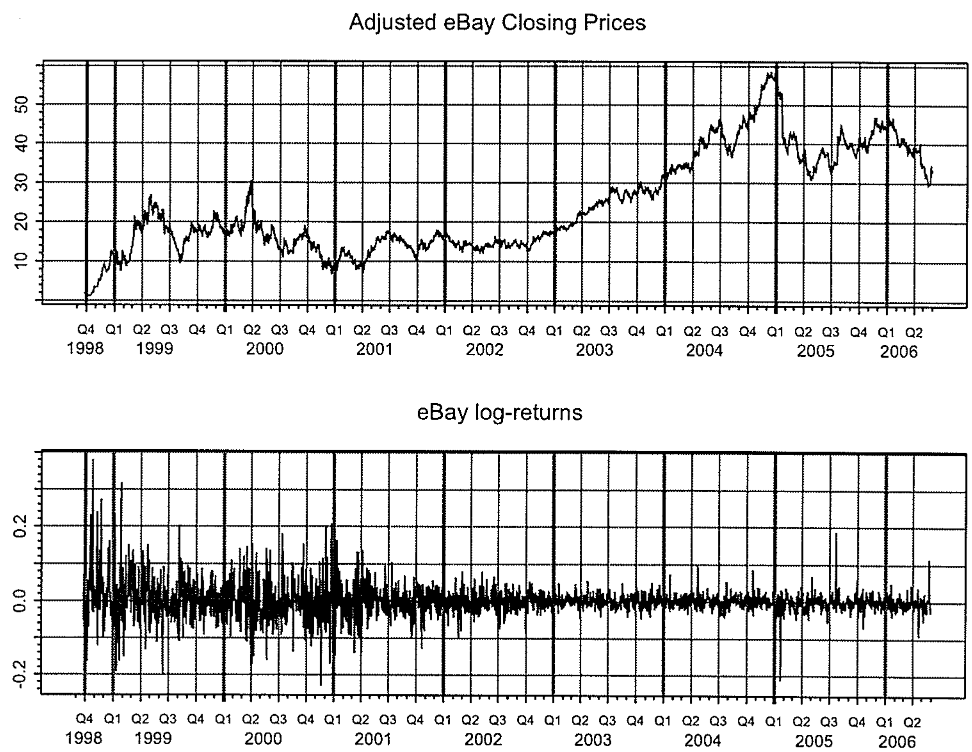

Figure 5.1: eBay^{TM} daily adjusted closing prices and daily log returns.

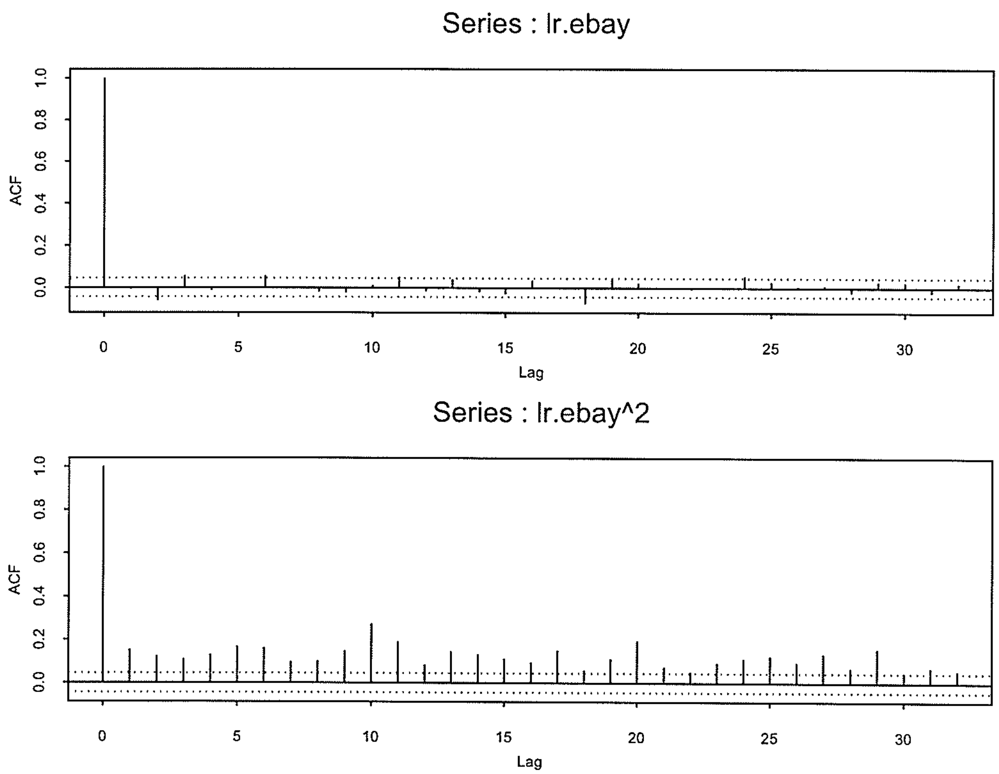

Figure 5.2: Sample ACF of eBay^{TM} daily log returns and squared daily log returns.

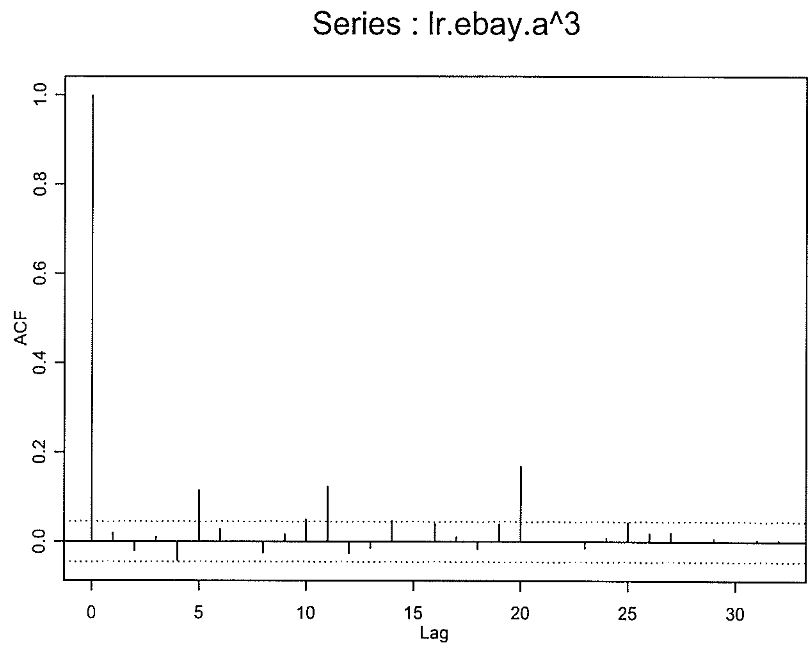

Figure 5.3: Sample ACF of cubed eBay^{TM} daily log returns.

the GARCH based estimate volatility.

5.3.2 Analysis of eBay^{TM} Call Option Data

On September 28, 2006, eBay^{TM} closed at \$28.36 and the three-month T-bill rate closed at 4.75%. Assuming 253 trading days per year, the risk-free rate is $r = 0.0475/253 = 0.000187747$. The sample standard deviation of eBay^{TM} daily log returns is 0.04647978. Call option prices for eBay^{TM} stock traded on September 28, 2006 are calculated using the sample standard deviation of log returns (or historical volatility) and an estimate of $\sqrt{\widehat{E(h_t)}} = \sqrt{\dfrac{\widehat{\omega}}{1-\widehat{\alpha_1}-\widehat{\beta}_1}}$ based on fitting the t_5-distribution GARCH(1, 1) model discussed in the previous section. The predicted call prices reported in Table 5.7 indicate that using the GARCH model based esti-

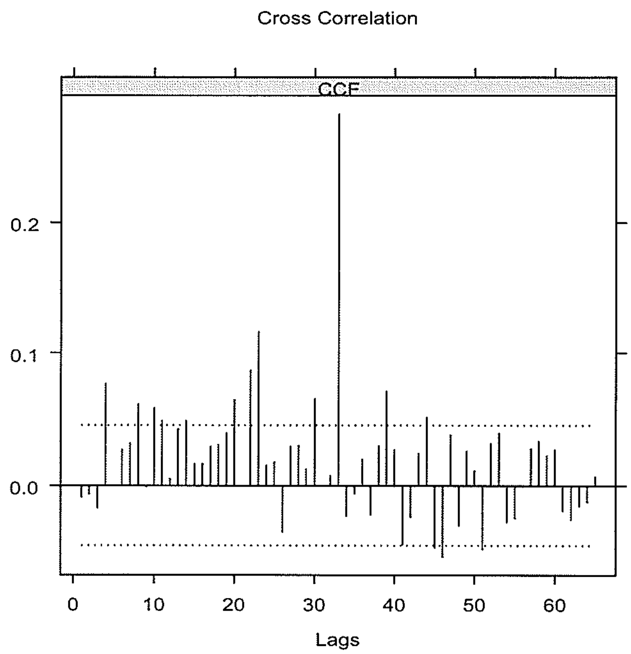

Figure 5.4: Sample CCF between eBay^{TM} daily log returns (y_t) and y_t^2.

mate of volatility yields improved prediction of the call prices. The same conclusion can be drawn from Figure 5.7.

The superiority of using a volatility estimate based on GARCH modeling over the historical volatility based method can further be seen from the p-values of the following goodness of fit tests. If we view the call prices reported on September 28, 2006 as expected call prices in the goodness of fit framework, and the call prices obtained by using σ_H and σ_G from the data as observed call prices, then the χ^2 goodness of fit test statistic that measures the deviation of observed prices from expected prices in each case is 22.36 and 4.23, respectively with 9 degrees of freedom. The p-value of the test that uses historical volatility (σ_H) is 0.0079; whereas the p-value of the test that uses $\widehat{E(h_t)}(\sigma_G)$ from the adequate GARCH model is 0.8955.

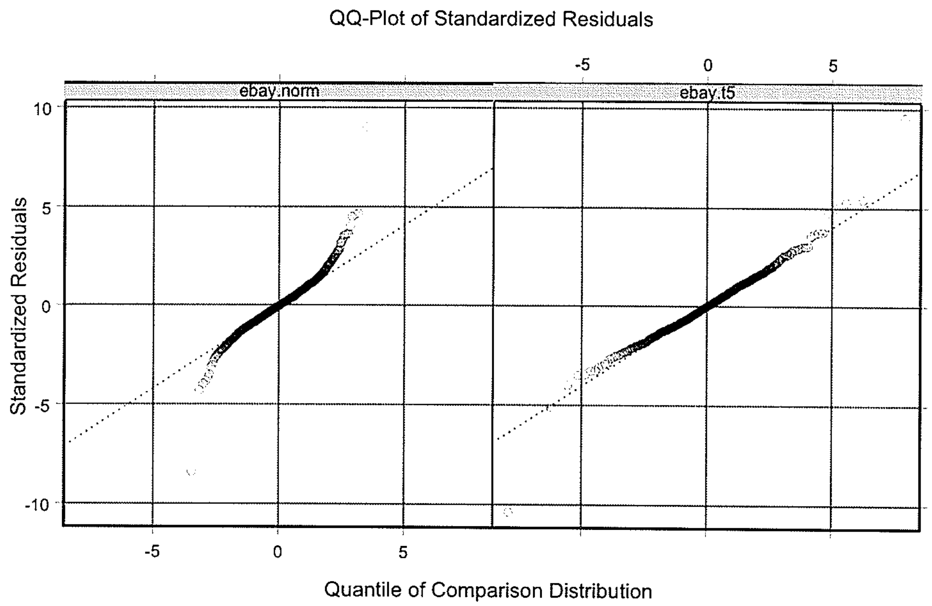

Figure 5.5: Quantile-quantile plots of standardized residuals of a normal-GARCH(1, 1)(left panel) and t_5-distribution GARCH(1, 1) model fit to eBay^{TM} daily log returns.

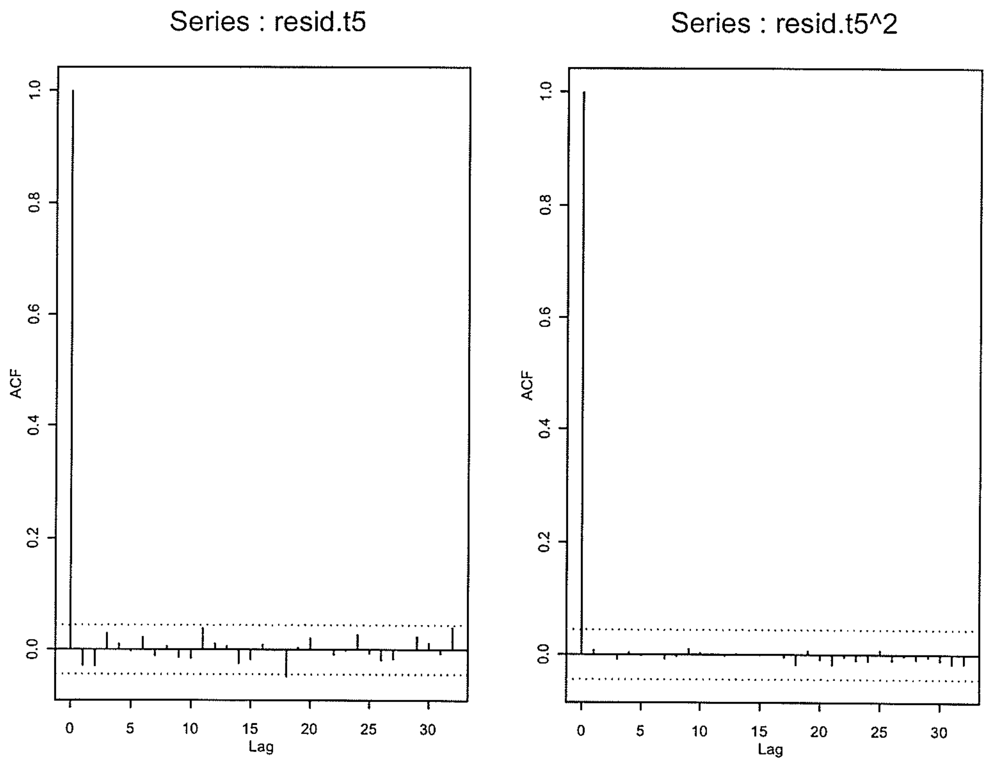

Figure 5.6: Sample ACF of residuals and squared residuals from t_5-distribution GARCH(1, 1) model fit to eBay^{TM} daily log returns.

Distribution	Kurtosis	sample Kurtosis
t_5	9	13.81
$N(0,1)$	3	10.63
GED	4.94	12.87
Laplace	6	12.97

Table 5.5: Comparison of empirical kurtosis of standardized residuals with implied kurtosis.

Parameter	Estimate	S.E.	Statistic	p-value
ω	2.027×10^{-6}	1.356×10^{-6}	1.494	0.06766
α_1	0.03182	0.006029	5.277	< 0.0001
β_1	0.9666	0.005759	167.833	< 0.0001

Table 5.6: Parameter estimates from a t_5-distribution GARCH$(1,1)$ model fit to eBay^{TM} daily log returns.

The goodness-of-fit test statistic in historical volatility case is found in Table 5.8, whereas the goodness-of-fit test statistic for the method that uses GARCH modeling is found in Table 5.9 for more details about the tests). Using GARCH modeling is an improvement over using historical volatility in this case.

K	Month of Expiration	T (in days)	Call Price	B&S price $\widehat{\sigma}_H = 0.0465$	B&S price $\widehat{\sigma}_G = 0.0359$
22.50	Oct	15	5.90	6.13	5.99
25	Oct	15	3.60	4.09	3.78
27.50	Oct	15	1.80	2.50	2.06
27.50	Jan	76	3.20	5.10	4.11
30	Oct	15	0.65	1.40	0.96
30	Nov	35	1.10	2.51	1.81
32.50	Oct	15	0.20	0.72	0.38
32.50	Jan	76	1.10	3.22	2.20
40	Oct	15	0.05	0.07	0.01
42.50	Jan	76	0.10	1.23	0.55

Table 5.7: Call prices and prices determined by the Black-Scholes(B&S) formula for eBay^{TM} options on September 28, 2006. K is the exercise price. T is the maturity. $\widehat{\sigma}_H$ and $\widehat{\sigma}_G$ are the historical volatility and GARCH based volatility, respectively.

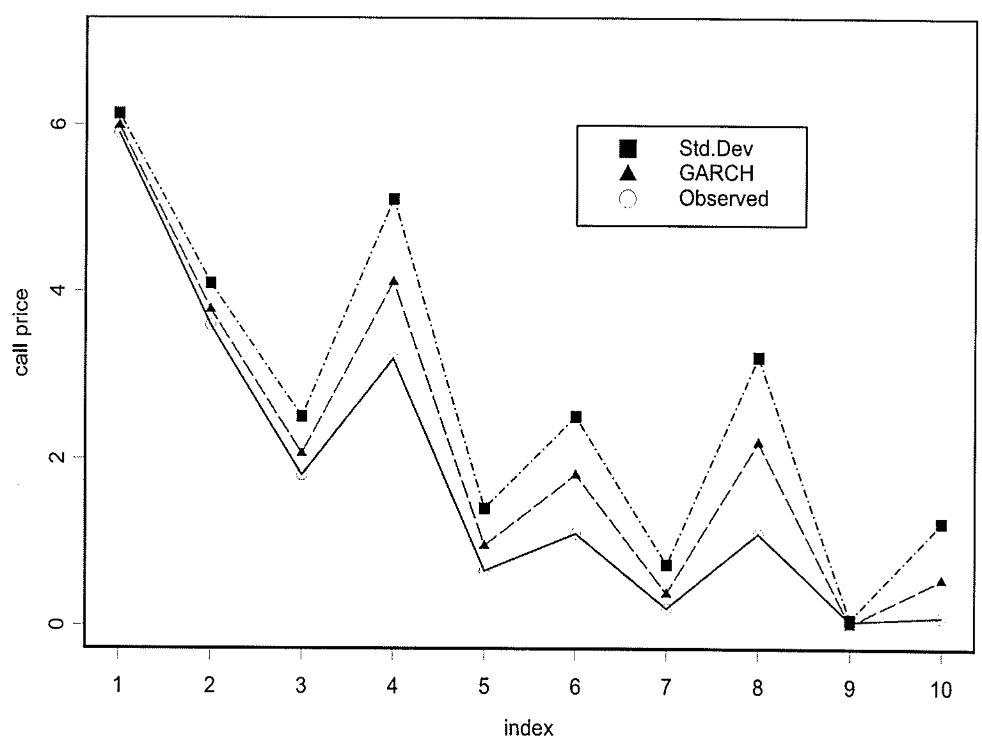

Figure 5.7: Time series plots of observed and predicted eBay^{TM} call prices.

	Price	Expected price	O−E	$(O-E)^2/E$
	6.13	5.90	0.23	0.00897
	4.09	3.60	0.49	0.06669
	2.50	1.80	0.7	0.27222
	5.10	3.20	1.9	1.12812
	1.40	0.65	0.75	0.86538
	2.51	1.10	2.12	4.08582
	0.72	0.20	0.52	1.35200
	3.22	1.10	2.12	4.08582
	0.07	0.05	0.02	0.00800
	1.23	0.10	1.13	12.76900
Sum				22.36357

Table 5.8: Comparison of prices predicted by using historical volatility(σ_H).

	Price	Expected price	O−E	$(O-E)^2/E$
	5.99	5.90	0.09	0.001372
	3.78	3.60	0.18	0.009000
	2.06	1.80	0.26	0.037556
	4.11	3.20	0.91	0.258781
	0.96	0.65	0.31	0.147846
	1.81	1.10	0.71	0.458273
	0.38	0.20	0.18	0.162000
	2.20	1.10	1.10	1.100000
	0.01	0.05	-0.04	0.032000
	0.55	0.10	0.45	2.025000
Sum				4.231829

Table 5.9: Comparison of prices predicted by using GARCH(σ_G).

5.4 Interval Estimation of Volatility

Standard errors for prediction intervals for l-steps ahead forecasts are typically calculated as $2E(h_{T+l}|\mathcal{F}_T^y)$ even though the l-steps ahead forecast is not normally distributed. Recently, Pascual et al. (2006) proposed bootstrap prediction intervals for returns and volatilities in GARCH models. In this section, we propose a bootstrap based method for obtaining a confidence interval for $E(h_t)$ in GARCH models.

5.4.1 Bootstrap Confidence Intervals

Once the error distribution of a GARCH model is identified, the confidence interval for $E(h_t)$ in a GARCH model can be calculated in the following manner.

Without loss of generality, suppose that we are interested in obtaining a confidence interval for $E(h_t)$ in a GARCH(1,1) model. That is, suppose the observed process is given by the following:

$$y_t = \sqrt{h_t}\, Z_t, \tag{5.4.1}$$

$$h_t = \omega + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1}, \tag{5.4.2}$$

where $Z_t \sim (0,1)$ is an uncorrelated sequence with density f. Let $\hat{\omega}$, $\hat{\alpha}_1$, and $\hat{\beta}_1$ be the estimated model parameters obtained through conditional least squares. Calculate the standardized residuals as

$$\hat{Z}_t = y_t/\sqrt{\hat{h}_t}, \tag{5.4.3}$$

where $t = 1, \ldots, n$ and $\hat{h}_1 = \dfrac{\hat{\omega}}{1 - \hat{\alpha}_1 - \hat{\beta}_1}$ is the estimate of $E(h_t)$.

(i) Obtain B replicates of a random sample $Z_1^*, \ldots, Z_n^*$, by resampling $\hat{Z}_1, \ldots, \hat{Z}_n$, B times. From each replicate, obtain a bootstrap sample $y_1^*, \ldots, y_n^*$ using the following:

$$y_t^* = \sqrt{h_t^*} Z_t^* \tag{5.4.4}$$

$$h_t^* = \hat{\omega} + \hat{\alpha}_1 y_{t-1}^{*2} + \hat{\beta}_1 h_{t-1}^*, \quad t = 1, \ldots, n. \tag{5.4.5}$$

Let $h_1^* = \hat{h}_1$ for each of the B replicates.

(ii) Using each replicate, estimate the GARCH model parameters. Denote the estimated parameters by ω^*, α_1^* and β_1^* respectively. Calculate $\hat{E}^*(h_t) = \dfrac{\omega^*}{1 - \alpha_1^* - \beta_1^*}$.

(iii) Order the estimates of $E(h_t)$ in the previous step and use the 2.5% and 97.5% quantiles of the distribution of estimates of $E(h_t)$ to form the 95% confidence interval for $E(h_t)$.

(iv) Taking the square root of endpoints of the interval gives an interval for volatility.

The procedure just described is a non-parametric bootstrap procedure. A parametric bootstrap procedure may also be used to calculate a confidence interval for $E(h_t)$ in a GARCH process. We modify the nonparametric procedure in the following manner.

(i) Identify the error distribution of the GARCH process using the procedure described in Chapter 3 and estimate the model parameters ω, α, and β assuming the identified error distribution.

(ii) Draw a random sample with replacement $Z_1^*, \ldots, Z_n^*$ from the distribution of Z_t.

(iii) Let $h_1^* = \dfrac{\hat{\omega}}{1 - \hat{\alpha}_1 - \hat{\beta}_1}$ be the estimated $E(h_t)$.

(iv) Obtain B replicates of bootstrap samples of size n using the following:

$$y_t^* = \sqrt{h_t^*} Z_t^* \quad (5.4.6)$$

$$h_t^* = \hat{\omega} + \hat{\alpha}_1 y_{t-1}^{*2} + \hat{\beta}_1 h_{t-1}^*, \quad t = 1, \ldots, n. \quad (5.4.7)$$

(v) Using $y_1^*, \ldots, y_n^*$, estimate a GARCH$(1,1)$ model whose innovations are assumed to have the density identified in Step (i) and obtain the parameter estimates ω^*, α_1^* and β_1^*. Calculate the estimate of $E(h_t) = \dfrac{\omega^*}{1 - \alpha_1^* - \beta_1^*}$.

(vi) Order the estimates of $E(h_t)$ in the previous step and use the 2.5% and 97.5% quantiles of the distribution of estimates of $E(h_t)$ to form the 95% confidence interval for $E(h_t)$.

(vii) Take the square root of the endpoints of the interval in the previous step.

5.4.2 Simulation Studies

Asset pricing such as pricing of call options requires knowledge of volatility which is defined as the standard deviation of (log) returns. We propose a confidence interval for volatility associated with GARCH models by taking the square root of the endpoints of the confidence interval for $E(h_t)$ in a GARCH process. To study the small sample properties of such a procedure, we begin by studying interval estimation for the population variance of a set of normally distributed data. We study the coverage of the interval formed by taking the square root of the endpoints as a confidence interval for the population standard deviation. The following examples motivates our interval estimation procedure for $\sqrt{E(h_t)}$.

Example 5.4.1 *Suppose $X_1, \ldots, X_n$ is a random sample from a $N(\mu, \sigma^2)$ population. The confidence interval for the population variance is given by*

$$\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)s^2}{\chi^2_{1-(\alpha/2),n-1}} \right),$$

where $\chi^2_{\alpha/2,n-1}$ and $\chi^2_{1-(\alpha/2),n-1}$ are the upper $\alpha/2$ and lower $1-(\alpha/2)$ percentage points of the chi-square distribution with $n-1$ degrees of freedom, respectively. s^2 is the sample variance.

Interest centers on obtaining a confidence interval for σ, the population standard deviation. 100 samples of size 1000 were simulated from three different Gaussian distributions. For each sample, a 95% confidence interval for the population variance is constructed. The square root of the endpoints of each interval is taken and the fraction of intervals containing the population standard deviation is tallied. The average coverage along with the standard error of the estimate are tabulated in Table 5.10. We found that the procedure produces approximate 95% confidence intervals for the population standard deviation, σ.

In the next section, we study the small sample properties of bootstrap based confidence intervals for volatility.

5.4.3 Interval Estimation of GARCH Volatility

In this subsection, *parametric* (P) and *non-parametric* (NP) bootstrap interval estimates of volatility are calculated for a single sample of size $n = 5000$ drawn from a normal-ARCH(1), a t_5-distribution ARCH(1) and a double-exponential ARCH(1) process with parameters $\omega = 1$, and $\alpha_1 = 0.6$. All calculations were performed using S-Plus© software. The marginal variance of the process is $E(h_t) = \dfrac{\omega}{1-\alpha_1} = 2.5$.

Sample Size	Model	Coverage(s.e.)
n	$N(65, 5^2)$	
100		0.97(0.017)
250		0.97(0.011)
500		0.95(0.0097)
1000		0.96(0.0062)
n	$N(0, 1)$	Coverage(s.e.)
100		0.95(0.022)
250		0.94(0.015)
500		0.98(0.0063)
1000		0.95(0.0069)
n	$N(100, 15^2)$	Coverage(s.e.)
100		0.96(0.012)
250		0.97(0.011)
500		0.99(0.0044)
1000		0.94(0.0075)

Table 5.10: Confidence intervals for the population standard deviation of a random sample of Gaussian observations.

80%, 90% and 98% confidence intervals are calculated for $E(h_t)$ parameter $\dfrac{\omega}{1-\alpha_1}$ using $B = 100, 250, 500, 1000$ bootstrap samples in each case. Interval estimates of volatility can be obtained by taking the square root of the endpoints of the interval for $E(h_t)$. The results are reported in Tables 5.11, 5.12 and 5.13. In one instance, a negative estimate of $E(h_t)$ was obtained. This implies that the procedure needs to constrain the estimate of α_1 to lie in (0, 1). The intervals are markedly different; however the parametric bootstrap estimator of volatility converges faster to the maximum likelihood estimate of volatility than the non-parametric bootstrap estimator as the number of bootstrap samples B increases. Extensive simulation studies are needed in order to compare the the coverage properties of the nonparametric and parametric bootstrap intervals.

B	Method	80% C.I.	90% C.I.	98% C.I.	$\widehat{\sigma}_B^2$
100	P	(2.3892 2.8278)	(2.3462 2.8855)	(2.2412 3.1740)	2.6218
	NP	(2.1660 2.5623)	(2.1450 2.6024)	(2.1009 2.8003)	2.3457
250	P	(2.3697 2.8171)	(2.3126 2.8775)	(2.1769 3.0200)	2.5928
	NP	(2.1622 2.5437)	(2.1133 2.5790)	(2.0847 2.6804)	2.3386
500	P	(2.3897 2.8082)	(2.3325 2.8734)	(2.2209 3.0652)	2.5845
	NP	(2.1551 2.5360)	(2.1191 2.6042)	(2.0510 2.7210)	2.3311
1000	P	(2.3746 2.8141)	(2.3275 2.8924)	(2.2344 3.0210)	2.5901
	NP	(2.1595 2.5351)	(2.1230 2.6109)	(2.0641 2.8443)	2.3396

Table 5.11: Confidence intervals for $E(h_t)$ in a normal-ARCH(1) process. $\widehat{E(h_t)} = 2.6062$.

B	Method	80% C.I.	90% C.I.	98% C.I.	$\widehat{\sigma}_B^2$
100	P	(1.9867 2.8253)	(1.8751 3.0385)	(-1.8766 3.5793)	2.3445
	NP	(1.8130 2.4406)	(1.7567 2.6873)	(1.6336 3.1729)	2.0976
250	P	(2.0066 2.7353)	(1.9175 3.0398)	(1.7945 3.5370)	2.3878
	NP	(1.7811 2.3162)	(1.7376 2.4182)	(1.6380 3.0331)	2.0573
500	P	(2.0151 2.8546)	(1.9308 3.0888)	(1.8201 3.7969)	2.3994
	NP	(1.7797 2.3349)	(1.7346 2.5197)	(1.6869 3.2392)	2.0613
1000	P	(2.0188 2.8475)	(1.9404 3.1049)	(1.8218 3.6510)	2.3958
	NP	(1.7903 2.3622)	(1.7466 2.5326)	(1.6766 3.4470)	2.0878

Table 5.12: Confidence intervals for $E(h_t)$ in a t_5- distribution ARCH(1) process. $\widehat{E(h_t)} = 2.3843$.

Example 5.4.2 *Interval Estimation for eBay^{TM} Volatility*

Using $B = 1000$ samples, we obtained the 90% C.I. for the $E(h_t) = \dfrac{\omega}{1 - \alpha_1 - \beta_1}$ as (0.0005068, 0.01085). The 90% C.I. for volatility (obtained by taking the square root of the endpoints) is given by (0.02251, 0.1042). "An interval estimate" for the call price can be obtained by substituting the endpoints of (0.02251, 0.1042) into the Black-Scholes formula (See Table 5.14).

B	Method	80% C.I.	90% C.I.	98% C.I.	$\widehat{\sigma}_B^2$
100	P	(1.7559 2.3672)	(1.7034 2.5180)	(1.6188 2.6925)	2.0537
	NP	(1.5913 1.8709)	(1.5638 1.9306)	(1.5074 2.0670)	1.7378
250	P	(1.7898 2.2790)	(1.7295 2.3346)	(1.5629 2.5748)	2.0222
	NP	(1.6058 1.8965)	(1.5701 1.9880)	(1.5213 2.1585)	1.7570
500	P	(1.7978 2.3441)	(1.7291 2.4417)	(1.6225 2.6400)	2.0440
	NP	(1.5876 1.8878)	(1.5612 1.9379)	(1.4893 2.1777)	1.7410
1000	P	(1.8053 2.3121)	(1.7559 2.4335)	(1.6624 2.6498)	2.0546
	NP	(1.6000 1.8971)	(1.5629 1.9702)	(1.5003 2.1515)	1.7405

Table 5.13: Confidence intervals for $E(h_t)$ in a double-exponential ARCH(1) process. $\widehat{E(h_t)} = 2.0535$.

5.5 Conclusions

The usefulness of the correlation between the LS and LAD estimating functions as a GARCH model identification tool has been investigated through simulation studies. An application of GARCH volatility modeling in the area of European call options pricing has been discussed and has been shown to be superior over the option pricing formulas that use historical volatility. Comparison with asset pricing formulas that use implied volatility also needs to be investigated and this is the subject of future investigation. Parametric and nonparametric bootstrap based interval estimates of the marginal standard deviation of GARCH models have been proposed. All procedures have been investigated in simulation studies and have been illustrated using real financial data. The superiority of the optimal estimating function approach for hypothesis testing has been demonstrated through some simulation studies.

K	Month of Expiration	T (in days)	Call Price	B&S price $\widehat{\sigma}_G = 0.0359$	B&S price $\widehat{\sigma}_L = 0.02251$	B&S price $\widehat{\sigma}_U = 0.1042$
22.50	Oct	15	5.90	5.99	5.93	7.68
25	Oct	15	3.60	3.78	3.50	6.19
27.50	Oct	15	1.80	2.06	1.51	4.94
27.50	Jan	76	3.20	4.11	2.85	10.35
30	Oct	15	0.65	0.96	0.42	3.92
30	Nov	35	1.10	1.81	0.93	6.34
32.50	Oct	15	0.20	0.38	0.07	3.09
32.50	Jan	76	1.10	2.20	0.96	8.82
40	Oct	15	0.05	0.01	0.00	1.49
42.50	Jan	76	0.10	0.55	0.06	6.56

Table 5.14: Call prices and prices determined by the Black-Scholes(B&S) formula for eBay^{TM} options on September 28, 2006. K is the exercise price. T is the maturity. $\widehat{\sigma}_L$ and $\widehat{\sigma}_U$ are the endpoints of the 90% C.I. for σ. $\widehat{\sigma}_G$ is the GARCH based volatility.

Chapter 6

Summary & Future Work

In Chapter 2, the cross-correlation function and the sample ACF function have been shown to be useful for the purpose of identifying the presence of conditional skewness and conditional leverage, respectively. The inverse Gaussian model of Christoffersen et al. (2006) is a nonlinear GARCH model that adequately captures conditional skewness as well as other salient features of the class of linear GARCH models. The moment properites of *nonlinear* GARCH models that involve other skewed parametric distributions is the subject of future research.

In Chapter 2, moment properties of the ACP and the DACP process have been studied. Markov switching type ACP processes have applications in epidemiology. For example, by incorporating a Markov-switching structure in the class of ACP processes for a time series of influenza counts, the presence of epidemics can be adequately modelled. Moment properties and estimation of such processes using estimating function methodology is of interest.

In Chapter 3, several applications of combining estimating functions have been studied. By combining the least squares estimating function and least absolute deviation estimating function, a GARCH model identification technique for models with symmetric conditional distributions has been developed. Identification of the class of linear GARCH models with asymmetric conditional distributions is a possible extension. The nonlinear time series models in Abraham and Thavaneswaran (1991) have been extended by combining estimating functions for volatility. The combined estimating functions have been shown to have larger information than the component estimating function. Combining estimating functions for *non-linear* GARCH models will also be studied.

Tests for non-linear time series via estimating functions have been studied by Thavaneswaran (1991), Baswa (1991) and Thavaneswaran and Peiris (1998).(See also Chapter 9 in Heyde (1997) for more details). The results of Thavaneswaran and Peiris (1998) have been extended to ARMA models with GARCH errors in Chapter 3. Tests for *non-linear* GARCH models based on estimating functions is a possible extension.

The filtering problem for linear GARCH models and stochastic volatility models has been studied by Peiris and Thavaneswaran (2004) by combining estimating functions. The filtering problem for ARMA models with GARCH errors and the class of *non-linear* GARCH models using estimating function methodology is an extension of interest.

In Chapter 5, interval estimation of volatility was studied. In option pricing, interval estimates of volatility allow the analyst to gauge the sensitivity of call prices to periods of low and high volatility. Nonparametric and parametric bootstrap interval estimates have been proposed. The coverage properties of such intervals need to be studied using simulation studies.

Bibliography

[1] Abraham, B. and Thavaneswaran, A. (1991). A nonlinear time series model and estimation of missing observations, *Annals of the Institute of Statistical Mathematics*, 43(3). 493-504.

[2] Andersen, T.G. and Bollerslev, T. (1998). DM-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer- Run Dependencies , *Journal of Finance*, 53(1), 219-265 .

[3] Appadoo, S.S., Ghahramani, M. and Thavaneswaran, A.(2005). Moment properties of some time series models, *The Mathematical Scientist*, 30(1), 50-63.

[4] Aue, A. (2004). Strong approximation for RCA(1) time series with applications, *Statistics and Probability Letters*, 68, 369-382.

[5] Bai, X., Russell, J., Tiao, G. (2003). Kurtosis of GARCH and stochastic volatility models with non-normal innovations, *Journal of Econometrics*, 114, 349-360.

[6] Baillie, R.T., Bollerslev, T. and Mikkelsen, H.O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 74, 3-30.

[7] Baswa, I.V.(1991). Generalized score tests for composite hypotheses. In V.P. Godambe, Ed. Estimating Functions, Oxford Science Publications, Oxford, 121-131.

[8] Bera, A.K. and Lee, S. (1990). On the Formulation of a General Structure for Conditional Heteroskedasticity, (mimeo, University of Illinois).

[9] Black, F., and Scholes, M. (1972). The Valuation of Option Contracts and a Test of Market Efficiency, *Journal of Finance*, 27, 399-417.

[10] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity, *Journal of Econometrics*, 31, 307-327.

[11] Bollerslev, T. (1987). Modeling the coherence in short run nominal exchange rates: A multivariate generalized ARCH approach, *Review of Economics and Statistics*, 72, 498-505.

[12] Bollerslev, T., Chou, R. Y., and Kroner, K. F. (1992). ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence, *Journal of Econometrics*, 52, 5-59.

[13] Bollerslev, T., Engle R., Nelson D. (1994). ARCH Models, in Handbook of Econometrics (Volume IV). ed. R. Engle and D. McFadden, Amsterdam: North Holland, 2959-3038.

[14] Bollerslev, T. and Mikkelsen, H.O. (1996). Modeling and pricing long memory in stock market volatility, *Journal of Econometrics*, 73, 151 - 184.

[15] Box, G.P. and Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control*, Holden-day Inc., San Francisco, CA.

[16] Chen, G., Choi, Y.K. and Zhou, Y.(2005). Nonparametric estimation of structural change points in volatility models for time series. *Journal of Econometrics*, 126, 79-114.

[17] Chhikara, R.S. and Folks, J.L. (1989). *The inverse Gaussian distribution, theory, methodology and applications.* Marcel Dekker Inc. New York.

[18] Christoffersen, P., Heston, S. and Jacobs, K. (2006). Option Valuation with Conditional Skewness, *Journal of Econometrics*, 131, 253-284.

[19] Clements, M.P. and Hendry, D.F.(1998). Forecasting Economic Time Series, *Cambridge University Press*, Cambridge.

[20] Ding, Z. Granger, C.W.J. and Engle, R.F. (1993). A long memory property of stock market returns and a new model, *Journal of Empirical Finance*, 83-106.

[21] Ding, Z. and Granger, C.W.J. (1996). Modeling volatility persistence of speculative returns: A new approach. *Journal of Econometrics*, 1, 185-215.

[22] Duan, J.C. (1995). The GARCH option pricing model, *Mathematical Finance*, 5, 13-32.

[23] Duan, J.C. and Simonato, J.G. (2001). American option pricing under GARCH by a Markov chain approximation, *Journal of Economic Dynamics and Control*, 25, 1689-1718.

[24] Efron, B. (1982). The jackknife, the bootstrap, and other resampling plans, In Society of Industrial and Applied Mathematics CBMS-NSF Monographs, 38.

[25] Efron, B. (1986). Double exponential families and their use in generalized linear regression, *Journal of the American Statistical Association*, 81, 709-721.

[26] Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation, *Econometrica*, 50, 987-1008.

[27] Engle, R. (2002). New Frontiers for ARCH, *Journal of Applied Econometrics*, 17, 425-446.

[28] Engle, R.F. and Gonzalez-Rivéra, G. (1991). Semiparametric ARCH models, *Journal of Business and Economic Statistic*, 9(4), 345-359.

[29] Engle, R., and Ishida, I. (2002). Forecasting Variance of Variance: The Square-Root, the Affine, and the CEV Garch Models, Department of Finance Working Papers, New York University.

[30] Engle, R.F., Lilien, D.M. and Robins, R.P. (1987). Estimating time varying risk premia in the term structure: the arch-m model, *Econometrica*, 55, 391-407.

[31] Engle, R.F. and Russell, J.K. (1998). Autoregressive Conditional Duration: A new model for irregularly spaced transaction data, *Econometrica*, 66, 1127 - 1162.

[32] Ferguson, T.S. (1999). Asymptotic joint distribution of sample mean and a sample quantile, unpublished manuscript. Available at
<http://www.math.ucla.edu/ tom/papers/unpublished/meanmed.pdf>

[33] Fornari, F. and Mele, A. (1997). Sign-and volatility-switching ARCH models: Theory and applications to international stock markets, *Journal of Applied Econometrics*, 12, 49-65.

[34] Ghahramani, M. and Thavaneswaran, A. (2005). Hidden Semimartingale Models with GARCH Errors, *InterStat*, August 2005.

[35] Ghahramani, M. and Thavaneswaran, A. (2007). Identification of ARMA models with GARCH errors. *The Mathematical Scientist*, 32(1), 1-10.

[36] Giraitis, L. and Surgailis, D.(2002). ARCH-type bilinear models with double long memory. *Stochastic Processes and their applications*, 100, 275-300.

[37] Glosten, L., Jagannathan, R. and Runkle, D. (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, *Journal of Finance*, 48(5), 1779-1801.

[38] Godambe, V.P. (1985). The foundations of finite sample estimation in stochastic processes, *Biometrika*, 72(2), 419-428.

[39] Godambe,V.P. (1999). Linear Bayes and optimal estimation, *Annals of the Institute of Statistical Mathematics*, 51(2), 201-215.

[40] Gouriéroux, C. (1997). ARCH Models and Financial Applications, Springer-Verlag.

[41] Granger, C.W.J. (1998). Overview of Non-Linear time series specification in Economics. In *Econometrics Statist. Summer Symp.* (Berkeley, CA, August 1998). Available at http://emlab.berkeley.edu/symposia/nsf98/granger.pdf.

[42] Hamilton, J.D. and Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime, *Journal of Econometrics*, 64, 307-333.

[43] Hansen, P.R. and Lunde, A. (2004). A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?, *Brown University Economics Working Paper No. 01-04*, Available at SSRN: http://ssrn.com/abstract=264571

[44] He, C. and Teräsvirta, T.(1999a). Fourth moment structure of the GARCH(p,q) process, *Econometric Theory*, 15, 824-846.

[45] He, C. and Teräsvirta, T.(1999b). Properties of moments of a family of GARCH processes, *Journal of Econometrics*, 92, 173-192.

[46] Heinen, A.(2003). Modelling time series count data: an autoregressive conditional poisson model. Working paper 2003/62, Center for operations research and econometrics (CORE). Université catholique de Louvain, Belgium.

[47] Heston, S. and Nandi, S. (2000). A Closed-Form GARCH Option Pricing Model. *Review of Financial Studies*, 13, 586-625.

[48] Heyde, C.C. (1997). *Quasi-Likelihood and its Application: A General Approach to Optimal Parameter Estimation*, Springer, New York.

[49] Higgins, M.L. and Bera, A.K. (1992). A Class of Nonlinear ARCH Models, *International Economic Review*, 33(1), 137-158.

[50] Hsieh, K. C. and Ritchken, P. (2000). An empirical comparison of GARCH option pricing models, *Working paper.*

[51] Hull, J.C. (1995). *Introduction to Futures and Options Markets, 2nd Ed.*, Prentice-Hall, Englewood Cliffs, NJ.

[52] Jarrow, R. and Turnbull, S. (2000). *Derivative Securities, 2nd Ed.*, South Western College Publishing, Cincinnati, OH.

[53] Johnson, N.L. and Kotz, S. (1970). *Continuous Univariate Distributions, Volumes 1 and 2*, Wiley, New York.

[54] Lee, J. and Degennaro, R.P. (2000). Smooth Transition ARCH Models: Estimation and Testing, *Review of Quantitative Finance and Accounting*, 15, 5-20.

[55] Leipus, R. and Surgallis, D. (2003). Random Coefficient Autoregression, regime switching and long memory, *Advances in Applied Probability*, 35, 737-754.

[56] Letac G., Seshadri, V. and Whitmore, G.A. (1985). An exact chi-squared decomposition theorem for inverse Gaussian variates, *Journal of the Royal Statistical Society*, B47, 476-481.

[57] Liu, R.Y. and Singh, K. (1992). Efficiency and robustness in sampling, *Annals of Statistics*, 20, 370-384.

[58] Markowitz, H. M. (1952). Portfolio Selection, *Journal of Finance*, 7(1). 77-91.

[59] Merton, R. C. (1973). Theory of Rational Options Pricing, *Bell Journal of Economics and Management Science*, 4, 141-183.

[60] Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econmetrica*, 59, 347-370.

[61] Nelson, D. B., and Cao, C. Q. (1992). Inequality Constraints in the Univariate GARCH Model, *Journal of Business and Economic Statistics*, 10(2), 229-235.

[62] Nicholls, D.F. and Quinn, B.G. (1982). Random Coefficient Autoregressive Models: An Introduction, Lecture Notes in Statistics, Vol. 11. Springer, New York.

[63] Pascual, L., Romo, J., and Ruiz, E. (2006). Bootstrap prediction for returns and volatilities in GARCH models. *Computational Statistics & Data Analysis*, 50, 9, 2293-2312.

[64] Peiris, S. and Thavaneswaran, A. (2004). A note on the *filtering* for some time series models, *Journal of Time Series Analysis*, 25(3), 397-407.

[65] Pesaran, M.H. and Timmerman, A. (2004). How costly is it to ignore breaks when forecasting the direction of a time series?, *International Journal of Forecasting*, 20, 411-425.

[66] Ruppert, D. (2004). *Statistics and Finance: An Introduction*, Springer, New York.

[67] SAS Institute Inc. (2006). SAS/ETS User's Guide, Version 8, Cary, NC, USA: SAS Institute Inc.

[68] Samanta, M.(1985). Efficient estimation of a location parameter by linear combinations of order statistics, *Sankhya*, Series A., 47(1), 84-89.

[69] Schwert, W.G. (1989). Why does stock volatility change over time?, *Journal of Finance*, 44(5), 1115-1153.

[70] Sentana, E. (1995). Quadratic ARCH, *Review of Economic Studies*, 62, 639-661.

[71] Shalabh (2001). Estimation of bias and standard error of an improved estimator of normal mean, *Metrika*, 54, 43-51.

[72] Sharpe, W. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk, *Journal of Finance*, 19, 425-442.

[73] Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. (in D.R. Cox, David V. Hinkley and Ole E. Barndorff-Nielsen (eds.). *Time Series Models in Econometrics, Finance and Other Fields.* Chapman & Hall: London.)

[74] S-Plus 6.2.1 (2003). Seattle, WA, USA: Insightful Inc.

[75] Thavaneswaran, A. (1991). Tests based on an optimal estimate. Estimating Functions, 189-197. Oxford Statistical Science Series, 7, Oxford Univ. Press, New York.

[76] Thavaneswaran, A. and Abraham, B. (1988). Estimation for nonlinear time series models using estimating functions. *Journal of Time Series Analysis*, 9, 99-108.

[77] Thavaneswaran, A., Appadoo, S.S. and Peiris, S. (2005). Forecasting Volatility, *Statistics and Probability Letters*, 75(1), 1-10.

[78] Thavaneswaran, A., Appadoo, S.S., and Samanta, M. (2005). Random Coefficient GARCH models. *Mathematics and Computer Modelling*, 41(6-7). 723-733.

[79] Thavaneswaran, A. and Ghahramani, M. (2004). Applications of combining estimating functions. In *Proceedings of the International Sri Lankan Conference:*

Visions of Futuristic Methodologies, University of Peradeniya and Royal Melbourne Institute of Technology (RMIT), 515-532.

[80] Thavaneswaran, A. and Ghahramani, M.(2005). Forecasting with structural change, *InterStat*, March 2005.

[81] Thavaneswaran, A. and Heyde, C.C. (1999). Prediction via estimating functions, *Journal of Statistical Planning and Inference*, 77, 89-101.

[82] Thavaneswaran, A. and Peiris, S. (1998). Hypothesis testing for some time-series models: a power comparison, *Statistics and Probability Letters*, 38(2), 151-156.

[83] Thavaneswaran, A. and Thompson, M.E. (2002). Non-normal filtering via estimating function. *Advances on theoretical and methodology aspects of probability and statistics* (Hamilton, ON. 1998). 173-183, Taylor and Francis: London.

[84] Thompson, M.E. and Thavaneswaran, A. (1999). Filtering via estimating functions. *Applied Mathematics Letters*, 12(5), 61-67.

[85] Tobin, J. (1958). Liquidity Preference as Behavior Towards Risk, *Review of Economic Studies*, 25, 65-86.

[86] Tsay, R.S. (2001). *Analysis of Financial Time Series.* John Wiley & Sons, New York.

[87] Wei,W.S. (1994). *Time Series Analysis: Univariate and Multivariate Methods*, Addison Wesley, San Francisco.

[88] Wong, H., Ip, W., Li, Y. (2001). Detection of jumps by wavelets in a heteroscedastic autoregressive model, *Statistics & Probability Letters*, 52, 365-372.

[89] Zakoian, M. (1994). Threshold Heteroscedastic Models, *Journal of Economic Dynamics and Control*, 18, 931-955.

[90] Zeger, S. (1988). A regression model for time series of counts. *Biometrika*, 75, 621-629.

[91] Zivot, E. and Wang, J. (2002). Modeling Financial Time Series with S-PLUS, Springer-Verlag.

Appendix A

Some Properties of the Inverse Gaussian Distribution

A.1 Details in Subsection 2.7.3

The following proposition will be useful in random number generation from an $IG(\mu, \lambda)$ distribution.

Proposition A.1.1 *For $X \sim IG(\mu, \lambda)$, the transformed variable $Y^2 = \dfrac{\lambda(X-\mu)^2}{\mu^2 X}$ is distributed as a χ_1^2.*

Proof: The details are provided in Letac et al.(1985).

Remark A.1.1 *Chhikara and Folks(1989) provide the algorithm for generating random variates from the $IG(\mu, \lambda)$ distribution in the following manner. Consider the following variates.*

$$X_1 = \frac{\mu}{2\lambda}[2\lambda + \mu Y^2 - \sqrt{4\lambda\mu Y^2 + \mu^2 Y^4}] \tag{A.1.1}$$

and

$$X_2 = \frac{\mu^2}{X_1}. \tag{A.1.2}$$

1. *Generate random numbers from the chi-square distribution with one degree of freedom.*

2. *For each random value in step 1, compute the smaller root X_1 given above.*

3. *Perform a Bernoulli trial with probability of 'success' $p = \mu/(\mu + X_1)$.*

4. *If the trial results in a success, the root X_1 is chosen for the random observation from the inverse Gaussian distribution; otherwise the larger root X_2 is chosen.*

Note: The two 'roots' of the equation $Y^2 = \dfrac{\lambda(X-\mu)^2}{\mu^2 X}$ are the following:

$$\begin{aligned} X_1 &= \frac{\mu}{2\lambda}[2\lambda + \mu Y^2 - \sqrt{4\lambda\mu Y^2 + \mu^2 Y^4}] \\ X_2 &= \frac{\mu}{2\lambda}[2\lambda + \mu Y^2 + \sqrt{4\lambda\mu Y^2 + \mu^2 Y^4}] \end{aligned}$$

Simple algebra shows that $X_1 \times X_2 = \mu^2$ and that X_1 is the smaller root.

Theorem A.1.1, makes it possible to reduce the number of parameters and to obtain an inverse Gaussian distribution with a single parameter.

Theorem A.1.1 *Suppose $X \sim IG(\mu, \lambda)$. Let $Z = \lambda X/\mu^2$. Then $Z \sim IG(\delta, \delta^2)$(denoted by $IG(\delta)$ from herein) where $\delta = \lambda/\mu$. Furthermore, Z has the following properties:*

(i) Z has the following density: $f(z;\delta) = \dfrac{\delta}{\sqrt{2\pi z^3}} \exp\left[-\dfrac{1}{2}\left(\sqrt{z} - \dfrac{\delta}{\sqrt{z}}\right)\right]$,

(ii) $E(Z) = \delta$,

(iii) $Var(Z) = \delta$.

Proof: The details are provided in Chhikara and Folks(1989).

Remark A.1.2 *Since the model by Christoffersen et al.(2006) simulates from the $IG(\delta)$ model, we modify X_1 and X_2 in Remark A.1.1 in the following manner:*

$$X_1 = \frac{1}{2\delta}[2\delta^2 + \delta Y^2 - \sqrt{4\delta^3 Y^2 + \delta^2 Y^4}] \tag{A.1.3}$$

and

$$X_2 = \frac{\delta^2}{X_1} \tag{A.1.4}$$

Appendix B

Combining Estimating Functions Proofs in Chapter 3

B.1 Details in Subsection 3.3.1

The estimating functions based separately on m_t and k_t are

$$M = -\frac{1}{\sigma^2}\sum_{t=1}^{T} y_{t-1}m_t, \quad K = -\frac{2\dot{\sigma}}{(\kappa+2)\sigma^3}\sum_{t=1}^{T} k_t,$$

respectively. These were calculated using the following quantities.

$$\mathrm{E}\left(\frac{\partial m_t}{\partial \theta}|\mathcal{F}_{t-1}^y\right) = -y_{t-1} \ , \mathrm{E}\left(m_t^2|\mathcal{F}_{t-1}^y\right) = \sigma^2.$$

Thus,

$$\sum_{t=1}^{T} \mathrm{E}\left(\frac{\partial m_t}{\partial \theta}|\mathcal{F}_{t-1}^y\right) [\mathrm{E}\left(m_t^2|\mathcal{F}_{t-1}^y\right)]^{-1} m_t = -\frac{1}{\sigma^2}\sum_{t=1}^{T} y_{t-1}m_t.$$

To calculate K, we need

$$\mathrm{E}\left(\frac{\partial k_t}{\partial \theta}|\mathcal{F}_{t-1}^y\right) = -2\sigma\dot{\sigma}, \quad \mathrm{E}\left(k_t^2|\mathcal{F}_{t-1}^y\right) = (\kappa+2)\sigma^4.$$

The corresponding martingale informations of M and K follow from the definition.

B.2 Details in Proof of Part (a) of Theorem 3.3.3

Using $m_t = y_t - \theta y_{t-1}$ and $k_t = m_t^2 - \sigma^2$, $\mathrm{Cov}(m_t, k_t|\mathcal{F}_{t-1}^y) = \gamma\sigma^3$. Let

$$\psi_t = k_t - \frac{\mathrm{Cov}(m_t, k_t|\mathcal{F}_{t-1}^y)}{\mathrm{Var}(m_t|\mathcal{F}_{t-1}^y)} m_t.$$

Then, $\psi_t = k_t - \gamma\sigma^3 m_t$ will be uncorrelated with m_t. Consider the elementary estimating function $g_t = m_t + \psi_t$. Using Theorem 3.3.1, the optimal combination is given by the following:

$$\begin{aligned}
g(\underline{y};\theta) &= \sum_{t=1}^{T}\left\{\frac{E\left(\frac{dm_t}{d\theta}|\mathcal{F}_{t-1}^y\right)}{\mathrm{E}(m_t^2|\mathcal{F}_{t-1}^y)}m_t + \frac{E\left(\frac{d\psi_t}{d\theta}|\mathcal{F}_{t-1}^y\right)}{\mathrm{E}\left(\psi_t^2|\mathcal{F}_{t-1}^y\right)}\psi_t\right\} \\
&= \sum_{t=1}^{T}\left\{-\frac{y_{t-1}}{\sigma^2}m_t + \frac{(\gamma\sigma y_{t-1} - 2\sigma\dot{\sigma})}{(\kappa+2-\gamma^2)\sigma^4}\psi_t\right\}
\end{aligned}$$

which reduces to part (i) of (a) in Theorem 3.3.3.

Consider the elementary estimating function

$$g_t = \sigma(2\dot{\sigma}\gamma - (\kappa+2)y_{t-1})m_t + (y_{t-1}\gamma - 2\dot{\sigma})k_t.$$

In order to calculate the martinagle information of the estimating function, we need to compute the conditional expectation of $\dfrac{\partial g_t}{\partial\theta}$. It is easy to show the following:

$$\begin{aligned}
E\left(\frac{\partial}{\partial\theta}2\dot{\sigma}\sigma\gamma m_t\,|\mathcal{F}_{t-1}^y\right) &= -2\dot{\sigma}\sigma\gamma y_{t-1} \\
E\left(\frac{\partial}{\partial\theta} - (\kappa+2)\sigma y_{t-1}m_t\,|\mathcal{F}_{t-1}^y\right) &= (\kappa+2)\sigma y_{t-1}^2 \\
E\left(\frac{\partial}{\partial\theta}y_{t-1}\gamma k_t\,|\mathcal{F}_{t-1}^y\right) &= -2\sigma\dot{\sigma}\gamma y_{t-1}^2 \\
E\left(\frac{\partial}{\partial\theta} - 2\dot{\sigma}k_t\,|\mathcal{F}_{t-1}^y\right) &= 4\sigma(\dot{\sigma})^2
\end{aligned}$$

Therefore,

$$E\left(\frac{\partial g_t}{\partial\theta}\,|\mathcal{F}_{t-1}^y\right) = \sigma(4(\dot{\sigma})^2 - 4\dot{\sigma}\gamma y_{t-1} + (\kappa+2)y_{t-1}^2).$$

In order to calculate $E(g_t^2|\mathcal{F}_{t-1}^y)$, we need the following intermediate calculations.

$$\begin{aligned}
\mathrm{Var}\left(2\dot{\sigma}\sigma\gamma m_t|\mathcal{F}_{t-1}^y\right) &= 4(\dot{\sigma})^2\sigma^4\gamma^2 \\
\mathrm{Var}\left(-\sigma(\kappa+2)y_{t-1}m_t|\mathcal{F}_{t-1}^y\right) &= (\kappa+2)^2 y_{t-1}^2\sigma^4
\end{aligned}$$

$$\mathrm{Var}\left(y_{t-1}\gamma k_t|\mathcal{F}^y_{t-1}\right) = y^2_{t-1}\gamma^2\sigma^4(\kappa+2)$$

$$\mathrm{Var}\left(-2\dot{\sigma}k_t|\mathcal{F}^y_{t-1}\right) = 4(\dot{\sigma})^2\sigma^4(\kappa+2)$$

$$2\mathrm{Cov}\left(\sigma 2\dot{\sigma}\gamma m_t, -(\kappa+2)y_{t-1}m_t\sigma|\mathcal{F}^y_{t-1}\right) = -4\dot{\sigma}\sigma^4\gamma(\kappa+2)y_{t-1}$$

$$2\mathrm{Cov}\left(2\sigma\dot{\sigma}\gamma m_t, y_{t-1}\gamma k_t|\mathcal{F}^y_{t-1}\right) = 4\sigma^4\dot{\sigma}\gamma^3 y_{t-1}$$

$$2\mathrm{Cov}\left(\sigma 2\dot{\sigma}\gamma m_t, -2\dot{\sigma}k_t|\mathcal{F}^y_{t-1}\right) = -8(\dot{\sigma})^2\sigma^4\gamma^2$$

$$2\mathrm{Cov}\left(-(\kappa+2)\sigma y_{t-1}m_t, y_{t-1}\gamma k_t|\mathcal{F}^y_{t-1}\right) = -2(\kappa+2)\gamma^2\sigma^4 y^2_{t-1}$$

$$2\mathrm{Cov}\left(-(\kappa+2)y_{t-1}m_t\sigma, -2\dot{\sigma}k_t|\mathcal{F}^y_{t-1}\right) = 4\dot{\sigma}\sigma^4\gamma(\kappa+2)y_{t-1}$$

$$2\mathrm{Cov}\left(y_{t-1}\gamma k_t, -2\dot{\sigma}k_t|\mathcal{F}^y_{t-1}\right) = -4\dot{\sigma}\gamma(\kappa+2)\sigma^4 y_{t-1}.$$

Thus, $E(g^2_t|\mathcal{F}^y_{t-1}) = \{(\kappa+2)-\gamma^2\}\sigma^4(4(\dot{\sigma})^2-4\dot{\sigma}\gamma y_{t-1}+(\kappa+2)y^2_{t-1})$. The martingale information of the combined estimating function follows directly from the definition of martingale information and this reduces to the expression in (3.3.4).

B.3 Details in Proof of Part (a) of Theorem 3.3.4

$$\mathrm{E}\left(\frac{\partial m_t}{\partial\theta}\right) = -y_{t-1},\quad \mathrm{E}(m^2_t|\mathcal{F}^y_{t-1}) = \sigma^2 h^2_t.$$
$$\mathrm{E}\left(\frac{\partial k_t}{\partial\theta}\right) = -2\dot{\sigma}\sigma h_t,\quad \mathrm{E}(k^2_t|\mathcal{F}^y_{t-1}) = (\kappa+2)\sigma^4 h^2_t.$$

Put $\psi_t = k_t - \dfrac{\mathrm{Cov}(m_t,k_t|\mathcal{F}^y_{t-1})}{\mathrm{E}(m^2_t|\mathcal{F}^y_{t-1})}m_t = k_t - \gamma\sigma h^{1/2}_t m_t$. Then, $\mathrm{E}(\psi^2_t|\mathcal{F}^y_{t-1}) = (\kappa + 2-\gamma^2)\sigma^4 h^2_t$ and $\mathrm{E}\left(\dfrac{\partial\psi_t}{\partial\theta}|\mathcal{F}^y_{t-1}\right) = -2\sigma\dot{\sigma}h_t + \gamma y_{t-1}\sigma h^{1/2}_t$. Using Theorem 3.3.1, the optimal combined estimating function is given by the expression in (i) of part(a) in Theorem 3.3.4.

Let

$$g_t = \left\{ \left(2\sigma\dot{\sigma}\gamma - \sigma(\kappa+2)\frac{y_{t-1}}{\sqrt{h_t}} \right) m_t + \left(\frac{\gamma y_{t-1}}{\sqrt{h_t}} - 2\dot{\sigma} \right) k_t \right\}.$$

After more algebra,

$$\mathrm{E}(g_t^2|\mathcal{F}_{t-1}^y) = \frac{(\kappa+2-\gamma^2)}{\sigma^2}\left\{ 4(\dot{\sigma})^2 + (\kappa+2)\frac{y_{t-1}^2}{h_t} - \frac{4\dot{\sigma}\gamma y_{t-1}}{\sqrt{h_t}} \right\}.$$

$$\mathrm{E}\left(\frac{\partial g_t}{\partial\theta}|\mathcal{F}_{t-1}^y\right) = [4(\dot{\sigma})^2 + (\kappa+2)\frac{y_{t-1}^2}{h_t} - 4\dot{\sigma}\gamma\frac{y_{t-1}}{\sqrt{h_t}}].$$

B.4 Details in Subsection 3.3.3

Some of the details are as follows.

$$\mathrm{E}\left(\frac{\partial m_t}{\partial\theta}|\mathcal{F}_{t-1}^y\right) = -y_{t-1}$$

$$\mathrm{E}\left(\frac{\partial k_t}{\partial\theta}|\mathcal{F}_{t-1}^y\right) = -2\sigma\dot{\sigma}$$

$$\mathrm{E}(m_t^2|\mathcal{F}_{t-1}^y) = \sigma_b^2 y_{t-1}^2 + \sigma^2$$

$$\mathrm{E}(k_t^2|\mathcal{F}_{t-1}^y) = y_{t-1}^4(\kappa_b+2)\sigma_b^4 + 4y_{t-1}^2\sigma^2\sigma_b^2 + (\kappa+2)\sigma^4.$$

$$\begin{aligned}
\mathrm{Cov}(m_t, k_t|\mathcal{F}_{t-1}^y) &= \mathrm{Cov}(m_t, m_t^2|\mathcal{F}_{t-1}^y) \\
&= \mathrm{Cov}(y_{t-1}b_t + \varepsilon_t, y_{t-1}^2 b_t^2 + 2y_{t-1}b_t\varepsilon_t + \varepsilon_t^2|\mathcal{F}_{t-1}^y) \\
&= y_{t-1}^3\mathrm{Cov}(b_t, b_t^2) + \mathrm{Cov}(\varepsilon_t, \varepsilon_t^2) \\
&= y_{t-1}^3\sigma_b^3\gamma_b + \sigma^3\gamma.
\end{aligned}$$

B.5 Details in Proof of Part (b) of Theorem 3.3.3

In order to calculate the optimal estimating function, we need the following calculations.

$$\mathrm{E}\left(\frac{\partial m_t}{\partial \theta}|\mathcal{F}_{t-1}^y\right) = \frac{y_{t-1}^2}{\sigma^2}, \quad \mathrm{E}(m_t^2|\mathcal{F}_{t-1}^y) = \frac{y_{t-1}^2}{\sigma^2}.$$

$$\mathrm{E}\left(\frac{\partial l_t}{\partial \theta}|\mathcal{F}_{t-1}^y\right) = 2f(0)y_{t-1}, \quad \mathrm{E}(l_t^2|\mathcal{F}_{t-1}^y) = 4f^2(0)y_{t-1}^2.$$

$$\mathrm{Cov}(m_t, l_t|\mathcal{F}_{t-1}^y) = -y_{t-1}\rho\sigma$$

Put $\psi_t = l_t - \dfrac{\mathrm{Cov}(m_t, l_t|\mathcal{F}_{t-1}^y)}{\mathrm{E}(m_t^2|\mathcal{F}_{t-1}^y)}m_t$. Then, $\psi_t = l_t + y_{t-1}\dfrac{\rho}{\sigma}m_t$ is uncorrelated with m_t.

$$\mathrm{E}\left(\frac{\partial \psi_t}{\partial \theta}|\mathcal{F}_{t-1}^y\right) = -y_{t-1}^2(2f(0) + \frac{\rho}{\sigma}).$$

$$\mathrm{Var}(\psi_t|\mathcal{F}_{t-1}^y) = y_{t-1}^2(1-\rho^2).$$

The optimal combination is given by

$$\begin{aligned} g(\underline{y};\theta) &= \sum_{t=1}^{n}\left\{\frac{E\left(\frac{dm_t}{d\theta}|\mathcal{F}_{t-1}^y\right)}{E(m_t^2|\mathcal{F}_{t-1}^y)}m_t + \frac{E\left(\frac{d\psi_t}{d\theta}|\mathcal{F}_{t-1}^y\right)}{E(\psi_t^2|\mathcal{F}_{t-1}^y)}\psi_t\right\} \\ &= \frac{1}{\sigma^2(1-\rho^2)}\sum_{t=1}^{n}\left\{-\frac{y_{t-1}}{\sigma^2}m_t - \frac{1}{(1-\rho^2)}(2f(0)+\rho/\sigma)\psi_t.\right\} \end{aligned}$$

After much algebra, this reduces to the expression in (3.3.6).

Let g_t be the tth component in $g(\underline{y};\theta)$. It can be shown that

$$\mathrm{E}\left(\frac{\partial g_t}{\partial \theta}|\mathcal{F}_{t-1}^y\right) = \frac{y_{t-1}^2}{\sigma^2(1-\rho^2)}\{1 + 4f(0)\rho\sigma + 4f^2(0)\sigma^2\}$$

And,

$$\mathrm{Var}(g_t|\mathcal{F}^y_{t-1}) = \frac{y^2_{t-1}}{\sigma^2(1-\rho^2)}(1+4f(0)\rho\sigma+4f^2(0)\sigma^2).$$

So that the martingale information equals

$$\sum_{t=1}^{n} \mathrm{E}^2\left(\frac{\partial g_t}{\partial\theta}|\mathcal{F}^y_{t-1}\right)[\mathrm{E}(g_t^2|\mathcal{F}^y_{t-1})]^{-1}$$

and this reduces to I_g in (3.3.7).

B.6 Details in Proof of Part (b) of Theorem 3.3.4

$$E\left(\frac{\partial m_t}{\partial\theta}|\mathcal{F}^y_{t-1}\right) = -y_{t-1}, \quad E(m_t^2|\mathcal{F}^y_{t-1}) = \sigma^2 h_t.$$

$$E\left(\frac{\partial l_t}{\partial\theta}|\mathcal{F}^y_{t-1}\right) = 2f(0)y_{t-1}, \quad E(l_t^2|\mathcal{F}^y_{t-1}) = 1.$$

$$Cov(m_t, l_t|\mathcal{F}^y_{t-1}) = \rho\sigma\sqrt{h_t}$$

$$\psi_t = l_t - \frac{\rho}{\sigma\sqrt{h_t}}m_t$$

$$E\left(\frac{\partial\psi_t}{\partial\theta}|\mathcal{F}^y_{t-1}\right) = (2f(0)\sigma\sqrt{h_t}+\rho)\frac{y_{t-1}}{\sigma\sqrt{h_t}}, \quad E(\psi_t^2|\mathcal{F}^y_{t-1}) = (1-\rho^2).$$

$$g = \sum_{t=1}^{T} -\frac{y_{t-1}}{\sigma^2 h_t}m_t + \frac{(2f(0)\sigma\sqrt{h_t}+\rho)}{(1-\rho^2)}\frac{y_{t-1}}{\sigma\sqrt{h_t}}(l_t - \frac{\rho}{\sigma\sqrt{h_t}}m_t)$$

Let $g_t = \frac{y_{t-1}}{\sigma^2 h_t}\{(1 + 2f(0)\sigma\rho\sqrt{h_t})m_t - \sigma\sqrt{h_t}(2f(0)\sigma\sqrt{h_t} + \rho)l_t\}$.

$$E\left(\frac{\partial g_t}{\partial \theta}|\mathcal{F}^y_{t-1}\right) = \frac{y^2_{t-1}}{\sigma^2 h_t}[1 + 4f(0)\sigma^2 h_t + 4f(0)\rho\sigma\sqrt{h_t}]$$

$$E(g^2_t|\mathcal{F}^y_{t-1}) = \frac{y^2_{t-1}}{\sigma^4 h^2_t}[1 + 4f(0)\sigma^2 h_t + 4f(0)\rho\sigma\sqrt{h_t}]$$

B.7 Datasets used in Subsection 5.1.2

We illustrate our informal test procedure using several financial datasets. These are taken from the *finmetrics* financial analysis module of S-Plus© version 6.2, and Tsay(2001). We give a brief description of each datasets.

1. *bmw* represents the daily log returns on BMW share price from 1/2/1973 to 7/23/1996. (Note no trading takes place at the weekend.) The data is supplied by *finmetrics.*

2. *dell.s* is a data set with 1261 values, representing daily stock returns of Dell Corporation in percentage points from August 24, 1993, to August 19, 1998. The data is supplied by *finmetrics.*

3. *hp.s* is 2000 data points from stocks, representing the daily returns for Hewlett-Packard(HP) from Feburary 2, 1984, to December 31, 1991. The data is supplied by *finmetrics.*

4. *siemens* represents the daily log returns of Siemens share price from 1/2/1973 to 7/23/1996. Note no trading takes place at the weekend. The data is supplied by *finmetrics.*

5. *merck.s* is 2000 data points representing the daily returns of the pharmaceutical company Merck from Feburary 2, 1984, to December 31, 1991. The data is supplied by *finmetrics.*

6. *USCN.FP.lag1* represents the one month forward premium between U.S. Dollar and Canadian Dollar during March 1976 to June 1996. The data is supplied by *finmetrics.*

7. *USUK.FP.lag1* represents the one month forward premium between U.S. Dollar and British Pound during March 1976 to June 1996. The data is supplied by *finmetrics.*

8. *hkja.dat* represents 491 daily log returns of the Japan market and is taken from Tsay(2001).

9. *exch.perc.dat* represents 10-m log returns of FX (Mark-US) and is taken from Tsay(2001).

10. *d-ibmln.dat* represents Daily log returns of IBM stock and is taken from Tsay(2001).

11. The following datasets are columns of the dataset *DowJones30* representing the closing prices of thirty stocks in Dow Jones Industrial Average in *finmetrics.* This is a daily "timeSeries" object from January 2, 1991 to January 2, 2001, with thirty columns representing the closing prices of thirty stocks in Dow Jones Industrial Average including: Alcoa, Inc., AT&T, Boeing, Caterpillar, Inc., Honeywell, Wal-mart and Philip Morris.

B.8 Time Series Plots of Datasets in Subsection 5.1.2

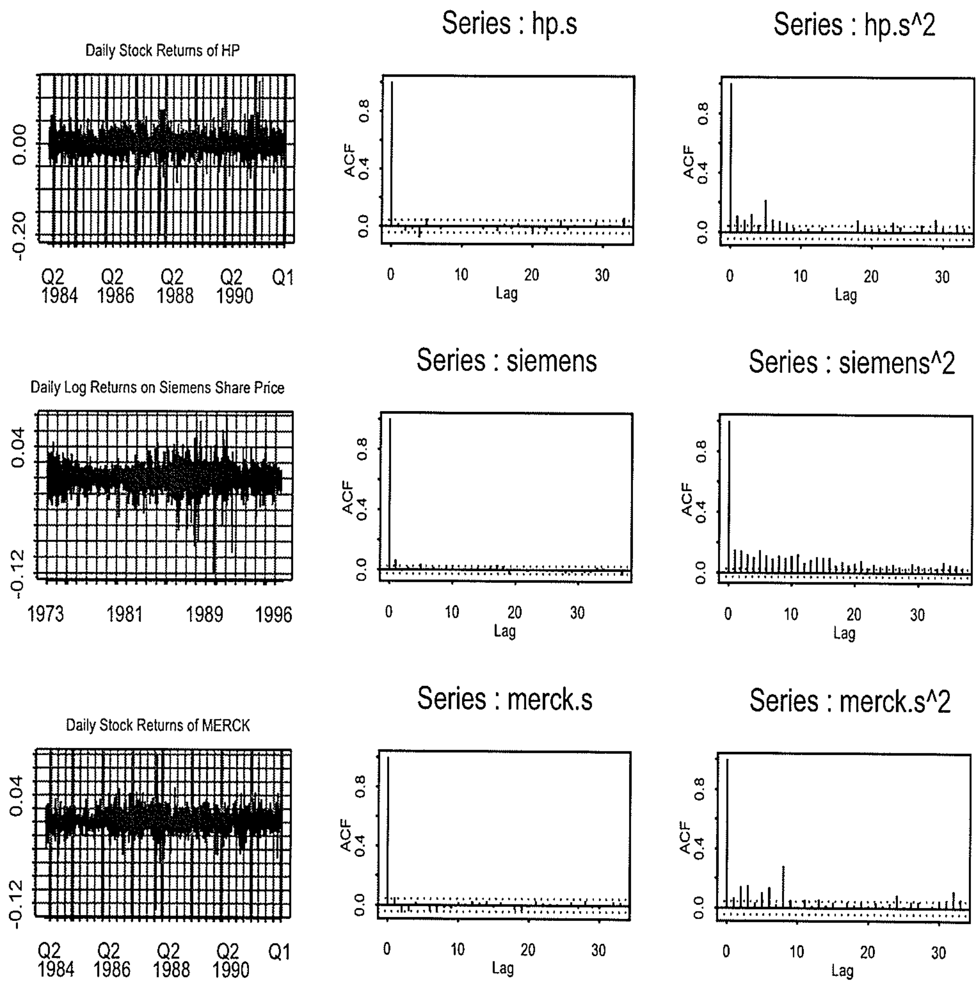

Figure B.1: Daily log returns.

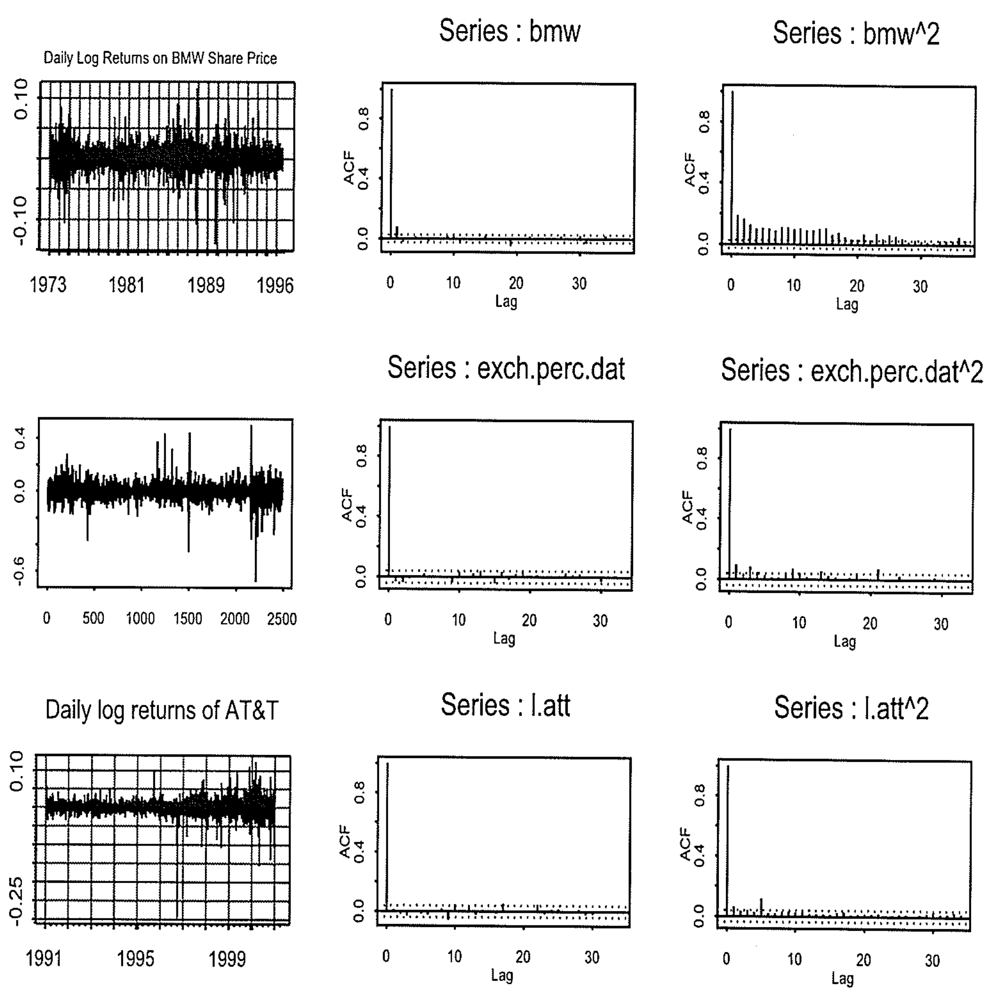

Figure B.2: Daily log returns.

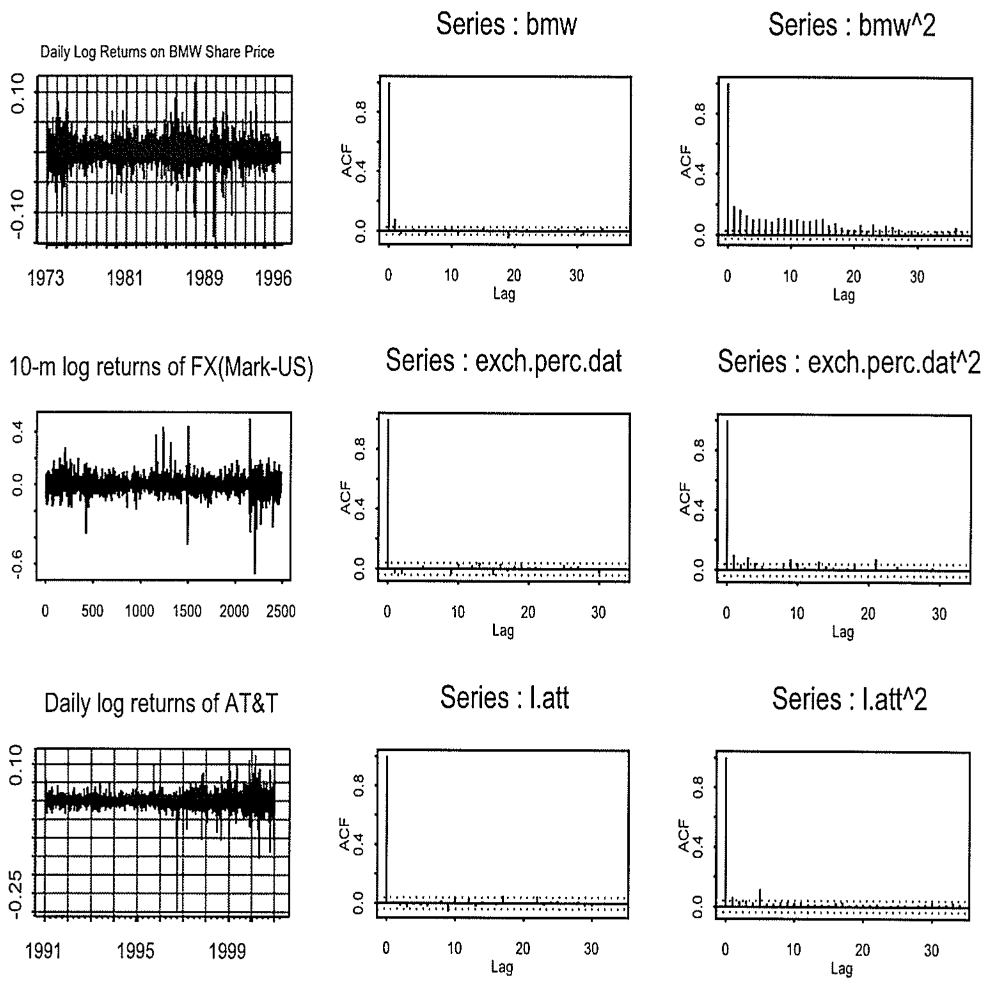

Figure B.3: Daily log returns.

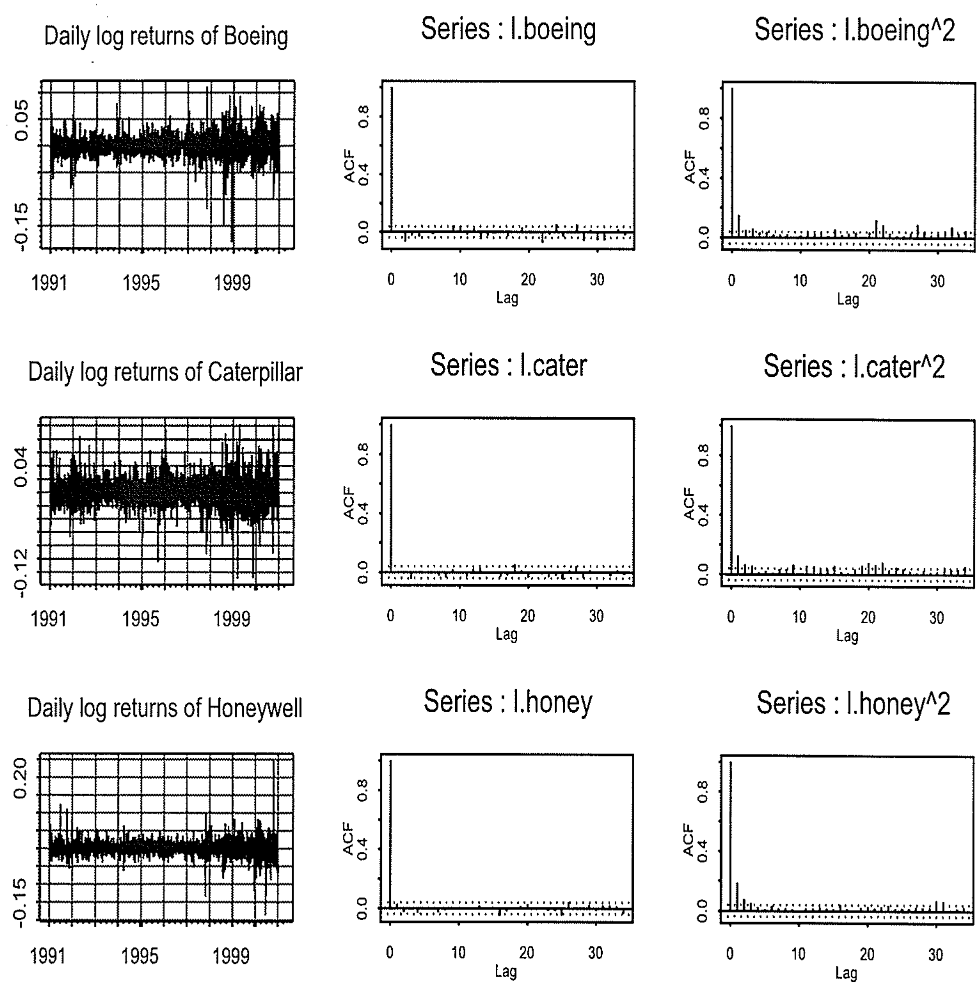

Figure B.4: Daily log returns.

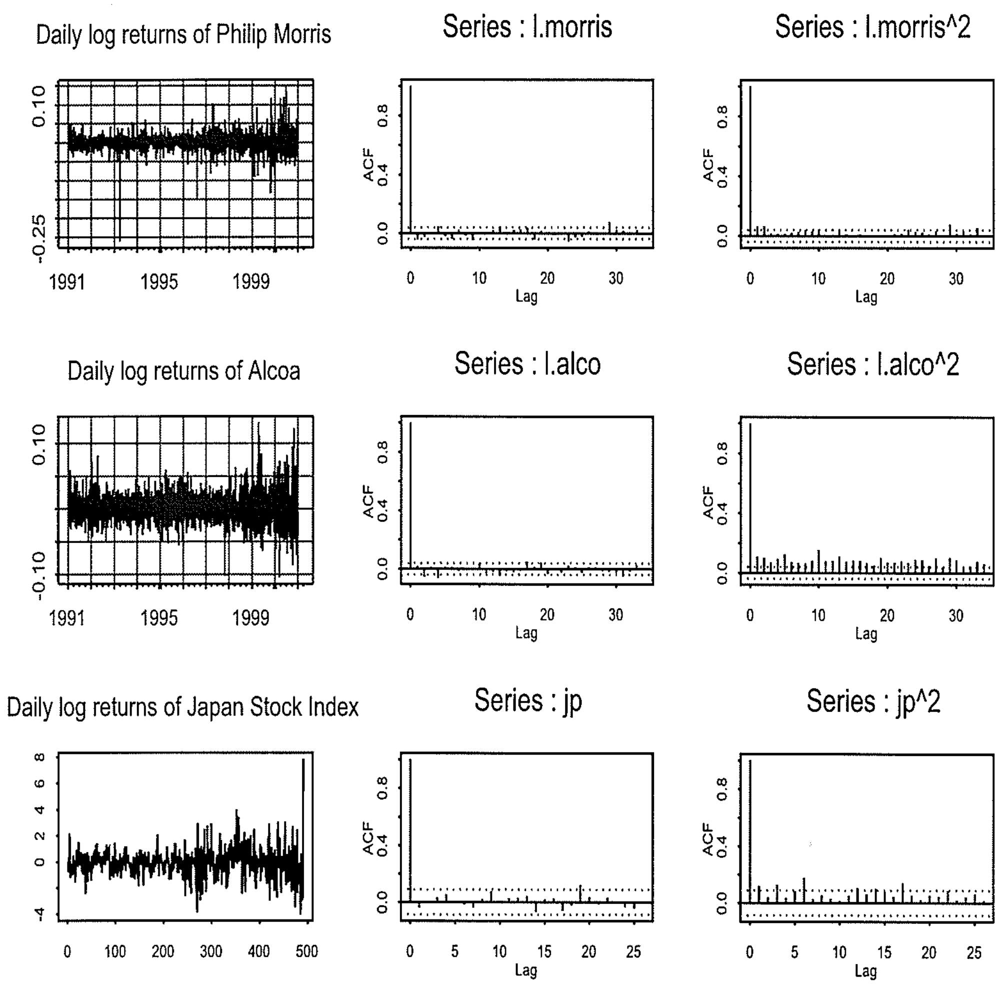

Figure B.5: Daily log returns.

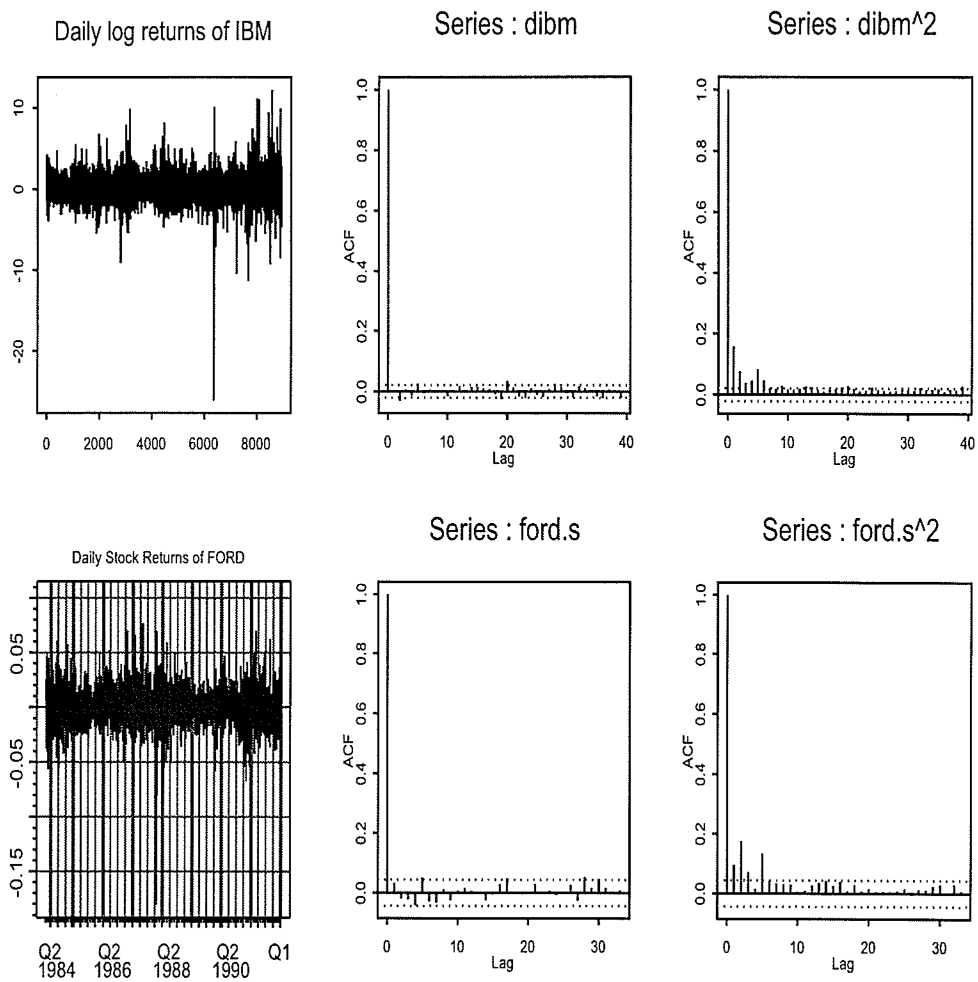

Figure B.6: Daily log returns.

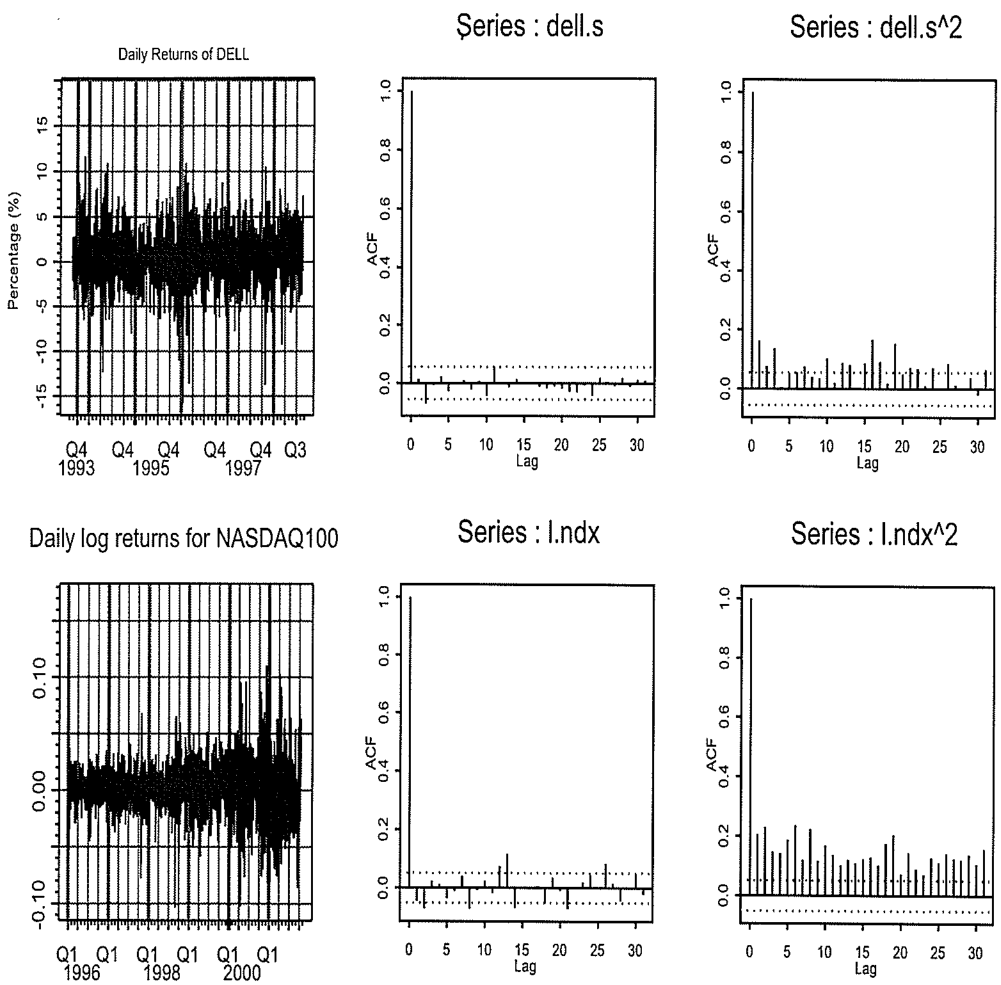

Figure B.7: Daily log returns.