Generalized Autoregressive Conditional Heteroscedastic Time Series Models

by

Michael S. Lo

B.Sc., Simon Fraser University 2000

A PROJECT SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in the Department

of

Statistics and Actuarial Science

SIMON FRASER UNIVERSITY

April 2003

APPROVAL

Name: Michael S. Lo

Degree: Master of Science

Title of project: Generalized Autoregressive Conditional Heteroscedastic Time Series Models

Examining Committee: Dr. K. Laurence Weldon
Chair

____

Dr. Richard A. Lockhart
Senior Supervisor
Simon Fraser University

____

Dr. Michael A. Stephens
Simon Fraser University

____

Dr. Carl J. Schwarz
External Examiner
Simon Fraser University

Date Approved: ______________________________

Abstract

Autoregressive and Moving Average time series models and their combination are reviewed. Autoregressive Conditional Heteroscedastic (ARCH) and Generalized Autoregressive Conditional Heteroscedastic (GARCH) models are extensions of these models. These are defined and compared to the class of Autoregressive Moving Average models. Maximum likelihood estimation of parameters is examined. Conditions for existence and stationarity of GARCH models are discussed and the moments of the observations and the conditional variance are derived. Characteristics of low order GARCH models are explored further through simulations with different initial parameter values. As examples, GARCH models with different orders are fitted to the Standard & Poor's 500 Stock Price Index.

Acknowledgments

First, I would like to thank Derek for encouraging me to apply for the M.Sc. program at SFU, Tim and Randy for explaining the benefits and essence of a master's degree and Julia for telling me what to expect from graduate school. I would also like to thank Rick for being persuaded to admit me, Charmaine for giving me an opportunity to fulfill my dreams, Carl for his confidence in me and Larry and Robin for showing me their enthusiasm in teaching. Of course, I thank Sylvia and Sadika for having the efficiency and accuracy next to no one!

Special thanks to my mom Yee-Ling and my dad Shong-Pak for supporting me throughout all these years both spiritually and financially, to my sister Linda for taking care of my parents during my absence in the house, to Mavis for once letting me be selfish and to my friend Tim for introducing the GARCH model to me.

I would like to thank all the graduate students whom I got to know:
Ellen (and George) for showing me how much one can do in a day, Wilson (and Helen) for showing me that guys also read online romance novels, Ruthie (and Jeff) for showing me *Joy*, Monica (and Rene) for always cliticizing my Spanish plonunciation, for serving me authentic Mexican food (Glacias!) and for making sure that I do not date Chinese girls with dyed hair and wearing Aritzia's clothes, Phil for demonstrating to me that "it's never too late", Milena (and Matthew) for giving me a warm hug besides the Thames, Grace for showing me that bigger does not always mean better, Farah for telling me that she is mean and I am average, Travis for showing me

the *dark* side of the world, Jason for keeping me awake in the office by screaming out loud, Jason for teaching me how to drink, Wanlin for letting me ride on her scooter in the heart of Taipei city, May for teaching me how to count properly by hand, Greg for sharing his thought "it doesn't have to be good, it just has to be done", Kelly for spending the cold, windy and no-fireworks Canada Day in front of the Parliament Building with me, Laura for showing me how things should be done, Jason for giving me two bus passes in Ottawa, Jean for training my ears' sensitivity by speaking softly (I can now hear her walking!), Simon (and Alyssa) for constantly raising the bar and for making me veggie pizzas, Jacquie for keeping me company on ICQ and for reminding me to take a look at who comes into K9501, Farouk for convincing me chips disappear mysteriously without reasons, Sandy for sharing coupons with me, Eugenia for downloading papers for me, for keeping me updated on gossip at UBC and for standing beside me against those whiny UBC students, Jeremy for proving that there are not only Chinese residing in Richmond and for showing me how much alcohol one can consume without passing out, Chunfang for understanding my horrible English and Steve for reminding me that there is a place called Mission. Thank you all!

How can I end this acknowledgment without thanking my girlfriend Wendy for patiently and quietly helping and supporting me to accomplish what I wanted to do.

Special thanks to my co-supervisor/friend Michael for showing me every facet of the world, for teaching me how to do research properly and for guiding me along as if I were his beloved son.

Last, but definitely not least, I would like to thank sincerely my senior supervisor Richard for having confidence in my teaching and patience in my sluggish progress and for showing me how little I knew on just about anything!

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Autoregressive (AR), Moving Average (MA) and the mixed autoregressive moving average (ARMA) models are often very useful in modelling general time series. However, they all have the assumption of homoscedasticity (or equal variance) for the errors; this is not appropriate when dealing with the financial market variables such as the stock price indices or currency exchange rates. These financial market variables typically have three characteristics which general time series models have failed to consider.

1. The unconditional distribution of financial time series such as the stock price returns X_t has heavier tails than the normal distribution.

2. Values of X_t do not have much correlation, but values of X_t^2 are highly correlated.

3. The changes in X_t tend to cluster. Large (small) changes in X_t tend to be followed by large (small) changes, as documented by Mandelbrot (1963).

One of the earliest time series models allowing for heteroscedasticity is the Autoregressive Conditional Heteroscedastic (ARCH) model introduced by Engle (1982). The

ARCH models have the ability to capture all the above characteristics in financial market variables. Bollerslev (1986) extended this idea into Generalized Autoregressive Conditional Heteroscedastic (GARCH) models which give more parsimonious results than ARCH models, similar to the situation where ARMA models are preferred over AR models.

In Chapter 2, we will describe the definitions and maximum likelihood estimation methods of general time series models like the AR, MA and ARMA models. Then the definitions of ARCH and GARCH models will be introduced; the maximum likelihood estimation for ARCH parameters will be discussed. The comparison between high order ARCH models with the GARCH$(1,1)$ models will also be addressed. In Chapter 3, we will describe the properties for GARCH$(1,1)$ models; existence and stationarity conditions are discussed and moments of the data will be derived. Typical simulation results will be used to discuss the properties of the estimates of the parameters. In Chapter 4, we fit the Standard & Poor's (S&P) 500 Stock Price Index for the period from January 2, 1990 to December 29, 2000 by ARCH(1) and GARCH$(1,1)$ models and some diagnostics will be applied. Alternative methods of diagnostic checking on the residuals of the fitted model will also be discussed.

Chapter 2

Time Series Concepts and Models

Let $\mathcal{S}$ be a subset of the real numbers. For every $t \in \mathcal{S}$, let $X_t(\omega)$ be a random variable defined on a probability space $\{\Omega : \omega \in \Omega\}$; then the stochastic process $\{X_t(\omega) : t \in \mathcal{S}\}$ is called a time series. For any given ω, X_t is the realization at time t. This will always be a time series with $\mathcal{S} \equiv \mathcal{I} \subset \{0, \pm 1, \cdots\}$; X_t is then called a time series in discrete time. Thus observations are $X_1, \cdots, X_T$ and we will assume each X_t, $t = 1, \cdots, T$, is real-valued. Often our models will require the existence of unobserved X_t values for $t \leq 0$ or $t > T$.

In Section 2.1, we describe the stationarity conditions used in general time series contexts. We then define some general time series models and the maximum likelihood method for estimating the parameters. In Section 2.3, the general ARCH(q) and GARCH(p, q) models are defined.

2.1 Stationarity

Suppose $\boldsymbol{X} = (X_1, \cdots, X_T)'$ has a multivariate normal distribution with mean vector $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_T)'$ and a $T \times T$ variance-covariance matrix Σ, where $'$ denotes the transpose of a vector or matrix. There are T data points but $T+T(T+1)/2$ parameters to estimate which is not feasible. Therefore, some assumptions on the process $\{X_t\}$ must be made which permits us to obtain reasonable estimates on the parameters; stationarity is the most commonly used such assumption in time series contexts.

Definition 2.1 (Joint Distribution). *The joint distribution function of $X_1, \cdots, X_T$ is given by*

$$F_{X_1,X_2,\cdots,X_T}(x_1, x_2, \cdots, x_T) = P(X_1 \leq x_1, X_2 \leq x_2, \cdots, X_T \leq x_T)$$

Definition 2.2 (Strict Stationarity). *A process is said to be strictly stationary if the joint distribution of $X_1, X_2, \cdots, X_k$ is the same as the joint distribution of $X_{t+1}, X_{t+2}, \cdots, X_{t+k}$, evaluated at the same set of points $x_1, x_2, \cdots, x_k$, i.e.*

$$F_{X_1,X_2,\cdots,X_k}(x_1, x_2, \cdots, x_k) = F_{X_{t+1},X_{t+2},\cdots,X_{t+k}}(x_1, x_2, \cdots, x_k)$$

for all t and for all k.

Definition 2.3 (Wide Sense Stationarity). *A process is said to be second order (or wide sense) stationary if*

$$E(X_t) = \mu \quad and \quad V(X_t) = \sigma^2$$

for all t and, for all k,

$$Cov(X_t, X_{t+k}) = Cov(X_{t+1}, X_{t+1+k}) = Cov(X_{t+2}, X_{t+2+k}) = \cdots$$

is a function of the time lag k only and does not depend on time t.

2.2 Standard Time Series Models

Historically, three basic time series models have been used to describe data. They are the Autoregressive, the Moving Average and the Autoregressive Moving Average models.

2.2.1 General Autoregressive Models

In the autoregressive (AR) time series model, an observation X_t is directly related to p previous observations by

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t. \tag{2.1}$$

This is the Autoregressive series of order p, AR(p). In this model, ϵ_t is called the error. When the errors are independent, have normal distributions with mean zero and constant variance σ_ϵ^2, they are called *white noise*. In our models, ϵ_t is assumed to be white noise.

The model (2.1) can then be expressed as:

$$\begin{aligned} X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} &= \epsilon_t \\ \boldsymbol{\phi}_p(B) X_t &= \epsilon_t \end{aligned}$$

where $\boldsymbol{\phi}_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ and B is the backshift operator on time t; for example, $BX_t = X_{t-1}$, $B^2 X_t = X_{t-2}$ and so on.

For the AR(p) process from (2.1) to be stationary, the roots of $\boldsymbol{\phi}_p(B) = 0$ must lie outside the unit circle. For illustration, an AR(1) process defined as

$$X_t = \phi_1 X_{t-1} + \epsilon_t \tag{2.2}$$

will be used but the results following can be generalized into AR processes with higher order. Rearranging (2.2) gives

$$(1 - \phi_1 B) X_t = \epsilon_t.$$

Hence, $1 - \phi_1 B = 0$ gives $B = 1/\phi_1$. Suppose that $|\phi_1| < 1$, then the root of $\boldsymbol{\phi}_1(B) = 0$ is greater than one, or lies outside the unit circle, and thus the AR(1) process from (2.2) is stationary.

Note that, from (2.2), X_t is defined recursively from its previous observations; it can be expressed as:

$$
\begin{aligned}
X_t &= \epsilon_t + \phi_1 X_{t-1} \\
&= \epsilon_t + \phi_1(\epsilon_{t-1} + \phi_1 X_{t-2}) \\
&= \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 X_{t-2} \\
&= \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 (\epsilon_{t-2} + \phi_1 X_{t-3}) \\
&= \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + \phi_1^3 X_{t-3} \\
&\quad \vdots \\
&= \sum_{j=0}^{k} \phi_1^j \epsilon_{t-j} + \phi_1^{k+1} X_{t-k-1}.
\end{aligned}
$$

Since $|\phi_1| < 1$, allow k to go to infinity; then

$$X_t = \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}.$$

It can be shown that if we define the process $\{X_t\}$ via this formula then X_t satisfies the recursive identity (2.2) provided $|\phi_1| < 1$; and $\{X_t\}$ is stationary.

The variance of X_t can then be found as:

$$
\begin{aligned}
V(X_t) &= \sum_{j=0}^{\infty} \phi_1^{2j} V(\epsilon_{t-j}) \\
&= \sigma_\epsilon^2 \sum_{j=0}^{\infty} \phi_1^{2j} \\
&= \frac{\sigma_\epsilon^2}{1 - \phi_1^2}
\end{aligned}
$$

This is finite and positive if $|\phi_1| < 1$.

We then examine how the parameters ϕ_1 and σ_ϵ^2 for the AR(1) model can be estimated. The joint density of the observations $X_1, \cdots, X_T$ can be written as the product of their conditional densities:

$$\begin{aligned} f_{X_1,\cdots,X_T}(x_1,\cdots,x_T) &= f_{X_T|X_1,\cdots,X_{T-1}}(x_T|x_1,\cdots,x_{T-1}) \times \\ &\quad f_{X_{T-1}|X_1,\cdots,X_{T-2}}(x_{T-1}|x_1,\cdots,x_{T-2}) \times \\ &\quad \cdots \times f_{X_2|X_1}(x_2|x_1) \times f_{X_1}(x_1). \end{aligned}$$

For any $k = 2, \cdots, T$, the conditional density of X_k, given $X_1, \cdots, X_{k-1}$, is

$$f_{X_k|X_1,\cdots,X_{k-1}}(x_k|x_1,\cdots,x_{k-1}) = \frac{1}{\sqrt{2\pi}\sigma_\epsilon} \exp\left\{-\frac{(x_k - \phi_1 x_{k-1})^2}{2\sigma_\epsilon^2}\right\}$$

and the marginal density for X_1 is

$$f_{X_1}(x_1) = \frac{1}{\sqrt{2\pi}\sigma_\epsilon/\sqrt{1-\phi_1^2}} \exp\left\{-\frac{x_1^2}{2\sigma_\epsilon^2/(1-\phi_1^2)}\right\}.$$

The marginal density of X_1 is usually dropped from the overall likelihood function for simplicity because its contribution on the likelihood function is negligible when the number of observations is large. The conditional likelihood, conditional on X_1, is therefore

$$\begin{aligned} L(\phi_1, \sigma_\epsilon^2) &= \prod_{j=2}^{T} f_{X_j|X_1,\cdots,X_{j-1}}(x_j|x_1,\cdots,x_{j-1}) \\ &= \prod_{j=2}^{T} \frac{1}{\sqrt{2\pi}\sigma_\epsilon} \exp\left\{-\frac{(x_j - \phi_1 x_{j-1})^2}{2\sigma_\epsilon^2}\right\} \end{aligned}$$

and the log likelihood function, neglecting the constant term, is

$$l(\phi_1, \sigma_\epsilon^2) = -(T-1)\log\sigma_\epsilon - \frac{1}{2\sigma_\epsilon^2}\sum_{j=2}^{T}(x_j - \phi_1 x_{j-1})^2.$$

We can then use the maximum likelihood to find the estimates $\hat{\phi}_1$ and $\hat{\sigma}_\epsilon^2$ by solving the derivatives of the log likelihood function $\dfrac{\partial l}{\partial \phi_1} = 0$ and $\dfrac{\partial l}{\partial \sigma_\epsilon^2} = 0$ respectively.

2.2.2 General Moving Average Models

Another common model is the Moving Average series of order q, MA(q), defined by:

$$X_t = \epsilon_t - \theta_1\, \epsilon_{t-1} - \cdots - \theta_q\, \epsilon_{t-q}, \tag{2.3}$$

where the ϵ_t's are white noise. Using the backshift operator B on time t, the model from (2.3) can also be expressed as:

$$X_t = \boldsymbol{\theta}_q(B)\, \epsilon_t$$

where $\boldsymbol{\theta}_q(B) = 1 - \theta_1\, B - \cdots - \theta_q\, B^q$. If the roots of $\boldsymbol{\theta}_q(B) = 0$ lie outside the unit circle, then the MA(q) process is said to be *invertible*, meaning that X_t can be written as an infinite order AR process in terms of θ_j for $j = 1, \cdots, q$.

For the case of MA(1),

$$X_t = \epsilon_t - \theta_1\, \epsilon_{t-1}, \tag{2.4}$$

the condition $|\theta_1| < 1$ is sufficient for the invertibility of the process. This can be shown by rearranging (2.4) to give

$$\begin{aligned}
\epsilon_t &= X_t + \theta_1\, \epsilon_{t-1} \\
&= X_t + \theta_1\, (X_{t-1} + \theta_1\, \epsilon_{t-2}) \\
&= X_t + \theta_1\, X_{t-1} + \theta_1^2\, \epsilon_{t-2} \\
&\quad\vdots \\
&= X_t + \sum_{j=1}^{k} \theta_1^j\, X_{t-j} + \theta_1^{k+1}\, \epsilon_{t-k-1}.
\end{aligned}$$

Since $|\theta_1| < 1$, we will let k go to infinity; ϵ_t becomes

$$\epsilon_t = X_t + \sum_{j=1}^{\infty} \theta_1^j\, X_{t-j};$$

this can be written as

$$X_t = -\sum_{j=1}^{\infty} \theta_1^j\, X_{t-j} + \epsilon_t.$$

This is an AR(∞) model $X_t = \sum_{j=1}^{\infty} \phi_j^*\, X_{t-j} + \epsilon_t$ where $\phi_j^* = -\theta_1^j$ for $j = 1, \cdots, \infty$.

By comparing with the AR(1) process, it follows that invertible moving average processes are always stationary. The definition of X from (2.3) is explicit so X is guaranteed to exist.

In order to estimate θ_1 and σ_ϵ^2 in the MA(1) model, we first find the joint density of the observations $X_1, \cdots, X_T$; this is a multivariate normal distribution. Therefore, the likelihood function can be written as:

$$\begin{aligned} L(\theta_1, \sigma_\epsilon^2) &= f_{\epsilon_0, X_1, \cdots, X_T}(\epsilon_0, x_1, \cdots, x_T) \\ &= f_{\epsilon_0, \epsilon_1, \cdots, \epsilon_T}(\epsilon_0, x_1 + \theta_1\, \epsilon_0, x_2 + \theta_1(x_1 + \theta_1\, \epsilon_0), \cdots) \times |J| \\ &= \prod_{j=0}^{T} \frac{1}{\sqrt{2\pi}\sigma_\epsilon} \exp\left\{-\frac{\epsilon_j^{*2}}{2\sigma_\epsilon^2}\right\} \times |J| \end{aligned} \tag{2.5}$$

where $|J|$ is the absolute value of the determinant of the appropriate Jacobian matrix connecting the X_t and the ϵ_t, and $\epsilon_j^* = x_j + \theta_1\, \epsilon_{j-1}^*$ for $j = 1, \cdots, T$, are obtained recursively.

This likelihood function from (2.5), however, cannot be evaluated since ϵ_0 is unknown. There are basically three ways to solve this problem. One way is to use its expected value $E(\epsilon_0) = 0$ and treat the result as if it were the true likelihood function. Another way is to treat ϵ_0 as a parameter and estimate it together with θ_1 and σ_ϵ^2 using maximum likelihood. The third way is the combination of the two methods:

1. Substitute ϵ_0 by its expected value $E(\epsilon_0) = 0$ into the likelihood function (2.5).

2. Obtain estimates $\hat{\theta}_1$ and $\hat{\sigma}_\epsilon^2$ using maximum likelihood.

3. Compute the expected value of ϵ_0 given all the observations and the estimated parameters at that iteration, i.e. $\hat{\epsilon}_0 = E(\epsilon_0|\hat{\theta}_1, \hat{\sigma}_\epsilon^2, X_1, \cdots, X_T)$.

4. Insert $\hat{\epsilon}_0$ back to the likelihood function (2.5).

5. Repeat Steps (2) to (4) until all three estimates, $\hat{\theta}_1$, $\hat{\sigma}_\epsilon^2$ and $\hat{\epsilon}_0$ converge.

2.2.3 General Autoregressive Moving Average Models

The process $\{X_t\}$ is an Autoregressive Moving Average process, ARMA(p, q), if it satisfies the formula

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} = \epsilon_t - \theta_1 \epsilon_{t-1} - \cdots - \theta_q \epsilon_{t-q} \quad (2.6)$$

where ϵ_t's are white noise. Formula (2.6) can be written as $\boldsymbol{\phi}_p(B) X_t = \boldsymbol{\theta}_q(B) \epsilon_t$ using backshift operator B, $\boldsymbol{\phi}_p(B)$ and $\boldsymbol{\theta}_q(B)$ defined above.

Because $1 + \sum_{j=1}^{q} \theta_j^2 < \infty$, the moving average terms on the right hand side of (2.6) will not affect the condition for stationarity of an autoregressive process, Wei (1994). Thus, equation (2.6) will define a stationary process provided that $\boldsymbol{\phi}_p(B) = 0$ has all the roots lying outside the unit circle. Similarly, the roots of $\boldsymbol{\theta}_q(B) = 0$ must lie outside the unit circle if the process is to be invertible.

Sometimes, models like autoregressive or moving average alone do not give parsimonious results when fitting the data. Therefore, the ARMA models with small p and q are preferred over AR models with high order, for example. An ARMA$(1, 1)$ model defined as

$$X_t - \phi_1 X_{t-1} = \epsilon_t - \theta_1 \epsilon_{t-1}$$

is used here for illustration and it can be expressed as:

$$\begin{aligned}
\epsilon_t &= X_t - \phi_1 X_{t-1} + \theta_1 \epsilon_{t-1} \\
&= X_t - \phi_1 X_{t-1} + \theta_1 (X_{t-1} - \phi_1 X_{t-2} + \theta_1 \epsilon_{t-2}) \\
&= X_t - (\phi_1 - \theta_1) X_{t-1} - \phi_1 \theta_1 X_{t-2} + \theta_1^2 \epsilon_{t-2} \\
&= X_t - (\phi_1 - \theta_1) X_{t-1} - \phi_1 \theta_1 X_{t-2} + \theta_1^2 (X_{t-2} - \phi_1 X_{t-3} + \theta_1 \epsilon_{t-3}) \\
&= X_t - (\phi_1 - \theta_1) X_{t-1} - (\phi_1 - \theta_1) \theta_1 X_{t-2} - \phi_1 \theta_1^2 X_{t-3} + \theta_1^3 \epsilon_{t-3} \\
&\;\;\vdots \\
&= X_t - (\phi_1 - \theta_1) \sum_{j=1}^{k} \theta_1^{j-1} X_{t-j} - \phi_1 \theta_1^k X_{t-k-1} + \theta_1^{k+1} \epsilon_{t-k-1}.
\end{aligned}$$

Suppose $|\theta_1| < 1$ and let k go to infinity; then

$$\epsilon_t = X_t - (\phi_1 - \theta_1) \sum_{j=1}^{\infty} \theta_1^{j-1} X_{t-j}$$

and therefore

$$X_t = (\phi_1 - \theta_1) \sum_{j=1}^{\infty} \theta_1^{j-1} X_{t-j} + \epsilon_t.$$

This is an AR(∞) model $X_t = \sum_{j=1}^{\infty} \phi_j^* X_{t-j} + \epsilon_t$ with $\phi_j^* = (\phi_1 - \theta_1)\, \theta_1^{j-1}$ for $j = 1, \cdots, \infty$. This suggests that an ARMA$(1,1)$ model may sometimes be a good approximation to higher order AR models.

The parameters ϕ_1, θ_1 and σ_ϵ^2 from an ARMA$(1,1)$ model can also be estimated by maximum likelihood. As in the AR(1) case, the joint density can be written as the product of their conditional densities:

$$\begin{aligned} f_{X_1,\cdots,X_T}(x_1,\cdots,x_T) &= f_{X_T|X_1,\cdots,X_{T-1}}(x_T|x_1,\cdots,x_{T-1}) \times \\ & \quad f_{X_{T-1}|X_1,\cdots,X_{T-2}}(x_{T-1}|x_1,\cdots,x_{T-2}) \times \\ & \quad \cdots \times f_{X_2|X_1}(x_2|x_1) \times f_{X_1}(x_1). \end{aligned}$$

The conditional density of X_k, for an arbitrary $k = 2, \cdots, T$, conditional on $X_1, \cdots, X_{k-1}$ and ϵ_1, is

$$f_{X_k|X_1,\cdots,X_{k-1},\epsilon_1}(x_k|x_1,\cdots,x_{k-1},\epsilon_1) = \frac{1}{\sqrt{2\pi}\sigma_\epsilon} \exp\left\{-\frac{(x_k - \phi_1\, x_{k-1} + \theta_1\, \epsilon_{k-1})^2}{2\sigma_\epsilon^2}\right\}.$$

The marginal density of X_1 is dropped for simplicity as in the AR(1) case. Therefore, the likelihood function, conditioning on X_1 and ϵ_1, is:

$$\begin{aligned} L(\phi_1, \theta_1, \sigma_\epsilon^2) &= f_{X_2,\cdots,X_T|X_1,\epsilon_1}(x_2,\cdots,x_T|x_1,\epsilon_1) \\ &= \prod_{j=2}^{T} f_{X_j|X_1,\cdots,X_{j-1},\epsilon_1}(x_j|x_1,\cdots,x_{j-1},\epsilon_1) \\ &= \prod_{j=2}^{T} \frac{1}{\sqrt{2\pi}\sigma_\epsilon} \exp\left\{-\frac{(x_j - \phi_1\, x_{j-1} + \theta_1\, \epsilon_{j-1}^*)^2}{2\sigma_\epsilon^2}\right\} \end{aligned}$$

where $\epsilon^*_{j-1} = x_{j-1} - \phi_1 \, x_{j-2} + \theta_1 \, \epsilon^*_{j-2}$ for $j = 3, \cdots, T$, are obtained recursively. As for MA(1), we substitute the expected value $E(\epsilon_1) = 0$ for ϵ_1 since it is unknown, and estimate the parameters iteratively until the estimates converge.

2.3 Financial Time Series Models

All three models described above are often very useful in modeling time series in general. However, they have the assumption of constant error variance, σ^2_ϵ. This is considered to be unrealistic in many areas of economics and finance. Therefore, Autoregressive Conditional Heteroscedastic (ARCH) models and Generalized ARCH (GARCH) models which allow variance to vary over time have been proposed, in particular to model financial market variables.

2.3.1 Autoregressive Conditional Heteroscedastic (ARCH) Models

Suppose $X_1, X_2, \cdots, X_T$ are the time series observations and let $\mathcal{F}_t$ be the set of X_t up to time t, including X_t for $t \leq 0$. As defined by Engle (1982), the process $\{X_t\}$ is an Autoregressive Conditional Heteroscedastic process of order q, ARCH(q), if:

$$\begin{aligned} X_t | \mathcal{F}_{t-1} &\sim N(0, h_t), \qquad \text{with} \\ h_t &= \alpha_0 + \alpha_1 \, X^2_{t-1} + \cdots + \alpha_q \, X^2_{t-q} \\ &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \, X^2_{t-i} \end{aligned} \tag{2.7}$$

where $q > 0$, $\alpha_0 > 0$ and $\alpha_i \geq 0$ for $i = 1, \cdots, q$. The conditions $\alpha_0 > 0$ and $\alpha_i \geq 0$ are needed to guarantee that the conditional variance $h_t > 0$.

It is obvious from (2.7) that the conditional expectation and variance of X_t are:

$$\begin{aligned} E(X_t|\mathcal{F}_{t-1}) &= 0 \\ V(X_t|\mathcal{F}_{t-1}) &= E(X_t^2|\mathcal{F}_{t-1}) = h_t. \end{aligned}$$

In the financial literature, the conditional variance h_t is called the volatility.

The simplest model is the ARCH(1) model:

$$\begin{aligned} X_t|\mathcal{F}_{t-1} &\sim N(0, h_t), \qquad \text{with} \\ h_t &= \alpha_0 + \alpha_1 X_{t-1}^2 \end{aligned} \tag{2.8}$$

and the parameters α_0 and α_1 can be estimated by maximum likelihood. The joint density of the observations $X_1, \cdots, X_T$ is

$$f_{X_1,\cdots,X_T}(x_1, \cdots, x_T) = \left\{ \prod_{j=2}^{T} f_{X_j|X_1,\cdots,X_{j-1}}(x_j|x_1, \cdots, x_{j-1}) \right\} \times f_{X_1}(x_1).$$

For $k = 2, \cdots, T$, the conditional density is

$$f_{X_k|X_1,\cdots,X_{k-1}}(x_k|x_1, \cdots, x_{k-1}) = \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 x_{k-1}^2)}} \exp\left\{ -\frac{x_k^2}{2(\alpha_0 + \alpha_1 x_{k-1}^2)} \right\}.$$

The marginal density of X_1 is again dropped for simplicity, as for the AR(1) model and the resulting likelihood function becomes

$$L(\alpha_0, \alpha_1) = \prod_{j=2}^{T} \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 x_{j-1}^2)}} \exp\left\{ -\frac{x_j^2}{2(\alpha_0 + \alpha_1 x_{j-1}^2)} \right\}.$$

The log likelihood function, neglecting the constant term, is

$$l(\alpha_0, \alpha_1) = -\frac{1}{2} \sum_{j=2}^{T} \left\{ \log\left(\alpha_0 + \alpha_1 x_{j-1}^2\right) + \frac{x_j^2}{\alpha_0 + \alpha_1 x_{j-1}^2} \right\}.$$

We can find the estimates $\hat{\alpha}_0$ and $\hat{\alpha}_1$ by solving the derivatives of the log likelihood function $\dfrac{\partial l}{\partial \alpha_0} = 0$ and $\dfrac{\partial l}{\partial \alpha_1} = 0$ respectively.

2.3.2 Generalized Autoregressive Conditional Heteroscedastic (GARCH) Models

The process $\{X_t\}$ is a Generalized Autoregressive Conditional Heteroscedastic model of order p and q, GARCH(p,q) (Bollerslev, 1986) if:

$$\begin{aligned} X_t|\mathcal{F}_{t-1} &\sim N(0,h_t), \qquad \text{with} \\ h_t &= \alpha_0 + \alpha_1 X_{t-1}^2 + \cdots + \alpha_q X_{t-q}^2 + \beta_1 h_{t-1} + \cdots + \beta_p h_{t-p} \\ &= \alpha_0 + \sum_{i=1}^{q} \alpha_i X_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \end{aligned} \tag{2.9}$$

where $q > 0$, $p \geq 0$, $\alpha_0 > 0$ and $\alpha_i \geq 0$ for $i = 1, \cdots, q$, $\beta_j \geq 0$ for $j = 1, \cdots, p$. Again, the conditions $\alpha_0 > 0$, $\alpha_i \geq 0$ and $\beta_j \geq 0$ are needed to guarantee that the conditional variance $h_t > 0$.

As for ARMA(p,q) models, the likelihood function for the GARCH(p,q) models is difficult to write out. Therefore, we will postpone the discussion of the maximum likelihood estimation until Section 3.4.1 for the special case of the GARCH$(1,1)$ model.

2.3.3 The ARCH(q) and the GARCH$(1,1)$ Models

The simplest and often most useful GARCH process is the GARCH$(1,1)$ process given by:

$$\begin{aligned} X_t|\mathcal{F}_{t-1} &\sim N(0,h_t), \qquad \text{with} \\ h_t &= \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 h_{t-1} \end{aligned} \tag{2.10}$$

where $\alpha_0 > 0$, $\alpha_1 \geq 0$ and $\beta_1 \geq 0$.

It is often found that when fitting ARCH models to financial data a high order is required to get a satisfactory fit (Bollerslev, 1986). We can see that this is expected for data which is really from a GARCH$(1,1)$ process by substituting h_{t-1} into the formula (2.10) recursively. This gives

$$\begin{aligned}
h_t &= \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 (\alpha_0 + \alpha_1 X_{t-2}^2 + \beta_1 h_{t-2}) \\
&= \alpha_0(1+\beta_1) + \alpha_1 X_{t-1}^2 + \alpha_1 \beta_1 X_{t-2}^2 + \beta_1^2 h_{t-2} \\
&= \alpha_0(1+\beta_1) + \alpha_1 X_{t-1}^2 + \alpha_1 \beta_1 X_{t-2}^2 + \beta_1^2 (\alpha_0 + \alpha_1 X_{t-3}^2 + \beta_1 h_{t-3}) \\
&= \alpha_0(1+\beta_1+\beta_1^2) + \alpha_1 X_{t-1}^2 + \alpha_1 \beta_1 X_{t-2}^2 + \alpha_1 \beta_1^2 X_{t-3}^2 + \beta_1^3 h_{t-3} \\
&\quad \vdots \\
&= \alpha_0 \sum_{j=1}^{k} \beta_1^{j-1} + \alpha_1 \sum_{j=1}^{k} \beta_1^{j-1} X_{t-j}^2 + \beta_1^k h_{t-k}.
\end{aligned}$$

We will see in Section 3.1 that in order to have a finite variance of X_t, the condition $\alpha_1 + \beta_1 < 1$ is needed. This means β_1 is strictly less than one. Thus, if $k \to \infty$, h_t becomes

$$h_t = \frac{\alpha_0}{1-\beta_1} + \alpha_1 \sum_{j=1}^{\infty} \beta_1^{j-1} X_{t-j}^2.$$

which corresponds to an ARCH(∞) model $h_t = \alpha_0^* + \sum_{j=1}^{\infty} \alpha_j^* X_{t-j}^2$ with $\alpha_0^* = \alpha_0/(1-\beta_1)$ and $\alpha_j^* = \alpha_1 \beta_1^{j-1}$ for $j = 1, \cdots, \infty$.

This result suggests that a GARCH$(1,1)$ model might replace a high order ARCH(q), giving a more parsimonious model. This is similar to the case in Section 2.2.3 when the ARMA$(1,1)$ model is written as an AR(∞) representation.

Chapter 3

The GARCH$(1,1)$ Model

In this chapter, we will consider the GARCH$(1,1)$ models. We first consider the existence of the GARCH$(1,1)$ process in Section 3.1. Moments of the observations and the conditional variance will be examined in Section 3.2 and the condition of stationarity for the GARCH$(1,1)$ models is studied in Section 3.3. In the subsequent sections, we look at the characteristics of the maximum likelihood estimates of the parameters from GARCH$(1,1)$ models.

3.1 Existence of the GARCH$(1,1)$ Process

The GARCH$(1,1)$ model, first mentioned in Section 2.3.3, is as follows:

$$\begin{aligned} X_t|\mathcal{F}_{t-1} &\sim N(0,h_t), \qquad \text{with} && (3.1)\\ h_t &= \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 h_{t-1} && (3.2) \end{aligned}$$

where $\alpha_0 > 0$, $\alpha_1 \geq 0$ and $\beta_1 \geq 0$.

As with the definition of AR(p) processes in Section 2.2.1, GARCH(p,q) processes are defined recursively and conditions are needed to guarantee the existence of stationary solutions. Here we derive such conditions for the GARCH(1, 1) process. Dividing by the square root of the conditional variance of X_t from (3.1), we obtain:

$$\frac{X_t}{\sqrt{h_t}}|\mathcal{F}_{t-1} \sim N(0,1)$$

and therefore the sequence $Z_1, \cdots, Z_T$ defined by $Z_t = X_t/\sqrt{h_t}$ should be independent and identically distributed (iid) $N(0,1)$. We can then construct a stationary solution of (3.2) starting from a sequence of iid $N(0,1)$ random variables $\{Z_t\}$.

Assuming that the process begins infinitely far in the past, h_t can be expressed as:

$$\begin{aligned}
h_t &= \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 h_{t-1} \\
&= \alpha_0 + (\alpha_1 Z_{t-1}^2 + \beta_1) h_{t-1} \\
&= \alpha_0 + (\alpha_1 Z_{t-1}^2 + \beta_1) \{\alpha_0 + (\alpha_1 Z_{t-2}^2 + \beta_1) h_{t-2}\} \\
&\quad \vdots \\
&= \alpha_0 + \alpha_0 \sum_{k=1}^{\infty} \left\{ \prod_{j=1}^{k} (\alpha_1 Z_{t-j}^2 + \beta_1) \right\}. \qquad (3.3)
\end{aligned}$$

Theorem 3.1. *If the expectation of an infinite sum of non-negative random variables is finite, then the sum converges almost surely.*

(See Lukacs, 1975, Theorem 4.2.1, p. 80.) We can use this theorem to find a condition under which the expression in (3.3) exists. Taking the unconditional expectation of both sides, we get

$$\begin{aligned}
E(h_t) &= \alpha_0 + \alpha_0 \sum_{k=1}^{\infty} \left\{ \prod_{j=1}^{k} (\alpha_1 E(Z_{t-j}^2) + \beta_1) \right\} \\
&= \alpha_0 + \alpha_0 \sum_{k=1}^{\infty} (\alpha_1 + \beta_1)^k \\
&= \frac{\alpha_0}{1 - \alpha_1 - \beta_1}.
\end{aligned}$$

Thus, the unconditional expected value of h_t is finite and the infinite series for h_t in (3.3) converges to $\alpha_0/(1-\alpha_1-\beta_1)$ provided that $\alpha_1+\beta_1<1$.

In summary, if $\alpha_1+\beta_1<1$ and $\alpha_1\geq 0, \beta_1\geq 0$, we can define h_t by (3.3) and $X_t = Z_t\sqrt{h_t}$. The resulting process $\{X_t\}$ is a stationary solution of (3.2).

3.2 Moments of X_t and h_t

After showing the existence of the GARCH(1, 1) process, we now examine the higher moments of h_t and then X_t. We have already seen

$$E(h_t) = \frac{\alpha_0}{1-\alpha_1-\beta_1}.$$

Squaring the equation (3.2), we get

$$\begin{aligned} h_t^2 &= \alpha_0^2 + \alpha_1^2 X_{t-1}^4 + \beta_1^2 h_{t-1}^2 + 2\alpha_1\beta_1 X_{t-1}^2 h_{t-1} + \\ &\quad 2\alpha_0\alpha_1 X_{t-1}^2 + 2\alpha_0\beta_1 h_{t-1}. \end{aligned}$$

Notice that Z_t and h_t are independent. Replacing X_t by $Z_t\sqrt{h_t}$ and taking the expectation, we see

$$\begin{aligned} E(h_t^2) &= \alpha_0^2 + \alpha_1^2 E(Z_{t-1}^4 h_{t-1}^2) + \beta_1^2 E(h_{t-1}^2) + 2\alpha_1\beta_1 E(Z_{t-1}^2 h_{t-1}^2) + \\ &\quad 2\alpha_0\alpha_1 E(Z_{t-1}^2 h_{t-1}) + 2\alpha_0\beta_1 E(h_{t-1}) \\ &= \alpha_0^2 + \alpha_1^2 E(Z_{t-1}^4)E(h_{t-1}^2) + \beta_1^2 E(h_{t-1}^2) + 2\alpha_1\beta_1 E(Z_{t-1}^2)E(h_{t-1}^2) + \\ &\quad 2\alpha_0\alpha_1 E(Z_{t-1}^2)E(h_{t-1}) + 2\alpha_0\beta_1 E(h_{t-1}) \\ &= \alpha_0^2 + (3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2) E(h_{t-1}^2) + 2\alpha_0(\alpha_1+\beta_1) E(h_{t-1}). \end{aligned} \tag{3.4}$$

Since the process is stationary, $E(h_t^2) = E(h_{t-1}^2)$. So

$$\begin{aligned} E(h_t^2) &= \frac{\alpha_0^2 + 2\alpha_0(\alpha_1+\beta_1) E(h_{t-1})}{1-3\alpha_1^2-2\alpha_1\beta_1-\beta_1^2} \\ &= \frac{\alpha_0^2 + 2\alpha_0^2(\alpha_1+\beta_1)/(1-\alpha_1-\beta_1)}{1-3\alpha_1^2-2\alpha_1\beta_1-\beta_1^2} \\ &= \frac{\alpha_0^2\,[1+(\alpha_1+\beta_1)]}{(1-\alpha_1-\beta_1)(1-3\alpha_1^2-2\alpha_1\beta_1-\beta_1^2)}. \end{aligned}$$

Therefore, the unconditional second moment of h_t is finite if $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$. If this condition is false, then there is no positive value for $E(h_t^2) = E(h_{t-1}^2)$ which satisfies the equation (3.4).

We now look at the moments of X_t. The first and the third moments of X_t are both zero:

$$\begin{aligned} E(X_t) &= E(Z_t\sqrt{h_t}) = E(Z_t)E(\sqrt{h_t}) = 0, \\ E(X_t^3) &= E(Z_t^3 h_t^{3/2}) = E(Z_t^3)E(h_t^{3/2}) = 0. \end{aligned}$$

The second and the fourth moments can be found by:

$$E(X_t^2) = E[E(X_t^2|\mathcal{F}_{t-1})] = E(h_t) = \frac{\alpha_0}{1-\alpha_1-\beta_1}$$

and

$$\begin{aligned} E(X_t^4) &= E(Z_t^4 h_t^2) = E(Z_t^4)E(h_t^2) = 3E(h_t^2) \\ &= 3\frac{\alpha_0^2[1+(\alpha_1+\beta_1)]}{(1-\alpha_1-\beta_1)(1-3\alpha_1^2-2\alpha_1\beta_1-\beta_1^2)} \\ &= 3\frac{\alpha_0^2[1+(\alpha_1+\beta_1)]}{(1-\alpha_1-\beta_1)(1-3\alpha_1^2-2\alpha_1\beta_1-\beta_1^2)} \times \left[\frac{1-(\alpha_1+\beta_1)}{1-(\alpha_1+\beta_1)}\right] \\ &= 3\left[\frac{1-(\alpha_1+\beta_1)^2}{1-3\alpha_1^2-2\alpha_1\beta_1-\beta_1^2}\right] \times \left[\frac{\alpha_0}{1-\alpha_1-\beta_1}\right]^2. \end{aligned} \tag{3.5}$$

Recall that the kurtosis, $K(\cdot)$, of a random variable Y with mean zero is defined as:

$$K(Y) = \frac{E(Y^4)}{[E(Y^2)]^2}$$

so that

$$E(Y^4) = K(Y)\,[E(Y^2)]^2.$$

The second term on the right hand side of (3.5) is the square of the unconditional variance of X_t, $E(X_t^2)$, thus the kurtosis of X_t is:

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

which is strictly greater than 3 unless $\alpha_1 = 0$.

The kurtosis for a standard normal random variable Z is 3. Thus, the kurtosis of X_t is greater than the kurtosis of a normal random variable, and the distribution of X_t has a heavier tail than the normal distribution, when $\alpha_1 > 0$. Some plots will be shown in Chapter 4.

3.3 Stationarity of the GARCH$(1,1)$ Process

It is interesting to examine in detail the conditions under which the second and fourth moments of the X_t are stationary, namely,

$$\begin{aligned} \alpha_1 + \beta_1 &< 1 \\ \text{and} \qquad 3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 &< 1, \end{aligned}$$

by looking at these regions on a graph. Let x be α_1 and y be β_1 to simplify the notation. Then, $x + y = 1$ is a straight line whereas

$$\begin{aligned} 3x^2 + 2xy + y^2 &= 1 \\ \Leftrightarrow \quad (x+y)^2 &= 1 - 2x^2 \end{aligned} \tag{3.6}$$

is an ellipse. However, it cannot be written as the general ellipse formula $x^2/a^2 + y^2/b^2 = 1$ on x- and y-axes. But it can be rotated geometrically and be expressed in the general ellipse formula on the rotated axes.

Let

$$R(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

be the rotation matrix which rotates the x- and y-axes counterclockwise into u- and v-axes by an angle θ. Then, using $R^{-1}(\theta)$, we have

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$

This gives

$$x = u\cos\theta - v\sin\theta$$

and

$$y = u\sin\theta + v\cos\theta.$$

Thus, we get

$$x + y = u(\sin\theta + \cos\theta) + v(\cos\theta - \sin\theta).$$

The left hand side of (3.6) is

$$(x+y)^2 = u^2(1+\sin 2\theta) + v^2(1-\sin 2\theta) + 2uv\cos 2\theta$$

whereas the right hand side of (3.6) is

$$1 - 2x^2 = 1 - 2u^2\cos^2\theta - 2u^2\cos^2\theta + 2uv\sin 2\theta.$$

Substituting in equation (3.6), we obtain

$$u^2(1+\sin^2\theta + 2\cos^2\theta) + v^2(1-\sin^2\theta+2\sin^2\theta) = 1 + 2uv(\sin 2\theta - \cos 2\theta).$$

If $\sin 2\theta = \cos 2\theta$, then we can eliminate the uv term and get a general ellipse formula on the u- and v-axes. Therefore, $\tan 2\theta = 1$ and $\theta = \pi/8$ or 22.5°. This gives $\sin 2\theta = \cos 2\theta = \sqrt{2}/2$. Recall that

$$\sin^2\theta = (1-\cos 2\theta)/2 \quad \text{and} \quad \cos^2\theta = (1+\cos 2\theta)/2.$$

We can then write (3.6) in the general ellipse formula in terms of u and v only:

$$\begin{aligned} u^2(1+\sqrt{2}/2+1+\sqrt{2}/2) &+ v^2(1-\sqrt{2}/2+1-\sqrt{2}/2) = 1 \\ u^2(2+\sqrt{2}) &+ v^2(2-\sqrt{2}) = 1 \\ \frac{u^2}{(1-\sqrt{2}/2)} &+ \frac{v^2}{(1+\sqrt{2}/2)} = 1 \end{aligned}$$

which gives the minor axis $a = (1-\sqrt{2}/2)^{1/2} = 0.54$ and major axis $b = (1+\sqrt{2}/2)^{1/2} = 1.31$. The corresponding graph is shown in Figure 3.1.

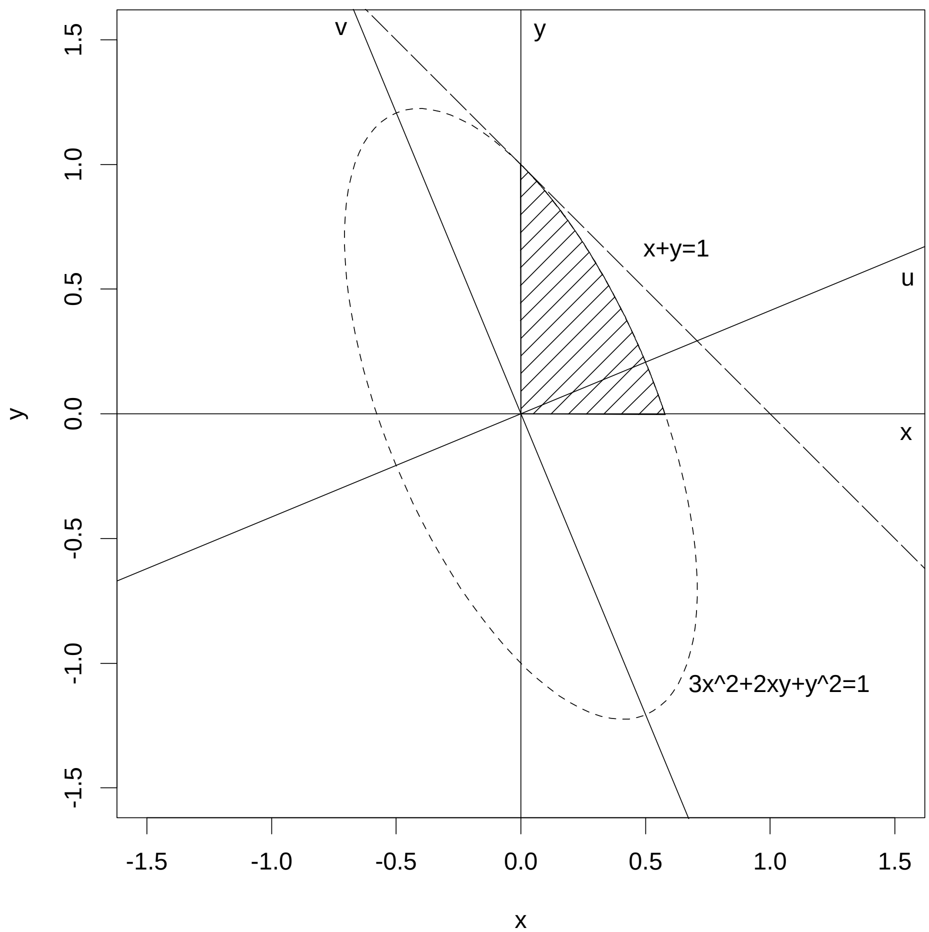

Figure 3.1: Regions giving finite second and fourth moments of X_t. The process is defined only for (x, y) in the first quadrant. The fourth moment of X_t is finite only inside the shaded region of the ellipse. A stationary solution of equation (3.2) exists only within the triangle bounded by the x-axis, the y-axis and the dashed line $x + y = 1$. Note that x and y are used instead of α_1 and β_1 respectively in this plot.

3.4 Data Simulation

From Section 3.1, X_t is defined to be a function of h_t and iid standard normal random variables Z_t; and the h_t defined in (3.2) can be expressed in terms of previous h_t and Z_t. Therefore, both X_t and h_t can be simulated recursively from the sequence of Z_t.

In the model, X_t are allowed to go infinitely into the past, but in reality start at $t = 1$. Therefore, we will use the expected values $E(X_t^2) = E(h_t) = \alpha_0/(1 - \alpha_1 - \beta_1)$ to substitute for past X_t^2 and h_t when needed. The expected values of the volatility in a stationary process, σ^2, is also called the long-term volatility of the process.

To simulate data from a GARCH(1, 1) process, we follow the steps:

1. Generate a sequence of iid N(0, 1) random variables, $\{Z_t\}$ for $t = 1, \cdots, T^*$.
2. Set $h_1 = \alpha_0/(1 - \alpha_1 - \beta_1)$ and $X_1 = Z_1 \sqrt{h_1}$.
3. For $t = 2, \cdots, T^*$ do the following recursively:

$$\begin{aligned} h_t &= \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 h_{t-1} \\ X_t &= Z_t \sqrt{h_t}. \end{aligned}$$

The sequence X_t is not exactly stationary but approaches stationarity as T^* becomes larger; therefore we take the last T values and regard them as stationary.

3.4.1 The Likelihood Function and Estimation of Parameters

For the GARCH(1, 1) model defined from (3.2), the joint density of the observations $X_1, \cdots, X_T$ can be written as the product of the conditional densities, conditioning on the previous observations:

$$f_{X_1,\cdots,X_T}(x_1,\cdots,x_T) = \left\{\prod_{j=2}^{T} f_{X_j|X_1,\cdots,X_{j-1}}(x_j|x_1,\cdots,x_{j-1})\right\} \times f_{X_1}(x_1);$$

for simplicity, the marginal density of X_1 will be dropped, as for the ARMA$(1,1)$ model. For $k = 2, \cdots, T$, the conditional density of X_k, conditioning on $X_1, \cdots, X_{k-1}$, is

$$f_{X_k|X_1,\cdots,X_{k-1}}(x_k|x_1,\cdots,x_{k-1}) = \frac{1}{\sqrt{2\pi}\sqrt{h_k}} \exp\left\{-\frac{x_k^2}{2h_k}\right\},$$

and the conditional likelihood function, given X_1 and h_1, is:

$$\begin{aligned} L(\alpha_0,\alpha_1,\beta_1) &= f_{X_2,\cdots,X_T|X_1,h_1}(x_2,\cdots,x_T|x_1,h_1) \\ &= \prod_{j=2}^{T} \frac{1}{\sqrt{2\pi}\sqrt{h_j^*}} \exp\left\{-\frac{x_j^2}{2h_j^*}\right\}, \end{aligned}$$

where $h_j^* = \alpha_0 + \alpha_1 X_{j-1}^2 + \beta_1 h_{j-1}^*$ are obtained recursively. We substitute h_1 by its expected value $E(h_1) = \alpha_0/(1-\alpha_1-\beta_1)$.

Taking the logarithm and neglecting the constant term, we find that the log likelihood function is:

$$l(\alpha_0,\alpha_1,\beta_1|\boldsymbol{X},\boldsymbol{h}) = -\frac{1}{2}\left\{\sum_{j=2}^{T} \log h_j^* + x_j^2/h_j^*\right\}$$

where $\boldsymbol{X} = (X_1,\cdots,X_T)'$ and $\boldsymbol{h} = (h_1,\cdots,h_T)'$.

The function *nlminb* from S-Plus is described as a local minimizer for smooth nonlinear functions subject to bound-constrained parameters. Since we want to maximize the log likelihood function, we will use the function *nlminb* to minimize its negative value. This function can have restrictions on the parameter values using the options *lower* and *upper*. However, we will not use these options in this section as we just want to see the general behaviour of the GARCH$(1,1)$ model and the function *nlminb*. In Section 3.6.2, we will make use of these options and compare the parameters estimated with and without using this constraint.

3.5 A typical GARCH(1, 1) Example

We now show how to generate data from the stationary GARCH(1, 1) model with initial values of parameters $\alpha_0 = 1, \alpha_1 = 0.2$ and $\beta_1 = 0.2$ and how to obtain the estimates using maximum likelihood.

3.5.1 Results from the Monte Carlo Simulations

Four typical plots of $T = 500$ observations X_t and the corresponding h_t from 100 Monte Carlo samples are shown in Figure 3.2. Note that large variation of X_t associates with large variation of h_t. This is expected as X_t are defined recursively from h_t.

A sample of the maximum likelihood estimates for the first 100 simulated samples and the estimated averages and standard errors are shown in Table 3.1, together with the values of $\hat{\alpha}_0/\sigma^2$, $\hat{\alpha}_0 + \hat{\beta}_1$, $\hat{\alpha}_0/\sigma^2 + \hat{\beta}_1$ and $\hat{\alpha}_0/\sigma^2 + \hat{\alpha}_1 + \hat{\beta}_1$ where $\sigma^2 = \alpha_0/(1 - \alpha_1 - \beta_1)$ is computed from their true values.

m	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\alpha}_0/\sigma^2$	$\hat{\alpha}_0 + \hat{\beta}_1$	$\hat{\alpha}_0/\sigma^2 + \hat{\beta}_1$	$\hat{\alpha}_0/\sigma^2 + \hat{\alpha}_1 + \hat{\beta}_1$
True Values	1.000	0.2000	0.200	0.600	1.200	0.8000	1.0000
1	0.907	0.2213	0.148	0.544	1.055	0.6922	0.9135
2	1.196	0.2133	-0.126	0.718	1.070	0.5916	0.8048
3	0.878	0.2156	0.288	0.527	1.166	0.8147	1.0303
4	1.179	0.1669	0.223	0.708	1.402	0.9303	1.0971
⋮				⋮			
98	1.211	0.2420	-0.012	0.727	1.199	0.7149	0.9569
99	1.073	0.2607	0.120	0.644	1.194	0.7643	1.0250
100	1.198	0.2229	0.085	0.719	1.283	0.8038	1.0266
Avg	1.041	0.1948	0.179	0.625	1.220	0.8034	0.9982
SE	0.039	0.0063	0.024	0.023	0.017	0.0068	0.0061

Table 3.1: Estimates of the parameters for GARCH(1, 1) model with values of parameters $\alpha_0 = 1, \alpha_1 = 0.2$ and $\beta_1 = 0.2$ from the $m = 100$ samples of Monte Carlo studies (incomplete table). Each sample has 500 observations. The estimated averages and estimated standard errors are shown in the last two rows.

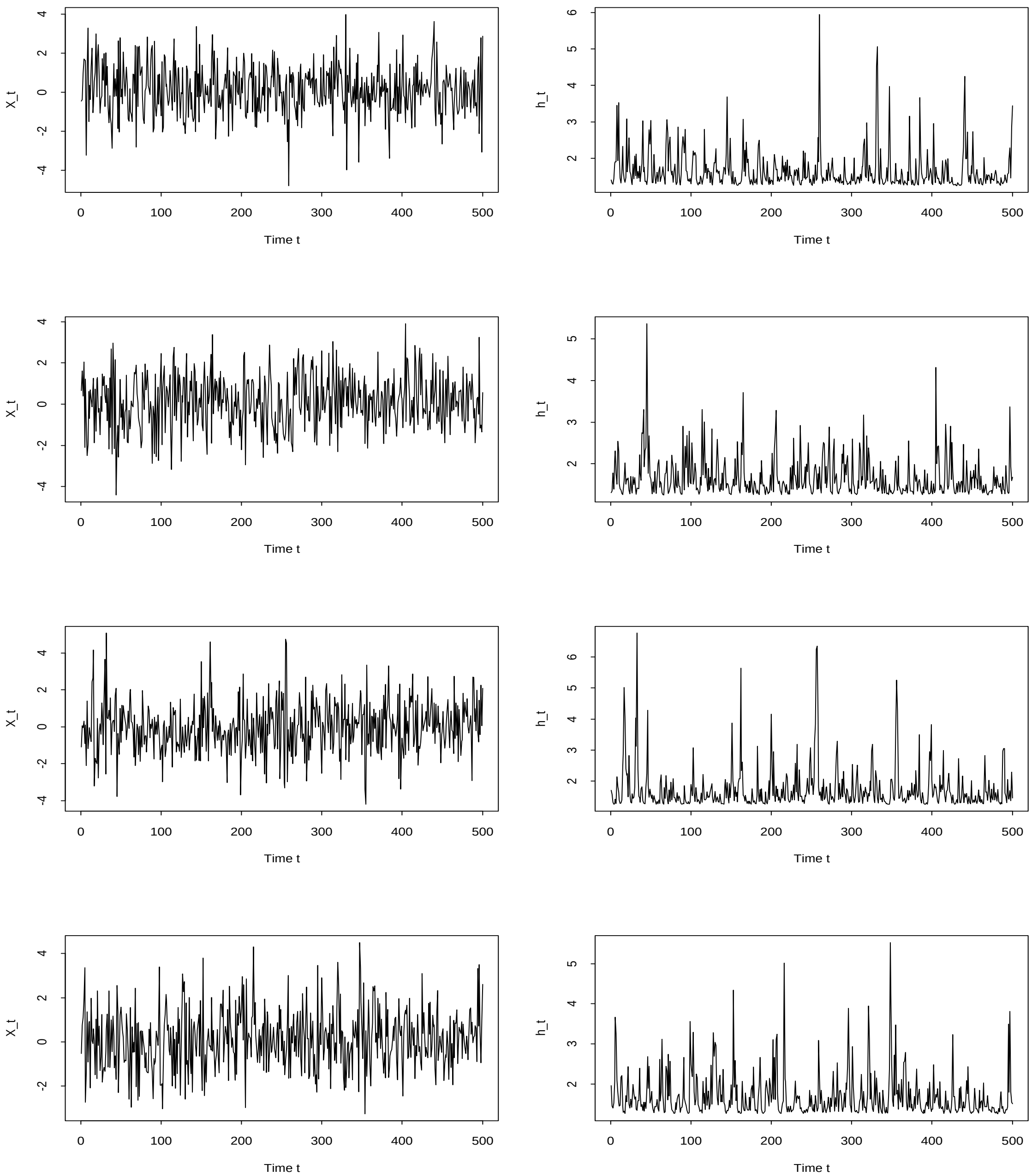

Figure 3.2: Four time series plots of typical observations X_t and the corresponding h_t from the stationary GARCH$(1, 1)$ model with the values of parameters $\alpha_0 = 1, \alpha_1 = 0.2$ and $\beta_1 = 0.2$.

Note that there are two negative estimates of $\hat{\beta}_1$ from samples 2 and 98. As mentioned above, this is possible because no restriction has been imposed on the range of the estimates during the "maximization". We will see how often this happens in Section 3.6.2.

The averages of the three estimates for GARCH$(1,1)$ with initial values of parameters $\alpha_0 = 1, \alpha_1 = 0.2$ and $\beta_1 = 0.2$ from the 100 Monte Carlo samples are shown in Table 3.2 together with their estimated standard errors and mean square errors (MSE). For large T, maximum likelihood estimates are normally distributed to high approximation. Using this, we see that all three estimates are not statistically different from their true values at the 5% significance level.

Estimates	True Values	Estimated Expected Values	Estimated Standard Errors	Estimated MSE
$\hat{\alpha}_0$	1.000	1.041	0.039	0.151
$\hat{\alpha}_1$	0.200	0.195	0.006	0.004
$\hat{\beta}_1$	0.200	0.179	0.024	0.059

Table 3.2: Averages, standard errors and mean squared errors (MSE) of the estimates for GARCH$(1,1)$ model with values of parameters $\alpha_0 = 1, \alpha_1 = 0.2$ and $\beta_1 = 0.2$.

3.5.2 Identifiability of parameters

The estimated variance-covariance matrix and the correlation matrix of the three parameter estimates and four sums are shown in Table 3.3. Note that the variance of $\hat{\alpha}_0 + \hat{\beta}_1$ is significantly smaller than the individual variances of $\hat{\alpha}_0$ and $\hat{\beta}_1$: 0.028 as opposed to 0.149 and 0.058. Recall that for any two arbitrary random variables U and V, the variance of their sum is:

$$Var(U + V) = Var(U) + Var(V) + 2Cov(U, V).$$

This suggests a negative association between the two estimates $\hat{\alpha}_0$ and $\hat{\beta}_1$ which can be seen from the value of the Pearson correlation coefficient, -0.96, in the correlation matrix, Table 3.3. This fact is further illustrated by the scatter plot between $\hat{\alpha}_0$ and

Variance-Covariance Matrix

$\hat{\alpha}_0$	**0.149**	0.005	-0.089	0.089	0.060	-0.000	0.004
$\hat{\alpha}_1$	0.005	0.004	-0.005	0.003	-0.001	-0.002	0.002
$\hat{\beta}_1$	-0.089	-0.005	**0.058**	-0.054	-0.031	0.005	-0.001
$\hat{\alpha}_0/\sigma^2$	0.089	0.003	-0.054	0.054	0.036	0.000	0.003
$\hat{\alpha}_0 + \hat{\beta}_1$	0.060	-0.001	-0.031	0.036	**0.028**	0.005	0.004
$\hat{\alpha}_0/\sigma^2 + \hat{\beta}_1$	-0.000	-0.002	0.005	0.000	0.005	0.005	0.002
$\hat{\alpha}_0/\sigma^2 + \hat{\alpha}_1 + \hat{\beta}_1$	0.004	0.002	-0.001	0.003	0.004	0.002	0.004

Correlation Matrix

$\hat{\alpha}_0$	1.00	0.18	**-0.96**	1.00	0.91	-0.00	0.19
$\hat{\alpha}_1$	0.18	1.00	-0.34	0.18	-0.07	-0.58	0.40
$\hat{\beta}_1$	**-0.96**	-0.34	1.00	-0.96	-0.76	0.28	-0.04
$\hat{\alpha}_0/\sigma^2$	1.00	0.18	-0.96	1.00	0.91	-0.00	0.19
$\hat{\alpha}_0 + \hat{\beta}_1$	0.91	-0.07	-0.76	0.91	1.00	0.40	0.39
$\hat{\alpha}_0/\sigma^2 + \hat{\beta}_1$	-0.00	-0.58	0.28	-0.00	0.40	1.00	0.52
$\hat{\alpha}_0/\sigma^2 + \hat{\alpha}_1 + \hat{\beta}_1$	0.19	0.40	-0.04	0.19	0.39	0.52	1.00

Table 3.3: Variance-covariance matrix and the correlation matrix for estimates from 100 Monte Carlo samples of size 500 from the GARCH$(1,1)$ model with parameters $\alpha_0 = 1, \alpha_1 = 0.2$ and $\beta_1 = 0.2$. Note that $\sigma^2 = \alpha_0/(1-\alpha_1-\beta_1)$.

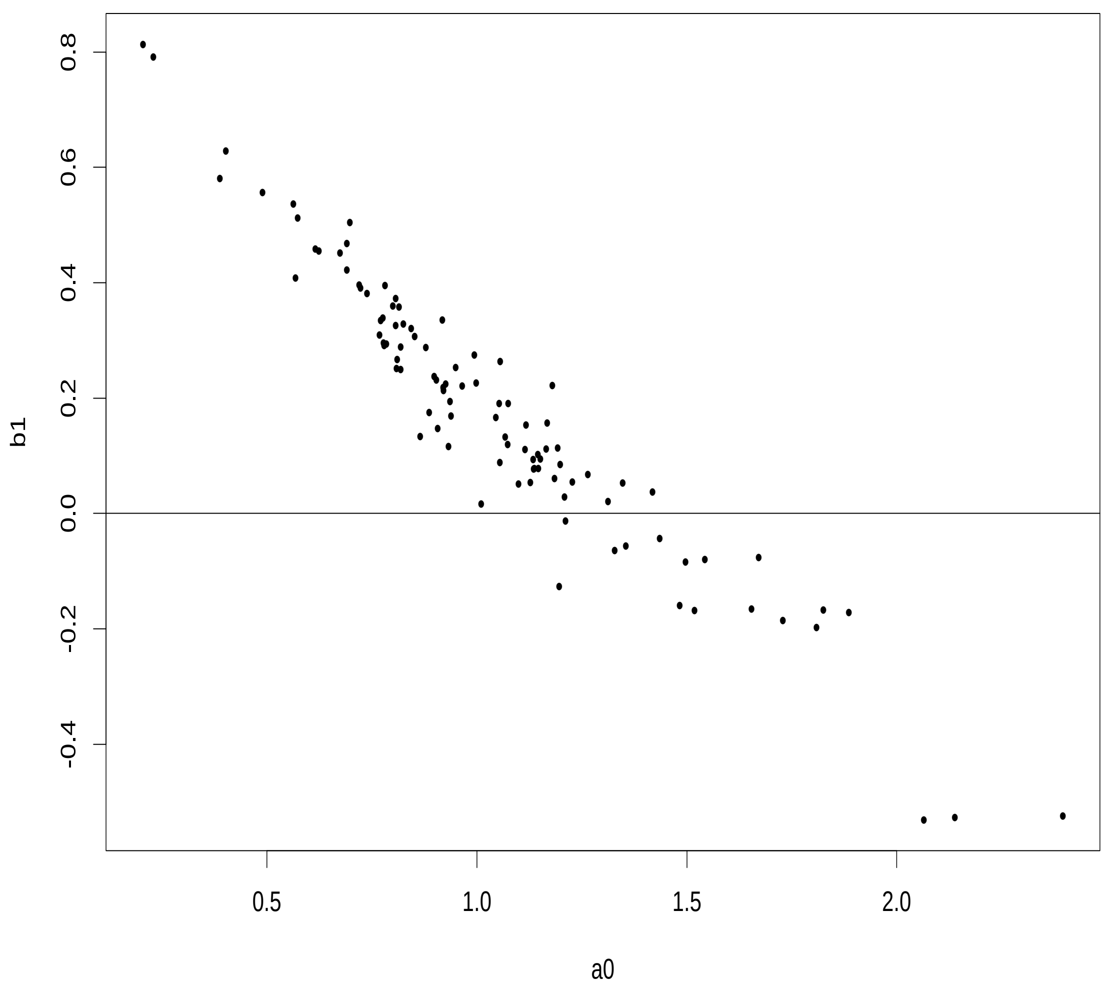

Figure 3.3: Scatter plot between $\hat{\alpha}_0$ and $\hat{\beta}_1$ from GARCH$(1, 1)$ model with values of parameters $\alpha_0 = 1, \alpha_1 = 0.2$ and $\beta_1 = 0.2$. Note that a_0 and b_1 are used in the plot instead of $\hat{\alpha}_0$ and $\hat{\beta}_1$ respectively.

$\hat{\beta}_1$ shown in Figure 3.3 where the points are close to a straight line with negative slope.

Notice also that the variance $V(\hat{\alpha}_0/\sigma^2+\hat{\beta}_1)$ is much less than the respective variances $V(\hat{\alpha}_0/\sigma^2)$ and $V(\hat{\beta}_1)$. This can be explained by the fact that these two parameters are approximately non-identifiable. Recall from (3.2) that h_t is expressed as

$$h_t = \alpha_0 + \alpha_1\, X_{t-1}^2 + \beta_1\, h_{t-1}.$$

Multiplying α_0 by h_{t-1} and dividing it by $E(h_{t-1})$ gives approximately the same value of α_0. Thus, h_t can be written as:

$$\begin{aligned} h_t &\approx \alpha_0\, \frac{h_{t-1}}{E(h_{t-1})} + \alpha_1\, X_{t-1}^2 + \beta_1\, h_{t-1} \\ &= \frac{\alpha_0}{\sigma^2}\, h_{t-1} + \alpha_1\, X_{t-1}^2 + \beta_1\, h_{t-1} \\ &= \left(\frac{\alpha_0}{\sigma^2} + \beta_1\right)\, h_{t-1} + \alpha_1\, X_{t-1}^2. \end{aligned}$$

The sum $\alpha_0/\sigma^2+\beta_1$ is a constant and can be considered as a single parameter. Therefore, the parameters α_0 and β_1 are approximately non-identifiable. (If $\alpha_1 = 0$, then α_0 and β_1 are exactly non-identifiable and $\alpha_0/\sigma^2 + \beta = 1$.)

Another interesting fact from Table 3.1 is that $\hat{\alpha}_0/\sigma^2 + \hat{\alpha}_1 + \hat{\beta}_1$ is very close to one (0.998) and the variance of this sum is very small (0.004) compared with the variances of the individual terms. Hull (1999) mentioned that the long-term volatility σ^2 could be incorporated directly into the constant term α_0 of the GARCH$(1,1)$ model. Suppose $\alpha_0 = \gamma\,\sigma^2$, then

$$h_t = \gamma\,\sigma^2 + \alpha_1\, X_{t-1}^2 + \beta_1\, h_{t-1};$$

taking expectation, we have that

$$\sigma^2 = E(h_t) = \gamma\,\sigma^2 + \alpha_1\,\sigma^2 + \beta_1\,\sigma^2$$

and therefore $\gamma + \alpha_1 + \beta_1 = 1$. Thus

$$\frac{\alpha_0}{\sigma^2} + \alpha_1 + \beta_1 = 1$$

which explains why the variance of $\dfrac{\hat{\alpha}_0}{\sigma^2} + \hat{\alpha}_1 + \hat{\beta}_1$ is so small.

3.6 Results from Further Simulations

3.6.1 Characteristics and Behaviour of the Estimates

We have discussed the GARCH(1, 1) model with the true values of parameters $\alpha_0 = 1, \alpha_1 = 0.2$ and $\beta_1 = 0.2$. Further Monte Carlo studies have been done with various combinations of the parameters α_1 and β_1. Table 3.4 gives the averages and

True Values		Estimated Averages			Estimated Root Mean Square		
α_1	β_1	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$
0.00	0.00	1.13	0.01	-0.14	0.53	0.05	0.55
0.00	0.20	1.22	0.00	0.02	0.81	0.05	0.66
0.00	0.30	1.10	0.01	0.24	0.79	0.05	0.54
0.00	0.50	1.42	-0.00	0.30	1.16	0.05	0.55
0.00	0.70	1.26	-0.00	0.62	1.37	0.05	0.42
0.10	0.00	1.06	0.10	-0.07	0.44	0.06	0.40
0.10	0.20	1.25	0.10	0.03	0.65	0.06	0.45
0.10	0.30	1.22	0.09	0.17	0.72	0.07	0.45
0.10	0.50	1.19	0.10	0.42	0.88	0.06	0.37
0.10	0.70	1.14	0.10	0.67	0.74	0.05	0.18
0.20	0.00	1.11	0.19	-0.08	0.38	0.07	0.29
0.20	0.20	1.08	0.19	0.16	0.39	0.07	0.25
0.20	0.30	1.06	0.18	0.27	0.42	0.07	0.25
0.20	0.50	1.10	0.20	0.47	0.46	0.06	0.17
0.20	0.70	1.14	0.21	0.68	0.54	0.05	0.08
0.30	0.00	1.04	0.29	-0.02	0.21	0.08	0.14
0.30	0.20	1.01	0.28	0.21	0.34	0.09	0.21
0.30	0.30	1.00	0.30	0.29	0.29	0.07	0.16
0.30	0.50	1.17	0.29	0.46	0.49	0.06	0.13
0.40	0.00	1.07	0.40	-0.03	0.21	0.08	0.10
0.40	0.20	1.12	0.40	0.15	0.30	0.09	0.12
0.40	0.30	0.97	0.39	0.31	0.23	0.10	0.12
0.40	0.50	1.09	0.40	0.47	0.26	0.07	0.08
0.50	0.00	1.05	0.47	-0.00	0.17	0.08	0.08
0.50	0.20	1.07	0.47	0.20	0.23	0.09	0.10

Table 3.4: Estimated averages and root mean squares of $m = 50$ Monte Carlo samples with different values of parameters. The value of α_0 is one.

the root mean squares of the estimates of the parameters with fixed $\alpha_0 = 1$, and α_1 and β_1 varying. When $\alpha_1 = 0$, we expect α_0 and β_1 to be non-identifiable. This is demonstrated in the table (top 5 rows), where in general there is a marked bias in $\hat{\alpha}_0$ and $\hat{\beta}_1$. The phenomenon occurs again for low α_1, but diminishes as $\alpha_1 \to 0.50$.

3.6.2 Negative estimates

It was seen from Table 3.4 that it is possible to obtain negative values for the estimates $\hat{\alpha}_1$ and $\hat{\beta}_1$, although this is not allowed from the definition (3.2) of the GARCH(1,1) model. Therefore, it seems appropriate to record how often such negative values occur.

Table 3.5 gives the probability of the negative estimates for various combinations of the parameters and for different numbers of the Monte Carlo samples. It is interesting to see that there is one Monte Carlo sample out of 200 for which the estimated mean of $\hat{\alpha}_0$ is less than zero when $\hat{\alpha}_1$ and $\hat{\beta}_1$ are set to 0.00 and 0.20. It is also noted that α_1 is positive in general, except when its true value is set to zero. In addition, it is very likely for β_1 to have negative estimates regardless of the true values used. This can also be seen from the scatter plot in Figure 3.3 where about 20% of the points falls below the line $\hat{\beta}_1 = 0$.

A second set of Monte Carlo samples was analyzed, using restrictions on the estimates at each stage of the iterative procedure, as follows:

$$\hat{\alpha}_0 > 0, \qquad \hat{\alpha}_1 \geq 0, \qquad \hat{\beta}_1 \geq 0 \qquad \text{and} \qquad \hat{\alpha}_1 + \hat{\beta}_1 < 1.$$

In Table 3.6, the two sets of estimates (those with no restrictions and those with restrictions) are compared. The table also shows the differences between the averages of the estimates. In general, restriction does not affect the estimate $\hat{\alpha}_0$ as much as it does $\hat{\alpha}_1$ and $\hat{\beta}_1$.

α_1	β_1	m	$P(\hat{\alpha}_0 < 0)$	$P(\hat{\alpha}_1 < 0)$	$P(\hat{\beta}_1 < 0)$
0.20	0.20	20	0.000	0.000	0.200
		50	0.000	0.000	0.180
		100	0.000	0.000	0.160
		200	0.000	0.000	0.175
0.20	0.00	20	0.000	0.000	0.500
		50	0.000	0.000	0.520
		100	0.000	0.010	0.610
		200	0.000	0.005	0.570
0.00	0.20	20	0.000	0.550	0.500
		50	0.000	0.500	0.360
		100	0.000	0.590	0.470
		200	0.005	0.600	0.300

Table 3.5: Probabilities of different estimates from GARCH$(1, 1)$ models having negative values for various combinations of true parameters α_1 and β_1 and different numbers m of Monte Carlo samples. Note that $\alpha_0 = 1$. Each sample has 500 observations.

α_1	β_1	m	$\hat{\alpha}_0$	$\hat{\alpha}_0^*$	D	$\hat{\alpha}_1$	$\hat{\alpha}_1^*$	D	$\hat{\beta}_1$	$\hat{\beta}_1^*$	D
0.20	0.20	20	0.90	1.09	-0.19	0.18	0.18	-0.00	0.27	0.14	0.13
		50	0.89	1.01	-0.12	0.18	0.20	-0.02	0.27	0.19	0.08
		100	1.00	0.96	0.04	0.20	0.19	0.01	0.20	0.23	-0.03
		200	1.00	1.04	-0.04	0.19	0.19	0.00	0.21	0.19	0.01
0.20	0.00	20	0.99	0.89	0.10	0.19	0.19	0.00	0.02	0.11	-0.09
		50	0.99	0.87	0.12	0.19	0.18	0.01	-0.00	0.10	-0.10
		100	1.03	0.91	0.12	0.19	0.19	0.00	-0.02	0.07	-0.10
		200	1.03	0.92	0.11	0.19	0.18	0.01	-0.02	0.08	-0.10
0.00	0.20	20	1.17	0.86	0.31	-0.01	0.02	-0.03	0.08	0.30	-0.22
		50	1.03	0.89	0.14	-0.00	0.02	-0.02	0.17	0.26	-0.10
		100	1.22	0.91	0.31	-0.01	0.02	-0.03	0.02	0.26	-0.24
		200	0.94	0.96	-0.02	-0.01	0.01	-0.02	0.26	0.22	0.04

Table 3.6: Average estimates of parameters from GARCH$(1, 1)$ models with and without restrictions imposed during the estimation and their differences for various combinations of true parameters α_1 and β_1. Each Monte Carlo sample has 500 observations. D is the difference between the two sets of estimates. The restricted estimates are denoted by "$*$".

Chapter 4

Data Analysis

We first discuss the dataset Standard & Poor's 500 Stock Price Index in Section 4.1. Then we fit the ARCH(1) model to the dataset in Section 4.2. The GARCH$(1,1)$ model is used to obtain a better fit to the dataset and subsequently, the diagnostic checking on the fit is discussed in Section 4.4.

4.1 Data Description

The dataset being considered in this chapter is taken from the Standard & Poor's (S&P) 500 Stock Price Index from Jan 2, 1990 to Dec 29, 2000 which has $T = 2780$ observations. It measures the performance of 500 of the largest companies in the U.S., diversified by different industries.

Let Y_t be the value of the S&P 500 Stock Price Index at time t for $t = 0, \cdots, T-1$. Figure 4.1 shows the time series plot of Y_t during the above period.

Let X_t be the continuously compounded returns for the S&P 500 Stock Price Index, defined by $X_t = \log(Y_{t+1}/Y_t)$ for $t = 0, \cdots, T-2$. This is sometimes referred to as

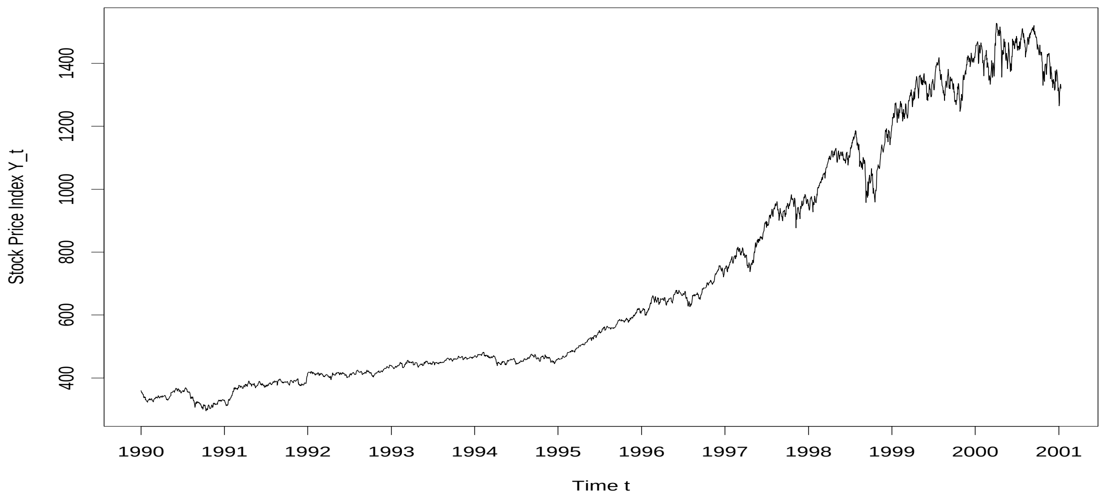

Figure 4.1: Time series plot of the S&P 500 Stock Price Index Y_t from Jan 2, 1990 to Dec 29, 2000.

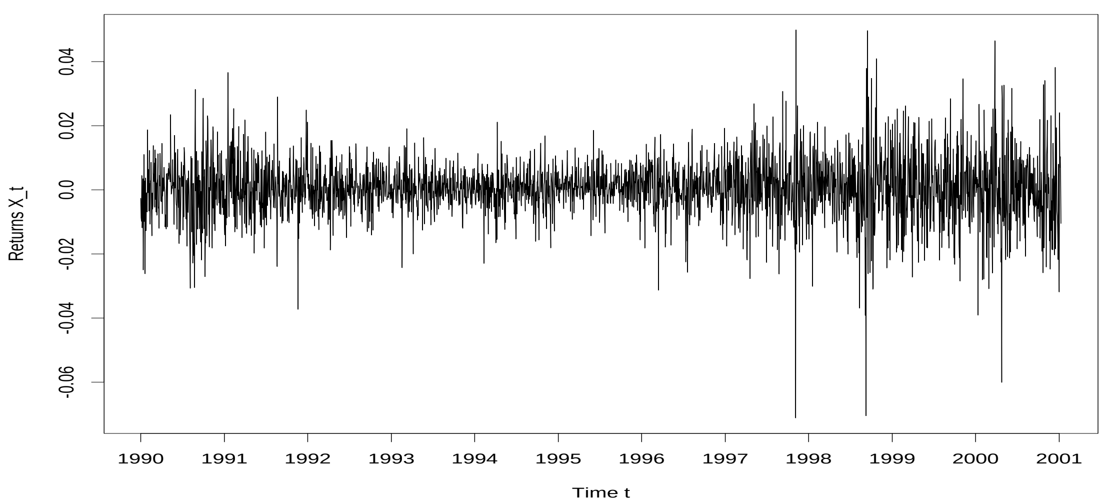

Figure 4.2: Time series plot of the continuously compounded returns X_t for S&P 500 Stock Price Index from Jan 2, 1990 to Dec 29, 2000.

the log difference of the stock price index; it has only $T-1 = 2779$ observations. The time series plot of X_t is given in Figure 4.2.

Some summary statistics of the X_t are shown in Table 4.1. As expected, the average of the returns X_t is slightly positive since the stock price index Y_t, from Figure 4.1, definitely has an upward trend. (In the model fitting which follows, we have not substracted the mean. However, analysis of the series $X_t - \bar{X}$ gives very similar results to those obtained below.) The skewness coefficient is -0.29 which suggests that X_t is slightly left skewed; the density plot of X_t in Figure 4.3 obtained by plotting the x- and y-coordinates from the output of S-Plus function *density* agrees with this skewness coefficient. The value of the kurtosis is greater than 3, meaning that it has an heavier tail than the standard normal distribution; this can also be shown from the normal probability plot in Figure 4.4.

Estimates	Values
Mean	0.00047
Standard Deviation	0.00946
Skewness	-0.29108
Kurtosis	7.72087

Table 4.1: Summary statistics of the continuously compounded returns X_t for S&P 500 Stock Price Index.

Figure 4.5 shows the plot of the autocorrelation function for X_t using the function *acf* in S-Plus. There is only weak dependence between the X_t since most of the autocorrelation coefficients at different time lags are within the approximate 95% limits (the dotted lines).

Miller (1979) mentioned that the residuals of a fitted ARMA model did not appear to be autocorrelated but the squared residuals seemed to be significantly correlated. Therefore, it seems reasonable to consider the volatility (or variability) of the X_t. Since the mean is almost zero, we consider the autocorrelation function plot of X_t^2; and this is shown in Figure 4.6. It shows that there is a substantial dependence

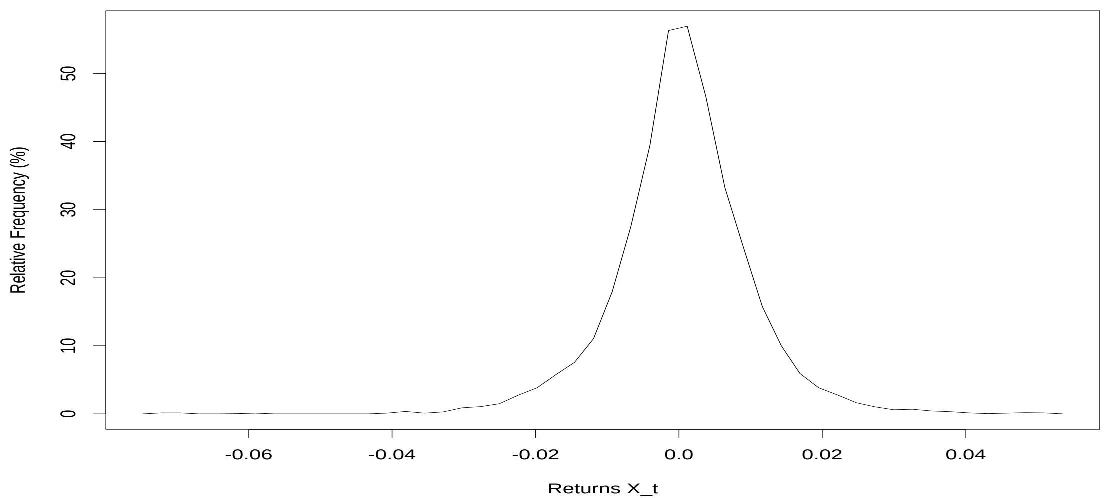

Figure 4.3: Density plot of the continuously compounded returns X_t for S&P 500 Stock Price Index.

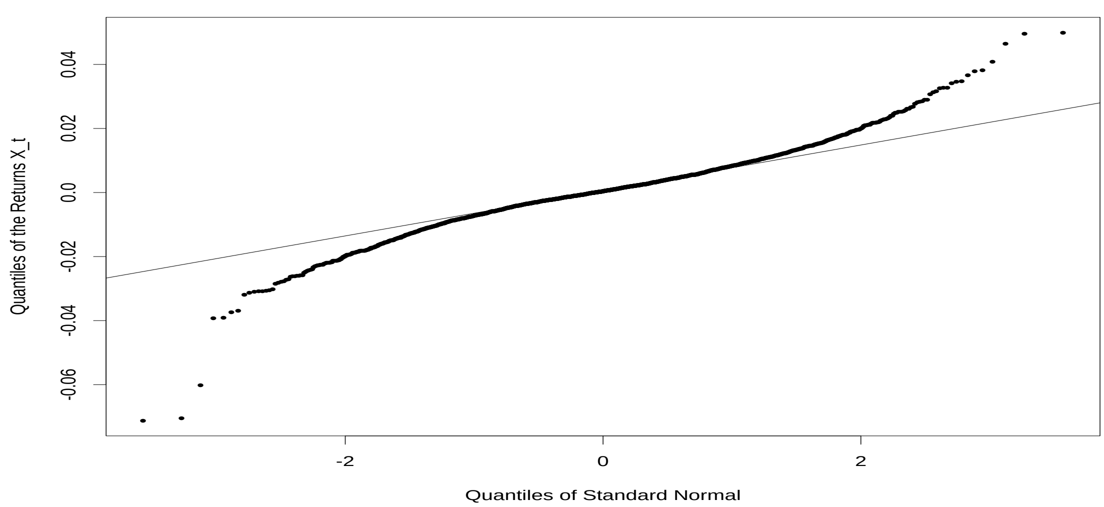

Figure 4.4: Normal probability plot of the continuously compounded returns X_t for S&P 500 Stock Price Index.

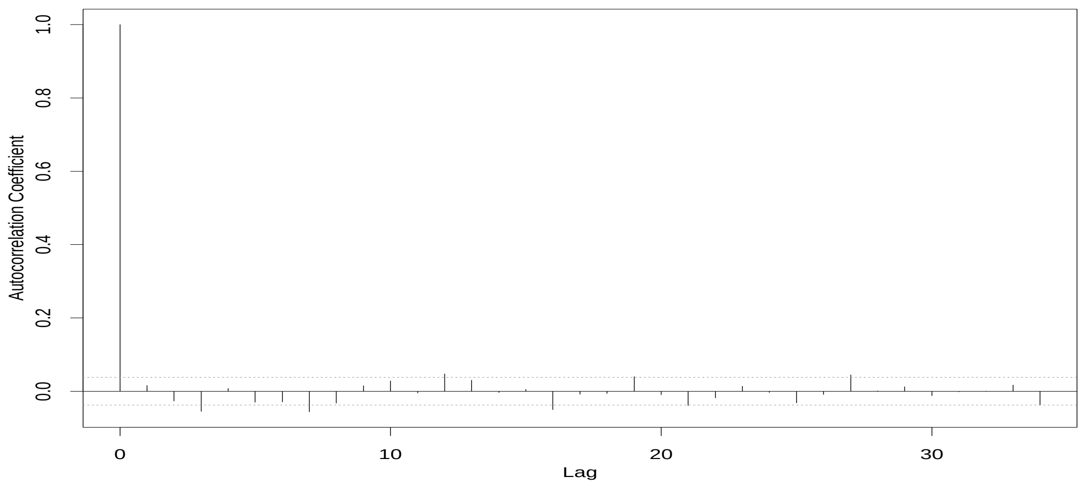

Figure 4.5: Autocorrelation function plot for the continuously compounded returns X_t of the S&P 500 Stock Price Index.

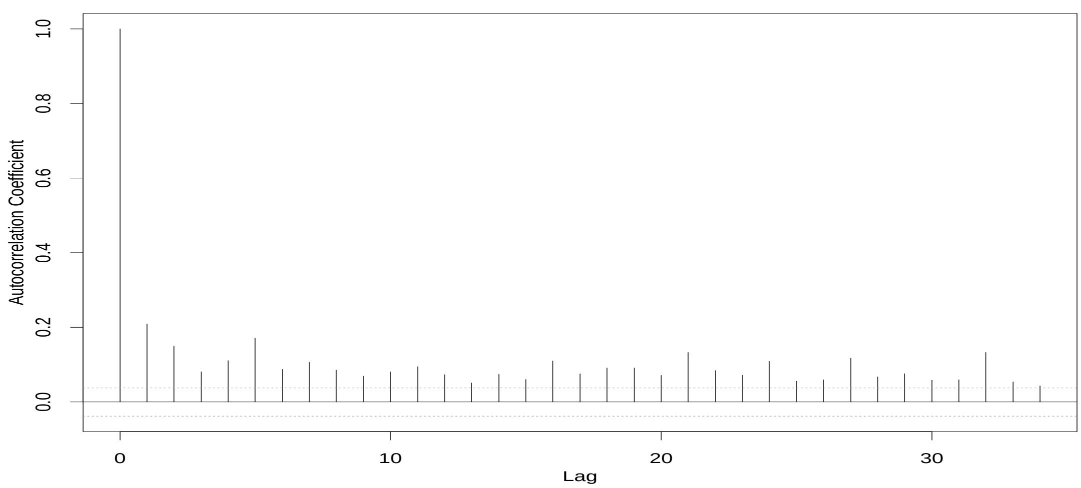

Figure 4.6: Autocorrelation function plot for the squared continuously compounded returns X_t^2 of the S&P 500 Stock Price Index.

between X_t^2 because all of the autocorrelation coefficients of X_t^2 are above the approximate 95% limit. Therefore, it seems appropriate to see how a GARCH$(1,1)$ model fits X_t.

4.2 Fitting ARCH(1) Models and Diagnostics

Before we fit the GARCH$(1,1)$ model, it is interesting to fit a simpler ARCH(1) model defined in (2.8). The discussion on the procedures used will be brief in this section, but it will be detailed in Sections 4.3 and 4.4.

The iterative method for estimation is similar to Section 3.4.1. The function *nlminb* from S-Plus is used to perform the maximum likelihood estimation. However, we will have to guess the initial values of the parameters to start the estimation; thus $\alpha_0 = 1.00$ and $\alpha_1 = 0.20$ are used[1]. The estimates obtained are $\hat{\alpha}_0 = 7.2 \times 10^{-5}$ and $\hat{\alpha}_1 = 2.1 \times 10^{-1}$.

After we have fit the ARCH(1) model, it is appropriate to examine how well the model fits the data. Recall that $X_t = Z_t \sqrt{h_t}$ where Z_t are iid standard normal random variables. Thus, if the ARCH(1) model is appropriate, then $X_t/\sqrt{h_t}$ should exhibit the behaviour of white noise, though the mean might not be zero.

Using the estimates $\hat{\alpha}_0$ and $\hat{\alpha}_1$, we can compute the estimated conditional variance $\hat{h}_t$ for ARCH(1) models defined as:

$$\hat{h}_t = \hat{\alpha}_0 + \hat{\alpha}_1 X_{t-1}^2 \quad \text{for} \quad t = 1, \cdots, T-2.$$

The estimated conditional variance $\hat{h}_t$ is then used to compute the *residuals* $\hat{Z}_t = X_t/\sqrt{\hat{h}_t}$ for diagnostic purposes.

[1]Different combinations of initial values of parameters are also used but they all give estimates which converge to the same estimates given above.

Figure 4.7 gives the normal probability plot for the residuals $\hat{Z}_t$ after fitting the ARCH(1) model. It shows substantial signs of heavier tails than the standard normal distribution. We examine further by looking at the autocorrelation function plot of the square of the residuals $\hat{Z}_t^2 = X_t^2/\hat{h}_t$ in Figure 4.8. It shows that there is still some autocorrelation among the squared residuals as more than half of the autocorrelation coefficients at different lags are outside the approximate 95% limits.

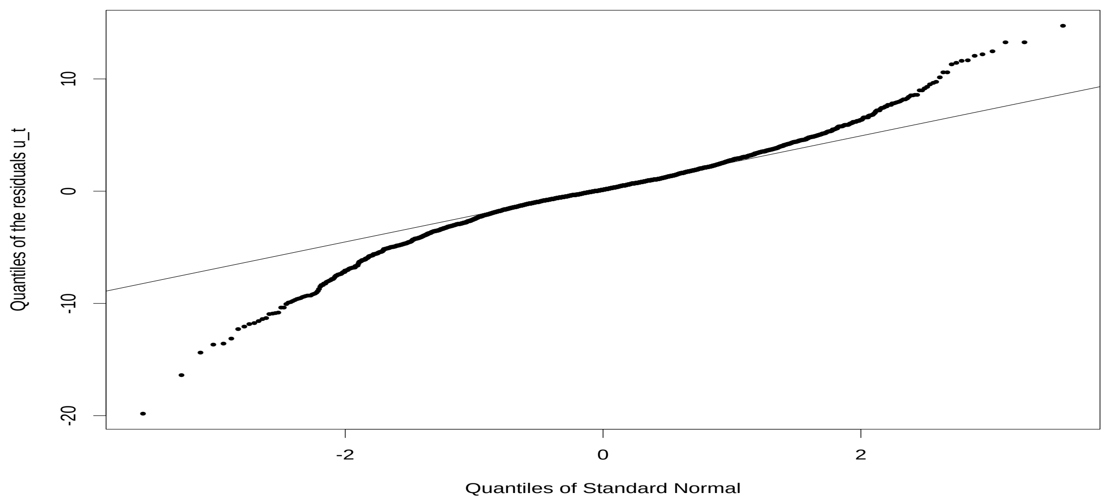

Figure 4.7: Normal probability plot for the residuals $\hat{Z}_t = X_t/\sqrt{\hat{h}_t}$ of the S&P 500 Stock Price Index using ARCH(1).

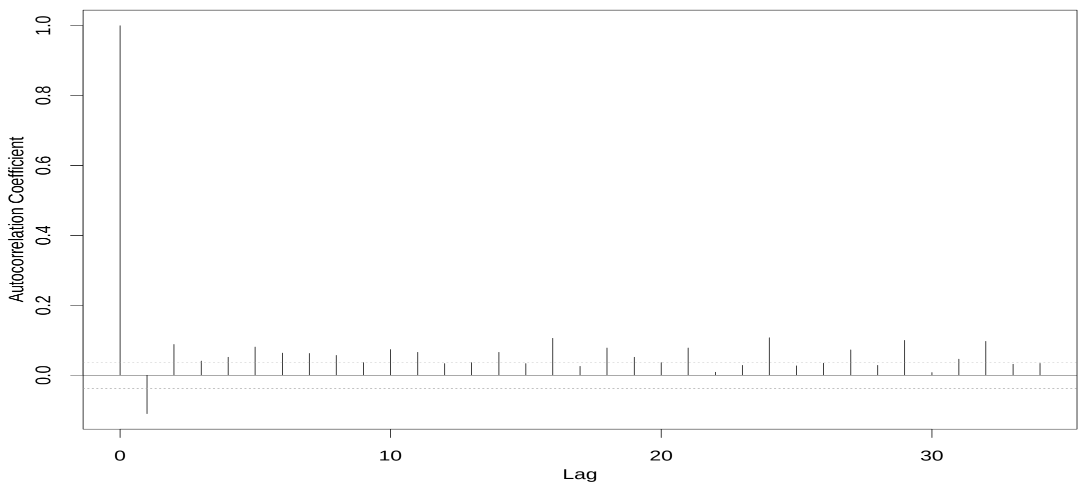

Figure 4.8: Autocorrelation function plot for the squared residuals $\hat{Z}_t^2 = X_t^2/\hat{h}_t$ of the S&P 500 Stock Price Index using ARCH(1) model.

4.2.1 Ljung-Box Q-Statistic

In addition to the visual inspection of the plotted autocorrelation function, the Ljung-Box Q-statistic is used for diagnostic checking. The Ljung-Box Q-statistic (Ljung and Box, 1978) is defined by:

$$Q = n(n+2)\sum_{j=1}^{K}\frac{r_j^2}{(n-j)}$$

where n is the number of observations, K is the largest lag used and r_j is the sample autocorrelation function at lag j of an appropriate time series X_t, for example. Statistic r_j for X_t is then defined as

$$r_j = \sum_{t=j+1}^{n}\left(X_t - \bar{X}\right)\left(X_{t-j} - \bar{X}\right)\Big/\sum_{t=1}^{n}(X_t - \bar{X})^2$$

where $\bar{X} = \dfrac{1}{n}\displaystyle\sum_{t=1}^{n} X_t$.

The number of observations n is $T-2=2778$ and largest lag K is 34 in this case. When fitting ARMA(p,q) models to data and testing the residuals to see if they are approximately white noise, under the null hypothesis that the series is ARMA(p,q), Q has approximately the χ^2 distribution with $(K-p-q)$ degrees of freedom.

The Q-statistic is a modification of the Box-Pierce test statistic (Box and Pierce, 1970); this was suggested for testing AR, MA and ARMA models. Both test statistics are based on the calculation of the sample autocorrelation function for the residuals $\hat{\epsilon}_t$ from those models. McLeod and Li (1983) argued that a similar test statistic based on different calculations using the autocorrelation function will be more useful for small sample applicability; it is defined as

$$Q^* = n(n+2)\sum_{j=1}^{K}\frac{r_j^{*2}}{(n-j)} \sim \chi^2_K$$

and r_j^* is

$$r_j^* = \sum_{t=j+1}^{n}(\hat{\epsilon}_t^2-\bar{\epsilon})(\hat{\epsilon}_{t-j}^2-\bar{\epsilon})\Big/\sum_{t=1}^{n}(\hat{\epsilon}_t^2-\bar{\epsilon})^2$$

where $\bar{\epsilon}=\frac{1}{n}\sum_{t=1}^{n}\hat{\epsilon}_t$. Li and Mak (1994) suggested the applicability of Q^* to the heteroscedastic time series models. For example, the test statistic for an ARCH(q) models is

$$\tilde{Q} = n\sum_{j=q+1}^{K}\tilde{r}_j^2 \sim \chi^2_{(K-q)}$$

where $\tilde{r}_j$ is a function of the squared residuals $\hat{Z}_t^2 = X_t^2/\hat{h}_t$ and is defined as:

$$\tilde{r}_j = \sum_{t=j+1}^{n}(\hat{Z}_t^2-\bar{Z})(\hat{Z}_{t-j}^2-\bar{Z})\Big/\sum_{t=1}^{n}(\hat{Z}_t^2-\bar{Z})^2$$

and

$$\bar{Z} = \frac{1}{n}\sum_{t=1}^{n}\hat{Z}_t^2 = \frac{1}{n}\sum_{t=1}^{n}X_t^2/\hat{h}_t.$$

Since $E(X_t^2/\hat{h}_t)=1$ and the sample autocorrelation function can also be defined as

$$\tilde{r}_j = \sum_{t=j+1}^{n}(\hat{Z}_t^2-1)(\hat{Z}_{t-j}^2-1)\Big/\sum_{t=1}^{n}(\hat{Z}_t^2-1)^2.$$

Ling and Li (1997) further extended the idea to the multivariate heteroscedastic time series context. Alternative diagnostic test for ARCH(q) models was proposed by Hong and Shehadeh (1999); it is based on the weighted sum of the squared sample autocorrelations of the squared residuals, with more weight being put in the terms of smaller lags. Horváth and Kokoszka (2001) developed the asymptotic theory for the linear statistic of sample autocorrelations of the squared residuals from an ARCH(q) model.

In our ARCH(q) context, we do not have enough knowledge about the large sample theory of the Ljung-Box Q-statistic and hence do not know much about its behaviour. Nonetheless, we proceed by analogy, and suppose the Ljung-Box Q-statistic has the χ^2 distribution with $(K-q)$ degrees of freedom. The critical value for $\chi^2_{(34-1)} = \chi^2_{33}$ with a 95^{th} percentile of 47.4.

Alternatively, Monte Carlo simulations could be used to find the exact p-value of the Ljung-Box Q-statistic for the ARCH(1) model. This can be done by generating, say, 1000 Monte Carlo samples from the ARCH(1) models using the estimated parameters $\hat{\alpha}_0$ and $\hat{\alpha}_1$ and then computing the Q-statistic for each sample. The sample p-value, or the probability of the number of samples which are as extreme or more extreme than our Q-statistic can be found, and we can see how well the ARCH(1) actually fits the data.

The Ljung-Box Q-statistic for X_t^2 is 893 (using the χ^2 approximation, the p-value ≈ 0) and this shows strong evidence of autocorrelation for the series X_t^2. Further, the Ljung-Box Q-statistic for the squared residuals $\hat{Z}_t^2 = X_t^2/\hat{h}_t$ after fitting the ARCH(1) is 375 (p-value ≈ 0). Clearly, this model does not adequately explain the S&P 500 Stock Price Index based on the normal probability plot and the Q-statistic, so we now try the GARCH$(1,1)$ model.

4.3 Fitting GARCH$(1,1)$ Models

As in Section 4.2, we first estimate the parameters, namely α_0, α_1 and β_1, for the GARCH$(1,1)$ model, then compute the series $\hat{h}_t$ and do some diagnostics on the fit.

The initial values of the estimates used are $\hat{\alpha}_0 = 0.01$, $\hat{\alpha}_1 = 0.05$ and $\hat{\beta}_1 = 0.90$, based on previous studies by Engle and Patton (2001), Hull (2000) and others. After 29 iterations, those estimates converge to the results shown in Table 4.2.

	Estimates
$\hat{\alpha}_0$	4.57×10^{-7}
$\hat{\alpha}_1$	5.00×10^{-2}
$\hat{\beta}_1$	9.46×10^{-1}

Table 4.2: Estimated parameters of X_t for GARCH$(1,1)$ model.

Since h_t is defined from the previous X_t and h_t, we will start the estimated conditional variance series $\hat{h}_t$ at $t = 1$. Also, we could have substituted h_1 by its own expected value. However, the expected value is not directly observed and we will use the observed X_0^2 instead, as the unconditional expected values of X_t^2 and h_t are equal. Then we compute $\hat{h}_t$ for $t = 2, \cdots, T-1$ recursively using the formula below with the estimates for α_0, α_1 and β_1:

$$\hat{h}_t = \hat{\alpha}_0 + \hat{\alpha}_1 \, X_{t-1}^2 + \hat{\beta}_1 \, \hat{h}_{t-1}.$$

Some of the X_t and estimated $\hat{h}_t$ are displayed in Table 4.3.

Figure 4.9 shows the time series plot for this estimated series of conditional variance $\hat{h}_t$. Notice that the estimated volatility is high for some periods and low for other periods. Recall that $\hat{\beta}_1$ is close to one and $\hat{\alpha}_0$ and $\hat{\alpha}_1$ are small. Since $\hat{h}_t = \hat{\alpha}_0 + \hat{\alpha}_1 \, X_{t-1}^2 + \hat{\beta}_1 \, \hat{h}_{t-1}$, we see that $\hat{h}_t$ tends to be close to $\hat{h}_{t-1}$. In other words, large values of h_t are clustered together and so are the small values of h_t.

Date	t	Y_t	X_t	$\hat{h}_t$	$\sqrt{\hat{h}_t}$
Jan-02-1990	0	359.69	-2.5889×10^{-3}		
Jan-03-1990	1	358.76	-8.6503×10^{-3}	6.70×10^{-6}	0.003
Jan-04-1990	2	355.67	-9.8041×10^{-3}	1.05×10^{-5}	0.003
Jan-05-1990	3	352.20	4.5043×10^{-3}	1.52×10^{-5}	0.004
Jan-08-1990	4	353.79	-1.1857×10^{-2}	1.59×10^{-5}	0.004
Jan-09-1990	5	349.62	-1.1482×10^{-2}	2.25×10^{-5}	0.005
	⋮		⋮		
	⋮		⋮		
Dec-26-2000	2776	1315.19	1.0385×10^{-2}	2.42×10^{-4}	0.016
Dec-27-2000	2777	1328.92	3.9803×10^{-3}	2.34×10^{-4}	0.015
Dec-28-2000	2778	1334.22	-1.0503×10^{-2}	2.23×10^{-4}	0.015
Dec-29-2000	2779	1320.28			

Table 4.3: Values of the S&P 500 Stock Price Index Y_t, the continuously compounded returns X_t, the estimated conditional variance $\hat{h}_t$ and standard deviation $\sqrt{\hat{h}_t}$ based on the estimates $\hat{\alpha}_0$, $\hat{\alpha}_1$ and $\hat{\beta}_1$ from Table 4.2 for GARCH$(1,1)$ models.

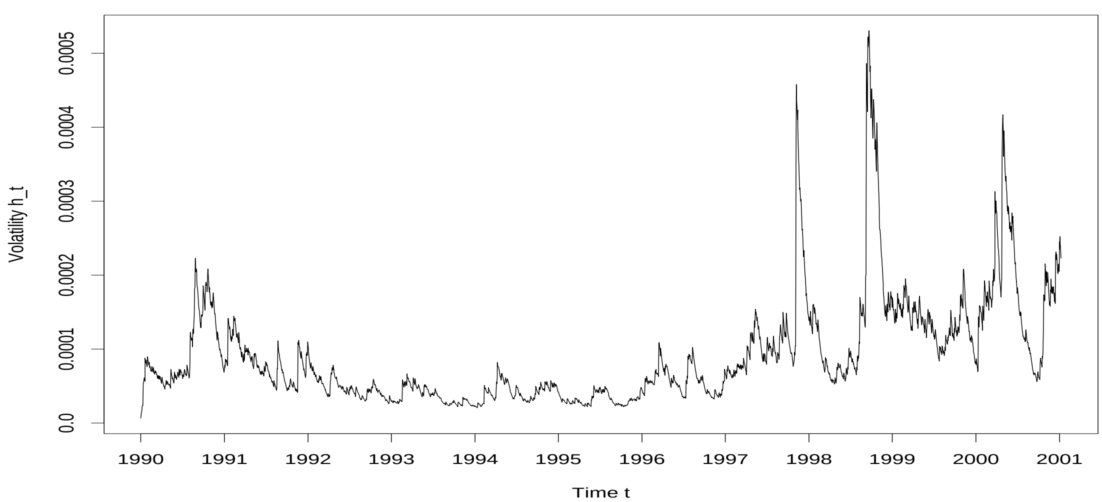

Figure 4.9: Time series plot for the estimated conditional variance $\hat{h}_t$ derived recursively from X_t and the estimated parameters $\hat{\alpha}_0$, $\hat{\alpha}_1$ and $\hat{\beta}_1$ from Table 4.2.

4.4 Diagnostics for the GARCH$(1,1)$ Models

After we have fit the model, it is appropriate to examine how well the GARCH$(1,1)$ model fits the data. Figure 4.10 gives the normal probability plot for the residuals $\hat{Z}_t = X_t/\sqrt{\hat{h}_t}$. It still shows that the new model has heavier tails than the standard normal distribution. As before, we then examine the autocorrelation function plot for the squared residuals $\hat{Z}_t^2 = X_t^2/\hat{h}_t$ shown in Figure 4.11. It seems there is not much dependence amongst $\hat{Z}_t^2$.

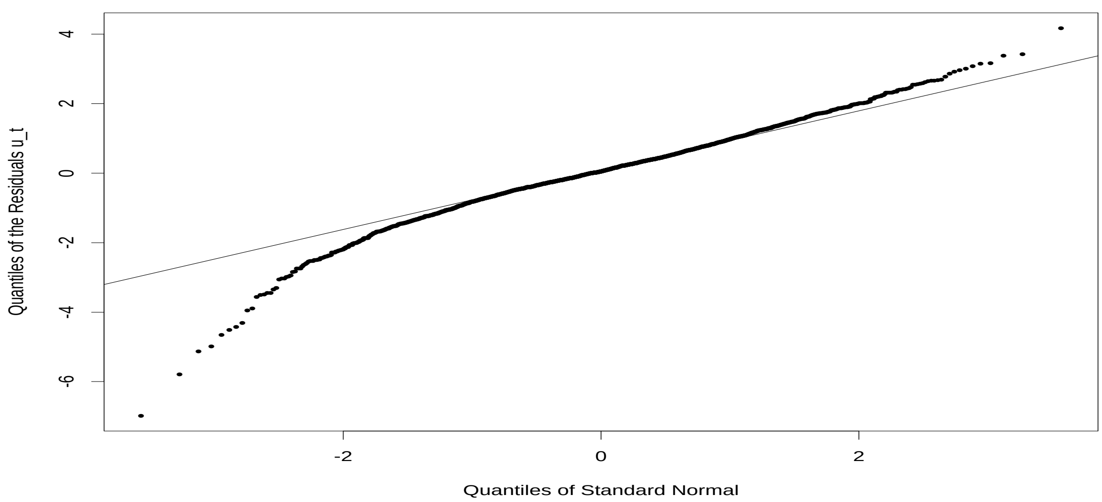

Figure 4.10: Normal probability plot for the residuals $\hat{Z}_t = X_t/\sqrt{\hat{h}_t}$ of the S&P 500 Stock Price Index using GARCH$(1,1)$.

In addtion, we also use the Ljung-Box Q-statistic to assess the fit. The number of observations n is $T - 2 = 2778$ and largest lag used K is 34 in this case. As for the ARCH model, modification of Q-statistic has been suggested, but we shall take the original form, and proceed by analogy to the ARMA(p,q) case and suppose the Ljung-Box Q-statistic for GARCH(p,q) models has the χ^2 distribution with $(K-p-q)$ degrees of freedom. The critical value for $\chi^2_{(34-1-1)} = \chi^2_{32}$ with a 95^{th} percentile of 46.2.

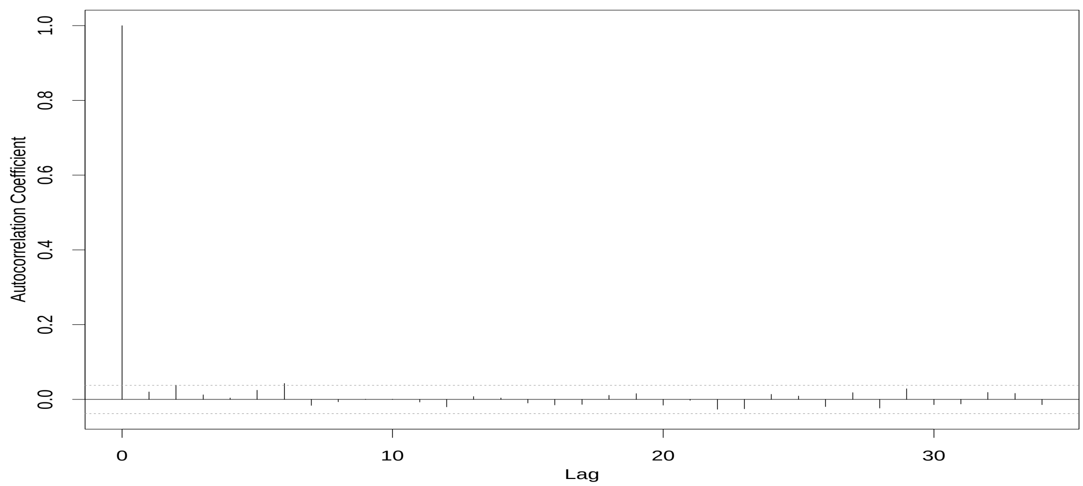

Figure 4.11: Autocorrelation function plot for the squared residuals $\hat{Z}_t^2 = X_t^2/\hat{h}_t$ of the S&P 500 Stock Price Index using GARCH$(1,1)$.

The Ljung-Box Q-statistic for the $\hat{Z}_t^2$ series after fitting the GARCH$(1,1)$ model is 31.5 (p-value of 0.49), suggesting that there is no significant correlation for the squared residuals $\hat{Z}_t^2$. The autocorrelation has been substantially removed by the GARCH$(1,1)$ model.

Moreover, Figure 4.12 shows the p-values obtained by calculating the Ljung-Box Q-statistics at different values of K, based on the χ^2 distribution with $(K - p - q)$ degrees of freedom. Note that at $K = 3, 6, 7$ and 8, the p-values are smaller than 0.05. The fact that $K = 3$ gives significance suggests that at least one more parameter is needed in the GARCH model. The higher p-values for $K \geq 9$ may be explained by the fact that using large values of K such as 34 dilutes the power of the Ljung-Box Q-statistic if the true correlation function at lag j is close to zero for higher j.

In summary, the GARCH$(1,1)$ model with estimated parameters $\hat{\alpha}_0 = 4.57 \times 10^{-7}$, $\hat{\alpha}_1 = 5.00 \times 10^{-2}$ and $\hat{\beta}_1 = 9.46 \times 10^{-1}$ fits the log difference X_t of the Standard & Poor's 500 Stock Price Index reasonably well. However, the significant value of the

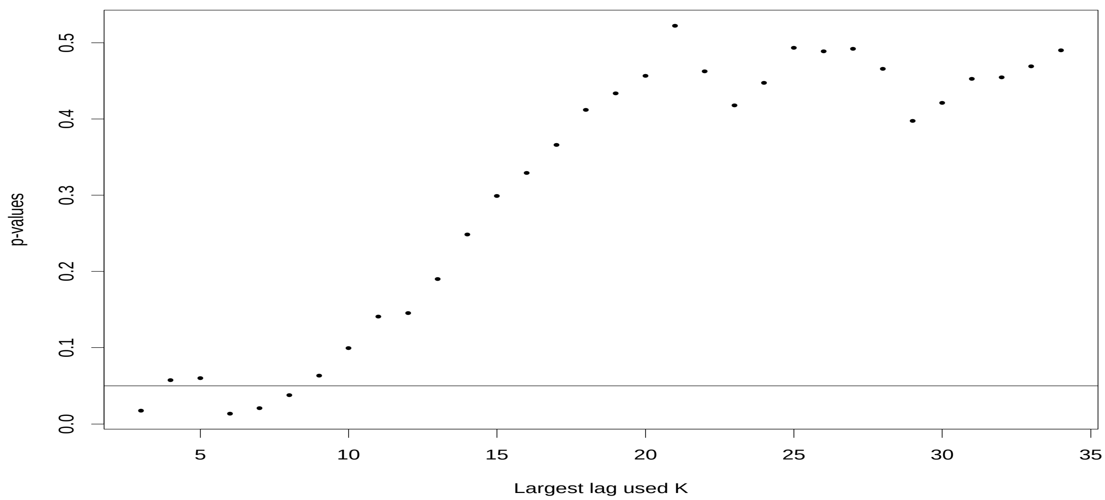

Figure 4.12: Diagnostic plot of p-values against the number of lags used K for the residuals X_t^2/h_t of the S&P 500 Stock Price Index.

autocorrelation of the residuals, for $K = 3$, suggests that models like GARCH$(1, 2)$, GARCH$(2, 1)$ or GARCH$(2, 2)$ might be still more successful in fitting the S&P index than GARCH$(1, 1)$ since the autocorrelation of the residuals would be better modelled.

4.5 Final Remarks

In this project, we have shown how some recently developed models for time series, particularly applicable to financial time series are used. The special feature of the models is that the series volatility is modelled as a function of the previous values of the variable. The simpler forms of the ARCH and GARCH models have been fitted to some financial data. However, many properties of these models are still to be investigated. To assess the fit of the models, various diagnostics have been suggested, but their properties, for example, their ability to detect a wrong model, have not yet been sufficiently studied. This also will provide a major topic of research.

Bibliography

Anderson, T. W. and Walker, A. M. (1964). On the Asymptotic Distribution of the Autocorrelation of a Sample from a Linear Stochastic Process. *Annals of Mathematical Statistics* **35**, 1296–1303.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics* **31**, 307–327.

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control.* 3rd edition. Prentice-Hall Inc., New Jersey.

Box, G. E. P. and Pierce, D. A. (1970). Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. *Journal of the American Statistical Association* **65**, 1509–1526.

Chan, N. H. (2002). *Time Series: Application to Finance.* John Wiley & Sons, Inc., New York.

Davies, N., Triggs, C. M. and Newbold, P. (1977). Significance Levels of the Box-Pierce Portmanteau Statistic in Finite Samples. *Biometrika* **64**, 517–522.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* **50**, 987–1008.

Engle, R. F. and Patton, A. J. (2001). What good is a volatility model? *Quantitative Finance* **1**, 237–245.

Gouriéroux, C. (1997). *ARCH Models and Financial Applications.* Springer-Verlag, New York.

Hong, Y. M. and Shehadeh, R. D. (1999). A New Test for ARCH Effects and Its Finite-Sample Performance. *Journal of Business & Economic Statistics* **17**, 99–108.

Horváth, L. and Kokoszka, P. (2001). Large Sample Distribution of Weighted Sums of ARCH(p) Squared Residual Correlations. *Econometric Theory* **17**, 283–295.

Hull, J. C. (1999). *Options, Futures & Other Derivatives.* 4th edition. Prentice-Hall Inc., New Jersey.

Li, W. K. (1992). On the Asymptotic Standard Errors of Residual Autocorrelations in Nonlinear Time Series Modelling. *Biometrika* **79**, 435–437.

Li, W. K. and Mak, T. K. (1994). On the Squared Residual Autocorrelations in Non-Linear Time Series with Conditional Heteroskedasticity. *Journal of Time Series Analysis* **15**, 627–636.

Li, W. K. and McLeod, A. I. (1981). Distribution of the Residual Autocorrelations in Multivariate ARMA Time Series Models. *Journal of the Royal Statistical Society, Series B (Methodological)* **43**, 231–239.

Ling, S. and Li, W. K. (1997). Diagnostic Checking of Nonlinear Multivariate Time Series with Multivariate ARCH Errors. *Journal of Time Series Analysis* **18**, 447–464.

Ljung, G. M. and Box, G. E. P. (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrika* **65**, 297–303.

Lukacs, E. (1975). *Stochastic Convergence* in *Probability and Mathematical Statistics: A Series of Monographs and Textbooks.* 2nd edition. Academic Press Inc., New York.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *Journal of Business* **36**, 394–419.

McLeod, A. I. and Li, W. K. (1983). Diagnostic Checking ARMA Time Series Models Using Squared-Residual Autocorrelations. *Journal of Time Series Analysis* **4**, 269–273.

Milhøj, A. (1985). The Moment Structure of ARCH Processes. *Scandinavian Journal of Statistics* **12**, 281–292.

Miller, R. B. (1979). Book Review on 'An Introduction to Bilinear Time Series Models' by C.W. Granger and A.P. Andersen. *Journal of the American Statistical Association* **74**, 927.

Wei, W. S. (1994). *Time Series Analysis: Univariate and Multivariate.* Addison-Wesley Publishing Company.

Weiss, A. A. (1984). ARMA Models with ARCH Errors. *Journal of Time Series Analysis* **5**, 129–143.