Contributions to Conditional Heteroscedastic Models: M-estimation and other methods.

Farhat Iqbal

Submitted for the degree of Doctor of Philosophy
at Lancaster University,
October 2009.

Abstract

This research makes contributions to conditional heteroscedastic models in financial time series. A class of M-estimators for time series models with asymmetric form of heteroscedasticity are developed. A weighted resampling method is used to approximate the sampling distribution of M-estimators. The primary finding is that there are estimators in this class that can perform better than the widely-used quasi-maximum likelihood estimator (QMLE) and even outperform the least absolute deviation estimator.

The asymptotic distributions of the squared and absolute residual autocorrelations for generalised autoregressive conditional heteroscedastic (GARCH) models estimated by M-estimators are derived. Diagnostic tests based on M-estimators are developed to check the adequacy of GARCH-type models.

The performance of M-estimators in the estimation and prediction of value-at-risk (VaR) is investigated. A wide range of summary statistics is used to evaluate and compare M-estimators in estimating the in-sample and predicting the out-of-sample VaR of three well-known stock indices. Some of the M-estimators are observed to show better performances in predicting the one-day-ahead VaR than the commonly-used QMLE.

The Linear Estimator (LE) for ARCH models is explored and results show that this estimator provides good estimates for the parameters of the ARCH model and also predicts the volatility better than the QMLE. Using a class of weighted resampling schemes, it is found that there are schemes that can match and even perform better than the commonly-used paired bootstrap scheme. Bootstrap prediction intervals for returns, volatilities and value-at-risk in ARCH models are also developed.

A weighted linear estimator (WLE) for the multivariate ARCH parameters is proposed. This estimator involves solving sets of linear equations and hence is very easy to compute. A weighted resampling method for multivariate ARCH models is also discussed. The accuracy of this estimator is compared with the QMLE in estimating the parameters of multivariate ARCH models. The WLE is also applied to real data sets and forecasts of volatilities and value-at-risk are obtained. Our study indicates that the forecasting performance of the WLE is not inferior to the QMLE and one-day-ahead risk estimates are found better.

M-estimators for multivariate GARCH models are discussed. Two different methods for the estimation of multivariate GARCH models using univariate GARCH specifications are proposed. These methods are easy to apply as these require several univariate GARCH estimations to estimate the full multivariate GARCH model. Results of Monte Carlo simulations and application to real data sets show that our methods provide better results in terms of estimating and predicting the conditional correlations and value-at-risk.

Acknowledgements

In the name of Allah, the Most Gracious, the Most Merciful.

In the first place I would like to thank my supervisor, Dr. Kanchan Mukherjee, for his academic guidance, his never-ending patience in answering questions, as well as giving me complete liberty in what I like to do.

I would also like to thank University of Balochistan Quetta, Pakistan for the financial support for this research.

Finally, I am deeply indebted to my parents and to my wife for all their support during my entire studies, encouraging me to continue my studies, helping me to overcome difficulties when things seemed fuzzy and being patient when I was nervous due to my work. They deserve more thanks that I can ever properly express. Without them this thesis would not have been completed.

Declaration

I declare that this thesis is composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified. The following publications are derived from this work:

i. Iqbal, F., and Mukherjee, K. (2009). M-estimators of some GARCH-type models; computation and application. *Statistics and Computing.* (in press) http://www.springerlink.com/content/5017077t01265408/.

ii. Iqbal, F., and Mukherjee, K. (2009). A study of value-at-risk based on M-estimators of GARCH models. (submitted)

iii. Iqbal, F. (2009). Diagnostic checking for GARCH-type models estimated by M-estimators. (submitted)

iv. Iqbal, F. A Weighted Linear Estimator of Multivariate ARCH Parameters. (under preparation)

v. Iqbal, F. M-estimators for multivariate GARCH models. (under preparation)

Farhat Iqbal

Contents

List of Figures

Chapter 1

Introduction

Many financial time series, such as stock returns and currency exchange rates are well described by *stylised facts* such as excess kurtosis and volatility clustering. In order to capture these features in the financial data, autoregressive conditional heteroscedastic (ARCH) model was introduced by Engle (1982). Since the introduction of the ARCH model many extensions have been proposed. Among them the generalised ARCH (GARCH) model of Bollerslev (1986) is the most popular. Many applications of the GARCH model to financial data sets have found that this model provides a good fit to the data. However, one weakness of this model is that it responds equally to positive and negative shocks and hence cannot capture the asymmetric feature common in many asset returns. For this purpose asymmetric GARCH models are proposed.

The estimation of GARCH-type models is often carried out using the quasi-maximum likelihood estimator (QMLE) where the Gaussian likelihood is used for the true but possibly unknown likelihood. The QMLE is consistent and asymptotically normal if the innovation has four finite moments. However, such stringent moment condition may not hold in many situations; an example is innovations with student-t distribution where the degree of freedom is at most four.

The main focus of this thesis is on GARCH-type models estimated by M-estimators which are applicable under weak moment assumptions. We address the issues of estimation, diagnostic testing, resampling and forecasting in both univariate and multivariate heteroscedastic models. In addition, we provide a detailed investigation of the Linear Estimator (LE), an alternative estimator to the QMLE for ARCH models. Thus the main aims of this thesis are (i) to develop computational algorithms for new and improved estimation techniques for ARCH-GARCH models (ii) to extend existing work on M-estimation of GARCH models and LE for ARCH models (iii) to develop testing, bootstrap and prediction techniques based on such estimators and (iv) to introduce new areas of application.

This thesis can be divided into three parts. In the first part of the thesis (Chapters 3–5), robust M-estimators for asymmetric GARCH models are proposed and a weighted resampling method for GARCH models is discussed. Diagnostic tests for GARCH models, when M-estimators are used for estimation, are developed. The performance of these estimators in terms of predicting risk estimates is evaluated using a wide range of summary statistics. In the second part of the thesis (Chapters 6–7), the Linear Estimator (LE) for ARCH models is explored empirically and weighted resampling schemes for LE are investigated. Bootstrap prediction intervals for returns, volatilities and VaR are also developed. A weighted version of LE (WLE) for multivariate ARCH models is proposed and a weighted resampling scheme for WLE in multivariate ARCH setup is defined. In the last part of the thesis (Chapter 8), some robust methods for multivariate GARCH models using univariate GARCH specification are introduced. All proposed methods are investigated through extensive Monte Carlo simulations and applications to real data sets. Some interesting results are found in this study and based on these results suggestions are made for the estimation, diagnostic checking, bootstrapping and prediction of ARCH-GARCH models.

Chapter 2 highlights some common characteristics of financial time series also known as stylised facts. We present a brief review of the ARCH, GARCH and asymmetric GARCH models. The estimation, testing and forecasting methods for these models are also discussed. The aim of this chapter is to make the reader familiar with the fundamental concepts and related work.

Our aim in Chapter 3 is to propose robust methods for GARCH models that can capture the asymmetric property of financial time series. We also aim to explore estimators that can perform better than the commonly-used QMLE in terms of parameters estimation and volatility forecasting. M-estimators for asymmetric GARCH models are defined. The class of estimators includes least absolute deviation (LAD), Huber's, Cauchy and B-estimator as well as the QMLE. Algorithms for the computation of these estimators are presented. Extensive simulations are used to check the relative performance of these estimators in both symmetric and asymmetric GARCH models. A weighted resampling method is used to approximate the sampling distribution of M-estimators. Our study indicates that there are estimators that can perform better than the QMLE and even outperform the robust LAD estimator when the error distribution is heavy-tailed. These estimators are applied to analyse real data sets.

There is a huge literature on the modelling of conditional heteroscedastic time series, but not much work has been done on model checking or model selection. Testing the adequacy of these heteroscedastic models is undoubtedly important for several economic and statistical reasons. Diagnostic is one of the important stages of model building. In Chapter 4, goodness-of-fit tests in the class of conditional heteroscedastic time series models are examined. Portmanteau statistics based on squared and absolute autocorrelations of residuals from GARCH-type models estimated by M-estimators are developed. The asymptotic distributions for these statistics are obtained and size and power analyses are conducted through Monte Carlo simulations. It is found that the asymptotic standard errors for both squared and absolute residual autocorrelations match the empirical standard errors quite satisfactorily for all estimators. Investigation of the sizes of these tests suggests that their empirical sizes are close to the nominal level. Analysis of the power of tests reveals that tests based on absolute residual autocorrelations outperform those based on squared residual autocorrelations. The power levels of tests, when Cauchy and B-estimators are used for estimation, are found superior to other estimators.

Risk management is one of the important tasks for financial institutions, non-financial corporations, regulators and asset managers. Value-at-risk (VaR) is a commonly-used statistic for measuring potential risk of economic losses in financial market. Our aim in Chapter 5 is to study the performance of M-estimator in estimating and predicting risk estimates. We propose evaluation measures and M-tests for this purpose. Symmetric and asymmetric GARCH models using M-estimators are fitted to three major stock indices and both the in-sample and out-of-sample VaR estimates are obtained. The predictive performances, when estimators other than the QMLE are applied for GARCH models, are investigated using various evaluation measures and M-tests. It is found that these estimators not only fit the data well, they also predict the VaR accurately and hence provide reliable estimates for risk. A comparison of results for both symmetric and asymmetric GARCH models reveals that an asymmetric model provides better forecasts for the data set used.

Bose and Mukherjee (2003) proposed the linear estimator (LE) estimator for the ARCH model. The computation of this estimator involves solving only two sets of linear equations. An advantage of LE over the widely-used QMLE is that its computation is very easy and requires less CPU time which enables one to perform computer intensive tasks on ARCH models in little time. Chapter 6 makes three contributions. First, a detailed investigation of the Linear Estimator (LE) for the ARCH model in terms of parameter estimation and volatility forecasting is provided. This estimator provides good results for the estimation of the parameters of ARCH models and also produces better volatility forecasts than the QMLE in almost all Monte Carlo simulations. These findings are further supported by application to three stock indices. Second, a weighted resampling method for the linear estimator is presented to approximate the distribution of the parameters of ARCH models. Results of our experiments show that alternative schemes such as Scheme E and Scheme U match the widely-used paired bootstrap and residual bootstrap and even perform better than these commonly-used methods. Third, bootstrap prediction intervals for returns, volatilities and VaR in ARCH models are developed. Monte Carlo results show that although both estimators provide good mean coverage, the LE can be considered superior in terms of its mean lengths close to the empirical with low standard errors.

Chapter 7 makes two contributions. First, a weighted linear estimator for multivariate ARCH parameters is proposed. The accuracy of this estimator is compared with the QMLE in estimating the parameters of multivariate ARCH models. This estimator is also applied to real data sets and forecasts of volatilities and value-at-risk are obtained. Results show that the forecasting performance of the WLE is not inferior to the QMLE. Also, one-day-ahead risk estimates are found to be better. As a second contribution of this chapter, a weighted resampling method for multivariate ARCH models is proposed. Using different weights for bootstrap it is shown that Scheme U and Scheme E provide better results than the commonly-used paired bootstrap.

Multivariate GARCH models focus on volatility and correlation analysis for more than one asset. In order to study the relations between the volatilities and co-volatilities of financial time series the development of multivariate GARCH-type models is very important. The problems in the application of multivariate GARCH models are that the number of parameters becomes large as the dimension of the system increases and many constraints need to be imposed to ensure the positive definiteness of the covariance matrix. The main objective of Chapter 8 is to propose robust methods for multivariate GARCH models that are easy to estimate and do not put additional constraints on the model. We propose M-estimators for multivariate GARCH models using univariate GARCH specification. Two different models are introduced and the results of simulations and real data analysis show that our robust estimators perform better than the widely-used QMLE in terms of estimating and predicting the conditional correlations and risk estimates. Our methods can be applied easily to high dimensional financial time series, since the number of parameters is relatively small.

Finally, in Chapter 9, a summary of the research is given and some possible directions for future research are discussed.

Chapter 2

Financial Time Series and Literature Review

2.1 Introduction

Autoregressive (AR), Moving Average (MA) and the Autoregressive Moving Average (ARMA) models are often very useful in modelling general time series. All these models are based on the assumption that the errors have equal variances. This is also known as homoscedasticity of the errors. When dealing with financial market variables such as daily quotes on a share, stock indices or currency exchange rates, this assumption is not appropriate due to some of the features of these financial data sets.

Let $\{P_t, t = 1, ..., T\}$, be a time series of prices of a financial asset. These prices are highly correlated and the variances of prices often increase with time. This makes the statistical analysis of prices difficult and often the prices P_t are transformed to log-returns

$$X_t = \log P_t - \log P_{t-1} = \log\left(\frac{P_t}{P_{t-1}}\right), \quad t = 1, ..., T.$$

The log-returns are easy to handle and have attractive statistical properties. These series are free of units and can be compared with each other.

2.2 Characteristics of Financial Time Series

Many financial time series have a number of characteristics in common. Cont (2001) presented a set of stylised empirical facts emerging from the statistical analysis of price variations in various types of financial markets. Some of the important 'stylised facts', in financial log-return series $\{X_t\}$, are described below.

Leptokurtosis:

The distribution of the financial asset returns is *leptokurtic*, i.e., exhibits excess kurtosis (heavy-tails) and have sharp peak. The frequency of large and small changes, relative to the range of data, is rather high which leads us to believe that the data do not come from a normal but a heavy-tailed (leptokurtic) distribution (relative high probability for extreme values).

Volatility Clustering:

Large and small values in a log-returns sample tend to occur in clusters. Extreme returns tend to be followed by other extreme returns, although not necessarily with the same sign. Mandelbrot (1963) quoted: "*... large changes tend to be followed by large changes -of either sign- and small changes by small changes...*". Fama (1965) also reported this behaviour. The implication of volatility clustering is that the volatility shock today influences the expectation of volatilities of many future periods ahead.

Leverage Effect:

The series $\{X_t\}$ responds differently to its own positive and negative movements. In other words, changes in stock prices tend to be negatively correlated with changes in volatility, i.e., volatility is higher after negative shocks than after positive shocks of same magnitude. This asymmetry, or "leverage" was first documented empirically by Black (1976). Some other studies that find evidence of leverage effect are Nelson (1991), Engle and Ng (1991) and Glosten et al. (1993).

Long-range Dependence:

Sample autocorrelations of the log-returns are small whereas the sample autocorrelations of the absolute and squared log-returns are significantly different from zero even for large lags. Absolute or squared log-returns exhibit significant positive autocorrelation or persistence (slow decay in autocorrelations).

Figure 2.1 below shows the monthly log-returns of IBM Stock from 1926 to 1999 (888 observations) and the *Quantile-Quantile* plot (Q-Q plot) of the marginal distribution of $\{X_t\}$ against the standard normal. Some of the properties mentioned above can be observed in this figure. It can be seen from the top panel of the figure that there is a time varying volatility (conditional heteroscedasticity) in the log returns. The Q-Q plot at the bottom reveals that the distribution is heavy-tailed.

Figure 2.2 contains three autocorrelation plots. The first plot is the autocorrelation plot of monthly log-returns. The autocorrelation plots of the squared and absolute log-returns can be seen in the second and third plot in Figure 2.2, respectively. These plots clearly indicate that although there is no significant autocorrelations in log-returns, the squared and absolute log-returns show profound dependence. This implies that large price variations are more likely to be followed by large price variations, and small price variations are more likely to be followed

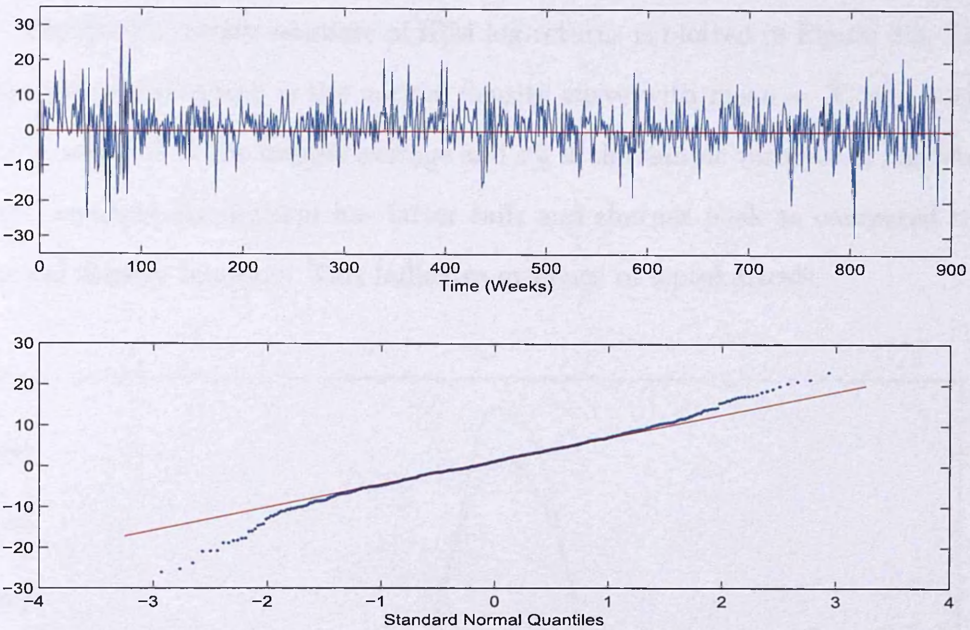

Figure 2.1: Monthly log-returns and Q-Q plot of IBM stock (1926-1999).

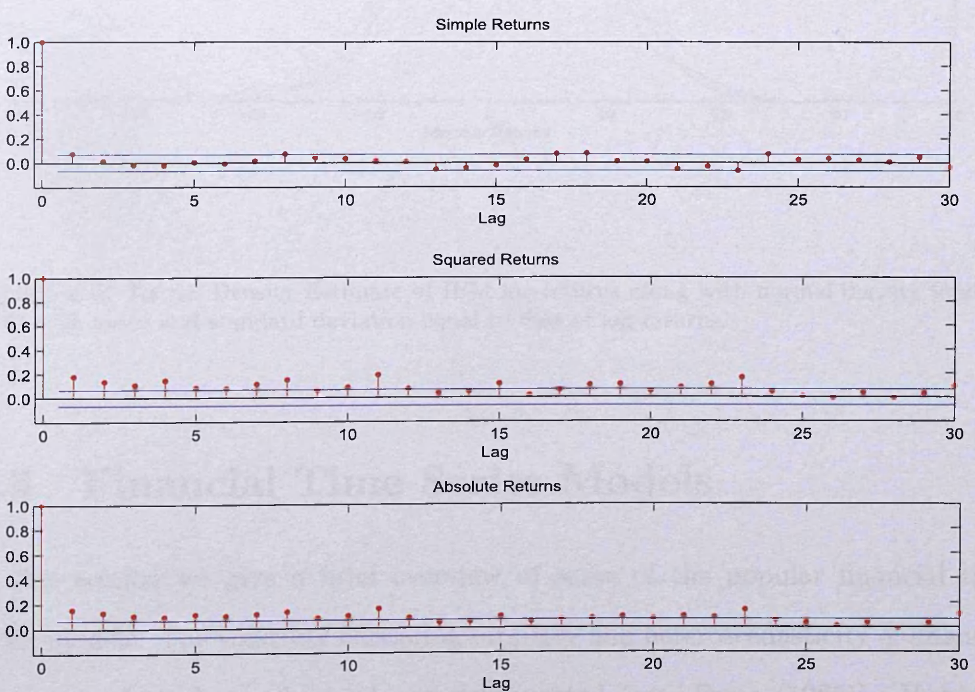

Figure 2.2: Autocorrelation plots of IBM log-returns, squared and absolute log-returns.

by small price variations.

The kernel density estimate of IBM log-returns is plotted in Figure 2.3. Superimposed on this graph is the normal density curve with mean $= \bar{X}$, and variance $= \sigma_X^2$, where $\bar{X}$ is the sample average and σ_X^2 is the sample variance of log-returns. The empirical distribution has fatter tails and sharper peak as compared to the normal density function. This indicates evidence of leptokurtosis.

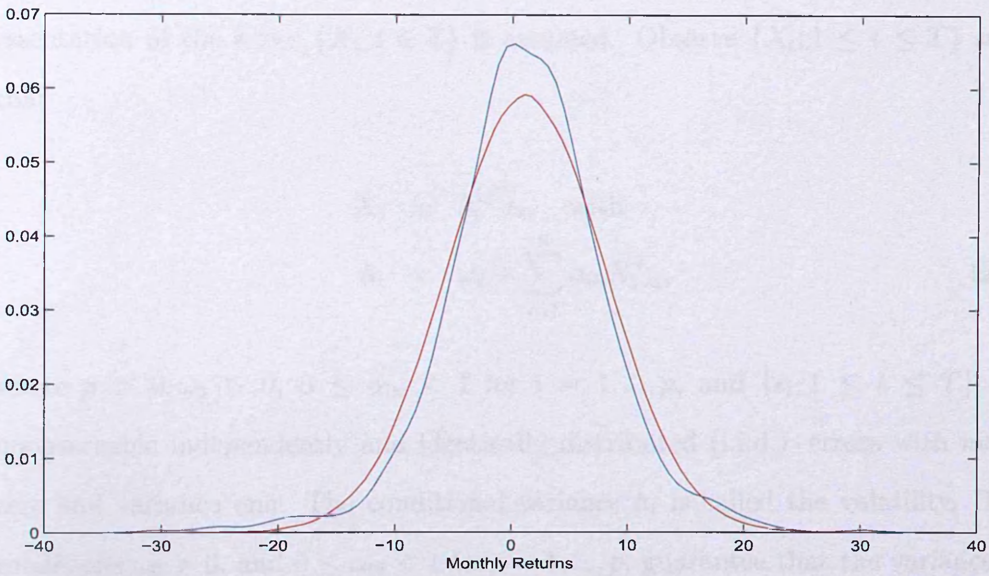

Figure 2.3: Kernel Density Estimate of IBM log-returns along with normal density function (red) with mean and standard deviation equal to that of log-returns.

2.3 Financial Time Series Models

In this section we give a brief overview of some of the popular financial time series models. The volatility clustering property and heteroscedasticity of financial time series have been observed and documented (see, Fama (1965)). However, traditional time series models cannot explain these properties well and new models are, therefore, required. In order to capture these stylised facts in the financial data, the autoregressive conditional heteroscedastic (ARCH) model was introduced by Engle (1982). This model for the conditional volatility has been extended by Bollerslev (1986), Nelson (1991) and Glosten et al. (1993), amongst many others.

2.3.1 The ARCH Model

Engle (1982) proposed the autoregressive conditional heteroscedastic (ARCH) model for the log-returns. The conditional variance in ARCH is specified as a linear function of past squared disturbances. In the ARCH(p) model, the following representation of the series $\{X_t; t \in \mathbb{Z}\}$ is assumed. Observe $\{X_t; 1 \leq t \leq T\}$ such that

$$\begin{aligned} X_t &= h_t^{1/2}\epsilon_t, \quad \text{with} \\ h_t &= \omega_0 + \sum_{i=1}^{p} \alpha_{0i} X_{t-i}^2, \end{aligned} \tag{2.1}$$

where $p > 0, \omega_0 > 0$, $0 \leq \alpha_{0i} < 1$ for $i = 1, ..., p$, and $\{\epsilon_t; 1 \leq t \leq T\}$ are unobservable independently and identically distributed (i.i.d.) errors with mean zero and variance one. The conditional variance h_t is called the volatility. The conditions $\omega_0 > 0$, and $0 \leq \alpha_{0i} < 1$ for $i = 1, ..., p$, guarantee that the variance of $\{X_t\}$ remains positive.

The ARCH model states that the conditional variance of X_t is an increasing function of the square of the shock that occurred in the previous time period. Therefore, if the absolute value of X_{t-1} is large, the absolute value of X_t is expected to be large as well. In this way the ARCH model can describe volatility clustering.

The conditional mean and variance of $\{X_t\}$ can be obtained as

$$\begin{aligned} E(X_t|\mathcal{F}_{t-1}) &= E(h_t^{1/2})E(\epsilon_t) = 0, \\ \text{Var}(X_t|\mathcal{F}_{t-1}) &= E(X_t^2|\mathcal{F}_{t-1}) = h_t, \end{aligned}$$

where $\mathcal{F}_{t-1}$ is the information set available up to time $t-1$. The unconditional variance of $\{X_t\}$, denoted by h, is given by

$$h = \frac{\omega_0}{1 - \sum_{i=1}^{p} \alpha_{0i}}.$$

The process $\{X_t\}$ is covariance stationary if and only if the sum of autoregressive parameters is less than one, i.e., $\sum_{i=1}^{p} \alpha_{0i} < 1$.

The ARCH(1) model is

$$\begin{aligned} X_t &= h_t^{1/2}\epsilon_t, \quad \text{with} \\ h_t &= \omega + \alpha_1 X_{t-1}^2. \end{aligned} \tag{2.2}$$

This model can be written as an AR(1) model of X_t^2. Let $\nu_t = X_t^2 - h_t$ so that $h_t = X_t^2 - \nu_t$. By plugging this into (2.2), we get

$$X_t^2 = \omega + \alpha_1 X_{t-1}^2 + \nu_t,$$

where $\{\nu_t\}$ is a martingale difference series.

2.3.2 The GARCH Model

To adequately describe the volatility process, the ARCH model requires estimation of many parameters. Since the introduction of ARCH model, many extensions of ARCH model are proposed and among these the generalized autoregressive conditional heteroscedastic (GARCH) model of Bollerslev (1986) where the volatility is not only a function of past observations but also past volatility, is certainly the most popular and successful.

In the GARCH(p,q) model, the following representation of the series $\{X_t; t \in \mathbb{Z}\}$ is assumed. Observe $\{X_t; 1 \le t \le T\}$ such that

$$\begin{aligned} X_t &= h_t^{1/2}\epsilon_t, \quad \text{with} \\ h_t &= \omega_0 + \sum_{i=1}^{p} \alpha_{0i} X_{t-i}^2 + \sum_{j=1}^{q} \beta_{0j} h_{t-j}, \end{aligned} \tag{2.3}$$

where $p > 0, q \ge 0, \omega_0 > 0$, $0 \le \alpha_{0i} < 1$ for $i = 1, ..., p$, $0 \le \beta_{0j} < 1$ for $j = 1, ..., q$, and $\{\epsilon_t; 1 \le t \le T\}$ are unobservable i.i.d. errors symmetric about zero. When $q = 0$, the GARCH model reduces to the ARCH model.

The unconditional variance of X_t is

$$h = \frac{\omega_0}{1 - \sum_{i=1}^{p} \alpha_{0i} - \sum_{j=1}^{q} \beta_{0i}}.$$

The process $\{X_t\}$ is covariance stationary if and only if $\sum_{i=1}^{p} \alpha_{0i} + \sum_{j=1}^{q} \beta_{0j} < 1$.

The GARCH model is a special case of an infinite-order (ARCH(∞)) model $X_t = h_t^{1/2}\epsilon_t$ with

$$h_t = \omega_0 + \sum_{i=1}^{\infty} \alpha_i X_{t-i}^2.$$

The ARCH(∞) representation is very useful when the existence of moments and long memory properties of ARCH and GARCH models are under consideration; see Giraitis et al. (2000).

Consider the GARCH(1,1) model

$$\begin{aligned} X_t &= h_t^{1/2}\epsilon_t, \quad \text{with} \\ h_t &= \omega + \alpha_1 X_{t-1}^2 + \beta_1 h_{t-1}, \ t \in \mathbb{Z}. \end{aligned} \tag{2.4}$$

The GARCH(1,1) model can be written as ARMA(1,1) model for X_t^2 as

$$X_t^2 = \omega + (\alpha_1 + \beta_1)X_{t-1}^2 + \nu_t - \beta_1 \nu_{t-1},$$

where $\nu_t = X_t^2 - h_t$. Notice that $E(\nu_t|\mathcal{F}_{t-1}) = 0$.

The one-step-ahead forecast of volatility is readily available and is given as

$$h_{T+1} = \omega + \alpha_1 X_T^2 + \beta_1 h_T.$$

The forecast of h_{T+k} for $k > 1$, make use of the fact that $E(X_{T+k}^2) = h_{T+k}$ and given by

$$h_{T+k} = \omega + (\alpha_1 + \beta_1)h_{T+k-1}.$$

By repeated substitutions, the k-step-ahead forecast can be written as

$$h_{T+k} = \frac{\omega[1 - (\alpha_1 + \beta_1)^{k-1}]}{1 - \alpha_1 - \beta_1} + (\alpha_1 + \beta_1)^{k-1} h_{T+1}.$$

This shows that as $k \to \infty$, the multi-step ahead volatility forecast of a GARCH(1,1) model converges to the unconditional variance, $h = \omega/(1 - \alpha_1 - \beta_1)$.

Usually a GARCH(1,1) model is adequate to obtain a good model fit for financial time series. For many financial time series the value of the GARCH coefficient β is found to be close to 0.9. This shows that large values of h_{t-1} will be followed by large values of h_t, and small values of h_{t-1} will be followed by small values of h_t and hence the GARCH model captures the volatility clustering in financial time series.

Many financial time series have fatter tails than a normal distribution. Bollerslev (1986) showed that the kurtosis implied by a GARCH(1,1) model with normal errors is greater than the kurtosis of a normal distribution assuming that the fourth order moment exists. Thus a GARCH model with normal errors can replicate some of the fat-tailed behaviour of financial time series.

When the GARCH(1,1) model of (2.4) is applied to high-frequency financial time series data, it is often observed that the estimate of α_1 and β_1 are such that their sum is close to or equal to one. Those models are called Integrated GARCH (IGARCH) that results $\alpha_1 + \beta_1 = 1$ (Engle and Bollerslev, 1986). Although IGARCH model is not weakly stationary, it is shown that the IGARCH(1,1) model may still be strictly stationary (Nelson, 1990).

2.3.3 Asymmetric GARCH Models

Standard GARCH models assume that positive and negative values of past observations have a symmetric effect on the volatility. In other words, good and bad news have the same effect on the volatility in the GARCH model and the sign of the shock is irrelevant. In practice this assumption is frequently violated, in particular by stock returns, in that the volatility increases more after bad news than after good news. This is so called the *leverage effect.* However, much applied research is still conducted assuming implicitly the existence of symmetric dynamics, which may lead to model misspecification if dynamic asymmetry is indeed present. The GARCH model is not a suitable choice for modelling the asymmetric effect in the returns.

The GJR Model

To capture the asymmetric effect of positive and negative shocks on volatility, a few variants of the basic GARCH model have been proposed. One such model is presented by Glosten, Jaganathan and Runkle (1993). This asymmetric GARCH model commonly known as GJR model is capable of capturing the asymmetric feature. The basic variant is the GJR(1, 1) model in which the following representation of the series $\{X_t; t \in \mathbb{Z}\}$ is assumed. Observe $\{X_t; 1 \leq t \leq T\}$ such that

$$\begin{aligned} X_t &= h_t^{1/2}\epsilon_t, \\ h_t &= \omega + \alpha_1 X_{t-1}^2 + \beta_1 h_{t-1} + \gamma_1 D_{t-1} X_{t-1}^2 \end{aligned} \tag{2.5}$$

where $D_{t-1} = 1$ if $X_{t-1} < 0$ and 0 otherwise, with

$$\omega > 0, \alpha_1, \beta_1, \gamma_1 \geq 0, \quad \text{and} \quad (\alpha_1 + \beta_1 + \frac{1}{2}\gamma_1) < 1,$$

and $\{\epsilon_t\}$ is a sequence of (i.i.d.) unobservable real-valued random variables. This model is based on the assumption that unexpected changes in the return have different effects on the conditional variance. An unexpected increase (good news) contributes to the variance in the model through multiplicator α_1 whereas an unforeseen fall (bad news) generates an increase in volatility through multiplicator $\alpha_1 + \gamma_1$. The non-negative value of the coefficient γ_1 indicates a 'leverage effect'. With $\gamma_1 = 0$, GJR model reduces to GARCH model.

The k-step-ahead forecast of the GJR(1,1) model can be obtained in a similar manner as of the GARCH(1,1). Assuming that

$$P(\epsilon_t \equiv h_t^{-1/2} X_t < 0) = \frac{1}{2},$$

the k-step-ahead volatility forecast of GJR(1,1) model is given by

$$h_{T+k} = \frac{\omega[1 - (\alpha_1 + \frac{1}{2}\gamma_1 + \beta_1)^{k-1}]}{1 - (\alpha_1 + \frac{1}{2}\gamma_1 + \beta_1)} + (\alpha_1 + \frac{1}{2}\gamma_1 + \beta_1)^{k-1} h_{T+1}.$$

Other volatility models that can capture the asymmetric property are the Exponential GARCH (EGARCH) model of Nelson (1991) and the Threshold GARCH (TGARCH) model of Zakoian (1994).

For review articles providing details of above GARCH models and their variants; see Bollerslev et al. (1992), Bollerslev et al. (1994) and Shephard (1996).

2.4 Testing ARCH Effects

Usually the autocorrelations of squared or absolute returns are used to check the volatility clustering in returns. The Ljung-Box statistic can be used to check the significance of these autocorrelations. A modified Q statistic is

$$Q_{LB}(M) = T(T+2)\sum_{k=1}^{M}\frac{\hat{r}_k^2}{T-k},$$

where $\hat{r}_k$ denote an estimate of k-lag sample autocorrelation of the absolute or squared returns. Under the null hypothesis that the data are white noise the $Q_{LB}(M)$ statistics has an asymptotic chi-square distribution with M degrees of freedom. A significant value for $Q_{LB}(M)$ provides evidence for time varying conditional volatility.

Engle (1982) proposed a Lagrange Multiplier (LM) test for ARCH effect. This test can be constructed based on the auxiliary regression

$$X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \cdots + \alpha_p X_{t-p}^2 + \nu_t, \tag{2.6}$$

where $\nu_t = X_t^2 - h_t$. Eq (2.6) is the AR(p) process for X_t^2. Under the null hypothesis that there are no ARCH effects, $\alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$, the LM test statistic

$$\text{LM} = T \cdot R^2$$

has an asymptotic chi-square distribution with p degrees of freedom, where T is the sample size and R^2 is computed from (2.6) using estimated residuals. This test can be used as a general specification test for GARCH effects.

2.5 Estimation

In practice, the values of the parameters in the GARCH models are not known and need to be estimated. In the following we discuss the quasi-maximum likelihood estimation (QMLE) which is generally used for the estimation of the parameters of the GARCH models. The QMLE is maximum likelihood applied to a model with the alteration that errors are presumed to be Gaussian. Under some regularity conditions, the QMLE is consistent and asymptotically normal.

Consider the GARCH model defined by (2.3). Let Θ be a compact subset of $(0,\infty)^{1+p} \times (0,1)^q$,. Let $\boldsymbol{\theta}_0 = (\omega_0, \alpha_{01}, \ldots, \alpha_{0p}, \beta_{01}, \ldots, \beta_{0q})'$ be the parameter vector to be estimated and $\boldsymbol{\theta} = (\omega, \alpha_1, \ldots \alpha_p, \beta_1, \ldots, \beta_q)'$ be any admissible value of $\boldsymbol{\theta}_0$. Let f denote the density function of $\epsilon_t(\boldsymbol{\theta}) = X_t(\boldsymbol{\theta})/h_t^{1/2}(\boldsymbol{\theta})$.

The negative log-likelihood functions is given by

$$\begin{aligned} L_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T} l_t(\boldsymbol{\theta}), \qquad (2.7) \\ \text{where} \quad l_t(\boldsymbol{\theta}) &= \left(\frac{1}{2}\right)\log(h_t(\boldsymbol{\theta})) - \log\{f(\epsilon_t(\boldsymbol{\theta}))\} \quad t = 1, 2, \ldots, T. \end{aligned}$$

The maximum likelihood estimator $\hat{\boldsymbol{\theta}}_T$ of the true parameter $\boldsymbol{\theta}_0$ is defined by minimising

$$\hat{\boldsymbol{\theta}}_T = \arg\min_{\theta \in \Theta} \hat{L}_T(\boldsymbol{\theta}),$$

where $\hat{L}_T(\boldsymbol{\theta})$ is an estimate of $L_T(\boldsymbol{\theta})$. The BHHH algorithm of Berndt et al. (1974) is often used to determine $\hat{\boldsymbol{\theta}}_T$. However, the BHHH algorithm faces serious convergence problem if the starting values are not sufficiently close to the solutions and instead a full Newton-Raphson algorithm may be used (Mak et al., 1997).

Several conditional distributions for the error density have been employed in the literature. The most common is the standard normal density for the errors

$$f(\epsilon_t(\boldsymbol{\theta})) = \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\epsilon_t^2(\boldsymbol{\theta}) \right\}$$

The log-likelihood function is then given as

$$l_t(\boldsymbol{\theta}) = \frac{1}{2}\log(2\pi) + \frac{1}{2}\log(h_t(\boldsymbol{\theta})) + \frac{1}{2}\epsilon_t^2(\boldsymbol{\theta})$$

The quasi-maximum likelihood estimates can be obtained as the solution of

$$\sum_{t=1}^{T} s_t(X_t; \boldsymbol{\theta}) = 0$$

where $s_t = \partial l_t / \partial \boldsymbol{\theta}$, is the vector of derivatives of the log-likelihood with respect to the parameters and is usually called the score function.

Throughout, for a function g, $\dot{g}$ and $\ddot{g}$ will denote the first and second derivatives, respectively, whenever they exist. The score function for the QMLE takes the form:

$$\begin{aligned} s_t = \frac{\partial l_t}{\partial \boldsymbol{\theta}} &= \frac{1}{2}\frac{\dot{h}_t(\boldsymbol{\theta})}{h_t(\boldsymbol{\theta})} - \frac{1}{2}\frac{X_t^2}{h_t(\boldsymbol{\theta})}\frac{\dot{h}_t(\boldsymbol{\theta})}{h_t(\boldsymbol{\theta})} \\ &= \frac{1}{2}\frac{\dot{h}_t(\boldsymbol{\theta})}{h_t(\boldsymbol{\theta})}\left(1 - \frac{X_t^2}{h_t(\boldsymbol{\theta})}\right) \end{aligned}$$

In order to capture the degree of tail fatness, several density functions for the normalized error have been proposed. After the Gaussian distribution originally used by Engle (1982), Engle and Bollerslev (1986) used the conditional student-t distribution. Nelson (1991) suggested the generalized error distribution. Hansen (1994) proposed the idea of using the skewed-t distribution to allow the shape and the skewness of distribution to change over time.

Some of the frequently used forms of f are:

Standardised Student-t density with ν degrees of freedom:

In this case, the density is given as

$$f(y) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi(\nu-2)}}\left(1+\frac{y^2}{\nu-2}\right)^{-\frac{\quad}{\quad}}$$

with $\nu > 2$ and $\Gamma(.)$ is the gamma function.

The log-likelihood is given as

$$l_t = \frac{1}{2}\log(h_t) - \log c(\nu) + \frac{\nu+1}{2}\log\left(1+\frac{y^2}{\nu-2}\right),$$

where $c(\nu) = \frac{\Gamma\left((\nu+1)/2\right)}{\Gamma(\nu/2)\sqrt{\pi(\nu-2)}}$.

Generalised Error Distribution (GED):

The GED is a symmetric distribution that can be both leptokurtic and platykurtic depending on the degree of freedom $\nu(\nu > 1)$. The GED has the following density function:

$$f(y) = \frac{\nu}{\lambda 2^{(1+1/\nu)}\Gamma(1/\nu)}\exp\left\{-\frac{1}{2}\left|\frac{y}{\lambda}\right|^{\nu}\right\},$$

where $\lambda = \left(2^{--}\Gamma(1/\nu)/\Gamma(3/\nu)\right)^{-}$ and $0 < \nu < 2$.

The log-likelihood for GED is

$$l_t = -\log C(\nu) + \frac{1}{2}\Big(\log(h_t) + \Big|\frac{y}{\lambda}\Big|^{\nu}\Big),$$

where $C(\nu) = \frac{\nu}{\lambda 2 \quad \Gamma(1/\nu)}$

The generalised error distribution reduces to the double exponential distribution when $\nu = 1$. The double exponential density is given as

$$f(y) = \exp\{-\sqrt{2}|y|\}/\sqrt{2}.$$

The tails of the GED distribution are thicker than the normal when $\nu < 2$ and thinner when $\nu > 2$.

Hansen's Skewed t-distribution:

It is often observed in application of GARCH-type models to financial time series that even asymmetric GARCH models fail to fully account for sample skewness and leptokurtosis of high frequency financial time series when they are assumed to follow normal or symmetric student-t distributions with ν degree of freedom. This has led to the use of asymmetric non-normal distributions. To better model conditional higher moments, Hansen (1994) assumed that the distribution of errors $\{\epsilon_t\}$ can be skewed and consider the following density function

$$f(\epsilon|\nu,\lambda) = \begin{cases} bc\left(1 + \frac{1}{\nu-2}\left(\frac{b\epsilon+a}{1-\lambda}\right)^2\right)^{-(\nu+1)/2} & \epsilon < a/b, \\ bc\left(1 + \frac{1}{\nu-2}\left(\frac{b\epsilon+a}{1+\lambda}\right)^2\right)^{-(\nu+1)/2} & \epsilon \geq a/b, \end{cases}$$

where $2 < \nu < \infty$, and $-1 < \lambda < 1$. The values of constants a, b and c are defined as

$$\begin{aligned} a &= 4\lambda c\Big(\frac{\nu-2}{\nu-1}\Big), \\ b^2 &= 1+3\lambda^2-a^2, \\ \text{and} \quad c &= \frac{\Gamma\Big((\nu+1)/2\Big)}{\sqrt{\pi(\nu-2)}\Gamma\Big(\nu/2\Big)}. \end{aligned}$$

We will denote the skewed-t distribution as $ST(\nu,\lambda)$. The skewed-t distribution takes the form of the student-t distribution when $\lambda = 0$ and nests the standard normal distribution when $\nu \to \infty$.

2.5.1 Asymptotics of the QMLE

The asymptotic normality for ARCH model was first presented by Weiss (1986). He proved the consistency and asymptotic normality for the QMLE in the linear ARCH(p) model under the existence of fourth-order moment of the ARCH process. These conditions are violated when GARCH models are fitted to financial data and also ruled out the IGARCH processes. The necessary and sufficient condition for the existence of the second-order moment of the GARCH(1,1) model and the necessary and sufficient condition for the fourth-order moments of GARCH(1, 2) and GARCH(2, 1) model were established by Bollerslev (1986). For the GARCH(1,1) and IGARCH(1,1) models, Lumsdaine (1996) established that the local QMLE is consistent and asymptotically normal assuming that $E[\log(\alpha\,\epsilon_t^2+\beta)] < 0$ and that the rescaled errors have 32-nd order moment. Lee and Hansen (1994) also considered the GARCH(1,1) model and required that $E[\epsilon_t^{2+k}] < \infty$ for some $k > 0$ and gave the proof of consistency of the QMLE under the assumption that $\{X_t\}$ is strictly stationary and ergodic.

Bougerol and Picard (1992) established the necessary and sufficient condition for the strict stationarity and ergodicity of GARCH(p, q) model in terms of top Lyapunov exponent. Ling and Li (1997) proved that the local QMLE is consistent and asymptotically normal if $E[\epsilon_t^4] < \infty$. Ling and McAleer (2002) proved the consistency of the global QMLE under the second-order moment condition. They derived the asymptotic normality of the global QMLE under the finiteness of the sixth-order moment condition. Hall and Yao (2003) established the asymptotic normality of the QMLE for the general GARCH(p, q) models under certain conditions. They discovered that the asymptotic normality may not be normal with an infinite fourth moment. Berkes, Horvath and Kokoszka (2003) extended the results to hold for linear stationary GARCH(p, q) model under weaker conditions on rescaled errors. The QMLE under general conditional heteroscedastic models based on stochastic recurrence equations was studied by Straumann and Mikosch (2006) and the asymptotic normality of the QMLE for the ARCH(∞) model which includes GARCH as a very special case was derived by Robinson and Zaffaroni (2006).

For the GJR model, Ling and McAleer (2002) pointed out that the regularity condition for the existence of the second moment of the GJR(1,1) is $(\alpha+\beta+\gamma/2) < 1$, provided that $\{\epsilon_t\}$ is symmetric. When $\gamma = 0$, this condition reduces to the second moment condition for the GARCH(1,1) model. They also showed that when $\epsilon_t \sim N(0, 1)$, the fourth moment condition for the asymmetric GJR(1,1) model is given by $\beta^2 + 2\alpha\beta + 3\alpha^2 + \beta\gamma + 3\alpha\gamma + \frac{3}{2}\gamma^2 < 1$. This condition reduces to the fourth moment condition for the GARCH(1,1) model in case $\gamma = 0$. Ling and McAleer (2002b) developed the weak long-moment condition for the GJR(1,1) model and showed that $E[\log(\alpha + \gamma D\epsilon_t^2 + \beta)] < 0$ is sufficient for the consistency and asymptotic normality of the QMLE. For a detailed review of some theoretical results for time series models with GARCH errors; see Li et al. (2002) and references therein.

The QMLE is consistent and asymptotically normal if the innovation has finite four moments. However, such stringent moment condition may not hold in many situations; an example is innovations with student-t distribution where the degree of freedom is at most four. To deal with such situations, several authors have proposed robust estimators for GARCH models and derived their asymptotic normality under less stringent moment conditions. Peng and Yao (2003) suggested that for heavy-tailed distributions, least absolute deviations (LAD) estimators should be used imposing an extra restriction such as median$(\epsilon_t^2) = 1$. Berkes and Horváth (2004) considered the QMLE and LAD however assuming $E(\epsilon^2) = 1$ or $E(|\epsilon|) = 1$ or $E\{|\epsilon|/(1+|\epsilon|)\}$ is known. Mukherjee (2008) considered M-estimator for GARCH model. The class of estimators include LAD, Huber's as well as the QMLE.

2.6 Forecasting

Forecasts from GARCH models are of great interest to researchers and practitioners. In this section we discuss some of the procedures used for volatility forecasting and also defined some commonly-used measures to evaluate these volatility forecasts.

Assume that $T = N + K$ observations are available for a time series X_t. The first N observations are used for the estimation of the model and the last K observations are left for evaluation of forecasts. There are three alternative ways to generate a sequence of K one-step ahead forecasts, namely the *recursive, rolling* and *fixed* schemes.

The *recursive scheme* uses the sample $\{1, \cdots N\}$ to estimate the model and generates the $(N+1)$-th forecast. The model is re-estimated using the sample $\{1, \cdots N, N+1\}$ and the second one-step ahead forecast is generated. These steps are repeated until the last one-step ahead forecast is generated based on the sample $\{1, \cdots, N+K-1\}$. In this scheme the sample size grows at each step and at the final step the one-step ahead forecast is based on $T-1$ observations. The recursive scheme uses all information available at each step to generate the forecasts but is not suitable for large K.

In the *rolling scheme*, the first one-step ahead forecast is produced in the same manner as in the *recursive scheme*. The model is re-estimated by dropping the first observation from the sample and including the $(N+1)$-th observation, i.e. the sample consists of $\{2, \cdots, N+1\}$ observations. Using this sample the second one-step ahead forecast is generated and the procedure continues till the last forecast is made. The *recursive scheme* uses the same window length N for estimating the model at each step. This scheme has an advantage over the recursive scheme as it uses a fixed sample size for any value of K and also omits information in the distant past giving more flexibility to time variations. For *fixed scheme*, the sample $\{1, \cdots, N\}$ is estimated and all forecasts are made based on this sample.

Poon and Granger (2003) provide an extensive review and references on forecasting volatility in financial markets. Hansen and Lunde (2005) provide a forecast comparison of the most important parametric formulations. Chuang et al. (2007) analyse the volatility forecasting performance of the GARCH models based on various distributional assumptions.

2.6.1 Comparing Forecasting Performance

There are various evaluation criteria available in the literature to compare the forecasting performance of two forecasting models. Let $\hat{h}_{i,t}$, denote the volatility forecast of h_t, for i-th GARCH model using any forecast scheme and $\varepsilon_{i,t} = \hat{h}_{i,t} - h_t$ be the forecast error at time t. Commonly used forecast evaluation statistics based on K out-of-sample forecast are the mean squared error (MSE), the mean absolute error (MAE), the root mean squared error (RMSE) and the mean absolute percentage error (MAPE) defined as

$$\begin{aligned}
\text{MSE}_i &= \frac{1}{K}\sum_{l=N+1}^{N+K} \varepsilon^2_{i,l|l-1}, \\
\text{MAE}_i &= \frac{1}{K}\sum_{l=N+1}^{N+K} |\varepsilon_{i,l|l-1}|, \\
\text{RMSE}_i &= \sqrt{\frac{1}{K}\sum_{l=N+1}^{N+K} \varepsilon^2_{i,l|l-1}}, \\
\text{MAPE}_i &= \frac{1}{K}\sum_{l=N+1}^{N+K} \frac{|\varepsilon_{i,l|l-1}|}{\sqrt{h_l}}.
\end{aligned}$$

These statistics are often used to evaluate volatility forecasts and the model which produces the smallest values of these statistics is considered the best model in terms of volatility forecasting. For real data, the true volatilities are unobservable and often the squared returns are used as proxies of the true volatilities. In case intra-day data are available, Andersen and Bollerslev (1998) showed that higher-frequency returns produce better approximations of realized volatilities than same-frequency returns.

Chapter 3

M-estimation of Heteroscedastic Models

3.1 Introduction

Commonly used statistical methods rely on a number of assumptions such as linearity of regression, independence and normality of errors, variance homogeneity, etc. However, often these assumptions are violated. Robust statistics investigates the effects of deviations from modelling assumptions and develops new, better procedures. There is a need of robust methods for financial time series due to the presence of large number of outliers in the financial data, heteroscedasticity and dependence of errors as discussed in Section 2.2.

One of the important concepts in many econometric models is volatility or the instantaneous variability of a financial time series. In a seminal paper, Engle (1982) introduced the autoregressive conditional heteroscedastic (ARCH) models to describe the volatility of the current return of an asset as a linear function of the squares of its past returns. This model explains some of the empirical stylised facts related to financial time series such as time-varying conditional volatility, volatility clustering and heavy-tailedness of the unconditional distribution of returns, among others. Since then many extensions of the ARCH model have been proposed. Among those, models where the volatility is not only a symmetric function of squares of past returns but also past conditional volatilities has turned out to be very useful. An important example is the generalized autoregressive conditional heteroscedastic (GARCH) model of order p, q, introduced by Bollerslev (1986).

A popular method for estimating the unknown parameters in GARCH models is to use the Gaussian likelihood of the innovations and the resulting estimator is called the quasi-maximum likelihood estimator (QMLE). The QMLE is consistent and asymptotically normal if the innovation has four finite moments. However, such stringent moment condition may not hold in many situations; an example is innovations with student-t distribution where the degree of freedom is at most four. To deal with such situations, several authors have proposed robust estimators for GARCH models and derived their asymptotic normality under less stringent moment conditions. See, for example, Peng and Yao (2003), Berkes and Horváth (2004), Muler and Yohai (2008) and Mukherjee (2008), among others.

In this chapter we propose a large class of M-estimators for estimating the parameters of GARCH-type models. The class of estimators include QMLE, least absolute deviation (LAD) estimator and the analogue of Huber's estimators as well as many other useful estimators. Computation of these estimators is a major issue in applications and one of the main contributions of this chapter is to describe general algorithm for the computation applicable to all score functions. We perform extensive simulations to compare the relative performance of M-estimators. Simulation study reveals that there are estimators such as Huber's and B-estimator that can perform better than the QMLE and even outperform the robust LAD estimator when the error distribution is heavy-tailed. Thus, as the second important contribution of this chapter, we propose the use of B-estimator as a desirable alternative to the QMLE and LAD. We also study the quality of approximation to the finite sample distribution of the M-estimators by a weighted resampling method. We compute several M-estimates for GARCH and GJR models when fitted to real data sets of IBM stock and S&P 500 index and use them to predict the value-at-risk (VaR).

The rest of the chapter is organised as follows. In Section 3.2, we briefly present M-estimators for location and scale models. In Section 3.3, we consider the class of M-estimators in both symmetric and asymmetric GARCH models, address the computational issues and state their distributional results. Results of Monte Carlo simulations and real data analysis are presented in Section 3.4. Section 3.5 presents the performance of M-estimators in estimating value-at-risk. A weighted resampling method for M-estimators is discussed in Section 3.6. Finally, Section 3.7 concludes the chapter.

3.2 M-estimators

Huber (1964) introduced the class of M-estimators for the location and regression models. For detailed studies on properties of M-estimators see Huber (1981) and Jurečková and Sen (1996). M-estimators are generalisations of the usual maximum-likelihood estimators. Consider the simple location model

$$y_i = \mu + \varepsilon_i \quad (i = 1, ..., n),$$

where ε_i's are independently and identically distributed (i.i.d.) random variables. Assume that F_0, the distribution function of ε_i, has a density $f_0 = \dot{F}_0$. The likelihood function is

$$L(y_1, ..., y_n; \mu) = \prod_{i=1}^{n} f_0(y_i - \mu)$$

The value that maximises $L(y_1, ..., y_n; \mu)$ is called the MLE of μ, i.e.,

$$\hat{\mu} = \arg\max L(y_1, ..., y_n; \mu) \tag{3.1}$$

If f_0 is positive everywhere, then taking the logarithm, $\hat{\mu}$ in (3.1) can be written as

$$\hat{\mu} = \arg\min \sum_{i=1}^{n} \rho(y_i - \mu), \tag{3.2}$$

where $\rho = -\log f_0$.

Differentiating (3.2) with respect to μ, provided that ρ is differentiable, we get

$$\sum_{i=1}^{n} \psi(y_i - \hat{\mu}) = 0 \tag{3.3}$$

where $\psi = \dot{\rho}$. Hence given a function ρ, an M-estimate of location is a solution of (3.3).

Similarly we can define an M-estimate of scale. Consider the multiplicative model

$$y_i = \sigma\varepsilon_i, \quad i = 1, ..., n, \tag{3.4}$$

where $\sigma > 0$ is the unknown parameter. The MLE of σ is

$$\hat{\sigma} = \arg\max \frac{1}{\sigma^n} \prod_{i=1}^{n} f_0\left(\frac{y_i}{\sigma}\right)$$

Now, taking the logarithm and differentiating the above with respect to σ gives

$$\frac{1}{n} \sum_{i=1}^{n} \rho\left(\frac{y_i}{\hat{\sigma}}\right) = 1$$

where $\rho(x) = -x(\dot{f}_0/f_0)$.

In general, any estimate satisfying an equation of the form

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{y_i}{\hat{\sigma}}\right) = \delta, \tag{3.5}$$

where δ is a positive constant is called an M-estimate of scale. M-estimator is determined by the choice of the criterion function ρ or of its derivative ψ. For many choices of ρ or ψ, no closed form expression exists. A few commonly used choices of ρ and ψ are given in Table 3.1. Standard optimisation algorithms such as Newton-Raphson can be used to compute the solution.

Table 3.1: Examples of ρ and ψ

	$\rho(x)$	$\psi(x)$
Least absolute deviations (L_1)	$\lvert x\rvert$	$\text{sign}(x)$
Least squares (L_2)	x^2	$2x$
Huber's: if $\lvert x\rvert \leq k$	$x^2/2$	x
if $\lvert x\rvert > k$	$k(\lvert x\rvert - k/2)$	$k\ \text{sign}(x)$
Cauchy	$k\log(1+(x/k))^2$	$\frac{2x}{(1+(x/k))}$
Tukey's: if $\lvert x\rvert \leq k$	$\frac{k}{6}(1-[1-(x/k)^2]^3)$	$x[1-(x/k)^2]^2$
if $\lvert x\rvert > k$	$k^2/6$	0

3.3 M-estimators for Heteroscedastic Models

Variant of M-estimation for the estimation of volatility parameters in GARCH model was first considered by Berkes and Horváth (2004). They define their estimators as follows:

Let $\{y_t, -\infty < k < \infty\}$ be a GARCH(1,1) process satisfying the equations

$$y_k = \sigma_k \varepsilon_k \tag{3.6}$$

and

$$\sigma_k^2 = \omega + \alpha y_{k-1}^2 + \beta \sigma_{k-1}^2. \tag{3.7}$$

Assume that

$$\omega > 0,\ \alpha \geq 0,\ \text{and}\ \beta \geq 0, \tag{3.8}$$

where $\boldsymbol{\theta} = (\omega, \alpha, \beta)$ is the parameter of the GARCH(1,1) process. Also assume that $\{\varepsilon_i, -\infty < i < \infty\}$ are i.i.d. random variables with $E(\varepsilon_0) = 0$ and $E(\varepsilon_0^2) = 1$.

If error density is f, then the log likelihood function is given by

$$\hat{L}_T(\mathbf{u}) = \frac{1}{T} \sum_{1<k\leq T} \log \left\{ \frac{1}{\hat{w}_k^{1/2}(\mathbf{u})} f(y_k / \hat{w}_k^{1/2}(\mathbf{u})) \right\},$$

and

$$\hat{w}_k(\mathbf{u}) = \frac{x}{1-t} + s \sum_{1\leq i\leq k-1} t^{i-1} y_{k-i}^2, \quad \mathbf{u} = (x, s, t),$$

where the definition of $\hat{w}_k(\mathbf{u})$ is motivated from the approximation of the variance function defined in (3.15) below. Hence for any error distribution, the quasi-likelihood estimator $\hat{\boldsymbol{\theta}}_T$ is defined as

$$\hat{\boldsymbol{\theta}}_T = \arg\max_{\mathbf{u}\in U} \hat{L}_T(\mathbf{u}),$$

where

$$U = \{\mathbf{u} = (x, s, t) : t \leq \rho_0 \text{ and } u_1 = \min(x, s, t) \leq \max(x, s, t) \leq u_2\},$$

with some $0 < u_1 < u_2$ and $0 < \rho_0 < 1$. Among other sets of conditions, Berkes and Horváth (2004) impose the following moment condition on ε_0:

$$E|\varepsilon_0^2|^\kappa < \infty \quad \text{with some} \quad \kappa > 0. \tag{3.9}$$

Also,

$$L(\mathbf{u}) = E\log\left\{\frac{1}{w_0^{1/2}(\mathbf{u})}f(y_0/w_0^{1/2}(\mathbf{u}))\right\}$$

exists for all $\mathbf{u} \in U$, where

$$w_k(\mathbf{u}) = \frac{x}{1-t} + s\sum_{1\le i<\infty} t^{i-1}y_{k-i}^2, \quad \mathbf{u} = (x, s, t)$$

and note that $w_k(\boldsymbol{\theta}) = \sigma_k^2$. The following condition implies that $L(\mathbf{u})$ has a unique maximum

$$Eg(\varepsilon_0, t) < Eg(\varepsilon_0, 1) \quad \text{for all} \quad 0 < t < \infty,\ t \neq 1, \tag{3.10}$$

where $g(y,t) = \log\{tf(yt)\}$. They showed that

$$\hat{\boldsymbol{\theta}}_T \to \boldsymbol{\theta} \quad a.s.$$

and

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}) \xrightarrow{\mathfrak{D}} \mathbf{N}(\mathbf{0}, 4\tau^2\mathbf{A}^{-1}),$$

where $\xrightarrow{\mathfrak{D}}$ denotes convergence in distribution,

$$0 < \tau^2 = \frac{E\dot{g}^2(\varepsilon_0, 1)}{(E\ddot{g}(\varepsilon_0, 1))^2} < \infty,$$

and

$$\mathbf{A} = E(\dot{w}_0(\boldsymbol{\theta})/w_0(\boldsymbol{\theta}))'(\dot{w}_0(\boldsymbol{\theta})/w_0(\boldsymbol{\theta})).$$

Note that $\dot{g}$ and $\ddot{g}$ represents the first and second derivative of g, respectively, whenever these exist and $\mathbf{B}'$ denotes the transpose of matrix $\mathbf{B}$.

The following three choices of f are considered by Berkes and Horváth (2004).

1. The standard normal density:

$$\begin{aligned} f(y) &= (2\pi)^{-1/2}\exp(-y^2/2) \\ \text{assuming} \qquad E\varepsilon_0^2 &= 1. \end{aligned} \tag{3.11}$$

2. Double exponential distribution:

$$\begin{aligned} f(y) &= (1/2)\exp(-|y|) \\ \text{assuming} \qquad E|\varepsilon_0| &= 1. \end{aligned} \tag{3.12}$$

3. The density of the following form:

$$\begin{aligned} f(y) &= \{(\nu-1)/2\}(1+|y|)^{-\nu},\ \nu>1 \\ \text{and assuming} \qquad E\{|\varepsilon_0|/(1+|\varepsilon_0|)\} &= 1/\nu \end{aligned} \tag{3.13}$$

Assumptions (3.11)-(3.13) are undesirable and impossible to verify as we do not know about the error distribution in advance. Therefore, other robust estimators need to be investigated that do not impose these restrictions. Also estimators with less moment conditions are very desirable especially when it is known that financial time series often have heavy tails.

3.3.1 M-estimators for Asymmetric GARCH Models

In this section we define M-estimators for asymmetric GARCH models. In particular, we consider the GJR model of Glosten et al. (1993) but M-estimators for other asymmetric models such as the TGARCH model of Zakoian (1994) can also be defined in a similar way. As mentioned earlier, the GARCH model is not a good choice for modeling when financial data has leverage effect. An asymmetric GARCH model such as GJR-GARCH model is designed to capture the asymmetric effect. Our aim is to define estimators for these GARCH-type models that require weaker moment assumption and also handle asymmetric data. Our second aim is to propose the use of other estimators as a desirable alternative to the QMLE and the least absolute deviations (LAD) estimators. Note that, hereafter the term 'QMLE' is used for the score function of type $H(x) = x^2$ defined below.

Mukherjee (2008) discussed M-estimators for GARCH(p, q) models. We start our discussion of M-estimators in asymmetric model with the simple GJR(1,1) model where the following representation of the return series $\{X_t; t \in \mathbb{Z}\}$ is assumed. Observe $\{X_t; 1 \leq t \leq T\}$ such that

$$\begin{aligned} X_t &= h_t^{1/2}\epsilon_t, \\ h_t &= \omega_0 + \alpha_0 X_{t-1}^2 + \beta_0 h_{t-1} + \gamma_0 D_{t-1} X_{t-1}^2, \end{aligned} \tag{3.14}$$

where $D_{t-1} = 1$ if $X_{t-1} < 0$ and 0 otherwise, with $\{\epsilon_t\}$ a sequence of independent and identically distributed (i.i.d.) unobservable real-valued random variables and the unknown parameter $\boldsymbol{\theta}_0 = [\omega_0, \alpha_0, \gamma_0, \beta_0]'$ is in the parameter space

$$\Theta = \{\boldsymbol{\theta} = [\omega, \alpha, \gamma, \beta]'; \omega > 0, \alpha, \beta, \gamma \geq 0, (\alpha + \beta + \frac{1}{2}\gamma) < 1\}.$$

Under these parameter constraints, model (3.14) is strictly stationary and hence covariance stationary under finite second moment of X_t. The GJR(1,1) model reduces to the GARCH(1,1) model when there is no leverage effect i.e., when $\gamma_0 = 0$.

By recursive substitution, we get

$$\begin{aligned}
h_t &= \omega_0 + \alpha_0 X_{t-1}^2 + \gamma_0 D_{t-1} X_{t-1}^2 + \beta_0\{\omega_0 + \alpha_0 X_{t-2}^2 + \gamma_0 D_{t-2} X_{t-2}^2 + \beta_0 h_{t-2}\} \\
&= \omega_0(1+\beta_0) + \alpha_0(X_{t-1}^2 + \beta_0 X_{t-2}^2) + \gamma_0(D_{t-1}X_{t-1}^2 + \beta_0 D_{t-2} X_{t-2}^2) \\
&\quad + \beta_0^2\{\omega_0 + \alpha_0 X_{t-3}^2 + \gamma_0 D_{t-3} X_{t-3}^2 + \beta_0 h_{t-3}\} \\
&= \omega_0(1+\beta_0+\beta_0^2) + \alpha_0(X_{t-1}^2 + \beta_0 X_{t-2}^2 + \beta_0^2 X_{t-3}^3) \\
&\quad + \gamma_0(D_{t-1}X_{t-1}^2 + \beta_0 D_{t-2}X_{t-2}^2 + \beta_0^2 D_{t-3} X_{t-3}^2) + \beta_0^3 h_{t-3} \\
&= \vdots \qquad\qquad \vdots \qquad\qquad\qquad \vdots \\
&= \frac{\omega_0}{(1-\beta_0)} + \alpha_0 \sum_{j=1}^{\infty} \beta_0^{j-1} X_{t-j}^2 + \gamma_0 \sum_{j=1}^{\infty} D_{t-j} \beta_0^{j-1} X_{t-j}^2.
\end{aligned}$$

For $\boldsymbol{\theta} \in \Theta$, define the variance function

$$v_t(\boldsymbol{\theta}) = \frac{\omega}{(1-\beta)} + \alpha \sum_{j=1}^{\infty} \beta^{j-1} X_{t-j}^2 + \gamma \sum_{j=1}^{\infty} D_{t-j} \beta^{j-1} X_{t-j}^2, \tag{3.15}$$

and note that $v_t(\boldsymbol{\theta}_0) = h_t$.

In (3.14), if f denotes the error density, then the conditional density of $\{X_t\}$ given past will be $v_t^{-1/2}(\boldsymbol{\theta}_0) f\{v_t^{-1/2}(\boldsymbol{\theta}_0) X_t\}$, $1 \le t \le T$. Now we can define a random quantity as a minimizer of the negative log-likelihood function

$$L_T(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^{T} \left[\left(\frac{1}{2}\right) \log v_t(\boldsymbol{\theta}) - \log f\{X_t / v_t^{1/2}(\boldsymbol{\theta})\} \right], \quad \boldsymbol{\theta} \in \Theta.$$

Then, its derivative is

$$\frac{\partial L_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{T} \left(\frac{1}{2}\right) \left[1 - H^*\{X_t / v_t^{1/2}(\boldsymbol{\theta})\}\right] \{\dot{v}_t(\boldsymbol{\theta}) / v_t(\boldsymbol{\theta})\},$$

where $H^*(x) := x\{-\dot{f}(x)/f(x)\}$.

More generally, we define a score function H as follows. Let $\psi : \mathbb{R} \to \mathbb{R}$

be a skew-symmetric function (that is $\psi(-x) = -\psi(x)$, $\forall x \in \mathbb{R} - \{0\}$) which is differentiable in all but finite number of points. Let $\mathcal{D} \subset \mathbb{R}$ denote the set of points where ψ is differentiable and let $\mathcal{D}^c$ denote its complement. Let $H(x) := x\psi(x)$, $x \in \mathbb{R}$. Note that $H(-x) = H(x)$, $\forall x$. We can then define $\boldsymbol{\theta}_T$ in the model (3.14) as a solution of the equation

$$\sum_{t=1}^{T} \left(\frac{1}{2}\right)\left\{1 - H\{X_t/v_t^{1/2}(\boldsymbol{\theta})\}\right\}\{\dot{v}_t(\boldsymbol{\theta})/v_t(\boldsymbol{\theta})\} = \mathbf{0}.$$

Since $\{X_t; t \leq 0\}$ are not observable, $\{v_t(\boldsymbol{\theta})\}$'s are non observable and hence $\boldsymbol{\theta}_T$'s are noncomputable. We define an observable approximation $\{\hat{v}_t(\boldsymbol{\theta}); t \geq 1\}$ to the variance functions $\{v_t(\boldsymbol{\theta}); t \geq 1\}$ as

$$\hat{v}_t(\boldsymbol{\theta}) = \frac{\omega}{(1-\beta)} + \left\{\alpha \sum_{j=1}^{t-1} \beta^{j-1} X_{t-j}^2 + \gamma \sum_{j=1}^{t-1} D_{t-j}\beta^{j-1} X_{t-j}^2\right\} I(t \geq 2). \tag{3.16}$$

Then an M-estimator $\hat{\boldsymbol{\theta}}_T$ based on the score function H or ψ is defined as a solution of the equation

$$\hat{\mathbf{M}}_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} \hat{m}_t(\boldsymbol{\theta}) = \mathbf{0}, \tag{3.17}$$

where

$$\hat{m}_t(\boldsymbol{\theta}) = \left(\frac{1}{2}\right)\left\{1 - H\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta})\}\right\}\{\dot{\hat{v}}_t(\boldsymbol{\theta})/\hat{v}_t(\boldsymbol{\theta})\},$$

and $\dot{\hat{v}}_t$ can be obtained by differentiating (3.16) w.r.t ω, α, γ and β, respectively, as

$$\begin{aligned}
\frac{\partial \hat{v}_t(\boldsymbol{\theta})}{\partial \omega} &= \frac{1}{1-\beta}, \\
\frac{\partial \hat{v}_t(\boldsymbol{\theta})}{\partial \alpha} &= \Big\{\sum_{j=1}^{t-1} \beta^{j-1} X_{t-j}^2\Big\} I(t \geq 2), \\
\frac{\partial \hat{v}_t(\boldsymbol{\theta})}{\partial \gamma} &= \Big\{\sum_{j=1}^{t-1} \beta^{j-1} D_{t-j} X_{t-j}^2\Big\} I(t \geq 2), \\
\frac{\partial \hat{v}_t(\boldsymbol{\theta})}{\partial \beta} &= \frac{\omega}{(1-\beta)^2} + \sum_{j=2}^{t-1} (j-1)\beta^{j-2}\Big\{\alpha + \gamma D_{t-j}\Big\} X_{t-j}^2 I(t \geq 2).
\end{aligned}$$

It turns out that $\hat{\boldsymbol{\theta}}_T$ actually estimates a function of the true parameter $\boldsymbol{\theta}_0$ defined by

$$\boldsymbol{\theta}_{0H} = [c_H\omega_0, c_H\alpha_0, c_H\gamma_0, \beta_0]', \tag{3.18}$$

for a constant $c_H > 0$ defined in (3.20) below, that depends on the underlying score function H. Note that β_0 is free from c_H.

Next, we discuss some examples of the score function H.

Example 1. *Least absolute deviations (LAD) score:*

Let $\psi(x) = \text{sign}(x)$. Then $\mathcal{D}^c = \{0\}$ and $H(x) = |x|$.

Example 2. *Huber's k-score:*

Let $\psi(x) = xI(|x| \leq k) + k\,\text{sign}(x)I(|x| > k)$, where $k > 0$ is known. Then $\mathcal{D}^c = \{-k, k\}$ and $H(x) = x^2 I(|x| \leq k) + k|x|I(|x| > k)$.

Example 3. *The QMLE:*

Let $\psi(x) = x$. Then $H(x) = x^2$.

Example 4. *The maximum likelihood estimation (MLE):*

Let $\psi(x) = -\dot{f}_0(x)/f_0(x)$, where f_0 is the true density of ϵ, assumed to be known. Then $H(x) = x\{-\dot{f}_0(x)/f_0(x)\}$.

Example 5. B-*estimator:*

Let $\psi(x) = B\,\text{sign}(x)/(1+|x|)$, where $B > 1$ is a user specified constant. Then $\mathcal{D}^c = \{0\}$ and $H(x) = B|x|/(1+|x|)$.

Example 6. *Cauchy-estimator:*

Let $\psi(x) = 2x/(1+x^2)$. Then $H(x) = 2x^2/(1+x^2)$.

Example 7. *The exponential pseudo-maximum likelihood estimation:*

Let $\psi(x) = a|x|^{b-1}\text{sign}(x)$, where $a > 0$ and $1 < b \leq 2$ are known constants. Such score can be motivated from the class of densities considered by Nelson (1991) and Robinson and Zaffaroni (2006) to model the innovations of the exponential GARCH model. Here $\mathcal{D}^c = \{0\}$ and $H(x) = a|x|^b$.

Example 8. *Score function for Hansen's skewed-t distribution:*

Let $\psi(x) = \frac{b(\nu+1)(bx+a)}{(\nu-2)(1-\lambda)\ +(bx+a)} I(x < -a/b) + \frac{b(\nu+1)(bx+a)}{(\nu-2)(1+\lambda)\ +(bx+a)} I(x \geq -a/b)$, see Section 2.5 for notations used in the definition of $\psi(x)$. Then $H(x) = \frac{bx(\nu+1)(bx+a)}{(\nu-2)(1-\lambda)\ +(bx+a)} I(x < -a/b) + \frac{bx(\nu+1)(bx+a)}{(\nu-2)(1+\lambda)\ +(bx+a)} I(x \geq -a/b)$.

Remark 3.1. Define a function ρ by $\rho(x) = \int_0^x \psi(t)dt$, for $x \geq 0$ and $\rho(x) = \rho(-x)$ for $x < 0$. Define

$$R_T(\boldsymbol{\theta}) := \sum_{t=1}^{n} \left[\left(\frac{1}{2}\right) \log \hat{v}_t(\boldsymbol{\theta}) + \rho\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta})\}\right], \ \boldsymbol{\theta} \in \Theta,$$

and notice that

$$\dot{R}_T(\boldsymbol{\theta}) = \sum_{t=1}^{n} \left(\frac{1}{2}\right)\left\{1 - H\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta})\}\right\}\{\dot{\hat{v}}_t(\boldsymbol{\theta})/\hat{v}_t(\boldsymbol{\theta})\}.$$

Hence $\hat{\boldsymbol{\theta}}_T$ from (3.17) can also be considered as an M-estimator that minimizes the criterion function R_T based on ρ. This helps to explain the intuition behind the above score functions. For Example 1, $\rho(x) = |x|$ and hence the corresponding estimator minimizes a variant of the sum of the absolute value of the residuals

$\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta})\}$ plus a penalty function arising out in the form of the logarithm of the approximating variance. For Example 3, $\rho(x) = x^2$ and hence the corresponding estimator minimizes a variant of the sum of the squared residuals plus a penalty function. Similar interpretations can be given for other examples.

As we mentioned earlier, an M-estimator based on a score function H consistently estimates $\boldsymbol{\theta}_{0H} = [c_H\omega_0, c_H\alpha_0, c_H\gamma_0, \beta_0]'$, where c_H is a constant that depends on the score function H through the error distribution. Using the QMLE and assuming $c_H = 1$, an M-estimator actually estimate $\boldsymbol{\theta}_0$.

Following the guidelines of Section (3.3.1), we can also define M-estimators for another asymmetric GARCH model, called the threshold GARCH (TGARCH) model.

3.3.2 Asymptotic Normality

To state the asymptotic distribution of the suitably normalized $\hat{\boldsymbol{\theta}}_T$, we make the following technical assumptions.

Model assumptions:

The parameter space Θ is a compact set and its interior Θ_0 contains both $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_{0H}$ of (3.18), respectively. Assume that for some $\kappa > 0$,

$$E[|\epsilon|^{\kappa}] < \infty. \tag{3.19}$$

Moreover, $\{X_t\}$ of (3.14) is stationary and ergodic.

Conditions on the score function:

(i) Identifiability condition:

Corresponding to the score function H, there exists a unique number $c_H > 0$ satisfying

$$E[H(\epsilon/c_H^{1/2})] = 1. \tag{3.20}$$

(ii) Moment conditions:

$$E[H(\epsilon/c_H^{1/2})]^2 < \infty \text{ and } 0 < E\{(\epsilon/c_H^{1/2})\dot{H}(\epsilon/c_H^{1/2})\} < \infty. \tag{3.21}$$

(iii) Smoothness conditions:

One can assume smoothness conditions of varying degree that are applicable to different score functions. One such (strong) assumption is that the score function is three times differentiable with bounded third derivative. It is possible to have weak smoothness conditions on $\mathcal{D}^c$ that are satisfied by all score functions of Examples 1-8.

Remark 3.2. Recall that if ϕ is a nondecreasing, odd score function satisfying some smoothness conditions, then there exists a point θ_ϕ such that $E[\phi(\varepsilon - \theta_\phi)] = 0$. Since a location estimation problem, say, $x_t = \theta + \varepsilon_t$ can be rewritten as $x_t = (\theta + \theta_\phi) + (\varepsilon_t - \theta_\phi)$, a location-invariant M-estimator based on ϕ estimates $\theta + \theta_\phi$. Condition (3.20) is a natural counterpart of the location model in the scale estimation problem. To illustrate, first note from (3.15) and (3.18) that

$$v_t(\boldsymbol{\theta}_{0H}) = c_H v_t(\boldsymbol{\theta}_0), \tag{3.22}$$

and hence (3.14) can be rewritten as

$$X_t = v_t^{1/2}(\boldsymbol{\theta}_{0H})(\epsilon_t/c_H^{1/2}). \tag{3.23}$$

Now suppose that the score function H in a scale estimation problem is nondecreasing on $[0, \infty)$ with $\lim_{c\to 0+} H(\epsilon/c^{1/2}) = H(\infty)$ and $\lim_{c\to\infty} H(\epsilon/c^{1/2}) = H(0)$ and suppose that the expectation function $g(c) = E\{H(\epsilon/c^{1/2})\}$ satisfying $H(0) \le g(c) \le H(\infty)$ has a range containing 1. Then, there exists a point $c_H > 0$ satisfying $E[H(\epsilon/c_H^{1/2})] = 1$ which is assumed unique in (3.20). Hence the scale invariant M-estimator based on H actually estimates $\boldsymbol{\theta}_{0H}$ by (3.23).

The following theorem states the asymptotic normality of $\hat{\boldsymbol{\theta}}_T$. Define the score function factor

$$\sigma^2(H) := 4 \operatorname{var}\{H(\epsilon/c_H^{1/2})\}/[E\{(\epsilon/c_H^{1/2})\dot{H}(\epsilon/c_H^{1/2})\}]^2,$$

where both the numerator and denominator are positive by (3.21). Also, define

$$\mathbf{G} = \mathbf{G}(\boldsymbol{\theta}_{0H}) := E\{\dot{v}_1(\boldsymbol{\theta}_{0H})\dot{v}_1'(\boldsymbol{\theta}_{0H})/v_1^2(\boldsymbol{\theta}_{0H})\}.$$

Theorem 3.1 *Under (3.20), (3.21) and mild smoothness conditions on the score function H*

$$T^{1/2}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_{0H}) \xrightarrow{\mathfrak{D}} \mathbf{N}(\mathbf{0}, \sigma^2(H)\mathbf{G}^{-1}). \tag{3.24}$$

Theorem 3.1 can be proved in a similar way as in Mukherjee (2008).

From (3.23) define residuals by

$$\hat{\epsilon}_{tH} = X_t/\{\hat{v}_t^{1/2}(\hat{\boldsymbol{\theta}}_T)\},\ 1 \le t \le T. \tag{3.25}$$

Since

$$\sigma^2(H) = 4[E\{H(\epsilon/c_H^{1/2})\}^2 - 1][E\{(\epsilon/c_H^{1/2})\dot{H}(\epsilon/c_H^{1/2})\}]^{-2},$$

it can be estimated based on residuals $\{\hat{\epsilon}_{tH}\}$ by

$$\begin{aligned}\widehat{\sigma^2(H)} &= 4\left[(1/T)\sum_{t=1}^{T}\{H(X_t/\hat{v}_t^{1/2}(\hat{\boldsymbol{\theta}}_T)\}^2 - [(1/T)\sum_{t=1}^{T}\{H(X_t/\hat{v}_t^{1/2}(\hat{\boldsymbol{\theta}}_T)\}]^2\right] \\ &\quad\times[(1/T)\sum_{t=1}^{T}\{(X_t/\hat{v}_t^{1/2}(\hat{\boldsymbol{\theta}}_T)\dot{H}(X_t/\hat{v}_t^{1/2}(\hat{\boldsymbol{\theta}}_T)\}]^{-2}.\end{aligned}$$

Also, $\mathbf{G}^{-1}$ can be estimated by $(\hat{\mathbf{G}})^{-1}$ where

$$\hat{\mathbf{G}} = \frac{1}{T}\sum_{t=1}^{T}\{\dot{\hat{v}}_t(\hat{\boldsymbol{\theta}}_T)\dot{\hat{v}}_t'(\hat{\boldsymbol{\theta}}_T)/\hat{v}_t^2(\hat{\boldsymbol{\theta}}_T)\}.$$

Using (3.24), $\widehat{\sigma^2(H)}$ and $(\hat{\mathbf{G}})^{-1}$, we can get confidence intervals of $\boldsymbol{\theta}_{0H}$.

Note that Theorem 3.1 is derived under weak moment assumptions on the error distribution. The identifiability condition determines what the M-estimator can consistently estimate.

Using Monte Carlo simulations and application to real data sets the parameters of the GARCH(1,1) and the GJR(1,1) models based on these score functions are estimated. Our results suggest that the performance of these estimators are better than the other commonly-used estimators especially in the case of heavy-tailed distributions.

3.3.3 Computation of M-estimators

Here we discuss the computation of M-estimators. Eq.(3.17) is a highly non linear equation and solving such equation is computationally very challenging. Although explicit algorithms for the QMLE are available in the literature and most statistical packages, but not for computing other score functions. Algorithms for estimating the parameters of GARCH and GJR models using M-estimators are developed.

As we showed, an M-estimator is a solution of

$$\hat{\mathbf{M}}_T(\boldsymbol{\theta}) = \sum_{t=1}^{T}\hat{m}_t(\boldsymbol{\theta}) = \mathbf{0},$$

where

$$\hat{m}_t(\boldsymbol{\theta}) = \left(\frac{1}{2}\right)\left[1 - H\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta})\}\right]\{\dot{\hat{v}}_t(\boldsymbol{\theta})/\hat{v}_t(\boldsymbol{\theta})\}.$$

To solve this, we use an iterative algorithm where $\hat{\boldsymbol{\theta}}_{(r+1)}$, the estimate at the $(r+1)$-th iteration, is computed by

$$\hat{\boldsymbol{\theta}}_{(r+1)} = \hat{\boldsymbol{\theta}}_{(r)} + \Big[-\mathbf{C}(\hat{\boldsymbol{\theta}}_{(r)}) \Big]^{-1} \hat{\mathbf{M}}_T(\hat{\boldsymbol{\theta}}_{(r)}), \;\; r = 0, 1, 2, \ldots$$

where $\mathbf{C}(\hat{\boldsymbol{\theta}}_{(r)})$ is an invertible matrix. The choice of $\mathbf{C}(\hat{\boldsymbol{\theta}}_{(r)})$ and the starting value of $\hat{\boldsymbol{\theta}}_{(0)}$ are crucial for the convergence of the above algorithm. In this paper, we propose the following modification of the popular BHHH type algorithm of Econometrics involving a sum of the modified products of gradients as

$$\mathbf{C}(\hat{\boldsymbol{\theta}}_{(r)}) = \frac{1}{T} \sum_{t=1}^{T} \{ \dot{\hat{v}}_t(\hat{\boldsymbol{\theta}}_{(r)}) \dot{\hat{v}}_t'(\hat{\boldsymbol{\theta}}_{(r)}) / \{\hat{v}_t(\hat{\boldsymbol{\theta}}_{(r)})\}^2 \},$$

and with this choice excellent convergence was observed. Our default choice of the initial estimator $\hat{\boldsymbol{\theta}}_{(0)}$ is guided by the rule followed by the software MATLAB for the computation of the QMLE. In particular, we chose $\hat{\alpha}_{(0)} = 0.05$, $\hat{\beta}_{(0)} = 0.85$, $\hat{\gamma}_{(0)} = 0$. For $\hat{\omega}_{(0)}$, we use $(1 - \hat{\alpha}_{(0)} - \hat{\beta}_{(0)} \times \hat{v}(X)) = (1 - 0.05 - 0.85)\, \hat{v}(X)$ where $\hat{v}(X) = \sum (X_i - \bar{X})^2/(T-1)$, the sample variance of the observed series $\{X_1, \ldots, X_T\}$. We use this choice not only for initiating the computation of the QMLE but also for initiating the computation based on any other score functions even though the resulting estimator is consistent for $\boldsymbol{\theta}_{0H}$ which varies with H.

3.4 Simulation and Empirical Results

In this section, first we investigate the relative performance based on the mean squared error (MSE) of each estimator in GARCH and GJR models through extensive Monte Carlo simulations assuming various distributions for errors. Then the finite sample properties of M-estimators are assessed when applied to two important financial data sets, IBM stock and S&P500 Index. All computations are performed using MATLAB software.

3.4.1 Monte Carlo Simulations

To compare relative performance of estimators based on different score functions, we define the mean squared errors (MSE) of an estimator for the GJR (1, 1) model as

$$E[\{(\hat{\omega}+\hat{\gamma})/\hat{\alpha}+\hat{\beta}\}-\{(\omega_0+\gamma_0)/\alpha_0+\beta_0)\}]^2.$$

From (3.18) and (3.24), the ratio of $(\hat{\omega}+\hat{\gamma})$ and $\hat{\alpha}$ is consistent to a quantity that is free from the underlying score function H used for the M-estimation and hence the above definition of MSE compares the relative performance of different M-estimators. When specialized to the GARCH (1, 1) model with $\gamma_0 = 0$, the corresponding MSE is defined as

$$E[\{(\hat{\omega}/\hat{\alpha})+\hat{\beta}\}-\{(\omega_0/\alpha_0)+\beta_0)\}]^2.$$

We use simulations to estimate these quantities corresponding to five different M-estimators (QMLE, LAD, Huber with $k = 1.5 \times 1.483\,\text{median}|\hat{\epsilon}_t|$, B-estimator with $B = 2.5$ and Cauchy). Our Monte Carlo simulations are based on K replicates each of sample size T. For both the GARCH(1,1) and the GJR(1,1) model, errors are generated from (i) the standard normal distribution (ii) contaminated normal distribution $(1-\epsilon)\Phi(x)+\epsilon\Phi(x/\sigma)$ with $\epsilon = 0.05$, and $\sigma^2 = 9$, Hansen's skewed-t distribution with $\nu = 5$ and $\lambda = 0.25$, and standardized student-t distributions with 3 and 4 degrees of freedom. We generate $500 + T$ observations in each replication and discard the first 500 observations to reduce the impact of initial values.

Estimating the GARCH model

In the first study we generate $K = 1000$ independent replicates each of sample size $T = 500$ from the GARCH(1,1) model assuming that errors come from the standard normal distribution, contaminated normal distribution and student-t distribution with 3 and 4 degrees of freedom. The values of true parameters considered are $\boldsymbol{\theta}_0 = (0.005, 0.2, 0.75)$. Errors generated from t-distributions are standardised to ensure that the first two moments are 0 and 1 respectively.

Table 3.2 shows the average MSE over 1000 replications for each score function with their standard deviations in parenthesis. Entries in bold are the smallest values among all estimators representing the best performing estimator in each case in terms of average MSE. As expected the QMLE performs well when normality is assumed for error distribution. But it can be seen that the QMLE is not a good choice when error density follows student-t distributions. Peng and Yao (2003) suggested that when $\{\varepsilon_t\}$ follows heavy-tailed distribution, for example $t(3)$, least absolute deviations (LAD) estimators should be used. But in these cases we observe that there are score function such as Huber's and B-estimator that can perform even better than LAD when $\{\varepsilon_t\} \sim t(3)$. It is worth mentioning here that we get these results without imposing extra restrictions such as median(ε^2)=1 and squaring the GARCH model which requires the condition of higher moments.

In the second study, 1000 replicates each of sample size 1000 are generated from the GARCH(1,1) model. Errors are generated from the same distributions as in previous experiment. The true values of parameters are set to $\boldsymbol{\theta}_0 = (0.05, 0.1, 0.85)$. Table 3.3 shows the MSE for each score function. Again the MSE for the QMLE is found the least in case of normal density but when errors are generated from the contaminated normal distribution and $t(3)$, B-estimators provides the best estimates with Huber's score also competing well. For students-t distribution with

Table 3.2: Mean Squared Error of the GARCH(1,1) model using M-estimators. (sample size=500)

$\boldsymbol{\theta}$ = (0.005, 0.2, 0.75)	QMLE	LAD	Huber's	B-estimator	Cauchy
$T = 500$			*Standard Normal Distribution*		
MSE	**0.0202**	0.0677	0.0272	0.0441	0.0812
	(0.0631)	(0.1468)	(0.0758)	(0.1094)	(0.1428)
$T = 500$			*Contaminated Normal Distribution*		
MSE	0.0720	0.0440	0.0434	**0.0408**	0.0909
	(0.1106)	(0.0883)	(0.0879)	(0.0839)	(0.1450)
$T = 500$			*Student-t distribution (3)*		
MSE	0.0302	0.0163	**0.0119**	0.0133	0.0204
	(0.0745)	(0.0595)	(0.0392)	(0.0488)	(0.0604)
$T = 500$			*Student-t distribution (4)*		
MSE	0.0241	0.0153	**0.0145**	0.0153	0.0351
	(0.0535)	(0.0448)	(0.0430)	(0.0398)	(0.0925)

4 df, LAD estimate shows good result. These results indicate that B-estimators can be considered for estimation of GARCH models when data has heavy tails or there is evidence of outliers.

In order to check the effect of both asymmetric and non-normal errors on the estimated parameters, we perform another experiment. For this experiment we generate random draws from the skewed-t distribution proposed by Hansen (1994) with skewness parameter $\lambda = 0.25$ and degrees of freedom $\nu = 3$ and 5. Errors are also generated from the standard normal, contaminated normal and student-t distribution with 3 df. In this experiment the values of true parameters are set as $\boldsymbol{\theta}_0 = (0.005, 0.2, 0.3)$. We choose this set of parameter values to study the behaviour of different estimators under small value for the coefficient of the GARCH term. A sample of size $T = 1000$ is chosen and parameters of the GARCH(1,1) model are estimated by M-estimators. The results of MSE and their standard errors based on $K = 1000$ replications for each estimator are reported in Table 3.4.

Table 3.3: Mean Squared Error of the GARCH(1,1) model using M-estimators. (sample size=1000)

$\boldsymbol{\theta}$ = (0.05, 0.1, 0.85)	QMLE	LAD	Huber's	B-estimator	Cauchy
$T = 1000$			*Standard Normal Distribution*		
MSE	**0.0238**	0.0414	0.0257	0.0410	0.0697
	(0.0439)	(0.0896)	(0.0441)	(0.0772)	(0.1198)
$T = 1000$			*Contaminated Normal Distribution*		
MSE	0.0223	0.0052	0.0048	**0.0043**	0.0059
	(0.0309)	(0.0082)	(0.0074)	(0.0081)	(0.0092)
$T = 1000$			*Student-t distribution (3)*		
MSE	0.0343	0.0317	0.0226	**0.0225**	0.3230
	(0.1075)	(0.1207)	(0.1088)	(0.0945)	(0.1109)
$T = 1000$			*Student-t distribution (4)*		
MSE	0.0089	**0.0033**	0.0090	0.0061	0.0124
	(0.0405)	(0.0150)	(0.0648)	(0.0417)	(0.0670)

It can be noticed from the results of Table 3.4 that although the QMLE performs better than other estimators in the case of the standard normal assumption for errors, it is not a good choice for any other cases. In the case of the contaminated normal distribution, B-estimator produces the best result as it did in previous studies for this type of distribution with Cauchy estimator also performing well. LAD and Huber's estimators outperform other estimators when errors were generated from a skewed-t distribution. Huber's estimator shows very good results for $\{\epsilon_t\} \sim t(3)$. We set a small value for the GARCH coefficient and results in Table 3.4 show that the performance of these estimators remain consistent. These results suggest that there are other estimators such as Huber's and B-estimator that can perform better than the QMLE and LAD when there is evidence of asymmetry, heavy-taildness and outliers in the data.

Table 3.4: Mean Squared Error of the GARCH(1,1) model using M-estimators. (sample size=1000)

θ = (0.005, 0.2, 0.3)	QMLE	LAD	Huber's	B-estimator	Cauchy
$T = 1000$			*Standard Normal Distribution*		
MSE	**0.0207**	0.0279	0.0235	0.0370	0.0575
	(0.0380)	(0.0516)	(0.0432)	(0.0646)	(0.0904)
$T = 1000$			*Contaminated Normal Distribution*		
MSE	0.1234	0.0571	0.0515	**0.0397**	0.0490
	(0.1730)	(0.0828)	(0.0780)	(0.0674)	(0.0726)
$T = 1000$			*Student-t distribution (3)*		
MSE	0.0711	0.0379	**0.0337**	0.0354	0.0575
	(0.1026)	(0.0620)	(0.0538)	(0.0608)	(0.0849)
$T = 1000$			*Skewed-t distribution (3,0.25)*		
MSE	0.0756	0.0414	0.0361	**0.0355**	0.0546
	(0.1046)	(0.0703)	(0.0620)	(0.0614)	(0.0835)
$T = 1000$			*Skewed-t distribution (5,0.25)*		
MSE	0.0457	0.0312	**0.0268**	0.0304	0.0492
	(0.0740)	(0.0558)	(0.0468)	(0.0533)	(0.0761)

Estimating the GJR model

Next we consider estimating the parameters of the GJR(1,1) model using M-estimators. First, we generate 1000 replicates each of sample size 500 from the GJR(1,1) model. The four choices for error distribution are same as in the case of the GARCH(1,1) model, i.e. the standard normal distribution, contaminated normal distribution with $\epsilon = 0.05$, $\sigma^2 = 9$ and standardised student-t distribution with 3 and 4 degrees of freedom. The true parameters values are $\omega_0 = 1.0, \alpha_0 = 0.1$, $\gamma_0 = 0.1$ and $\beta_0 = 0.5$. Next, 1000 independent replicates each of sample size 1000 are generated from the GJR(1,1) model with same four choices for error distributions. The true parameters values in this case are $\omega_0 = 0.5, \alpha_0 = 0.3$, $\gamma_0 = 0.25$ and $\beta_0 = 0.4$.

Table 3.5 below shows the average MSE over 1000 replications for each score functions with their standard errors in parentheses. The results show that the QMLE is again not a good choice in the GJR(1,1) model when errors are generated from heavy-tailed and contaminated distributions. In the presence of outliers, the QMLE performs badly. B-estimator outperforms other estimators in all cases except for normal assumption. Cauchy score function may be considered as the second best.

Table 3.5: Mean Squared Error of the GJR(1,1) model using M-estimators. (sample size=500)

$\boldsymbol{\theta}$ = (0.5, 0.3, 0.25, 0.4)	QMLE	LAD	Huber's	B-estimator	Cauchy
$T = 500$			*Standard Normal Distribution*		
MSE	**0.0227**	0.0304	0.0303	0.0373	0.0371
	(0.0155)	(0.0178)	(0.0187)	(0.0337)	(0.0268)
$T = 500$			*Contaminated Normal*		
MSE	0.1081	0.0601	0.0593	**0.0454**	0.0566
	(0.0925)	(0.0782)	(0.0811)	(0.0175)	(0.0177)
$T = 500$			*Student-t distribution (3)*		
MSE	0.0786	0.0400	0.0503	**0.0294**	0.0322
	(0.0625)	(0.0378)	(0.0853)	(0.0154)	(0.0160)
$T = 500$			*Student-t distribution (4)*		
MSE	0.0815	0.0598	0.0600	**0.0547**	0.0597
	(0.0430)	(0.0315)	(0.0132)	(0.0316)	(0.0727)

Table 3.6 states the same results for large sample size $T = 1000$ and again B-estimator seems to be the best choice among all competing estimators. From our simulation studies we conclude that for GJR models estimator such as B-estimator performs better than the QMLE when errors are contaminated with outliers and even outperform the robust estimate such as LAD when the tail of the error distribution is heavier than the normal. The performance of this estimator is also good when data comes from asymmetric distribution. B-estimator and Huber's estimator show good results in both the GARCH(1,1) and the GJR(1,1) cases for both small and large sample sizes.

Table 3.6: Mean Squared Error of the GJR(1,1) model using M-estimators. (sample size=1000)

$\boldsymbol{\theta}$ = (0.05, 0.3, 0.25, 0.4)	QMLE	LAD	Huber's	B-estimator	Cauchy
$T = 1000$			*Standard Normal Distribution*		
MSE	**0.0554**	0.0564	0.0564	0.0581	0.0590
	(0.0199)	(0.0189)	(0.0266)	(0.0215)	(0.0221)
$T = 1000$			*Contaminated Normal*		
MSE	0.0769	0.0641	0.0513	**0.0511**	0.0513
	(0.0955)	(0.0381)	(0.2532)	(0.0214)	(0.0222)
$T = 1000$			*Student-t distribution (3)*		
MSE	0.0805	0.0601	0.0597	**0.0574**	0.0575
	(0.0365)	(0.0275)	(0.0276)	(0.0254)	(0.0258)
$T = 1000$			*Student-t distribution (4)*		
MSE	0.0624	0.0562	0.0566	**0.0551**	0.0553
	(0.0302)	(0.0234)	(0.0231)	(0.0234)	(0.0241)

3.4.2 Empirical Illustration

In this section, we fit the GARCH(1,1) and the GJR(1,1) models using M-estimators to real data sets. The results of M-estimates with the QMLE estimated by MATLAB are also compared.

Tsay (2005, Chapter 3) analyzed two important data sets, namely, (A) The monthly log-returns of IBM stock from 1926 to 1999 (888 observations with first value 1.0434 and last value 4.5633) and (B) The monthly excess returns of S&P 500 from 1926 to 1991 (792 observations with first value 0.0225 and last value 0.1116) and fitted various types of conditional heteroscedastic models to them. These data, denoted by $\{y_t; 1 \leq t \leq T\}$ can be found in

$$http://faculty.chicagogsb.edu/ruey.tsay/teaching/fts2/.$$

We compute M-estimates of the parameters by fitting GARCH (1, 1) and GJR (1, 1) models for the centered IBM stock and the centered S&P 500 index, denoted

Table 3.7: Descriptive statistics for Data Sets.

Statistics	IBM Stock	S&P 500 Index
Sample size	888	792
Mean	1.2042	0.0061
Median	1.2230	0.0087
Minimum	-30.3676	-0.2994
Maximum	30.0971	0.4220
Std. Dev.	6.7287	0.0585
Skewness	-0.2369	0.4113
Kurtosis	4.9278	12.3002
Jarque-Bera	145.81	2876.70
$Q^2(10)$	2128.03	387.47

Q (10) is the Ljung-Box statistic at lag 10 of the squared log-returns.

by $\{X_t = y_t - \bar{y}; 1 \leq t \leq T\}$. Table 3.7 shows the summary statistics. Both data sets are asymmetric and having kurtosis greater than that of normal distribution. Jarque-Bera, a formal test statistic for testing whether the returns are normally distributed is also calculated and high values for both data confirm the nonnormality in data. The high values for Ljung-Box, $Q^2(10)$, statistics for the squared returns up to lag 10 indicate dependence in squared returns (ARCH effect).

The results of using MATLAB's GARCH Toolbox, for estimation of the parameters of the GARCH(1,1) and the GJR(1,1) models, are reported in Table 3.8 below. We need these results to check the consistency and accuracy of our algorithm for M-estimation.

First, we estimate the parameters of the GARCH(1,1) model for the IBM data. Table 3.9 below shows estimated parameters of the GARCH(1,1) model and their standard errors (SE's) in parentheses using five different M-estimators. The Ljung-Box statistics for the squared standardised shocks $\{\hat{\epsilon_t}^2\}$ are also computed to check the adequacy of the volatility equation. High p-values of Ljung-Box statistics for lag 10 suggest that the GARCH(1,1) model is adequate for the data at 5% significance level. As mentioned earlier an M-estimator based on a score

Table 3.8: Estimated parameters for the GARCH(1,1) and the GJR(1,1) with SE's using MATLAB

Data set	IBM Stock		S&P 500 Index	
	GARCH(1,1)	GJR(1,1)	GARCH(1,1)	GJR(1,1)
ω	2.9987	3.3579	0.00008	0.00009
	(0.9415)	(0.9810)	(0.00002)	(0.00002)
α	0.0953	0.0667	0.1211	0.0727
	(0.0201)	(0.0238)	(0.0199)	(0.0210)
γ	-	0.0558	-	0.0822
	-	(0.0256)	-	(0.0283)
β	0.8376	0.8293	0.8556	0.8543
	(0.0365)	(0.0380)	(0.0190)	(0.0185)

function H consistently estimates $\boldsymbol{\theta}_{0H} = (c_H\omega_0, c_H\alpha_0, \beta_0)'$. The QMLE estimates are approximately same as those computed by MATLAB as in this case $c_H = 1$. Since β_0 does not depend on the constant c_H, all M-estimates should give approximately the same estimate for this parameter and from Table 3.9 it can also be seen that the estimates of β_0 for all estimators are close to each other.

Table 3.9: Estimated parameters for the GARCH(1,1) with SE's and Ljung-Box statistic for $\hat{\epsilon}_t^2$ (IBM Stock)

Parameters	QMLE	LAD	Huber's	B-estimator	Cauchy
$c_H\omega$	3.0045	1.6319	1.9419	2.0021	0.8984
	(1.4277)	(0.7314)	(0.8795)	(1.0151)	(0.4722)
$c_H\alpha$	0.0950	0.0542	0.0680	0.0717	0.0297
	(0.0307)	(0.0162)	(0.0201)	(0.0236)	(0.0105)
β	0.8378	0.8475	0.8557	0.8502	0.8473
	(0.0535)	(0.0465)	(0.0435)	(0.0502)	(0.0547)
$Q^2(10)$	2.8528	3.0512	3.2429	3.1591	3.0479
p-value	0.9847	0.9802	0.9751	0.9774	0.9803

Q (10) is the Ljung-Box statistic at lag 10 of the standardised squared residuals.

Second, the parameters of the GJR(1,1) model is estimated for the IBM data. M-estimators are used for estimation and results are reported in Table 3.10. Standard errors for these estimated parameters are reported in parentheses. High p-values of Ljung-Box statistics for lag 10 of the standardised squared residuals suggest that the GJR(1,1) model is also adequate for this data set. M-estimators for the GJR(1,1) based on a score function H consistently estimates $\boldsymbol{\theta}_{0H} = (c_H\omega_0, c_H\alpha_0, c_H\gamma_0, \beta_0)'$. Again it can be seen that the estimates of β_0 for different estimators are 0.83, 0.84, 0.84, 0.84 and 0.84, respectively. These values are close to each other as expected and also close to the estimated value of β_0 in Table 3.8. The estimates of the QMLE are approximately same as those reported in Table 3.8. These findings confirm that M-estimators estimate the unknown parameters of the model correctly.

Table 3.10: Estimated parameters for the GJR(1,1) with SE's and Ljung-Box statistic for $\hat{\epsilon}_t^2$ (IBM Stock)

Parameters	QMLE	LAD	Huber's	B-estimator	Cauchy
$c_H\omega$	3.4542	1.7702	2.2448	2.2262	0.9251
	(1.5490)	(0.7512)	(0.3227)	(1.0468)	(0.4538)
$c_H\alpha$	0.0676	0.0377	0.0471	0.0490	0.0187
	(0.0333)	(0.0173)	(0.0074)	(0.0249)	(0.0105)
$c_H\gamma$	0.0570	0.0373	0.0489	0.0552	0.0255
	(0.0429)	(0.0232)	(0.0100)	(0.0346)	(0.0153)
β	0.8257	0.8383	0.8431	0.8381	0.8412
	(0.0569)	(0.0477)	(0.0156)	(0.0514)	(0.0528)
$Q^2(10)$	2.8068	3.0582	3.1182	3.2097	3.2548
p-value	0.9856	0.9800	0.9785	0.9761	0.9748

Q (10) is the Ljung-Box statistic at lag 10 of the standardised squared residuals.

Next, we consider estimating the parameters of the GARCH(1,1) model for the S&P 500 index. The results of estimated parameters of the GARCH(1,1) model using five different M-estimators and their standard errors are displayed in Table 3.11. Again the estimates of β, for all estimators are approximately same whereas the QMLE estimates are nearly equal to MATLAB's estimates which shows that M-estimators correctly estimate the parameters.

Table 3.11: Estimated parameters for the GARCH(1,1) with SE's and Ljung-Box statistic for $\hat{\epsilon}_t^2$ (S&P 500 Index)

Parameters	QMLE	LAD	Huber's	B-estimator	Cauchy
$c_H\omega$ (x10^{-3})	0.0737	0.0651	0.0859	0.0951	0.0444
	(0.0327)	(0.0245)	(0.0325)	(0.0392)	(0.0192)
$c_H\alpha$	0.1201	0.0616	0.0814	0.0676	0.0280
	(0.0279)	(0.0166)	(0.0221)	(0.0223)	(0.0102)
β	0.8590	0.8545	0.8549	0.8587	0.8575
	(0.0280)	(0.0334)	(0.0333)	(0.0400)	(0.0438)
$Q^2(10)$	10.1338	11.6089	11.6036	13.9245	15.1043
p-value	0.4288	0.3121	0.3125	0.1765	0.1283

Q (10) is the Ljung-Box statistic at lag 10 of the standardized squared residuals.

The GJR(1,1) model is also fitted to S&P 500 index. Table 3.12 reports the estimated parameters and standard errors. After the application of M-estimator to both real data sets and comparing the results with MATLAB estimated parameters we can confidently use M-estimators for the estimation of the parameters of GARCH and GJR models.

3.5 Value-at-Risk

Next we consider prediction of Value-at-Risk (VaR) based on M-estimates. VaR is the p-th conditional quantile of the distribution of the change in value of an asset over a certain period of time where p is known and close to zero. It is an estimate of the maximal loss associated with a given probability p and is used by

Table 3.12: Estimated parameters for the GJR(1,1) with SE's and Ljung-Box statistic for $\hat{\epsilon}_t^2$ (S&P 500 Index)

Parameters	QMLE	LAD	Huber's	B-estimator	Cauchy
$c_H\omega$(x10^{-3})	0.0882	0.0788	0.1050	0.1070	0.0465
	(0.0400)	(0.0277)	(0.0373)	(0.0416)	(0.0190)
$c_H\alpha$	0.0732	0.0232	0.0318	0.0186	0.0063
	(0.0374)	(0.0176)	(0.0236)	(0.0227)	(0.0099)
$c_H\gamma$	0.0786	0.0710	0.0911	0.1002	0.0449
	(0.0468)	(0.0260)	(0.0346)	(0.0372)	(0.0170)
β	0.8581	0.8491	0.8486	0.8526	0.8543
	(0.0341)	(0.0370)	(0.0377)	(0.0424)	(0.0442)
$Q(10)$	9.3685	9.3373	9.4392	9.3887	9.3030
p-value	0.4975	0.5004	0.4910	0.4957	0.5036

financial institutions and regulators for risk measurement. Existing approaches for predicting this are historical simulation, extreme value theory and quantile regression, among others.

For the returns of a portfolio $\{X_t; 1 \leq t \leq T\}$, the VaR $q_t = q_t(p)$ at time $t > 1$ is defined by

$$q_t = \inf\{x; p \leq P_{t-1}(X_t \leq x)\},$$

where P_{t-1} is the conditional distribution of X_t given the information available upto time $t-1$. When returns are of the form (3.14), we get

$$q_t = v_t^{1/2}(\boldsymbol{\theta}_0)F^{-1}(p),$$

where F^{-1} is the quantile function of the errors $\{\epsilon_t\}$. Using (3.22),

$$q_t = \frac{1}{c_H^{1/2}} v_t^{1/2}(\boldsymbol{\theta}_{0H})F^{-1}(p), \tag{3.26}$$

where notice that $F^{-1}(p)$ is the p-th quantile of the scaled errors $\{\epsilon_t/c_H^{1/2}\}$. Estimating $v_t^{1/2}(\boldsymbol{\theta}_{0H})$ by $\hat{v}_t^{1/2}(\hat{\boldsymbol{\theta}}_T)$ and $F^{-1}(p)$ by the p-th quantile of the residuals $\{X_t/\{\hat{v}_t(\hat{\boldsymbol{\theta}}_T)\}^{1/2}; 1 \leq t \leq T\}$, we obtain from (3.26) the predicted value $\hat{q}_t$ of q_t. Clearly $\hat{q}_t$ depends on the underlying M-estimates.

The statistic $T_* = \sum_{t=1}^{T} I(X_t \leq \hat{q}_t)$ denoting the number of violations, can be used to assess the overall predictive performance of the underlying conditional heteroscedastic model and the M-estimates used for computing $\hat{q}_t$. First the number of violations is assessed by the unconditional likelihood ratio test statistic, proposed by Kupiec (1995) when the QMLE is used as $\hat{\boldsymbol{\theta}}_T$, by

$$\mathrm{LR}_{uc} = 2\Big[\ln\left\{(1-\hat{p})^{T-T}\,\hat{p}^{T}\,\right\} - \ln\left\{(1-p)^{T-T}\,p^{T}\,\right\}\Big],\ \ \hat{p} = T_*/T.$$

This is asymptotically $\chi^2_{(1)}$.

Note that in a reasonable model of VaR, the previous history of violations should not convey any information about whether or not additional VaR violations may occur in future. Towards that, using the QMLE as $\hat{\boldsymbol{\theta}}_T$, Christoffersen (1998) defined the independence coverage test statistic, denoted by LR_{ind}, which characterizes the ways in which these violations occur as follows. For $1 \leq t \leq T$, let $I_t = I(X_t \leq \hat{q}_t)$. For $i, j = 0, 1$, let T_{ij} be the number of time points $\{t; 2 \leq t \leq T\}$ for which $I_t = i$ is followed by $I_{t+1} = j$. Let $\hat{\pi}_{ij} = T_{ij}/(T_{i0}+T_{i1})$, $\hat{\pi} = (T_{01}+T_{11})/T$. Then

$$\mathrm{LR}_{ind} = 2\Big[\ln\left((1-\hat{\pi}_{01})^{T}\ \hat{\pi}_{01}^{T}\ (1-\hat{\pi}_{11})^{T}\ \hat{\pi}_{11}^{T}\ \right) - \ln\left((1-\hat{\pi})^{(T\ \ +T\ \ )}\hat{\pi}^{(T\ \ +T\ \ )}\right)\Big].$$

It is important to recognize that both the unconditional coverage and the independence properties should be satisfied for an accurate VaR model. Hence Christof-

Table 3.13: Conditional likelihood statistics of IBM data based on M-estimates for the GARCH (1, 1) and GJR (1, 1) models

	QMLE	LAD	Huber's	B-estimator	Cauchy
$p = 10\%$			*GARCH(1,1)*		
T_*	89	89	89	89	89
LR_{cc}	7.5242*	7.5242*	7.5242*	5.8783	5.8783
$p = 10\%$			*GJR(1,1)*		
T_*	89	89	89	89	89
LR_{cc}	3.1422	3.1422	2.0684	2.0684	2.0684

* shows significant at 5%.

fersen (1998) proposed the statistic

$$\text{LR}_{cc} = \text{LR}_{uc} + \text{LR}_{ind}$$

which is asymptotically $\chi^2_{(2)}$.

In this study, we have computed $\{\hat{q}_t\}$ based on different M-estimates and Table 3.13 presents the number of violations T_* at coverage probability $p = 10\%$ and the corresponding LR_{cc} for IBM data fitted with both the GARCH (1, 1) and GJR (1, 1) models. Note that in this case, the expected number of violations is $10\% \times 888 \approx 89$ and this is same as the observed number of violations T_* for both models based on all five M-estimates. Using $\chi^2_{0.05,2} = 5.991$, Table 3.13 shows that the LR_{cc} based on the QMLE, LAD and Huber's estimates are statistically significant whereas those based on the Cauchy and B-estimator are not significant at 5% level for fitting GARCH (1, 1) model. On the other hand GJR (1, 1) model turns out to be not significant based on all M-estimates. Thus we conclude that M-estimators based on the Cauchy and B-estimator also provide good predictions of VaR especially when the GARCH (1, 1) model is fitted with this data.

3.6 A Weighted Resampling for M-estimators

Efron (1979) introduced the idea of bootstrapping which is a general approach to statistical inference based on building a sampling distribution for a statistic by resampling from the data at hand. For a comprehensive introduction of bootstrap methods, see Efron and Tibshirani (1993).

Chatterjee and Bose (2005) introduced a bootstrap technique for estimators obtained by solving estimating equations. They call it generalized bootstrap (GBS) because classical bootstrap, the delete-d jackknife and variations of the Bayesian bootstrap are shown to be some special cases of GBS. Examples of GBS weights and their implementation in heteroscedastic time series, generalized linear models and nonlinear regression models are also discussed.

Using the idea of weighted resampling we develop suitable bootstrap versions for M-estimators. We bootstrap M-estimators and our goal is to approximate the sampling distribution of the parameters with this new approach to resampling.

Recall that for a GARCH(1,1) model, the M-estimator is a solution of

$$\begin{aligned} \hat{\mathbf{M}}_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T} \hat{m}_t(\boldsymbol{\theta}) = \mathbf{0}, \\ \text{where} \quad \hat{m}_t(\boldsymbol{\theta}) &= \left(\frac{1}{2}\right)\Big\{1 - H\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta})\}\Big\}\{\dot{\hat{v}}_t(\boldsymbol{\theta})/\hat{v}_t(\boldsymbol{\theta})\}, \end{aligned} \tag{3.27}$$

with $\boldsymbol{\theta} = (\omega, \alpha, \beta)'$.

In order to estimate the sampling distribution and the asymptotic variance of $\hat{\boldsymbol{\theta}}_T$, define resampling estimator $\hat{\boldsymbol{\theta}}_T^*$ as the solution of

$$\hat{\mathbf{M}}_T^*(\boldsymbol{\theta}) = \sum_{t=1}^{T} w_{Tt}\hat{m}_t(\boldsymbol{\theta}) = \mathbf{0}, \tag{3.28}$$

where $\{w_{Tt}; 1 \leq t \leq T, T \geq 1\}$ is a triangular array of r.v.'s such that for each

$T \geq 1, \{w_{Tt}; 1 \leq t \leq T\}$ are exchangeable, independent of $\{X_t\}$. These are called the bootstrap weights. Minor modifications of the algorithm for computing $\hat{\boldsymbol{\theta}}_T$ are used to compute $\hat{\boldsymbol{\theta}}^*_T$.

We assume the following basic conditions (Conditions BW of Chatterjee and Bose (2005)) where $\sigma_t^2 = V_B(w_{Tt})$ and $k_1 > 0$ is a constant. The conditions on weights are as under:

$$\begin{aligned} E_B(w_{T1}) &= 1, \quad 0 < k_1 < \sigma_T^2 = o(T), \\ \text{and} \quad \text{corr}_B(w_{T1}, w_{T2}) &= O(T^{-1}). \end{aligned} \tag{3.29}$$

Let $\{W_{Tt} := (w_{Tt} - 1)/\sigma_t\}$ be the standardized weights satisfying

$$E_B(W_{T1}^4 < \infty), \quad \text{and} \quad \lim_{T\to\infty} E_B(W_{T1}^2 W_{T2}^2) = 1. \tag{3.30}$$

We are interested in approximating the distribution of $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_{0H})$. We approximate such distribution via weighted bootstrap.

Similar to Chatterjee and Bose (2005, Theorem 3.2) it can be shown that under some technical assumptions on the correlation structure of the bootstrap weights, the distribution of $T^{1/2}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_{0H})$ can be approximated by the distribution of $\sigma_T^{-1} T^{1/2}(\hat{\boldsymbol{\theta}}^*_T - \hat{\boldsymbol{\theta}}_T)$ outside a set of probability zero, where σ_T^2 denotes the variance of w_{T1}.

Three different schemes for weights are considered. These are

(i) Scheme M when weights have a multinomial $(T, 1/T, ..., 1/T)$ distribution.

(ii) Scheme G when $w_{Tt} = G_t/\overline{G}$, where G_t's are i.i.d Gamma(1,3) and $\overline{G} = T^{-1}\sum_{t=1}^{T} G_t$.

(iii) Scheme E when $w_{Tt} = E_t/\overline{E}$, where E_t's are i.i.d Exponential(1) $\overline{E} = T^{-1}\sum_{t=1}^{T} E_t$.

Note that Scheme M corresponds to the commonly-used paired-bootstrap in heteroscedastic models. We empirically study Schemes G and E as possible alternatives to the paired-bootstrap. It is possible to obtain quantiles of the bootstrap distribution of $\sigma_T^{-1}T^{1/2}(\hat{\boldsymbol{\theta}}_T^* - \hat{\boldsymbol{\theta}}_T)$ using simulation and then using the bootstrap approximation, we can construct the bootstrap confidence intervals of $\boldsymbol{\theta}_{0H}$.

3.6.1 Results

This section reports the results of resampling study. We are interested in studying the quality of bootstrap approximation to the finite sample distribution of $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_{0H})$. We use a sample of size T, and assume that the underlying error distributions of $\{\epsilon_t\}$ is standard normal. GARCH model is fitted using three M-estimators, the QMLE, LAD, and B-estimator.

Accordingly, we generate $K = 1000$ samples each of size $T = 500$ from the GARCH(1,1) model with $\boldsymbol{\theta}_0 = (0.05, 0.15, 0.65)$. Let $\hat{\boldsymbol{\theta}}_{T(k)}$ denote the estimate computed from the k-th sample, $1 \leq k \leq K$. The estimated marginal means and variances of $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_{0H})$ by the (entrywise) average and sample variance of these three sets of K numbers are reported in Table 3.14.

Table 3.14: Estimated mean and variance of $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_{0H})$ for GARCH(1,1) model.

$\boldsymbol{\theta} = (0.05, 0.15, 0.65)$	QMLE		LAD		B-estimator	
$T = 500$	Mean	Var	Mean	Var	Mean	Var
ω_0	0.0162	0.0032	-0.0475	0.0030	-0.0076	0.0032
α_0	0.1601	1.8087	-1.6823	3.6652	-0.5562	2.8326
β_0	-0.9836	12.9870	1.2091	10.1425	-0.3921	17.4046

Next, we generate $B^* = 1000$ bootstrap samples based on weights $\{w_{Tt}; 1 \leq t \leq T\}$ under M-, G- and E-schemes. Fixing k in $\{1, \ldots, K\}$, we obtain $\sigma_T^{-1}\sqrt{T}(\hat{\boldsymbol{\theta}}_T^* - \hat{\boldsymbol{\theta}}_{T(k)})$ from the b-th sample, $1 \leq b \leq B$, and compute (entrywise) average and sample average of these three sets of B numbers and finally took average of these quantities over $R = 100$. The results are shown in Table 3.15 where entries in bold represent the closest match of bootstrapped variance to that of Table 3.14.

From Table 3.15, in case of the QMLE, Scheme M, which is commonly-used paired bootstrap, seems to approximate the MSE of estimating ω_0 and β_0 more accurately whereas it cannot perform well for other parameters. Although Scheme G does not do well for the QMLE, it produces best approximation of MSE for LAD and also has a good performance for B-estimator. Scheme E turns out to be the second best choice in nearly all cases. It gives results close to Scheme M for the QMLE and also matches Scheme G for other estimators. Thus Scheme G is a good competitor to the Scheme M. Overall, none of these bootstrap schemes has any distinctly better performance compared to each other or normal approximation in approximating the mean and variance of the standardized M-estimators. Nevertheless, the bootstrap schemes do capture the shape of the standardized M-distributions reasonably well.

Table 3.15: Estimated mean and variance of $\sigma_T^{-1}\sqrt{T}(\hat{\theta}_T^* - \hat{\theta}_T)$ for GARCH(1,1) model.

	Scheme M		Scheme G		Scheme E	
$B^* = 1000$	Mean	Var	Mean	Var	Mean	Var
QMLE						
ω_0	-0.0021	**0.0031**	-0.0053	0.0026	-0.0025	0.0030
α_0	0.2065	2.4984	0.1709	**2.3781**	0.1760	2.4091
β_0	-0.0251	**12.4979**	0.1624	10.8873	-0.0294	12.4184
LAD						
ω_0	-0.0016	0.0007	-0.0014	**0.0008**	-0.0017	0.0007
α_0	0.1061	0.9536	0.1424	**1.0486**	0.0786	0.9583
β_0	0.0319	7.7821	-0.0495	**8.6255**	0.0147	8.2876
B-estimator						
ω_0	-0.0028	**0.0031**	-0.0033	0.0035	-0.0031	0.0034
α_0	0.2422	2.7450	0.3134	**2.8542**	0.2272	2.7402
β_0	0.0647	16.2140	0.0294	18.7581	0.0202	**18.1217**

3.7 Conclusion

We conclude this chapter by pointing out our main contributions and findings.

We discussed the robust estimation of heteroscedastic time series models. In particular, M-estimators for asymmetric GARCH models are defined. Algorithms for computing these estimators were also discussed. The class of M-estimators considered are the QMLE, LAD, Huber's, Cauchy and B-estimators. Monte Carlo simulations were conducted to check the relative performance of these estimators under different distributional assumptions for errors.

It was found that some of the M-estimators provide good alternatives to the widely-use QMLE. Estimators such as B-estimator, Cauchy, and Huber's produced very good results in terms of mean squared error and outperformed the QMLE. These estimators even outperformed the robust LAD estimator in cases when errors are generated from heavy-tailed distributions. Performances of these estimators were investigated under different set of parameter values and also under heavy-tailed asymmetric errors. We also noted that under the existence of some fractional error moments, it is still possible to estimate consistently scalar multiples of the GARCH and GJR parameters. Application to real data sets and simulations revealed the better performance of the alternative estimators. We conjecture that it is possible to investigate M-estimators under more general asymmetric GARCH model, namely, the asymmetric power ARCH or APARCH model, which nests a number of important symmetric and asymmetric models.

Using the idea of weighted resampling for estimating equations, suitable bootstrap versions for M-estimators were developed. We bootstrapped M-estimators and approximate the sampling distribution of the parameters with this new approach to resampling. In addition to the commonly-used Scheme M, we used Scheme G and Scheme E and results showed that these schemes show good results.

Overall, none of these bootstrap schemes had any distinctly better performance compared to each other or normal approximation in approximating the mean and variance of the standardized M-estimators. Nevertheless, the bootstrap schemes did capture the shape of the standardized M-distributions reasonably well and there is a need of further investigation on weighted resampling for M-estimators in GARCH models.

Chapter 4

Diagnostic Checking for GARCH-type Models

4.1 Introduction

The autoregressive conditional heteroscedastic (ARCH) model of Engle (1982) and the generalized ARCH (GARCH) model of Bollerslev (1986) have been found to be successful in capturing the volatility or the conditional variance structure of many financial time series. There is a huge literature on modeling these conditional heteroscedastic time series, but not much work has been done on model checking or model selection. Testing the adequacy of these heteroscedastic models is undoubtedly important for several economic and statistical reasons. Diagnostic is one of the important stages of model building. Generally, misspecification in the mean and variance results in inconsistency and loss of efficiency in the estimated parameters. Residual autocorrelations are used to identify possible departure from the assumption that the white noise disturbances in the specified model are uncorrelated (see Box and Jenkins, 1970).

To check the model adequacy, the asymptotic distribution of the squared and absolute residual autocorrelations derived from such models might be useful. One option is to look at the graphs of autocorrelation function (ACF) and partial autocorrelation function (PACF) of the residuals. The graphs will show which lags of the ACF and PACF display significant values and also reveal some remaining structures. However, these graphs only show linear dependent structures, and it is well known that in many cases we are likely to have different non-linear structures.

The second option is to build a test statistic to test the null hypothesis that the residuals are independent up to a lag M. The test statistic can be applied to check for non-linearity in mean and also for nonlinearity in variance. In particular, we are interested in finding not only linear, but also non-linear structures. The test statistics usually used are called portmanteau statistics. Next, we discussed some of the frequently used statistics in time series for diagnostic checking.

One of the widely used portmanteau statistic is the one proposed by Box and Pierce (1970). This statistic is used to test the null hypothesis that the first M autocorrelations of a covariance stationary time series are zero. If significant autocorrelation is not found in the residuals from the model, then the model is declared to be adequate. Ljung and Box (1978) discussed the finite sample properties and conservative behavior of the Box-Pierce statistic. In financial time series analysis, it is particularly important to check serial correlations of squared series. McLeod and Li (1983) derived a portmanteau test for model adequacy based on the squared residual autocorrelations in ARMA models.

In practice, many researchers apply the Ljung-Box or McLeod-Li tests to the squares of the estimated standardised residuals when testing the adequacy of an ARCH/GARCH model. A χ^2 distribution with M degrees of freedom, as the large sample distribution for these statistics is found misleading and using the squared residual autocorrelations a correct portmanteau test is proposed by Li and Mak (1994). They derived the asymptotic variance of the residual correlation coefficients, and suggested some diagnostics for the ARCH/GARCH models.

Wong and Li (1995) presented a portmanteau test using ranks of squared residuals and showed through simulations that their test using ranks is a more robust alternative to the McLeod-Li statistics. Ling and Li (1997b) further generalised the Li-Mak work and derived the asymptotic distribution of the portmanteau statistic in multivariate case. Tse and Zuo (1997) reported some Monte Carlo results for the finite sample performance of some commonly used diagnostics used in literature and found that the Li-Mak test based on the asymptotic variance under the Gaussian assumption performs favorably among other versions of statistics.

Asymptotic theory for quadratic forms of the autocorrelation of squared residuals from a GARCH(p, q) model was developed by Berkes et al. (2003b). Kwan et al. (2005) carried out a comparative study of the finite-sample performance of some well-known portmanteau tests. Based on their Monte Carlo results they reported that when the data generating process is skewed then the empirical size of these tests are severely undersized and that the non-parametric test is more powerful than the portmanteau tests.

Li and Li (2005) derived the asymptotic distributions of absolute residual autocorrelations and squared residual autocorrelations from the GARCH model estimated by the least absolute deviation method proposed by Peng and Yao (2003). They also develop diagnostic tools for checking the adequacy of GARCH models fitted by least absolute deviation method.

Tests used to check the adequacy of GARCH models estimated by the quasi-maximum likelihood method assume that the innovations have at least finite fourth moment. Li and Mak (1994) assume the conditional normality and the existence of fourth-order moment of the observations. However, these assumptions are not satisfied by many financial time series. Many heavy-tailed distributions are excluded as the existence of squared residual autocorrelations needs a finite fourth moment. Moreover, a drawback of this approach is the lack of robustness because it is sensitive to outliers and error distributions. Therefore, it is important to investigate the robustness of portmanteau tests. Tests derived under the assumption of Gaussian errors should be used only for such errors and hence it is also important to develop tests for GARCH-type models estimated by estimators other than the QMLE.

In the previous chapter we suggested using Cauchy and B-estimator for fitting GARCH-type model. Our Monte Carlo simulation and application to real data sets showed that these estimators perform better than the QMLE and even outperform the robust estimator such as the LAD when the error distribution is heavy-tailed. A significant contribution to the existing literature on the subject would be to develop some diagnostic tests that can be used to check the adequacy of GARCH-type models estimated by M-estimators. For this, it is important to derive the asymptotic distribution of absolute and squared residual autocorrelations for GARCH models. It is also important to examine the finite sample behaviour of these tests in the presence of outliers, under heavy-tailed and skewed distributions.

In this chapter we derive the asymptotic distributions of absolute and squared residual autocorrelations from GARCH-type models when M-estimators are used for estimation. We propose two new diagnostic tools based on the correct large sample distributions of the squared and absolute standardised residual autocorrelations. Since these tests are based on M-estimators, we call these robust portmanteau tests for GARCH models.

An advantage of using the absolute values of the residuals is that in order to obtain the asymptotic distribution we need to assume the existence of only the second-order moment of residuals, whereas for the squared residuals the existence of the fourth-order moment is required. Our results are valid under very weak conditions on the errors and hence robust under heavy-tailed distributions. The asymptotic distributions for these statistics are obtained and size and power analysis are conducted through Monte Carlo simulations. It is found that the asymptotic standard errors for both squared and absolute residual autocorrelations match the empirical standard errors quite satisfactory for all estimators. Investigation of the size of these tests suggests that empirical sizes of these tests are close to the nominal level. Analysis of the power of tests reveal that tests based on absolute residual autocorrelations outperform those based on squared residual autocorrelations. The power levels of tests, when Cauchy and B-estimators are used for estimation, are found superior than other estimators used.

The plan of the rest of the chapter is as follows. In the next section, we give a brief introduction to some commonly used portmanteau statistics. Diagnostic checking for GARCH models estimated by M-estimators are explained in Section 4.3 where asymptotic of squared and absolute residual autocorrelations are discussed in detail. Simulation results of the size and the power of the tests are reported in Section 4.4. Finally, Section 4.5 concludes the chapter.

4.2 Portmanteau Statistics

In this section we discuss some of the frequently used statistics in time series for diagnostic checking.

4.2.1 Box-Pierce Statistic

The Box-Pierce statistic is used to test the null hypothesis that the first M autocorrelations of a covariance stationary time series are zero. Under the assumption that the observations are independent and identically distributed, the asymptotic covariance matrix of the vector of sample autocorrelations is the inverse of the sample size times the identity matrix. This test is generally called the classical portmanteau statistic. The lag-k residual autocorrelation is defined as

$$\hat{r}_{1k} = \frac{\sum_{t=k+1}^{T}(\hat{\epsilon}_t - \bar{\epsilon})(\hat{\epsilon}_{t-k} - \bar{\epsilon})}{\sum_{t=1}^{T}(\hat{\epsilon}_t - \bar{\epsilon})^2} \quad \text{for} \quad k = 1, 2, ..., M, \tag{4.1}$$

where $\{\hat{\epsilon}_t\}$ are residuals from an autoregressive moving average, ARMA(p, q) model, $\bar{\epsilon} = \frac{1}{T}\sum \hat{\epsilon}_t$ and T is the sample size. The Box-Pierce statistic is defined as

$$Q_{BP}(M) = T\sum_{k=1}^{M} \hat{r}_{1k}^2$$

where $\hat{r}_{1k}$ is the sample residual autocorrelation of order $k = 1, ..., M$. Under the null hypothesis that ARMA(p, q), model is adequate, $Q_{BP}(M)$ is asymptotically distributed as a χ^2 with $(M - p - q)$ degrees of freedom.

4.2.2 Ljung-Box Statistic

A modified test proposed by Ljung and Box is

$$Q_{LB}(M) = T(T+2)\sum_{k=1}^{M} \frac{\hat{r}_{1k}^2}{T-k}.$$

It has been shown that the finite sample distribution of this statistic is much closer to that of the $\chi^2_{(M-p-q)}$, however its variance could be substantially larger than that of its asymptotic distribution.

4.2.3 McLeod-Li Statistic

The lag-k squared residual autocorrelation is defined as

$$\hat{r}_{2k} = \frac{\sum_{t=k+1}^{T}(\hat{\epsilon}_t^2 - \bar{\epsilon})(\hat{\epsilon}_{t-k}^2 - \bar{\epsilon})}{\sum_{t=1}^{T}(\hat{\epsilon}_t^2 - \bar{\epsilon})^2} \quad \text{for} \quad k = 1, 2, ..., M, \tag{4.2}$$

where $\bar{\epsilon} = \frac{1}{T}\sum \hat{\epsilon}_t^2$ and T is the sample size. The McLeod-Li statistic is

$$Q_{ML}(M) = T(T+2)\sum_{k=1}^{M} \frac{\hat{r}_{2k}^2}{T-k}.$$

They showed that, if the eighth order moment of the returns exists, $Q_{ML}(M)$ is distributed asymptotically as $\chi^2_{(M)}$. This test is asymptotically equivalent to the Langrange Multiplier (LM) test of Engle (1982). When the $Q_{ML}(M)$ statistic is implemented with absolute values, only the fourth order moment of returns should be finite for the asymptotic distribution to hold.

4.2.4 Li-Mak Statistic

Li and Mak (1994) derived the asymptotic variance of the correlation coefficients, and suggested some diagnostics for the ARCH/GARCH models. The lag-k correlation coefficient $\hat{r}_{3k}$ is defined as

$$\hat{r}_{3k} = \frac{\sum_{t=k+1}^{T}(\hat{\epsilon}_t^2 - 1)(\hat{\epsilon}_{t-k}^2 - 1)}{\sum_{t=1}^{T}(\hat{\epsilon}_t^2 - 1)^2}, \tag{4.3}$$

where $\hat{\epsilon}_t$ are the standardised residuals from GARCH model estimated by QMLE. Li and Mak (1994) showed that $\sqrt{T}\hat{\mathbf{r}}_3$ is asymptotically normally distributed with mean $\mathbf{0}$ and covariance matrix $\mathbf{V}$, where $\hat{\mathbf{r}}_3$ denotes the vector of sample correlation coefficients defined by $\hat{\mathbf{r}}_3 = (\hat{r}_{31}, ..., \hat{r}_{3M})'$ and $\mathbf{V}$ can be consistently estimated by $\hat{\mathbf{V}} = \mathbf{I}_M - (1/4)\mathbf{X}\hat{\mathbf{G}}^{-1}\mathbf{X}'$ where $\mathbf{I}_M$ is the $M \times M$ identity matrix, $\hat{\mathbf{G}}^{-1}$ is a consistent estimate of the asymptotic variance of $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ and $\mathbf{X} = (X_1, ..., X_M)'$ with $X_k = -T^{-1}\sum_{t=k+1}^{T}(\hat{\epsilon}_{t-k}^2 - 1)(\dot{\hat{h}}_t/\hat{h}_t)$, and $\hat{h}_t$ is the estimate of the conditional variance of the GARCH model. The Li-Mak statistic is

$$Q_{LM}(M) = T\hat{\mathbf{r}}'\hat{\mathbf{V}}^{-1}\hat{\mathbf{r}}.$$

If the model is correct, $Q_{LM}(M)$ asymptotically follows χ^2 distribution with M degrees of freedom.

4.3 Diagnostic Checking for GARCH-type Models Estimated by M-estimators

In this section we discuss the diagnostic checking for GARCH-type models when M-estimators are used for estimation. We derive the asymptotic distributions of autocorrelations of squared and absolute residuals from GARCH models estimated by M-estimators. Based on these results we develop portmanteau statistics that can be used to check the adequacy of GARCH models.

We start our discussion of M-estimators in the GJR model and introduce some notations and definitions (see Section 3.3.1 for detailed discussion on M-estimators for GARCH-type models). For simple GJR(1,1) model, the following representation of the return series $\{X_t; t \in \mathbb{Z}\}$ is assumed. Observe $\{X_t; 1 \leq t \leq T\}$ such that

$$\begin{aligned} X_t &= h_t^{1/2}\epsilon_t, \\ h_t &= \omega_0 + \alpha_0 X_{t-1}^2 + \beta_0 h_{t-1} + \gamma_0 D_{t-1} X_{t-1}^2 \end{aligned} \tag{4.4}$$

where $D_{t-1} = 1$ if $X_{t-1} < 0$ and 0 otherwise, with $\{\epsilon_t\}$ is a sequence of independent and identically distributed (i.i.d.) unobservable real-valued random variables and the unknown parameter $\boldsymbol{\theta}_0 = [\omega_0, \alpha_0, \gamma_0, \beta_0]'$ is in the parameter space

$$\Theta = \{\boldsymbol{\theta} = [\omega, \alpha, \gamma, \beta]'; \omega > 0, \alpha, \beta, \gamma \geq 0, (\alpha + \beta + \frac{1}{2}\gamma) < 1\}.$$

Under these parameter constraints, model (4.4) is strictly stationary and hence covariance stationary under finite second moment. The GJR(1,1) model reduces to the GARCH(1,1) model when there is no leverage effect i.e. when $\gamma_0 = 0$.

By recursive substitution, we get

$$h_t = \frac{\omega_0}{(1-\beta_0)} + \alpha_0 \sum_{j=1}^{\infty} \beta_0^{j-1} X_{t-j}^2 + \gamma_0 \sum_{j=1}^{\infty} D_{t-j} \beta_0^{j-1} X_{t-j}^2.$$

For $\boldsymbol{\theta} \in \Theta$, define the variance function

$$v_t(\boldsymbol{\theta}) = \frac{\omega}{(1-\beta)} + \alpha \sum_{j=1}^{\infty} \beta^{j-1} X_{t-j}^2 + \gamma \sum_{j=1}^{\infty} D_{t-j} \beta^{j-1} X_{t-j}^2, \tag{4.5}$$

and note that $v_t(\boldsymbol{\theta}_0) = h_t$.

In (4.4), if f denotes the error density, then the conditional density of $\{X_t\}$ given past will be $v_t^{-1/2}(\boldsymbol{\theta}_0) f\{v_t^{-1/2}(\boldsymbol{\theta}_0) X_t\}$, $1 \le t \le T$. Now we can define a random quantity as a minimizer of the negative log-likelihood function

$$L_T(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^{T} l_t(\boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \Theta,$$

where

$$l_t(\boldsymbol{\theta}) = \left[\left(\frac{1}{2}\right) \log v_t(\boldsymbol{\theta}) - \log f\{X_t / v_t^{1/2}(\boldsymbol{\theta})\} \right].$$

Then, the derivative of the log-likelihood is

$$\frac{\partial L_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{T} \left(\frac{1}{2}\right) \left[1 - H^*\{X_t / v_t^{1/2}(\boldsymbol{\theta})\}\right] \{\dot{v}_t(\boldsymbol{\theta}) / v_t(\boldsymbol{\theta})\},$$

where $H^*(x) := x\{-\dot{f}(x)/f(x)\}$.

More generally, we define a score function H as follows. Let $\psi : \mathbb{R} \to \mathbb{R}$ be a skew-symmetric function (that is $\psi(-x) = -\psi(x)$, $\forall x \in \mathbb{R} - \{0\}$) which is differentiable in all but finite number of points. Let $H(x) := x\psi(x)$, $x \in \mathbb{R}$. Note that $H(-x) = H(x)$, $\forall x$. We can then define $\boldsymbol{\theta}_T$ in the model (4.4) as a solution of the equation

$$\mathrm{M}_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} m_t(\boldsymbol{\theta}) = \mathbf{0},$$

where

$$m_t(\boldsymbol{\theta}) = \frac{\partial l_t}{\partial \boldsymbol{\theta}} = \left(\frac{1}{2}\right)\left\{1 - H\{X_t/v_t^{1/2}(\boldsymbol{\theta})\}\right\}\{\dot{v}_t(\boldsymbol{\theta})/v_t(\boldsymbol{\theta})\}. \tag{4.6}$$

Since $\{X_t; t \leq 0\}$ are not observable, $\{v_t(\boldsymbol{\theta})\}$'s are non observable and hence $\boldsymbol{\theta}_T$'s are noncomputable. We define an observable approximation $\{\hat{v}_t(\boldsymbol{\theta}); t \geq 1\}$ to the variance functions $\{v_t(\boldsymbol{\theta}); t \geq 1\}$ as

$$\hat{v}_t(\boldsymbol{\theta}) = \frac{\omega}{(1-\beta)} + \left\{\alpha \sum_{j=1}^{t-1} \beta^{j-1} X_{t-j}^2 + \gamma \sum_{j=1}^{t-1} D_{t-j} \beta^{j-1} X_{t-j}^2\right\} I(t \geq 2). \tag{4.7}$$

Then an M-estimator $\hat{\boldsymbol{\theta}}_T$ based on the score function H or ψ is defined as a solution of the equation

$$\hat{\mathrm{M}}_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} \hat{m}_t(\boldsymbol{\theta}) = \mathbf{0}, \tag{4.8}$$

where

$$\hat{m}_t(\boldsymbol{\theta}) = \left(\frac{1}{2}\right)\left\{1 - H\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta})\}\right\}\{\dot{\hat{v}}_t(\boldsymbol{\theta})/\hat{v}_t(\boldsymbol{\theta})\},$$

In Section 3.3.1, we showed that an M-estimator $\hat{\boldsymbol{\theta}}_T$, based on a score function H consistently estimates

$$\boldsymbol{\theta}_{0H} = [c_H\omega_0, c_H\alpha_0, c_H\gamma_0, \beta_0]', \tag{4.9}$$

where $c_H > 0$ is a constant that depends on the score function H through the error distribution. From (4.5) and (4.9), notice that

$$v_t(\boldsymbol{\theta}_{0H}) = c_H v_t(\boldsymbol{\theta}_0).$$

Hence

$$X_t/\{v_t(\boldsymbol{\theta}_{0H})\}^{1/2} = X_t/\{c_H v_t(\boldsymbol{\theta}_0)\}^{1/2} = \epsilon_t/c_H^{1/2}. \tag{4.10}$$

By Theorem 3.1.

$$T^{1/2}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_{0H}) \xrightarrow{\mathfrak{D}} \mathbf{N}[\mathbf{0}, \sigma^2(H)\mathbf{G}^{-1}]. \tag{4.11}$$

where

$$\sigma^2(H) := 4 \text{ var}\{H(\epsilon/c_H^{1/2})\}/[E\{(\epsilon/c_H^{1/2})\dot{H}(\epsilon/c_H^{1/2})\}]^2,$$

and

$$\mathbf{G} = \mathbf{G}(\boldsymbol{\theta}_{0H}) := E\{\dot{v}_1(\boldsymbol{\theta}_{0H})\dot{v}_1'(\boldsymbol{\theta}_{0H})/v_1^2(\boldsymbol{\theta}_{0H})\}.$$

Based on the above theorem, it can be shown that

$$\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_{0H} = (T\mathbf{B})^{-1}\sum_{t=1}^{T}\left(\frac{1}{2}\right)\left(1 - H\{\epsilon_t/c_H^{1/2}\}\right)\frac{\dot{v}_t(\boldsymbol{\theta}_{0H})}{v_t(\boldsymbol{\theta}_{0H})} + o_p(T^{-1/2}) \tag{4.12}$$

where $\mathbf{B} = -E(T^{-1}\dot{\mathbf{M}}_T(\boldsymbol{\theta}_{0H}))$ and

$$\begin{aligned}\dot{\mathbf{M}}_T(\boldsymbol{\theta}_{0H}) &= (1/4)\sum_{t=1}^{T}[\dot{H}(\epsilon_t/c_H^{1/2})(\epsilon_t/c_H^{1/2})\{\dot{v}_t(\boldsymbol{\theta}_{0H})\dot{v}_t'(\boldsymbol{\theta}_{0H})/v_t^2(\boldsymbol{\theta}_{0H})\}] \\ &+ (1/2)\sum_{t=1}^{T}[\{1 - H(\epsilon_t/c_H^{1/2})\}\{v_t(\boldsymbol{\theta}_{0H})\ddot{v}_t(\boldsymbol{\theta}_{0H}) - \dot{v}_t(\boldsymbol{\theta}_{0H})\dot{v}_t'(\boldsymbol{\theta}_{0H})\}/v_t^2(\boldsymbol{\theta}_{0H})].\end{aligned}$$

As shown in Mukherjee (2008, A.23)

$$\dot{\mathbf{M}}_T(\boldsymbol{\theta}_{0H})/(2T) \to \mathbf{V},$$

where

$$\mathbf{V} = \tilde{a}\,\mathbf{G} \text{ and } \tilde{a} = E\{(\epsilon/c_H^{1/2})\dot{H}(\epsilon/c_H^{1/2})\}/8 > 0. \tag{4.13}$$

Therefore, for M-estimators

$$\mathbf{B}^{-1} = -\frac{4\mathbf{G}^{-1}}{k(H)},$$

where $k(H) = E\{(\epsilon/c_H^{1/2})\dot{H}(\epsilon/c_H^{1/2})\}$.

We can also write $\sigma^2(H)$ in (4.11) as

$$\sigma^2(H) = \frac{4\sigma_H^2}{k^2(H)},$$

where $\sigma_H^2 = \text{var}\{H(\epsilon/c_H^{1/2})\}$.

Next, the asymptotic distributions of squared residual autocorrelations and absolute residual autocorrelations for GARCH models are derived. Using these results two new diagnostic tests are developed that can be used to test the adequacy of GARCH-type models when M-estimators are used for estimation.

4.3.1 Asymptotic Distribution of the Squared Residual Autocorrelations

In this section we derive the asymptotic distribution of the squared residual autocorrelation. This leads us to build a useful portmanteau test for checking the adequacy of GARCH models fitted by M-estimators.

Define the estimated residuals by

$$\hat{\epsilon}_{tH} = X_t/\{\hat{v}_t(\boldsymbol{\theta}_T)\}^{1/2},\ 1 \le t \le T. \tag{4.14}$$

Following Li and Mak (1994), the lag-k standardized squared residual autocorrelation can be defined as

$$\tilde{r}_k = \frac{\sum_{t=k+1}^{T}(X_t^2/\hat{v}_t - \bar{\epsilon}_H)(X_{t-k}^2/\hat{v}_{t-k} - \bar{\epsilon}_H)}{\sum_{t=1}^{T}(X_t^2/\hat{v}_t - \bar{\epsilon}_H)^2} \quad \text{for} \quad k = 1, ..., M, \tag{4.15}$$

where $\bar{\epsilon}_H = \frac{1}{T}\sum_{t=1}^{T} X_t^2/\hat{v}_t$, and $\hat{v}_t = \hat{v}_t(\boldsymbol{\theta}_T)$.

If the model is correct, by the ergodic theorem,

$$\bar{\epsilon}_H = \frac{1}{T}\sum_{t=1}^{T} X_t^2/\hat{v}_t \xrightarrow{\text{a.s}} E(X_t^2/v_t) \qquad \text{as} \quad T \to \infty,$$

and $E(X_t^2/v_t) = E(\epsilon_t^2/c_H) = \mu_\epsilon$. So $\tilde{r}_k$ can be replaced by

$$\hat{r}_k = \frac{\sum_{t=k+1}^{T}(X_t^2/\hat{v}_t - \mu_\epsilon)(X_{t-k}^2/\hat{v}_{t-k} - \mu_\epsilon)}{\sum_{t=1}^{T}(X_t^2/\hat{v}_t - \mu_\epsilon)^2} \quad \text{for} \quad k = 1, ..., M. \tag{4.16}$$

In particular $\mu_\epsilon = 1$, if $\{\epsilon_t\}$ follows the standard normal distribution. We consider the asymptotic distributions of the squared residual autocorrelations $(\hat{r}_1, ..., \hat{r}_M)'$ for some integer $M > 0$.

If the model is correct,

$$\frac{1}{T}\sum_{t=1}^{T}(X_t^2/\hat{v}_t - \mu_\epsilon)^2 \xrightarrow{\text{a.s}} E(X_t^2/v_t - \mu_\epsilon)^2 \qquad \text{as} \quad T \to \infty,$$

and $E(X_t^2/v_t - \mu_\epsilon)^2 = E(\epsilon_t^2/c_H - \mu_\epsilon)^2 = \text{var}(\epsilon_t^2/c_H) = \sigma_\epsilon^2$. The constant $\sigma_\epsilon^2 = 2$, in case of the standard normal distribution.

Hence we only need to consider the asymptotic distribution of

$$\hat{C}_k = \frac{1}{T}\sum_{t=k+1}^{T}\left(\frac{X_t^2}{\hat{v}_t} - \mu_\epsilon\right)\left(\frac{X_{t-k}^2}{\hat{v}_{t-k}} - \mu_\epsilon\right),$$

and $\hat{r}_k$ can be written as $\hat{C}_k/\hat{C}_0$. Denote by C_k the counterpart of $\hat{C}_k$ when $\hat{v}_t$ is replaced by v_t. Let $\mathbf{C} = (C_1, C_2, ..., C_M)'$ and $\hat{\mathbf{C}} = (\hat{C}_1, \hat{C}_2, ..., \hat{C}_M)'$, for some integer $M > 0$. We can define $\mathbf{r} = (r_1, r_2, ..., r_M)'$ and $\hat{\mathbf{r}} = (\hat{r}_1, \hat{r}_2, ..., \hat{r}_M)'$ in a similar way. By expanding $\hat{\mathbf{C}}$ in a Taylor series expansion, we obtain

$$\hat{\mathbf{C}} \approx \mathbf{C} + \frac{\partial \mathbf{C}}{\partial \boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_{0H}), \tag{4.17}$$

where $\partial \mathbf{C}/\partial \boldsymbol{\theta} = (\partial C_1/\partial \boldsymbol{\theta}, ..., \partial C_M/\partial \boldsymbol{\theta})'$, and for $k = 1, ..., M$,

$$\frac{\partial C_k}{\partial \boldsymbol{\theta}} = -\frac{1}{T}\sum_{t=k+1}^{T} \frac{X_t^2}{v_t^2}\left(\frac{X_{t-k}^2}{v_{t-k}} - \mu_\epsilon\right)\dot{v}_t - \frac{1}{T}\sum_{t=k+1}^{T} \frac{X_{t-k}^2}{v_{t-k}^2}\left(\frac{X_t^2}{v_t} - \mu_\epsilon\right)\dot{v}_{t-k}.$$

By the ergodic theorem, we obtain

$$\frac{\partial C_k}{\partial \boldsymbol{\theta}} \xrightarrow{\text{a.s}} -Y_k, \qquad \text{as } T \to \infty,$$

where

$$Y_k = \mu_\epsilon E\left[\left(\frac{X_{t-k}^2}{v_{t-k}} - \mu_\epsilon\right)\frac{\dot{v}_t}{v_t}\right]$$

Then $\hat{\mathbf{C}}$ in (4.17) can be approximated by

$$\hat{\mathbf{C}} \approx \mathbf{C} - \mathbf{Y}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_{0H}). \tag{4.18}$$

where $\mathbf{Y} = (Y_1, ..., Y_M)'$.

The following lemma may be shown by straight forward calculation.

Lemma 4.3.1 *For any constant vector* $\mathbf{Z} = (Z_1, Z_2, ..., Z_M)'$

$$\sqrt{T}\mathbf{Z}'\mathbf{C} = \frac{1}{\sqrt{T}}\sum_{t=M+1}^{T} U_t + O_p(1), \tag{4.19}$$

where

$$U_t = \sum_{k=1}^{M} Z_k\left(\frac{X_t^2}{v_t} - \mu_\epsilon\right)\left(\frac{X_{t-k}^2}{v_{t-k}} - \mu_\epsilon\right). \tag{4.20}$$

and

$$E\{U_t^2\} = \sigma_\epsilon^4 \mathbf{Z}'\mathbf{Z} < \infty. \tag{4.21}$$

Lemma 4.3.2

$$E\{(\partial l_t/\partial \boldsymbol{\theta})U_t\} = \frac{d(H)}{2\mu_\epsilon}\mathbf{Y}'\mathbf{Z} \tag{4.22}$$

where U_t is defined as in (4.20) and $d(H) = \mu_\epsilon - E\{(X_t^2/v_t)H(X_t/v_t^{1/2})\}$.

Proof. By (4.4), (4.6) and (4.20), we have

$$\begin{aligned}\frac{\partial l_t}{\partial \boldsymbol{\theta}} U_t &= \left[\left(\frac{1}{2}\right)\left\{1 - H\{X_t/v_t^{1/2}\}\right\}\{\dot{v}_t/v_t\}\right] \\ &\times \left[\sum_{k=1}^{M} Z_k\left(\frac{X_t^2}{v_t} - \mu_\epsilon\right)\left(\frac{X_{t-k}^2}{v_{t-k}} - \mu_\epsilon\right)\right].\end{aligned} \tag{4.23}$$

It follows that

$$\begin{aligned}E\Big(\frac{\partial l_t}{\partial \boldsymbol{\theta}}U_t\Big) &= E\Big\{E\Big(\frac{\partial l_t}{\partial \boldsymbol{\theta}}U_t\Big|\mathcal{F}_{t-1}\Big)\Big\} \\ &= \frac{1}{2\mu_\epsilon}\left[\mu_\epsilon - E\{(X_t^2/v_t)H(X_t/v_t^{1/2})\}\right]\sum_{k=1}^{M} Z_k E\left[\mu_\epsilon\left(\frac{X_{t-k}^2}{v_{t-k}} - \mu_\epsilon\right)\frac{\dot{v}_t}{v_t}\right] \\ &= \frac{1}{2\mu_\epsilon}\left[\mu_\epsilon - E\{(X_t^2/v_t)H(X_t/v_t^{1/2})\}\right]\sum_{k=1}^{M} Z_k Y_k \\ &= \frac{d(H)}{2\mu_\epsilon}\mathbf{Y}'\mathbf{Z}.\end{aligned}$$

Lemma 4.3.3 *The asymptotic joint distribution of $\sqrt{T}\mathbf{C}$ and $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_{0H})$ is normal with mean zero and covariance*

$$\begin{pmatrix} \sigma_\epsilon^4 \mathbf{I}_M & d(H)\mathbf{Y}\mathbf{B}^{-1}/(2\mu_\epsilon) \\ d(H)\mathbf{B}^{-1}\mathbf{Y}'/(2\mu_\epsilon) & \sigma^2(H)\mathbf{G}^{-1} \end{pmatrix}.$$

Proof. Let $\tilde{\mathbf{Z}} = (\mathbf{Z}', \mathbf{V}')'$ be any constant vector and $\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}} \neq \mathbf{0}$, where $\mathbf{V}$ is of the same dimension as $\boldsymbol{\theta}$.

By (4.12) and (4.19), we have

$$\begin{aligned}\sqrt{T}\tilde{\mathbf{Z}}'(\mathbf{C}', \hat{\boldsymbol{\theta}}'_T - \boldsymbol{\theta}'_{0H})' &= \frac{1}{\sqrt{T}}\left(\sum_{t=M+1}^{T} U_t + \sum_{t=1}^{T} \mathbf{V}'\mathbf{B}^{-1}\frac{\partial l_t}{\partial \boldsymbol{\theta}}\right) + O_p(1) \\ &= \frac{1}{\sqrt{T}}\sum_{t=M+1}^{T}\left(U_t + \mathbf{V}'\mathbf{B}^{-1}\frac{\partial l_t}{\partial \boldsymbol{\theta}}\right) + O_p(1) \qquad (4.24)\end{aligned}$$

It can be easily shown that $(1/\sqrt{T})\sum_{t=M+1}^{T}\left(U_t + \mathbf{V}'\mathbf{B}^{-1}\partial l_t/\partial\boldsymbol{\theta}\right)$ is a martingale.

Now, by (4.21), Lemma 4.3.1 and Lemma 4.3.2,

$$\begin{aligned}&E\left(U_t + \mathbf{V}'\mathbf{B}^{-1}\partial l_t/\partial\boldsymbol{\theta}\right)^2 \\ &= \sigma_\epsilon^4 \mathbf{Z}'\mathbf{I}_M\mathbf{Z} + d(H)\mathbf{Z}'\mathbf{Y}\mathbf{B}^{-1}\mathbf{V}/(2\mu_\epsilon) + d(H)\mathbf{V}'\mathbf{B}^{-1}\mathbf{Y}'\mathbf{Z}/(2\mu_\epsilon) + \mathbf{V}'\sigma^2(H)\mathbf{G}^{-1}\mathbf{V} \\ &= \tilde{\mathbf{Z}}'\begin{pmatrix} \sigma_\epsilon^4\mathbf{I}_M & d(H)\mathbf{Y}\mathbf{B}^{-1}/(2\mu_\epsilon) \\ d(H)\mathbf{B}^{-1}\mathbf{Y}'/(2\mu_\epsilon) & \sigma^2(H)\mathbf{G}^{-1}\end{pmatrix}\tilde{\mathbf{Z}} < \infty.\end{aligned}$$

Hence, by (4.24) and Billingsley's (1961) martingale central limit theorem, the proof is completed.

Theorem 4.3.4

$$\begin{aligned}\sqrt{T}\hat{\mathbf{C}} &\xrightarrow{\mathfrak{D}} \mathbf{N}[\mathbf{0}, \sigma_\epsilon^4\mathbf{W}] \qquad \textit{as} \quad T \to \infty, \\ \sqrt{T}\hat{\mathbf{r}} &\xrightarrow{\mathfrak{D}} \mathbf{N}[\mathbf{0}, \mathbf{W}] \qquad \textit{as} \quad T \to \infty,\end{aligned}$$

where

$$\mathbf{W} = \mathbf{I}_M + \frac{4\mathbf{Y}\mathbf{G}^{-1}\mathbf{Y}'}{\sigma_\epsilon^4 k(H)}\left\{\frac{d(H)}{\mu_\epsilon} + \frac{\sigma_H^2}{k(H)}\right\}.$$

Proof. This follows from (4.18) and Lemma 4.3.3.

Remark:

The proofs of these lemmas and theorem are shown by Ling and Li (1997b) for multivariate case when QMLE is used for estimation.

The correct asymptotic standard errors for the squared residual autocorrelations can be obtained from the above and these will give more accurate asymptotic standard errors than $1/\sqrt{T}$ for the squared residual autocorrelations.

In general, the matrix $\mathbf{W}$ is not an idempotent matrix even asymptotically. Therefore, $T\hat{\mathbf{r}}'\hat{\mathbf{r}}$ is not asymptotically distributed as a χ^2. However, based on our results, if the model is correct, the statistic

$$Q(M) = T\hat{\mathbf{r}}'\mathbf{W}^{-1}\hat{\mathbf{r}}$$

will be asymptotically χ^2 distributed with M degrees of freedom. Hence, a large value of the statistic Q will imply that there is a temporal dependence in variance of the series under investigation. Equivalently, this gives an indication of the presence of conditional heteroscedasticity and thus ARCH/GARCH models should be considered. This new portmanteau statistic may be useful for checking the adequacy of GARCH type models that are estimated by M-estimators.

If the distribution of $\{\epsilon_t\}$ is known, the exact values of $k(H), \sigma_H^2$, and μ_ϵ can be obtained. For example when ϵ_t follows the standard normal distribution, we have $c_H = 1$, $k(H) = 2$, $\sigma_H^2 = 2$, $\mu_\epsilon = 1$, $\sigma_\epsilon^4 = 4$ and $d(H) = -2$. Hence, the asymptotic covariance matrix of $\sqrt{T}\hat{\mathbf{r}}$ is

$$\mathbf{W} = \mathbf{I}_M - \frac{1}{4}\mathbf{Y}\mathbf{G}_1^{-1}\mathbf{Y}',$$

where $\mathbf{G}_1^{-1} = 2\mathbf{G}^{-1}$, and we have the Li and Mak (1994) result. If v_t is constant over time, then $\mathbf{Y} = \mathbf{0}$; the asymptotic standard error of $\hat{r}_k$ is exactly $1/\sqrt{T}$ and we get the McLeod and Li (1983) result.

In general, often the error distribution is unknown. We can estimate these quantities as

$$\begin{aligned}
\hat{k}(H) &= \frac{1}{T}\sum_{t=1}^{T}\{(X_t/\hat{v}_t^{1/2})\dot{H}(X_t/\hat{v}_t^{1/2})\}, \quad \hat{\mu}_\epsilon = \frac{1}{T}\sum_{t=1}^{T}(X_t^2/\hat{v}_t), \\
\hat{\sigma}_H^2 &= \frac{1}{T}\sum_{t=1}^{T}\{H(X_t/\hat{v}_t^{1/2})\}^2 - \left[\frac{1}{T}\sum_{t=1}^{T}\{H(X_t/\hat{v}_t^{1/2})\}\right]^2, \\
\text{and} \quad \hat{d}(H) &= \hat{\mu}_\epsilon - \frac{1}{T}\sum_{t=1}^{T}\{(X_t^2/\hat{v}_t)H(X_t/\hat{v}_t^{1/2})\}.
\end{aligned}$$

Also, $\mathbf{G}^{-1}$ can be estimated by $(\hat{\mathbf{G}})^{-1}$ where

$$\hat{\mathbf{G}} = \frac{1}{T}\sum_{t=1}^{T}\left\{\frac{\dot{\hat{v}}_t\dot{\hat{v}}_t'}{\hat{v}_t^2}\right\},$$

$\mathbf{Y}$ can be estimated by $\hat{\mathbf{Y}} = (\hat{Y}_1, ..., \hat{Y}_M)'$, where

$$\hat{Y}_k = \frac{1}{T}\sum_{t=k+1}^{T}\hat{\mu}_\epsilon\left(\frac{X_{t-k}^2}{\hat{v}_{t-k}} - \hat{\mu}_\epsilon\right)\frac{\dot{\hat{v}}_t}{\hat{v}_t}.$$

Also σ_ϵ^4 can be replaced by $(\hat{C}_0)^2$. Using these sample estimates, we can define an estimate of $\mathbf{W}$ as

$$\hat{\mathbf{W}} = \mathbf{I}_M + \frac{4\hat{\mathbf{Y}}\hat{\mathbf{G}}^{-1}\hat{\mathbf{Y}}'}{\hat{\sigma}_\epsilon^4\,\hat{k}(H)}\left\{\frac{\hat{d}(H)}{\hat{\mu}_\epsilon} + \frac{\hat{\sigma}_H^2}{\hat{k}(H)}\right\}.$$

4.3.2 Asymptotic Distribution of the Absolute Residual Autocorrelations

Next, the asymptotic distribution of the absolute standardised residual autocorrelation is discussed. The lag-k standardized absolute residual autocorrelation is defined as

$$\tilde{\rho}_k = \frac{\sum_{t=k+1}^{T}(|X_t|/\hat{v}_t^{1/2} - \bar{\varepsilon}_H)(|X_{t-k}|/\hat{v}_{t-k}^{1/2} - \bar{\varepsilon}_H)}{\sum_{t=1}^{T}(|X_t|/\hat{v}_t^{1/2} - \bar{\varepsilon}_H)^2} \quad \text{for} \quad k = 1, ..., M, \tag{4.25}$$

where $\bar{\varepsilon}_H = \frac{1}{T}\sum_{t=1}^{T}|X_t|/\hat{v}_t^{1/2}$.

If the model is correct, by the ergodic theorem,

$$\bar{\varepsilon}_H = \frac{1}{T}\sum_{t=1}^{T}|X_t|/\hat{v}_t^{1/2} \xrightarrow{\text{a.s}} E(|X_t|/v_t^{1/2}) \qquad \text{as} \quad T \to \infty,$$

and $E(|X_t|/v_t^{1/2}) = E(|\varepsilon_t|/c_H^{1/2}) = \nu_\varepsilon$. So $\tilde{\rho}_k$ can be replaced by

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{T}(|X_t|/\hat{v}_t^{1/2} - \nu_\varepsilon)(|X_{t-k}|/\hat{v}_{t-k}^{1/2} - \nu_\varepsilon)}{\sum_{t=1}^{T}(|X_t|/\hat{v}_t^{1/2} - \nu_\varepsilon)^2} \quad \text{for} \quad k = 1, ..., M. \tag{4.26}$$

Now consider the asymptotic distributions of the absolute residual autocorrelations $(\hat{\rho}_1, ..., \hat{\rho}_M)'$ for some integer $M > 0$. If the model is correct,

$$\frac{1}{T}\sum_{t=k+1}^{T}(|X_t|/\hat{v}_t^{1/2} - \nu_\varepsilon)^2 \xrightarrow{\text{a.s}} E(|X_t|/\hat{v}_t^{1/2} - \nu_\varepsilon)^2 \qquad \text{as} \quad T \to \infty,$$

and $E(|X_t|/\hat{v}_t^{1/2} - \nu_\varepsilon)^2 = E(|\varepsilon_t|/c_H^{1/2} - \nu_\varepsilon)^2 = \varsigma_\varepsilon^2$.

Hence we only need to consider the asymptotic distribution of

$$\hat{C}_k^a = \frac{1}{T}\sum_{t=k+1}^{T}\left(\frac{|X_t|}{\hat{v}_t^{1/2}} - \nu_\varepsilon\right)\left(\frac{|X_{t-k}|}{\hat{v}_{t-k}^{1/2}} - \nu_\varepsilon\right),$$

$\hat{\rho}_k$ can be written as $\hat{C}_k^a/\hat{C}_0^a$. Let $\hat{\mathbf{C}}^a = (\hat{C}_1^a, ..., \hat{C}_M^a)'$ and $\mathbf{C}^a = (C_1^a, ..., C_M^a)'$, for some integer $M > 0$. $\hat{\boldsymbol{\rho}}$ and $\boldsymbol{\rho}$ can be defined in a similar way. By Taylor's expansion of $\hat{\mathbf{C}}^a$ about $\boldsymbol{\theta}_{0H}$ and evaluated at $\hat{\boldsymbol{\theta}}_T$, we have

$$\hat{\mathbf{C}}^a \approx \mathbf{C}^a + \frac{\partial \mathbf{C}^a}{\partial \boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_{0H}), \tag{4.27}$$

where $\partial\mathbf{C}^a/\partial\boldsymbol{\theta} = (\partial C_1^a/\partial\boldsymbol{\theta}, ..., \partial C_M^a/\partial\boldsymbol{\theta})'$, and for $k = 1, ..., M$,

$$\frac{\partial C_k^a}{\partial\boldsymbol{\theta}} = -\frac{1}{2T}\sum_{t=k+1}^{T}\frac{|X_t|}{v_t^{1/2}}\left(\frac{|X_{t-k}|}{v_{t-k}^{1/2}} - \nu_\varepsilon\right)\frac{\dot{v}_t}{v_t} - \frac{1}{2T}\sum_{t=k+1}^{T}\frac{|X_{t-k}|}{v_{t-k}^{1/2}}\left(\frac{|X_t|}{v_t^{1/2}} - \nu_\varepsilon\right)\frac{\dot{v}_{t-k}}{v_{t-k}}.$$

By the ergodic theorem, we obtain

$$\frac{\partial C_k^a}{\partial\boldsymbol{\theta}} \xrightarrow{\text{a.s}} -Y_k^a, \qquad \text{as } T \to \infty,$$

where

$$Y_k^a = \frac{\nu_\varepsilon}{2}E\left[\left(\frac{|X_{t-k}|}{v_{t-k}^{1/2}} - \nu_\varepsilon\right)\frac{\dot{v}_t}{v_t}\right]$$

Then $\hat{\mathbf{C}}$ in (4.27) can be approximated by

$$\hat{\mathbf{C}}^a \approx \mathbf{C}^a - \mathbf{Y}^a(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_{0H}). \tag{4.28}$$

where $\mathbf{Y}^a = (Y_1^a, ..., Y_M^a)'$. By simple calculation the following lemma may be shown.

Lemma 4.3.5 *For any constant vector* $\mathbf{Z} = (Z_1, Z_2, ..., Z_M)'$

$$\sqrt{T}\mathbf{Z}'\mathbf{C}^a = \frac{1}{\sqrt{T}}\sum_{t=M+1}^{T} U_t^a + O_p(1), \tag{4.29}$$

where

$$U_t^a = \sum_{k=1}^{M} Z_k\left(\frac{|X_t|}{v_t^{1/2}} - \nu_\varepsilon\right)\left(\frac{|X_{t-k}|}{v_{t-k}^{1/2}} - \nu_\varepsilon\right). \tag{4.30}$$

and

$$E\{(U_t^a)^2\} = \varsigma_\varepsilon^4\mathbf{Z}'\mathbf{Z} < \infty. \tag{4.31}$$

Lemma 4.3.6

$$E\{(\partial l_t/\partial\boldsymbol{\theta})U_t^a\} = \frac{1}{\nu_\varepsilon}d^a(H)\mathbf{Y}^a\,\mathbf{Z} \tag{4.32}$$

where U_t^a is defined in (4.30) and $d^a(H) = \nu_\varepsilon - E\{(|X_t|/v_t^{1/2})H(X_t/v_t^{1/2})\}$.

Proof. By (4.4), (4.6) and (4.30), we have

$$\begin{aligned}\frac{\partial l_t}{\partial \boldsymbol{\theta}} U_t^a &= \left[\left(\frac{1}{2}\right)\left\{1 - H\{X_t/v_t^{1/2}\}\right\}\{\dot{v}_t/v_t\}\right] \\ &\times \left[\sum_{k=1}^{M} Z_k \left(\frac{|X_t|}{v_t^{1/2}} - \nu_\varepsilon\right)\left(\frac{|X_{t-k}|}{v_{t-k}^{1/2}} - \nu_\varepsilon\right)\right]. \end{aligned} \quad (4.33)$$

Hence, it follows that

$$\begin{aligned} E\left(\frac{\partial l_t}{\partial \boldsymbol{\theta}} U_t^a\right) &= E\left\{E\left(\frac{\partial l_t}{\partial \boldsymbol{\theta}} U_t^a \Big| \mathcal{F}_{t-1}\right)\right\} \\ &= \frac{1}{2\nu_\varepsilon}\left[\nu_\varepsilon - E\{(|X_t|/v_t^{1/2})H(X_t/v_t^{1/2})\}\right]\sum_{k=1}^{M} Z_k E\left[\nu_\varepsilon\left(\frac{|X_{t-k}|}{v_{t-k}^{1/2}} - \nu_\varepsilon\right)\frac{\dot{v}_t}{v_t}\right] \\ &= \frac{1}{\nu_\varepsilon}\left[\nu_\varepsilon - E\{(|X_t|/v_t^{1/2})H(X_t/v_t^{1/2})\}\right]\sum_{k=1}^{M} Z_k Y_k^a \\ &= \frac{1}{\nu_\varepsilon} d^a(H)\mathbf{Y}^a \mathbf{Z}. \end{aligned}$$

Lemma 4.3.7 *The joint distribution of $\sqrt{T}\mathbf{C}^a$ and $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_{0H})$ is asymptotically normal with mean zero and covariance*

$$\begin{pmatrix} \varsigma_\varepsilon^4 \mathbf{I}_M & d^a(H)\mathbf{Y}^a\mathbf{B}^{-1}/\nu_\varepsilon \\ d^a(H)\mathbf{B}^{-1}\mathbf{Y}^a/\nu_\varepsilon & \sigma^2(H)\mathbf{G}^{-1} \end{pmatrix}.$$

Proof. Let $\tilde{\mathbf{Z}} = (\mathbf{Z}', \mathbf{V}')'$ be any constant vector and $\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}} \neq \mathbf{0}$, where the dimension of $\mathbf{V}$ is same as that of $\boldsymbol{\theta}$.

Now, by (4.12) and (4.29), we have

$$\begin{aligned}\sqrt{T}\tilde{\mathbf{Z}}'(\mathbf{C}^{a\prime},\hat{\boldsymbol{\theta}}_T'-\boldsymbol{\theta}')' &= \frac{1}{\sqrt{T}}\left(\sum_{t=M+1}^{T}U_t^a+\sum_{t=1}^{T}\mathbf{V}'\mathbf{B}^{-1}\frac{\partial l_t}{\partial\boldsymbol{\theta}}\right)+O_p(1)\\ &= \frac{1}{\sqrt{T}}\sum_{t=M+1}^{T}\left(U_t^a+\mathbf{V}'\mathbf{B}^{-1}\frac{\partial l_t}{\partial\boldsymbol{\theta}}\right)+O_p(1) \qquad (4.34)\end{aligned}$$

It can be shown that $(1/\sqrt{T})\sum_{t=M+1}^{T}\left(U_t^a+\mathbf{V}'\mathbf{B}^{-1}\partial l_t/\partial\boldsymbol{\theta}\right)$ is a martingale and by (4.31), Lemma 4.3.5 and Lemma 4.3.6,

$$\begin{aligned}&E\left(U_t^a+\mathbf{V}'\mathbf{B}^{-1}\partial l_t/\partial\boldsymbol{\theta}\right)^2\\ &= \varsigma_\epsilon^4\mathbf{Z}'\mathbf{I}_M\mathbf{Z}+d^a(H)\mathbf{Z}'\mathbf{Y}^a\mathbf{B}^{-1}\mathbf{V}/\nu_\epsilon+d^a(H)\mathbf{V}'\mathbf{B}^{-1}\mathbf{Y}^a\,\mathbf{Z}/\nu_\varepsilon+\mathbf{V}'\sigma^2(H)\mathbf{G}^{-1}\mathbf{V}\\ &= \tilde{\mathbf{Z}}'\begin{pmatrix}\varsigma_\varepsilon^4\mathbf{I}_M & d^a(H)\mathbf{Y}^a\mathbf{B}^{-1}/\nu_\epsilon\\ d^a(H)\mathbf{B}^{-1}\mathbf{Y}^a\,/\nu_\epsilon & \sigma^2(H)\mathbf{G}^{-1}\end{pmatrix}\tilde{\mathbf{Z}}<\infty.\end{aligned}$$

Hence, the proof completes by using martingale central limit theorem and (4.34).

Theorem 4.3.8

$$\begin{aligned}\sqrt{T}\hat{\mathbf{C}}^a &\xrightarrow{\mathfrak{D}} \mathbf{N}[\mathbf{0},\varsigma_\varepsilon^4\mathbf{W}^a] \qquad as \quad T\to\infty,\\ \sqrt{T}\hat{\boldsymbol{\rho}} &\xrightarrow{\mathfrak{D}} \mathbf{N}[\mathbf{0},\mathbf{W}^a] \qquad as \quad T\to\infty,\end{aligned}$$

where

$$\mathbf{W}^a=\mathbf{I}_M+\frac{4\mathbf{Y}^a\mathbf{G}^{-1}\mathbf{Y}^a}{\varsigma_\varepsilon^4\,k(H)}\left\{\frac{2d^a(H)}{\nu_\varepsilon}+\frac{\sigma_H^2}{k(H)}\right\}.$$

Proof. This follows from (4.28) and Lemma 4.3.7.

Hence we obtain the correct asymptotic standard errors for the absolute residual autocorrelations. These will give more accurate standard errors for the absolute residual autocorrelations.

In general, $\mathbf{W}^a$ is not an idempotent matrix even asymptotically, therefore asymptotically $T\hat{\boldsymbol{\rho}}'\hat{\boldsymbol{\rho}}$ is not distributed as a chi-squared. However, if the model is correct the portmanteau statistic

$$Q^a(M) = T\hat{\boldsymbol{\rho}}'[\mathbf{W}^a]^{-1}\hat{\boldsymbol{\rho}}$$

will be asymptotically χ^2 distributed with M degrees of freedom. The adequacy of GARCH type models that are estimated by M-estimators can be checked using this new portmanteau statistic. It is worth mentioning here that only the existence of a second-order moment is required in this case. Since the distribution of ϵ_t is not known, the values of $k(H)$, σ_H^2 and others can be estimated by their sample counter parts. Also, Y_k^a can be estimated by

$$\hat{Y}_k^a = \frac{1}{2T}\sum_{t=k+1}^{T} \hat{\nu}_\varepsilon \left(\frac{|X_{t-k}|}{\hat{v}_{t-k}^{1/2}} - \hat{\nu}_\varepsilon \right) \frac{\dot{\hat{v}}_t}{\hat{v}_t},$$

where

$$\hat{\nu}_\varepsilon = \frac{1}{T}\sum_{t=1}^{T}(|X_t|/\hat{v}_t^{1/2}),$$

and ς_ε^4 can be replaced by $(\hat{C}_0^a)^2$.

4.4 Results

In this section we report results of Monte Carlo simulations. First, the usefulness of the asymptotic results are assessed using simulations and results of empirical and large sample standard errors are presented. Then, the empirical size and power of the tests are investigated through Monte Carlo simulations.

4.4.1 Result of Empirical and Large Sample Standard Errors

First we perform simulations to assess the usefulness of the asymptotic results obtained in previous sections. For our first experiment the time series X_t satisfies the GJR-GARCH(1,1) model,

$$\begin{aligned} X_t &= h_t^{1/2}\epsilon_t, \\ h_t &= 0.01 + 0.4X_{t-1}^2 + 0.5h_{t-1} + 0.1X_{t-1}^2 \mathrm{I}(X_{t-1} < 0), \end{aligned}$$

where $\{\epsilon_t\}$ is an independently identically distributed sequence. Data are also generated from the GARCH(1,1) model and for this purpose we set $\gamma = 0$. The errors are generated from the standard normal distribution, standardised student-t distribution with 3 and 5 degrees of freedom and contaminated normal distribution $(1 - \varepsilon)\Phi(x) + \Phi(x/\sigma)$ with $\varepsilon = 0.05$, and $\sigma^2 = 9$. The sample size T =100, 250, and 500 are considered for simulation purpose. In all the experiments, $R =$ 1000 independent replications are used. We are interested to investigate under different error distributions: the asymptotic and empirical standard errors of the squared and absolute residual autocorrelations and the finite sample performance of portmanteau tests, $Q(M)$ and $Q^a(M)$ when M-estimators are used for fitting GARCH models.

The parameters are estimated using the algorithms developed for M-estimators. All simulations are performed on MATLAB software. The asymptotic standard errors $A_i(i = 1, ..., 10)$, of the squared residual autocorrelations, $\hat{\mathbf{r}} = (\hat{r}_1, ..., \hat{r}_{10})'$, and absolute residual autocorrelations, $\hat{\boldsymbol{\rho}} = (\hat{\rho}_1, ..., \hat{\rho}_{10})$ are obtained, respectively, from the results in Section 4.3.1 and 4.3.2. The empirical standard errors $S_i(i = 1, ..., 10)$, of $\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\rho}}$ over 1000 replications are also obtained and considered as the "true" standard errors.

Both GJR(1,1) and GARCH(1,1) models are fitted and the results for the empirical standard errors and the averages of the asymptotic standard errors for lags 1, 2, 3, 7 and 10 when the errors are generated by standard normal distribution are presented in Table 4.1 and 4.2. For both models we find that the asymptotic standard errors for squared and absolute residual autocorrelations match the empirical standard errors quite satisfactory for all estimators.

The results for student-t distribution with 3 and 5 degrees of freedom and contaminated normal distribution when both GARCH(1,1) and GJR(1,1) models are fitted using M-estimators are not reported as we did not find any significant difference between the asymptotic and empirical standard errors in those cases.

4.4.2 Analysis of the Size and Power of Tests

We conduct Monte Carlo simulations to examine the empirical sizes of the portmanteau statistics $Q(M)$ and $Q^a(M)$. We use 1000 replications and the sample size of $T = 100$, 250, and 500 for all experiments. The data are generated from the following data generating processes (DGPs), denoted by M1 for GARCH(1,1), M2 and M3 for ARCH(2) model:

$$\begin{aligned}
\text{M1:}\quad X_t = h_t^{1/2}\epsilon_t,\quad h_t &= 0.4 + 0.4\,X_{t-1}^2 + 0.1\,h_{t-1}.\\
\text{M2:}\quad X_t = h_t^{1/2}\epsilon_t,\quad h_t &= 0.01 + 0.2\,X_{t-1}^2 + 0.7\,X_{t-2}^2.\\
\text{M3:}\quad X_t = h_t^{1/2}\epsilon_t,\quad h_t &= 0.01 + 0.7\,X_{t-1}^2 + 0.2\,X_{t-2}^2.
\end{aligned}$$

These DGPs with different values of the parameters are also used by Tsui (2004) to study the size and power of Li-Mak and other diagnostic tests. The conditional mean of each data generating process (DGP) is assumed to be zero. For each DGP, ϵ_t are generated from the standard normal distribution, student-t distribution with

Table 4.1: The empirical (S_i) and the large sample (A_i) standard errors of squared and absolute residual autocorrelations for the GJR(1,1) model using standard normal distribution

			lag i									
$\boldsymbol{\theta} = (0.01, 0.4, 0.1, 0.5)$			Squared Residuals					Absolute Residuals				
$N(0,1)$	T		1	2	3	7	10	1	2	3	7	10
	100	A	0.1130	0.1924	0.1025	0.1165	0.1357	0.3313	0.1924	0.1125	0.1151	0.1259
		S	0.0583	0.0862	0.0908	0.0913	0.0992	0.0561	0.0862	0.0928	0.0955	0.1021
QMLE	250	A	0.0961	0.0937	0.0880	0.0969	0.0980	0.0982	0.0942	0.0885	0.0976	0.0983
		S	0.0565	0.0852	0.0884	0.0996	0.0989	0.0551	0.0864	0.0905	0.1001	0.1023
	500	A	0.0287	0.0390	0.0406	0.0430	0.0439	0.0268	0.0388	0.0407	0.0432	0.0440
		S	0.0287	0.0396	0.0396	0.0430	0.0438	0.0268	0.0398	0.0403	0.0437	0.0437
	100	A	0.0861	0.0909	0.0855	0.0964	0.0979	0.0841	0.0917	0.0836	0.0963	0.0976
		S	0.0603	0.0866	0.0912	0.0893	0.0987	0.0485	0.0845	0.0930	0.0945	0.1020
LAD	250	A	0.0880	0.0897	0.0854	0.0959	0.0974	0.0866	0.0907	0.0835	0.0954	0.0972
		S	0.0601	0.0876	0.0892	0.0986	0.0980	0.0473	0.0864	0.0901	0.1001	0.1015
	500	A	0.0313	0.0395	0.0411	0.0433	0.0440	0.0233	0.0376	0.0398	0.0430	0.0439
		S	0.0316	0.0408	0.0403	0.0433	0.0439	0.0236	0.0400	0.0400	0.0435	0.0436
	100	A	0.0874	0.0894	0.0854	0.0960	0.0979	0.0867	0.0898	0.0844	0.0957	0.0978
		S	0.0576	0.0849	0.0881	0.0970	0.0979	0.0512	0.0824	0.0896	0.0993	0.1024
Huber's	250	A	0.0974	0.0932	0.0875	0.0972	0.0989	0.0984	0.0941	0.0873	0.0980	0.0996
		S	0.0600	0.0870	0.0878	0.0994	0.0988	0.0529	0.0869	0.0894	0.1003	0.1023
	500	A	0.0305	0.0394	0.0409	0.0432	0.0440	0.0247	0.0381	0.0402	0.0431	0.0440
		S	0.0318	0.0405	0.0401	0.0431	0.0441	0.0266	0.0400	0.0400	0.0435	0.0438
	100	A	0.1007	0.0980	0.0966	0.0995	0.0991	0.1026	0.0983	0.0918	0.0986	0.0979
		S	0.0800	0.0920	0.0947	0.0898	0.0979	0.0644	0.0873	0.0948	0.0946	0.1022
B-Estimator	250	A	0.0982	0.0962	0.0967	0.0988	0.0991	0.0972	0.0950	0.0921	0.0972	0.0980
		S	0.0823	0.0934	0.0924	0.0990	0.0977	0.0640	0.0891	0.0919	0.1008	0.1018
	500	A	0.0413	0.0432	0.0436	0.0443	0.0445	0.0309	0.0397	0.0412	0.0435	0.0442
		S	0.0428	0.0441	0.0430	0.0446	0.0445	0.0323	0.0422	0.0415	0.0440	0.0438
	100	A	0.1075	0.1016	0.0987	0.1005	0.0995	0.1241	0.1047	0.0927	0.1004	0.0993
		S	0.0834	0.0929	0.0944	0.0888	0.0977	0.0669	0.0880	0.0940	0.0949	0.1019
Cauchy	250	A	0.1088	0.1004	0.0999	0.1010	0.1005	0.1161	0.0998	0.0935	0.0991	0.0993
		S	0.0849	0.0967	0.0934	0.0984	0.0981	0.0645	0.0904	0.0922	0.1004	0.1019
	500	A	0.0436	0.0441	0.0443	0.0446	0.0446	0.0325	0.0399	0.0412	0.0436	0.0442
		S	0.0465	0.0446	0.0436	0.0444	0.0443	0.0336	0.0421	0.0414	0.0438	0.0436

Table 4.2: The empirical (S_i) and the large sample (A_i) standard errors of squared and absolute residual autocorrelations for the GARCH(1,1) model using standard normal distribution

			lag i									
$\theta = (0.01, 0.4, 0.5)$			Squared Residuals					Absolute Residuals				
$N(0,1)$	T		1	2	3	7	10	1	2	3	7	10
	100	A	0.1116	0.0991	0.0907	0.1001	0.1030	0.1145	0.0997	0.0916	0.1009	0.1029
		S	0.0610	0.0862	0.0900	0.0985	0.0959	0.0589	0.0890	0.0912	0.1023	0.1017
QMLE	250	A	0.0415	0.0528	0.0534	0.0608	0.0622	0.0404	0.0527	0.0539	0.0611	0.0624
		S	0.0401	0.0560	0.0553	0.0588	0.0619	0.0383	0.0552	0.0580	0.0609	0.0623
	500	A	0.0123	0.0261	0.0285	0.0308	0.0313	0.0142	0.0267	0.0290	0.0310	0.0313
		S	0.0190	0.0282	0.0279	0.0309	0.0303	0.0208	0.0283	0.0296	0.0310	0.0318
	100	A	0.0812	0.0875	0.0824	0.0954	0.0975	0.0787	0.0885	0.0810	0.0947	0.0973
		S	0.0653	0.0856	0.0900	0.0980	0.0947	0.0507	0.0869	0.0898	0.1024	0.1007
LAD	250	A	0.0448	0.0533	0.0541	0.0611	0.0623	0.0378	0.0516	0.0518	0.0608	0.0622
		S	0.0327	0.0546	0.0570	0.0609	0.0619	0.0417	0.0568	0.0558	0.0595	0.0616
	500	A	0.0165	0.0269	0.0290	0.0309	0.0313	0.0113	0.0260	0.0286	0.0309	0.0313
		S	0.0216	0.0287	0.0287	0.0310	0.0305	0.0183	0.0278	0.0295	0.0308	0.0318
	100	A	0.0961	0.0931	0.0859	0.0968	0.0996	0.0979	0.0947	0.0855	0.0970	0.0998
		S	0.0631	0.0859	0.0900	0.0976	0.0949	0.0530	0.0881	0.0906	0.1022	0.1009
Huber's	250	A	0.0469	0.0541	0.0540	0.0609	0.0623	0.0425	0.0530	0.0526	0.0608	0.0623
		S	0.0417	0.0568	0.0558	0.0595	0.0616	0.0338	0.0553	0.0577	0.0612	0.0620
	500	A	0.0165	0.0267	0.0287	0.0309	0.0313	0.0137	0.0263	0.0285	0.0310	0.0314
		S	0.0209	0.0285	0.0283	0.0312	0.0303	0.0196	0.0279	0.0294	0.0311	0.0317
	100	A	0.1039	0.0986	0.0965	0.0999	0.0993	0.1089	0.0998	0.0912	0.0975	0.0991
		S	0.0873	0.0901	0.0916	0.0975	0.0939	0.0658	0.0887	0.0904	0.1020	0.1002
B-Estimator	250	A	0.0587	0.0600	0.0609	0.0627	0.0630	0.0489	0.0557	0.0557	0.0616	0.0624
		S	0.0600	0.0619	0.0588	0.0598	0.0611	0.0454	0.0573	0.0583	0.0611	0.0615
	500	A	0.0285	0.0302	0.0309	0.0315	0.0315	0.0198	0.0274	0.0294	0.0312	0.0314
		S	0.0307	0.0312	0.0307	0.0312	0.0310	0.0230	0.0285	0.0303	0.0307	0.0321
	100	A	0.1322	0.1076	0.1011	0.1042	0.1053	0.1735	0.1188	0.0977	0.1029	0.1067
		S	0.0902	0.0907	0.0933	0.0966	0.0942	0.0686	0.0886	0.0909	0.1022	0.1003
Cauchy	250	A	0.0618	0.0620	0.0622	0.0630	0.0631	0.0521	0.0563	0.0554	0.0616	0.0625
		S	0.0624	0.0627	0.0600	0.0600	0.0609	0.0471	0.0582	0.0590	0.0612	0.0618
	500	A	0.0305	0.0311	0.0313	0.0316	0.0316	0.0209	0.0274	0.0293	0.0312	0.0314
		S	0.0332	0.0322	0.0313	0.0315	0.0312	0.0246	0.0290	0.0304	0.0308	0.0321

3 and 5 degrees of freedom and the contaminated normal distribution. We also generate error from Hansen's skewed-t distribution with 4 degrees of freedom and skewness parameter 0.25. The skewed-t distribution is chosen to observe the effect of asymmetry on the size of tests. The GARCH(1,1) model is fitted to each DGP using M-estimators and the portmanteau statistics $Q(M)$ and $Q^a(M)$, of squared and absolute autocorrelation of residuals, respectively, are computed. The rejection frequency represents the estimated size of the test when the underlying DGP is M1, while for M2 and M3, it represents the estimated power.

Although, like residual-based diagnostics, these portmanteau tests have no specific alternatives, it should not be construed that the portmanteau tests are consistent against all model misspecifications. For empirical power, we use ARCH(2) models as our DGPs, other DGPs such as high order ARCH models may also be used.

Table 4.3 reports the proportion of rejection of $Q(M)$ and $Q^a(M)$, based on the upper 5th percentile of the corresponding asymptotic χ^2 distribution. We chose a lag length, $M = 6$ for this study. It can be seen from Table 4.3 that $Q(M)$ test generally has a reasonable size under the standard normal distribution and student-t distribution with 5 df. The size of $Q^a(M)$ is also found close to the nominal level under these distributions.

Under the heavy-tailed distribution, student-t distribution with 3 degrees of freedom in this case, the empirical significance level of $Q(M)$ is found slightly greater than the nominal size of 5%. This is found for all sample size considered in this experiment. Both the tests over reject the null hypothesis when QMLE is used to fit the GARCH model. The size of $Q^a(M)$ is close to the nominal size in case of $t(5)$ for all the estimators other than the QMLE. In the case of the contaminated normal distribution, both tests over reject for all sample sizes but again some of the estimators show better results than the QMLE.

When errors are generated from skewed-t distribution, the test based on squared residual autocorrelation is found to be consistently undersized as compared to the test based on absolute residual autocorrelation. The $Q^a(M)$ test, when estimators other than the QMLE is used for GARCH model fitting, show better empirical size. Increasing the sample size improves the size slightly.

In general, based on the empirical sizes reported in Table 4.3, we conclude that although the size of both tests may be considered reliable, $Q^a(M)$ has a slight edge over $Q(M)$ when the data are generated from a heavy-tailed or skewed distribution. Moreover, estimators other than the QMLE can be considered adequate for GARCH models under non-normal distributions.

Next, in our second experiment we check the empirical power of Q and Q^a. The empirical power of both diagnostic tests are obtained from the rejection frequency when data are generated from M2 and the GARCH(1,1) model using M-estimators is fitted. Table 4.4 below shows the result of this study. The results for standard normal errors suggest that QMLE has the highest power as compared to other estimators. For student-t distribution with 3 degrees of freedom, B-estimator shows the best results with LAD and Cauchy estimator also provide good empirical powers. For the case of student-t distribution with 5 degrees of freedom, LAD outperforms other estimators.

Both tests show low power when errors are generated from contaminated normal distribution. In this case, B-estimator can be considered as the best estimator for fitting GARCH model. When errors are generated from skewed-t distribution, the feature we observed for non-normal distributions seems to hold here as well with the QMLE showing the lowest power and other M-estimators provide reasonable results. Another feature we notice from this study is that the test based on absolute residual autocorrelations show better empirical powers than the test

Table 4.3: The empirical sizes (in percentage) of $Q(M)$ and $Q^a(M)$ for $M = 6$; $\phi = 5\%$

	T	QMLE $Q(M)$	QMLE $Q\ (M)$	LAD $Q(M)$	LAD $Q\ (M)$	Huber's $Q(M)$	Huber's $Q\ (M)$	B-estimator $Q(M)$	B-estimator $Q\ (M)$	Cauchy $Q(M)$	Cauchy $Q\ (M)$
$N(0,1)$											
	100	4.1	3.3	3.8	3.3	3.1	3.5	3.8	3.0	3.5	4.4
	250	4.3	3.9	3.7	3.6	3.5	3.3	3.6	3.7	3.6	4.3
	500	4.4	4.5	4.0	4.5	4.3	4.6	3.4	3.6	3.9	5.5
$t(3)$											
	100	8.7	7.7	5.4	4.3	5.3	4.5	5.7	4.8	5.5	4.7
	250	8.8	7.7	5.6	4.2	5.4	5.3	5.2	4.5	6.1	4.6
	500	7.4	7.0	5.1	5.5	5.9	5.8	5.4	5.3	5.6	5.2
$t(5)$											
	100	6.1	5.7	4.6	3.8	5.0	4.2	4.6	4.0	4.3	4.7
	250	5.7	5.6	5.1	4.6	5.9	4.5	4.4	4.9	4.4	5.4
	500	5.7	5.3	6.0	4.9	6.0	4.7	5.4	5.2	5.1	5.0
$CN(0.05, 9)$											
	100	7.9	7.2	6.7	6.3	6.7	6.0	6.3	6.1	6.6	6.1
	250	6.4	6.1	6.4	6.2	6.6	6.1	5.9	6.1	6.6	6.0
	500	6.5	6.1	6.4	6.0	6.1	5.8	5.6	5.6	6.5	5.7
$ST(4, 0.25)$											
	100	2.5	3.2	2.9	4.3	2.5	3.0	3.3	3.4	3.1	3.2
	250	3.4	4.1	3.4	4.4	3.6	4.1	3.9	4.1	3.4	4.1
	500	3.5	4.1	3.4	4.5	4.1	4.8	3.9	4.2	3.5	4.5

$Q(M)$ and $Q\ (M)$ are portmanteau statistics for squared and absolute residual autocorrelations, respectively, for lag M.

Table 4.4: The empirical powers (in percentage) of $Q(M)$ and $Q^a(M)$ for $M = 6$; $\phi = 5\%$

	T	QMLE $Q(M)$	QMLE $Q\ (M)$	LAD $Q(M)$	LAD $Q\ (M)$	Huber's $Q(M)$	Huber's $Q\ (M)$	B-estimator $Q(M)$	B-estimator $Q\ (M)$	Cauchy $Q(M)$	Cauchy $Q\ (M)$
$N(0,1)$											
	100	25.6	37.1	23.2	28.3	24.1	30.0	23.1	33.3	21.1	33.4
	250	66.4	67.1	61.7	62.2	60.1	60.7	59.1	60.0	55.6	55.4
	500	74.4	77.2	70.1	74.2	70.9	74.1	71.1	73.6	68.8	71.1
$t(3)$											
	100	6.9	7.1	12.4	13.1	12.5	14.1	13.7	14.2	13.5	13.1
	250	13.3	26.3	16.1	54.1	17.1	55.9	15.5	56.1	15.3	55.8
	500	15.1	47.2	20.2	65.1	19.9	61.3	22.2	66.0	20.6	65.0
$t(5)$											
	100	11.0	11.7	18.4	25.1	17.5	24.9	17.1	23.8	16.1	22.7
	250	25.5	45.2	44.4	56.6	41.4	55.1	44.5	54.0	42.1	53.8
	500	33.3	55.1	55.5	67.9	51.5	65.4	55.1	65.5	55.0	63.5
$CN(0.05,9)$											
	100	5.9	6.1	8.9	11.3	6.1	6.2	9.4	12.4	8.8	12.1
	250	5.4	6.2	11.1	11.4	6.4	7.3	13.1	14.4	11.1	13.6
	500	8.5	8.4	12.3	13.1	8.0	10.8	16.5	17.22	15.2	15.8
$ST(4,0.25)$											
	100	17.2	19.1	22.1	25.4	18.1	25.0	18.2	26.1	16.4	23.1
	250	24.1	26.3	33.4	44.2	31.1	42.2	35.3	52.2	43.4	48.7
	500	33.3	45.6	51.3	59.1	48.9	54.1	52.2	64.1	53.5	59.8

$Q(M)$ and $Q\ (M)$ are portmanteau statistics for squared and absolute residual autocorrelations, respectively, for lag M.

based on squared residual autocorrelations especially for non-normal errors.

We note that the powers of the proposed tests, when Cauchy and B-estimators are fitted to data, are found higher than other estimators. These findings indicate that estimators such as Cauchy and B-estimator are more adequate than the competing estimators especially when errors are non-normal. These results support the suggestion of using B-estimators for fitting GARCH-type models.

The finite sample performance of the size and power of the tests may vary with the number of lag correlation coefficients taken. To examine this effect, we perform another study and consider $M = 1, 2, ..., 25$. The rejection frequency represents the estimated power of the tests when the underlying DGP is M3 and the GARCH(1,1) is fitted using M-estimators. For the errors, we use the same distributions as in previous studies and the graphs of empirical powers of both tests are plotted against M. The sample sizes considered are 100 and 500.

Fig. 4.1 below presents the empirical powers of Q and Q^a for $T = 100$, when errors are normally distributed. It is easy to see that the empirical power of the test based on absolute residual is greater than the test based on squared residuals. The power levels of Q^a under Cauchy and B-estimators are very close to the power level of the QMLE for normal errors. We also observe that both tests under the LAD and Huber's estimator show low power as compared to other three estimators considered. Another feature we notice is that the power of both the tests reaches its highest value at $M = 2$ and decreases for higher lags.

For $T = 500$, Fig. 4.2 below shows the same features we observed for small sample. The empirical power of both tests decline when the value of M increases and both tests produce the highest empirical power at $M = 2$. This suggests that the choice of the lag of sample autocorrelation, (M), can affect the power of these tests. We again note for this sample size that the power of tests when Cauchy and

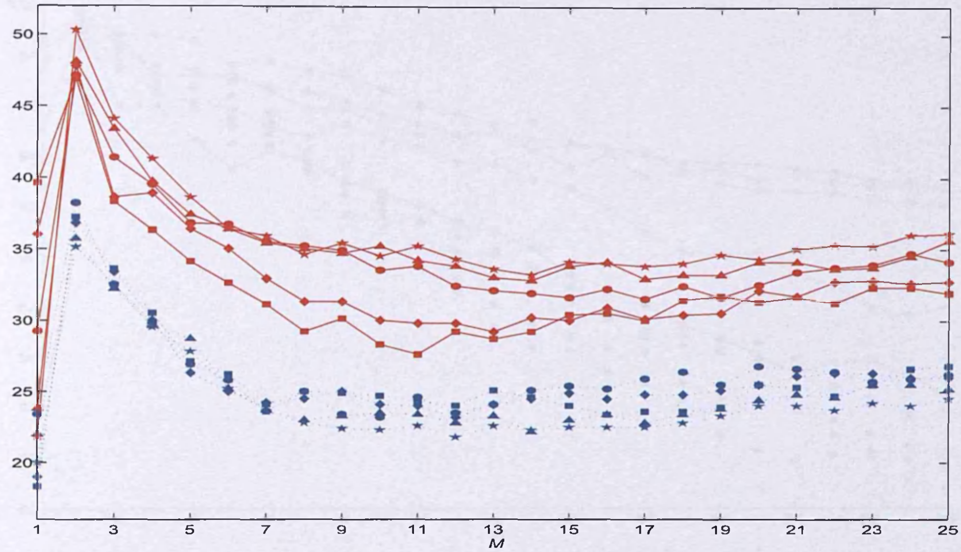

Figure 4.1: Power levels of Q (dotted) and Q^a (solid) under QMLE (•), LAD (■), Huber's (♦), B-estimator (▲) and Cauchy (★): $\epsilon_t \sim N(0,1)$, $T = 100$.

B-estimators are used for estimation match that of the QMLE.

In case where the errors are generated from the standardised student-t distribution with 3 df, the results presented in Figs. 4.3 and 4.4 reveal that both tests posses low empirical powers. For $T = 100$ the test based on absolute residuals, under Cauchy and B-estimators, shows empirical power greater than the power of other estimators. The power levels of this test slightly increases for higher lags. This feature is not observed when sample size is increased to 500. In this case the power levels of both tests decline as M increases. For this sample size, The QMLE and LAD show slightly lower powers than other estimators. Moreover, the important evidence is the big difference between the power levels of Q and Q^a. This show the superiority of the test based on absolute residuals over the test based on squared residuals for heavy-tailed distributions.

The findings in the case of student-t distribution with 3 degrees of freedom remain valid for the same distribution for 5 degrees of freedom. Again for small sample size, the empirical power of Q^a under Cauchy and B-estimators are found

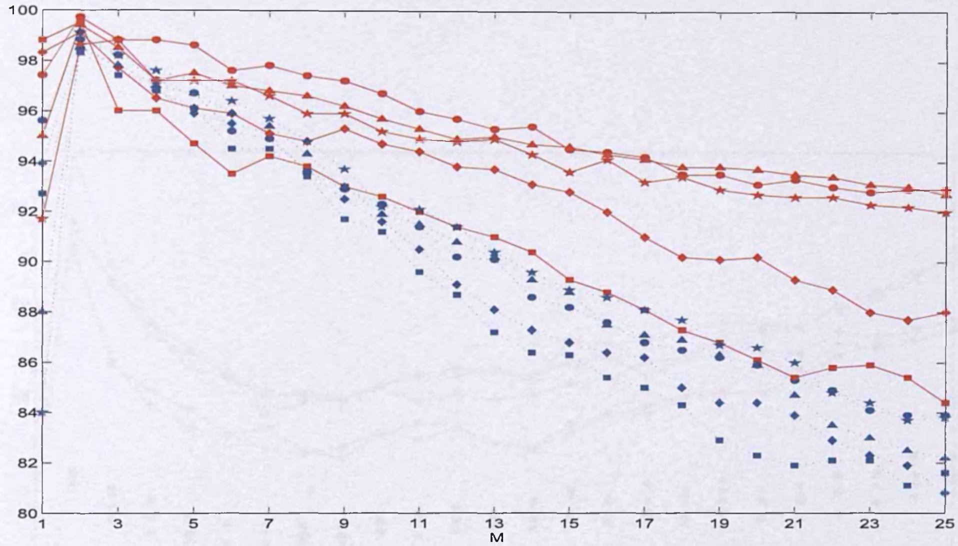

Figure 4.2: Power levels of Q (dotted) and Q^a (solid) under QMLE (•), LAD (■), Huber's (♦), B-estimator (▲) and Cauchy (★): $\epsilon_t \sim N(0,1)$, $T = 500$.

greater and for large sample size, the dominance of Q^a over Q is significant. The only difference we observed is the gain in power. In case of $t(5)$, for large sample size, the power at $M = 2$, is found close to 96% for Q^a, whereas it dropped to 83% for $t(3)$.

The power results of Q and Q^a are displayed in Figs. 4.7 and 4.8 when errors are generated from the contaminated normal distribution. For $T = 100$, the empirical powers of Q^a based on B-estimator and LAD are found better than those of other estimators with the QMLE showing very low power. When the sample size is increased this feature can be observed more clearly. The power levels of Q^a drop as in the previous cases and reach the level of Q for higher lags.

In case of skewed-t distribution, the power plots show the same characteristics as observed in case of $t(5)$, with Q^a out performing Q in terms of empirical power and Cauchy and B-estimator slightly producing better results. Hence, we do not plot those graphs.

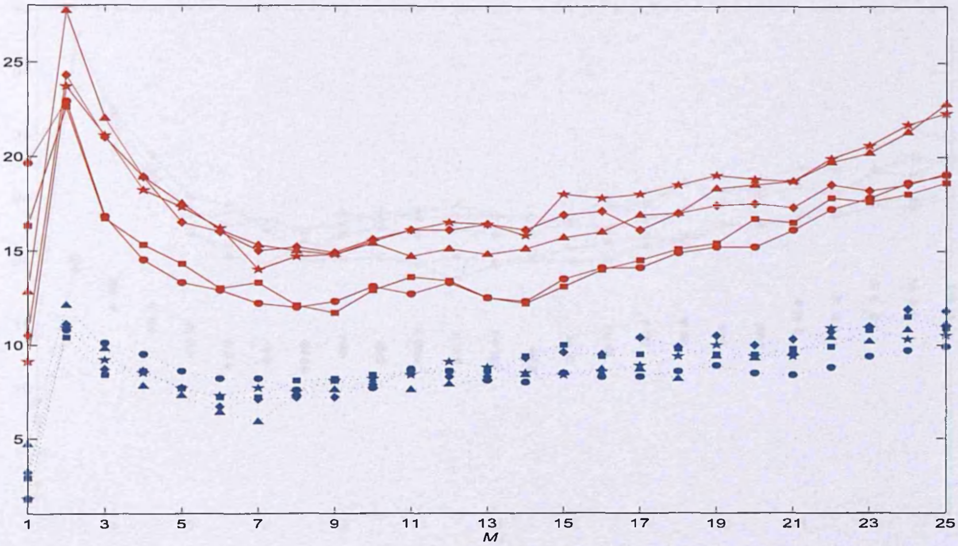

Figure 4.3: Power levels of Q (dotted) and Q^a (solid) under QMLE (●), LAD (■), Huber's (♦), B-estimator (▲) and Cauchy (★): $\epsilon_t \sim t(3)$, $T = 100$.

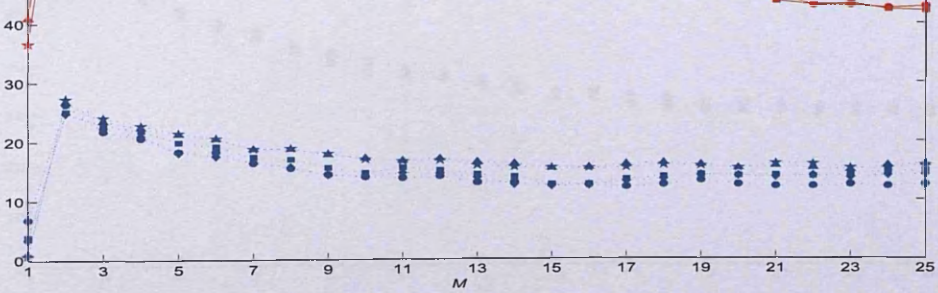

Figure 4.4: Power levels of Q (dotted) and Q^a (solid) under QMLE (●), LAD (■), Huber's (♦), B-estimator (▲) and Cauchy (★): $\epsilon_t \sim t(3)$, $T = 500$.

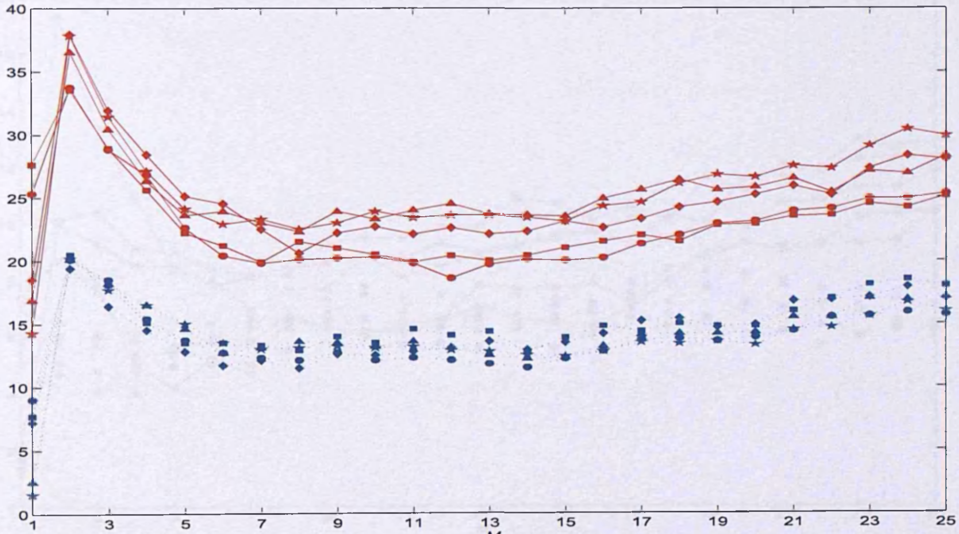

Figure 4.5: Power levels of Q (dotted) and Q^a (solid) under QMLE (•), LAD (■), Huber's (♦), B-estimator (▲) and Cauchy (★): $\epsilon_t \sim t(5)$, $T = 100$.

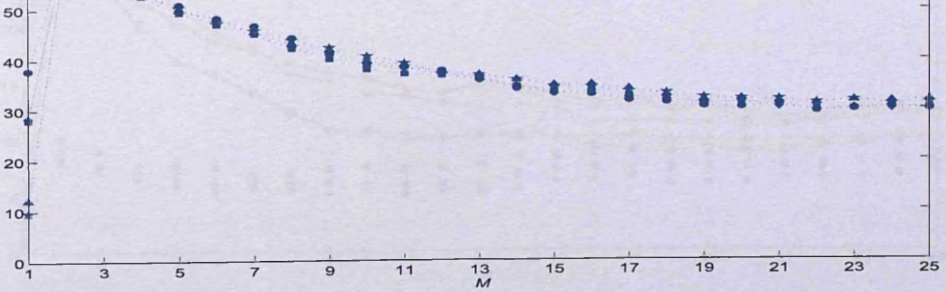

Figure 4.6: Power levels of Q (dotted) and Q^a (solid) under QMLE (•), LAD (■), Huber's (♦), B-estimator (▲) and Cauchy (★): $\epsilon_t \sim t(5)$, $T = 500$.

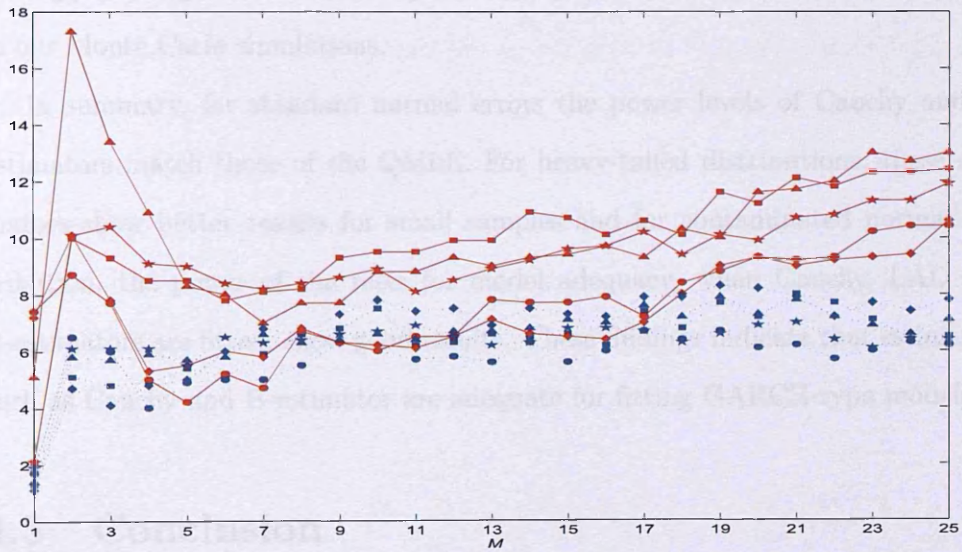

Figure 4.7: Power levels of Q (dashed line) and Q^a (solid line) under QMLE (•), LAD (■), Huber's (♦), B-estimator (▲) and Cauchy (★): $\epsilon_t \sim CN(0.05, 9)$, $T = 100$.

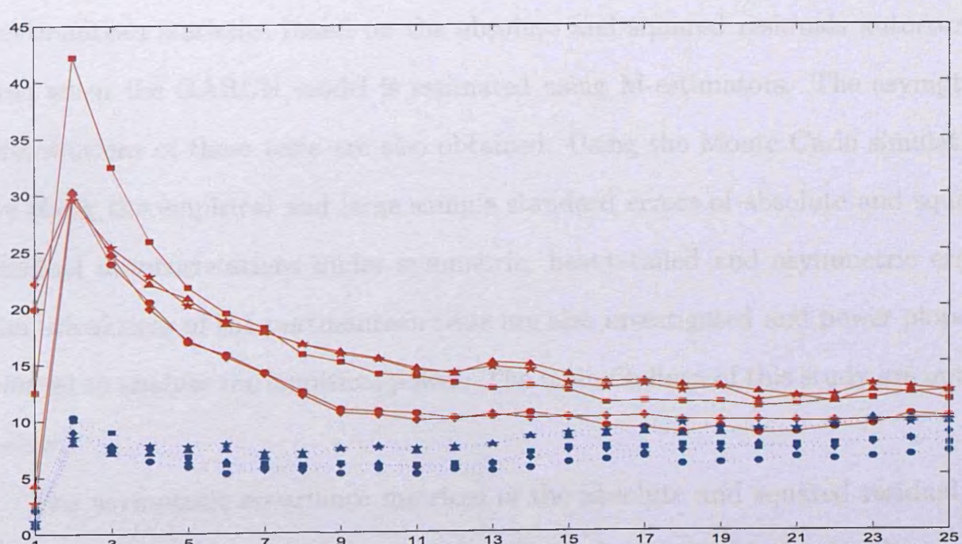

Figure 4.8: Power levels of Q (dashed line) and Q^a (solid line) under QMLE (•), LAD (■), Huber's (♦), B-estimator (▲) and Cauchy (★): $\epsilon_t \sim CN(0.05, 9)$, $T = 500$.

We also note from this study that for size accuracy a fairly large M is required, although a smaller M may be desirable for power improvements. This shows that the choice of M induces a trade-off between size and power. Tse and Zuo (1997) suggested $M = p+q+1$ as an appropriate choice. Based on our limited experience we suggest using $M = 6$ as this lag value provides reasonable size and good power in our Monte Carlo simulations.

In summary, for standard normal errors the power levels of Cauchy and B-estimators match those of the QMLE. For heavy-tailed distributions, these estimators show better results for small samples and for contaminated normal distribution, the power of the tests for model adequacy, when Cauchy, LAD and B-estimators are fitted, show good results. These findings indicate that estimators such as Cauchy and B-estimator are adequate for fitting GARCH-type models.

4.5 Conclusion

This chapter deals with the diagnostic checking of GARCH models. We develop portmanteau statistics based on the absolute and squared residuals autocorrelation when the GARCH model is estimated using M-estimators. The asymptotic distributions of these tests are also obtained. Using the Monte Carlo simulations we check the empirical and large sample standard errors of absolute and squared residual autocorrelations under symmetric, heavy-tailed and asymmetric errors. Empirical sizes of the portmanteau tests are also investigated and power plots are plotted to analyse the empirical power. The main findings of this study are outline below.

The asymptotic covariance matrices of the absolute and squared residual autocorrelations are derived. The correct asymptotic standard errors for these autocorrelations are obtained and these give more accurate standard errors than $1/\sqrt{T}$

for the residual autocorrelations. Based on these results, new portmanteau tests are developed for diagnostic checking of GARCH models.

Simulation results show that the asymptotic standard errors for both squared and absolute residual autocorrelations match the empirical standard errors quite satisfactory for all estimators.

The empirical sizes of both tests are found close to the nominal sizes in most of the cases. When errors are generated from heavy-tailed or asymmetric distribution, the empirical power of the test based on absolute residual autocorrelations is found slightly better than the test based on squared residual autocorrelations. Furthermore, the empirical powers of competing estimators are found higher than the QMLE for non-normal errors. Our study suggests the use of robust estimators such as Cauchy and B-estimator for fitting GARCH-type models.

Chapter 5

Value-at-Risk Based on M-estimators for GARCH-type Models

5.1 Introduction

Risk management is one of the important tasks for financial institutions, nonfinancial corporations, regulators and asset managers. Value-at-Risk (VaR) and Expected Shortfall (ES) are commonly-used statistics for measuring potential risk of economic losses in financial market. VaR is the quantile of the loss that can occur within a given portfolio during a specified time period while ES is the expected loss, given that loss is at least as large as some given VaR. A precise quantile estimate far out in the left tail of the return distribution is needed for univariate VaR measures; see Jorion (2000) for a general introduction and exposition of VaR. Value-at-Risk has been widely used for financial risk management by institutions including banks, regulators and portfolio managers.

Existing approaches for estimating VaR may be classified into three approaches.

These are the nonparametric, semi-parametric and parametric approaches. One of the popular nonparametric approaches is historical simulation (see, e.g., Hendricks, 1996), which computes the empirical quantiles of historical portfolio returns. In the semi-parametric approaches are the GARCH based extreme value theory (EVT) approach (McNeil and Frey, 2000), which specifically model the tails of the distribution of residuals and quantile regression approach (Engle and Manganelli, 2004), which directly models a specific quantile rather than the whole distribution. See Kuester et al. (2006) and the references therein for an overview and comparison of these and further models.

Among the parametric approaches are the J.P. Morgan's RiskMetrics (1996) and GARCH models. Under the RiskMetrics approach the variance is estimated based on the exponentially weighted moving average method. The GARCH-based approach is also popular as it provides a parsimonious model with few parameters which usually fit econometric time series very well. The GARCH approach first fits the GARCH-type models for financial returns series and then models the residuals form these models based on the assumptions of the conditional distributions of the residuals. Angelidis et al. (2004) evaluated the performance of an extensive family of GARCH-type models in modelling daily VaR.

Empirical evidences show that often the distribution of asset returns is skewed and heavy-tailed. This implies that extreme events are much more likely to occur in practice than would be predicted by the normal distribution. Hence, value-at-risk estimates produced by the model under normality assumption may not be reliable and accurate.

The main aim of this chapter is to propose not only robust measures of VaR based on GARCH-type models, but also more reliable information on risk estimates. To achieve this aim. we present, develop and empirically test VaR estimates from GARCH-type models when M-estimators are used for estimation. Risk estimates from both symmetric and asymmetric GARCH models are obtained. The performance of the proposed VaR estimates is extensively studied for three important financial data sets (S&P500, FTSE100, NIKKEI225). Both in-sample and out-of-sample VaR estimates are evaluated. The accuracy of the proposed VaR estimates is discussed using a number of newly-introduced M-test statistics.

Our study is important from a number of different angles. Since QMLE is a member of the class of M-estimators, in many senses, our method is applicable to most of the previous studies on the VaR evaluations using the QMLE. In addition, since we use nonparametric setup for the error distributions and some of the M-estimators of the GARCH parameters used for the VaR evaluation is consistent and asymptotically normal under minimal moment assumption related to merely finite second moment of the innovations, our estimators are expected to perform well for those financial data for which the use of the QMLE cannot be justified due to lack of fourth moment. In fact, our empirical study indicates that in most of the cases estimators such as Cauchy and B-estimators predict the VaR more accurately than the frequently-used QMLE. These findings strengthen the use of M-estimators for fitting GARCH models and predicting VaR. A comparison of the results for both GARCH and GJR model shows that for data sets used in this study, the GJR model provides better forecasts for risk. Our study may help risk managers to select appropriate estimator for predicting VaR and to use range of summary statistics for the evaluation of risk estimates.

Section 5.2 describes value-at-risk estimation for GARCH-type models fitted by M-estimators. Some evaluation measures used to assess the VaR predicting performance are introduced in Section 5.3. In Section 5.5, the in-sample and out-of-sample VaR estimates of three major stock indices are obtained using both GARCH and GJR models. Backtesting and evaluation methods are applied to check the accuracy and efficiency of these estimates. Section 5.6 concludes this chapter.

5.2 Value-at-Risk (VaR)

Value-at-Risk or VaR is the p-th conditional quantile of the distribution of the change in value of an asset over a certain period of time where p is known and close to zero. It is an estimate of the maximal loss associated with a given probability p and is used by financial institutions and regulators for risk measurement.

Hence for the returns $\{X_t; 1 \leq t \leq T\}$ of a portfolio, the VaR at time $t > 1$, denoted by $q_t = q_t(p)$, is defined by

$$q_t = \inf\{x; p \leq P_{t-1}(X_t \leq x)\},$$

where P_{t-1} is the conditional distribution of X_t given the information available up to time $t-1$. When returns are of the form (3.14), we get

$$q_t = v_t^{1/2}(\boldsymbol{\theta}_0)F^{-1}(p),$$

where F^{-1} is the quantile function of the errors $\{\epsilon_t\}$. Hence, using (3.15) and (3.18) we get

$$q_t = \frac{1}{c_H^{1/2}} v_t^{1/2}(\boldsymbol{\theta}_{0H})F^{-1}(p), \tag{5.1}$$

where notice that $F^{-1}(p)$ is the p-th quantile of the scaled errors $\{\epsilon_t/c_H^{1/2}\}$. Estimating $v_t^{1/2}(\boldsymbol{\theta}_{0H})$ by $\hat{v}_t^{1/2}(\hat{\boldsymbol{\theta}}_T)$ and $F^{-1}(p)$ by the p-th quantile of the residuals $\{X_t/\{\hat{v}_t(\hat{\boldsymbol{\theta}}_T)\}^{1/2}; 1 \leq t \leq T\}$, we obtain (5.1) the predicted value $\hat{q}_t$ of q_t as

$$\hat{q}_t = \hat{v}_t^{1/2}(\hat{\boldsymbol{\theta}}_T) \times \big([Tp]+1\big)\text{-th order statistics of } \{X_t/\{\hat{v}_t(\hat{\boldsymbol{\theta}}_T)\}^{1/2}\}, 2 \le t \le T. \tag{5.2}$$

Clearly $\hat{q}_t$ depends on the underlying M-estimates.

5.3 M-tests of VaR

Next, we define some performance measures and M-tests for the evaluation of risk estimates produced by M-estimators. Due to the importance of VaR estimates to banks and regulators, evaluating the accuracy of VaR estimates is a necessary exercise. One approach to improve the accuracy of VaR is through backtesting, which is continual statistical testing of the accuracy of VaR estimates. Backtesting helps to identify the advantages and disadvantages in each model. It can also be used to detect incorrectly specified VaR models and sources of inaccuracy of VaR forecasts. Value-at-Risk models and their accuracy can be compared through backtesting to select adequately accurate models in market risk management. Backtesting is also required by the regulatory bodies such as the Basel Committee on Banking Supervision of the Bank of International Settlements. Basel Committee on Banking Supervision (1996) has set up standards for the quality of VaR data in the backtesting framework.

5.3.1 Coverage M-Test

Let

$$T_* = \sum_{t=2}^{T} I_t \text{ with } I_t = I(X_t \le \hat{q}_t)$$

denote the total number of violations. The closeness of the empirical rejection probability

$$\hat{p} = T_*/T \tag{5.3}$$

to 'p' can be used to assess the overall predictive performance of the underlying conditional heteroscedastic model and the M-estimates used for computing $\hat{q}_t$. We propose below two statistical tests for the null hypothesis $E(T_*/T) = p$ against its negation, as they are related to the model validity. Note that the following statistics are also defined in Section 3.5.

According to Kupiec (1995), the number of violations T_* follows a binomial distribution with probability p. The probability of experiencing T_* violations in a sample of T observations, if the model is correct, is given by:

$$\mathrm{P}(T_*; T, p) = \begin{pmatrix} T \\ T_* \end{pmatrix} p^{T} \, (1-p)^{T-T}$$

Under the null hypothesis the unconditional likelihood ratio test statistic is given by

$$\mathrm{LR}_{uc} = 2\Big[\ln\left\{ (1-\hat{p})^{T-T} \, \hat{p}^{T} \right\} - \ln\left\{ (1-p)^{T-T} \, p^{T} \right\} \Big].$$

Kupiec (1995) proposed this statistic when the QMLE is used as $\hat{\boldsymbol{\theta}}_T$ and the test statistics is asymptotically $\chi^2_{(1)}$. The inspection of LR_{uc} reveals that if the proportion of VaR violations, $\hat{p} \times 100\%$, is exactly equal to $p \times 100\%$ then the test statistic takes the value zero, indicating no evidence of any inadequacy in the underlying VaR measure. A problem with this test is that it may fail to detect VaR measures that exhibit correct unconditional coverage but exhibit dependent VaR violations. VaR models that violate the independence property may result in losses that exceed the reported VaR in clusters.

Note however that in a reasonable model of VaR, the previous history of violations should not convey any information about whether or not additional VaR violations may occur in future. Towards that, using the QMLE as $\hat{\boldsymbol{\theta}}_T$, Christoffersen (1998) defined the independence coverage test statistic, denoted by LR_{ind}, which characterizes the ways in which these violations occur as follows.

For $i, j = 0, 1$, let T_{ij} be the number of time points $\{t; 2 \leq t \leq T\}$ for which $I_t = i$ is followed by $I_{t+1} = j$. Let

$$\hat{\pi}_{ij} = T_{ij}/(T_{i0} + T_{i1}), \;\; \hat{\pi} = (T_{01} + T_{11})/T.$$

Then

$$\text{LR}_{ind} = 2\Big[\ln\big((1-\hat{\pi}_{01})^{T} \;\; \hat{\pi}_{01}^{T} \; (1-\hat{\pi}_{11})^{T} \;\; \hat{\pi}_{11}^{T} \;\big) - \ln\big((1-\hat{\pi})^{(T \;\; +T \;\;)}\hat{\pi}^{(T \;\; +T \;\;)}\big)\Big].$$

Since both the unconditional coverage and the independence properties should be satisfied for an accurate VaR model, Christoffersen (1998) proposed the conditional coverage statistic

$$\text{LR}_{cc} = \text{LR}_{uc} + \text{LR}_{ind}$$

which is asymptotically $\chi^2_{(2)}$. We propose the same test statistics when $\{\hat{q}_t\}$'s are evaluated using M-estimates.

There are two limitations of Christoffersen's test. First, independence is tested against a very particular form which does not take into account dependences of order higher than one. Second, the use of a Markov chain makes it possible only to measure the influence of past violations and not that of any other exogenous variable. The tests proposed by Engle and Manganelli (2004) overcome these two limitations.

5.3.2 Dynamic Quantile M-Test

Since the LR_{cc} test only checks the first order dependence in the risk estimates, it is also desirable to check the high order dependence. Another test to check the high order dependence among $\{I_t\}$'s when the QMLE is used as $\hat{\boldsymbol{\theta}}_T$ is the Dynamic Quantile (DQ) test of Engle and Manganelli (2004). This test can be used to check the high order dependence in risk estimates. In this case the indicator variable Hit_t, $2 \le t \le T$, is defined by

$$\mathrm{Hit}_t = \begin{cases} 1-p & \text{if } X_t \le \hat{q}_t, \\ -p & \text{if } X_t > \hat{q}_t \end{cases}$$

and $\mathrm{Hit}_1 = -p$. Engle and Manganelli (2004) suggest to test jointly that $E(\mathrm{Hit}_t) = 0$, and that Hit_t is uncorrelated with variables included in the information set. This can be conducted by using the regression $\mathrm{Hit}_t = \mathbf{X}\boldsymbol{\beta} + e_t$, where $\mathbf{X}$ is a $T \times k$ matrix with ones in the first column, and the remaining columns are lagged values of Hit_t and some additional explanatory variables such as the current VaR. It is shown that under the null hypothesis, the dynamic quantile test is

$$\mathrm{DQ} = \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}}{p(1-p)},$$

where $\hat{\boldsymbol{\beta}}$ is the ordinary least squares (OLS) estimate of $\boldsymbol{\beta}$. The DQ test has an asymptotic chi-square distribution with k degrees of freedom.

5.3.3 Sign M-Test

Sarma et al. (2003) applied a one-sided nonparametric test. The null hypothesis of this test is that both models under consideration have same forecasting accuracy against a one-sided alternative hypothesis of superiority of one model over the other. Consider two VaR models, model i and model j. The null hypothesis is:

$$H_0 : \delta = 0$$

against the one-sided alternative hypothesis:

$$H_A : \delta < 0,$$

where δ is defined as the median of the distribution of the loss deferential, $\{dL_t;\ 1 \leq t \leq T\}$, where $dL_t = L_{it} - L_{jt}$, with L_{it} and L_{jt} are the values of a loss function for model i and model j, respectively, for day t. Negative values of dL_t indicates the superiority of model i over j.

Now, define an indicator variable s_t, such as

$$s_t = \begin{cases} 1 & \text{if } dL_t \geq 0, \\ 0 & \text{if } dL_t < 0. \end{cases}$$

The sign statistic is then given by

$$S_{ij} = \sum_{t=1}^{T} s_t$$

Under the null hypothesis, if dL_t is i.i.d., S_{ij} is binomially distributed with parameters $(T, 0.5)$. The standardized S_{ij} is asymptotically standard normal and given as

$$S^a_{ij} = \frac{(S_{ij} - 0.5T)}{\sqrt{0.25T}} \sim N(0,1) \text{ asymptotically.}$$

We can reject the null hypothesis at the 5% level of significance if $S^a_{ij} < -1.645$. Rejection of the null hypothesis means that model i is significantly better than model j in terms of the given loss function.

5.4 Comparisons among competing M-estimators

After assessing model validity using above tests based on different M-estimators, we make pairwise comparisons of only the competing M-estimators in terms of the following criteria, namely the mean relative bias and the average quadratic loss.

5.4.1 Mean Relative Bias

Suppose there are c number of competing estimators $\{\hat{q}_{jt}; 1 \leq t \leq T, 1 \leq j \leq c\}$. Hendricks (1996) defined the mean relative bias (MRB) of the j-th estimator $(1 \leq j \leq c)$ as

$$\text{MRB}_j = \frac{1}{T}\sum_{t=1}^{T} \frac{\hat{q}_{jt} - \bar{q}_t}{\bar{q}_t}, \quad \text{where} \quad \bar{q}_t = \frac{1}{c}\sum_{j=1}^{c} \hat{q}_{jt}.$$

The mean relative bias assesses the relative size, and hence the average conservatism, of VaR estimates produced by various models.

5.4.2 Average Quadratic Loss

The LR_{uc} test counts the number of violations and does not take into account the magnitude of losses. One might be interested, for example, in the magnitude of the violation rather than simply whether or not a violation occurred. Lopez (1999) introduced regulatory loss functions that assign a numerical score, which reflects specific regulatory concerns, to VaR estimates. A model that has the minimum value of the loss function is preferred to the other models. The overall quadratic loss of a VaR is estimated by $\sum_{t=1}^{T} L_t/T$ where

$$L_t = \begin{cases} 1 + (r_t - \hat{q}_t)^2 & \text{if } X_t \leq \hat{q}_t, \\ 0 & \text{if } X_t > \hat{q}_t. \end{cases}$$

Thus, as before, a score of one is imposed when an exception occurs, but now, an additional term based on its magnitude is included. The above loss function takes into consideration the magnitude of the failure, i.e. by how much the actual loss exceeds VaR estimate and thus penalizes the model that produces higher VaR.

5.5 Application to Stock Market Indices

We conduct a detailed study to check the accuracy and reliability of the value-at-risk estimates using M-estimators. Both in-sample and one-day-ahead out-of-sample VaR estimates are calculated. We fit the GARCH(1,1) and the GJR(1,1) model to data sets. Backtesting methods are used to check the accuracy of the in-sample VaR estimates and various evaluation criteria are applied to assess the reliability of the out-of-sample VaR estimates. We study the accuracy of VaR predictions at 10%, 5% and 1% confidence levels for one-day-ahead forecast horizon in real data applications.

5.5.1 Data description and preliminary analysis

The data sets used in the empirical application are daily closing indices of three major stock indices of US, Europe and Asia, namely S&P500 Index, FTSE100 Index, and NIKKEI225 Index, respectively. The data sets are obtained for the period of January 1990 to December 2005 (sixteen years data) from yahoo web site (http://www.finance.yahoo.com). For each of the three indices, the log-return at time t is defined as

$$r_t = (\ln P_t - \ln P_{t-1}) \times 100\%, \quad t = 1, 2, ..., T,$$

where P_t is the closing index at time t. Next using $\{X_t = r_t - \bar{r}; 1 \leq t \leq T\}$ (with $\bar{r} = \sum_{t=1}^{T} r_t/T$) as our observations, the whole data range is divided into two parts; the estimation period and the validation period. Initial $T - K$ values are used as the initial sample for estimation purpose, where T is the total sample size and K is the number of forecast step. Last K observations are left for out-of-sample evaluation. For this study we set $K = 2000$ (nearly eight years data). Table 5.1 summarises the basic descriptive statistics for data sets.

Table 5.1: Descriptive statistics for daily log-returns

Statistics	S&P500 Index	FTSE100 Index	NIKKEI225 Index
Estimation period			
Sample size	2042	2042	1938
Mean	0.0077	0.0078	-0.0104
Median	0.0055	0.0053	-0.0092
Minimum	-3.1028	-1.8069	-2.9546
Maximum	2.1528	2.3534	5.4070
Std. Dev.	0.3440	0.3537	0.6460
Skewness	-0.3555	0.1503	0.3754
Kurtosis	8.3752	5.2804	7.4802
Jarque-Bera	2501.29	450.17	1666.38
$Q^2(20)$	258.42	223.59	296.27
Forecast period (2000 observations)			
Mean	-0.0078	-0.0080	0.0100
Median	0.0031	0.0096	0.0045
Minimum	-3.0728	-2.4362	-3.1320
Maximum	2.4072	2.5550	3.3343
Std. Dev.	0.5197	0.5209	0.6470
Skewness	0.0043	-0.1114	-0.0114
Kurtosis	5.4035	5.2426	4.9091
Jarque-Bera	481.41	423.25	303.78
$Q^2(20)$	792.55	205.49	367.17

Q (20) is the Ljung-Box statistic at lag 20 of the squared log-returns.

S&P500 and FTSE100 indices consist of total $T = 4042$ values. The data set for NIKKEI225 index consists of $T = 3938$ observations. The mean and the standard deviation of the S&P500 index are close to that of the FTSE100 index. The values of kurtosis of the S&P500 index and the NIKKEI225 index are close to each other and higher than the FTSE100 index for estimation period. It can be seen that all data set exhibit excess kurtosis and skewness for both estimation and forecasting period. S&P500 index is slightly negatively skewed where as other two indices show sign of positive skewness in estimation period, S&P500 index shows positive skewness while other two indices are negatively skewed in forecasting period. For both estimation and forecasting periods, the Jarque-Bera (JB) statistics are significantly large, rejecting the hypothesis of normality for three indices. High values of Ljung-Box ($Q^2(20)$) statistic for the squared returns up to lag 20 in both periods indicate dependence in squared returns (ARCH effect).

Descriptive graphs (density of daily returns along with normal density with mean and standard deviation equal to that of raw returns and QQ-plot against the standard normal) for S&P500, FTSE100 and NIKKEI225 indices for in-sample periods are illustrated in Fig. 5.1. By comparing density graphs against the standard normal show that each data set exhibits non-normal characteristics. The empirical distributions have fatter tails and sharper peaks as compared to the normal density function. This indicates evidence of leptokurtosis. The QQ-plot against the standard normal distribution for all data sets exhibits fat-tailedness. Thus, the preliminary analysis of the data and graphs suggests the use of a GARCH model to capture the time-varying volatility and fat-tailedness in these stock indices.

5.5.2 Results of GARCH model

The GARCH(1,1) model is fitted and the parameters are estimated using M-estimators. We allow the corresponding parameters to change over time using moving window of size $w = 2042$ for S&P500 and FTSE100 indices and $w = 1938$ for NIKKEI225 index. This corresponds to roughly eight years of trading data. After predicting the one-step-ahead VaR forecast, the parameters are re-estimated

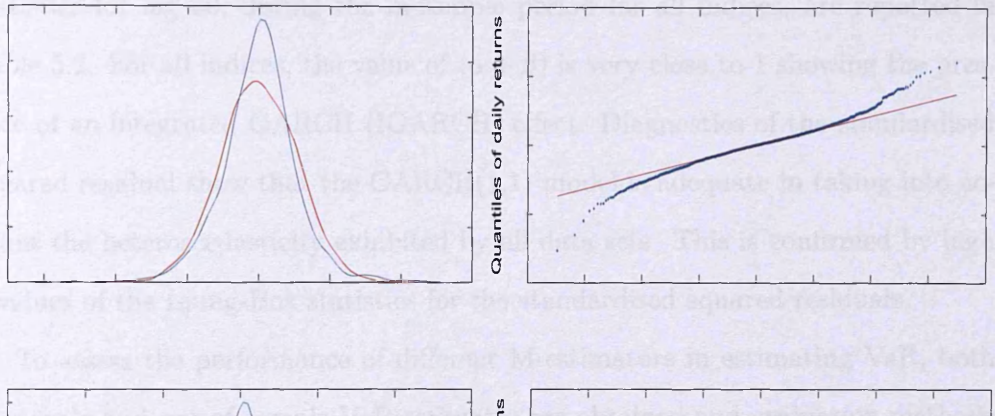

Figure 5.1: Density estimate of daily log-returns along with normal density (red) with mean and standard deviation equal to that of returns and QQ-plot of sample data against the standard normal for three indices.

for each moving window with an increment of one day, discarding the first observation. This process is repeated till the end of the data is reached. In this way 2000 one-day-ahead VaR estimates based on commonly-used rejection probabilities $p = [0.01, 0.05, 0.10]$ are calculated.

The estimated parameters along with their standard errors and Ljung-Box statistics for lag 20, during the in-sample period for all indices, are reported in Table 5.2. For all indices, the value of $(\alpha + \beta)$ is very close to 1 showing the presence of an integrated GARCH (IGARCH) effect. Diagnostics of the standardised squared residual show that the GARCH(1,1) model is adequate in taking into account the heteroscedasticity exhibited by all data sets. This is confirmed by high p-values of the Ljung-Box statistics for the standardised squared residuals.

To assess the performance of different M-estimators in estimating VaR, both in-sample and out-of-sample VaR estimates are obtained and evaluation methods for VaR described above are used.

In-sample VaR performance

Range of summary statistics is used to assess the performance of VaR models. The first performance criterion we apply is the mean relative bias (MRB). The MRB is a measure of size for each value-at-risk estimator that is relative to the average of all five estimators. This quantity is measured in percentage terms, so that a value of 0.10 implies that the value-at-risk estimate for a given estimator is 10% larger, on average, than the average of all five value-at-risk estimates.

After assessing the relative size of risk estimates, we check the accuracy of the VaR estimates produced by each M-estimator. We are concerned to check whether the VaR estimates are large enough to cover the true underlying risk? For this we use the likelihood statistics for conditional coverage (LR_{uc}) proposed by Kupiec (1995). The LR_{uc} statistic tests the null hypothesis that the actual

Table 5.2: Estimated parameters for the GARCH(1,1) with SE's and Ljung-Box statistic for $\hat{\epsilon}_t^2$ (S&P500, FTSE100, NIKKEI225)

	QMLE	LAD	Huber's	B-estimator	Cauchy
S&P500 Index					
$c_H\omega$	0.0005	0.0003	0.0002	0.0002	0.0001
	(0.0004)	(0.0002)	(0.0001)	(0.0002)	(0.0001)
$c_H\alpha$	0.0323	0.0183	0.0193	0.02163	0.0103
	(0.0088)	(0.0040)	(0.0038)	(0.0051)	(0.0024)
β	0.9642	0.9639	0.9652	0.9669	0.9615
	(0.0102)	(0.0081)	(0.0075)	(0.0081)	(0.0087)
$Q^2(20)$	10.5497	10.5100	10.4879	10.6582	10.0032
p-value	0.9571	0.9580	0.9584	0.9546	0.96811
FTSE100 Index					
$c_H\omega$	0.0026	0.0006	0.0006	0.0010	0.0006
	(0.0011)	(0.0003)	(0.0003)	(0.0003)	(0.0002)
$c_H\alpha$	0.0550	0.0195	0.0202	0.0276	0.0127
	(0.0134)	(0.0043)	(0.0044)	(0.0057)	(0.0024)
β	0.9250	0.9609	0.9621	0.9570	0.9515
	(0.0185)	(0.0136)	(0.0084)	(0.0111)	(0.0135)
$Q^2(20)$	21.2076	20.9067	21.3215	21.5500	21.6972
p-value	0.3850	0.4026	0.3784	0.3654	0.3572
NIKKEI225 Index					
$c_H\omega$	0.01414	0.0054	0.0064	0.0048	0.0021
	(0.0047)	(0.0017)	(0.0019)	(0.0018)	(0.0008)
$c_H\alpha$	0.1118	0.0595	0.0746	0.0665	0.0270
	(0.0236)	(0.0100)	(0.0093)	(0.0129)	(0.0059)
β	0.8623	0.8755	0.8843	0.8878	0.8770
	(0.0278)	(0.0191)	(0.0178)	(0.0191)	(0.0201)
$Q^2(20)$	11.0322	10.9808	10.4957	10.7182	11.0189
p-value	0.9454	0.9467	0.9583	0.9532	0.9457

Q (20) is the Ljung-Box statistic at lag 20 of the standardised squared residuals.

and expected number of exceptions are statistically same. Rejection of the null hypothesis indicates that the computed VaR estimates are not accurate enough.

The LR_{uc} is based on the number of violations and does not take into account the magnitude of losses. The Quadratic Loss Function (QLF) of Lopez (1999) takes account of the magnitude of the exceptions and thus provides a more powerful measure of model accuracy. The QLF penalizes large failures more severely than the small failures. Average quadratic loss (AQL) is also calculated for each estimator. If two or more models pass the LR_{uc} test, then the model with the least AQL can be considered more accurate than others.

In the presence of volatility clustering, the LR_{uc} test will classify inaccurate VaR estimates as acceptably accurate since it cannot examine whether the exceptions are randomly distributed. It is of great importance that VaR exceptions be uncorrelated. The likelihood ratio statistic for conditional coverage (LR_{cc}) is used to examine the serial independence of VaR estimated. Since the LR_{cc} test checks the first order dependence in risk estimates, it is also desirable to check dependence of high order. The dynamic quantile (DQ) test is applied to jointly test that the expectation of Hit_t is zero and that Hit_t is uncorrelated with its lagged values or/and lagged values of VaR estimates. For this study, we use five lags of Hit_t and the current VaR as the explanatory variables.

Table 5.3 reports the results of in-sample VaR estimates for three stock indices at rejection probabilities p of 1%, 5%, and 10%. The values of $\hat{p}$ provide a good estimate of rejection probabilities in each case. The average VaR estimates for all estimators are close to each other and we cannot find any significant difference in these estimates. For the 99% VaR confidence level, the mean relative bias for the QMLE falls between -4 and 3 percent, whereas for other estimators, not much variation is observed in this quantity. This shows that the QMLE, on average, produces VaR estimates that are slightly below or above than the average of all five estimators. Results for other confidence levels suggest that the difference in the average sizes is very small.

For all coverage probabilities we consider, the likelihood ratio statistics for unconditional coverage LR_{uc} are found statistically insignificant for all estimators at 5% significance level and indicate that the expected and the actual number of observations falling below the VaR threshold are statistically the same. Also, the conditional coverage test statistic LR_{cc} obtained from all estimators fails to reject the null hypothesis of no serial dependence in in-sample VaR estimates at 5% significance level. The LR_{cc} generated by LAD estimators are found the lowest in many cases.

The mean quadratic loss (AQL) for each estimator is also reported in Table 5.3. Cauchy estimator produces the least AQL for S&P500 Index at all coverage probabilities. It also provide good estimate at 90% VaR confidence level for FTSE100 index. B-estimator provides the least AQL for NIKKEI225 index at $p = 5\%$ and 10% and also for FTSE100 index at $p = 1\%$. Although, AQL for the QMLE is the smallest for FTSE100 index at $p = 5\%$, other estimators also provide very close estimates.

Out-of-sample VaR performance

Next, we look at the performance of M-estimator in producing one-day-ahead VaR estimates. Table 5.4 provides result of out-of-sample VaR estimates. The estimated rejection probabilities are close to the expected values except for S&P500 index at $p = 1\%$. In this case, both likelihood ratio statistics, the LR_{uc} and the LR_{cc} are rejected at 1% and 5% significance levels, respectively. The LR_{cc} test is also rejected in case of FTSE100 index at 95% VaR confidence level. In other cases, all M-estimators pass these tests showing that the number of violation produced

Table 5.3: In-sample VaR evaluation for the GARCH(1,1) model using M-estimators

	$\hat{p}$	Mean VaR	MRB	LR	LR	AQL
			99% VaR confidence level			
S&P500 Index						
QMLE	0.0098	-0.9326	0.0280	0.0088	0.4243	0.0123
LAD	0.0098	-0.9016	-0.0087	0.0088	0.4243	0.0125
Huber's	0.0098	-0.9022	-0.0110	0.0088	0.4243	0.0124
B-estimator	0.0098	-0.9131	0.0003	0.0088	0.4243	0.0124
Cauchy	0.0093	-0.9072	-0.0085	0.1021	0.4779	**0.0120**
FTSE100 Index						
QMLE	0.0098	-0.8270	-0.0415	0.0088	0.4243	0.0105
LAD	0.0098	-0.8621	-0.0020	0.0088	0.4243	0.0104
Huber's	0.0093	-0.8678	0.0052	0.1021	0.4779	0.0099
B-estimator	0.0088	-0.8777	0.0159	0.3017	0.6398	**0.0094**
Cauchy	0.0088	-0.8841	0.0225	0.3017	0.6398	0.0094
NIKKEI225 Index						
QMLE	0.0103	-1.5780	0.0058	0.0198	0.4579	**0.0126**
LAD	0.0119	-1.5681	-0.0058	0.6445	1.2211	0.0144
Huber's	0.0114	-1.5704	-0.0029	0.3428	0.8712	0.0139
B-estimator	0.0108	-1.5805	0.0005	0.1332	0.6153	0.0134
Cauchy	0.0108	-1.5840	0.0025	0.1332	0.6153	0.0133
			95% VaR confidence level			
S&P500 Index						
QMLE	0.0490	-0.5206	0.0038	0.0458	1.0359	0.0567
LAD	0.0480	-0.5207	0.0015	0.1756	0.2939	0.0556
Huber's	0.0480	-0.5194	-0.0040	0.1756	0.6382	0.0557
B-estimator	0.0480	-0.5189	0.0055	0.1756	0.6382	0.0557
Cauchy	0.0480	-0.5253	0.0042	0.1756	0.6382	**0.0555**
FTSE100 Index						
QMLE	0.0485	-0.5574	0.0043	0.1000	0.3555	**0.0517**
LAD	0.0485	-0.5544	-0.0023	0.1000	0.2084	0.0518
Huber's	0.0485	-0.5560	0.0011	0.1000	0.2084	0.0518
B-estimator	0.0485	-0.5578	0.0036	0.1000	0.2084	0.0518
Cauchy	0.0490	-0.5525	-0.0067	0.0458	1.0359	0.0524
NIKKEI225 Index						
QMLE	0.0531	-1.0121	-0.0039	0.3964	2.7431	0.0648
LAD	0.0537	-1.0190	-0.0024	0.5354	2.7278	0.0651
Huber's	0.0547	-1.0188	-0.0012	0.8741	2.7763	0.0662
B-estimator	0.0521	-1.0308	0.0076	0.1802	1.7335	**0.0632**
Cauchy	0.0542	-1.0233	0.0000	0.6947	2.7390	0.0656
			90% VaR confidence level			
S&P500 Index						
QMLE	0.0950	-0.3923	0.0063	0.5747	5.1378	0.1069
LAD	0.0955	-0.3926	0.0046	0.4668	4.8136	0.1073
Huber's	0.0955	-0.3907	-0.0033	0.4668	4.8136	0.1074
B-estimator	0.0970	-0.3890	-0.0082	0.2111	4.8616	0.1089
Cauchy	0.0945	-0.3934	0.0006	0.6940	4.5303	**0.1062**
FTSE100 Index						
QMLE	0.0955	-0.4329	0.0140	0.4668	1.1401	0.1019
LAD	0.0955	-0.4253	-0.0049	0.4668	0.6940	0.1021
Huber's	0.0940	-0.4261	-0.0026	0.8247	1.0225	0.1007
B-estimator	0.0945	-0.4247	-0.0064	0.6940	0.8968	0.1012
Cauchy	0.0926	-0.4278	-0.0001	1.2860	1.4974	**0.0992**
NIKKEI225 Index						
QMLE	0.1042	-0.7350	-0.0007	0.3808	5.9871	0.1311
LAD	0.1022	-0.7405	0.0014	0.1005	3.7972	0.1286
Huber's	0.1022	-0.7361	-0.0031	0.1005	3.7972	0.1291
B-estimator	0.1017	-0.7423	0.0023	0.0584	3.9585	**0.1280**
Cauchy	0.1017	-0.7408	0.0001	0.0584	3.9585	0.1282

The smallest AQL for each data set at each confidence level is bold faced to highlight the best performance.

by M-estimators are statistically same as expected number of violations and these violations are independent of each other. Not much variation in mean relative bias is observed in out-of-sample VaR estimates. As in the case of in-sample VaR, the AQL of Cauchy estimator is again the smallest in most of the cases.

Table 5.5 presents the results of dynamic quantile test statistics of Engle and Manganelli (2004) on the out-of-sample VaR performance of M-estimators. For each data set, at $p = 10\%$, the DQ statistics for all estimators fail to accept the null hypothesis of no higher order dependence in VaR violations. The test also fails for S&P500 at $p = 1\%$ and FTSE100 indices at $p = 5\%$. This shows that although in some cases the LR_{cc} accepts the null hypothesis of no serial dependence in VaR estimates for different M-estimates, the DQ test confirms the existence of high order dependence in these estimates. As an example, all M-estimators pass the conditional coverage statistics at both $p = 5\%$ and 10% but fails to accept the dynamic quantile test at these coverage probabilities.

To further examine the superiority of one estimator to another in predicting one-day-ahead VaR, we use the one sided sign statistics. Table 5.6 presents the sign statistics on quadratic loss functions between pairs of M-estimators which pass the dynamic quantile test. We only report the result for 95% VaR confidence level as those of others confidence levels produce the same results. The test statistic applied to the quadratic loss function indicates that none of the estimator is significantly better than the others for all data set at different confidence levels.

5.5.3 Results of GJR model

In our second study we fit the GJR(1,1) model to those three stock indices and estimate the parameters using M-estimators. We predict the one-day-ahead VaR using the rolling window method. The size of window length, the number of out-

Table 5.4: Out-of-sample VaR evaluation for the GARCH(1,1) model using M-estimators

	$\hat{p}$	Mean VaR	MRB	LR	LR	AQL
			99% VaR confidence level			
S&P500 Index						
QMLE	0.0045	-1.3340	-0.0037	7.6879**	7.7783*	0.0069
LAD	0.0045	-1.3312	-0.0081	7.6879**	7.7783*	0.0070
Huber's	0.0045	-1.3381	-0.0022	7.6879**	7.7783*	0.0069
B-estimator	0.0045	-1.3473	-0.0014	7.6879**	7.7783*	**0.0068**
Cauchy	0.0045	-1.3637	0.0154	7.6879**	7.7783*	0.0069
FTSE100 Index						
QMLE	0.0120	-1.2000	-0.0003	0.7595	1.3670	0.0128
LAD	0.0110	-1.2019	0.0001	0.1957	0.7074	0.0118
Huber's	0.0110	-1.1995	-0.0007	0.1957	0.7074	0.0118
B-estimator	0.0110	-1.1995	-0.0007	0.1957	0.7074	0.0118
Cauchy	0.0110	-1.1961	-0.0043	0.1957	0.7074	**0.0117**
NIKKEI225 Index						
QMLE	0.0085	-1.5267	-0.0135	0.4789	0.7876	0.0118
LAD	0.0080	-1.5406	-0.0063	0.8675	1.1417	**0.0113**
Huber's	0.0095	-1.5449	-0.0035	0.0514	0.4351	0.0128
B-estimator	0.0090	-1.5512	0.0009	0.2090	0.5542	0.0123
Cauchy	0.0090	-1.5782	0.0224	0.2090	0.5542	0.0122
			95% VaR confidence level			
S&P500 Index						
QMLE	0.0540	-0.8034	-0.0076	0.6573	1.0143	0.0635
LAD	0.0520	-0.8037	-0.0080	0.1663	0.7529	0.0614
Huber's	0.0515	-0.8103	0.0016	0.0939	0.2965	0.0608
B-estimator	0.0520	-0.8142	0.0024	0.1663	1.4925	0.0610
Cauchy	0.0505	-0.8186	0.0116	0.0105	0.1165	**0.0597**
FTSE100 Index						
QMLE	0.0600	-0.7927	0.0008	3.9684	7.1741*	0.0665
LAD	0.0595	-0.7934	-0.0000	3.5915	8.3236*	0.0659
Huber's	0.0605	-0.7923	-0.0004	4.3631	8.6490*	0.0670
B-estimator	0.0605	-0.7971	0.0038	4.3631	8.6490*	0.0668
Cauchy	0.0595	-0.7900	-0.0042	3.5915	8.3236*	**0.0659**
NIKKEI225 Index						
QMLE	0.0515	-1.0462	-0.0053	0.0939	0.7505	0.0623
LAD	0.0490	-1.0520	-0.0020	0.0424	1.1279	0.0597
Huber's	0.0495	-1.0522	-0.0023	0.0106	0.3674	0.0602
B-estimator	0.0490	-1.0529	-0.0009	0.0424	0.4500	0.0599
Cauchy	0.0485	-1.0626	0.0106	0.0956	2.3394	**0.0595**
			90% VaR confidence level			
S&P500 Index						
QMLE	0.1120	-0.6111	-0.0045	3.0927	3.7433	0.1296
LAD	0.1120	-0.6107	-0.0058	3.0927	4.0705	0.1295
Huber's	0.1110	-0.6144	0.0014	2.6058	4.2349	0.1284
B-estimator	0.1140	-0.6139	-0.0022	4.1867	4.6202	0.1312
Cauchy	0.1090	-0.6209	0.0110	1.7541	3.3588	**0.1262**
FTSE100 Index						
QMLE	0.1055	-0.6049	0.0022	0.6616	4.7973	0.1210
LAD	0.1050	-0.6048	0.0004	0.5475	4.0543	0.1204
Huber's	0.1055	-0.6030	-0.0011	0.6616	3.9799	0.1211
B-estimator	0.1045	-0.6067	0.0031	0.4441	4.1451	**0.1198**
Cauchy	0.1060	-0.6016	-0.0046	0.7862	3.9221	0.1215
NIKKEI225 Index						
QMLE	0.1005	-0.7689	-0.0007	0.0055	0.2202	0.1253
LAD	0.1005	-0.7683	-0.0033	0.0055	0.2552	0.1254
Huber's	0.0995	-0.7672	-0.0051	0.0056	0.2175	0.1246
B-estimator	0.0995	-0.7675	-0.0042	0.0056	0.2566	0.1247
Cauchy	0.0990	-0.7789	0.0132	0.0223	0.3497	**0.1241**

The smallest AQL for each data set at each confidence level is bold faced to highlight the best performance. *,** denote significant at the 5% and 1% level, respectively.

Table 5.5: Dynamic quantile test statistics for the GARCH(1,1) model

p	QMLE	LAD	Huber's	B-estimator	Cauchy
S&P500 Index					
1%	16.8290*	16.8252*	16.8177*	16.8887*	16.8703*
5%	13.2744	12.5658	12.6396	20.7608**	10.3658
10%	16.2397*	21.7491**	22.6867**	18.4834**	14.7659*
FTSE100 Index					
1%	4.1285	4.1423	4.6675	4.8795	4.9002
5%	24.4414**	19.9735**	19.9962**	20.4048**	28.1486**
10%	16.1487*	20.4942**	23.0526**	26.9402**	28.3048**
NIKKEI225 Index					
1%	4.4172	5.2616	5.6683	5.9421	5.8800
5%	8.9755	11.2762	11.4320	11.3691	12.3032
10%	19.0729**	23.4811**	23.3008**	24.3819**	24.4052**

The dynamic quantile (DQ) test statistics on the out-of-sample VaR performance of M-estimators. The DQ test statistic is asymptotically χ (7). *,** denote significant at the 5% and 1% level, respectively.

of-sample forecast and the rejection probabilities are kept same as in the case of the GARCH(1,1) model. The purpose of this study is to check the predictive performance of M-estimators in asymmetric GARCH models. We also aim to compare the results of M-estimators in both GARCH and GJR model and to investigate which estimator and model fits these data sets well.

Table 5.7 reports the estimated parameters of the GJR(1,1) model for M-estimators along with their standard errors and the Ljung-Box statistics for the standardised squared residuals for lag 20. The p-values of the Ljung-Box statistics show that the GJR(1,1) model is adequate to model the volatility.

In-sample VaR performance

Using evaluation methods, the in-sample VaR performance of M-estimators is analysed, when GJR(1,1) model is fitted to these data sets. The results for three stock indices at rejection probabilities p of 1%, 5%, and 10% are reported in Table 5.8. As we can see from the table, the values of $\hat{p}$ provide good estimates of

Table 5.6: Sign test for out-of-sample VaR performance for the GARCH(1,1) model using M-estimators

	(i, j)	S^a_{ij}	S^a_{ji}
	95% VaR confidence level		
S&P500 Index			
	(1, 2)	-42.0381**	-42.5300**
	(1, 3)	-41.3673**	-43.1561**
	(1, 5)	-41.2331**	-43.2903**
	(2, 3)	-41.0542**	-43.6928**
	(2, 5)	-41.5461**	-43.1561**
	(3, 5)	-42.1722**	-42.5300**
NIKKEI225 Index			
	(1, 2)	-42.1722**	-42.6195**
	(1, 3)	-42.1275**	-42.5747**
	(1, 4)	-42.1275**	-42.4853**
	(1, 5)	-42.3511**	-42.2617**
	(2, 3)	-42.1275**	-42.8431**
	(2, 4)	-42.2617**	-42.6195**
	(2, 5)	-42.5747**	-42.3064**
	(3, 4)	-42.3958**	-42.5300**
	(3, 5)	-42.9772**	-41.9039**
	(4, 5)	-43.1114**	-41.8145**

The QMLE, LAD, Huber's, B-estimator, and Cauchy estimators are numbered 1, 2, 3, 4, and 5, respectively. The critical values of the sign statistic at 5% significance level is -1.6345.*,** denote significant at the 5% and 1% level, respectively.

rejection probabilities in each case. The average VaR estimates for all estimators are found close to each other. For FTSE100 Index the mean relative biases of Cauchy estimator are found larger as compare to other estimators for this data sets showing that on average the Cauchy estimator produce VaR estimates that are higher then the average of other estimates.

Both the likelihood ratio statistics for unconditional coverage LR_{uc} and the likelihood ratio statistics for conditional coverage LR_{cc} are not found statistically significant at 5% level of significance for all coverage probabilities. This is an indication that the expected and the actual number of observations falling below the VaR threshold are statistically the same and that there are no serial dependence between two VaR estimates produce by an M-estimator.

Table 5.7: Estimated parameters for the GJR(1,1) with SE's and Ljung-Box statistic for $\hat{\epsilon}_t^2$ (S&P500, FTSE100, NIKKEI225)

	QMLE	LAD	Huber's	B-estimator	Cauchy
S&P500 Index					
$c_H\omega$	0.0041	0.0023	0.0027	0.0028	0.0013
	(0.0006)	(0.0003)	(0.0003)	(0.0004)	(0.0002)
$c_H\alpha$	0.0284	0.0170	0.0216	0.0226	0.0099
	(0.0099)	(0.0046)	(0.0058)	(0.0062)	(0.0028)
$c_H\gamma$	0.1007	0.0567	0.0710	0.0689	0.0327
	(0.0153)	(0.0071)	(0.0088)	(0.0095)	(0.0044)
β	0.8863	0.8805	0.8843	0.8751	0.8648
	(0.0119)	(0.0101)	(0.0096)	(0.0112)	(0.0121)
$Q^2(20)$	10.5497	10.5100	10.4879	10.6582	10.0032
p-value	0.9571	0.9580	0.9584	0.9546	0.96811
FTSE100 Index					
$c_H\omega$	0.0019	0.0011	0.0013	0.0010	0.0006
	(0.0006)	(0.0002)	(0.0003)	(0.0002)	(0.0001)
$c_H\alpha$	0.0206	0.0076	0.0188	0.0063	0.0017
	(0.0074)	(0.0023)	(0.0034)	(0.0031)	(0.0013)
$c_H\gamma$	0.0504	0.0343	0.0391	0.0491	0.0238
	(0.0121)	(0.0041)	(0.0068)	(0.0060)	(0.0028)
β	0.9406	0.9475	0.9444	0.9547	0.9518
	(0.0119)	(0.0056)	(0.0097)	(0.0054)	(0.0059)
$Q^2(20)$	18.9195	19.5729	24.1036	19.7610	21.1915
p-value	0.5271	0.4849	0.2379	0.4730	0.3859
NIKKEI225 Index					
$c_H\omega$	0.0085	0.0044	0.0053	0.0060	0.0026
	(0.0013)	(0.0006)	(0.0007)	(0.0009)	(0.0004)
$c_H\alpha$	0.0148	0.0083	0.0109	0.0096	0.0040
	(0.0067)	(0.0032)	(0.0041)	(0.0046)	(0.0021)
$c_H\gamma$	0.1391	0.0886	0.1137	0.1304	0.0564
	(0.0140)	(0.0069)	(0.0088)	(0.0108)	(0.0048)
β	0.8976	0.8964	0.9003	0.8870	0.8873
	(0.0091)	(0.0072)	(0.0069)	(0.0084)	(0.0086)
$Q^2(20)$	14.9329	15.8846	15.8154	16.6741	16.7256
p-value	0.7802	0.7238	0.7280	0.6740	0.6707

Q (20) is the Ljung-Box statistic at lag 20 of the standardised squared residuals.

The summary of results of the mean quadratic loss (AQL) for each estimator is as follows. For S&P500 Index, Cauchy estimator produces the least AQL at 99% and 95% VaR confidence level whereas at 90% confidence level the AQL of Huber's estimator is found the least. Huber's estimator also produces the lowest AQL for FTSE100 Index at 99% and 90% confidence interval and B-estimator outperform other estimator based on AQL at 95% confidence level. Similarly for NIKKEI225 Index B-estimator shows the least value for AQL at 99% and 90% and LAD's AQL is found the lowest at 95%. From these findings we may conclude that based on AQL, Huber's and B-estimator provide better results than the competing estimators.

Out-of-sample VaR performance

The results of out-of-sample VaR estimate are displayed in Table 5.9. The estimated rejection probabilities are close to the expected values except for S&P500 index at $p = 1\%$. In this case, both likelihood ratio statistics, the LR_{uc} and the LR_{cc} for B-estimator are rejected at 5% significance levels. The LR_{uc} and the LR_{cc} tests are not rejected in any other cases we consider. Again we see that all M-estimators pass these tests showing that the number of violation produced by M-estimators are statistically same as expected number of violations and these violations are independent of each other.

The mean relative biases of all estimators do not show much variation in out-of-sample VaR estimates. The AQL's of B-estimator are the smallest for FTSE100 index at all coverage probabilities. B-estimators also produces the least AQLs for S&P500 and NIKKEI225 indices at 95% VaR confidence level. The values of AQL for QMLE are found the lowest at 90% for S&P500 Index. These results suggests that B-estimator again produces better out-of-sample VaR estimates as compared to other competing estimators when the GJR(1,1) model is fitted to these stock

Table 5.8: In-sample VaR evaluation for the GJR(1,1) model using M-estimators

	$\hat{p}$	Mean VaR	MRB	LR	LR	AQL
			99% VaR confidence level			
S&P500 Index						
QMLE	0.0093	-0.8890	0.0089	0.1021	0.4779	0.0120
LAD	0.0098	-0.8775	-0.0111	0.0088	0.4243	0.0126
Huber's	0.0098	-0.8773	-0.0095	0.0088	0.4243	0.0126
B-estimator	0.0098	-0.8909	0.0052	0.0088	0.4243	0.0125
Cauchy	0.0093	-0.8952	0.0064	0.1021	0.4779	**0.0120**
FTSE100 Index						
QMLE	0.0098	-0.8124	-0.0165	0.0088	0.4243	0.0107
LAD	0.0093	-0.8229	-0.0070	0.1021	0.4779	0.0102
Huber's	0.0088	-0.8233	-0.0033	0.3017	0.6398	**0.0097**
B-estimator	0.0098	-0.7906	-0.0342	0.0088	0.4243	0.0110
Cauchy	0.0098	-0.8611	0.0610	0.0088	0.4243	0.0107
NIKKEI225 Index						
QMLE	0.0114	-1.4921	0.0008	0.3428	0.8712	0.0139
LAD	0.0114	-1.5013	-0.0017	0.3428	0.8712	0.0139
Huber's	0.0124	-1.4914	-0.0083	1.0341	1.6613	0.0150
B-estimator	0.0103	-1.5181	0.0074	0.0198	0.4579	**0.0127**
Cauchy	0.0103	-1.5104	0.0018	0.0198	0.4579	0.0128
			95% VaR confidence level			
S&P500 Index						
QMLE	0.0495	-0.5306	0.0065	0.0125	0.3242	0.0567
LAD	0.0490	-0.5370	0.0117	0.0458	1.0359	0.0560
Huber's	0.0495	-0.5255	-0.0084	0.0125	0.9124	0.0567
B-estimator	0.0480	-0.5274	-0.0052	0.1756	1.3620	0.0553
Cauchy	0.0475	-0.5296	-0.0046	0.2725	0.7953	**0.0547**
FTSE100 Index						
QMLE	0.0480	-0.5511	-0.0066	0.1756	0.2939	0.0513
LAD	0.0480	-0.5506	-0.0106	0.1756	0.2939	0.0514
Huber's	0.0480	-0.5474	-0.0133	0.1756	0.2939	0.0515
B-estimator	0.0455	-0.5560	0.0113	0.8789	1.1152	**0.0490**
Cauchy	0.0480	-0.5554	0.0191	0.1756	2.4220	0.0516
NIKKEI225 Index						
QMLE	0.0542	-0.9881	-0.0051	0.6947	0.8316	0.0649
LAD	0.0516	-1.0124	0.0106	0.1034	2.8868	**0.0618**
Huber's	0.0526	-1.0091	0.0073	0.2780	1.7017	0.0627
B-estimator	0.0521	-1.0029	-0.0009	0.1802	1.5040	0.0624
Cauchy	0.0521	-0.9924	-0.0119	0.1802	3.1032	0.0629
			90% VaR confidence level			
S&P500 Index						
QMLE	0.0965	-0.3889	0.0036	0.2851	1.9906	0.1083
LAD	0.0975	-0.3933	0.0079	0.1483	3.6884	0.1089
Huber's	0.0955	-0.3882	-0.0035	0.4668	2.4323	**0.1072**
B-estimator	0.0989	-0.3862	-0.0090	0.0264	2.3053	0.1108
Cauchy	0.0960	-0.3916	0.0011	0.3703	2.2031	0.1076
FTSE100 Index						
QMLE	0.0960	-0.4261	0.0001	0.3703	1.1124	0.1024
LAD	0.0960	-0.4264	-0.0026	0.3703	0.7939	0.1025
Huber's	0.0950	-0.4240	-0.0051	0.5747	0.7873	**0.1016**
B-estimator	0.0955	-0.4215	-0.0019	0.4668	3.8900	0.1023
Cauchy	0.0960	-0.4226	0.0096	0.3703	4.4786	0.1031
NIKKEI225 Index						
QMLE	0.1037	-0.7419	0.0109	0.2940	2.6928	0.1272
LAD	0.1037	-0.7403	0.0001	0.2940	1.5418	0.1275
Huber's	0.1032	-0.7400	-0.0003	0.2183	2.1247	0.1266
B-estimator	0.1027	-0.7399	-0.0025	0.1538	1.1521	**0.1259**
Cauchy	0.1042	-0.7360	-0.0082	0.3808	1.1304	0.1279

The smallest AQL for each data set at each confidence level is bold faced to highlight the best performance.

indices.

The dynamic quantile test statistics on the out-of-sample VaR performance of M-estimators when GJR(1,1) model is fitted to data sets are provided in Table 5.10. For S&P500 Index at $p = 5\%$, the DQ tests is rejected at 5% level of significance for Huber's Cauchy and B-estimator. For FTSE100 Index the test is rejected for all estimators at 5% level of significance except Cauchy estimator. For all other case, all estimators pass the DQ test. These results show the DQ test confirms the existence of high order dependence in the above mentioned cases. In all other cases the acceptance of the DQ test means that there is not any high order dependence among the out-of-sample VaR estimates.

5.5.4 Comparison of Results

Next, we compare the results of fitting GARCH(1,1) and GJR(1,1) model to S&P500, FTSE100, and NIKKEI225 indices. First we compare results of Table 5.3 and Table 5.8 for in-sample VaR estimates. The average VaR estimates produced by M-estimators when GARCH(1,1) model is fitted to these data sets are found higher than the GJR(1,1) model at 99% VaR confidence level. This indicates that the GARCH(1,1) model produces high VaR estimates as compared to the GJR(1,1) model. For other confidence levels we do not find any significant difference. For GARCH(1,1), the cases where the AQLs of the QMLE were found the least are replaced by B-estimator when GJR(1,1) model is fitted. The values of the AQL of Huber's estimate were not found the least at any occasion in Table 5.3 for GARCH(1,1) model but in Table 5.8 when GJR(1,1) model is fitted, these values are found the least on four occasions.

Finally, we compare results of Table 5.4 and Table 5.9 for out-of-sample VaR estimates. The first observation is that, at 99% VaR confidence level when the

Table 5.9: Out-of-sample VaR evaluation for the GJR(1,1) model using M-estimators

	$\hat{p}$	Mean VaR	MRB	LR	LR	AQL
			99% VaR confidence level			
S&P500 Index						
QMLE	0.0075	-1.2610	-0.0193	1.3822	1.6240	0.0094
LAD	0.0075	-1.2790	-0.0076	1.3822	1.6240	0.0096
Huber's	0.0075	-1.2979	0.0026	1.3822	1.6240	0.0096
B-estimator	0.0050	-1.3080	0.0119	6.1875*	6.2981*	0.0069
Cauchy	0.0060	-1.3117	0.0124	3.7725	3.9294	**0.0080**
FTSE100 Index						
QMLE	0.0110	-1.2015	0.0088	0.1957	0.7074	0.0118
LAD	0.0105	-1.1781	-0.0144	0.0497	0.5167	0.0114
Huber's	0.0095	-1.1944	-0.0045	0.0514	0.4351	0.0103
B-estimator	0.0095	-1.2157	0.0120	0.0514	0.4351	**0.0102**
Cauchy	0.0100	-1.2000	-0.0019	0.0000	0.4244	0.0108
NIKKEI225 Index						
QMLE	0.0105	-1.5456	-0.0022	0.0497	0.5167	0.0143
LAD	0.0110	-1.5432	-0.0048	0.1957	0.7074	0.0142
Huber's	0.0105	-1.5593	0.0057	0.0497	0.5167	0.0137
B-estimator	0.0105	-1.5533	0.0012	0.0497	0.5167	**0.0136**
Cauchy	0.0105	-1.5484	-0.0019	0.0497	0.5167	0.0137
			95% VaR confidence level			
S&P500 Index						
QMLE	0.0520	-0.8201	-0.0037	0.1663	3.3534	0.0599
LAD	0.0520	-0.8235	-0.0011	0.1663	3.3534	0.0602
Huber's	0.0500	-0.8292	0.0027	0.0000	2.6333	0.0582
B-estimator	0.0485	-0.8241	-0.0010	0.0956	0.9707	**0.0565**
Cauchy	0.0490	-0.8285	0.0030	0.0424	0.9974	0.0570
FTSE100 Index						
QMLE	0.0550	-0.8070	0.0065	1.0210	1.7778	0.0606
LAD	0.0560	-0.7975	-0.0091	1.4616	1.6678	0.0615
Huber's	0.0570	-0.8044	-0.0023	1.9779	2.1374	0.0623
B-estimator	0.0535	-0.8135	0.0067	0.5048	0.6291	**0.0568**
Cauchy	0.0575	-0.8069	-0.0018	2.2640	2.4072	0.0628
NIKKEI225 Index						
QMLE	0.0490	-1.0302	0.0023	0.0424	0.9974	0.0619
LAD	0.0490	-1.0318	-0.0003	0.0424	0.9974	0.0607
Huber's	0.0490	-1.0354	0.0028	0.0424	0.9974	0.0611
B-estimator	0.0490	-1.0335	0.0015	0.0424	0.9974	**0.0606**
Cauchy	0.0500	-1.0283	-0.0043	0.0000	1.1257	0.0618
			90% VaR confidence level			
S&P500 Index						
QMLE	0.1100	-0.6048	-0.0058	2.1595	2.5563	**0.1268**
LAD	0.1105	-0.6055	-0.0059	2.3776	2.7369	0.1275
Huber's	0.1110	-0.6111	0.0005	2.6058	3.1171	0.1278
B-estimator	0.1110	-0.6150	0.0097	2.6058	3.1171	0.1271
Cauchy	0.1130	-0.6107	0.0015	3.6197	3.9627	0.1296
FTSE100 Index						
QMLE	0.1060	-0.6069	0.0053	0.7862	3.9221	0.1207
LAD	0.1090	-0.6015	-0.0080	1.7541	2.8918	0.1237
Huber's	0.1070	-0.6059	-0.0029	1.0671	3.8556	0.1211
B-estimator	0.1055	-0.6141	0.0081	0.6616	3.2526	**0.1190**
Cauchy	0.1070	-0.6077	-0.0024	1.0671	3.8556	0.1211
NIKKEI225 Index						
QMLE	0.0995	-0.7706	0.0011	0.0056	0.7268	0.1244
LAD	0.1005	-0.7706	-0.0013	0.0055	0.8766	0.1254
Huber's	0.0980	-0.7751	0.0042	0.0894	0.9901	**0.1228**
B-estimator	0.0985	-0.7726	0.0009	0.0502	1.0344	0.1234
Cauchy	0.0995	-0.7688	-0.0050	0.0056	0.4027	0.1245

The smallest AQL for each data set at each confidence level is bold faced to highlight the best performance. *,** denote significant at the 5% and 1% level, respectively.

GARCH(1,1) model is fitted for S&P500 Index, both the LR_{uc} and the LR_{cc} tests were rejected at 5% significance level for all M-estimators. But, when the GJR(1,1) model is fitted to the same data set both likelihood ratio statistics are not found significantly different at the same significance level except for B-estimator. Similarly for FTSE100 Index, at 95% VaR confidence, the GARCH(1,1) model rejected the LR_{cc} test at 5% significance level but it is accepted for the GJR(1,1) model. This suggests that for those data sets the expected number of violation for out-of-sample VaR estimates in the GJR(1,1) case are equal to the true number of violation and also these violations are independent of each other. In Table 5.4 the AQL of Cauchy and B-estimators were found the least in most of the cases whereas in Table 5.9 we observe that the B-estimator overall shows good results.

Comparing the results of Table 5.5 with Table 5.10, it can be seen that the dynamic quantile test for no high order dependence is accepted mostly by the GJR(1,1) model. This confirms that there exists high order dependence in out-of-sample VaR estimates of GARCH(1,1) models but in GJR(1,1) case these estimates are independent in most of the cases. Hence, we conclude that for data sets we analyse in this study, GJR(1,1) model provides a better fit as compared to the GARCH(1,1) model in terms of predicting one-day-ahead VaR estimates.

We analyse in detail the results of using M-estimators for in-sample and out-of-sample VaR estimates. Both symmetric and asymmetric GARCH models are fitted and various performance measures and M-tests are used to evaluate the risk estimates. Our results show that M-estimators can provide reliable estimates for risk. These results suggest that estimators such as Huber's, Cauchy and B-estimators outperform the widely-used QMLE. This strengthens our suggestion of using robust M-estimators for GARCH-type models. Our study may help risk managers to select appropriate estimator for predicting VaR and to use range of

Table 5.10: Dynamic quantile test statistics for the GJR(1,1) model

p	QMLE	LAD	Huber's	B-estimator	Cauchy
S&P500 Index					
1%	7.1346	7.2829	7.3984	14.5561*	10.8300
5%	12.5005	12.5284	15.6235*	14.4700*	14.8701*
10%	9.3568	11.9913	13.8734	13.3806	14.9990*
FTSE100 Index					
1%	4.2564	2.1902	2.1268	2.1393	2.3754
5%	6.6261	3.0925	5.7603	1.4448	3.4796
10%	19.7151**	11.0671	14.9138*	14.2534*	11.9589
NIKKEI225 Index					
1%	1.3981	4.6199	5.2873	5.6648	5.2949
5%	1.5635	2.3098	2.2481	1.4319	2.3191
10%	7.4248	6.4519	5.6769	5.4136	4.8600

The dynamic quantile (DQ) test statistics on the out-of-sample VaR performance of M-estimators. The DQ test statistic is asymptotically χ (7). *,** denote significant at the 5% and 1% level, respectively.

summary statistics to evaluate their predicted risk estimates.

5.6 Conclusion

In this chapter we proposed robust measures of VaR in the GARCH setup using M-estimators. We applied these to financial data sets and used backtesting methods to assess the in-sample VaR performance of M-estimators. We also assessed the accuracy of the proposed VaR estimates with out-of-sample VaR analysis. We compare the relative performance of the competing M-estimators using a number of newly-introduced M-test statistics.

From our empirical analysis it turned out that the average quadratic losses of the Cauchy and B-estimator were the least among the five M-estimators considered for the cited data sets. The mean relative bias (MRB) of the QMLE was also found to be higher than other estimators in most of the cases, indicating that the risk estimate of the QMLE was slightly larger than the average of other risk estimates.

These findings confirmed the superiority of the Cauchy and B-estimator over the QMLE for fitting the GARCH model and VaR calculations. In fact, in many occasions, the QMLE is routinely used without paying attention to the fact that the finite fourth moment assumption is not tenable for that data. In those cases, such alternatives to QMLE for which a well-developed asymptotic theory exist, provide strong justification for their use. A comparison of the results of both symmetric and asymmetric models reveals that asymmetric model provides better forecasts for these data sets.

A number of interesting extensions and questions emerge naturally from this research which needs further investigation. In (5.2) we can probably get a better estimator of q_t if we can use extreme-value theory for estimating the second term involving quantile of the error distribution. Recent work on the skewed t-distribution will be also useful for that purpose. It will be also of theoretical interest to investigate different rates of approximation of the quantity $|\hat{q}_t - q_t|$.

Chapter 6

The Linear Estimator for ARCH Models

6.1 Introduction

The most frequently used method for the estimation of the parameters of ARCH model is the quasi-maximum likelihood estimation (QMLE). The QMLE is consistent and asymptotically normal even when the conditional error density is non-normal. Weiss (1986) established the asymptotic properties of the QMLE for ARCH model under the existence of fourth-order moment on the ARCH process. The QMLE does not admit a closed form expression and numerical optimization methods must be used to obtain the estimates. Numerical optimization methods are not easy to compute and require the choice of the initial values of the parameters. These procedures may face convergence problem when the sample size is small.

Bose and Mukherjee (2003) proposed an estimator whose computation involves solving only two sets of linear equations. The asymptotics of this two-stage least squares estimator of ARCH models were discussed and it was shown that the estimator has the same asymptotic normal distribution as that of QMLE. Result of a small simulation study showed that the linear estimator (LE) performed better than the QMLE.

The linear estimator has a closed form and is obtained by solving linear equations. Hence, it can be easily implemented and does not require the use of any numerical optimization methods or the choice of initial values of parameters. Another advantage of LE is that it requires very little computational time for the estimation of the parameters of ARCH model. This enables one to perform other computer intensive tasks such as volatility forecasting using recursive scheme and bootstrapping volatility models in little time.

Although linear estimator for ARCH models was introduced few years ago but to the best of our knowledge no applications to real data sets exist to date. Moreover, the accuracy under non-normal errors and the volatility forecasting performance of the estimator have not been investigated. We provide a detailed empirical analysis of the LE and compare this with the QMLE in terms of estimating, volatility forecasting, and bootstrapping ARCH models. The purpose of this study is to address all these issues and compare the results of both estimators. Based on these comparisons some useful suggestions are made for practitioners. Our aims for this study are as follows.

Our first aim is to check the accuracy of the LE in estimating the parameters of ARCH models. Monte Carlo experiments are performed to compare the accuracy of these estimators with the QMLE. Sample of size 50 is used along with other choices for sample sizes to check the performance of these estimators in very small samples. Errors are generated from normal and non-normal distributions and different orders for ARCH models are considered. The results show that the linear estimator provides accurate estimates for the parameters of ARCH models and outperform the QMLE in most of the cases.

Our second aim is to compare the predictive ability of the LE with the QMLE in terms of forecasting volatility. The forecasting performance of both linear estimator and the QMLE are checked using various evaluation measures. It is found that the performance of the LE in estimating and forecasting volatility is comparable to QMLE and better in small samples. Recursive and rolling schemes for volatility forecasts are used for generating one-day ahead forecasts of SP500, FTSE100 and NIKKEI225 indices and results of the study of real data sets support simulations.

Third, to approximate the true distribution of the LE in finite sample using a weighted bootstrap method. The weighted bootstrapped LE is obtained by solving linear equations and hence the approach is easy to implement. Two schemes (Scheme E and Scheme U) are used for bootstrapping besides widely used paired bootstrap and residual bootstrap. We observe that weighted bootstrap schemes work well for ARCH models when LE is used. It is found that scheme U and scheme E are good alternative to paired bootstrap. Moreover, using the LE instead of the QMLE for fitting ARCH models enables us to obtain these results in very little time.

Our final aim is to develop bootstrap prediction intervals for returns, volatility and value-at-risk. These prediction intervals are developed for both the LE and the QMLE using a simple bootstrap method. These prediction intervals along with point estimate will help practitioners to evaluate the forecasting performance of their models. Monte Carlo results showed that our bootstrap method generates reliable prediction intervals. We found that although both estimators provide good mean coverage, the LE can be considered superior in terms of its mean lengths close to empirical with low standard errors. The bootstrapped prediction intervals for volatilities and VaR capture the asymmetry commonly present in real data sets.

This chapter is organised as follows. In Section 6.2 we define the LE for ARCH model and Monte Carlo simulations are performed to check the accuracy in estimating the parameters of ARCH models. The forecasts of volatility are also obtained and application to real data sets is presented. In Section 6.4 using a weighted bootstrap method the true distribution of the LE in finite sample is approximated. Section 6.5 describes methods for constructing bootstrap prediction intervals for returns, volatilities and value-at-risk in ARCH models. Finally, Section 6.6 concludes the chapter.

6.2 The Linear Estimator

Consider the following ARCH model where one observes $\{X_t; 1-p \leq t \leq T\}$ satisfying

$$X_t = h_t^{1/2}(\boldsymbol{\beta})\epsilon_t; \qquad 1 \leq t \leq T, \tag{6.1}$$

where $\boldsymbol{\beta} = [\beta_0, \beta_1, ..., \beta_p]'$ is the unknown parameter to be estimated with $\beta_0 > 0$, $\beta_j \geq 0$, $1 \leq j \leq p$,

$$h_t = \beta_0 + \beta_1 X_{t-1}^2 + ... + \beta_p X_{t-p}^2,$$

with $\{\epsilon_t;\ 1 \leq t \leq T\}$ are i.i.d with mean zero and unit variance. It is assumed that $\{\epsilon_t;\ 1 \leq t \leq T\}$ are independent of $\{X_t;\ 1-p \leq t \leq 0\}$. It is also assumed that (6.1) holds, $\{X_t; t \geq 1-p\}$ is a stationary and ergodic and $E(\epsilon^4) < \infty$. These will be called model assumptions.

Let $Y_t = X_t^2, 1-p \leq t \leq T$,

$$\mathbf{Z}_{t-1} = [1, Y_{t-1}, ..., Y_{t-p}]' = [1,\ X_{t-1}^2, ..., X_{t-p}^2]',$$

and $\eta_t = \epsilon_t^2 - 1$, $1 \leq t \leq T$. Then squaring both sides of (6.1) and using the form

$h_{t-1}(\boldsymbol{\beta}) = \mathbf{Z}'_{t-1}\boldsymbol{\beta}$, we get

$$Y_t = \mathbf{Z}'_{t-1}\boldsymbol{\beta} + h_{t-1}(\boldsymbol{\beta})\eta_t \quad 1 \le t \le T, \tag{6.2}$$

where $E\{h_{t-1}(\boldsymbol{\beta})\eta_t\} = E\{h_{t-1}(\boldsymbol{\beta})\}E(\eta_t) = 0$, $1 \le t \le T$.

Bose and Mukherjee (2003) define a preliminary least squares estimator $\hat{\boldsymbol{\beta}}_{pr}$ of $\boldsymbol{\beta}$ as the solution of

$$\sum_{t=1}^{T} \left[\mathbf{Z}_{t-1}\{Y_t - \mathbf{Z}'_{t-1}\boldsymbol{\beta}\}\right] = \mathbf{0}, \tag{6.3}$$

which yields the estimator

$$\hat{\boldsymbol{\beta}}_{pr} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y},$$

where $\mathbf{Z}$ is the $T \times (1+p)$ matrix whose t-th row equals $\mathbf{Z}'_{t-1}$ and $\mathbf{Y}$ is the vector with t-th entry Y_t, $1 \le t \le T$.

An improved estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ can be obtained as follows. Dividing (6.2) by $h_{t-1}(\boldsymbol{\beta})$, we get

$$\frac{Y_t}{\mathbf{Z}'_{t-1}\boldsymbol{\beta}} = \left\{\frac{\mathbf{Z}_{t-1}}{\mathbf{Z}'_{t-1}\boldsymbol{\beta}}\right\}'\boldsymbol{\beta} + \eta_t$$

Now replacing $\mathbf{Z}'_{t-1}\boldsymbol{\beta}$ by $\mathbf{Z}'_{t-1}\hat{\boldsymbol{\beta}}_{pr}$ yields

$$\frac{Y_t}{\mathbf{Z}'_{t-1}\hat{\boldsymbol{\beta}}_{pr}} \approx \left\{\frac{\mathbf{Z}_{t-1}}{\mathbf{Z}'_{t-1}\hat{\boldsymbol{\beta}}_{pr}}\right\}'\boldsymbol{\beta} + \eta_t.$$

Therefore, a linear estimator of $\boldsymbol{\beta}$ is defined as the solution of

$$\sum_{t=1}^{T} \left[\left\{\mathbf{Z}_{t-1}/(\mathbf{Z}_{t-1}\hat{\boldsymbol{\beta}}_{pr})\right\}\left\{Y_t/(\mathbf{Z}'_{t-1}\hat{\boldsymbol{\beta}}_{pr}) - \{\mathbf{Z}_{t-1}/(\mathbf{Z}'_{t-1}\hat{\boldsymbol{\beta}}_{pr})\}'\boldsymbol{\beta}\right\}\right] = \mathbf{0}, \tag{6.4}$$

yielding the linear estimator

$$\hat{\boldsymbol{\beta}}_T = \left[\sum_{t=1}^{T}\left\{\frac{\mathbf{Z}_{t-1}\mathbf{Z}'_{t-1}}{(\mathbf{Z}'_{t-1}\hat{\boldsymbol{\beta}}_{pr})^2}\right\}\right]^{-1}\left[\sum_{t=1}^{T}\left\{\frac{\mathbf{Z}_{t-1}Y_t}{(\mathbf{Z}'_{t-1}\hat{\boldsymbol{\beta}}_{pr})^2}\right\}\right]. \tag{6.5}$$

It is shown in Bose and Mukherjee (2003) that under the model assumptions

$$T^{1/2}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) \rightarrow N\Big[\mathbf{0}, \mathrm{Var}(\epsilon_1^2)\Big\{\mathrm{E}\{\mathbf{Z}_0\mathbf{Z}_0'(\boldsymbol{\beta}'\mathbf{Z}_0)^{-2}\}\Big\}^{-1}\Big]. \tag{6.6}$$

The linear estimator has a closed form expression. It does not require any optimisation method for solution and hence is easy to compute. Although this method is very fast and has better or at least the same performance as QMLE, the estimated parameters are not guaranteed to be nonnegative. Applying the nonnegativity constraint on parameters of ARCH models may solve this problem. We do not pursue in this study and leave it for future research.

Bose and Mukherjee (2003) performed a small simulation study to check the finite sample performance of these estimators. In next sections, we explore these estimators in detail. We check the performance of these estimators both in terms of estimating the parameters and forecasting the volatility under different error distributions and small sample sizes.

6.3 Simulation and Empirical Results

In this section we use Monte Carlo simulations to check the accuracy of estimated parameters of ARCH model estimated by the LE. We also evaluate the volatility forecasts of LE and QMLE. Application to real data set is also presented.

6.3.1 Monte Carlo Simulations

First we report the results of simulations performed to check the accuracy of the estimated parameters.

Parameters Estimation

We investigate the relative performance based on two frequently used accuracy measures, the mean squared error (MSE) and the mean absolute error (MAE) where estimates of the MSE and the MAE are obtained as

$$\text{MSE} = \frac{1}{K}\sum_{k=1}^{K}\Big\{\sum_{j=0}^{p}(\hat{\beta}_j - \beta_j)^2\Big\},$$

$$\text{MAE} = \frac{1}{K}\sum_{k=1}^{K}\Big\{\sum_{j=0}^{p}|\hat{\beta}_j - \beta_j|\Big\},$$

where $\hat{\beta}_j$ and β_j for $0 \leq j \leq p$, are the estimated and true parameters, respectively, of the ARCH(p) model at the k-th replication.

We use simulations to estimate these quantities for linear estimator and the QMLE. All Monte Carlo simulations are based on $K = 1000$ independent replicates each of sample size T, from ARCH(p) model. The errors are generated from three different distributions (i) the standard normal distribution (ii) contaminated or mixture normal distribution $(1-\epsilon)\Phi(x) + \epsilon\Phi(x/\sigma)$ with $\epsilon = 0.05$, and $\sigma^2 = 9$, and standardised student-t distribution with 3 degrees of freedom. We discard the initial 500 simulated values and only the last T observations are kept to reduce the impact of initial values.

In the first experiment, 1000 replicates each of sample size T = 50, 100 and 500 are generated from the following ARCH(2) model.

$$\begin{aligned} \text{Model A:} \quad X_t &= h_t^{1/2}\epsilon_t, \\ h_t &= 0.1 + 0.4X_{t-1}^2 + 0.2X_{t-2}^2. \end{aligned}$$

Errors are generated from three distributions mentioned above. Student-t distributions are standardised to have mean 0 and variance 1. We use 'garchfit', a function of MATLAB GARCH Toolbox for estimating ARCH models and write our own code for the estimation of the LE. The results of the MSE and the MAE for both estimators along with their standard errors in parentheses are shown in Table 6.1. Entries in bold represent best performing estimator.

Table 6.1: Mean Squared Error and Mean Absolute Error of the ARCH(2) model.

	QMLE	LE	QMLE	LE	QMLE	LE
Standard Normal	$T = 50$		$T = 100$		$T = 500$	
MSE	0.0776	**0.0674**	0.0526	**0.0507**	0.0127	**0.0126**
	(0.0789)	(0.0577)	(0.0548)	(0.0450)	(0.0204)	(0.0130)
MAE	0.3425	**0.3318**	**0.2783**	0.2800	**0.1335**	0.1370
	(0.1718)	(0.1519)	(0.1440)	(0.1314)	(0.0853)	(0.0698)
Mixture Normal	$T = 50$		$T = 100$		$T = 500$	
MSE	0.1026	**0.0814**	0.0887	**0.0709**	0.0822	**0.0560**
	(0.1674)	(0.0739)	(0.1188)	(0.0652)	(0.1276)	(0.0627)
MAE	0.3892	**0.3652**	0.3515	**0.3351**	0.3146	**0.2861**
	(0.2147)	(0.1725)	(0.2115)	(0.1703)	(0.2246)	(0.1659)
Student-t(3)	$T = 50$		$T = 100$		$T = 500$	
MSE	0.1673	**0.1091**	0.1461	**0.1087**	0.0829	**0.0676**
	(0.2405)	(0.0857)	(0.1499)	(0.0831)	(0.1023)	(0.0625)
MAE	0.5033	**0.4297**	0.4701	**0.4274**	0.3403	**0.3279**
	(0.2596)	(0.1804)	(0.2356)	(0.1781)	(0.1868)	(0.1506)

It can be seen from Table 6.1 that for all sample sizes and all errors distributions considered in this study, the MSEs of the LE are found smaller than the widely-used QMLE. Results from the table also show that the LE provides better estimates than the QMLE for very small sample size such as $T = 50$. The only case when the MSE of the QMLE is found closer to that of the LE, but with large standard errors, is when errors are generated from the standard normal distribution and the sample size of $T = 500$ is considered. This trend does not hold for other errors distributions although increasing sample sizes improve the MSE and the MAE of both estimators. The MAEs of the QMLE are found better than the MAEs of the LE for standard normal distribution only at sample sizes $T = 250$ and 500 but again with large standard errors. On all other occasions the MAEs of the LE are found better than that of the QMLE.

In the second experiment, 1000 replicates each of sample size $T = 250$, 500 and 1000 are generated from the following ARCH(3) model

$$\text{Model B:} \quad X_t = h_t^{1/2}\epsilon_t,$$
$$h_t = 0.01 + 0.2X_{t-1}^2 + 0.2X_{t-2}^2 + 0.1X_{t-3}^2,$$

and errors are generated from the standard normal distribution, contaminated normal distribution and standardised student-t-distribution with 3 degrees of freedom. Table 6.2 reports the results of 1000 independent replications of this experiment.

Table 6.2: Mean Squared Error and Mean Absolute Error of the ARCH(3) model.

	QMLE	LE	QMLE	LE	QMLE	LE
Standard Normal	$T = 250$		$T = 500$		$T = 1000$	
MSE	0.0216	**0.0207**	0.0119	**0.0119**	**0.0066**	0.0066
	(0.0180)	(0.0177)	(0.0100)	(0.0098)	(0.0055)	(0.0056)
MAE	0.2029	**0.2010**	**0.1519**	0.1530	**0.1122**	0.1125
	(0.0899)	(0.0870)	(0.0674)	(0.0660)	(0.0504)	(0.0511)
Mixture Normal	$T = 250$		$T = 500$		$T = 1000$	
MSE	0.0621	**0.0345**	0.0646	**0.0381**	0.0626	**0.0389**
	(0.1224)	(0.0417)	(0.1148)	(0.0483)	(0.1006)	(0.0498)
MAE	0.2902	**0.2569**	0.2918	**0.2625**	0.3030	**0.2646**
	(0.1972)	(0.1207)	(0.2012)	(0.1370)	(0.1950)	(0.1358)
Student-t(3)	$T = 250$		$T = 500$		$T = 1000$	
MSE	0.0920	**0.0525**	0.0753	**0.0465**	0.0472	**0.0355**
	(0.1279)	(0.0578)	(0.1096)	(0.0546)	(0.0695)	(0.0401)
MAE	0.3804	**0.3232**	0.3452	**0.3031**	0.2800	**0.2633**
	(0.2113)	(0.1401)	(0.1877)	(0.1311)	(0.1431)	(0.1116)

Table 6.2 reveals same features observed in Table 6.1. When errors are generated from the standard normal distribution and sample size becomes larger, the MSEs of the QMLE get closer to the MSEs of the LE and the MAEs of the QMLE

outperforms the MAEs of the LE but with larger standard errors. In cases of mixture normal distribution and student-t distribution with 3 degrees of freedom the LE performs better both in terms of the MSE and the MAE. It can be noticed that increasing the sample size improves both measures but the LE remains the best choice.

Financial time series are often heavy-tailed and have outliers. These results show that the LE can be applied to financial data sets as they provide better results than the QMLE in terms of the MSE and the MAE. Even for very small sample sizes where QMLE fails to provide good estimates the LE produces reasonable estimates of the parameters of the ARCH model.

Volatility Forecasts

Forecasting the volatility of financial time series is of great importance to practitioners and researchers. The issue of volatility forecasting is also crucial for policy makers, option traders and investors, since volatility forecasts can be used for calculation of the risk measures such as the value-at-risk. Financial time series often exhibits volatility clustering and leptokurtosis. The most popular class of econometric models for describing these empirical facts are the ARCH/GARCH models.

Next we turn our attention to the volatility forecasting performance of linear estimator. The purpose of this study is to compare the volatility forecasts of the LE with that of the QMLE. Although results from previous section show that the LE estimates the parameters as accurately as the QMLE and even better in the cases of small sample size and non-normal distribution, it is worth checking their predictive ability under different error distributions and small sample size.

First, we describe the procedure of obtaining the forecasts of volatility from an

ARCH model. Consider an ARCH(p) model

$$X_t = h_t^{1/2}\epsilon_t \quad \text{and} \quad h_t = \beta_0 + \beta_1 X_{t-1}^2 + \cdots + \beta_p X_{t-p}^2.$$

The one-step ahead volatility forecast based on the ARCH(p) is defined as

$$\hat{h}_{T+1} = \hat{\beta}_0 + \hat{\beta}_1 X_T^2 + \cdots + \hat{\beta}_p X_{T+1-p}^2, \tag{6.7}$$

where $\hat{\beta}_j$, $0 \leq j \leq p$ are estimated parameters of ARCH(p) model.

The s-step ahead forecast h_{T+s} for volatility is

$$\hat{h}_{T+s} = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j \hat{h}_{T+s-j}, \tag{6.8}$$

where $\hat{h}_{T+s-j} = X_{T+s-j}^2$ if $s - j \leq 0$.

Initially the ARCH(p) model is estimated over the in-sample period. The estimated parameters $\hat{\beta}_j$, $0 \leq j \leq p$ are then used to obtain one-step ahead forecasts. For simulation study we use the *rolling scheme* to generate volatility forecasts. In the rolling scheme the sample is rolled forward one day, the model parameters are re-estimated, and these new estimates are used to forecast one-step ahead forecast of volatility. In this way the volatility forecasts for the remaining out-of-sample period is generated. The moving window modelling and forecasting procedure is repeated until the end of the data is reached.

The parameter vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)'$ is estimated by the QMLE and LE. After generating one-step-ahead forecasts of volatilities of ARCH model based on the estimated parameters, the forecasting performance of both QMLE and LE are evaluated. Evaluation criteria used in this study are the mean absolute prediction error (MAPE) and the root mean squared error (RMSE). These are defined as follows:

$$\text{MAPE} = \frac{1}{l}\sum_{t=T+1}^{T+l}\left|\hat{h}_t - h_t\right|$$

$$\text{RMSE} = \sqrt{\frac{1}{l}\sum_{t=T+1}^{T+l}(\hat{h}_t - h_t)^2},$$

where l is the number of out-of-sample data, $h_t\,(T+1 \leq t \leq T+l)$ are the volatilities generated form the ARCH model using the true parameter values and $\hat{h}_t$ are the forecasted volatility. The MAPE are used as this measure is more robust to outliers than the mean squared prediction error (MSPE). The RMSE is a standard measure used for evaluation.

We use Model A and Model B from previous studies to compare the volatility forecasts of the LE with the QMLE. From both models we generate $T+l$ observation, where the number of out-of-sample observations are set to 20 i.e., $l = 20$. The errors are generated from three different distributions as in the earlier study. These are the standard normal, mixture normal and the standardised student-t distribution with 3 df. The rolling window of size T is used to estimate the parameter and one-day ahead volatility forecast is made. Then the first observation is dropped and $(T+1)$-th observation is included in the sample and parameters are re-estimated and again next day volatility is predicted. This process is repeated till we reach at the $(T+l)$-th observation. Table 6.3 presents the results of the MAPE and the RMSE of both the LE and the QMLE. All results are based on 1000 replications.

Results of the the Table 6.3 indicate that for Model A, when sample size is very small $T = 50$, LE forecasts the volatility better than the QMLE both in terms of the MAPE and RMSE. This hold for all error distributions considered in this study. This suggests that for very small sample sizes, LE is a better choice than the QMLE not only for estimating the parameters of the ARCH model but also for forecasting one-step-ahead volatility. When the sample size is increased to $T = 250$ the results from Model B show that the predictive performance of the QMLE gets better although the LE still performs well for the mixture normal distribution. The study for volatility forecasts shows that for small sample sizes and non-normal

Table 6.3: Performance of volatility forecast in terms of mean absolute prediction error and root mean squared error.

	QMLE	LE	QMLE	LE	QMLE	LE
Model A: $T = 50$	*Standard Normal*		*Mixture Normal*		*Student-t(3)*	
MAPE	0.2043	**0.1642**	0.1596	**0.1523**	0.1491	**0.1437**
	(0.1628)	(0.1038)	(0.0308)	(0.0311)	(0.0375)	(0.0283)
RMSE	0.4097	**0.3698**	**0.5963**	0.5957	0.5442	**0.5412**
Model B: $T = 250$	*Standard Normal*		*Mixture Normal*		*Student-t(3)*	
MAPE	**0.0130**	0.0202	0.0078	**0.0077**	**0.1186**	0.1199
	(0.0101)	(0.0021)	(0.0017)	(0.0016)	(0.1176)	(0.1233)
RMSE	**0.0204**	0.0241	0.0301	**0.0301**	**0.1674**	0.1721

distributions, the LE may be preferred over the widely-used QMLE. This estimator also performs well for large sample size in the presence of outliers in the data. These findings suggest the use of the LE in those cases.

6.3.2 Empirical Illustration

The forecasting performance of both the QMLE and the LE is also checked through applications to real data sets. The SP500, FTSE100 and NIKKEI225 indices are used in this study (see section 5.5 for summary statistics of these indices). Both *rolling window* and *recursive scheme* are used to generate 2000 one-step ahead volatility forecasts. By looking at the sample partial autocorrelation plots of all three indices we choose ARCH models of order $p = 5$, 3 and 6 for SP500 index, FTSE100 index and NIKKEI225 index, respectively as the higher lag correlation are not found significant.

In addition to the mean absolute prediction error (MAPE) and the root mean square error (RMSE), the mean squared prediction error (MSPE) is also used to evaluate predictive performance of both estimators. The MSPE is another criterion often used to estimate the error between estimates and true values. The MSPE is defines as

$$\text{MSPE} = \frac{1}{l}\sum_{t=T+1}^{T+l}\left(\hat{h}_t - h_t\right)^2,$$

where $\hat{h}_t$ are the estimated volatilities from the ARCH model. Since the true volatilities, h_t, are unobservable, the squared returns can be used as proxies for these true volatilities to evaluate one-day ahead forecasts. The daily squared return is conditionally an unbiased estimator of the daily conditional variance. Another conditionally unbiased estimator for the daily conditional volatility is the 'realised volatility' computed from high-frequency intra-day returns (see Andersen and Bollerslev, 1998). In this study we use the squared returns as proxies for volatilities as the intra-day observations are not available. Table 6.4 reports the results of these evaluation criteria using both forecasting schemes.

Table 6.4: Evaluating volatility forecast for each stock price index.

	QMLE	LE	QMLE	LE	QMLE	LE
Recursive Scheme	*S&P500*		*FTSE100*		*NIKKEI225*	
MSPE	0.2973	**0.2955**	0.2772	**0.2711**	0.6579	**0.6564**
MAPE	0.2702	**0.2671**	0.2677	**0.2627**	0.4489	**0.4476**
RMSE	0.5452	**0.5436**	0.5265	**0.5207**	0.8111	**0.8102**
Rolling Scheme	*S&P500*		*FTSE100*		*NIKKEI225*	
MSPE	0.3023	**0.3004**	0.2730	**0.2718**	0.6555	**0.6540**
MAPE	0.2817	**0.2795**	0.2786	**0.2720**	0.4469	**0.4442**
RMSE	0.5498	**0.5481**	0.5225	**0.5213**	0.8096	**0.8087**

As we can see from Table 6.4 that the LE provides better forecasts than the QMLE based on the standard evaluation measures. Not only MSPE but also MAPE are found the least for the LE. These results hold for both recursive and rolling schemes and for all three data sets used in this study. Based on these results and also looking at the results of the RMSE, we can say that the LE may be used instead of the widely-used QMLE for the estimation of ARCH models as this estimator not only estimate the parameters well but also provides better forecasts.

6.4 A Weighted Resampling for the Linear Estimator in ARCH Models

This section deals with resampling methods for ARCH models estimated by linear estimator. Chatterjee and Bose (2005) developed the idea of weighted bootstrap of estimators that have been obtained as minimisers via solution of equations in general dependent models. We use the idea of weighted resampling to develop suitable bootstrap version of the linear estimator.

Let $\{w_{Tt}; 1 \leq t \leq T, T \geq 1\}$ be a triangular array of r.v.'s such that for each $T \geq 1, \{w_{Tt}; 1 \leq t \leq T\}$ are exchangeable, independent of $\{X_t; t \geq 1-p\}$ and $\{\epsilon_t; t \geq 1\}$ and $E(w_{Tt}) = 1$. These are called the bootstrap weights. The bootstrap preliminary least squares estimator $\hat{\boldsymbol{\beta}}^*_{pr}$ of $\boldsymbol{\beta}$ is defined by mimicking (6.3), as the solution of

$$\sum_{t=1}^{T} w_{Tt}\Big[\mathbf{Z}_{t-1}\{Y_t - \mathbf{Z}'_{t-1}\boldsymbol{\beta}\}\Big] = \mathbf{0}. \tag{6.9}$$

Similarly, as in (6.4), the bootstrapped linear estimator $\hat{\boldsymbol{\beta}}^*_T$, may be defined as a solution of

$$\sum_{t=1}^{T} w_{Tt}\Big[\Big\{\mathbf{Z}_{t-1}/(\mathbf{Z}_{t-1}\hat{\boldsymbol{\beta}}^*_{pr})\Big\}\Big\{Y_t/(\mathbf{Z}'_{t-1}\hat{\boldsymbol{\beta}}^*_{pr}) - \{\mathbf{Z}_{t-1}/(\mathbf{Z}'_{t-1}\hat{\boldsymbol{\beta}}^*_{pr})\}'\boldsymbol{\beta}\Big\}\Big] = \mathbf{0}, \tag{6.10}$$

which gives

$$\hat{\boldsymbol{\beta}}_T^* = \left[\sum_{t=1}^{T} w_{Tt}\left\{\frac{\mathbf{Z}_{t-1}\mathbf{Z}'_{t-1}}{(\mathbf{Z}'_{t-1}\hat{\boldsymbol{\beta}}_{pr})^2}\right\}\right]^{-1}\left[\sum_{t=1}^{T} w_{Tt}\left\{\frac{\mathbf{Z}_{t-1}Y_t}{(\mathbf{Z}'_{t-1}\hat{\boldsymbol{\beta}}_{pr})^2}\right\}\right]. \tag{6.11}$$

We are interested in approximating the distribution of $\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})$ by the distribution of $\sqrt{T}(\hat{\boldsymbol{\beta}}_T^* - \hat{\boldsymbol{\beta}}_T)$. We approximate such distribution via weighted bootstrap. Three different schemes for weights are considered. These are

(i) Scheme M when weights have a multinomial $(T, 1/T, ..., 1/T)$ distribution.

(ii) Scheme U when $w_{Tt} = U_t/\overline{U}$, where U_t's are i.i.d Uniform(0.5, 1.5) and $\overline{U} = T^{-1}\sum_{t=1}^{T} U_t$.

(iii) Scheme E when $w_{Tt} = E_t/\overline{E}$, where E_t's are i.i.d Exponential(1) $\overline{E} = T^{-1}\sum_{t=1}^{T} E_t$.

We also consider residual bootstrap when standardised residuals are bootstrapped to form a new bootstrapped return series. Using this bootstrapped series the bootstrapped parameters are estimated and the bootstrapped distributions of the parameters are obtained. It is also possible to obtain quantiles of the bootstrap distribution of $\sigma_T^{-1}\sqrt{T}(\hat{\boldsymbol{\beta}}_T^* - \hat{\boldsymbol{\beta}}_T)$ using simulation and then using the bootstrap approximation, we can construct the bootstrap confidence intervals of $\boldsymbol{\beta}$.

6.4.1 Simulation Results

This section reports the results of a Monte Carlo simulation. We investigate the quality of bootstrap approximation to the finite sample distribution of $\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})$. We use a sample of size T, and assume that the underlying error distributions of $\{\epsilon_t\}$ be standard normal. An ARCH(p) model is fitted to the data set using linear estimator.

In our first experiment, we generate $K = 10,000$ samples each of size $T = 50, 250$, and 500 from the ARCH(2) model with $\boldsymbol{\beta} = (0.1, 0.2, 0.3)'$. Let $\hat{\boldsymbol{\beta}}_{T(k)} =$

$(\hat{\beta}_{T0}, \hat{\beta}_{T1}, \hat{\beta}_{T3})'$ denote the vector of estimated parameters computed from the k-th sample, $1 \leq k \leq K$. For each replication we compute $\sqrt{T}(\hat{\beta}_{Tj} - \beta_j)$, $0 \leq j \leq p = 2$. The mean and average of the squares of the three sets over K replications represent the mean and the mean squared error (MSE) of $\sqrt{T}(\hat{\beta}_{Tj} - \beta_j)$, $0 \leq j \leq p = 2$. The corresponding histograms approximate the marginal distributions of $\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})$.

The estimates of means under normal approximation are zero. The estimate of MSE using the normal approximation is obtained by averaging over K estimated MSEs where the k-th $(1 \leq k \leq K)$ estimate is obtained from the diagonals of the matrix $\hat{V}[T^{-1}\sum_{t=1}^{T}\{\mathbf{Z}_{t-1}\mathbf{Z}_{t-1}'(\hat{\boldsymbol{\beta}}_T'\mathbf{Z}_{t-1})^{-2}\}]^{-1}$; here $\hat{V}$ is the variance of $\{\epsilon_1^2, \cdots, \epsilon_T^2\}$, where $\hat{\epsilon}_t = X_t/(\hat{\boldsymbol{\beta}}_T'\mathbf{Z}_{t-1})^{1/2}$, $1 \leq t \leq T$, $\hat{\boldsymbol{\beta}}_T$ being the estimate based on the k-th replication.

Table 6.5 reports the results of first experiment. The true means are found significantly different from the normal approximation. The MSE for small samples are also different from normal approximation values. It can be seen from the table that $\hat{\beta}_1$ and $\hat{\beta}_2$ underestimate β_1 and β_2, respectively. For large sample size the MSE of the distributions of $\hat{\beta}_1$ and $\hat{\beta}_2$ are very close to the MSE of the distribution of these estimates under normal approximations.

Table 6.5: Means and the MSEs of the distributions of $\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})$ for ARCH(2) model and the MSE due to normal approximation of the distribution.

	$T = 50$			$T = 250$			$T = 500$		
	Mean	MSE	MSE_N	Mean	MSE	MSE_N	Mean	MSE	MSE_N
$\beta_0 = 0.1$	0.0846	0.0919	0.0953	0.0918	0.0889	0.0775	0.0749	0.0862	0.0755
$\beta_1 = 0.2$	-0.0558	0.9734	2.2851	-0.2615	1.9965	2.0761	-0.2445	2.2800	2.0912
$\beta_2 = 0.3$	-0.2997	1.4903	2.7870	-0.4635	3.0102	2.6626	-0.3664	3.1427	2.7431

All results are based on 10,000 replications. MSE is the MSE under normal approximation.

Next we turn our attention to bootstrap approximations. To approximate the distribution of $\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})$, we proceed as follows. We choose and fix $\hat{\boldsymbol{\beta}}_{T(r)}$ $1 \leq r \leq R$, $R \leq K$. In this study we generate $B = 999$ bootstrap samples. Bootstrap results are based on $R = 100$ replications. For weighted resampling, these bootstrap samples are generated based on weights under three schemes, Scheme M, Scheme U and Scheme E, after fixing $\hat{\boldsymbol{\beta}}_{T(r)}$ $1 \leq r \leq R$. For the b-th sample, $1 \leq b \leq B$, we compute $\sigma_T^{-1} T^{1/2}(\hat{\boldsymbol{\beta}}^*_{T(b)} - \hat{\boldsymbol{\beta}}_{T(r)})$. For residual bootstrap, we generate $B = 999$ bootstrap samples and for the b-th sample, $1 \leq b \leq B$, compute $T^{1/2}(\hat{\boldsymbol{\beta}}^*_{T(b)} - \hat{\boldsymbol{\beta}}_{T(r)})$, after fixing $\hat{\boldsymbol{\beta}}_{T(r)}$ $(1 \leq r \leq R)$.

Table 6.6 reports the results of means and the MSEs of the distribution of the standardised bootstrap estimators under residual bootstrap and three different schemes. Entries in bold represent MSEs that provide the closest approximations to those in Table 6.5. It can be observed from the results in Table 6.6 that the residual bootstrap approximate the means better than the other bootstrap schemes especially for the large sample size. The MSEs of Scheme M shows good approximations. The bootstrap means are different from true means in all cases. For sample size $T = 250$, the means of all three schemes are significantly different from true means. The means from residual bootstrap provide better approximation as compared to weighted resampling schemes.

In terms of MSEs, Scheme E provides good bootstrapped approximations for sample size of 50 and Scheme M for sample size 500. Scheme U shows better results for two cases and the residual bootstrap outperform other schemes just once. These results show that although there is no clear cut winner in terms of MSEs, Scheme E and Scheme U can be considered as alternatives to Scheme M and residual bootstrap.

The kernel density plots of the distributions of $\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})$ and their bootstrap approximations $\sigma_T^{-1}\sqrt{T}(\hat{\boldsymbol{\beta}}^*_T - \hat{\boldsymbol{\beta}}_T)$ under different schemes for sample size $T = 500$ are shown in Fig. 6.1. For residual bootstrap we fix $\sigma = 1$. Fig. 6.1 shows that the density plots of the distributions of $\sigma_T^{-1}\sqrt{T}(\hat{\boldsymbol{\beta}}^*_T - \hat{\boldsymbol{\beta}}_T)$ based on B bootstrap samples are very similar in shape and characteristics to the plots of the distributions of $\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})$. All bootstrap schemes show good approximations for β_0. A

Table 6.6: Means and the MSEs of the distribution of the standardised bootstrap estimators for ARCH(2) model under different schemes.

$B = 999$	Scheme M		Scheme U		Scheme E		Residual Boot	
$T = 50$	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
$\beta = 0.1$	-0.0632	0.1898	-0.0210	0.1263	-0.0680	**0.0808**	0.0369	0.1053
$\beta = 0.2$	0.3063	1.1677	0.1548	1.7945	0.2792	**1.0724**	0.1370	1.0903
$\beta = 0.3$	-0.0130	1.4138	-0.0464	2.6676	-0.0492	1.3081	-0.1928	**1.4259**
$T = 250$								
$\beta = 0.1$	0.0118	**0.0872**	0.0097	0.0909	0.0115	0.0825	0.0837	0.0975
$\beta = 0.2$	0.0087	2.0606	-0.0304	2.3983	-0.0096	**1.9393**	-0.1827	2.0128
$\beta = 0.3$	-0.1300	2.6268	-0.0463	**2.8504**	-0.1363	2.3938	-0.3203	2.5703
$T = 500$								
$\beta = 0.1$	0.0290	**0.0869**	0.0100	0.0851	0.0273	0.0831	0.0737	0.0883
$\beta = 0.2$	-0.0649	2.3234	-0.0525	**2.3112**	-0.0708	2.0862	-0.1863	2.1176
$\beta = 0.3$	-0.1770	**3.0550**	-0.0476	2.9881	-0.1614	2.8640	-0.3400	2.9880

All results are based on 100 replications.

close inspection of the plots reveals that both Scheme U Scheme E serve as good alternatives to other commonly-used bootstrap methods.

In the second experiment, we generate $K = 10,000$ samples each of size $T = 50, 250$, and 500 from the ARCH(3) model with $\boldsymbol{\beta} = (0.01, 0.1, 0.2, 0.2)'$. For each replication we compute $\sqrt{T}(\hat{\beta}_{Tj} - \beta_j)$, $0 \leq j \leq p = 3$. The mean and average of the squares of the four sets over K replications represent the mean and the mean squared error (MSE) of $\sqrt{T}(\hat{\beta}_{Tj} - \beta_j)$, $0 \leq j \leq p = 3$.

Table 6.7 shows results of means, MSEs and MSE under normal approximations. It can be seen that the true means of the distributions of all parameters except β_0, are significantly different from the normal approximation values. The MSEs for small sample sizes are also found different than the normal approximations. For $T = 500$, the values of the MSE match that of the MSE under normal approximations.

Using weighted resampling method and residual bootstrap, we generate $B = 999$ bootstrap samples. Bootstrap approximations of means and MSEs are computed. All results are based on 100 replications. These results are reported in Table 6.8.

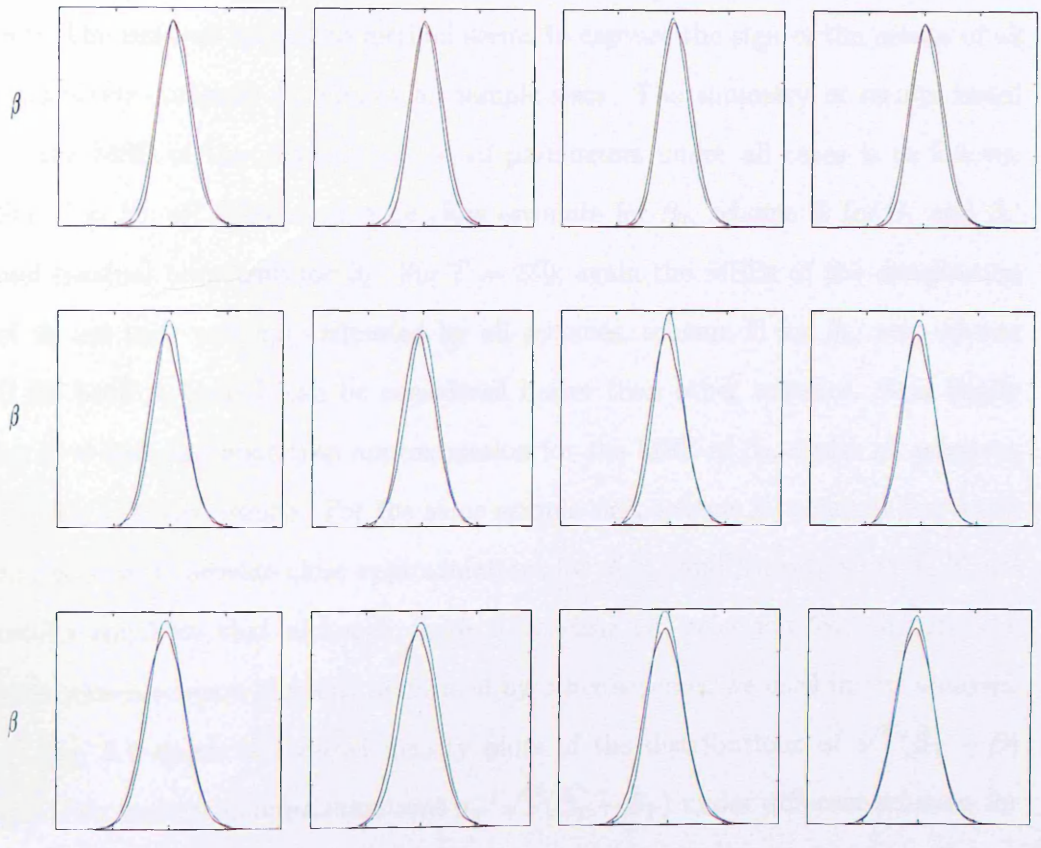

Figure 6.1: Kernel density plots of the distributions of $\sqrt{T}(\hat{\beta}_T - \beta)$ (red) along with their bootstrap approximations (blue) for ARCH(2) model ($T = 500$).

Table 6.7: Means and the MSEs of the distributions of $\sqrt{T}(\hat{\beta}_T - \beta)$ for ARCH(3) model and the MSE due to normal approximation of the distribution.

	$T = 50$			$T = 250$			$T = 500$		
	Mean	MSE	MSE	Mean	MSE	MSE	Mean	MSE	MSE
β = 0.01	0.0043	0.0010	0.0009	0.0096	0.0011	0.0010	0.0096	0.0011	0.0010
β = 0.1	0.3106	0.7253	2.3588	0.1155	1.1062	1.6271	0.0126	1.3101	1.5818
β = 0.2	-0.1028	0.8923	2.6574	-0.4087	2.0168	2.0555	-0.3579	2.2632	2.0880
β = 0.2	-0.1119	0.9461	2.6099	-0.3669	2.0833	2.0862	-0.3583	2.2986	2.1263

All results are based on $K = 10,000$ replications. MSE is the MSE under normal approximation.

The bootstrapped approximations of means do not match the corresponding estimated means in most of the cases except under the residual bootstrap at $T = 500$. The residual bootstrap method seems to capture the sign of the means of all parameters correctly for almost all sample sizes. The summary of results based on the MSE of the distributions of all parameters under all cases is as follows: For $T = 50$, all schemes provide close estimate for β_0, scheme E for β_1 and β_3, and residual bootstrap for β_2. For $T = 250$, again the MSEs of the distribution of β_0 are very well approximated by all schemes, scheme E for β_1, and scheme U for both β_2 and β_3 can be considered better than other schemes. And finally for $T = 500$, the bootstrap approximation for the MSE of β_0, under all schemes, provide accurate results. For the same sample size, scheme E, residual bootstrap and scheme U provide close approximations for β_1,β_2, and β_3, respectively. These results conclude that although there is no clear cut selection for schemes, the widely used scheme M is out performed by other schemes we used in our analysis.

Fig. 6.2 shows the kernel density plots of the distributions of $\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})$ and their bootstrap approximations $\sigma_T^{-1}\sqrt{T}(\hat{\boldsymbol{\beta}}_T^* - \hat{\boldsymbol{\beta}}_T)$ under different schemes for sample size $T = 500$. Again for the residual bootstrap we fix $\sigma_T = 1$. It can be noticed from the figure that although all schemes capture the shape of the distribution reasonably well, the bootstrap approximations based on the residual bootstrap and Scheme U seem to provide better fit.

We observe from this study that weighted bootstrap schemes work well for

Table 6.8: Means and the MSEs of the distribution of the standardised bootstrap estimators for ARCH(3) model under different schemes.

$B = 999$	Scheme M		Scheme U		Scheme E		Residual Boot	
$T = 50$	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
$\beta = 0.1$	-0.0079	**0.0009**	-0.0044	0.0013	-0.0086	0.0008	0.0076	0.0035
$\beta = 0.3$	0.2084	0.8097	0.1008	**1.4651**	0.1756	0.7690	0.0759	0.8191
$\beta = 0.2$	0.1654	0.9205	0.0769	**1.7891**	0.1307	0.8601	0.0736	0.8906
$\beta = 0.1$	0.0227	1.0108	0.0317	**2.1055**	0.0044	1.0082	-0.0898	1.1000
$T = 250$								
$\beta = 0.1$	-0.0052	0.0010	-0.0008	**0.0011**	-0.0045	0.0009	0.0047	0.0010
$\beta = 0.3$	0.2303	1.2915	0.0588	1.4598	0.1945	**1.2032**	0.1082	1.2416
$\beta = 0.2$	0.0982	1.6402	0.0106	**1.8967**	0.0613	1.5235	-0.1197	1.5939
$\beta = 0.1$	-0.0288	1.8742	-0.0430	**2.1741**	-0.0551	1.7376	-0.2123	1.8346
$T = 500$								
$\beta = 0.1$	-0.0003	0.0010	0.0007	0.0010	0.0001	0.0010	0.0079	**0.0010**
$\beta = 0.3$	0.1956	1.3506	0.0411	1.5165	0.1721	**1.3066**	0.0711	1.2910
$\beta = 0.2$	-0.0966	2.0314	-0.0494	2.1212	-0.1060	1.9078	-0.3006	**2.1318**
$\beta = 0.1$	-0.1216	1.9385	-0.0555	**2.0342**	-0.1213	1.8245	-0.2604	1.9985

All results are based on $R = 100$ replications.

ARCH models when LE is used for estimation. We also found that schemes such as scheme U and scheme E are good alternative to scheme M. Finally, using LE instead of the QMLE for fitting ARCH models enables us to obtain these results in very quick time.

6.5 Bootstrap Prediction Intervals for ARCH Models

Predicting the distribution of the future returns has become an increasingly interesting area of research among financial practitioners and researchers. Accurate prediction of future volatilities are important for the implementation and evaluation of asset and derivative pricing (Pascual et al., 2006). Measuring the financial risk such as value-at-risk (VaR) is also very important and an accurate measure of this risk estimate is desired. Most of the surveys deal with predicting point forecast of returns, volatilities and VaR, see Baillie and Bollerslev (1992), Andersen et al. (2001) and Tsay (2005) among others for discussion on forecasting.

These studies focus on point forecasts and most importantly ignore parameter uncertainty. Beran (1990) first discussed the parametric bootstrap prediction intervals. Thombs and Schucany (1990) gave non-parametric bootstrap intervals for AR models. Miguel and Olave (1999) proposed a bootstrap method for prediction intervals of future observation in ARMA models with ARCH errors without considering parameters uncertainty. Pascual et al. (2004) extended bootstrap methods to ARIMA models. Reeves (2005) compared nonparametric and parametric bootstrap with Baillie and Bollerslev (BB) Gaussian asymptotic prediction interval in a Monte Carlo experiment. Christoffersen and Concalves (2005) used different VaR estimation methods and develop confidence intervals for VaR when QMLE is used for the estimation.

In this study we use bootstrap to obtain prediction intervals for returns, volatilities and VaR. The bootstrap prediction intervals are obtained using both the LE and the QMLE. These prediction intervals along with point estimate will help practitioners to evaluate the forecasting performance of their models. We investigate the difference in bootstrap prediction intervals of both estimators. We are also interested in assessing the loss of accuracy from estimation error when estimating risk estimates and quantifying this error by confidence intervals around the VaR. It is important to mention again that LE can be estimated in quick time and thus developing bootstrapped confidence intervals using LE requires very small processing time as compared to the QMLE.

6.5.1 Bootstrap Prediction Intervals

Consider an ARCH(p) model where one observes $\{X_t; 1-p \leq t \leq T\}$ satisfying

$$X_t = h_t^{1/2}(\boldsymbol{\beta})\epsilon_t \quad \text{and} \quad h_t(\boldsymbol{\beta}) = \beta_0 + \beta_1 X_{t-1}^2 + \cdots + \beta_1 X_{t-p}^2.$$

Our aim is to estimate the distribution of s-steps ahead returns X_{T+s}, volatilities h_{T+s}, and value-at-risk q_{T+s}, where

$$q_t = h_t^{1/2} F^{-1}(\gamma), \tag{6.12}$$

where F^{-1} is the quantile function of the errors $\{\epsilon_t\}$ and $h_t = h_t(\boldsymbol{\beta})$.

The bootstrap methods are described in the following steps.

Fitting the ARCH model

Fit an ARCH model to the given data set and estimate the parameters of the model $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)'$. We use both the QMLE and the LE for the estimation of the parameter vector. Let the estimated parameter vector be $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_p)'$ and the estimated residuals $\hat{\epsilon}_t$ are computed as

$$\hat{\epsilon}_t = X_t / \hat{h}_t^{1/2}, \tag{6.13}$$

where $\hat{h}_t = \hat{\beta}_0 + \hat{\beta}_1 X_{t-1}^2 + \cdots + \hat{\beta}_p X_{t-p}^2$.

Bootstrapping

Use the fitted model to generate bootstrap draws of the parameter. First we generate ϵ_t^*, random draws with replacement from $\hat{F}_T$, where $\hat{F}_T$ is the empirical distribution function of the centered residuals $(\hat{\epsilon}_t - \sum_{t=1}^T \hat{\epsilon}_t / T)$. Then the followings replicates are generated:

$$\begin{aligned} \hat{h}_t^* &= \hat{\beta}_0 + \hat{\beta}_1 X_{t-1}^{2*} + \cdots + \hat{\beta}_p X_{t-p}^{2*}, \\ X_t^* &= \hat{h}_t^{*1/2} \epsilon_t^*, \quad \text{for} \quad t = 1, 2, ..., T. \end{aligned} \tag{6.14}$$

The parameters of this generated series are estimated and the estimated parameters of this bootstrap series $\hat{\boldsymbol{\beta}}^* = (\hat{\beta}_0^*, \hat{\beta}_1^*, \cdots, \hat{\beta}_p^*)'$ are used to obtain future values.

Future realisations

Generate a future realisation of returns, volatilities and VaR. We want to estimate the distribution of future returns X_{T+s}, future volatilities h_{T+s} and future VaR q_{T+s} for $s > 0$, where s is the forecast step. In order to get these future realisations, we need $X_{T+1-i}^* = X_{T+1-i}$, $(1 \leq i \leq p)$, and ϵ_{T+s}^* (random draws with replacement from $\hat{F}_T$). Using the above, the future realisations of returns are generated recursively as:

$$\begin{aligned} \hat{h}_{T+s}^* &= \hat{\beta}_0^* + \hat{\beta}_1^* X_{T+s-1}^{2*} + \cdots + \hat{\beta}_p^* X_{T+s-p}^{2*}, \\ X_{T+s}^* &= \epsilon_{T+s}^* \hat{h}_{T+s}^{*1/2}, \quad \text{for} \quad s = 1, 2, ... \end{aligned} \tag{6.15}$$

Similarly the estimate of VaR at step $T + s$ can be obtained as

$$\hat{q}_{T+s}^* = \sqrt{\hat{h}_{T+s}^*} F^{-1*}(\gamma), \tag{6.16}$$

where F^{-1*} is the quantile function of estimated bootstrap centered residuals $\{\hat{\epsilon}_t^* - \overline{\hat{\epsilon}^*}\}_{t=1}^T$, where $\hat{\epsilon}_t^* = X_t / \sqrt{\hat{h}_t^*}$ and $\overline{\hat{\epsilon}^*} = \frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_t^*$. The centering of the bootstrap residuals ensures that the estimated bootstrap residuals have the zero mean property.

Prediction intervals

Once the set of B bootstrap future values, $(X_{T+s}^{*(1)}, ..., X_{T+s}^{*(B)})$ are obtained, the prediction intervals are defined as quantiles of the bootstrapped cumulative distribution function (cdf) of X_{T+s}^*. More specifically, we define the bootstrapped cdf of X^*_{T+s} by $G^*_X(l) = \Pr\{X^*_{T+s} \leq l\}$ and its Monte Carlo estimate by $G^*_X(l) = \#\{X^{*b}_{T+s} \leq l\}$, where $\#(\cdot)$ counts the number of cases where the condition within brackets is satisfied. Then for a given ϕ, a $100(1-\phi)\%$ prediction interval for X^*_{T+s} is given by

$$\Big[L^*_{X,B}(X), U^*_{X,B}(X)\Big] = \Big[Q^*_{X,B}\Big(\frac{\phi}{2}\Big), Q^*_{X,B}\Big(1-\frac{\phi}{2}\Big)\Big], \tag{6.17}$$

where $Q^*_{X,B} = G^{*-1}_{X,B}$.

Similarly we can define the bootstrap prediction intervals for volatilities and VaR. For future volatilities $(\hat{h}^{*(1)}_{T+s}, ..., \hat{h}^{*(B)}_{T+s})$, the prediction intervals are defined as quantiles of the bootstrap cdf of $\hat{h}^*_{T+s}$. The bootstrap cdf of $\hat{h}^*_{T+s}$ is given by $G^*_h(l) = \Pr\{\hat{h}^*_{T+s} \leq l\}$ and its Monte Carlo estimate by $G^*_h(l) = \#\{\hat{h}^{*b}_{T+s} \leq l\}$. Then, a $100(1-\phi)\%$ prediction interval for $\hat{h}^*_{T+s}$ is given by

$$\Big[L^*_{h,B}(h), U^*_{h,B}(h)\Big] = \Big[Q^*_{h,B}\Big(\frac{\phi}{2}\Big), Q^*_{h,B}\Big(1-\frac{\phi}{2}\Big)\Big], \tag{6.18}$$

where $Q^*_{h,B} = G^{*-1}_{h,B}$.

Finally, for VaR a $100(1-\phi)\%$ prediction interval for $\hat{q}^*_{T+s}$ is given by

$$\Big[L^*_{q,B}(q), U^*_{q,B}(q)\Big] = \Big[Q^*_{q,B}\Big(\frac{\phi}{2}\Big), Q^*_{q,B}\Big(1-\frac{\phi}{2}\Big)\Big], \tag{6.19}$$

where $Q^*_{q,B} = G^{*-1}_{q,B}$.

6.5.2 Simulation Results

Two different studies are conducted to develop bootstrap prediction intervals for returns, volatilities and VaR and to compare the results of the QMLE and LE.

The first model is an ARCH(2) model

$$\text{Model A:} \quad X_t = h_t^{1/2}\epsilon_t,$$
$$h_t = 0.1 + 0.4X_{t-1}^2 + 0.2X_{t-2}^2.$$

The second model considered is an ARCH(3) model

$$\text{Model B:} \quad X_t = h_t^{1/2}\epsilon_t,$$
$$h_t = 0.01 + 0.2X_{t-1}^2 + 0.2X_{t-2}^2 + 0.1X_{t-3}^2.$$

For both studies errors are generated from the standard normal and student-t distribution with 3 degrees of freedom. The sample sizes considered are $T = 50$ and 500. ARCH models are simulated from Model A and Model B and the parameters are estimated using the QMLE and LE. For both models, under particular sample size and error distribution, $R = 1000$ future values of X_{T+s}, h_{T+s}, and $q_{T+s}(\gamma)$ are generated with true parameter values, where the forecast step $s = 1$, 10 and 20. Using the bootstrap method with B=999, a $100(1-\phi)\%$ prediction intervals for returns denoted by (L_X^*, U_X^*), volatilities denoted by (L_h^*, U_h^*), and VaR denoted by (L_q^*, U_q^*) are obtained.

The conditional coverage and length for returns are computed as $\widehat{1-\phi_X^*} = \#\{L_X^* \leq X_{T+s}^T \leq U_X^*\}/R$. Choices of nominal coverage considered are 80%, 95% and 99% although only result for 99% prediction intervals are considered as this interval could be of interest in risk management. The length is defined as $LEN_X = U_X^* - L_X^*$. Similarly the conditional coverage and length for volatilities and VaR are obtained. The coverage of the left and right tail of the distribution of returns, volatilities and VaR are also obtained. The average and the standard deviation for coverage and length and the average proportion of observation lying out of the left and right quantiles are computed based on $K = 100$ Monte Carlo replicates. For empirical lengths $R = 10,000$ independent replication each of size 500 are generated. The root mean squared error (RMSE) for both estimators at each step length is calculated, where RMSE for returns is defined as

$$\text{RMSE}_X = \sqrt{\frac{1}{R}\sum_{i=1}^{R}(\hat{\phi}^*_{i,X} - \phi)^2}.$$

Table 6.9 reports the mean coverage and the corresponding standard errors together with the mean length with its corresponding standard errors and the mean coverage on the left and right tails and the RMSE when ARCH(2) models are generated with standard normal and student-t distribution with 3 df for predicting intervals for returns for $s = 1$, 10 and 20 steps ahead. It can be seen that the mean coverage and their corresponding standard errors for both estimators are close to each other and provide good match to the empirical coverage with the QMLE having slightly high probability. By examining the results of mean length we found that the lengths for LE are close to empirical length and their standard errors are below than those of QMLE. These findings become more prominent in case of student-t distribution. The mean lengths of QMLE for all step lengths are found greater than both the empirical lengths and LE. This shows that prediction intervals of QMLE are on average larger than the mean length of LE and this may be one of the reasons of high coverage probabilities of QMLE. The mean coverages on the left and right tails of both estimators show similar results. The root mean squared errors of QMLE are found slightly smaller than the LE.

Next, we analyse the performance of both LE and QMLE prediction intervals for future volatilities. Using same DGP as in the previous case, we develop 99% bootstrapped prediction intervals for $s = 1$, 10 and 20 steps ahead volatilities. The results when errors are generated form Gaussian and student-t distribution with

Table 6.9: Prediction intervals for returns of ARCH(2) model with nominal coverage of 99%.

$T = 500$ $B = 999$	Mean coverage	S.D of coverage	Mean length	S.D of length	Mean coverage below/above	RMSE
			Standard Normal Distribution			
1-step ahead						
Empirical	0.9900		2.8933		0.50%/0.50%	
QMLE	0.9869	0.0071	3.0967	0.4891	0.57%/0.74%	0.0107
LE	0.9852	0.0074	2.9622	0.4284	0.63%/0.86%	0.0123
10-steps ahead						
Empirical	0.9900		3.0550		0.50%/0.50%	
QMLE	0.9828	0.0096	3.0816	0.4992	0.98%/0.75%	0.0155
LE	0.9816	0.0080	2.9936	0.4833	1.02%/0.82%	0.0156
20-steps ahead						
Empirical	0.9900		3.0416		0.50%/0.50%	
QMLE	0.9919	0.0051	3.0938	0.4647	0.28%/0.54%	0.0059
LE	0.9914	0.0047	3.0161	0.4280	0.28%/0.58%	0.0060
			Student-t Distribution (3)			
1-step ahead						
Empirical	0.9900		2.7157		0.50%/0.50%	
QMLE	0.9907	0.0066	3.3364	0.9949	0.57%/0.36%	0.0078
LE	0.9889	0.0057	2.9192	0.5496	0.66%/0.46%	0.0084
10-steps ahead						
Empirical	0.9900		2.8917		0.50%/0.50%	
QMLE	0.9909	0.0053	3.3347	0.8500	0.39%/0.52%	0.0067
LE	0.9874	0.0067	2.9554	0.6913	0.64%/0.62%	0.0071
20-steps ahead						
Empirical	0.9900		3.1312		0.50%/0.50%	
QMLE	0.9861	0.0074	3.3538	0.9669	0.60%/0.79%	0.0116
LE	0.9836	0.0065	3.0282	0.6815	0.77%/0.87%	0.0121

3 df are tabulated in Table 6.10. The mean coverage for QMLE are found greater than LE with low standard errors. The mean lengths of LE are close to empirical lengths where as that of QMLE are larger in size with large standard errors. Again this feature can be seen in the case of $t(3)$. The results of the average coverage on the left and right tails reveal that the shape of the volatility is asymmetric which is often observes in real data sets. The RMSEs of QMLE are found smaller than the LE and this can be due to the wider lengths of QMLE.

Finally, we develop prediction intervals for 1% VaR of ARCH(2) model with nominal coverage of 99%. Results of mean coverage along with their corresponding standard errors, mean length and their standard errors, mean coverage on the left and right tails and the root mean squared errors are shown in Table 6.11. The mean

Table 6.10: Prediction intervals for volatilities of ARCH(2) model with nominal coverage of 99%.

$T = 500$ $B = 999$	Mean coverage	S.D of coverage	Mean length	S.D of length	Mean coverage below/above	RMSE
			Standard Normal Distribution			
1-step ahead						
Empirical	0.9900		1.7390		0.50%/0.50%	
QMLE	0.9857	0.0211	2.0875	1.2881	0.48%/0.94%	0.0229
LE	0.9730	0.0448	1.6939	0.8186	1.45%/1.24%	0.0497
10-steps ahead						
Empirical	0.9900		1.7957		0.50%/0.50%	
QMLE	0.9846	0.0276	2.0545	1.0999	0.65%/0.89%	0.0294
LE	0.9735	0.0433	1.8277	1.2007	1.57%/1.09%	0.0482
20-steps ahead						
Empirical	0.9900		1.9711		0.50%/0.50%	
QMLE	0.9891	0.0268	2.2128	1.1523	0.74%/0.35	0.0273
LE	0.9777	0.0441	1.8338	1.1777	1.66%/0.57%	0.0472
			Student-t Distribution (3)			
1-step ahead						
Empirical	0.9900		1.7759		0.50%/0.50%	
QMLE	0.9887	0.0150	3.4645	4.2050	0.00%/1.13%	0.0162
LE	0.9748	0.0373	1.3863	0.8745	0.51%/2.01%	0.0422
10-steps ahead						
Empirical	0.9900		1.8344		0.50%/0.50%	
QMLE	0.9947	0.0108	3.5826	4.1048	0.00%/0.53%	0.0108
LE	0.9835	0.0300	1.4372	1.1011	0.37%/1.28%	0.0320
20-steps ahead						
Empirical	0.9900		2.3055		0.50%/0.50%	
QMLE	0.9894	0.0127	3.5116	4.1143	0.00%/1.06%	0.0138
LE	0.9743	0.0410	1.4704	1.4169	0.55%/2.01%	0.0457

coverage of both estimators, under Gaussian and student-t distribution with 3 df are found similar. The significant difference between both estimators can be found by comparing their mean lengths with the empirical length. LE seems to provide a reasonable approximation with low standard errors. Mean lengths of QMLE are found slightly greater than the empirical length for Gaussian distribution but for heavy-tailed distribution the bootstrap prediction intervals for 1% VaR of QMLE are too wider with large standard errors. The asymmetric feature of prediction intervals of VaR is also captured well by both estimators.

We conclude this section by highlighting our contributions and findings. We defined bootstrap prediction intervals for returns, volatilities and value-at-risk for ARCH models. We showed that our method is easy to apply especially if LE is

Table 6.11: Prediction intervals of 1% VaR of ARCH(2) model with nominal coverage of 99%.

$T = 500$ $B = 999$	Mean coverage	S.D of coverage	Mean length	S.D of length	Mean coverage below/above	RMSE
			Standard Normal Distribution			
1-step ahead						
Empirical	0.9900		2.6774		0.50%/0.50%	
QMLE	0.9906	0.0100	3.0416	1.0363	0.88%/0.06%	0.0109
LE	0.9881	0.0120	2.7215	0.8034	1.12%/0.06%	0.0138
10-steps ahead						
Empirical	0.9900		2.7307		0.50%/0.50%	
QMLE	0.9909	0.0112	3.0210	0.9539	0.88%/0.03%	0.0119
LE	0.9897	0.0097	2.8344	0.9777	0.98%/0.05%	0.0110
20-steps ahead						
Empirical	0.9900		2.9759		0.50%/0.50%	
QMLE	0.9961	0.0088	3.1582	1.0039	0.31%/0.08%	0.0089
LE	0.9943	0.0100	2.7689	0.9267	0.52%/0.04%	0.0100
			Student-t Distribution (3)			
1-step ahead						
Empirical	0.9900		4.1052		0.50%/0.50%	
QMLE	0.9910	0.0145	6.0102	4.4658	0.90%/0.00%	0.0150
LE	0.9843	0.0168	3.7793	1.7953	1.57%/0.00%	0.0199
10-steps ahead						
Empirical	0.9900		4.4615		0.50%/0.50%	
QMLE	0.9958	0.0113	6.1197	4.4186	0.42%/0.00%	0.0113
LE	0.9920	0.0131	3.8266	1.7752	0.80%/0.00%	0.0134
20-steps ahead						
Empirical	0.9900		4.3195		0.50%/0.50%	
QMLE	0.9912	0.0123	6.0465	4.1538	0.88%/0.00%	0.0128
LE	0.9845	0.0137	3.8093	1.8256	1.55%/0.00%	0.0172

used for the estimation of ARCH models. Results of our simulations indicated that the proposed bootstrap method is appropriate for predicting interval forecasts. We found that LE provides better prediction intervals than the QMLE in most of the cases.

6.6 Conclusion

This chapter explores the linear estimator for the parameters of ARCH models. An advantage of the LE over the widely-used QMLE is that its computation is very easy and requires less CPU time which enables one to do computer intensive tasks on ARCH model in quick time.

We showed in this study that this estimator provides very good results for the estimation of the parameters of ARCH models. This estimator also predicts volatility better than the QMLE in almost all our Monte Carlo simulations. These findings were further supported by application to three stock indices.

A weighted resampling for the linear estimator is used to approximate the distribution of the parameters of ARCH models. Results of our experiments showed that there are other schemes such as Scheme E and Scheme U that can match the paired bootstrap and residual bootstrap and even perform better than these commonly used methods in some cases.

Prediction intervals for returns, volatilities and value-at-risk are developed using a simple bootstrap method. Monte Carlo results showed that although both estimators provide good mean coverage, the LE can be considered favourable in terms of its mean lengths close to the empirical with low standard errors.

Chapter 7

A Weighted Linear Estimator for the Multivariate ARCH Parameters

7.1 Introduction

The autoregressive conditional heteroscedasticity (ARCH) model of Engle (1982) and the generalised ARCH (GARCH) model of Bollerslev (1986) have been widely used for capturing the time-varying variances in financial time series. The successful application in the univariate case has motivated many researchers to extend these models to multivariate volatility models. Bollerslev et al. (1988) extended the GARCH representation to the vectorized conditional-variance matrix. The number of parameters increases with the dimension of data and estimation becomes difficult. Cecchetti et al. (1988) introduced a bivariate ARCH model with constant conditional correlation. Bollerslev (1990) proposed the constant conditional correlation GARCH (CCC-GARCH) model. In CCC-GARCH, under the assumption of constant correlations, the maximum likelihood estimate of the correlation matrix is equal to the sample correlation matrix and further simplification is achieved in the optimisation when the correlation matrix is concentrated out of the log-likelihood function. The CCC-GARCH model has become very popular among the applied researchers due to its computational simplicity and many empirical studies exist in literature (see Bollerslev (1990), Kroner and Claessens (1991), Park and Switzer (1995) and Lien and Tse (1998), among others).

The quasi-maximum likelihood (QML) estimation is generally used for the estimation of the parameters of the multivariate ARCH model. Assuming the errors process to be Gaussian, the log-likelihood function is maximised. The QMLE does not admit a closed form expression and needs numerical optimisation methods to compute the solution. One major problem with estimating these multivariate volatility models is that the number of parameters in the variance covariance matrix increases rapidly with the dimension of the system, the likelihood function becomes very flat and consequently the optimisation of the likelihood function becomes infeasible.

Bose and Mukerjee (2003) proposed the linear estimators (LE) of the ARCH parameters. The computation of the LE involves solving only two sets of linear equations and it has the same asymptotic normal distribution as that of QMLE. Using the idea of the linear estimators, Mousazadeh and Karimi (2009) extended the LE to multivariate ARCH (MARCH) model with constant correlation. They also discuss the asymptotic properties of the two-stage least-squares (TSLS) estimator of the parameters of MARCH model.

To derive the limiting distribution of the LE, the crucial assumption is that all ARCH parameters must be strictly positive. This assumption restricts the application of the result. In order to tackle these problems, Bose and Mukherjee (2009) proposed a weighted linear estimator (WLE) of the ARCH parameter and derive its limiting distribution. The limit distribution turns out to be multivariate normal even when some of the parameters are zero.

In this chapter, we propose a weighted linear estimator for the parameters of multivariate ARCH models. The estimator involves solving set of linear equations and hence very easy to obtain. Using Monte Carlo simulations, we evaluate and compare the performance of the WLE with QMLE under various error distributions. We find that the WLE produces as accurate results as the QMLE for large sample size and outperforms QMLE in small samples and under heavy-tailed errors both in terms of the mean squared error and the mean absolute error. The forecasting performance of the WLE is found not inferior to the QMLE and one-day risk estimates are found better. This estimator is also applied to real data sets and forecasts of volatility and value-at-risk are obtained.

As a second contribution we consider a weighted bootstrapped version of the weighted linear estimator (BWLE) for the parameters of multivariate ARCH models. This estimator is proposed using the idea of weighted resampling for estimating equations. The proposed bootstrap estimator resulting as solution of certain linear equations is simple and easy to calculate and covers several different resampling approaches including classical bootstrap, bootstrap clone methods and so on. Simulation results show that the BWLE provides better approximations than the normal. We proposed various bootstrap schemes and found these bootstrap schemes outperform the widely-used paired bootstrap method.

An advantage of using the WLE for multivariate ARCH parameters over the commonly used QMLE is that the former is very easy to compute and requires less CPU time. In high dimensions where the QMLE is very difficult to compute, the WLE can be easily applied to estimates the parameters in quick time. Bootstrapping is a computationally intensive task even in the univariate case. The quick computation of the WLE enables one to use the idea of bootstrapping in multivariate ARCH models with ease.

The plan for the rest of the chapter is as follows. In Section 7.2 we propose a weighted linear estimator of the multivariate ARCH parameters. Using simulations we compare the accuracy of this estimator with the QMLE under different error distributions and varying sample sizes. The WLE is also applied to real data sets where multivariate ARCH model is fitted to the daily log returns of SP500 index and the stocks of Cisco Systems and Intel Corporation. One-day-ahead volatility forecasts and value-at-risk estimates are obtained using both the WLE and the QMLE. Evaluation measure are used to assess the forecasting performance of these estimators. The weighted resampling for the WLE in multivariate ARCH model is introduced in Section 7.3. The distribution of the parameters of multivariate ARCH model is approximated empirically using three different bootstrapped schemes. Section 7.3 concludes the chapter.

7.2 A Weighted Linear Estimator of the Multivariate ARCH Model

A sequence of $\{\mathbf{X}_t, t \in \mathbb{Z}\}$ of random variables with values in $\mathbb{R}^N$ follows a multivariate ARCH(p) process with constant correlation if

$$\mathbf{X}_t = \mathbf{\Sigma}_t \boldsymbol{\epsilon}_t, \tag{7.1}$$

where $\mathbf{X}_t = (X_{1,t}, X_{2,t}, \cdots, X_{N,t})'$, $\boldsymbol{\epsilon}_t = (\epsilon_{1,t}, \epsilon_{2,t}, \cdots, \epsilon_{N,t})'$ and $\mathbf{\Sigma}_t$ is a diagonal matrix of conditional standard deviations of $\mathbf{X}_t$ defined as

$$\mathbf{\Sigma}_t = diag(\sigma_{1,t}, \sigma_{2,t}, \cdots, \sigma_{N,t}), \tag{7.2}$$

$$\boldsymbol{\sigma}_t^2 = (\sigma_{1,t}^2, \sigma_{2,t}^2, \cdots, \sigma_{N,t}^2)' = \mathbf{A}_0 + \sum_{i=1}^{p} \mathbf{A}_i(\mathbf{X}_{t-i} \odot \mathbf{X}_{t-i}), \tag{7.3}$$

$\mathbf{A}_0$ is a $(N \times 1)$ vector with positive elements, $\mathbf{A}_i$ are $(N \times N)$ diagonal matrices with non-negative elements and

$$\mathbf{B} = [\mathbf{A}_0 \vdots \mathbf{A}_1 \vdots \mathbf{A}_2 \vdots \cdots \vdots \mathbf{A}_p] \tag{7.4}$$

is the matrix of parameters. Also note that N and p are the dimension and the known order of the ARCH model, respectively and $\odot$ is the Hadamard product of two matrices. The hadamard product of two matrices $\mathbf{U} = [u_{ij}]_{i,j=1,\ldots,N}$ and $\mathbf{V} = [v_{ij}]_{i,j=1,\ldots,N}$ is defined as the element wise product

$$\mathbf{U} \odot \mathbf{V} = [u_{ij}v_{ij}]_{i,j=1,\ldots,N}.$$

Furthermore, we have the following conditions on $\boldsymbol{\epsilon}_t$:

1. $\{\boldsymbol{\epsilon}_t, t \in \mathbb{Z}\}$ is a sequence of i.i.d. $\mathbb{R}^N$-valued random variables with mean $\mathbf{0}$ and positive definite covariance matrix $\boldsymbol{\Gamma}$ such that

$$\boldsymbol{\Gamma} = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1N} \\ \rho_{12} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_{(N-1)N} \\ \rho_{1N} & \cdots & \rho_{(N-1)N} & 1 \end{pmatrix},$$

2. Also, $\boldsymbol{\epsilon}_t$ is independent of $\mathcal{F}_{t-1}$ (σ-field generated by $\{\mathbf{X}_{t-k}, k \geq 1\}$).

With these assumptions, $\mathrm{E}[\mathbf{X}_t|\mathcal{F}_{t-1}] = \mathbf{0}$, and

$$\mathrm{E}[\mathbf{X}_t\mathbf{X}_t|\mathcal{F}_{t-1}] = \mathbf{H}_t = \begin{cases} \sigma_{i,t}^2 & i = j \\ \rho_{ij}\sigma_{i,t}\sigma_{j,t} & i \neq j \end{cases}$$

where $\mathbf{H}_t = \boldsymbol{\Sigma}_t\boldsymbol{\Gamma}\boldsymbol{\Sigma}_t$ is the conditional covariance matrix of $\{\mathbf{X}_t\}$.

Let $\mathbf{Y}_t = (Y_{1,t}, Y_{2,t} \cdots Y_{N,t})' = \mathbf{X}_t \odot \mathbf{X}_t$, for $1 - p \leq t \leq T$, is a vector of length N. For $1 \leq t \leq T$, let $\mathbf{Z}_t = (1, \mathbf{Y}'_{t-1}, \cdots, \mathbf{Y}'_{t-p})'$, is a $((1 + pN) \times 1)$ vector and $\boldsymbol{\eta}_t = (\eta_{1,t}, \eta_{2,t}, \cdots, \eta_{N,t})'$ is a $(N \times 1)$ vector, where $\eta_{j,t} = \epsilon^2_{j,t} - 1$, for $1 \leq j \leq N$. Then

$$\boldsymbol{\sigma}^2_t = (\sigma^2_{1,t}, \sigma^2_{2,t}, \cdots, \sigma^2_{N,t})' = \mathbf{A}_0 + \sum_{i=1}^{p} \mathbf{A}_i \mathbf{Y}_{t-i} = \mathbf{B}\mathbf{Z}_t. \tag{7.5}$$

Now squaring both sides of (7.1) and using the form of (7.5), we get

$$\mathbf{Y}_t = \mathbf{B}\mathbf{Z}_t + \boldsymbol{\Sigma}^2_t \boldsymbol{\eta}_t, \quad 1 \leq t \leq T, \tag{7.6}$$

where $\mathrm{E}\{\boldsymbol{\Sigma}^2_t \boldsymbol{\eta}_t\} = \mathrm{E}\{\boldsymbol{\Sigma}^2_t\}\,\mathrm{E}\{\boldsymbol{\eta}_t\} = \mathbf{0}, \quad 1 \leq t \leq T.$

Let $\{(u_{i,t}, v_{i,t}); 1 \leq t \leq T, 1 \leq i \leq N\}$ be a sequence of non-negative random variables called weights and let $\mathbf{U}_t$ and $\mathbf{V}_t$ be $(N \times N)$ diagonal matrices with elements $u_{i,t}$ and $v_{i,t}$, respectively, in main diagonals. Now, in (7.6), ignoring the randomness of $\boldsymbol{\Sigma}^2_t$ and the presence of parameter $\mathbf{B}$ in it, we obtain a preliminary weighted least squares estimator $\hat{\mathbf{B}}_{pr}$ as the solution of

$$\sum_{t=1}^{T} \left[\mathbf{U}_t \left\{\mathbf{B}\mathbf{Z}_t - \mathbf{Y}_t\right\} \mathbf{Z}'_t\right] = \mathbf{0}. \tag{7.7}$$

Thus, guided by (7.7), we can define our preliminary estimator

$$\hat{\mathbf{b}}_{i,pr} = \left(\sum_{t=1}^{T} u_{i,t} \left\{Y_{i,t} \mathbf{Z}'_t\right\}\right) \left(\sum_{t=1}^{T} u_{i,t} \left\{\mathbf{Z}_t \mathbf{Z}'_t\right\}\right)^{-1}, \quad i = 1, 2, ..., N, \tag{7.8}$$

where $\hat{\mathbf{b}}_{i,pr}$ is the i-th row of matrix $\mathbf{B}$ and hence we get $\hat{\mathbf{B}}_{pr}$, the preliminary estimate of $\mathbf{B}$. Using this preliminary estimate we can get an estimate of $\boldsymbol{\Sigma}^2_t$ as

$$\hat{\boldsymbol{\Sigma}}^2_t = diag(\hat{\sigma}^2_{1,t}, \hat{\sigma}^2_{2,t}, \cdots, \hat{\sigma}^2_{N,t}),$$

where

$$(\hat{\sigma}^2_{1,t}, \hat{\sigma}^2_{2,t}, \cdots, \hat{\sigma}^2_{N,t})' = \hat{\mathbf{B}}_{pr}\mathbf{Z}_t.$$

Now, from (7.6), we get

$$\hat{\boldsymbol{\Sigma}}_t^{-2}\mathbf{Y}_t \simeq \hat{\boldsymbol{\Sigma}}_t^{-2}\mathbf{B}\mathbf{Z}_t + \boldsymbol{\eta}_t, \quad 1 \le t \le T. \tag{7.9}$$

Using (7.9), we can find another estimate of $\mathbf{B}$, by solving the following set of equations

$$\sum_{t=1}^{T}\left[\mathbf{V}_t\Big\{(\hat{\boldsymbol{\Sigma}}_t^{-2})'\hat{\boldsymbol{\Sigma}}_t^{-2}\Big\}\Big\{\mathbf{B}\mathbf{Z}_t - \mathbf{Y}_t\Big\}\mathbf{Z}_t'\right] = \mathbf{0}, \tag{7.10}$$

yielding the estimator

$$\hat{\mathbf{b}}_{i,T} = \left(\sum_{t=1}^{T} v_{i,t}\Big\{Y_{i,t}\mathbf{Z}_t'/(\hat{\sigma}^2_{i,t})^2\Big\}\right)\left(\sum_{t=1}^{T} v_{i,t}\Big\{\mathbf{Z}_t\mathbf{Z}_t'/(\hat{\sigma}^2_{i,t})^2\Big\}\right)^{-1}, \; i = 1, 2, ..., N, \tag{7.11}$$

Hence, we get $\hat{\mathbf{B}}_T$, the final estimator of $\mathbf{B}$. Note that the preliminary and the final estimator of $\mathbf{B}$ are not guaranteed to have all positive elements when sample size is small. We found only 4 negative estimates when 1000 independent samples of size 2000 are estimated by WLE. This problem can be solved by using a constrained two-stage least squares method.

7.2.1 Asymptotics of the WLE

Besides ergodicity, we assume that the errors satisfy

(C1) $\mathrm{E}(\epsilon_i^4) < \infty$, $i = 1, ..., N$. These assumptions will be referred as model assumptions.

Assume the following conditions on weights. Consider the increasing sequence of sigma-field $\{\mathcal{F}_t = \sigma < \mathbf{Z}_1, \cdots, \mathbf{Z}_t >; t \geq 1\}$, and that

(C2) the weights $\{(u_{i,t}, v_{i,t}; t, i \geq 1)\}$, are stationary and for each, $t \geq 1$, $(u_{i,t}, v_{i,t})$ is $\mathcal{F}_t$-measurable.

Also assume that for the asymptotic normality of the preliminary estimator

$$\begin{aligned} &\mathrm{E}(u_{i,1}^2) < \infty,\ \mathrm{E}\{u_{i,1}Y_{i,-j}Y_{i,-k}\} < \infty, \\ \text{and}\quad &\mathrm{E}\{u_{i,1}^2 Y_{i,-j}Y_{i,-k}Y_{i,-l}Y_{i,-m}\} < \infty,\ \forall 1 \leq j,k,l,m \leq p, i = 1,\cdots,N. \end{aligned} \tag{7.12}$$

Condition (7.12) ensure that $\mathrm{E}\{u_{i,1}(\mathbf{Z}_1\mathbf{Z}_1')\}$ and $\mathrm{E}\{u_{i,1}^2(\mathbf{b}_i\mathbf{Z}_1)^2(\mathbf{Z}_1\mathbf{Z}_1')\}$, for $i = 1, ..., N$, are all finite.

Then under model assumptions and (7.12)

$$\begin{aligned} T^{1/2}(\hat{\mathbf{b}}_{i,pr} - \mathbf{b}_i) \xrightarrow{\mathfrak{D}} N[\mathbf{0}, \mathrm{var}(\epsilon_{i,1}^2)\{\mathrm{E}(u_{i,1}(\mathbf{Z}_1\mathbf{Z}_1'))\}^{-1} \times \\ \mathrm{E}(u_{i,1}^2(\mathbf{b}_i\mathbf{Z}_1)^2(\mathbf{Z}_1\mathbf{Z}_1'))\{\mathrm{E}(u_{i,1}(\mathbf{Z}_1\mathbf{Z}_1'))\}^{-1}]. \end{aligned} \tag{7.13}$$

We also assume the following conditions

$$\begin{aligned} &\mathrm{E}(v_{i,1}^2) < \infty,\ \mathrm{E}\{v_{i,1}Y_{i,-j}Y_{i,-k}\} < \infty,\ \mathrm{E}\{v_{i,1}^2 Y_{i,-j}Y_{i,-k}\} < \infty, \\ \text{and}\quad &\mathrm{E}\{v_{i,1}Y_{i,-j}Y_{i,-k}Y_{i,-l}\} < \infty,\ \forall j,k,l \in \mathcal{P}_0, i = 1,\cdots,N, \end{aligned} \tag{7.14}$$

where $\mathcal{P}_0 = \{j; b_{i,j} = 0, 0 < j < p, i = 1, \cdots, N\}$.

Condition (7.14) ensures that

$$\mathrm{E}\{v_{i,1}(\mathbf{Z}_1\mathbf{Z}_1')(\mathbf{b}_i\mathbf{Z}_1)^{-2}\} < \infty \quad \text{and} \quad \mathrm{E}\{v_{i,1}^2(\mathbf{Z}_1\mathbf{Z}_1')(\mathbf{b}_i\mathbf{Z}_1)^{-2}\} < \infty.$$

The proof of (7.13) for univariate case is shown by Bose and Mukherjee (2009) and for LE in multivariate case is shown by Mousazadeh and Karimi (2009). Therefore, from those we conjecture that using the property

$$T^{1/2}(\hat{\mathbf{b}}_{i,pr} - \mathbf{b}_i) = O_p(1), \tag{7.15}$$

of $\hat{\mathbf{b}}_{i,pr}$ where O_p denotes the boundedness in probability, and suppose that model assumptions, (7.14) and (7.15) hold, then

$$\begin{aligned} T^{1/2}(\hat{\mathbf{b}}_{i,T} - \mathbf{b}_i) \xrightarrow{\mathfrak{D}} \mathrm{N}\Big[\mathbf{0}, \mathrm{var}(\epsilon_{i,1}^2)\{\mathrm{E}(v_{i,1}(\mathbf{Z}_1\mathbf{Z}_1')(\mathbf{b}_i\mathbf{Z}_1)^{-2})\}^{-1} \times \\ \mathrm{E}(v_{i,1}^2(\mathbf{Z}_1\mathbf{Z}_1')(\mathbf{b}_i\mathbf{Z}_1)^{-2})\{\mathrm{E}(v_{i,1}(\mathbf{Z}_1\mathbf{Z}_1')(\mathbf{b}_i\mathbf{Z}_1)^{-2})\}^{-1}\Big]. \end{aligned} \tag{7.16}$$

7.2.2 Simulation and Empirical Results

In this section we report the results of Monte Carlo simulations performed to evaluate and compare the performance of WLE with the QMLE in terms of parameters estimation and volatility and value-at-risk forecasting. Application to real data sets is also presented.

Monte Carlo Simulations

We use our own MATLAB code for the QMLE since no multivariate volatility modelling routines are available in MATLAB software for the QMLE. We also write the complete program for estimating multivariate ARCH models using WLE. In this section we investigate the performance of WLE with the QMLE based on Mean Squared Error (MSE) in estimating MARCH(p) models of dimension N through extensive simulations. We assume various distributions for errors. Each experiment is repeated K times and estimates of the MSE and the MAE are obtained as

$$\begin{aligned} \text{MSE} &= \frac{1}{K}\sum_{k=1}^{K}\Big\{\sum_{i=1}^{N}\sum_{j=0}^{p}(\hat{b}_{ij}-b_{ij})^2\Big\}, \\ \text{MAE} &= \frac{1}{K}\sum_{k=1}^{K}\Big\{\sum_{i=1}^{N}\sum_{j=0}^{p}|\hat{b}_{ij}-b_{ij}|\Big\}, \\ \text{MAE}_\rho &= \frac{1}{K}\sum_{k=1}^{K}\Big\{\sum_{i=1}^{N-1}\sum_{i<j}^{N}|\hat{\rho}_{ij}-\rho_{ij}|\Big\}, \end{aligned}$$

where $\hat{b}_{ij}$ and b_{ij} for $1 \le i \le N$, $0 \le j \le p$ are the elements of the estimated and the true parameters matrices, $\hat{\mathbf{B}}_T$ and $\mathbf{B}$, respectively, of the N-dimensional multivariate ARCH(p) model and $\hat{\rho}_{ij}$ is an estimate of constant conditional correlation ρ_{ij}.

We use Monte Carlo experiments to estimate these quantities for weighted linear estimator and the QMLE. All results are based on K replicates each of sample size T. The errors are generated from the standard normal distribution and student-t distribution with 3 degrees of freedom. Student-t distributions are standardized to have mean 0 and variance 1. Initial 500 simulated values in each replication are discarded to reduce the impact of initial values.

In first experiment, 1000 independent replicates each of sample size $T = 100$, 250 and 1000 are generated from the following bivariate ARCH(2) model.

$$\text{Model A:} \quad \begin{pmatrix} \sigma_{1,t}^2 \\ \sigma_{2,t}^2 \end{pmatrix} = \begin{pmatrix} 0.10 + 0.05X_{1,t-1}^2 + 0.10X_{1,t-2}^2 \\ 0.20 + 0.07X_{2,t-1}^2 + 0.20X_{2,t-2}^2 \end{pmatrix},$$

where the value of ρ_{12}, the constant conditional correlation coefficient, for this experiment is set to 0.95. The results of MSE and MAE for both estimators along with their standard errors in parentheses are shown in Table 7.1.

The results in Table 7.1 are found similar to those in the univariate case where we compare LE with QMLE in ARCH models. For standard normal distribution the MSE and MAE of WLE are found better than QMLE for sample sizes $T = 100$ and 250. This shows that WLE in multivariate case provide better estimates than the QMLE for small sizes. For large sample $T = 1000$ in our case, both estimators shows approximately same results in terms of MSE and MAE with QMLE slightly out performing WLE.

When errors are generated from standardised student-t distribution with 3 df, we can see from Table (7.1) that WLE may be considered as a better choice than the QMLE even for large sample size. The MAEs for correlation for both estimators are found similar in all cases.

Table 7.1: Mean Squared Error and Mean Absolute Error of the parameters and constant conditional correlation of a two dimensional ARCH(2) model.

	QMLE	WLE	QMLE	WLE	QMLE	WLE
	$T = 100$		$T = 250$		$T = 1000$	
Standard Normal						
MSE	0.0638	**0.0543**	0.0234	**0.0228**	**0.0065**	0.0067
	(0.0781)	(0.0634)	(0.0238)	(0.0222)	(0.0060)	(0.0062)
MAE	0.4234	**0.3921**	0.2618	**0.2594**	**0.1402**	0.1431
	(0.2212)	(0.1998)	(0.1223)	(0.1196)	(0.0659)	(0.0661)
MAE_ρ	0.0084	**0.0082**	0.0050	**0.0050**	**0.0025**	0.0025
	(0.0069)	(0.0069)	(0.0039)	(0.0039)	(0.0019)	(0.0019)
Student-t(3)						
MSE	0.2058	**0.1113**	0.0657	**0.0490**	0.0489	**0.0381**
	(0.2685)	(0.1250)	(0.1279)	(0.0760)	(0.1084)	(0.0670)
MAE	0.6930	**0.5532**	0.3781	**0.3560**	0.3195	**0.3094**
	(0.3945)	(0.2689)	(0.2449)	(0.2009)	(0.2159)	(0.1809)
MAE_ρ	0.0087	**0.0084**	**0.0038**	0.0040	**0.0026**	0.0027
	(0.0074)	0.0079)	(0.0031)	(0.0037)	(0.0021)	(0.0030)

MAE is the mean absolute error for constant conditional correlation. Standard errors in parenthesis.

In second experiment, 1000 replicates each of sample size $T = 100$, 250 and 1000 are generated from the following DGP:

$$\text{Model B:} \quad \begin{pmatrix} \sigma_{1,t}^2 \\ \sigma_{2,t}^2 \\ \sigma_{3,t}^2 \end{pmatrix} = \begin{pmatrix} 0.10 + 0.10X_{1,t-1}^2 + 0.20X_{1,t-2}^2 + 0.20X_{1,t-3}^2 \\ 0.25 + 0.25X_{2,t-1}^2 + 0.10X_{2,t-2}^2 + 0.05X_{2,t-3}^2 \\ 0.05 + 0.20X_{3,t-1}^2 + 0.15X_{3,t-2}^2 + 0.10X_{3,t-3}^2 \end{pmatrix},$$

We consider $\rho_{12} = 0.75, \rho_{13} = 0.50$, and $\rho_{23} = 0.20$ for this experiment. Model B is a three dimensional ARCH(3) model. The MSE and MAE for the parameters of the above model and the MAE for ρ is calculated. The results are displayed in Table 7.2.

Table 7.2: Mean Squared Error and Mean Absolute Error of the parameters and constant conditional correlation of a three dimensional ARCH(3) model.

	QMLE	WLE	QMLE	WLE	QMLE	WLE
	$T = 100$		$T = 250$		$T = 1000$	
Standard Normal						
MSE	0.1446	**0.1137**	0.0629	**0.0600**	**0.0177**	0.0185
	(0.0863)	(0.0608)	(0.0328)	(0.0281)	(0.0089)	(0.0093)
MAE	0.9347	**0.8433**	0.6325	**0.6243**	**0.3369**	0.3459
	(0.0956)	(0.2067)	(0.1562)	(0.1477)	(0.0851)	(0.0865)
MAE_ρ	**0.1700**	0.1703	0.1114	**0.1113**	**0.0541**	0.0541
	(0.0956)	(0.0958)	(0.0617)	(0.0618)	(0.0311)	(0.0312)
Student-t(3)						
MSE	0.2113	**0.1777**	0.1419	**0.1035**	0.1002	**0.0799**
	(0.1515)	(0.1287)	(0.1344)	(0.0643)	(0.1010)	(0.0523)
MAE	1.0207	**0.9898**	0.8699	**0.7950**	0.7273	**0.6945**
	(0.3016)	(0.2690)	(0.2661)	(0.1936)	(0.2200)	(0.1809)
MAE_ρ	**0.1035**	0.1043	**0.0802**	0.0802	0.0574	**0.0573**
	(0.06616)	(0.0668)	(0.0451)	(0.0457)	(0.0338)	(0.0339)

MAE is the mean absolute error for constant conditional correlation. Standard errors in parenthesis.

The features we noticed for a 2-dimensional ARCH(2) model in Table 7.1 seem to hold for a 3-dimensional ARCH(3) model in Table 7.2. Again, WLE providing accurate estimates of the model than the QMLE for both error distributions and all sample sizes considered. The only occasion when the MSE and MAE of QMLE show better result than WLE is when sample size in 1000 and errors are normally distributed. Not much difference is observed in the MAE for correlation.

In our final experiment we set some of the ARCH parameters to zero. We are interested to find how well the WLE accounts for zero parameters. Again 1000 replicates each of sample size $T = 100$, 250 and 1000 are generated from the following DGP:

$$\text{Model C:} \quad \begin{pmatrix} \sigma_{1,t}^2 \\ \sigma_{2,t}^2 \end{pmatrix} = \begin{pmatrix} 0.10 + 0.10X_{1,t-1}^2 + 0.20X_{1,t-3}^2 \\ 0.25 + 0.10X_{2,t-2}^2 + 0.05X_{2,t-3}^2 \end{pmatrix},$$

We consider $\rho_{12} = 0.85$ for this experiment. This is a bivariate ARCH(3) model with one ARCH parameter in each series is set to zero. The results of the MSE and MAE for the parameters and the MAE for ρ is tabulated in Table 7.3. The results in Table 7.3 are similar to previous experiments where WLE perform better than the QMLE for non-normal distributions and small sample sizes. Hence, the same conclusion can be drawn from these results that WLE provide better estimates than the QMLE for multivariate ARCH models.

Using our own MATLAB and Fortran code we checked the CPU time (in sec) taken by both WLE and the QMLE for estimating a three dimensional MARCH(2) model. Experiment was performed on a Pentium CPU with Intel Core 2 Duo process running at 2 Ghz and having 2 GB of random access memory (RAM). The sample size used is $T = 10,000$ and the experiment was repeated $K = 1000$ times. WLE took *583.61 sec* where as the QMLE took *2478.20 sec* for estimating the same data sets. We also computed the MSE and the MAE for the parameters and the difference between the two estimators for this large sample size were negligible. This clearly reveals the advantage of using the WLE for estimating the parameters of MARCH models. The WLE take less than one-fourth of the time than the QMLE and also is not only efficient but also estimates the parameters as accurately

Table 7.3: Mean Squared Error and Mean Absolute Error of the parameters and constant conditional correlation of a two dimensional ARCH(3) model when some ARCH parameters are zero.

	QMLE	WLE	QMLE	WLE	QMLE	WLE
	$T = 100$		$T = 250$		$T = 1000$	
Standard Normal						
MSE	0.1002	**0.0803**	0.0357	**0.0340**	**0.0090**	0.0095
	(0.0782)	(0.0577)	(0.0239)	(0.0215)	(0.0060)	(0.0064)
MAE	0.6455	**0.5924**	0.3955	**0.3917**	**0.2009**	0.2070
	(0.2385)	(0.2039)	(0.1359)	(0.1281)	(0.0672)	(0.0686)
MAE_ρ	0.0227	**0.0221**	**0.0145**	0.0146	**0.0071**	0.0071
	(0.0184)	(0.0182)	(0.0115)	(0.0115)	(0.0055)	(0.0055)
Student-t(3)						
MSE	0.1121	**0.1015**	0.0976	**0.0669**	0.0590	**0.0474**
	(0.1244)	(0.0812)	(0.1220)	(0.0701)	(0.0912)	(0.0586)
MAE	0.6961	**0.6591**	0.5587	**0.5028**	0.4298	**0.4159**
	(0.3222)	(0.2451)	(0.2541)	(0.1904)	(0.2073)	(0.1738)
MAE_ρ	**0.0191**	0.0199	**0.0103**	0.0105	**0.0073**	0.0073
	(0.0103)	(0.0105)	(0.0082)	(0.0083)	(0.0057)	(0.0059)

MAE_ρ is the mean absolute error for constant conditional correlation. Standard errors in parenthesis.

as the QMLE.

7.2.3 Application to Real Data Sets

The data sets consist of the daily log-returns of SP500 index and the stocks of Cisco Systems and Intel Corporations from January 2, 1991 to December 31, 1999, a sample of $T = 2275$ observations. These data sets have been used by Tsay (2005, Ch. 10) for modeling higher dimension volatility models. The data were obtained from

http://faculty.chicagobooth.edu/ruey.tsay/teaching/fts/.

The log returns are in percentages and we denote the log-return vector at time t by $\mathbf{X}_t = (X_{1,t}, X_{2,t}, X_{3,t})'$, where $X_{1,t}$, $X_{2,t}$, and $X_{3,t}$, are the log returns at time t of SP500, Cisco and Intel, respectively.

Initial $N = 1275$ observations are used for model estimation and the remaining $K = 1000$ observations are retained for out-of-sample forecasting of volatility. Summary of a few descriptive statistics of the mean corrected returns are provided in Table 7.4. SP500 index is slightly positively skewed where as Cisco and Intel returns show negative skewness. All return series exhibit excessive kurtosis. The Jarque-Bera tests are highly significant with p-values close to zero. The correlation matrix confirms that the returns are correlated with each other. We also calculate the multivariate portmanteau statistic of Hosking (1980) to detect any serial dependence in the return series. This test statistic in multivariate form is defined as:

$$Q_N(M) = T^2 \sum_{l=1}^{M} \frac{1}{T-l} \mathrm{tr}\left[C_{0l} C_{00}^{-1} C_{0l}' C_{00}^{-1}\right],$$

where

$$C_{0j} = \frac{1}{T} \sum_{t=j+1}^{T} \mathbf{X}_t \mathbf{X}_{t-j}',$$

where N is the dimension of $\mathbf{X}_t$, M is the lag length, T is the sample size, and $\mathrm{tr}(\mathbf{A})$ is the trace of the matrix $\mathbf{A}$. Under the null hypothesis (of independence) the test statistic is approximately distributes as a chi-squared with $(N^2 \times M)$ degrees of freedom.

The values of the test statistics obtained up to lag 4 and 8 are $Q(4) = 60.66\,(0.0001)$ and $Q(8) = 100.11\,(0.0006)$, respectively, where p-values are given in parentheses. These statistics are highly significant as compared to chi-squared distribution with 36 and 72 degrees of freedom, respectively. The highly significant p-values suggest that there is indeed some serial dependence in the data.

Parameter Estimation

Now, we turn our attention to volatility modelling. By examining the sample partial autocorrelation plots, we choose a multivariate ARCH model of order $p = 4$ for these return series. Note that for simplicity we choose a diagonal constant

Table 7.4: Descriptive statistics for the mean subtracted return series over the in-sample period.

	SP500 Index	Cisco Systems	Intel Corporation
Minimum	-3.7749	-22.3506	-14.7196
Maximum	3.6161	15.3254	8.9134
Std.dev	0.6523	2.9439	2.3777
Skewness	0.0282	-0.4444	-0.4149
Kurtosis	5.6517	7.8297	6.2800
JB	373.73	1281.17	608.12
	(0.0001)	(0.0001)	(0.0001)
Correlation			
SP500	1.0000	-	-
Cisco	0.4275	1.0000	-
Intel	0.4360	0.4185	1.0000

JB is the Jarque-Bera test for normality and the number in parenthesis are p-values.

conditional correlation multivariate volatility model. Using both the QMLE and WLE, we estimate the model parameters and correlations.

The result of parameter estimation of multivariate ARCH(4) model for the return series are displayed in Table 7.5. The standard errors of the parameters are given in parentheses. All estimates are found significant at 5% significance level. For the QMLE, the Ljung-Box statistics, up to lag 4 and 8, of the squared standardised residual are $Q^2(4) = 12.99\,(0.99)$ and $Q^2(8) = 40.66\,(0.99)$, respectively. For WLE, we have $Q^2(4) = 12.43\,(0.99)$ and $Q^2(8) = 35.68\,(0.99)$. Therefore, both fitted models appear to be adequate in modeling the conditional volatilities.

Table 7.6 represents the estimates of correlations among the return series. By comparing these results with that of Table 7.4, we can see that both models estimates the correlations accurately.

Forecast Evaluation

We evaluate the volatility forecasts generated by both the QMLE and WLE. The rolling window scheme is used in which initial $N = 1275$ sample is used to estimate the model and one-step ahead conditional variance-covariance matrix

Table 7.5: Estimated parameters of multivariate ARCH(4) model using the QMLE and WLE.

Method	Return	A		A			A			A			A	
	SP500	0.3047 (0.0010)	0.0553 (0.0026)			0.04514 (0.0010)			0.1250 (0.0047)			0.0710 (0.0013)		
QMLE	Cisco	4.9877 (0.4782)		0.1893 (0.0053)			0.0766 (0.0021)			0.0796 (0.0061)			0.1160 (0.0080)	
	Intel	4.6876 (0.1983)			0.0089 (0.0002)			0.0372 (0.0008)			0.0541 (0.0008)			0.0756 (0.0041)
	SP500	0.3238 (0.0011)	0.04122 (0.0013)			0.0345 (0.0021)			0.0892 (0.0018)			0.0660 (0.0016)		
WLE	Cisco	5.664 (0.5291)		0.1447 (0.0026)			0.0885 (0.0024)			0.0948 (0.0026)			0.0327 (0.0011)	
	Intel	4.794 (0.2379)			0.0060 (0.0008)			0.0315 (0.0013)			0.0528 (0.0021)			0.0670 (0.0017)

Table 7.6: Estimated correlation of multivariate ARCH(4) model using the QMLE and WLE.

Method	Return	$\hat{\Gamma}$		
		SP500	Cisco	Intel
	SP500	1.0000		
		(-)		
QMLE	Cisco	0.4411	1.0000	
		(0.0231)	(-)	
	Intel	0.4410	0.4237	1.0000
		(0.02390)	(0.0241)	(-)
	SP500	1.0000		
		(-)		
WLE	Cisco	0.4382	1.0000	
		(0.0290)	(-)	
	Intel	0.4400	0.4240	1.0000
		(0.0240)	(0.0249)	(-)

$\hat{\mathbf{H}}_{N+1}$ is generated. The sample is rolled forward one day by including the $(N+1)$-th observation and discarding the first observation. The model is re-estimated and again one-step-ahead forecast is made. This process is repeated till we get to the end of the data. In this way we obtain $K = 1000$ one-step ahead forecasts of variances and covariances.

Two standard evaluation measures used for forecasts evaluation, the root mean squared error (RMSE) and the mean absolute prediction error (MAPE), where

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{1}{K}\sum_{t=1}^{K}(X_{i,t}X_{j,t} - \hat{\sigma}^2_{ij,t})^2}, \\ \text{MAPE} &= \frac{1}{K}\sum_{t=1}^{K}\left|X_{i,t}X_{j,t} - \hat{\sigma}^2_{ij,t}\right|, \quad i,j = 1,2. \end{aligned}$$

Table 7.7 reports results of the MAE and the RMSE for one-step ahead forecasts of conditional variances and covariances of three data sets analysed. Both WLE and QMLE show similar characteristics in forecasting the variance-covariance matrix. The MAPEs and RMSEs of WLE are found less than the QMLE in almost all occasions. The only exception is the RMSE of covariance of SP500 index and

Intel Corporation. In this case the RMSE of QMLE is slightly lower than that of WLE. Hence we can conclude that the forecasting ability of WLE in multivariate ARCH models is not inferior to the QMLE.

Table 7.7: Mean absolute prediction errors and root mean square errors for one-step-ahead forecasts of variances and covariances.

	Variances			Covariances		
	SP500	Cisco	Intel	SP500-Cisco	SP500-Intel	Cisco-Intel
MAPE						
QMLE	1.2356	7.8080	7.3265	2.3209	2.0564	5.3610
WLE	**1.2181**	**7.7342**	**6.3089**	**2.3052**	**2.0016**	**5.1716**
RMSE						
QMLE	3.0754	14.1134	12.8339	5.4377	4.2727	**9.4823**
WLE	**3.0617**	**14.0851**	**12.6451**	**5.4330**	**4.2867**	9.4975

VaR Diagnostics

In the econometric literature, models are often evaluated by their out-of sample forecast performance using standard measures such as the MAPE and the RMSE as used above. Models can also be evaluated from the prospective of their use in risk management. We employ an alternative approach by considering both the QMLE and WLE in terms of their performance in risk management.

We choose equally weighted portfolio weights $\boldsymbol{\omega}$, and estimate the $\alpha\%$ one-step ahead VaR estimate for model i as

$$\hat{q}_{i,t}(\alpha) = c_i(\alpha)\hat{\sigma}^*_{i,t}, \quad i = 1, 2,$$

where $c_i(\alpha)$ is the $\alpha\%$ critical values of the distribution of $\hat{\epsilon}^*_{i,t}$, conditional of information set available up to time $t-1$ and model i, with

$$\hat{\sigma}^{*2}_{i,t} = \boldsymbol{\omega}' \hat{\mathbf{H}}_{i,t} \boldsymbol{\omega},$$

and

$$\hat{\epsilon}^*_{i,t} = \frac{\omega' \mathbf{X}_t}{\hat{\sigma}^*_{i,t}}.$$

Figure 7.1 shows the 1% one-day ahead VaR estimates of both models. It can be noticed that both models approximately produce similar results. The VaR estimates of the QMLE are slightly larger in absolute values as compared to those of WLE. This reveals that the QMLE produces higher risk estimates. We also calculate the number of violations of each model and test the hypothesis $H_0 : f = \alpha$ against $H_1 : f \neq \alpha$, where f is the failure rate estimated by the empirical failure rate. The Kupiec's unconditional likelihood ratio statistic is defined as

$$LR_{uc} = 2 \ln [\hat{f}(1 - \hat{f})^{K-K}] - 2 \ln [\alpha^{K} (1 - \alpha)^{K-K}],$$

where K is the total number of out-of-sample observations and K_* is the number of VaR violations (Kupiec, 1995). Under the null hypothesis, LR_{uc} is asymptotically distributed as a $\chi^2_{(1)}$. For the QMLE, we get $LR_{uc} = 1.8862\,(0.1696)$ and for WLE, we have $LR_{uc} = 0.4337\,(0.5102)$, where numbers in parentheses are p-values. Although both estimators pass the Kupiec's test, WLE has large p-value than the QMLE.

The results of volatility forecasts and risk estimates, based on the evaluations measure we used, show that the forecasting performance of WLE is not inferior to the QMLE. This suggests that one may use WLE for estimating multivariate ARCH models instead of the QMLE. We again mention the advantage of using the WLE that it is very easy to compute and require less CPU time for estimation and forecasting and by using WLE for multivariate ARCH models we are not losing anything in terms of volatility forecasts and risk estimation.

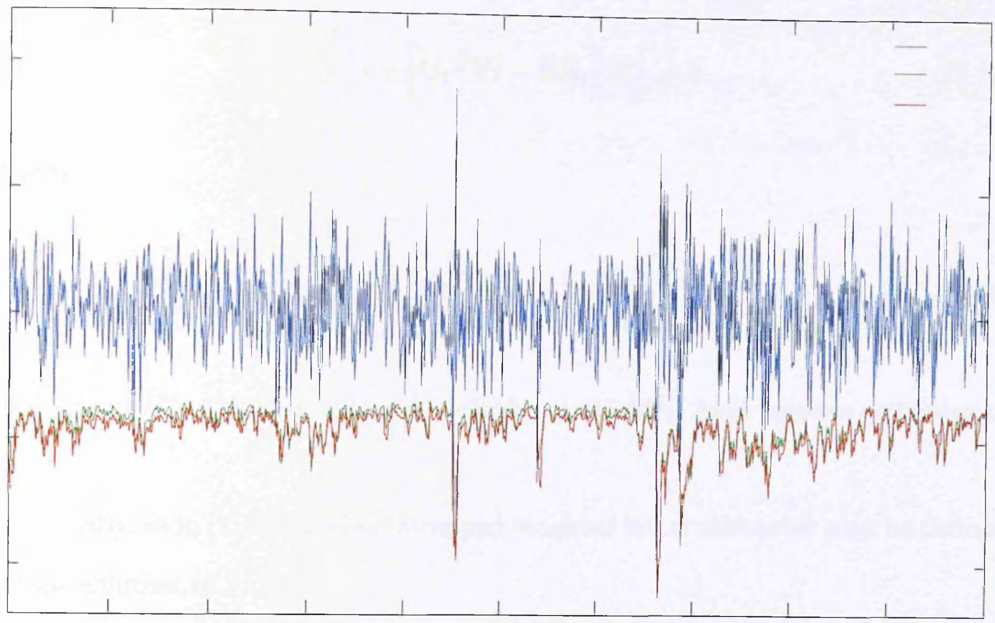

Figure 7.1: Portfolio returns (blue) of SP500, Cisco and Intel stock indices along with one-step ahead 1% VaR estimates of the QMLE and WLE.

7.3 A Weighted Resampling for the WLE in Multivariate ARCH models

In this section we define weighted resampling method for weighted linear estimator in multivariate ARCH models. A bootstrap technique proposed by Chatterjee and Bose (2005) is used for this purpose. Using the idea of weighted resampling we develop suitable bootstrap versions for WLE. We bootstrap WLE and our goal is to approximate the sampling distribution of the parameters of multivariate ARCH models with this new approach to resampling.

Let $\{w_{Tt}; 1 \leq t \leq T, T \geq 1\}$ be a triangular array of r.v.'s such that for each $T \geq 1$, $\{w_{Tt}; 1 \leq t \leq T\}$ are exchangeable, independent of $\{\mathbf{X}_t; t \geq 1-p\}$. These are called the bootstrap weights.

From (7.7), the bootstrapped preliminary weighted least squares estimator $\hat{\mathbf{B}}^*_{pr}$ of $\mathbf{B}$ is defined as the solution of

$$\sum_{t=1}^{T} w_{Tt}\Big[\mathbf{U}_t\Big\{\mathbf{Y}_t - \mathbf{B}\mathbf{Z}_t\Big\}\mathbf{Z}_t'\Big] = \mathbf{0} \tag{7.17}$$

Hence,

$$\hat{\mathbf{b}}^*_{i,pr} = \left(\sum_{t=1}^{T} w_{Tt}u_{i,t}\Big\{Y_{t,i}\mathbf{Z}_t'\Big\}\right)\left(\sum_{t=1}^{T} w_{Tt}u_{i,t}\Big\{\mathbf{Z}_t\mathbf{Z}_t'\Big\}\right)^{-1}, \quad i = 1, 2, ..., N, \tag{7.18}$$

and we get $\hat{\mathbf{B}}^*_{pr}$, the bootstrapped preliminary weighted least squares estimator of $\mathbf{B}$.

Finally, as in (7.10), the bootstrapped weighted linear estimator may be defined as the solution of

$$\sum_{t=1}^{T} w_{Tt}\Big[\mathbf{V}_t\Big\{(\hat{\boldsymbol{\Sigma}}_t^{-2})'\hat{\boldsymbol{\Sigma}}_t^{-2}\Big\}\Big\{\mathbf{Y}_t - \mathbf{B}\mathbf{Z}_t\Big\}\mathbf{Z}_t'\Big] = \mathbf{0}. \tag{7.19}$$

This gives us

$$\hat{\mathbf{b}}^*_{i,T} = \left(\sum_{t=1}^{T} w_{Tt}v_{i,t}\Big\{Y_{i,t}\mathbf{Z}_t'/(\hat{\sigma}^2_{i,t})^2\Big\}\right)\left(\sum_{t=1}^{T} w_{Tt}v_{i,t}\Big\{\mathbf{Z}_t\mathbf{Z}_t'/(\hat{\sigma}^2_{i,t})^2\Big\}\right)^{-1}, \; i = 1, 2, ..., N, \tag{7.20}$$

and the final bootstrapped weighted linear estimator $\hat{\mathbf{B}}^*_T$ is obtained.

We study these approximations via simulations based on three different schemes of weights. These are

(i) Scheme M when $\{w_{Tt}\}$ have a multinomial $(T, 1/T, ..., 1/T)$ distribution.

(ii) Scheme U when $w_{Tt} = U_t/\bar{U_T}$, where U_t's are i.i.d Uniform (0.5,1.5), $1 \leq t \leq T$ and $\bar{U_T} = (\sum_{i=1}^{T} U_i)/T$.

(iii) Scheme E when $w_{Tt} = E_t/\bar{E_T}$, where E_t's are i.i.d Exponential (1), $1 \leq t \leq T$ and $\bar{E_T} = (\sum_{i=1}^{T} E_i)/T$.

Note that Scheme M corresponds to the commonly-used paired-bootstrap in heteroscedastic models. In this chapter we empirically study Schemes U and E as possible alternatives to the paired-bootstrap. It is possible to obtain quantiles of the bootstrap distribution of $\sigma_T^{-1}T^{1/2}(\hat{\mathbf{b}}^*_{i,T} - \hat{\mathbf{b}}_{i,T})$ using simulation and then using the bootstrap approximation, we can construct the bootstrap confidence intervals of $\mathbf{B}$.

7.3.1 Asymptotics of the Bootstrapped WLE

For all $T \geq 1$,

$$\{w_{Tt}; 1 \leq t \leq T\} \text{ are exchangeable and independent of} \{\mathbf{Z}_t, \epsilon_{i,t}, u_{i,t}, v_{i,t}, 1 \leq t \leq T\}. \tag{7.21}$$

The weights are assumed to satisfy the following basic conditions of Chatterjee and Bose (2005),

$$\begin{aligned} & \mathrm{E}_B(w_{T1}) = 1,\ 0 < k_1 < \sigma_T^2 = o(T), \\ \text{and} \quad & \mathrm{corr}_B(w_{T1}, w_{T2}) = O(T^{-1}), \end{aligned} \tag{7.22}$$

where $\sigma_T^2 = V_B(w_{Tt})$, $k_1 > 0$ is a constant, E_B denotes the expectation with respect to the bootstrap distribution and $o(1)$ denotes the convergence in probability to zero.

Lemma 1. Suppose that model assumptions, (7.21) and (7.22) hold, then for some $\delta > 2$,

$$\mathrm{E}\{(u_{i,1}Y_{i,-j}Y_{i,-k})^\delta\} < \infty, \forall j, k,$$

then

$$\sigma_T^{-1}T^{1/2}(\hat{\mathbf{b}}^*_{i,pr} - \mathbf{b}_i) = O_B(1).$$

The standardised exchangeable weights $\{W_{Tt} := (w_{Tt} - 1)/\sigma_T\}$ satisfy

$$\mathrm{E}_B(W_{Tt}^4) < \infty, \text{ and } \lim_{T\to\infty} \mathrm{E}_B(W_{T1}^2 W_{T2}^2) = 1. \tag{7.23}$$

In addition, we assume that

$$\sigma_T^{-1} T^{1/2}(\hat{\mathbf{b}}_{i,pr}^* - \mathbf{b}_i) = O_B(1). \tag{7.24}$$

Also, For some $\delta > 0$,

$$\begin{aligned} &\mathrm{E}(v_{i,1}^2) < \infty,\ \mathrm{E}\{(v_{i,1}Y_{i,-j})^\delta\} < \infty,\ \mathrm{E}\{(v_{i,1}Y_{i,-j}Y_{i,-k})^\delta\} < \infty, \\ &\mathrm{E}\{(v_{i,1}^2 Y_{i,-j}Y_{i,-k})^\delta\} < \infty, \text{ and } \mathrm{E}\{(v_{i,1}Y_{i,-j}Y_{i,-k}Y_{i,-l})^\delta\} < \infty, \\ &\forall j,k,l \in \mathcal{P}_0, i = 1,\cdots,N. \end{aligned} \tag{7.25}$$

Suppose that model assumptions, (7.21)-(7.25) hold, heuristically then

$$\begin{aligned} \sigma_T^{-1} T^{1/2}(\hat{\mathbf{b}}_{i,T}^* - \hat{\mathbf{b}}_{i,T}) = \quad & \left[\mathrm{E}(v_{i,1}(\mathbf{Z}_1\mathbf{Z}_1')(\mathbf{b}_i\mathbf{Z}_1)^{-2})\right]^{-1} \times \\ & T^{-1/2}\sum_{t=1}^{T} \eta_{i,t} W_t v_{t,i} \mathbf{Z}_1(\mathbf{b}_i\mathbf{Z}_1)^{-1} + o_B(1), \end{aligned}$$

and hence

$$\sup\{|F_B(\mathbf{x}) - F_T(\mathbf{x})|; \mathbf{x} \in \mathbb{R}^{1+p}\} = o_p(1),$$

where

$F_B(\mathbf{x}) = P_B[\sigma_T^{-1} T^{1/2}(\hat{\mathbf{b}}_{i,T}^* - \hat{\mathbf{b}}_{i,T}) \le \mathbf{x}]$ and $F_T(\mathbf{x}) = P[T^{1/2}(\hat{\mathbf{b}}_{T,i} - \mathbf{b}_i) \le \mathbf{x}]$.

7.3.2 Results

We consider Model B, a three dimensional ($N = 3$) ARCH model of order $p = 3$ as our DGP and 10,0000 samples of size $T = 250$ and 1000 are generated to approximate the distribution of $\sqrt{T}(\hat{\mathbf{B}}_T - \mathbf{B})$, where $\hat{\mathbf{B}}_T$ is the matrix of estimated parameter with the i-th row given as $\hat{\mathbf{b}}_{i,T} = (\hat{b}_{i0}, \hat{b}_{i1}, \hat{b}_{i2}, \hat{b}_{i3})$, $1 \le i \le N$. The mean and the averages of the squares of these sets of K numbers are computed and these represent the means and the mean squared error (MSE) of $\sqrt{T}(\hat{\mathbf{b}}_{i,T} - \mathbf{b}_{i,T})$, $1 \le i \le N$.

The MSE under normal approximations are also computed. The estimated MSE under normal approximation at the k-th replication, $1 \le k \le K$, is obtained from the diagonals of matrices

$$\hat{V}_{\epsilon,i}\Big[T^{-1}\sum_{t=1}^{T}\{v_{t,i}(\mathbf{Z}_t\mathbf{Z}_t')(\hat{\mathbf{b}}_i\mathbf{Z}_t)^{-2}\}\Big]^{-1}\Big[T^{-1}\sum_{t=1}^{T}\{v_{t,i}^2(\mathbf{Z}_t\mathbf{Z}_t')(\hat{\mathbf{b}}_i\mathbf{Z}_t)^{-2}\}\Big]$$
$$\Big[T^{-1}\sum_{t=1}^{T}\{v_{t,i}(\mathbf{Z}_t\mathbf{Z}_t')(\hat{\mathbf{b}}_i\mathbf{Z}_t)^{-2}\}\Big]^{-1},$$

where $\hat{V}_{\epsilon,i}$ is the variance of the squared residuals $\{\hat{\epsilon}_{i,1}^2, \cdots, \hat{\epsilon}_{i,T}^2\}$, with residuals $\hat{\epsilon}_{i,t} = X_{i,t}/(\hat{\mathbf{b}}_i\mathbf{Z}_t)^{1/2}$, $1 \le t \le T$, $1 \le i \le N$, and $\hat{\mathbf{b}}_i$ being the estimate based on the k-th replication.

Table 7.8 reports the mean, mean squared error, and the mean squared error under normal approximation of $\sqrt{T}(\hat{\mathbf{B}}_T - \mathbf{B})$. The means are significantly different from normal approximations. The MSEs are found close to MSE under normal approximations.

Next we use weighted resampling to approximate the distribution of $\sqrt{T}(\hat{\mathbf{B}}_T - \mathbf{B})$. From many different choices of bootstrap weights, We choose Scheme M, Scheme U and Scheme E for this study. The weighted resampling procedure is described as follows. We generate $B^* = 1000$ bootstrap samples on the exchangeable weights $\{w_{t,i}\}$'s and compute $\sigma_{T,i}^{-1}\sqrt{T}(\hat{\mathbf{b}}_{i,T}^* - \hat{\mathbf{b}}_{i,T}^{(k)})$, where $\sigma_{T,i}$ depends on the underlying scheme, $\hat{\mathbf{b}}_{i,T}^*$ is the bootstrapped estimated parameter vector and $\hat{\mathbf{b}}_{i,T}^{(k)}$ is the estimated parameter vector of i-th series chosen at k-th replication, $1 \le k \le K$.

Table 7.8: Means and the MSEs of the distributions of $\sqrt{T}(\hat{\mathbf{B}}_T - \mathbf{B})$ for a three dimensional ARCH(3) model and the MSE due to normal approximation of the distribution.

	$T = 250$			$T = 1000$		
	Mean	MSE	MSE$_N$	Mean	MSE	MSE$_N$
b_{10}	0.0496	0.0997	0.1066	0.0578	0.1058	0.1017
b_{11}	0.2634	1.3616	1.7719	0.0067	1.6247	1.7031
b_{12}	-0.2857	1.9369	1.9912	-0.2907	2.3895	2.1118
b_{13}	-0.1926	1.9898	2.0927	-0.1488	2.2786	2.2043
b_{20}	-0.0506	0.4419	0.5133	0.0557	0.4911	0.5078
b_{21}	-0.1780	2.8131	2.7706	-0.0636	3.0078	3.0064
b_{22}	0.0239	1.0227	1.7419	-0.2073	1.6712	1.7307
b_{23}	0.3920	0.8381	1.4718	0.1220	0.9533	1.3131
b_{30}	0.0298	0.0224	0.0231	0.0313	0.0233	0.0222
b_{31}	-0.2769	2.1815	2.1069	-0.1557	2.4894	2.3200
b_{32}	-0.1922	1.4920	1.8385	-0.2227	2.0739	1.9256
b_{33}	0.1033	1.0788	1.6383	-0.1213	1.5311	1.5891

All results are based on 10,000 replications. MSE is the MSE under normal approximation.

We set $K = 100$ for this experiment. The means and the averages of the squares of $\sigma_{T,i}^{-1}\sqrt{T}(\hat{\mathbf{b}}_{i,T}^{*} - \hat{\mathbf{b}}_{i,T}^{(k)})$ are computed and these represent the bootstrapped mean and mean squared errors.

Table 7.9 shows the result of means and the MSEs of the distribution of the standardized bootstrap estimators for a three dimensional ARCH(3) model under different schemes. The entries in bold represent schemes providing closest approximations of MSE for corresponding parameters. Note that none of the scheme provide very good approximation of means although in few cases each scheme estimate the means for b_{i1} quite well. In estimating the MSE, for small sample size $T = 250$, scheme E shows good performance overall with scheme M following. Scheme U also approximate the MSEs once in each series. For large sample size $T = 1000$ it can be seen that scheme U dominates other schemes in terms of approximating MSEs. Scheme E shows better results than scheme U in three instances where as scheme M does not provide results better than other two schemes.

These findings suggest that our bootstrap weighted linear estimator shows nice results in approximating the shape of the distribution. We also observe that other schemes such as scheme U and scheme E can be used for weighted resampling as they provide better results the the commonly used paired bootstrap, scheme M.

Table 7.9: Means and the MSEs of the distribution of the standardized bootstrap estimators for a three dimensional ARCH(3) model under different schemes.

$B^* = 1000$	Scheme M		Scheme U		Scheme E	
$T = 250$	Mean	MSE	Mean	MSE	Mean	MSE
b_{10}	-0.0459	**0.0934**	-0.0114	0.1053	-0.0454	0.0843
b_{11}	0.2967	1.6269	0.0083	2.2369	0.2409	**1.5007**
b_{21}	-0.0674	**2.0174**	-0.0508	2.3931	-0.1096	1.8179
b_{31}	-0.0222	1.8346	-0.0231	**2.0920**	-0.0540	1.6480
b_{20}	-0.2000	0.4909	-0.0120	0.5195	-0.1974	**0.4452**
b_{21}	-0.0366	2.3938	-0.0222	**2.7544**	-0.0488	2.2025
b_{22}	0.1875	1.3947	-0.0146	1.6949	0.1363	**1.2868**
b_{23}	0.3359	1.1111	0.0618	1.3523	0.2924	**1.0421**
b_{30}	-0.0277	**0.0230**	0.0042	0.0250	-0.0252	0.0211
b_{31}	0.0406	1.9639	-0.0376	**2.3332**	-0.0014	1.8208
b_{32}	0.1229	**1.5190**	-0.0290	1.8537	0.0860	1.4429
b_{33}	0.1870	1.3577	-0.0134	1.6576	0.1568	**1.2662**
$T = 1000$						
b_{10}	0.0101	0.1036	0.0079	**0.1038**	0.0072	0.0999
b_{11}	0.1115	1.7413	0.0083	1.8280	0.0986	**1.6888**
b_{12}	-0.1451	2.2548	-0.0508	**2.2693**	-0.1395	2.1386
b_{13}	-0.0940	2.3134	-0.0231	**2.3098**	-0.0806	2.1733
b_{20}	-0.0659	0.4977	-0.0120	**0.4915**	-0.0668	0.4830
b_{21}	-0.0697	3.1201	-0.0222	**3.0679**	-0.0670	2.9488
b_{22}	-0.0084	1.4900	-0.0146	**1.6566**	-0.0200	1.4556
b_{23}	0.2238	1.1155	0.0618	**1.0631**	0.2069	1.0868
b_{30}	0.0090	0.0231	0.0042	**0.0233**	0.0084	0.0223
b_{31}	-0.1068	2.5472	-0.0376	2.5653	-0.1091	**2.4550**
b_{32}	-0.0800	2.0016	-0.0290	**2.0108**	-0.0705	1.9183
b_{33}	0.0033	1.5630	-0.0134	1.5676	0.0031	**1.5097**

All results are based on 100 replications.

7.4 Conclusion

Weighted linear estimator for the multivariate ARCH parameters is proposed in this chapter. The accuracy of this estimators is compared empirically with the QMLE in estimating the parameters of multivariate ARCH models. Using normal errors we found that WLE produced as accurate results as the QMLE for large sample size and outperforming QMLE in small samples. Under heavy-tailed errors, WLE always outperform QMLE in terms of the mean squared error and the mean absolute error. These estimators are also applied to real data sets and forecasts of volatilities and value-at-risk are obtained. Our results showed that the forecasting performance of WLE is not inferior to the QMLE and one-day risk estimates are also found better.

We also proposed weighted resampling method for multivariate ARCH models. Using different weights for bootstrap we showed that other schemes for weights such as scheme U and scheme E provide better results than commonly used paired bootstrap.

We suggest using WLE for the estimation of multivariate ARCH models as this estimator has a closed form expression and easy to estimate. WLE takes less computational time than the QMLE where some optimisation methods are used to reach the convergence. The quick estimation of multivariate ARCH models using WLE allows researchers to apply very heavy computational methods such as bootstrapping in multivariate models.

Chapter 8

Robust Methods for Multivariate GARCH Models

8.1 Introduction

Multivariate GARCH models focus on volatility and correlation analysis for more than one asset. In order to study the relations between the volatilities and co-volatilities of financial time series, the development of multivariate GARCH-type models is very important. Modelling and predicting the time varying dynamics of conditional covariances of asset returns is also crucial for asset pricing, portfolio allocation and risk management; see, for example, Bollerslev et al. (1988).

One major problem with the application of multivariate GARCH models is that the number of parameters to be estimated increases rapidly with the dimension of the system. The likelihood function becomes very flat and consequently the optimisation of the likelihood function becomes infeasible. Another problem is that many constraints need to be imposed to ensure the positive definiteness of the conditional covariance matrix. To deal with these problems, different specifications of multivariate GARCH models have been proposed. Bauwens et al. (2006) provide an extensive survey on multivariate GARCH models.

Bollerslev et al. (1988) first introduced the multivariate GARCH model in the familiar half-vec (*vech*) form for the conditional covariance matrices where every conditional variance and covariance is a function of all lagged conditional variances and covariances, as well as lagged squared returns and cross-products of returns. The estimation of the *vech* model is computationally demanding because of the large number of parameters in the model and also there are restrictive conditions on the conditional covariance matrices to be positive definite. A simplified version of this model called diagonal *vech* was proposed by Bollerslev et al. (1988) assuming the parameters matrices to be diagonal. A special case of the *vech* model is the Baba-Engle-Kraft-Kroner (BEKK) model of Engle and Kroner (1995). Further simplification can be achieved by using the diagonal BEKK (DBEKK) model assuming that parameters matrices are diagonal. BEKK models generate positive definite conditional covariances matrices but estimation of these models still need heavy computation due to several matrix inversions.

Bollerslev (1990) proposed the constant conditional correlation multivariate GARCH models (CCC-GARCH) where the conditional correlation matrix is time-invariant. The CCC-GARCH model is very popular amongst the practitioner because of its simplicity. An extended CCC-GARCH (ECCC-GARCH) model was introduced by Jeantheau (1998) in which the assumption that parameters matrices are diagonal is relaxed. The ECCC-GARCH allows a considerably richer autocorrelation structure for the squared observed returns than the standard CCC-GARCH model (He and Teräsvirta, 2004). Although CCC-GARCH models are computationally attractive, many empirical studies suggested that the assumption of constant conditional correlation is too restrictive. By making the conditional correlation matrix time-dependent, Engle (2002) proposed the Dynamic Conditional Correlation (DCC) and Tse and Tsui (2002) proposed Time Varying Correlation (TVC) GARCH models. Under both these specifications, the correlation is time varying and is able to capture the changes over time. A two-step estimation approach can be used for DCC models by writing the log-likelihood as the sum of a mean and volatility part and a correlation part (Engle and Sheppard, 2001). Ledoit et al. (2003) proposed a two-step approach by estimating each variance and covariance equation separately.

Another approach is to use univariate GARCH models to obtain the multivariate GARCH estimates. Engle et al. (1990) proposed a method to reduce the dimensionality of the problem by assuming that the observations are generated by underlying factors that are conditionally heteroscedastic and posses a GARCH-type structure. This model is called the factor GARCH (F-GARCH) model. Alexander and Chibumba (1997) proposed the Orthogonal GARCH (O-GARCH) model which is a generalization of the F-GARCH model to a multi-factor model with orthogonal factors. Alexander (2000, 2000b) further discussed that this method can be used for obtaining large positive semi-definite conditional covariance matrices by modeling the principal components of the financial returns' unconditional covariance matrix as univariate GARCH processes. The GO-GARCH model by van der Weide (2002) generalised the Orthogonal GARCH approach by allowing for the linear map that links components and observed data to be non orthogonal. Vrontos et al. (2003) introduced a variant of the factor models called the full-factor multivariate GARCH model.

Harris et al. (2007) suggested a method of estimating the elements of the conditional covariance matrix using univariate GARCH models and called it a simplified multivariate GARCH (S-GARCH) model. For each pair of variables, S-GARCH model estimates four univariate GARCH models - one for each variable, one on the sum of the variables and one on the difference of the variables. The difference of the later two yields four times the covariance between the two variables.

The objective of this research is to propose robust estimators for multivariate GARCH models that are easy to estimate and do not put additional constraints on the model. We present two new methods for estimating the multivariate GARCH models. These methods can be considered as robust versions of the S-GARCH model of Harris et al. (2007) and the O-GARCH model of Alexander (2000). Both models use univariate GARCH specification and estimates of these univariate GARCH models are obtained using M-estimators. The proposed methods have many advantages. First, these are easy to estimate as compared to other multivariate GARCH specifications as these require estimation of some univariate GARCH models. Second, they provide robust estimate of the parameters of multivariate GARCH models that according to our knowledge has not been proposed in these settings. Third, besides multivariate normal distribution and multivariate Student-t distribution not many multivariate distributions are used in practice for multivariate GARCH models. Although, the multivariate normal distribution is easy to use it is not consistent with the well-known asymmetry and excess kurtosis in financial data. The use of univariate GARCH models allow us to use many univariate densities for errors to capture these stylised facts which for multivariate case would be extremely difficult if not impossible.

Volatility forecasting has been the main focus of most of the previous studies. In this study we focus on correlation estimation and forecasting of multivariate GARCH models which have not received significant attention in the literature. In addition we investigate the forecasting performance of our methods and M-estimators in predicting value-at-risk using various evaluation measures.

Using Monte Carlo simulations the in-sample and out-of-sample performances of M-estimators for both models are measured in terms of estimating and predicting the time varying correlations between the returns. We estimate our models in two distinct applications to stock indices and currency exchange rates and focus on predictions for conditional correlations and value-at-risk. Several in-sample and out-of-sample performance measures are used to evaluate the predicted VAR of both models. Using these performance measures, we collect empirical evidences of the better predictive potential of B-estimator over other competing M-estimators. Our results suggest the use of robust estimators such as LAD and B-estimator for the estimation and prediction of multivariate GARCH models. We also discuss and compare both multivariate GARCH models with univariate GARCH specifications and find that the O-GARCH model has some advantages over the S-GARCH model.

The plan of the rest of the chapter is as follows. In Section 8.2, we propose the robust version of the S-GARCH model for estimating the parameters of the multivariate GARCH models using M-estimators. In Section 8.3, using M-estimators a robust version of the O-GARCH model is proposed. For both methods, Monte-Carlo simulations are performed to obtain the in-sample and out-of-sample correlation estimates. These methods are also applied to real data sets and out-of-sample correlations and risk estimates are obtained. The out-of-sample forecasts are evaluated using various performance measures. Section 8.4 discusses and compares both models and Section 8.5 concludes the chapter.

8.2 A Robust Method for the Simplified GARCH Model

The simplified multivariate GARCH (S-GARCH) model was proposed by Harris, Stoja, and Tucker (2007) (hereafter HST). This method involves estimating the elements of the conditional covariance matrix using univariate GARCH models. For each pair of variables, the S-GARCH model estimates four univariate GARCH models - one for each variable, one on the sum of the variables and one on the difference of the variables. The difference of the later two yields four times the covariance between the two variables.

Let $x_{1,t}$ and $x_{2,t}$ are two return series at time t, with

$$E[x_{i,t}|\mathcal{F}_{t-1}] = 0, \quad i,j = 1,2,$$

where $\mathcal{F}_{t-1}$ is the conditioning information set available at time $t-1$. The conditional covariance matrix of $x_{1,t}$ and $x_{2,t}$ is given by

$$\mathbf{H}_t = \begin{bmatrix} \sigma_{1,t}^2 & \sigma_{12,t} \\ \sigma_{12,t} & \sigma_{2,t}^2 \end{bmatrix}, \tag{8.1}$$

where $\sigma_{1,t}^2 = \text{Var}[x_{1,t}|\mathcal{F}_{t-1}]$, $\sigma_{2,t}^2 = \text{Var}[x_{2,t}|\mathcal{F}_{t-1}]$ are the variance elements and $\sigma_{12,t} = \sigma_{21,t} = \text{Cov}[x_{1,t}x_{2,t}|\mathcal{F}_{t-1}]$ are the covariance elements of $\mathbf{H}_t$, respectively. HST proposed a simple procedure for estimating the elements of the conditional covariance matrix, $\mathbf{H}_t$, that involves estimating only univariate GARCH models. First, using the GARCH(1,1) specification, the conditional variances, $\sigma_{1,t}^2$ and $\sigma_{2,t}^2$, are given by

$$\sigma_{1,t}^2 = \omega_1 + \alpha_1 x_{1,t-1}^2 + \beta_1 \sigma_{1,t-1}^2 \tag{8.2}$$

$$\sigma_{2,t}^2 = \omega_2 + \alpha_2 x_{2,t-1}^2 + \beta_2 \sigma_{2,t-1}^2 \tag{8.3}$$

HST used the quasi-maximum likelihood to obtain the estimates of model parameters and corresponding estimates of the conditional variances in (8.2) and (8.3). The QMLE is consistent and asymptotically normal if the innovation has finite four moments. This estimate is very sensitive to the presence of a few outliers in the sample and may not be considered a good choice when there is evidence of heavy-taildness. In this study we propose estimating the model (8.2) and (8.3) using robust M-estimators. It was shown that these estimators provides good estimates for non-normal data (see Section 3.4 for results and discussion).

Next, by adding and subtracting the series $x_{1,t}$ and $x_{2,t}$, two new series $x_{+,t} = x_{1,t} + x_{2,t}$ and $x_{-,t} = x_{1,t} - x_{2,t}$ are constructed. Using a GARCH(1,1) specification, the conditional variances of these two series are given by

$$\sigma^2_{+,t} = \omega_+ + \alpha_+ \, x^2_{+,t-1} + \beta_+ \, \sigma^2_{+,t-1} \tag{8.4}$$

$$\sigma^2_{-,t} = \omega_- + \alpha_- \, x^2_{-,t-1} + \beta_- \, \sigma^2_{-,t-1} \tag{8.5}$$

where $\sigma^2_{+,t} = \mathrm{Var}[x_{+,t}|\mathcal{F}_{t-1}]$ and $\sigma^2_{-,t} = \mathrm{Var}[x_{-,t}|\mathcal{F}_{t-1}]$. Again, we use M-estimators to obtain the estimates of the parameters of the model and corresponding estimates of the conditional variances. An estimate of the conditional covariance can then be base on the following identities.

$$\sigma^2_{+,t} \equiv \sigma^2_{1,t} + \sigma^2_{2,t} + 2\sigma_{12,t} \tag{8.6}$$

$$\sigma^2_{-,t} \equiv \sigma^2_{1,t} + \sigma^2_{2,t} - 2\sigma_{12,t} \tag{8.7}$$

Combining (8.6) and (8.7) gives the conditional covariance

$$\sigma_{12,t} \equiv (1/4)(\sigma^2_{+,t} - \sigma^2_{-,t}) \tag{8.8}$$

The conditional correlation between two returns can be easily calculated using (8.2), (8.3), and (8.8) as

$$\rho_t = \frac{\sigma_{12,t}}{\sigma_{1,t}\,\sigma_{2,t}}. \tag{8.9}$$

Our method consist of estimating several GARCH(1,1) models using a class of M-estimators as defined is Chapter 3. The method is easy to implement, does not impose extra restriction on model and most importantly robust in the presence of outliers. Instead of the GARCH model, any other univariate GARCH specification can be used to estimate the multivariate GARCH models. Asymmetric GARCH model such as GJR model can be used to capture the asymmetric feature of financial data set. Also, high order GARCH models can be fitted although empirical studies have shown that the GARCH(1,1) model provides a good fit to the data. See HST for some properties of the S-GARCH model under the univariate GARCH specification.

As discussed by HST, a potential problem with this approach is that the resulting estimate of the conditional correlation matrix is not necessarily positive semidefinite. There are many approaches to handle this problem. We use the simplest approach where the eigenvalues and eigenvectors of the estimated conditional correlation matrix are extracted and negative eigenvalues are truncated to some small positive numbers. Then, these truncated eigenvalues and the original eigenvectors are used to construct the pseudo-correlation matrix. Other approaches are to use the nearest correlation matrix as proposed by Higham (2002) or to transform the matrices of the parameters estimates such that the resulting matrices of parameters estimates are positive semidefinite (Ledoit at el. 2003).

Next, the method of forecasting volatilities and co-volatilities from the S-GARCH models is described. For a bivariate case, k-step ahead forecasts of the variances are derived as

$$\hat{\sigma}^2_{i,T+k} = \frac{\hat{\omega}_i[1-(\hat{\alpha}_i+\hat{\beta}_i)^{k-1}]}{(1-\hat{\alpha}_i-\hat{\beta}_i)} + (\hat{\alpha}_i+\hat{\beta}_i)^{k-1}\hat{\sigma}^2_{i,T+1}, \quad k>1,\ i=1,2, \tag{8.10}$$

where $\hat{\sigma}^2_{i,T+1} = \hat{\omega}_i + \hat{\alpha}_i\, x^2_{i,T} + \hat{\beta}_i\, \hat{\sigma}^2_{i,T}$.

Similarly, k-step ahead forecasts of conditional covariances can be obtained from

$$\hat{\sigma}_{12,T+k} = (1/4)(\hat{\sigma}^2_{+,T+k} - \hat{\sigma}^2_{-,T+k}), \qquad (8.11)$$

where

$$\hat{\sigma}^2_{\pm,T+k} = \frac{\hat{\omega}_\pm[1-(\hat{\alpha}_\pm+\hat{\beta}_\pm)^{k-1}]}{(1-\hat{\alpha}_\pm-\hat{\beta}_\pm)} + (\hat{\alpha}_\pm+\hat{\beta}_\pm)^{k-1}\hat{\sigma}^2_{\pm,T+1}, \quad k>1,$$

with $\hat{\sigma}^2_{\pm,T+1} = \hat{\omega}_\pm + \hat{\alpha}_\pm\, x^2_{\pm,T} + \hat{\beta}_\pm\, \hat{\sigma}^2_{\pm,T}$.

Hence, using (8.10) and (8.11), the k-step ahead forecast of conditional correlation form the S-GARCH model can be obtained as

$$\hat{\rho}_{T+k} = \frac{\hat{\sigma}_{12,T+k}}{\hat{\sigma}_{1,T+k}\,\hat{\sigma}_{2,T+k}}. \qquad (8.12)$$

8.2.1 Simulation and Empirical Results

In this section we perform Monte Carlo simulations to evaluate the in-sample and out-of-sample performance of our proposed method. Applications to real data sets are also presented.

Monte Carlo Simulations

A bivariate GARCH(1,1) model is simulated 1000 times using the sample size of $T = 1000$ observations. The data generating process is given by

$$\sigma^2_{1,t} = 0.01 + 0.05\, x^2_{1,t-1} + 0.85\, \sigma^2_{1,t-1},$$
$$\sigma^2_{2,t} = 0.10 + 0.01\, x^2_{2,t-1} + 0.60\, \sigma^2_{2,t-1},$$

with $x_{1,t} = \sqrt{h_{1,t}}\,\epsilon_{1,t}$, $x_{2,t} = \sqrt{h_{2,t}}\,\epsilon_{2,t}$, and

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{1,t} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_t \\ \rho_t & 1 \end{pmatrix}\right], \quad t = 1, ..., T.$$

We also consider simulating errors from Student-t distribution with 3 degrees of freedom and Skewed-t distribution with 4 df and skewness of 0.25. For correlations, we consider the following processes:

1. Constant: $\rho_t = 0.90$

2. Sine: $\rho_t = 0.50 + 0.40\cos(2\pi t/200)$

3. Fast sine: $\rho_t = 0.50 + 0.40\cos(2\pi t/20)$

4. Step: $\rho_t = 0.90 - 0.50(t > 500)$

5. DCC:

$$\begin{aligned} \rho_t &= \frac{q_{12,t}}{\sqrt{q_{11,t}q_{22,t}}}, \\ q_{12,t} &= 0.02 + 0.30\,\epsilon_{1,t-1}\,\epsilon_{2,t-1} + 0.60\,q_{12,t-1}, \\ q_{ii,t} &= 0.10 + 0.30\,\epsilon_{i,t-1}^2 + 0.60\,q_{ii,t-1}, \quad i = 1, 2. \end{aligned}$$

6. TVC:

$$\begin{aligned} \rho_t &= 0.07 + 0.10\,\psi_{t-1} + 0.80\,\rho_{t-1}, \\ \psi_{t-1} &= \frac{\sum_{h=1}^{2} \epsilon_{1,t-h}\epsilon_{2,t-h}}{\sqrt{(\sum_{h=1}^{2} \epsilon_{1,t-h}^2)(\sum_{h=1}^{2} \epsilon_{2,t-h}^2)}}. \end{aligned}$$

The correlation processes (1–4) exhibit constant, gradual changes and rapid changes and are also used by Engle (2002). DCC is the dynamic conditional correlation model of Engle (2002) where the time-varying correlation has three components with each component having an autoregressive moving average structure. TVC is the time varying correlation model of Tse and Tsui (2002) in which a strong dependence in the time-varying conditional correlations exist.

The S-GARCH model is fitted using M-estimators and the estimate of the correlations are obtained. In this study we use the QMLE, LAD and B-estimators for the estimation, estimators such as Cauchy and Huber's estimators can also be used, though. The performance of these estimators in estimating the in-sample correlations are measured using

$$\text{In-MAE}_{\rho} = \frac{1}{T}\sum_{t=1}^{T}\left|\hat{\rho}_t - \rho_t\right|,$$

where $\hat{\rho}_t$ and ρ_t are the estimated and true correlations, respectively. The average of these in-MAEs over 1000 replications are used as a performance measure.

Table 8.1 reports results of the average in-sample MAEs for each estimator. Entries in bold represent best performing estimators in terms of the MAE for correlations. When errors are generated from the standard normal distribution, the mean absolute errors produced by the QMLE under all correlation processes are found the least. For student-t distribution with 3 df, B-estimator shows excellent results. The MAEs of correlation estimates for this estimator are found better than the competing estimators for all correlation processes except the DCC process where the MAE of LAD is found smaller than the B-estimator but the difference is not significant. When errors have heavy tails and asymmetric, the B-estimator again outperforms other competing estimators and shows the least MAEs for nearly all correlation processes. The only occasion when LAD beats B-estimator is the DCC processes. This suggests that B-estimator can be considered as the best choice for in-sample correlation estimates when errors are non-normal.

Table 8.1: Mean Absolute Error for the In-sample Correlation Estimates for the S-GARCH Model.

	QMLE	LAD	B-estimator
Standard Normal			
Constant	**0.1010**	0.1027	0.1094
Sine	**0.1817**	0.1824	0.1865
Fast Sine	**0.2447**	0.2451	0.2481
Step	**0.1585**	0.1617	0.1663
DCC	**0.2530**	0.2539	0.2574
TVC	**0.1560**	0.1574	0.1639
Student-t(3)			
Constant	0.1783	0.1311	**0.1260**
Sine	0.2909	0.2655	**0.2635**
Fast Sine	0.3200	0.2972	**0.2950**
Step	0.2634	0.2388	**0.2339**
DCC	0.2884	**0.2611**	0.2613
TVC	0.2464	0.2114	**0.2091**
Skewed-t(4,0.25)			
Constant	0.1320	0.1167	**0.1165**
Sine	0.2665	0.2512	**0.2498**
Fast Sine	0.3002	0.2868	**0.2862**
Step	0.2392	0.2239	**0.2216**
DCC	0.2782	**0.2663**	0.2673
TVC	0.2167	0.1972	**0.1971**

The out-of-sample correlation forecasts of each estimator are also evaluated. The out-of-sample correlation forecasts are important for investors, portfolio and risk managers who want to evaluate the predictive performance of their models. At each replication, $K = 20$-step ahead correlation forecasts are produced using each M-estimator and the mean absolute error for out-of-sample correlation is calculated as

$$\text{Out-MAE}_\rho = \frac{1}{K}\sum_{k=1}^{K} \left|\hat{\rho}_{T+k} - \rho_{T+k}\right|.$$

The average over 1000 replications is calculated and the results are reported in Table 8.2. The out-of-sample correlation forecasts can be considered more important than the in-sample because of their use in decision making.

Table 8.2: Mean Absolute Error for the Out-of-sample Correlation Estimates for the S-GARCH Model.

	QMLE	LAD	B-estimator
Standard Normal			
Constant	**0.0689**	0.0732	0.0802
Sine	0.2465	0.2346	**0.2336**
Fast Sine	**0.2552**	0.2586	0.2602
Step	0.3417	**0.2989**	0.3178
DCC	0.2616	**0.2612**	0.2651
TVC	**0.1366**	0.1374	0.1433
Student-t(3)			
Constant	0.1616	0.0979	**0.0909**
Sine	0.3394	0.2367	**0.2069**
Fast Sine	0.3193	**0.2884**	0.2888
Step	0.4123	0.3054	**0.2800**
DCC	0.2911	0.2295	**0.2294**
TVC	0.2189	**0.1590**	0.1624
Skewed-t(4,0.25)			
Constant	0.1052	**0.0834**	0.0865
Sine	0.3144	0.2409	**0.2157**
Fast Sine	0.2844	**0.2660**	0.2679
Step	0.3753	0.3064	**0.2860**
DCC	0.2736	**0.2484**	0.2494
TVC	0.1883	**0.1552**	0.1601

For the standard normal errors, the results of the average out-of-sample MAEs for each estimator is as follows: the QMLE produces the least MAEs for constant correlation, fast sine and TVC processes with LAD showing good results for step and DCC processes whereas for sine correlation process, B-estimator outperforms others. Both LAD and B-estimator outperform the expected best estimator QMLE for normal errors on three correlation processes.

From the results of Table 8.1, we found that B-estimator was the best choice in terms of in-sample MAE of correlation estimates, this estimator provides very good results for out-of-sample correlation estimates as well. Only for fast sine and TVC correlation processes, LAD produces MAEs that are smaller than the B-estimator. On all other cases, B-estimator is the clear winner. For skewed-t distribution both LAD and B-estimator provide results better than the QMLE with LAD showing good results for constant, fast sine, DCC and TVC correlation processes and B-estimator for sine and step processes.

We used the MAE as the forecast error statistic to evaluate and compare forecast errors in correlation estimates. This error statistic has a shortcoming that the underlying loss function is symmetric. Brailsford and Faff (1996) suggested two different error statistics that account for the potential asymmetry in the loss function. Their Mean Mixed Error of Under-prediction (MME_U) penalizes under-predictions more heavily and defined as

$$\text{MME}_U = \frac{1}{K}\left(\sum_{k=1}^{K}\sqrt{|\hat{\rho}_{T+k} - \rho_{T+k}|}I(\hat{\rho}_{T+k} \leq \rho_{T+k}) + \sum_{k=1}^{K}|\hat{\rho}_{T+k} - \rho_{T+k}|I(\hat{\rho}_{T+k} > \rho_{T+k})\right).$$

It can be seen that the above statistic places a heavier weighting on under-predictions by taking the square roots of the absolute values of forecast errors. Similarly, the Mean Mixed Error of Over-prediction (MME_O) which penalizes over-predictions more heavily can be defined as

$$\text{MME}_O = \frac{1}{K}\left(\sum_{k=1}^{K}\sqrt{|\hat{\rho}_{T+k} - \rho_{T+k}|}I(\hat{\rho}_{T+k} > \rho_{T+k}) + \sum_{k=1}^{K}|\hat{\rho}_{T+k} - \rho_{T+k}|I(\hat{\rho}_{T+k} \leq \rho_{T+k})\right).$$

From the point of view of decision making, many investors and risk managers are sometime, interested to know how much their model under- or over-predicts the volatilities or correlations. These statistics based on asymmetric loss functions can answer their questions and help them decide accordingly.

Table 8.3 reports the results of asymmetrical statistical evaluation methods for under- and over-prediction for M-estimators under various correlation processes and error distributions. All experiments are repeated 1000 times and the average over these replications are reported. For sine and step correlation processes all estimators under-predict the correlations heavily with QMLE showing the largest values. The QMLE produces good results under normal errors. Both LAD and B-estimator have smaller mixed mean errors than the QMLE for non-normal distributions with B-estimator producing the best results overall.

Table 8.3: Mixed Mean Error of Under-Prediction and Over-prediction for the S-GARCH Model.

	QMLE		LAD		B-estimator	
	MME_U	MME_O	MME_U	MME_O	MME_U	MME_O
Standard Normal						
Constant	**0.1727**	**0.1503**	0.1791	0.1583	0.1865	0.1627
Sine	0.4727	0.2627	**0.4600**	**0.2471**	0.4648	0.2642
Fast Sine	**0.3648**	**0.3717**	0.3673	0.3775	0.3652	0.3780
Step	0.6251	0.4074	**0.5999**	**0.3787**	0.6118	0.3955
DCC	**0.3753**	**0.3692**	0.3759	0.3718	0.3800	0.3738
TVC	**0.2451**	**0.2411**	0.2468	0.2490	0.2553	0.2528
Student-t(3)						
Constant	0.2931	0.2589	0.1704	**0.1953**	**0.1556**	0.2044
Sine	0.5419	0.4071	0.4565	0.2649	**0.4112**	**0.2351**
Fast Sine	0.4280	0.4373	0.3713	**0.4034**	**0.3703**	0.4156
Step	0.5803	0.4438	0.4999	0.3092	**0.4721**	**0.2878**
DCC	0.4226	0.4140	**0.3384**	0.3398	0.3402	**0.3396**
TVC	0.3632	0.3551	0.2601	0.2742	**0.2598**	**0.2738**
Skewed-t(4,0.25)						
Constant	0.2272	0.2083	0.1641	**0.1882**	**0.1560**	0.1956
Sine	0.5199	0.3346	0.4512	0.2485	**0.4166**	**0.2316**
Fast Sine	0.3899	0.3942	0.3677	**0.3935**	**0.3669**	0.4034
Step	0.5810	0.4077	0.5162	0.3168	**0.5050**	**0.3075**
DCC	0.3898	0.3884	**0.3492**	**0.3588**	0.3506	0.3618
TVC	0.3087	0.3065	0.2496	**0.2687**	**0.2475**	0.2731

MME and MME are the mixed mean error of under- and over-prediction, respectively.

Based on the results of Monte Carlo experiment, it can be concluded that estimators such as LAD and B-estimator are good alternatives to the commonly-used QMLE for estimating the multivariate GARCH models. The results provide evidence that these estimators outperform the QMLE in terms of both in-sample and out-of-sample correlation forecasts. Using asymmetric loss functions we find that when errors are non-normal, the QMLE systematically under- and over-predicts correlation estimates more than other estimates. Overall, B-estimator shows results better then the competing estimators. Finally, the robust method used for the estimation of multivariate GARCH models is very easy to implement as it requires estimation of some univariate GARCH models.

Empirical Illustration

In this section the procedure outline above is applied to stock indices and currency exchange rates. The out-of-sample forecasts of the conditional correlations and value-at-risk are obtained. Various evaluation measures are used to assess the performance of competing estimators in predicting risk estimates.

Application to stock indices:

The data set used in this study are the daily log returns of SP500 index and the stocks of Cisco Systems and Intel Corporations from January 2, 1991 to December 31, 1999, with $T = 2275$ observations in each set. The same data set was used for empirical investigation of weighted linear estimator for multivariate ARCH models in previous chapter. Basic descriptive statistics for these data sets can be found in Section 7.2.3. The log returns are in percentages and we denote the return vector at time by $\mathbf{x}_t = (x_{1,t}, x_{2,t}, x_{3,t})'$, where $x_{1,t}$, $x_{2,t}$, and $x_{3,t}$, are the log returns at time t, of SP500, Cisco and Intel, respectively.

We divide the data set in two parts: initial $N = 1275$ observations are used for model estimation and the remaining $K = 1000$ observations are retained for out-of-sample forecast evaluation. Each data set over the in-sample period is estimated using the S-GARCH model. The univariate GARCH(1,1) model is fitted using M-estimators to each data sets and $\{\hat{\sigma}^2_{i,t}; 1 \leq t \leq N,\ i = 1, 2, 3\}$, estimates of conditional variances of $x_{i,t}$ are obtained.

For the conditional covariances, new series, $x_{+ij,t} = x_{i,t} + x_{j,t}$ and $x_{-ij,t} = x_{i,t} - x_{j,t}$, for $1 \leq i < j \leq 3$ are constructed. By fitting GARCH(1,1) models, conditional variances of these new series are obtained and these are used to get $\hat{\sigma}_{ij,t}$, estimates of conditional covariances between each pair of series.

Forecast Evaluation:

Accurate estimates and reliable forecasts of correlations are of paramount importance in risk management (see Skintzi et al. (2005), for the importance of correlation forecasting on risk management). The forecasting performance of correlation has not received significant attention in the literature.

To evaluate the predictive performance of M-estimators, the out-of-sample forecasts of correlation estimates between each pair of data sets are compared. The k-step ahead forecasts of variances are generated from (8.10) and then aggregated over the forecast horizon, K^* as

$$\hat{\sigma}^{2*}_{i,l} = \sum_{k=1}^{K} \hat{\sigma}^2_{i,N+k+K\ \cdot(l-1)}, \quad l = 1, ..., L,\ \ i = 1, 2, 3,$$

where $L = K/K^*$, is the total number of aggregated variances obtained from rolling window scheme. For this study we use $K^* = 5, 10$, and 20, these correspond to weekly, bi-weekly and monthly forecasts. Similarly, we obtain

$$\hat{\sigma}^*_{ij,l} = \sum_{k=1}^{K} \hat{\sigma}_{ij,N+k+K\ \cdot(l-1)},$$

the aggregated forecasts of covariances. Finally, forecasts of conditional correlations are obtained as

$$\hat{\rho}_{ij,l} = \frac{\hat{\sigma}^*_{ij,l}}{\hat{\sigma}^*_{i,l}\hat{\sigma}^*_{j,l}}, \quad l = 1, ..., L, \ \ 1 \leq i < j \leq 3.$$

Since the true correlations are not observable, we use realized correlations as a proxy of these unknown correlations. Andersen and Bollerslev (1998) showed that higher-frequency returns produce better approximations of realized volatilities than same-frequency returns. In the absence of high-frequency data, we construct realized correlation from daily asset returns over the forecast horizon as:

$$\rho_{ij,l} = \frac{\sum_{k=1}^{K} x_{i,N+k+K\ \cdot(l-1)}\ x_{j,N+k+K\ \cdot(l-1)}}{\sqrt{\sum_{k=1}^{K} x^2_{i,N+k+K\ \cdot(l-1)}}\sqrt{\sum_{k=1}^{K} x^2_{j,N+k+K\ \cdot(l-1)}}}, \quad l = 1, ..., L, \quad 1 \leq i < j \leq 3. \tag{8.13}$$

The mean absolute error (MAE) is used to evaluate the correlation forecasts of M-estimators where MAE is defined as

$$\text{MAE}_\rho = \frac{1}{L}\sum_{l=1}^{L}\left[\sum_{1\leq i<j\leq 3}\left|\hat{\rho}_{ij,l} - \rho_{ij,l}\right|\right],$$

where $\hat{\rho}_{ij,l}$ and $\rho_{ij,l}$ are the estimated and realized correlations, respectively.

In addition to the MAE_ρ, we evaluate the performance of our proposed method in predicting value-at-risk. We choose equally weighted portfolio weights and estimate the p% VaR for each forecast horizon (see Section 7.2.3 for the construction of equally weighted portfolio VaR). Some useful evaluation measures are used to assess these forecasts. In this study the average number of VaR violations, $\hat{p}$, Kupiec's likelihood ratio statistic for unconditional coverage, LR_{uc} (Kupiec, 1995), the Dynamic Quantile (DQ) test of Engle and Manganelli (2004) and the average quadratic loss function (AQL) of Lopez (1999) are used to measure the out-of-sample performance of competing M-estimators. See Section 5.3 for the explanation of these terms. The results are summarised in Table 8.4.

It can be seen from Table 8.4 that B-estimator produces the least MADs for correlation forecasts for 5-step and 10-step forecast horizons whereas for 20-step horizon, LAD shows better results. The MADs of QMLE are not found the least in any of the cases considered. This is consistent with our Monte Carlo results where LAD and B-estimator produced very good results for out-of-sample conditional correlation estimates. The average number of violations for all M-estimators are found reasonable and each estimator pass the LR_{uc} test for all forecast horizons. B-estimator passes the DQ test for high order dependence in VaR violations at 5 and 10-steps forecast horizons where as both the QMLE and LAD fail to pass this test. Finally, the AQLs of B-estimator are found smaller than the other two competing estimators across all three forecast horizons. These findings clearly suggest that B-estimator is the best choice among the competing estimators for the S-GARCH model.

Application to exchange rates:

In the second empirical example we analyse the currency exchange rates of the US Dollar (US$) against the British Pound (GBP), Euro (EUR), and Australian Dollar (AUD) from January 4, 2000 until December 31, 2007. The data sets are obtained from *http://www.ukforex.co.uk/*. Let $r_{i,t}$ $(i = 1, 2, 3)$, denotes the exchange rate for US$ against GBP, EUR and AUD, respectively, the log-exchange rates are defined as $x_{it} = (\ln r_{i,t} - \ln r_{i,t-1}) \times 100\%$. The data set consist of $T = 2169$ observations. Again initial $N = 1169$ observations are used for estimation and the remaining $K = 1000$ observations are left for forecasts evaluation.

First row of Figure 8.1 displays the raw exchange rates of US$ against three

Table 8.4: Out-of-sample forecasting performance of M-estimators using the S-GARCH Model for stock indices.

Forecast horizon	QMLE	LAD	B-estimator
5-step			
$\hat{p}$	0.0100	0.0100	0.0080
LR_{uc}	0.00	0.00	0.44
	[1.0000]	[1.0000]	[0.5102]
DQ Test	40.65**	40.10**	12.76
	[0.0000]	[0.0000]	[0.0781]
AQL	0.0737	0.0747	*0.0687*
MAE_ρ	0.8884	0.8530	**0.8469**
10-step			
$\hat{p}$	0.0100	0.0100	0.0080
LR_{uc}	0.00	0.00	0.44
	[1.0000]	[1.0000]	[0.5102]
DQ Test	41.49**	42.06**	13.16
	[0.0000]	[0.0000]	[0.0683]
AQL	0.0743	0.0772	*0.0707*
MAE_ρ	0.6926	0.6532	**0.6513**
20-step			
$\hat{p}$	0.0110	0.0150	0.0100
LR_{uc}	0.098	2.19	0.00
	[0.7544]	[0.1390]	[1.0000]
DQ Test	38.26**	37.52**	19.58**
	[0.0000]	[0.0000]	[0.0065]
AQL	0.0743	0.0843	*0.0731*
MAE_ρ	0.5889	**0.5468**	0.5484

currencies. It can be seen that all three graphs show same trend and hence it can be deduced that these exchange rates may be correlated with each other. The sample correlation estimates between log-exchange rates are found as $\rho_{12} = 0.6833$, $\rho_{13} = 0.4607$, and $\rho_{23} = 0.5497$, where 1, 2 and 3 are used for US$/GBP, US$/EUR, and US$/AUD, respectively. The log-exchange rates are also shown in the second row of the figure and the last row shows the histogram of daily log-exchange rates. A slight departure from normality can be observed from these histograms.

The results of MAE and other evaluation measures for exchange rate data are

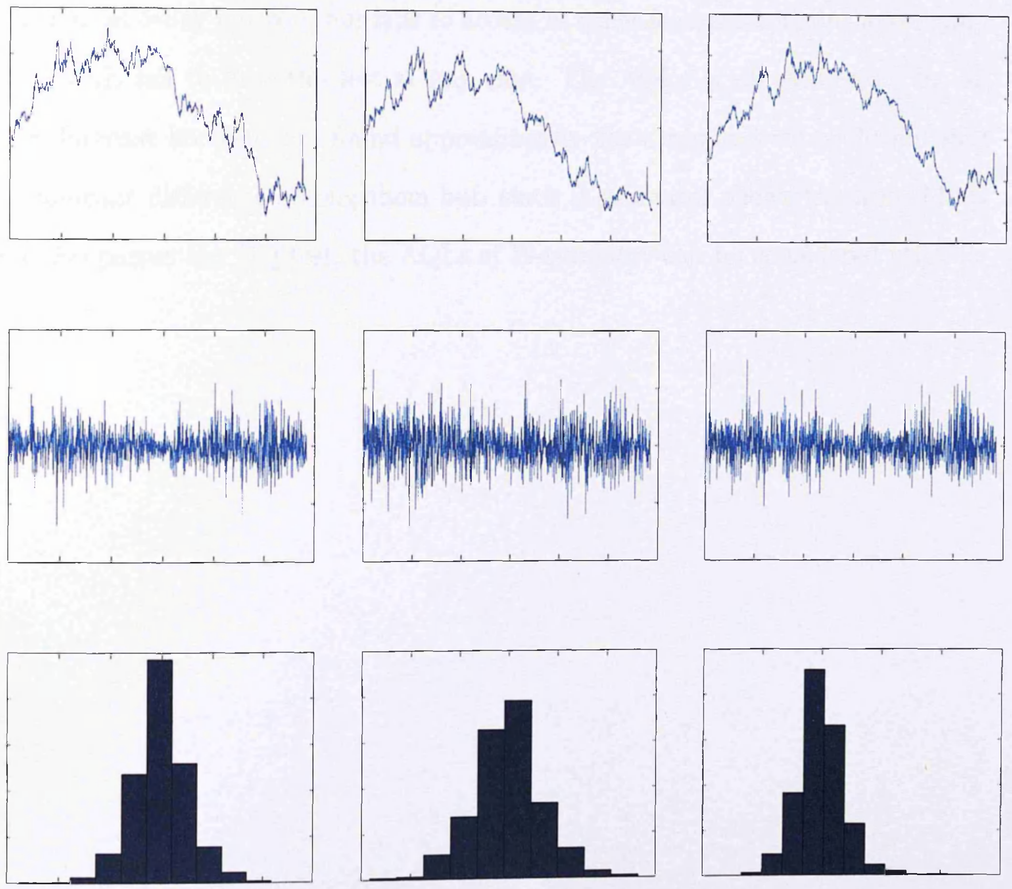

Figure 8.1: Daily exchange rates of US Dollars against GBP, EUR and AUD form January 4, 2000 to November 3, 2004. First row: daily exchange rates, second row: log-exchange rates, third row: histograms of log-exchange rates.

reported in Table 8.5. The mean absolute errors for correlation forecasts of B-estimator again outperform other estimator in two out of three forecast horizons. The LAD shows the least MAD for 10-day horizon correlation forecasts and QMLE fails to compete these estimator based on this performance measure. All estimators produce the same average VaR violations and hence pass the unconditional coverage test. B-estimator passes the dynamic quantile tests in first two cases but rejects this test at 20-step forecast horizon. LAD passes the DQ test at only one occasion, at 5-day horizon, but fails to accept at other forecast horizons and finally the QMLE fail to pass the test at any case. The AQLs of all estimators for all three forecast horizons are found approximately the same and we could not find a significant difference among them but, since B-estimator shows the least MAD and also passes the DQ test, the AQLs of B-estimator can be considered reliable.

Table 8.5: Out-of-sample forecasting performance of M-estimators using the S-GARCH Model for exchange rates.

Forecast horizon	QMLE	LAD	B-estimator
5-step			
$\hat{p}$	0.0110	0.0100	0.0100
LR_{uc}	0.10	0.00	0.00
	[0.7544]	[1.0000]	[1.0000]
DQ Test	16.43*	9.13	9.64
	[0.0215]	[0.2432]	[0.2102]
AQL	0.0111	0.0101	*0.0100*
MAE_ρ	0.7749	0.7546	**0.7265**
10-step			
$\hat{p}$	0.0120	0.0110	0.0110
LR_{uc}	0.38	0.10	0.10
	[0.5377]	[0.7544]	[0.7544]
DQ Test	37.96**	33.56**	34.15**
	[0.0000]	[0.0000]	[0.0000]
AQL	0.0121	0.0112	*0.0110*
MAE_ρ	0.5551	**0.5395**	0.5411
20-step			
$\hat{p}$	0.0100	0.0100	0.0100
LR_{uc}	0.00	0.00	0.00
	[1.0000]	[1.000]	[1.0000]
DQ Test	46.27**	37.12**	37.07**
	[0.0000]	[0.0000]	[0.0000]
AQL	*0.0102*	*0.0102*	*0.0102*
MAE_ρ	0.4486	0.4540	**0.4248**

8.3 A Robust Method for the Orthogonal GARCH Model

Engle et al. (1990) proposed the parameterisation of the conditional covariance matrix using the idea that co-movements of the stock returns are driven by a small number of common underlying variables, called factors. Ding (1994) first introduced the use of factor GARCH model using orthogonal factors and later Alexander (2000) suggested the construction of unconditionally uncorrelated linear combinations of the observed series based on principal component analysis.

It is well known that the correlations between returns make the modelling and estimation of the multivariate GARCH modeling more difficult. Therefore, the basic idea of the orthogonal model is that in the first step all unconditional correlations are removed by taking principal components of the standardised returns and some of the principal components are modelled by univariate GARCH models. In the second step, the inverse of the principal components construction is used to transform the conditional moments of the principal components into the conditional variance of the returns themselves. This approach makes the estimation of the multivariate GARCH models very easy as only some univariate GARCH estimations are required in the first step only.

In our model we use as many factors as the number of variables in the model. This solves the problem of choosing the correct number of factors, or equivalently, the number of principal components for the model. Since we are using only univariate GARCH models for each factors, use of all factors is not a serious computational issue. Also the use of all factors ensures the positive definiteness of the conditional covariance matrix.

More specifically, suppose there are M return series with T observations each. Let $x_{i,t}$ be the return of variable i at time t and $\mathbf{X}$ be the $T \times M$ matrix of these return series. Define the standardised return series as

$$y_{i,t} = \frac{(x_{i,t} - \mu_i)}{\sigma_i}, \quad i = 1, ..., M,$$

where μ_i and σ_i are the estimated mean and standard deviation for return series $x_{i,t}$. Denoted by $\mathbf{Y}$, the $T \times M$ matrix of standardised returns, the principal component matrix $\mathbf{P}$, can be defined as

$$\mathbf{P} = \mathbf{YW} \tag{8.14}$$

where $\mathbf{W} = [w_{ij}]$ for $i, j = 1, ..., M$, is the $M \times M$ orthogonal matrix of eigenvectors of $\mathbf{Y}'\mathbf{Y}$.

Since $\mathbf{W}$ is orthogonal, the principal component representation of the system can be obtained by inverting (8.14) as $\mathbf{Y} = \mathbf{PW}'$, that is

$$Y_i = w_{i1}P_1 + w_{i2}P_2 + ... + w_{iM}P_M, \tag{8.15}$$

where Y_i and P_i denote the i-th columns of $\mathbf{Y}$ and $\mathbf{P}$, respectively. Thus each data vector in $\mathbf{Y}$ is a linear combination of the principal components. In terms of the original variables $\mathbf{X}$ the representation (8.15) is equivalent to

$$X_i = \mu_i + w_{i1}^* P_1 + w_{i2}^* P_2 + ... + w_{iM}^* P_M, \tag{8.16}$$

where X_i denotes the i-th column of $\mathbf{X}$, and $w_{ij}^* = w_{ij}\sigma_i$.

The covariance matrix of $\mathbf{X}_t$ at time t, denoted by $\mathbf{H}_t$ can be obtained as

$$\mathbf{H}_t = \mathbf{W}^* \mathbf{D}_t \mathbf{W}^* \tag{8.17}$$

where $\mathbf{W}^* = [w^*_{ij}]$, for $i, j = 1, ..., M$, is the matrix of normalized factor weights and $\mathbf{D}_t$ is the diagonal matrix of the variances of the principal components at time t. Since $\mathbf{D}_t$ has positive diagonal elements, the variance-covariance matrix $\mathbf{H}_t$ is always positive definite when all principal components are used.

In order to get $\hat{\mathbf{H}}_t$, an estimate of the conditional variance-covariance matrix, only the eigenvectors of $\mathbf{Y}'\mathbf{Y}$, and the diagonal elements of $\mathbf{D}_t$ need to be estimated. These diagonal elements can be modelled, independently in a univariate setting, using a GARCH framework.

The GARCH model is proposed for the estimation of the variances of the principal components. Let $p_{i,t}$ $(i = 1, ..., M)$ be the elements of the i-th column of matrix $\mathbf{P}$, the conditional variances of the principal components are modelled as GARCH(1,1):

$$\sigma^2_{i,t} \;=\; \omega_{i,0} + \alpha_{i,0} p^2_{i,t-1} + \beta_{i,0} \sigma^2_{i,t-1}, \quad i = 1, ..., M, \quad t = 1, ..., T, \qquad (8.18)$$

where $\sigma^2_{i,t}$ is the conditional variance of $p_{i,t}$, and $\omega_i > 0$, $\alpha_i, \beta_i \geq 0$. Note that other GARCH-type models can also be used for the estimation of the variances of the principal components.

We propose estimating the parameters in (8.18) using M-estimators. Hence, we obtain $\hat{\sigma}^2_{i,t}$ $(i = 1, ..., M)$, estimates of the conditional variances, $\sigma^2_{i,t}$ $(i = 1, ..., M)$, at time t. The estimated variance-covariance matrix $\hat{\mathbf{H}}_t$ of return series $\mathbf{X}$ at time t are obtained using (8.17) with matrix $\mathbf{D}_t$ replaced by its estimate, $\hat{\mathbf{D}}_t = diag[\hat{\sigma}^2_{1,t}, ..., \hat{\sigma}^2_{M,t}]$.

The k-step ahead forecast from the O-GARCH models are generated as follows: From (8.17), we get

$$\hat{\mathbf{H}}_{T+k} = \mathbf{W}^*_{T+k} \hat{\mathbf{D}}_{T+k} \mathbf{W}^*_{T+k}, \qquad (8.19)$$

where $\hat{\mathbf{D}}_{T+k} = diag[\hat{\sigma}^2_{1,T+k}, ..., \hat{\sigma}^2_{M,T+k}]$, and $\mathbf{W}^*_{T+k}$ can be approximated by $\mathbf{W}^*_T$ without introducing large errors in the covariance matrix, since $\mathbf{W}$ does not change much from day to day. For the i-th principal component, the forecast at time $T+k$ can be generated from the GARCH(1,1) model as

$$\hat{\sigma}^2_{i,T+k} = \frac{\hat{\omega}_i(1-(\hat{\alpha}_i+\hat{\beta}_i)^{k-1})}{1-(\hat{\alpha}_i+\hat{\beta}_i)} + (\hat{\alpha}_i+\hat{\beta}_i)^{k-1}\hat{\sigma}^2_{T+1}, \quad k > 1,$$

where $\hat{\sigma}^2_{i,T+1} = \hat{\omega}_i + \hat{\alpha}_i p^2_{i,T} + \hat{\beta}_i\hat{\sigma}^2_{i,T}$.

The advantage of using the orthogonal GARCH specification is that only univariate GARCH models for all principal components need to be estimated to obtain the robust estimates for the full multivariate GARCH model. This reduces the computational complexity without imposing other constraints on the model. We suggest using all principal components in the first step of estimating the univariate GARCH models. This solves the problem of choosing the correct number of factors in the model. Also, as the matrix of normalized factor weights $\mathbf{W}^*$ has already been estimated in the first step, the second step, the inverse transformation (8.17), requires no further estimation.

8.3.1 Simulation and Empirical Results

Monte Carlo Experiments

Using the same data generating and correlation processes as defined in Section 8.2.1, we fit the O-GARCH model using the QMLE, LAD and B-estimators. The performances of these estimators are measured using the in-sample mean absolute error for correlation estimates. From (8.19), 20-steps ahead forecasts of variance-covariance matrix is generated and using this forecast matrix the out-of-sample mean absolute error for correlation estimates are obtained. All experiments are repeated 1000 times.

Table 8.6 reports the results of average in-sample MAE for correlation estimates over 1000 independent replications. When errors are generated from the standard normal distribution, the in-sample MAEs for conditional correlations of the QMLE show good results for all correlation processes except for fast sine correlation process. In this case LAD produces slightly better results. For student-t distribution with 3 df, B-estimators dominates over other two estimators. In this case B-estimator has the least MAD for constant, sine, fast sine and step correlation processes and LAD for DCC and TVC correlation processes. Mixed results are found for skewed-t distribution where both B-estimator and LAD produce the least MAD for three correlation cases each. For fast sine, DCC and TVC, LAD is considered as a better choice and for all other cases B-estimator show good results. Again QMLE is not found better than the competing estimators for heavy-tailed and heavy-tailed asymmetric distributions in terms of estimating the conditional correlations.

The out-of-sample results for estimating conditional correlations are reported in Table 8.7. Although the average out-of-sample MADs for LAD are found the least in two correlation cases, QMLE clearly perform well for normal errors. For heavy-tailed distributions for errors, both LAD and B-estimators beat the QMLE and produce the least MAD for three cases each. B-estimator show good results for constant, sine and DCC correlation process and LAD for others. LAD dominates the lower part of the table, where results of the out-of-sample MAEs for conditional correlations are displayed. For constant and sine correlation process B-estimator, and for other four cases LAD can be considered as the best choice. These results confirm that QMLE may not produce better in-sample and out-of-sample results for conditional correlations when errors are non-normal. Also, we conclude that

Table 8.6: Mean Absolute Error for the In-sample Correlation Estimates for the O-GARCH model.

	QMLE	LAD	B-estimator
Standard Normal			
Constant	**0.0096**	0.0099	0.0115
Sine	**0.1291**	0.1293	0.1313
Fast Sine	0.2248	**0.2233**	0.2244
Step	**0.0615**	0.0627	0.0636
DCC	**0.1415**	0.1419	0.1452
TVC	**0.1212**	0.1261	0.1258
Student-t(3)			
Constant	0.0279	0.0180	**0.0171**
Sine	0.1835	0.1704	**0.1699**
Fast Sine	0.2468	0.2393	**0.2390**
Step	0.1137	0.0991	**0.0965**
DCC	0.1738	**0.1696**	0.1721
TVC	0.1234	**0.1114**	0.1125
Skewed-t(4,0.25)			
Constant	0.0196	0.0147	**0.0145**
Sine	0.1625	0.1582	**0.1562**
Fast Sine	0.2384	**0.2343**	0.2365
Step	0.0923	0.0850	**0.0846**
DCC	0.1788	**0.1701**	0.1739
TVC	0.1100	**0.1037**	0.1073

for the Orthogonal GARCH model, LAD and B-estimator should be used for estimation and prediction when there is an evidence of departure from normality. Finally, LAD has a slight edge over B-estimator in terms of out-of-sample MAD for conditional correlations in O-GARCH models. The mixed mean errors of under- and over-prediction for each estimator is also calculated and results are displayed in Table 8.8. The results of LAD and B-estimator are found better than the QMLE in all cases of non-normal errors and also in few cases under normal errors. Another feature we notice from this table is that when there is a sudden change in correlation, as in the case of step correlation process, all estimators under-predicts

Table 8.7: Mean Absolute Error for the Out-of-sample Correlation Estimates for the O-GARCH Model.

	QMLE	LAD	B-estimator
Standard Normal			
Constant	**0.0060**	0.0096	0.0090
Sine	0.2758	**0.0728**	0.1205
Fast Sine	**0.2452**	0.2542	0.2545
Step	0.4796	**0.2806**	0.3962
DCC	**0.2507**	0.2526	0.2525
TVC	**0.1233**	0.1263	0.1267
Student-t(3)			
Constant	0.0294	0.0121	**0.0117**
Sine	0.4222	0.1495	**0.1318**
Fast Sine	0.2779	**0.2533**	0.2572
Step	0.4895	**0.2862**	0.3021
DCC	0.2379	0.2113	**0.2102**
TVC	0.1462	**0.1304**	0.1324
Skewed-t(4,0.25)			
Constant	0.0178	0.0123	**0.0115**
Sine	0.3666	0.1173	**0.1112**
Fast Sine	0.2630	**0.2538**	0.2587
Step	0.4723	**0.2629**	0.2952
DCC	0.2415	**0.2341**	0.2370
TVC	0.1339	**0.1293**	0.1316

the correlation heavily with QMLE producing the largest error. Also, for sine correlation process, the QMLE under-predicts the correlations more than other estimators.

Empirical Illustration

As an empirical application, the O-GARCH model is fitted to stock indices and currency exchange rates.

Application to stock indices:

The same data sets from the previous section, i.e. the daily log returns of SP500

Table 8.8: Mixed Mean Error of Under-Prediction and Over-prediction for the O-GARCH Model under Normal errors.

	QMLE		LAD		B-estimator	
	MME_U	MME_O	MME_U	MME_O	MME_U	MME_O
Standard Normal						
Constant	**0.0507**	**0.0273**	0.0727	0.0281	0.0628	0.0331
Sine	0.5096	0.2747	**0.2503**	**0.0762**	0.3339	0.1191
Fast Sine	0.3609	**0.3559**	0.3561	0.3781	**0.3530**	0.3799
Step	0.6886	0.4788	**0.4851**	**0.2848**	0.6219	0.3995
DCC	**0.3679**	**0.3629**	0.3783	0.3677	0.3690	0.3671
TVC	0.2308	**0.2175**	**0.2211**	0.2339	0.2258	0.2301
Student-t(3)						
Constant	0.0871	0.0444	0.0812	**0.0299**	**0.0750**	0.0365
Sine	0.6183	0.4196	0.3683	0.1461	**0.3382**	**0.1313**
Fast Sine	0.3859	0.3768	0.3599	**0.3743**	**0.3575**	0.3817
Step	0.6749	0.4798	**0.4716**	**0.2829**	0.4957	0.2986
DCC	0.3389	0.3333	**0.3245**	0.3182	0.3264	**0.3160**
TVC	0.2646	0.2504	0.2305	**0.2356**	**0.2279**	0.2408
Skewed-t(4,0.25)						
Constant	0.0885	0.0443	0.0834	**0.0298**	**0.0755**	0.0333
Sine	0.5866	0.3676	0.3261	0.1159	**0.3159**	**0.1130**
Fast Sine	0.3718	0.3657	0.3580	0.3271	**0.3551**	**0.3257**
Step	0.6717	0.4630	**0.4559**	**0.2672**	0.5018	0.2924
DCC	0.3476	0.3613	**0.3420**	0.3316	0.3423	**0.3487**
TVC	0.2418	0.3298	0.2212	**0.2377**	**0.2169**	0.2481

MME and MME are the mixed mean error of under- and over-prediction, respectively.

index and the stocks of Cisco Systems and Intel Corporations from January 2, 1991 to December 31, 1999, with $T = 2275$ observations, are used in this section. The principal components are obtained using the initial $N = 1275$ observations. The O-GARCH model is fitted to all of these principal components and k-step ahead forecasts for correlations are generated where $k = 1, ..., K^*$ and the forecast horizon K^* =5, 10, and 20-day. The out-of-sample performances of M-estimators are compared using various test statistics and performance measured from previous section.

Table 8.9 displays the results of out-of-sample forecasting performance of M-estimators using the O-GARCH model. First we look at the 5-step forecast horizon. It is found that all M-estimators provide reasonable estimates of average violations, $\hat{p}$ and pass the likelihood ratio test for unconditional coverage. The QMLE and LAD fail to pass the dynamic quantile test for high order independence in VaR violations. B-estimator passes this test although with a low p-value. The average quadratic loss and the mean absolute error for correlation forecasts of B-estimator are found the least. B-estimator also show better results than the other estimators in terms of AQL and MAE_ρ for 10 and 20-step forecast horizons. It passes the DQ test again at 10-day horizon where both QMLE and LAD fail to pass the test. For 20-step forecast horizon all estimator fail to pass the DQ test. In summary, B-estimator provides excellent results and outperform the other competing estimators when the O-GARCH model is fitted to stock indices data.

Application to exchange rates:
We fit the O-GARCH models to the exchange rate data of the US Dollar (US$) against the British Pound (GBP), Euro (EUR), and Australian Dollar (AUD) from January 4, 2000 until December 31, 2007. The out-of-sample results for M-estimators are reported in Table 8.10.

The results of Table 8.10 are summarised as follows: All M-estimators considered in this study provide good estimate of average number of VaR violation and pass the Kupiec's Likelihood ratio test. For this data set, none of the estimator fail to reject the DQ test for all three forecast horizons. This shows that the VaR violations produced by each estimator, when O-GARCH models are fitted, are not dependent even at higher lags. The AQLs of B-estimator are again found the least in all cases. At 20-day horizon, LAD shows better results for the out-of-sample mean absolute error of correlations but for 5 and 10-step forecast horizons, B-estimator performs well.

The results of our Monte-Carlo experiments and real data analysis show that

Table 8.9: Out-of-sample forecasting performance of M-estimators using the O-GARCH Model for stock indices.

Forecast horizon	QMLE	LAD	B-estimator
5-step			
$\hat{p}$	0.0090	0.0100	0.0080
LR_{uc}	0.10	0.00	0.44
	[0.7465]	[1.0000]	[0.5102]
DQ Test	45.46**	39.71**	12.63
	[0.0000]	[0.0000]	[0.0816]
AQL	0.0756	0.0773	*0.0722*
MAE_ρ	0.9222	0.9016	**0.8878**
10-step			
$\hat{p}$	0.0090	0.0100	0.0080
LR_{uc}	0.10	0.00	0.44
	[0.7465]	[1.0000]	[0.5102]
DQ Test	45.51**	39.99**	13.05
	[0.0000]	[0.0000]	[0.0708]
AQL	0.0800	0.0757	*0.0744*
MAE_ρ	0.6730	0.6546	**0.6470**
20-step			
$\hat{p}$	0.0100	0.0150	0.0100
LR_{uc}	0.00	2.19	0.00
	[1.0000]	[0.1390]	[1.0000]
DQ Test	52.21**	86.79**	19.23**
	[0.0000]	[0.0000]	[0.0075]
AQL	0.0789	0.0872	*0.0765*
MAE_ρ	0.5393	0.5216	**0.5073**

estimator such as LAD and especially B-estimator can produce better results than the frequently-used estimator, the QMLE. Both in-sample and out-of-sample results confirm that when errors are generated from a heavy-tailed distribution such as Student-t with 3 df or a heavy-tailed-skewed distribution like Hansen's Skewed-t distribution, the QMLE is not a right choice. Both LAD and B-estimator comprehensively outperformed the QMLE in terms of standard performance measures. The QMLE is clearly the worst performer under non-normal errors and these observations are confirmed by the two asymmetric error statistics. Applications to real

Table 8.10: Out-of-sample forecasting performance of M-estimators using the O-GARCH Model for exchange rates.

Forecast horizon	QMLE	LAD	B-estimator
5-step			
$\hat{p}$	0.0090	0.0100	0.0070
LR_{uc}	0.10	0.00	1.02
	[0.7544]	[1.0000]	[0.3136]
DQ Test	1.32	1.47	1.30
	[0.9878]	[0.9832]	[0.9884]
AQL	0.0103	0.0115	*0.0083*
MAE_{ρ}	0.7418	0.7415	**0.7300**
10-step			
$\hat{p}$	0.0090	0.0110	0.0080
LR_{uc}	0.10	0.10	0.44
	[0.7465]	[0.7544]	[0.5102]
DQ Test	2.15	2.60	1.28
	[0.9920]	[0.9187]	[0.9891]
AQL	0.0111	0.0133	*0.0102*
MAE_{ρ}	0.5163	0.5101	**0.5049**
20-step			
$\hat{p}$	0.0080	0.0110	0.0080
LR_{uc}	0.43	0.10	0.43
	[0.5102]	[0.7544]	[0.5102]
DQ Test	0.97	2.39	0.94
	[0.9953]	[0.9352]	[0.9958]
AQL	0.0112	0.0136	*0.0103*
MAE_{ρ}	0.4001	**0.3936**	0.3974

data sets further support these findings. B-estimator shows excellent results and stand out as the best estimator for the estimation and prediction of multivariate GARCH models.

8.4 Comparison of both models

In previous two sections we defined M-estimators for two multivariate GARCH models namely, the S-GARCH and the O-GARCH models. We investigated the performance of our estimators with commonly used estimation method, QMLE. Our results showed that LAD and B-estimators produce better results than the QMLE both in estimating and predicting value-at-risk and correlations between assets. In this section we compare both these multivariate GARCH models. More specifically, we discuss problem of estimation, positive definiteness of conditional variance-covariance matrix and also empirical results of our real data analysis.

First, the S-GARCH model not only involves estimation of univariate GARCH models for the individual return series but also for the sum and difference of each pair of series whereas the O-GARCH model requires estimation of univariate GARCH models for all principal components. For a M-dimensional problem, the S-GARCH involves estimation of $M \times M$ univariate GARCH models but the O-GARCH needs estimation of at most M models. This shows that as the number of variable increases the O-GARCH model can be preferred over S-GARCH as the former requires less computational time and also easy to manage.

Second, The conditional variance-covariance matrices in O-GARCH model is always positive definite by construction if all orthogonal factors are used but this is not guaranteed in S-GARCH model. Although few techniques, described in previous sections, to ensure the positive semi-definiteness of covariance matrix can be employed but these may introduce some error in estimation and prediction. Hence, we prefer the O-GARCH model with all orthogonal factors over the S-GARCH model. Note that the principal components are only unconditionally uncorrelated, so a covariance matrix of principal components is not necessarily diagonal and the assumption of zero conditional correlation has to be made (Alexander 2000).

Finally, we compare the out-of-sample performance of M-estimators in both the S-GARCH and O-GARCH models. By examining the results of Tables 8.4 and 8.9 (applications to stock indices), we find that both models produce more or less same results with B-estimators showing the best results for out-of-sample correlations in most of the cases and also producing the least AQL. For 5-step forecast horizon, the mean absolute error for correlations of S-GARCH is found less than those of O-GARCH but for 10 and 20-step horizons, the O-GARCH shows smaller values. For both models, B-estimators pass the DQ test at 5 and 10-step forecast horizons while other competing estimators fail to pass this test at all three horizons. When both models are applied to currency exchange rates, comparison of Tables 8.5 and 8.10 reveal some interesting results. All estimators pass the DQ test at all three forecast horizons when the O-GARCH model is fitted but when the S-GARCH model is fitted to the same data set, only LAD and B-estimator pass this test at only 5-step forecast horizon. This gives some evidence that the number of VaR violations produced by each estimator, when the O-GARCH model is fitted, are independent of each other even at higher lags. Also, the mean absolute errors for out-of-sample correlations are found smaller for O-GARCH model.

It can be concluded from the comparison of both these models that the O-GARCH has some advantages over the S-GARCH model. The former is easy to estimate, does not suffer from non-positive definite conditional covariance matrices when all orthogonal factors are used and predicts conditional correlations and risk estimates better.

8.5 Conclusion

Using univariate GARCH specification, two robust methods for the estimation of multivariate GARCH models are proposed. In the S-GARCH model four univariate GARCH models are estimated by M-estimators - one for each variable, one on the sum of the variables and one on the difference of the variables. The difference of the later two yields four times the covariance between the two variables. In the O-GARCH model the data are transformed into their principal components and for *all* components, univariate GARCH models are fitted using M-estimators. Our models are easy to estimate as these only requires several univariate GARCH estimation to estimate the full multivariate GARCH model without imposing extra restrictions on the model.

Monte Carlo simulations show that use of LAD and B-estimator provide better estimates than the QMLE and the out-of-sample performance of these estimators are also found superior. These estimators are applied to stock indices and currency exchange rates by focusing on predictions for conditional correlations and VaR. Our results suggest the use of robust estimators such as LAD and B-estimator for the estimation and prediction of multivariate GARCH models. Using different performance measures, we collect empirical evidence of the better predictive potential of B-estimators over other competing M-estimators.

We also discuss and compare both multivariate GARCH models and suggest using M-estimators for *all* components of the O-GARCH model as this model is easy to estimate, produces positive definite conditional covariance matrices and predicts better conditional correlations and risk estimates. Our approach is not restricted to GARCH model only, in fact, any univariate GARCH-type model can be used for the estimation of the multivariate model.

Chapter 9

Conclusions

Overview of the contributions

The contributions of this thesis on conditional heteroscedastic time series models are as follows:

A class of robust estimators for asymmetric GARCH models was proposed. We explored estimators such as Cauchy and B-estimator and our findings showed that these estimators not only outperform the commonly-used QMLE but also the robust LAD estimator when errors are non-normal. A weighted resampling method for GARCH models was discussed and bootstrapped M-estimators were used to approximate the sampling distribution of the parameters with this approach to resampling.

The asymptotic distributions of absolute and squared residual autocorrelations were obtained and new portmanteau tests were developed. These tests can be used to check whether or not a GARCH model fitted by using M-estimators is adequate. Our results are applicable under weak error moment assumptions. These diagnostic tests can help practitioners to use the correct statistics for checking the adequacy of GARCH-type models.

Some useful evaluation measure and M-tests were developed to assess the performance of M-estimators in predicting value-at-risk. Our results confirmed that Cauchy and B-estimator not only provide robust estimate for GARCH models but also provide reliable risk estimates. Based on our results we suggest the use of these robust estimators for estimation and prediction of GARCH-type models. Our findings can help risk managers to use these estimators to predict risk estimates with confidence. A comparison of results of both symmetric and asymmetric models revealed that for data sets analysed in the study, the asymmetric model provides better forecasts for risk.

Another contribution was a detailed investigation of linear estimators (LE) for ARCH models. Through extensive simulations and real data analysis we found that this estimator shows accurate results for parameter estimation and predicts the conditional volatility better than the QMLE. Moreover, using a weighted resampling approach, other bootstrap schemes were found that can match the widely-used paired and residual bootstrap. We also developed predictions intervals for returns, volatilities and VaR and observed that LE can be considered favourable. Based on our empirical findings we prefer LE to QMLE for ARCH models and suggest its use for computer intensive tasks.

Promising results for LE in univariate settings encouraged us to generalise this estimator to multivariate models. We proposed a weighted linear estimator (WLE) for the parameters of multivariate ARCH (MARCH) models and showed that this estimator estimates and forecasts MARCH models better than the competing QMLE. We generalised the weighted resampling to multivariate setup and showed the fast and easy way of approximating the sampling distribution of the parameters of MARCH model. Our method of estimating MARCH models using the WLE allows researchers to apply heavy computational methods such as bootstrapping in multivariate models in less time.

As our final and important contribution, we developed robust methods for the estimation of multivariate GARCH (MGARCH) models. We proposed M-estimators for MGARCH models using univariate GARCH specifications. Two different methods that are robust and easy to estimate were suggested for these models. Monte Carlo simulations and empirical analyses showed that B-estimator performs well in terms of estimating and predicting multivariate models. Our robust methods are easy to apply and do not impose extra restrictions on model. A comparison of both models showed that the orthogonal GARCH model with *all* orthogonal factors has an edge over the simplified GARCH model.

Possible directions for future research

Some of the interesting questions and extensions that emerge naturally in course of this research but are not addressed in this thesis and remain topics for future research and investigation are as follows:

Selection of M-estimator among competing estimators is not addressed in this research. This problem of when to use what can be solved by defining the goodness-of-fit statistics for each estimator that measure the distances between the empirical distribution of error and the uniform distribution on $(0, 1)$.

One important future work is to provide a package for an open-source software such as R. The package can be used for fitting GARCH-type models using a class of M-estimators. This package will also include other features such as diagnostic testing, bootstrapping, forecasting value-at-risk and multivariate GARCH modelling using M-estimators, to name a few.

Using weighted resampling, a higher order analysis and consequent demonstration of any inaccuracy of the paired-bootstrap under the asymmetric GARCH models such as the GJR model is another interesting area of research.

Another approach to estimate the value-at-risk and related risk measures is to combine M-estimation and extreme value theory (EVT). In this approach the GARCH-type models are fitted using M-estimators to obtain the estimate of the conditional volatility and EVT is used to estimate the tail of the distribution of the residuals. We believe that use of robust estimators for the conditional volatility and EVT for the tail of the innovation distribution will provide more reliable risk estimates.

We used M-estimation of GARCH(1,1) specification for modelling multivariate GARCH models and checked the performance of M-estimator in predicting the conditional correlations and risk estimates. It will be interesting to use asymmetric models in multivariate settings and compare the predictive performance of both symmetric and asymmetric GARCH models.

An important extension is to develop M-estimators for MGARCH models such as CCC, DCC, VEC, and BEKK GARCH models, among others. Moreover, if M-estimators of these MGARCH models are developed, the next research is to compare these robust models with those we proposed. The comparison should be on out-of-sample value-at-risk and correlations predictions as in-sample comparisons hold little if any information.

Outlier correction in MGARCH model is another area that needs to be investigated. Using univariate GARCH specification, methods of detecting and correcting outliers can be developed.

Not much work has been done in bootstrapping multivariate GARCH models due to high computational cost. Using univariate specifications, bootstrap methods in multivariate setup can be developed.

Bibliography

[1] Alexander, C. O. (2000). A primer on the orthogonal GARCH model. ISMA Centre Discussion Paper in Finance, University of Reading, UK.

[2] Alexander, C. O. (2000b). Orthogonal Methods for Generating Large Positive Semi-Definite Covariance Matrices. Discussion Paper in Finance, University of Reading, UK.

[3] Alexander, C. O., and Chibumba, A. M. (1996). Multivariate orthogonal factor GARCH. Discussion Papers in Mathematics, University of Sussex, UK.

[4] Andersen, T. G., and Bollerlsev, T. (1998). Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review,* **39** (4), 885–905.

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

[6] Angelidis, T., Benos, A., and Degiannakis, S. A. (2004). The use of GARCH models in VaR estimation. *Statistical Methodology*, **1** (2), 105–128.

[7] Baillie, R. T., and Bollerslev, T. (1992). Prediction in dynamic models with time-dependent conditional variances. *Journal of Econometrics*, **52** (1-2), 91–113.

[8] Basel Committee on Banking Supervision.: Supervisory Framework for the Use of Backtesting in Conjunction with the Internal Model-Based Approach to Market Risk Capital Requirements. *BIS*, Basel, Switzerland, (1996).

[9] Bauwens, L., Laurent, S., and Rombouts, J. V. K. (2006). Multivariate GARCH models: A survey. *Journal of Applied Econometrics*, **21** (1), 79-109.

[10] Beran, R. (1990). Calibrating prediction regions. *Journal of the American Statistical Association*, **85**, 715–723.

[11] Berkes, I., and Horváth, L. (2004). The efficiency of the estimators of the parameters in GARCH processes. *Annals of Statistics*, **32** (2), 633–655.

[12] Berkes, I., Horváth, L., and Kokoszka, P. (2003). GARCH processes: structure and estimation. *Bernoulli,* **9** (2), 201–227.

[13] Berkes, I., Horváth, L., and Kokoszka, P. (2003b). Asymptotics for GARCH squared residual correlations. *Econometric Theory*, **19**, 515–540.

[14] Berndt, E. K., Hall, B. H., Hall, R. E., and Hausman, J. A. (1974). Estimation and Inference in Nonlinear Structural Models. *Annals of Economic and Social Measurement*, **3/4**, 653–665.

[15] Billingsley, P. (1961). The Lindeberg-Levy theorem for martingales. *Proceedings of the American Mathematical Society*, **12**, 788–792.

[16] Black, F. (1976). Studies in Stock Price Volatility Changes, in *Proceedings of the 1976 Meeting of the Business and Economic Statistics Section, American Statistical Association*, 177–181.

[17] Bollerslev, T. (1986). Generalised autoregressive conditional heteroscedasticity. *Journal of Econometrics*, **31** (3), 307–327.

[18] Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *Review of Economics and Statistics*, **72** (3), 498–505.

[19] 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** (1-2), 5–59.

[20] Bollerslev, T., Engle, R. F., and Nelson, D. (1994). ARCH models. In: Engle, R., McFadden, D. (Eds.), *Handbook of Econometrics*, vol. **4**. North-Holland, Amsterdam, 2959–3038.

[21] Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988). A Capital Asset Pricing Model With Time-Varying Covariances. *Journal of Political Economy*, **96** (1), 116-131.

[22] Bose, A., and Mukherjee, K. (2003). Estimating the ARCH parameters by solving linear equations. *Journal of Time Series Analysis*, **24** (2), 127–136.

[23] Bose, A., and Mukherjee, K. (2009). Bootstrapping a Weighted Linear Estimator of the ARCH Parameters. *Journal of Time Series Analysis*, **30** (3), 315–331.

[24] Bougerol, P., and Picard, N. (1992). Stationarity of GARCH processe and of some nonnegative time series. *Journal of Econometrics*, **52** (1-2), 115–127.

[25] Box, G. E. P., and Jenkins, G. M. (1970). Time Series Analysis-Forecasting and Control. Holden-Day, San Francisco.

[26] Box, G. E. P., and Pierce, D. (1970). Distribution of the residual autocorrelations in autoregressive integrated moving average time series models. *Journal of the American Statistical Association*, **65**, 1509-1526.

[27] Brailsford, T. J., and Faff, R. W. (1996). An evaluation of volatility forecasting technique. *Journal of Banking and Finance*, **20**, 419–438.

[28] Cecchetti, S. G., Cumby, R. E., and Figlewski, S. (1988). Estimation of the optimal futures hedge. *Review of Economics and Statistics*, **70** (4), 623–630.

[29] Chatterjee, S., and Bose, A. (2005). Generalized bootstrap for estimating equations. *The Annals of Statistics*, **33** (1), 414–436.

[30] Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economics Review*, **39** (4), 841–862.

[31] Christoffersen, P. F., and Concalves, S. (2005). Estimation Risk in Financial Risk Management. *Journal of Risk*, **7** (3), 1–28.

[32] Chuang, I-Y., Lu, J-R., and Lee, P-H. (2007). Forecasting volatility in the financial markets: a comparison of alternative distributional assumption. *Applied Financial Economics*, **17** (13), 1051–1060.

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

[34] Ding, Z. (1994). Time Series Analysis of Speculative Returns. PhD Thesis, UCSD.

[35] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, **7** (1), 1-26.

[36] Efron, B., and Tibshirani, R. J. (1993). An Introduction to the Bootstrap. Chapman & Hall, New York.

[37] Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of U.K. Inflation. *Econometrica*, **50**, 987–1008.

[38] Engle, R. F. (2002). Dynamic Conditional Correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics*, **20** (3), 339-350.

[39] Engle, R. F., and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, **5** (1), 1–50.

[40] Engle, R. F., and Kroner, K. F. (1995): Multivariate simultaneous generalized ARCH. *Econometric Theory*, **11**, 122-150.

[41] Engle, R. F., and Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics*, **22**, 367–381.

[42] Engle, R. F., and Ng, V. K. (1991). Measuring and Testing the Impact of News on Volatility. *Journal of Finance*, **48**, 1749–1778.

[43] Engle, R. F., Ng, V. K., and Rothschild, M. (1990). Asset pricing with a factor-ARCH covariance structure: empirical estimates for treasury bills. *Journal of Econometrics*, **45** (1/2), 213-238.

[44] Engle, R. F., and Sheppard, K. (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. National Bureau of Economic Research Working Paper 8554.

[45] Fama, E. F. (1965). The behavior of stock market prices. *Journal of Business*, **38** (1), 34–105.

[46] Giraitis, L., Kokoszka, P., and Leipus, R. (2000). Stationary ARCH Models: Dependence Structure and Central Limit Theorem. *Econometric Theory*, **16**, 3-22.

[47] Glosten, L., Jagannathan, R., and Runkle, D. (1993). On the relation between the expected value and the volatility on the nominal excess return on stocks. *Journal of Finance*, **48** (5), 1779–1801.

[48] Hall, P., and Yao, Q. (2003). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica*, **71** (1), 285–317.

[49] Hansen, B. E. (1994). Autoregressive Conditional Density Estimation. *International Economics Review*, **35** (3), 705–730.

[50] Hansen, P. R., and Lunde, A. (2005). A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)? *Journal of Applied Econometrics*, **20** (7), 873–889.

[51] Harris, R. D. F., Stoja, E., and Tucker, J. (2007). A simplified approach to modelling the comovement of asset returns. *Journal of Future Markets*, **27** (6), 575–598.

[52] He, C., and Teräsvirta, T. (2004). An extended constant conditional correlation GARCH model and its fourth-moment structure. *Econometric Theory*, **20** (5), 904–926.

[53] Hendrick, D. (1996). Evaluation of Value-at-Risk Models using Historical Data. *Economic Policy Review*, **2** (1), 39–69.

[54] Higham, N. (2002): Computing the nearest correlation matrix - A problem from finance. *IMA Journal of Numerical Analysis*, **22**, 329–343.

[55] Huber, P. J. (1964). Robust estimation of location parameter. *The Annals of Mathematical Statistics*, **35** (1), 73–101.

[56] Huber, P. J. (1981). Robust Statistics. John Wiley, New York

[57] Hosking, J. R. M. (1980). The multivariate portmanteau statistics. *Journal of the American Statistical Association*, **75**, 602–608.

[58] Jeantheau, T. (1998): Strong consistency of estimators of multivariate ARCH models. *Econometric Theory,* **14**, 70–86.

[59] Jorion, P. (2000). Valuc-at-Risk: The New Benchmark for Managing Financial Risk. McGraw-Hill, New York.

[60] Jurečková, J., and Sen, P. K. (1996). Robust Statistical Inferences: Asymptotics and Interrelations. John Wiley, New York.

[61] Keuster, K., Mittinik, S., and Paolella, M. S. (2006). Value-at-Risk Prediction: A Comparison of Alternative Strategies. *Journal of Financial Econometrics*, **4** (1), 53–89.

[62] Kroner, K. F., Claessens, S. (1991). Optimal dynamic hedging portfolios and the currency composition of external debt. *Journal of International Money and Finance*, **10** (1), 131–148.

[63] Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, **3** (2), 73–84.

[64] Kwan, A. C. C., Sim, A-B., and Wu, Y. (2005). A comparative study of the finite-sample performance of some portmanteau tests of randomness of a time series. *Computational Statistics & Data Analysis*, **48**, 391–413.

[65] Ledoit, O., Santa-Clara P., and Wolf, M. (2003). Flexible multivariate GARCH modeling with an application to international stock markets. *The Review of Economics and Statistics,* **85** (3), 735-747.

[66] Lee, S., and Hansen, B. (1994). Asymptotic theory of the GARCH(1, 1) quasi-maximum likelihood estimator. *Econometric Theory,* **10**, 29–52.

[67] Li, G., and Li, W. K. (2005). Diagnostic checking for time series models with conditional heteroscedasticity estimated by the least absolute deviation approach. *Biometrika*, **92** (3), 691–701.

[68] Li, W. K., Ling, S., and McAleer, M. (2002). Recent theoretical results for time series models with GARCH errors. *Journal of Economic Surveys*, **16** (3), 245–269.

[69] 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** (6), 627–636.

[70] Lien, D., and Tse, Y. K. (1998). Hedging time-varying downside risk. *Journal of Futures Markets*, **18**, 705–722.

[71] Ling, S., and Li, W. K. (1997). On Fractionally Integrated Autoregressive Moving-Average Time Series Models with Conditional Heteroscedasticity. *Journal of the American Statistical Association*, **92** (439), 1184-1194

[72] Ling, S. and Li, W. K. (1997b). Diagnostic checking of nonlinear multivariate time series with multivariate ARCH errors. *Journal of Time Series Analysis*, **18** (5), 447–464.

[73] Ling, S., and McAleer, M. (2002). Stationarity and the existence of moments of a family of GARCH processes. *Journal of Econometrics*, **106** (1), 109–117.

[74] Ling, S., and McAleer, M. (2002b). Necessary and Sufficient Moment Conditions for the GARCH(r, s) and Asymmetric Power GARCH(r, s) Models. *Econometric Theroy*, **18** (3), 722–729.

[75] Ljung, G. M., and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, **65**, 297–303.

[76] Lopez, J. A. (1999). Methods for evaluating Value-at-Risk estimates. *Economic Review*, 3-17.

[77] Lumsdaine, R. L. (1996). Consistency and Asymptotic Normality of the Quasi-maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models. *Econometrica*, **64** (3), 575–596.

[78] Mak, T. K., Wong, H., and Li, W. K. (1997). Estimation of nonlinear time series with conditional heteroscedastic variances by iteratively weighted least squares. *Computational Statistics & Data Analysis*, **24** (2), 169–178.

[79] Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *Journal of Business*, University of Chicago Press, **36**, 394.

[80] McLeod, A. I., and Li, W. K. (1983). Diagnostic checking ARMA time series models using squared residual autocorrelations. *Journal of Time Series Analysis*, **4** (4), 269–273.

[81] McNeil, A. J., and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance* **7** (3-4), 271–300.

[82] Miguel, J. A., and Olave, P. (1999). Bootstrapping forecast intervals in ARCH models. *Test,* **8** (2), 345-364.

[83] Mousazadeh, K., and Karimi, M. (2009). Estimating multivariate ARCH parameters by two-stage least-squares method. *Signal Processing*, **89**, 921–932.

[84] Mukherjee, K. (2008). M-estimation in GARCH model. *Econometric Theory*, **24**, 1530–1553.

[85] Muler, N., and Yohai, Y. J. (2008). Robust estimates for GARCH models. *Journal of Statistical Planning and Inference*, **138** (10), 2918–2940.

[86] Nelson, D. B. (1990). Stationarity and Persistence in the GARCH(1,1) Model. *Econometric Theory*, **6** (3), 318–334.

[87] Nelson, D. B. (1991). Conditional Heteroskedasticity in Assest Returns: A New Approach. *Econometrica*, **59** (2), 347–370.

[88] Park, T. H., and Switzer, L. N. (1995). Bivariate GARCH estimation of the optimal hedge ratios for stock index futures: a note. *Journal of Futures Markets*, **15**, 61–67.

[89] Pascual, L., Romo, J., and Ruiz, E. (2004). Bootstrap predictive inference for ARIMA processes. *Journal of Time Series Analysis*, **25** (4), 449–465.

[90] 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.

[91] Peng, L., and Yao, Q. (2003). Least absolute deviations estimation for ARCH and GARCH models. *Biometrika*, **90** (4), 967–975.

[92] Poon, S-H., and Granger C. W. J. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, **41** (2), 478–539.

[93] Reevs, J. J. (2005). Bootstrap prediction intervals for ARCH models. *International Journal of Forecasting*, **21** (2), 237–248.

[94] RiskMetrics (1996). Technical Document. *Morgan Guaranty Trust Company of New York.*

[95] Robinson, P. M., and Zaffaroni, P. (2006). Pseudo-maximum likelihood estimation of ARCH(∞) models. *Annals of Statistics*, **34** (3), 1049–1074.

[96] Sarma, M., Thomas, S., and Shah, A. (2003). Selection of Value-at-Risk Models. *Journal of Forecasting*, **22** (4), 337–358.

[97] Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In D. R. Cox, D. V. Hinkley, and O. E. Barndorff-Nielsen (Eds.). *Time Series Models in Econometrics, Finance and other Fields.* Chapman & Hall, London, 1–67.

[98] Skintzi, V., Skiadopoulos, G., and Refenes, A. P. (2005). The effect of misestimating correlation on value-at-risk. *Journal of Alternative Investments,* **4**, 66–82.

[99] Straumann, D., and Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *Annals of Statistics*, **34** (5), 2449–2495.

[100] Thombs, L. A., and Schucany, W. R. (1990). Bootstrap prediction intervals for autoregression. *Journal of the American Statistical Association*, **85**, 486–492.

[101] Tsay, R. S. (2005). Analysis of Financial Time Series. 2nd Edition. Wiley. New York.

[102] Tse, Y. K., and Tsui, A. K. C. (1999). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business and Economic Statistics*, **20** (3), 351-362.

[103] Tse, Y. K., and Zuo, X. L. (1997). Testing for conditional heteroscedasticity: some monte carlo results. *Journal of Statistical Computation and Simulation*, **58** (3), 237–253.

[104] Tsui, A. K. C. (2004). Diagnostic for conditional heteroscedastic models: some simulation results. *Mathematics and Computers in Simulation*, **64**, 113–119.

[105] van der Weide, R. (2002). GO-GARCH: A multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics*, **17**, 549-564.

[106] Vrontos, I. D., Dellaportas, P., and Politis, D. N. (2003). A full-factor multivariate GARCH model. *Econometrics Journal*, **6**, 312-334.

[107] Weiss, A. A. (1986). Asymptotic Theory for ARCH models: Estimation and testing. *Econometric Theory*, **2**, 107–131.

[108] Wong, H., and Li, W. K. (1995). Portmanteau test for conditional heteroscedasticity, using ranks of squared residuals. *Journal of Applied Statistics*, **22** (1), 121–134.

[109] Zakoian, J. M. (1994). Threshold Heteroskedastic Models. *Journal of Economic Dynamics and Control*, **18** (5), 931–955.