FINAL REPORT OF MINOR RESEARCH PROJECT

(*UGC letter No. MRP(S)-0243/12-13/KLKE005/UGC-SWRO dated 27/03/2014*)

SHOCK DETECTION IN INDIAN STOCK SENSEX BY CONDITIONAL HETEROSCEDASTIC MODELS

Principal Investigator

ANJANA V

Assistant Professor in Statistics

Sree Narayana College, Chempazhanthy

Thiruvananthapuram-695 587, Kerala

FINAL REPORT OF A MINOR RESEARCH PROJECT SUBMITTED TO

UNIVERSITY GRANTS COMMISSION

SOUTH WESTERN REGIONAL OFFICE

BANGALORE

2017

ACKNOWLEDGEMENTS

I take this opportunity to express my gratitude to University Grants Commission, India for providing me the financial assistance in the form of Minor Research Project to complete the present work.

I am grateful to Dr. L. Thulaseedharan, Principal, Sree Narayana College, Chempazhanthy, Smt. Dhanya S. R., Smt. Bijila B. R., Assistant Professors in Mathematics and other colleagues of my college for their encouragement and constant support for completing this work.

I express my sincere thanks to Dr. P. Yageen Thomas, UGC-Emeritus Scientist, University of Kerala, Thiruvananthapuram and Dr. N. Balakrishna, Professor, Department of Statistics, Cochin University of Science and Technology, Ernakulam for their valuable advices and suggestions during the work.

ANJANA V

Assistant Professor in Statistics

Sree Narayana College

Chempazhanthy, Thiruvananthapuram

Chempazhanthy

28-05-2017

CONTENTS

CHAPTER 1
INTRODUCTION

A time series is a sequence of measurements of some variable collected over a time period. Most often measurements are made at regular time intervals. Time series data often arise when monitoring the industrial processes or tracking corporate business metrics. One difference from standard linear regression is that the data are not necessarily independent and not necessarily identically distributed. One defining characteristic of time series is that this is a list of observations where the ordering matters. Ordering is very important because there is dependency and changing the order could change the meaning of the data.

Alternately, a time series is a set of observations generated sequentially in time. A statistical phenomenon that evolves in time according to probabilistic laws is called a stochastic process. So by analyzing a time series one can realize a stochastic process.

1.1 Stationary Stochastic Process

A class of stochastic process is called stationary if the process is in a state of statistical equilibrium. It is called strictly stationary if its properties are unaffected by a change of time origin. Thus the joint distribution of any set of observations must be unaffected by shifting all the times of observation forward and backward by an integer.

1.1.1 Autocorrelation Function

The stationary assumption implies that the joint probability distribution $P(z_{t1}, z_{t2})$ is same for all times (t1, t2). The covariance between z_t and z_{t+k} separated by k intervals of time is called autocovariance at lag k and is defined by

$$\gamma_k = Cov(z_t, z_{t+k}) = E[(z_t - \mu)(z_{t+k} - \mu)] \tag{1.1}$$

Similarly autocorrelation at lag k is

$$\rho_k = \frac{E[(z_t - \mu)(z_{t+k} - \mu)]}{\sqrt{E(z_t - \mu)^2 \, \mathrm{E}(z_{t+k} - \mu)^2}} \tag{1.2}$$

For a stationary process, autocorrelation is

$$\rho_k = \frac{\gamma_k}{\gamma_0} \tag{1.3}$$

Both autocovariance matrix and autocorrelation matrix are positive definite for any stationary process. In particular for n=2, the conditions satisfied by the autocorrelations of a stationary process is

$$-1 < \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} < 1 \tag{1.4}$$

The autocorrelations considered as a function of k are referred to as the autocorrelation function (ACF) or correlogram. The ACF plays a major role in modelling the dependencies among observations of a stationary stochastic process, since ACF characterizes the process together with the process mean E(z) and variance γ_0 equal to V(z).

The correlation between two random variables is often due only to the fact that both variables are correlated with the same third variable. In the time series context, a large portion of the correlation between z_t and z_{t-k} can be due to the correlation of the intermittent variables z_{t-1}, $z_{t-2}, \ldots, z_{t-k-1}$. To adjust for this correlation, one can calculate partial autocorrelations.

1.2 Stochastic Difference Equation Models

The models we consider in this section are based on the observation by Yule (1921, 1927), where he noticed that a time series in which successive values are autocorrelated can be represented as a linear combination of a sequence of uncorrelated random variables. This representation was later confirmed by Wold (1938), who showed that every weakly stationary non-deterministic stochastic process (z_t-μ) can be written as a linear combination of a sequence of uncorrelated random variables. The linear combination is given by

$$z_t - \mu = a_t + \psi_1 a_{t-1} + ... = \sum_{j=0}^{\infty} \psi_j a_{t-j}, where\ \psi_0 = 1. \tag{1.5}$$

The random variables a_t are a sequence of uncorrelated random variables from a fixed distribution with mean $E(a_t) = 0$, variance $V(a_t) = E(a_t^2) = \sigma^2$ and $Cov(a_t, a_{t-k}) = E(a_t, a_{t-k}) = 0$ for all $k \neq 0$. Such a sequence is usually referred to as a white-noise process. Occasionally we will also call these random variables as random shocks. The ψ_j weights are the coefficients in this linear combination and their number can be either finite or infinite.

1.2.1 Autoregressive Process

In autoregressive process of order p [AR(p)] model, treat z_t, the deviation from the mean at time t as being regressed on the p previous deviations $z_{t-1}, z_{t-2,\ldots}, z_{t-p.}$

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + ... + \phi_p z_{t-p} + a_t \tag{1.6}$$

$$\phi(B) z_t = \theta(B) a_t \tag{1.7}$$

where $\phi(B) = 1 - \phi_1\, B - \phi_2 B^2 - ... - \phi_p B^p$ and B is the backshift operator, $B(z_t)=z_{t-1}$

1.2.2 Moving Average Process

Another class of stochastic model is obtained by specifying only ψ weights.

$$z_t - \phi_1 z_{t-1} - \phi_2 z_{t-2} - ... - \phi_p z_{t-p} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - ... - \theta_q a_{t-q} \tag{1.8}$$

$$z_t = \theta(B) a_t, \tag{1.9}$$

where $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q$

1.2.3 Autoregressive Moving Average Process

The Autoregressive Moving Average model is

$$z_t - \phi_1 z_{t-1} - \phi_2 z_{t-2} - ... - \phi_p z_{t-p} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - ... - \theta_q a_{t-q} \tag{1.10}$$

$$\phi(B) z_t = \theta(B) a_t \tag{1.11}$$

1.3 Non-Stationary Process : ARIMA

Let z_t be non-stationary series, but it should be transformed into a stationary process by considering relevant differences, i.e., $(1\text{-}B)^d z_t$. Then the autoregressive integrated moving average model of order (p, d, q) denoted by ARIMA(p, d, q) is defined as

$$\phi(B)(1-B)^d z_t = \theta(B)a_t \tag{1.12}$$

The model is called "integrated", since z_t can be thought of as the summation of a stationary series.

1.4 Financial Time Series

Financial time series analysis is concerned with the theory and practice of asset valuation over time. It is a highly empirical discipline, but like other scientific fields theory forms the foundation for making inference. Volatility is an important factor in options trading. Here volatility means the conditional standard deviation of the underlying asset return. Financial markets are an important channel in the transmission mechanism of monetary policy. Developments in financial markets reflect the expectations about future economic and financial developments which might have an influence on monetary policy decisions and vice-versa. Two empirical characteristics commonly found in financial return series are the presence of high excess kurtosis and persistence in volatility. These characteristics led to the development of the ARCH model of Engle (1982) and GARCH model of Bollerslev (1986). Many authors have argued that the high persistence in volatility of empirical asset returns is due to sudden shocks to the series associated with extraordinary economic events such as financial crises, recessions and changes in policy. These shocks may well be described by structural breaks in the GARCH model governing the conditional volatility of the return series. A special feature of stock volatility is that it is not directly observable.

1.4.1 The ARCH Model

The first model that provides a systematic framework for volatility modeling is the ARCH model of Engle (1982). The basic idea of ARCH models is that (a) the shock of an asset return is serially uncorrelated, but dependent and (b) the dependence of a_t can be described by a simple quadratic function of its lagged values. Specifically, an ARCH (m) model assumes that

$$a_t=\sigma_t\epsilon_t, \text{ where } \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \ldots + \alpha_m a_{t-m}^2 \tag{1.13}$$

where $\{\epsilon_t\}$ is a sequence of independent and identically distributed random variables with mean zero and variance 1, $\alpha_0 > 0$ and $\alpha_i \geq 0$ for $i > 0$. The coefficients α_i must satisfy some regularity conditions to ensure that the unconditional variance of it is finite. In practice, ϵ_t is often assumed to follow the standard normal or a standardized student-t or a generalized error distribution.

1.4.2 The GARCH Model

Although the ARCH model is simple, it often requires many parameters to adequately describe the volatility process of an asset return. Bollerslev (1986) proposes a useful extension to ARCH model known as the generalized ARCH (GARCH) model. For a log return series r_t, let $a_t = r_t - \mu_t$ be the innovation at time t. Then it follows a GARCH (m, s) model if

$$a_t=\sigma_t\epsilon_t, \text{ where } \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \ldots + \alpha_m a_{t-m}^2 + \beta_1\sigma_{t-1}^2 + \beta_2\sigma_{t-2}^2 + \ldots + \beta_s\sigma_{t-s}^2 \tag{1.14}$$

where again $\{\epsilon_t\}$ is a sequence of iid random variables with mean =0 and variance =1, $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, and $\sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) < 1$. Here it is understood that $\alpha_i = 0$ for $i > m$ and $\beta_j = 0$ for $j > s$. The latter constraint on $\alpha_i+\beta_i$ implies that the unconditional variance of a_t is finite, whereas its conditional variance σ_t^2 evolves over time. As before, ϵ_t is often assumed to follow a standard normal or standardized student-t distribution or generalized error distribution. Equation (1.14) reduces to a pure ARCH(m) model if s = 0. The α_i and β_j are referred to as ARCH and GARCH parameters, respectively.

1.5 Objective of the Work

As a market index, we will use a long time-series of returns on the BSE sensex, a widely used market index on the BSE. The dataset has 1000 observations of daily returns over the period from July 1997 to July 2012 was used to find strong evidence of heteroscedasticity in weekly and monthly returns. In the present work, the dataset is fitted with appropriate conditional model and proposed a test method to determine the regime shift in the level of unconditional variance of the data. The unobserved volatility, moments and parameters of the data set were estimated by analyzing the effects of structural changes on the observed data series. A test statistic for testing changes in individual parameters of the volatility equation of GARCH model using estimation function approach was carried out.

CHAPTER 2
INTERVENTION ANALYSIS

Time series observations may sometimes be affected by unusual events, disturbances or errors that create spurious effects in the series and result in extraordinary patterns in the observations that are not in accord with most observations in the time series. Such unusual observations may be referred to as outliers. They may be the result of unusual external events such as strikes, sudden political or economic changes, sudden changes in a physical system or simply due to recording or gross errors in measurement. The presence of such outliers in time series can have substantial effects on the behaviour of sample autocorrelations, partial autocorrelations, estimates of ARMA model parameters, forecasting and can even affect the specification of the model.

Intervention analysis, introduced by Box and Tiao (1975), provides a framework for assessing the effect of an intervention on a time series under study. It is assumed that the intervention affects the process by changing the mean function or trend of a time series. Interventions can be natural or man-made. For example, increase of the speed limit from 65 miles per hour to 70 miles per hour on an interstate highway. This may make driving on the highway more dangerous. On the other hand, drivers may stay on the highway for a shorter length of time because of the faster speed, so the net effect of the increased speed limit change is unclear. The effect of increase in speed limit may be studied by analyzing the mean function of some accident time series data. For example, the quarterly number of fatal car accidents on some segment of an interstate highway.

2.1 Outlier in Time Series

Outliers refer to typical observations that may arise because of measurement and/or copying errors or because of abrupt, short-term changes in the underlying process. For time series, two kinds of outliers can be distinguished, namely additive outliers (AO) and innovative outliers (IO). These two kinds of outliers are often abbreviated as AO and IO, respectively.

2.1.1 The Model

Abraham and Box (1979), Fox (1972) and Martin (1980) discussed two characterizations of outliers in the context of time series models:

1. Aberrant observation model (AO: additive outlier):

$$y_t = z_t + \omega\, I_t(k);\ \ \Phi(B)\, z_t = \Theta(B)\, a_t \tag{2.1}$$

2. Aberrant innovation model (IO: innovational outlier):

$$y_t = z_t + \omega\, \Phi^{-1}(B)\, \Theta(B)\ I_t(k);\ \ \Phi(B)\, z_t = \Theta(B)\, a_t \tag{2.2}$$

Here y_t denotes the observed time series, z_t the underlying process without the impact of outliers and $I_t(k) = 1$ if $t = k$ and zero otherwise. In the first model, only the level of the k^{th} observation is affected. In the second model, the outlier affects the shock at time k, which in turn influences z_k, $z_{k+1}, \ldots$.

2.1.2 Estimates of Outlier Effects

Let us assume that the time T of the intervention and the time series parameters φ and θ are known. Let $e_t = \Pi(B)y_t$; $\Pi(B) = \Phi(B)\, \Theta^{-1}(B)$. Then,

For AO $\qquad e_t = a_t + \omega\, x_{1t} \qquad (2.3)$

IO $\qquad e_t = a_t + \omega\, x_{2t} \qquad (2.4)$

where $x_{1t} = \Pi(B)\, I_t(T)$ and $x_{2t} = I_t(T)$.

The least squares estimates of the intervention impact ω and their variances can be obtained as follows:

for AO $$\hat{W}_{ao} = \frac{\sum e_t x_{1t}}{\sum x_{1t}^2} \tag{2.5}$$

$$V(\hat{W}_{ao}) = \frac{\sigma^2}{\sum x_{1t}^2} \tag{2.6}$$

for IO

$$\hat{W}_{io} = \frac{\sum e_t x_{2t}}{\sum x_{2t}^2} = e_T \tag{2.7}$$

$$V(\hat{W}_{io}) = \frac{\sigma^2}{\sum x_{2t}^2} = \sigma^2 \tag{2.8}$$

The notation reflects the fact that the estimates depend upon T.

In the AO model, $\hat{w}_{ao}$ is a linear combination of the shocks future to T and its variance can be much smaller than σ^2. In the IO model, $\hat{w}_{io}$ is the residual at T with variance σ^2.

Significance tests for outliers can be performed based on the standardized estimates $\lambda_{ao} = \frac{\tau \hat{w}_{ao}}{\sigma}$; $\lambda_{io} = \frac{\hat{w}_{io}}{\sigma}$; *where* $\tau = \sum_0^{n-T} \pi_i^2$, for AO and IO types respectively. Under the null hypothesis that $\hat{w} = 0$, both statistics will have standard normal distribution.

In practice, T as well as the time series parameters are unknown and have to be replaced by the estimates.

2.1.3 Iterative Procedure for Outlier Detection

In practice, the time point T of a possible outlier as well as the model parameters are unknown. To address the problem of detection of outliers at unknown times, iterative procedures that are relatively convenient computationally had been proposed by Chang, Tiao and Chen to identify and adjust for the effects of outliers. It consists of *specification, estimation, detection and removal* cycles to build a time series model in the presence of exogenous disturbances. In each iteration, the maximum of a given test statistic is selected as the candidate for that type of disturbance and the grand maximum across the tests is identified as the most likely exogenous disturbance. This grand maximum is then compared with a pre-specified critical value so that the existence of an exogenous disturbance can be judged. It consists of the following steps:

Step 1: Specify an ARMA model for the observed series Y_t and obtain parameter estimates and residuals of the specified model.

Step 2: The statistics λ_{ao} and λ_{io} are computed for each time t= 1, 2,…, n as well as $\lambda_T = \max_t \left(\max(\lambda_{ao}, \lambda_{io})\right)$, where T denotes the time when this maximum occurs.

Step 3: The possibility of an outlier type IO identified at time T if $\lambda_T = |\lambda_{io}| > c$, where c is a pre-specified value. If $\lambda_T = |\lambda_{ao}| > c$, the possibility of an AO outlier is identified.

Step 4: The effect of IO can be eliminated from the residuals by defining $\tilde{e}_t = e_t - \hat{w}_{io} = 0$ at T. The effect of AO can be removed from the residuals by defining $\tilde{e}_t = e_t - \hat{w}_{ao}\pi_{t-T}$ for t ≥ T. In either case, a new estimate $\hat{\sigma}_a^2$ is computed from the modified residuals $\tilde{e}_t$.

Step 5: Repeat the above procedure until all the outliers are identified.

2.1.4 Illustrative Examples

As an example, the value of unfilled orders of radio and TV (UNFTV) from the U.S. Bureau of the Census is reanalysed. This is a monthly series for the period from January 1958 to October 1980 giving a total of 274 data points. This data set was used by Martin, Samarov and Vandaele (1983) to demonstrate the effectiveness of their approximate conditional mean (ACM) type robust filter model and Tsay (1988) used to describe the method of detection of different types of outliers. Figure 2.1 plots the logged series Y_t and first differenced series $X_t = (1\text{-}B)Y_t$. It is clear from the differenced series plot of figure 2.1 that the series has changing variance.

Following the proposed procedure, a tentative model specification for X_t was first performed and the multiplicative seasonal IMA (0, 0, 1) x (0, 1, 1), model appears to be reasonable. This is also the model used by Martin *et al*. Table 2.1 summarizes the detection results of the procedure. A change in the model at t = 224 was detected with a variance ratio of 5.608. Figure 2.2 plots the adjusted series X_t*. The series now appears to be stable. Some outlying observations, however are present. Procedure described in section 2.1.3 was then applied to X_t^*. Table 2.2 gives the results of outlier detection. No level change, permanent or transient was detected. There are, however, two innovational outliers at times t = 125 and 256, respectively. An additive outlier appears at time t = 77 if the critical value is reduced to 3.0. Figure 2.3 shows the residual plot the outlier-adjusted series z_t^*.

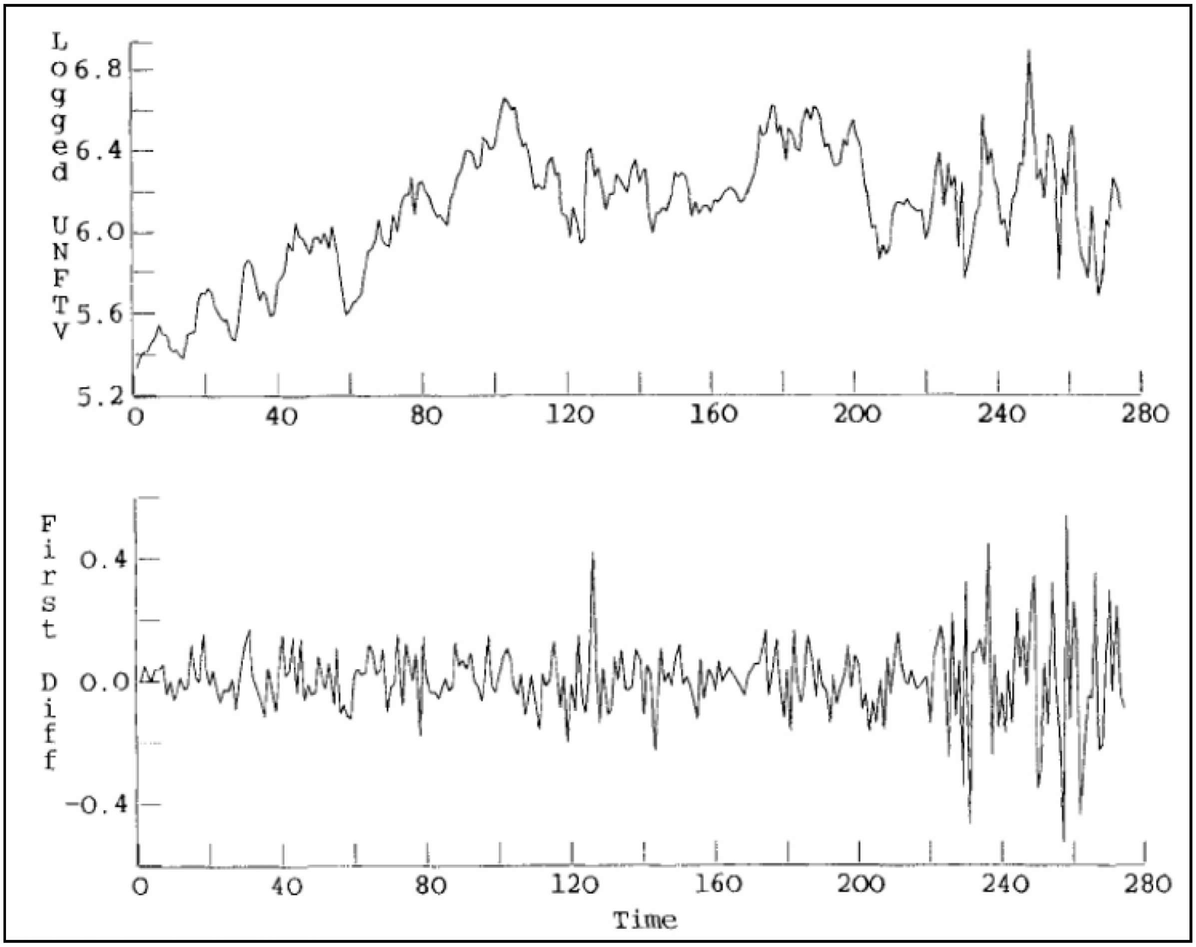

Figure 2.1: Plots of logged UNFTV (a) Original series and (b) First differenced series.

Table 2.1: Summary of variance detection of UNFTV series.

Iteration	Parameter Estimates		Variance Ratio
	Θ_1	Θ_{12}	
1	0.3735	0.7535	
2	0.2313	0.8739	5.608

If the variance change at t = 224 was ignored, then the procedure described in section 2.1.3 would identify more than 15 outliers or level changes for the data. On the other hand, when one takes into account of the variance change, there are only two outliers left. Thus, ignoring the variance change can be troublesome. Second, the final parameter estimates given in Table 2.2 are different from those obtained by the ACM-type robust filter model of Martin *et al*. (1983), who used the original data instead of the logged series. The seasonal MA(1, 2) parameter estimates are close but the MA(1) coefficients are different. The MA(1) estimate in Table 2.2 is significant. As a final note, when the original data were used, procedure detected two variance changes, one at t= 44 and the other at t = 243 and also identified three IO's at t = 235, 125 and 248.

The adjusted series has MA coefficients θ_1 = 0.2343 (0.0602) and θ_{12} = 0.8552 (0.0358).

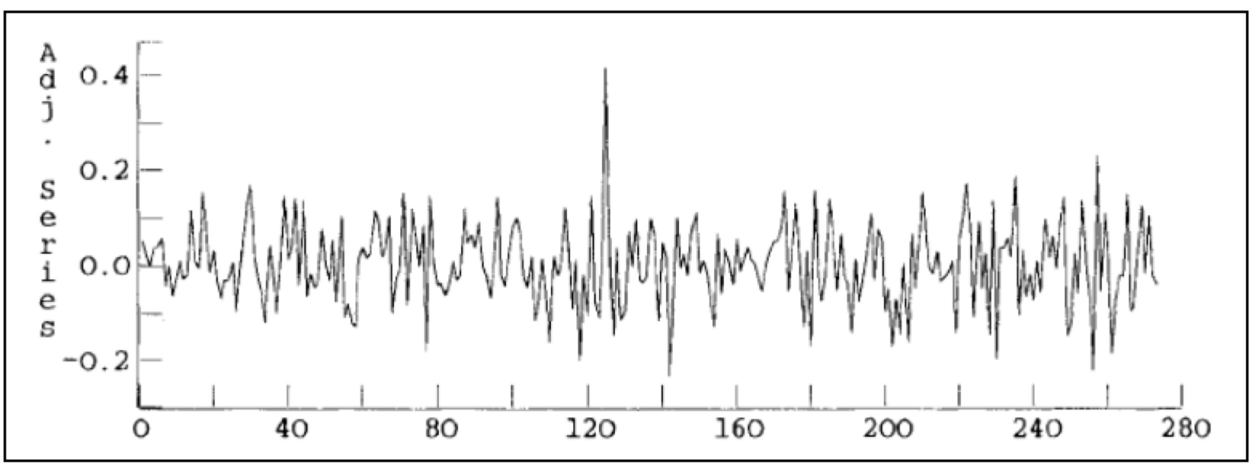

Figure 2.2: Plot of first differenced log UNFTV series after variance change is adjusted.

Table 2.2 Summary of outlier detection of UNFTV series.

Iteration	MA Parameters		Type	Time	Magnitude
	θ_1	θ_{12}			
1	0.2323	0.8739	IO	125	0.3815
2	0.2071	0.8556	IO	256	-0.2671
3	0.1991	0.8600	Insignificant		

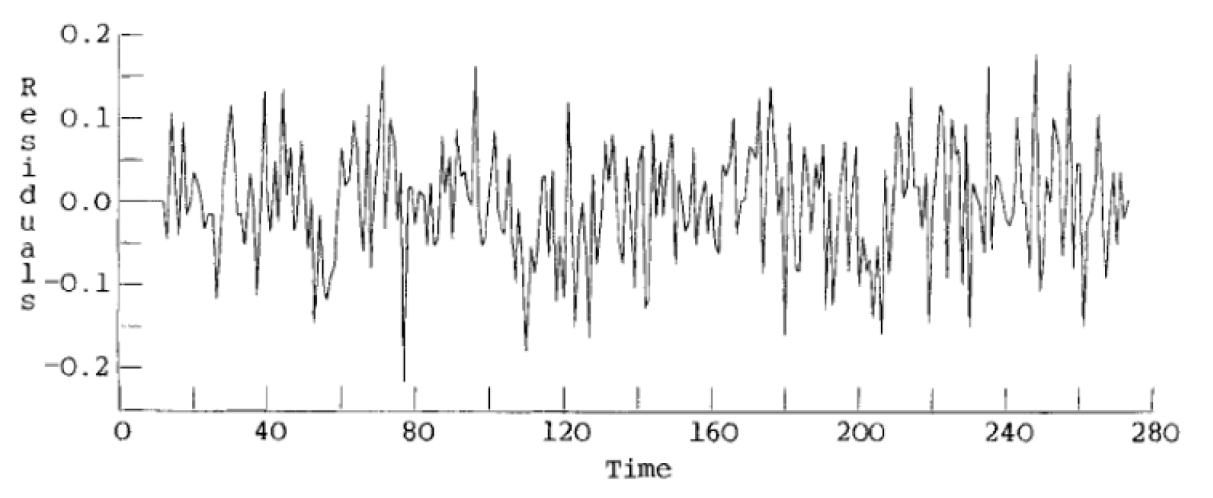

Figure 2.3: Residual plot of logged UNFTV series after disturbances removed.

2.2 Outliers in Financial Time Series

Outliers are aberrant observations that are away from the rest of the data. They can be caused by recording errors or unusual events such as changes in economic policies, wars, disasters, financial crises and so on. They are also likely to occur if errors have fat-tailed distributions as in the case of financial time series. These observations may take several forms in time series. The first and most usually studied is the additive outlier (AO), which only affects a single observation. In contrast, an innovative outlier (IO) affects several observations. Balke and Fomby (1994) found that many of the detected outliers in financial time series are IO's, especially for data at a high frequency.

2.2.1 The Model

Charles and Darné (2005) extended the additive-outlier detection method in GARCH models developed by Franses and Ghijsels (1999) to innovative outliers, using the Chen and Liu (1993) approach.

Consider the returns series e_t, which is defined by e_t =log p_t - log p_{t-1}, where p_t is the observed price at time t and consider the GARCH(1,1) model

$$\begin{aligned} &\varepsilon_t = z_t\sqrt{h_t} \\ &\varepsilon_t \sim N(0,h_t) \\ &z_t \sim iid\ N(0,1) \\ &h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1} \end{aligned} \tag{2.9}$$

$$\begin{aligned} &where \\ &\alpha_0 > 0, \alpha_1 \geq 0, \beta_1 \geq 0\ and\ \alpha_1 + \beta_1 < 1 \end{aligned}$$

The GARCH(1, 1) model can be rewritten as an ARMA(1, 1) model for e_t^2.

$$\varepsilon_t{}^2 = \alpha_0 + \left(\alpha_1 + \beta_1\right)\varepsilon_{t-1}^2 + v_t \text{ - } \beta_1 v_{t\text{-}1}, where, v_t = \varepsilon_t^2 - h_t \tag{2.10}$$

This analogy of the GARCH model with an ARMA model allows one to directly adapt the method of Chen and Liu (1993) to detect and correct AO's and IO's in GARCH models. Specifically, suppose that instead of the true series e_t one observes the series e_t, which is defined as

$$e_t^2 = \varepsilon_t^2 + \omega_i\xi_i(B)I_t(\tau), with\ i = 1,2. \tag{2.11}$$

where $I_t(\tau)$ is the indicator function defined as $I_t(\tau)$ =1 if τ =T and zero otherwise where τ is the date of outlier occurring, ω_i is the magnitude of the outlier effect and $\xi_i(B)$ represents their dynamic pattern with

$$\xi_1(B) = 1\ for\ \text{AO, and}\ \xi_2(B) = (1-\beta_1 B)(1-(\alpha_1+\beta_1)B)^{-1}\ for\ IO \tag{2.12}$$

An AO is related to an exogenous change that directly affects the series and only its level of the given observation at time t =T. An IO is possibly generated by an endogenous change in the series and affects all the observations after time t through the memory of the process.

The residuals h_t of the observed series e_t^2 are given by

$$\eta_t = \frac{-\alpha_0}{1-\beta_1 B} + \pi(\mathrm{B})\,\mathrm{e}_t^2 = v_t + \pi(\mathrm{B})\xi_i(B)\omega_i I_t(\tau), \quad \pi(B) = (1-(\alpha_1+\beta_1)B)(1-\beta_1 B)^{-1} \tag{2.13}$$

The expression (2.13) can be interpreted as a regression model. Hence $\eta_t = \omega_i x_{it} + v_t$. Outlier detection is based on the maximum value of the standardized statistics of the outliers effects with x_{it}=0 for i=1, 2 and t<τ , x_{it}=1 for i=1, 2 and t = τ , $x_{1,\tau+k}=-\pi_k$ (for AO) and $x_{2,\tau+k}$ =0 (for IO) for t >τ and k>0.

$$\begin{aligned}
AO \quad \hat{\tau}_1 &= \left(\widehat{\omega_1(\tau)} \big/ \widehat{\sigma_v}\right)\left(\sum_{t=\tau}^{n} x_{1t}^2\right)^{1/2} \\
&= \left[\left(\sum_{t=\tau}^{n} x_{1t}\eta_t\right)\left(\sum_{t=\tau}^{n} x_{1t}\right)^{-1} / \widehat{\sigma_v}\right]\left(\sum_{t=\tau}^{n} x_{1t}^2\right)^{1/2} \\
IO \quad \hat{\tau}_1 &= \left(\widehat{\omega}_2(\tau) \big/ \widehat{\sigma}_v\right) = \eta_t / \widehat{\sigma}_v
\end{aligned} \tag{2.14}$$

where $\widehat{\sigma_v^2}$ denotes the estimated variance of the residual process.

2.2.2 Iterative Procedure

The outlier detection method for GARCH(1,1) models then consists of following steps:

1. Estimate a GARCH(1, 1) model for the observed series e_t and obtain estimates of the conditional variance $\widehat{h_t}$ and $\widehat{\eta_t} = e_t^2 - \widehat{h_t}$. (2.15)

2. Obtain estimates $\hat{\omega}_i$ (i =1, 2) for all possible t =1, . . . , n, and compute $\widehat{\tau_{\max}} = \max_{1\le\tau\le n}\left|\widehat{\tau_i}\right|$. If the value of the test statistic exceeds the critical value C, an outlier is detected at the observation for which $\hat{t}$ is maximized.

3. Replace e_t^2 with

$$\begin{aligned}
&AO: e_\tau^{*2} = e_\tau^2 - \widehat{\omega_1}, \\
&IO: e_\tau^{*2} = e_{\tau+j}^2 - \widehat{\omega_2}\psi_j,
\end{aligned} \quad \text{with } j > 0 \tag{2.16}$$

where $\psi(B) = \pi^{-1}(\mathrm{B})$. The outlier-corrected series e_t* is defined as

$$
AO: e_t^* = \begin{cases} e_t & for\, t \neq \tau \\ sign(\mathrm{e}_t)\sqrt{e_t^{*2}} & for\, t = \tau \end{cases}
$$

$$
IO: e_t^* = \begin{cases} e_t & for\, t < \tau \\ sign(\mathrm{e}_t)\sqrt{e_t^{*2}} & for\, t = \tau + j, \mathrm{j} > 0 \end{cases} \qquad (2.17)
$$

4. Return to step 1 to estimate a GARCH(1, 1) model for the series e_t* and repeat all steps until no $\hat{t}$ maximum test-statistic exceeds the critical value C.

A critical value C =10 is used, which considered a low-sensitivity value for small sample size (Verhoeven and McAleer, 2000). This choice for C is based on simulation experiments proposed by Franses and Van Dijk (2002). The authors simulate some percentiles of the distribution of the $\hat{t}$ maximum statistic under the null hypothesis that no outliers are present for several values of ARCH and GARCH parameters and for two sample sizes (250 and 500). It is seen that the value of C =10 is reasonably close to the 90$^{\text{th}}$ percentile of this distribution for any parameter combinations.

CHAPTER 3
SOME DIAGNOSTIC CHECKING MEASURES

Some statistics used in regression analysis are considered for detection of outliers in time series. Approximations and asymptotic distributions of these statistics are derived. In this work, a method is proposed for distinguishing an observational outlier from an innovational one. Andrews and Pregibon (1980), Cook and Weisberg (1982), and Draper and John (1981) discussed detection of outliers and influential points in regression models. Their approach was basically to delete suspicious observations and to build a measure of the resulting change in the features of the model such as the estimated parameter values and residuals.

3.1 Q - Statistics in Time Series

Let z, be a stationary AR process of order p [AR(p)],

$$\phi(\mathrm{B})\,\mathrm{Z}_t = a_t\,, \tag{3.1}$$

where $\phi(\mathrm{B}) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p$

Given a set of observations $z_1, z_2, \ldots, z_n$, we can write

$$Z = X\phi + a \tag{3.2}$$

where $Z' = (z_{p+1,}.....z_n), \phi' = (\phi_{1,...,}\phi_p), \mathrm{a}' = (\mathrm{a}_{p+1,...,}\mathrm{a}_n)$ and

$$X = \begin{bmatrix} z_p & z_{p-1} & & z_1 \\ z_{p+1} & z_p & & z_2 \\ . & . & . & . \\ z_{n-1} & z_{n-2} & \cdots & z_{n-p} \end{bmatrix}$$

Then the conditional least squares (CLS) estimate of φ is given by $\hat{\phi} = (X'X)^{-1} X'Z$ (3.3)

the fitted values are $\hat{Z} = X\hat{\phi} = X(X'X)X'Z = HZ$ (3.4)

where $H = X(X'X)X'$ and the residuals are e = (I-H)Z. In a linear regression model, X is assumed to be a constant matrix, which is no longer true in the present situation. Moreover, the

z_t's are assumed to be independent in the linear regression model. An observation could be deleted without affecting the consecutive ones and the deletion of an equation is equivalent to the deletion of an observation. In the time series context, however, this is no longer true either. A suspect observation, Z_T, is involved not only in one equation but also in p + 1 consecutive equations. Thus it is necessary to delete not only one equation but p + 1 equations.

Suppose that there is one suspected observation at t = T. The matrix X and vectors Z and e can be partitioned as follows

$$X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \begin{array}{c} (T-p) \text{ x p} \\ k \times p \\ (\text{n}-T-k) \text{ x p} \end{array}$$

$$Z = \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} \begin{array}{c} (T-p) \text{ x } 1 \\ k \times 1 \\ (\text{n}-T-k) \text{ x } 1 \end{array}$$

$$e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} \begin{array}{c} (T-p) \text{ x } 1 \\ k \times 1 \\ (\text{n}-T-k) \text{ x } 1 \end{array} \tag{3.5}$$

where k is the number of equations that are to be deleted. The residuals, e, can be expressed in the partitioned form as

$$e = \begin{bmatrix} I-H_{11} & -H_{12} & -H_{13} \\ -H_{21} & I-H_{22} & -H_{23} \\ -H_{31} & -H_{32} & I-H_{33} \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} \tag{3.6}$$

Where $H_{ij} = X_i(X'X)X_j', \quad \text{i, j} = 1,2,3.$ Following the suggestion of Draper and John (1981) for regression situations, we consider the statistics

$$Q_{k(T)} = e'_2(I-H_{22})^{-1}e_2 \tag{3.7}$$

When k = 1, $e_2 = e_T$, and when k = p + 1, e' = (e_T, . .. , e_{T-p}). Now $Q_{k(T)}$ could be decomposed into two terms

$$Q_{k(T)} = e_2' e_2 + (\widehat{\phi} - \widehat{\phi}_*)(X_1'X_1 + X_3'X_3)(\widehat{\phi} - \widehat{\phi}_*) \\ = Q_{k1(T)} + Q_{k2(T)} \tag{3.8}$$

$\phi_*' = (X_1'X_1 + X_3'X_3)^{-1}(X_1' Z_1 + X_3' Z_3)$ is the estimate of φ after deleting k equations. It is found in simulations that the statistics $Q_{k(T)}$, $Q_{k1(T)}$ and $Q_{k2(T)}$ are useful indicators for outliers.

3.1.1 Asymptotic Distribution

We consider the statistics maxQ_k(t), maxQ_{kl}(t), and maxQ_{k2}(t) for identifying outlier locations for which the sampling properties of these statistics are required. The exact sampling distributions are difficult to come by, however, and hence we appeal to large-sample theory. If there are no outliers, it is well known that $\hat{\varphi}$ converge in probability to φ. Then

$$Q_{k1(t)} \xrightarrow{p} Q^*_{k(t)} \\ Q^*_{k(t)} = \sum_{i=t}^{t+k-1} a_i^2 \sim \sigma^2 \chi^2_{(k)}, \tag{3.9}$$

$$Q_{k(t)} \xrightarrow{p} Q^*_{k(t)}, Q_{k2(t)} \xrightarrow{p} 0,$$

$$\max_t Q_{k1(t)} \xrightarrow{p} \max_t Q^*_{k(t)},$$

$$\max_t Q_{k(t)} \xrightarrow{p} \max_t Q^*_{k(t)}$$

Then it can be shown that

$$P(\max_t Q^*_{k(t)} \le C_m(\tau)) \to e^{-\nu\tau}, \nu - \text{constant} \tag{3.10}$$

For a given significance level α and F_k(.) denote the cumulative distribution function of chisquare, then

$$C_m(\tau) = F_k^{-1}[1 + \frac{\ln(1-\alpha)}{m\nu}] \tag{3.11}$$

3.1.2 Simulation Study

Here a simulation study to gain some understanding of the accuracy of the extreme-value type of approximations made was conducted. Samples of size n = 100 and 200 were generated from the following models:

$$\begin{aligned}
&\textit{Model } 1: z_t = 0.5 z_{t-1} + a_t \\
&\textit{Model } 2: z_t = z_{t-1} - 0.5 z_{t-2} + a_t \\
&\textit{where } a_t \sim \mathrm{N}(0,1)
\end{aligned}$$

For Model 1, the statistics $\max Q_k(.)$ and $\max Q_{kl}(.)$ were calculated for k = 1 and 2, whereas for Model 2 these values were calculated for k = 1 and 3. This process was repeated 1,000 times. The 1%, 2.5%, 5%, and 10% significance points were estimated. These are shown in Tables 3.1to 3.4, together with those obtained from the extreme-value approximations. Since calculations were done in double precision and large storage was required, it was decided to limit the number of simulations to 1,000. The estimated 2.5% significance points from 500 simulations were almost the same as those from 1,000 simulations. We also felt that 1,000 repetitions would yield fairly accurate estimates of the 1% significance points.

The results indicate good agreement between the simulated values and the extreme value approximations. For example, when n = 100 and k = 2 (Model 1, Table 3.3), the 10% significance point from the approximation is 13.34, the simulated value for $\max_t Q_{kl}(t)$ is 13.25, and that for $\max_t Q_k(t)$ is 13.53. When n = 100, k = 3 (Model 2, Table 3.4), the 10% significance point from the approximation is 15.11, the simulated value for $\max_t Q_{kl}(t)$ is 14.85, and that for $\max Q_k(t)$ is 15.30. Similar agreement is seen for other significance levels and other sample sizes also. In our experience, the plots of the statistics introduced are usually sufficient to spot the outlier locations and types.

Table 3.3 Significance points for Q when p=1 and k=1 of Model 1.

n	Type	α			
		0.1	**0.05**	**0.025**	**0.01**
100	EV	10.71	12.05	13.38	15.13
	Q_{11}	10.73	12.36	13.67	15.72
	Q_1	10.88	12.44	14.03	15.73
200	EV	12.01	13.36	14.68	16.47
	Q_{11}	11.85	13.42	14.68	16.12
	Q_1	11.98	13.43	14.75	16.17

Table 3.2 Significance points for Q when p=2 and k=1 of Model 2.

n	Type	α			
		0.1	**0.05**	**0.025**	**0.01**
100	EV	10.69	12.04	13.36	15.11
	Q_{11}	10.70	12.20	13.61	15.51
	Q_1	10.87	12.36	13.85	15.7371
200	EV	12.00	13.35	14.67	16.45
	Q_{11}	11.80	13.46	14.44	15.99
	Q_1	11.99	13.56	14.53	16.11

Table 3.3 Significance points for Q when p=1 and k=2 of Model 1.

n	Type	α			
		0.1	**0.05**	**0.025**	**0.01**
100	EV	13.34	14.83	16.26	18.12
	Q_{k1}	13.25	14.49	15.98	17.71
	Q_1	13.53	14.68	16.34	18.04
200	EV	14.80	16.25	17.67	19.51
	Q_{k1}	14.38	15.79	17.00	18.99
	Q_1	14.53	15.83	17.22	19.31

Table 3.4 Significance points for Q when p=2 and k=3 of Model 2.

n	Type	α			
		0.1	**0.05**	**0.025**	**0.01**
100	EV	15.11	16.73	18.25	20.21
	Q_{11}	14.84	16.24	17.95	19.69
	Q_1	15.30	16.91	18.57	20.67
200	EV	16.70	18.38	19.81	21.85
	Q_{11}	16.16	17.73	18.92	21.12
	Q_1	16.44	18.09	19.26	21.39

The patterns of the diagnostic statistics for an AR(p) model are discussed and indicated that these patterns can be used to detect and identify outlier types even when the model is over fitted. In general, the process may not be autoregressive. Suppose that the true process is ARMA(p, q) as given in equation 1.11. Often such processes can be approximated by an AR(p + q) model as described in Box and Jenkins (1976). In practice, for outlier detection, we found this to be a good approximation. Our model building strategy then starts with the fitting of a sufficiently large AR process. Based on the outlier detection methods discussed in the previous sections, a model building procedure is put forward.

Table 3.5 Patterns of Q statistics assuming an outlier at t=T.

Statistic	IO outlier	AO outlier
Q_{11}, Q_1, k=1	Only high value at t=T	All the values at t= T, T+1,…,T+p are affected
Q_{12}, k=1	All the values at t= T, T+1,…,T+p may be affected	All the values at t= T, T+1,…,T+p are affected
$Q_{(P+1)1}$, $Q_{(p+1)}$, k=p+1	Large values at t=T-p,T-p+1, …T.	The values at t=T-p, T-p+1, …T. are affected and maximum value at t=T
$Q_{(P+1)2}$, k=p+1	The values at t=T-p,T-p+1, …T. may affected	The values at t=T-p,T-p+1, …T. are affected and maximum value at t=T

We now illustrate the four-step model building procedure. Consider an out-procedure using the first 100 observations from Series- A (chemical process concentration readings) of Box and Jenkins (1976), where the model employed by the authors is ARMA(1, 1). An AO be the position outlier at t = 43 by subtracting 1 unit from the original observation. The data is analysed with mean, $\hat{\mu} = 17$ is subtracted to adjust. Tentative model selection suggests $p^* = 3$. Then the estimates, $\hat{\varphi}^* = (0.228, 0.263, 0.140)'$, are computed by fitting the model. The plots of $Q_{1(t)}$ and $Q_{4(t)}$ are shown in figure 3.1 and they indicate that the observations 43 and 64 may be discrepant.

Now

$$Q_{1(t)} / \widehat{\sigma^2} = Q_{1(43)} / \widehat{\sigma^2} = 22.72$$

and
$$P(\max_t Q_{1(t)} / \widehat{\sigma} > 22.72) < 0.01$$

The patterns of the statistics suggest that the 43rd observation is an AO outlier. Deletion of equations corresponding to Z_{43}, Z_{44}, Z_{45} and Z_{46} leads to $\varphi^* = (0.326, 0.271, 0.145)'$ and $y_{43} = 0.515$. Then the second iteration with the cleaned series gives $\varphi^* = (0.325, 0.271, 0.155)'$. It now appears that Q_1 has a single large value at t = 64, $Q_l(64)/\sigma^2 = 23.90$, indicating the presence of an IO outlier as shown in figure 3.2. After deleting the equation for Z_{63} and estimating the parameters, the series is cleaned. Now the third iteration leads to small values for $\max Q_1(t)$ and $\max Q_4(t)$. Hence no more outliers are suspected and the iteration is terminated. Further model specification suggests an ARMA(1, 1) process and the maximum likelihood estimates are $\hat{\varphi} = 0.94, \hat{\theta} = 0.58 \textit{ and } \hat{\sigma}^2 = 0.087$.

The proposed outlier detection strategy is simple and intuitively appealing, and it seems to work reasonably well in a number of examples. It can be made easily as a part of the existing time series softwares.

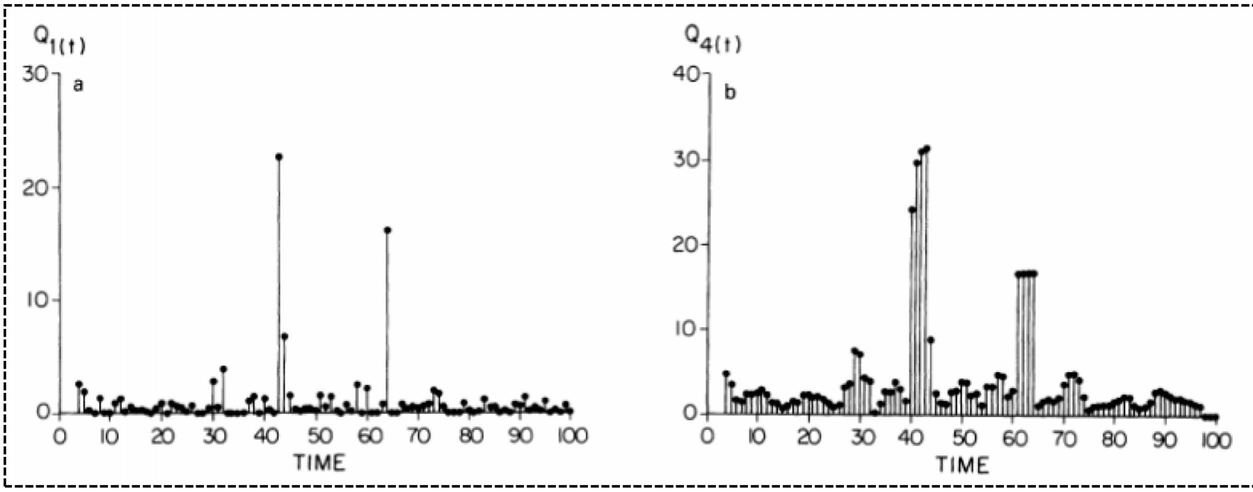

Figure 3.1 Q_1(t) and Q_4(t) after the First iteration.

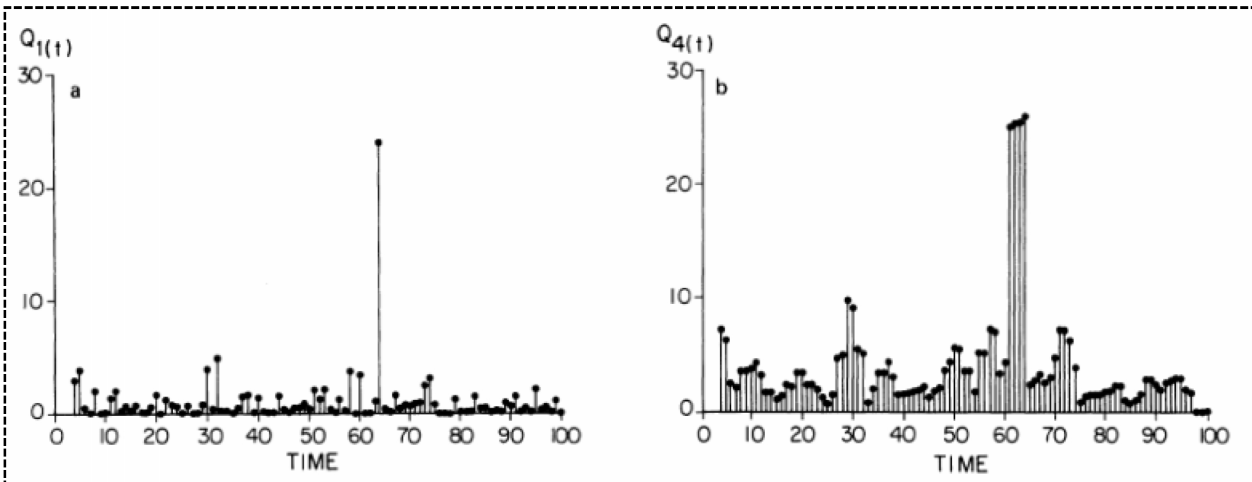

Figure 3.2 Q_1(t) and Q_4(t) after the Second iteration.

3.2 Q-Statistics in Financial Time Series

Consider an ARCH(1) model described in section 1.4.1. As in Tsay (2002) the ARCH(1) model can be converted to AR model. Then the diagnostic measures described in section 3.1 can be directly applied to the series.

The statistics $Q_T(k)$ and $Q_{T1}(k)$ are useful indictors of outliers.

$$Q_T(k) = e_2'(I-H_{22})^{-1}e_2 \qquad (3.12)$$

They may be used to distinguish an AO outlier from an IO one. If the outlier at t=T is AO, then $Q_1(T)$ and $Q_1(T+1)$ are large compared with the rest. But only $Q_{12}(T)$ is larger than the others. If the outlier at t=T is IO, then $Q_1(T)$ are large compared with the others. The probability distribution of the test statistic is difficult to determine in the financial time series set up since the error distribution is chisquare. Hence the critical value of the test statistic is determined by simulation.

We conducted a simulation study to observe the pattern of Q-statistics with a sample size 100 for an ARCH(1) model with an additive outlier at t=22. The graphical outcome of the test statistic is shown in figure 3.3.

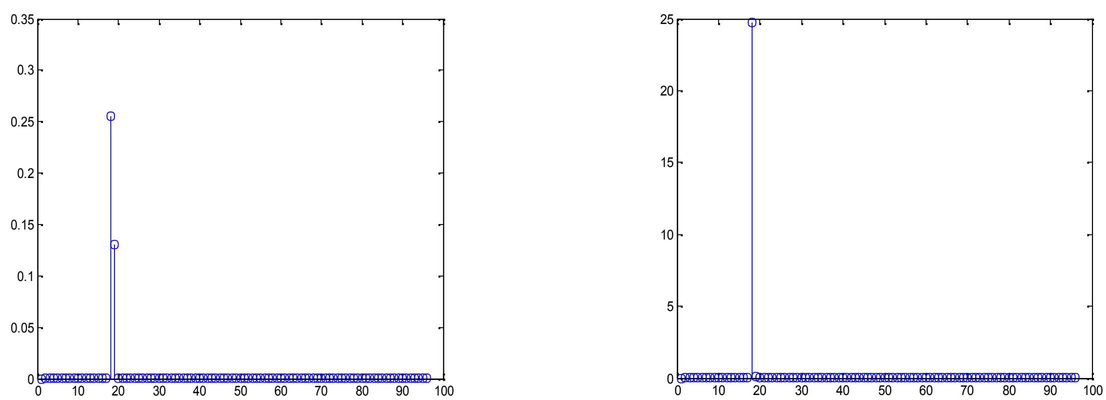

Figure 3.3 $Q_1(T)$ and $Q_{12}(T)$ versus time data for the simulated data. This directly indicates that the outlier is AO.

3.3 Forecasting

Let T be the point at which outlier is identified and if it is AO, then delete p equations (T-p) to T from the data series. Then forecasts of the ARCH model can be obtained recursively as those of an AR model.

Consider an ARCH(m) model. At the forecast origin T, the 1-step-ahead forecast

$$h_t(1) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \ldots + \alpha_m \varepsilon_{t+1-m}^2$$

and the k-step-ahead forecast

$$\widehat{h_{t+1}} = \alpha_0 + \alpha_1 e_t^2 + \beta_1 h_t$$

The optimal predictor, h_{t+s}, of the conditional variance for forecast horizon 's' is the conditional expected value of h_{t+s}, that is $E(h_{t+s})$. It is easy to show that:

$$\widehat{h_{t+s}} = E(h_{t+s}) = \alpha_0 \sum_{i=1}^{s-1} \left[(\alpha_1 + \beta_1) + (\alpha_1 + \beta_1)^{s-1} h_{t+1} \right]$$

where $\widehat{h_{t+1}} = \alpha_0 + \alpha_1 e_t^2 + \beta_1 h_t$ is known at time t.

The necessary and sufficient condition for the existence of the unconditional variance is

$$Var(e_t) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1} \text{ of } e_t \text{ is } \alpha_1 + \beta_1 < 1$$

3.4 Conclusions

We have considered some statistics available in the regression analysis to detect outliers in time series and financial time series. Here we have considered the Q statistics to detect outliers in ARCH (1) models. Specification of outlier with the same statistic is also described. After detecting the outlier, traditional methods can be used to clean the series and the m-step ahead forecast can be used to correct the affected observations. The use of this statistics in the outlier detection of GARCH process is used for further study.

CHAPTER 4
SHOCK DETECTION IN STOCK EXCHANGE

BSE (Bombay Stock Exchange) Limited was established in 1875 and it is the Asia's fastest stock exchanges with a speed of 200 microseconds, and the world's third largest leading exchange for Index option trading (from March 2014 onwards, source: World Federation of Exchange). The total market capitalization is of USD 1.151 Trillion for the companies which listed on BSE Ltd as of May 2014, given in Wikipedia, and the Free Encyclopedia (2014). Standard & Poor's (S&P) BSE index consists of the following sector names as follows Auto, Banks, Consumer Durables, Capital Goods, FMCG, Healthcare, IT, Metal, Oil & Gas, Power, and Technology. These sectoral indices have significantly received a large amount of money from Foreign Institutional Investors (FIIs) and also have a large number of subsets contained in these broad sectoral indices, which provide a great trade-off platform for the intercontinental traders to invest their stocks in the Indian market. The highlight of the increasing SENSEX aids the sectoral indices that have outperformed others from 1 January 2013 to March 2014, by Priyanka (2014).

Shocks in sensex data may be associated with several extraordinary economic events such as financial crises, recessions, calamities on earth, weather and changes in Fed policy. Structure changes and outliers are very common in financial data and lead to erroneous conclusions due to important model misspecification. If the type and date of the disturbances are known, then their effects can easily be controlled.

In a recent study, Bologna and Cavallo (2002) investigated the stock market volatility in the post derivative period for the Italian stock exchange using Generalised Autoregressive Conditional Heteroscedasticity (GARCH) class of models. To eliminate the effect of factors other than stock index futures (i.e., the macroeconomic factors) determining the changes in volatility in the post derivative period, the GARCH model was estimated after adjusting the stock return equation for market factors, proxied by the returns on an index (namely Dax index). In the present study, following the method used by Bologna and Cavallo (2002), a GARCH model has used to empirically evaluate the effects on volatility of the Indian stock market and to see that what extent the change (if any) could be attributed to the introduction of index futures. Bologna and Covalla also found that in the post index-future period, the importance of 'present news' has gone up in comparison to the 'old news' in determining the stock price volatility.

4.1 Empirical Analysis

Daily data of BSE sensex has been used for the period from July 1997 to July 2012. The older S& P BSE sensex data can be downloaded from the website of Bombay Stock exchange of India (http://www.bseindia.com/). All share index and compared these results with the inflation rates for the period 1997-2012. This research work relied on information from documentary source or secondary data.

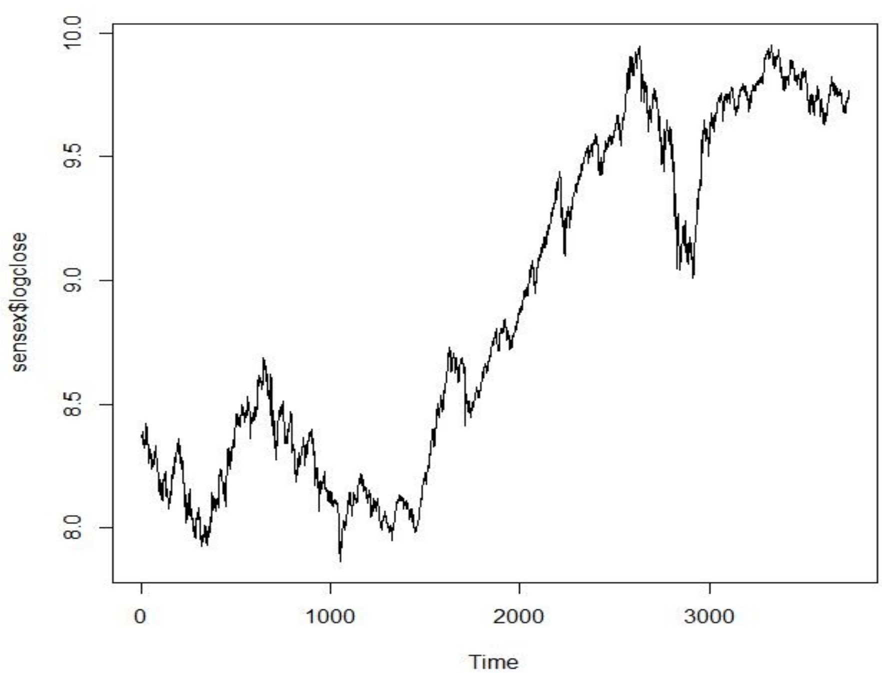

Figure 4.1 Plot of logaritham of close value of BSE sensex from July 1997 to July 2012.

Following Bologna and Cavallo (2002), this work uses Generalised Autoregressive Conditional Heteroscedasticity (GARCH) framework to model returns volatility. The GARCH model was developed by Bollerslev (1986) as a generalised version of Engle's (1982) Autoregressive Conditional Heteroscedasticity (ARCH). In the ARCH model, the conditional variance at time 't' depends on the past values of the squared error terms and the past conditional variances. From the time series plot of the original close values, it is clear that the process is not stationary. So the volatility has been estimated on return (R_t) which is defined as

$$R_t = \log (P_t / P_{t-1}) \tag{4.1}$$

where R_t is the logarithmic daily return at time t and P_{t-1} and P_t are the daily price of an asset at two successive days t-1 and t, respectively. The return series is represented in figure 4.1. The table 4.1 describes the summary statistics of the selected daily close and the log return which helps to understand the properties of the data.

Table 4.1 Descriptive statistics of the log return and daily close.

Summary Statistics	Log Return	Daily Close
Mean	$-7.5012\text{x } 10^{-06}$	9146.029
Standard Error	0.000649034	96.18594
Median	-0.000353566	6246.04
Standard Deviation	0.039665467	5880.73
Sample Variance	0.001573349	34582985
Kurtosis	4.530970047	-1.36605
Skewness	-0.124074112	0.501286
Range	0.54606627	18404.84
Minimum	-0.293557384	2600.12
Maximum	0.252508886	21004.96
Sum	-0.02801713	34187856
Count	3735	3738
Largest(1)	0.252508886	21004.96
Smallest(1)	-0.293557384	2600.12
Confidence level (95.0%)	0.001272495	188.5821

4.2 Fitting GARCH (1, 1) Model

To fit an appropriate GARCH model for the estimation of the conditional market volatility BSE Sensex close, firstly we considered the autocorrelation function and partial auto correlation functions. The same values are calculated and the following figure 4.2 shows their significance in the fixation of order of the model. However, squared residuals from the stock returns equation exhibit autocorrelation.

Using R software packages, we estimated the parameters of GARCH (1, 1) and test their significance. The programme used in R is illustrated below

```
library(tseries)
d = read.table ("mydata/logrt.dat")
 arch.d = garch (d,order=c(0,1))
summary (arch.d)
```

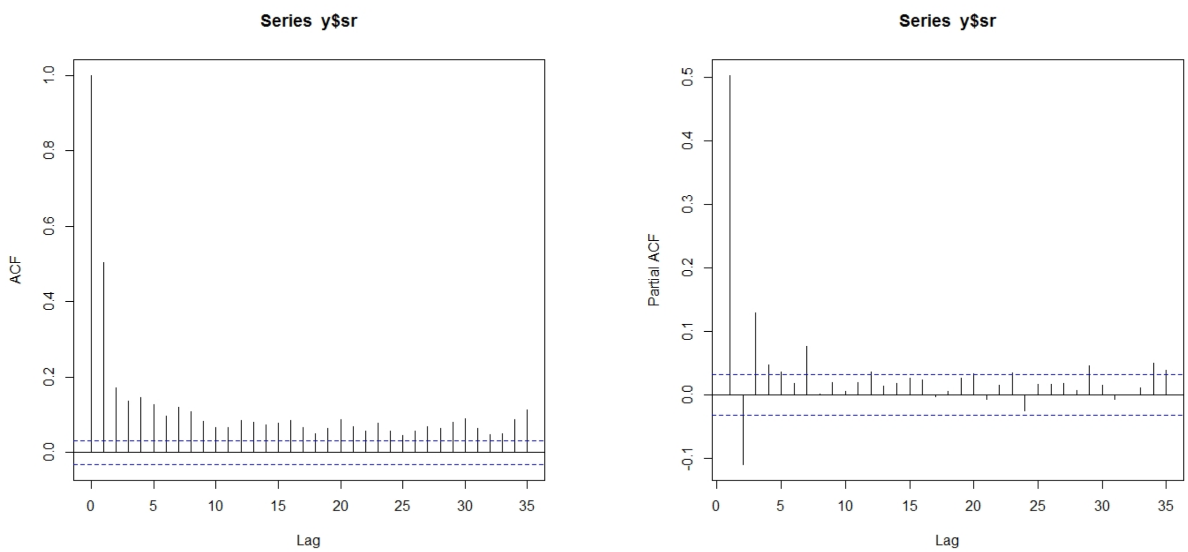

Figure 4.2 Plot of ACF & PACF of square of log returns.

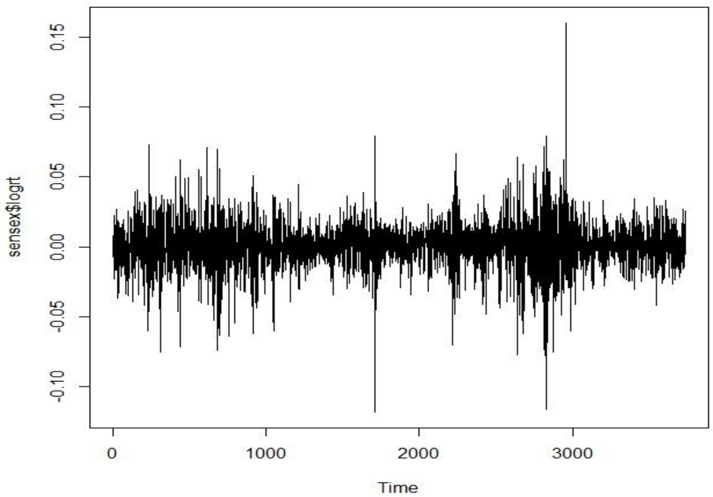

Figure 4.3 Plot of square of log returns.

The output is described below:

****** RELATIVE FUNCTION CONVERGENCE ******

FUNCTION	*-1.091169e+04*	*RELDX*	*8.382e-06*
FUNC. EVALS	*71*	*GRAD. EVALS*	*27*
PRELDF	*4.098e-11*	*NPRELDF*	*4.098e-11*

I	***FINAL X(I)***	***D(I)***	***G(I)***
1	*2.066426e-04*	*1.000e+00*	*-4.330e+00*
2	*5.319016e-01*	*1.000e+00*	*-4.996e-03*
3	*3.930037e-01*	*1.000e+00*	*-9.053e-03*

Call:
garch(x = y, order = c(1, 1))

Coefficient(s):

a0	*a1*	*b1*
0.0002066	*0.5319016*	*0.3930037*

Call: garch(x = d)
Model: GARCH(1,1)

Residuals:

Min	1Q	Median	3Q	Max
-4.69464	-0.62380	-0.01216	0.65214	5.78262

Coefficient(s):

	Estimate	Std. Error	t value	Pr(>|t|)	
a0	2.066e-04	1.406e-05	14.70	<2e-16	***
a1	5.319e-01	3.369e-02	15.79	<2e-16	***
b1	3.930e-01	2.254e-02	17.44	<2e-16	***

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Diagnostic Tests:Jarque Bera Test , data: Residuals

X-squared = 194.09, df = 2, p-value < 2.2e-16

Box-Ljung test

data: Squared.Residuals

X-squared = 24.473, df = 1, p-value = 7.536e-07

The fitted model for log returns of daily BSE Sensex close is as follows:

$$\begin{aligned} &\varepsilon_t = z_t\sqrt{h_t} \\ &\varepsilon_t \sim N(0,h_t),\ z_t \sim iidN(0,1), \\ &h_t = 0.0002066 + 0.5319016\varepsilon_{t-1}^2 + 0.3930037h_{t-1} \end{aligned} \tag{4.2}$$

Diagnostics were performed with the residuals from the two returns equations.

4.3 Diagnostics for the GARCH (1,1) Model

Once volatility clustering is confirmed, our focus is on determining the fitted GARCH model applicable to the return series. We first estimate the parameters, namely for the GARCH (1, 1) model. As the GARCH model is analogous with an ARMA model, often such processes can be approximated by an AR(p+q) model. Then the diagnostic measures proposed in the section 3.1 can be measured and the results of Q values obtained is displayed by the following plot for k=1, 2 and 3. From the graph, it is clear that at point t =1708 an outlier effect is present. As a result of this the 2 observations at the time point t=1707 and t=1708 has to be deleted. For replacing this value we can use one step ahead forecast of the model at the point t=1707 and two step ahead forecast at t=1708.

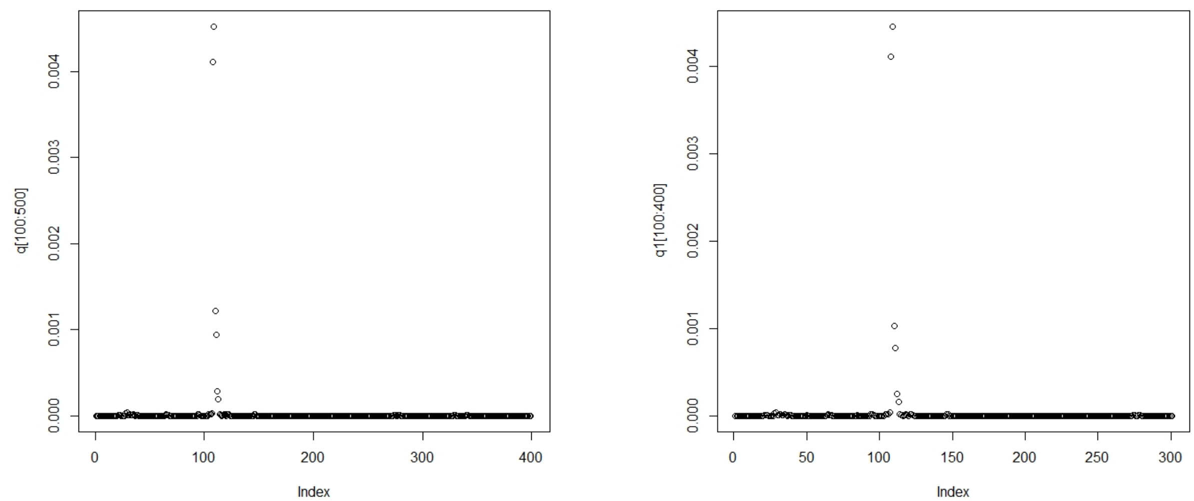

Figure 4.4 Q_{11} and Q_1 for squared log returns of daily close of BSE sensex.

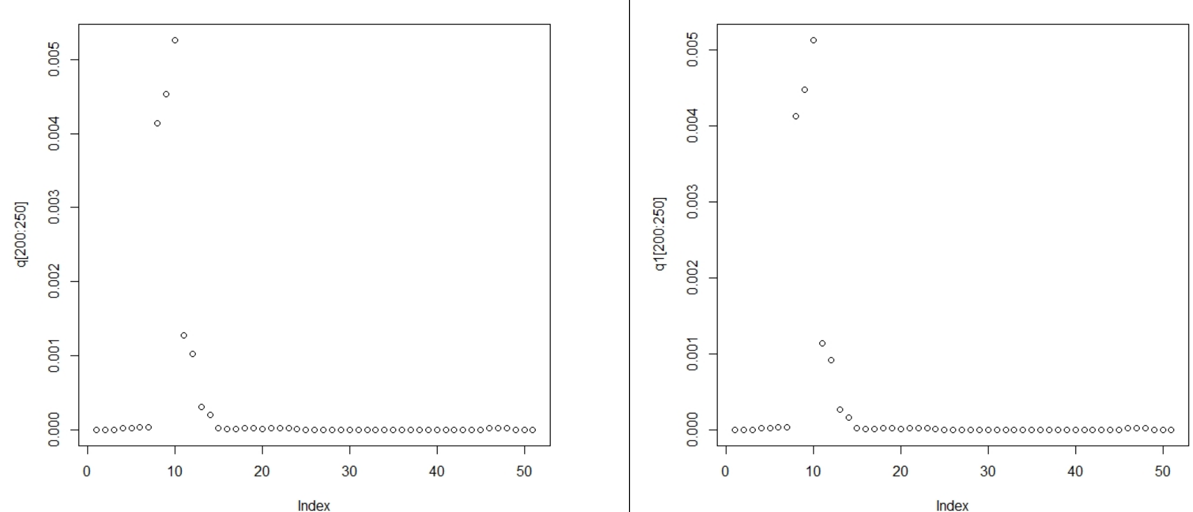

Figure 4.5 Q_2 and Q_{21} statistics for squared log returns of daily close of BSE sensex.

From the calculation of Q statistics, it is clear that there is a presence of additive outlier in the series and from diagnosing the original series, it is clear that the outlier occurred around the month, September 2001.

Using the observations up to t=1708, a garch model is fitted as

$$
\begin{aligned}
&\varepsilon_t = z_t \sqrt{h_t} \\
&\varepsilon_t \sim N(0, h_t),\ z_t \sim iidN(0,1), \\
&h_t = 0.0006870738 + 0.2111966304\varepsilon_{t-1}^2 + 0.6663586989 h_{t-1}
\end{aligned}
\tag{4.3}
$$

The prediction of the outlier values is done using this model at t=1708, 1709, 1710 and replaced the original values by this estimate and tested for outlier. But no further deviations in Q statistic is observed.

4.4 Conclusions

Using GARCH methodology, the present work evaluated the impact of outlier model in the conventional time series on volatility returns of Bombay Stock Exchange. Some regression diagnostic measures available in the time series setup are applied to the financial time series set up for detecting the extreme values. The simple GARCH (1, 1) model has been estimated for the stock exchange close value of BSE Sensex. In conclusion, the empirical results of this study indicated that there has been a change in the market environment since the year 2001, which is reflected as abnormality in the BSE Sensex from September 2001.

REFERENCES

1. Abraham, B., and Box, G. E. (1979). Bayesian analysis of some outlier problems in time series, *Biometrika*, **66(2)**, 229-236.
2. Andrews, D. F., and Pregibon, D. (1978). Finding the outliers that matter, *Journal of the Royal Statistical Society. Series B (Methodological)*, 85-93.
3. Balke, N. S., and Fomby, T. B. (1994). Large shocks, small shocks, and economic fluctuations: Outliers in macroeconomic time series, *Journal of Applied Econometrics*, **9(2)**, 181-200.
4. Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of econometrics*, **31(3)**, 307-327.
5. Bologna, P., and Cavallo, L. (2002). Does the introduction of stock index futures effectively reduce stock market volatility? Is the futures effect immediate? Evidence from the Italian stock exchange using GARCH, *Applied Financial Economics*, **12(3)**, 183-192.
6. Box, G. E., and Jenkins, G. M. (1976). Time series analysis, control, and forecasting, *San Francisco, CA: Holden Day*, **3226** (3228), 10.
7. Box, G. E., and Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems, *Journal of the American Statistical association*, **70** (349), 70-79.
8. Chang, I., Tiao, G. C., and Chen, C. (1988). Estimation of time series parameters in the presence of outliers, *Technometrics*, **30(2)**, 193-204.
9. Charles, A., and Darné, O. (2005). Outliers and GARCH models in financial data, *Economics Letters*, **86(3)**, 347-352.
10. Chen, C., and Liu, L. M. (1993). Joint estimation of model parameters and outlier effects in time series, *Journal of the American Statistical Association*, **88 (421)**, 284-297.
11. Cook, R. D., and Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression, *Biometrika*, **70(1)**, 1-10.

12. Draper, N. R., and John, J. A. (1981). Influential observations and outliers in regression, *Technometrics*, **23(1)**, 21-26.

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

14. Engle, R. F., and Bollerslev, T. (1986). Modelling the persistence of conditional variances, *Econometric reviews*, **5(1)**, 1-50.

15. Fox, A. J. (1972). Outliers in time series, *Journal of the Royal Statistical Society. Series B (Methodological)*, 350-363.

16. Franses, P. H., and Ghijsels, H. (1999). Additive outliers, GARCH and forecasting volatility, *International Journal of Forecasting*, **15(1)**, 1-9.

17. Martin, R. D. (1980). Robust estimation of autoregressive models, *Directions in time series*, **1**, 228-262.

18. Martin, R. D., and Samarov, A. Vandaele (1983).Robust methods for ARIMA models, *Applied Time Series Analysis of Economic Data*, 153-169.

19. Martin, R. D., and Yohai, V. J. (1986). Influence functionals for time series, *The annals of Statistics*, 781-818.

20. Priyanka, K. (2014). In dalal street investment journal, Top Five Performing S&P Sectoral Index since January 1, 2013. Retrieved 11:23, September 23, 2014 from http://www.dsij.in/article-details/articleid/9855/top-five-best-performing-s-p-bse-sectoralindex-since-january-1-2013.aspx.

21. Sheelapriya, G., and Murugesan, R. (2014). Asian Journal of Empirical Research, *Asian Journal of Empirical Research*, **4 (11)**, 503-513.

22. Tiao, G. C., and Box, G. E. (1981). Modeling multiple time series with applications, *Journal of the American Statistical Association*, **76 (376)**, 802-816.

23. Tiao, G. C., Box, G. E. P., and Hamming, W. J. (1975). Analysis of Los Angeles photochemical smog data: a statistical overview, *Journal of the Air Pollution Control Association*, **25(3)**, 260-268.

24. Tsay, R. S. (1986). Time series model specification in the presence of outliers, *Journal of the American Statistical Association*, **81(393)**, 132-141.

25. Tsay, R. S. (1988). Outliers, level shifts, and variance changes in time series, *Journal of forecasting*, **7(1)**, 1-20.

26. Tsay, R. S. (2005). *Analysis of financial time series* (Vol. 543), John Wiley & Sons.

27. Van Dijk, D., Franses, P. H., and Paap, R. (2002). A nonlinear long memory model, with an application to US unemployment, *Journal of Econometrics*, **110 (2)**, 135-165.

28. Verhoeven, P., and McAleer, M. (2004). Fat tails and asymmetry in financial volatility models, *Mathematics and Computers in Simulation*, **64(3)**, 351-361.

29. Verhoeven, P., Pilgram, B., McAleer, M., and Mees, A. (2002). Non-linear modelling and forecasting of S&P 500 volatility, *Mathematics and Computers in Simulation*, **59(1)**, 233-241.

30. Wold, H. (1938). *A study in the analysis of stationary time series* (Doctoral dissertation, Almqvist & Wiksell).

31. Yule, G. U. (1921). On the time-correlation problem, with especial reference to the variate-difference correlation method, *Journal of the Royal Statistical Society*, **84(4)**, 497-537.

32. Yule, G. U. (1927). On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers, *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, **226**, 267-298.

33. Yule, G. U. (1927). On reading a scale, *Journal of the Royal Statistical Society*, **90(3)**, 570-587.