Detecting Level Change Outliers (LC) in GARCH (1, 1) Processes

MOHAMMAD SAID ZAINOL\textsuperscript{1}, SITI MERIAM ZAHARI\textsuperscript{1}, KAMARUZAMAN IBRAHIM\textsuperscript{3},
AZAMI ZAHARIM\textsuperscript{2}, KAMARUZAMAN SOPIAN\textsuperscript{3}

\textsuperscript{1}Center of Studies for Decision Sciences, Faculty Computer Science and Mathematics, Universiti Teknologi MARA 40450 Shah Alam, MALAYSIA.
\textsuperscript{2}Head Centre for Engineering Education Research, Faculty of Engineering and Built Environment, 43600 UKM Bangi, MALAYSIA
\textsuperscript{3}Solar Energy Research Institute, Faculty of Engineering and Built Environment, 43600 UKM Bangi, MALAYSIA

said@tmsk.uitm.edu.my, smbz_ma@yahoo.com, azami.zaharim@gmail.com,

Abstract: This paper reports the results of a study of outlier detection in time series data for the outlier of level change (LC) type. The main objective is to derive a test statistic for detecting LC in GARCH(1,1) processes. Subsequently a procedure for testing the presence of outliers using the statistics was developed. In the derivation of the statistics, the method applied was based on an analogy of GARCH(1,1) as being equivalent to ARMA(1,1) for the residuals $\varepsilon_t^2$. Because of the difficulty in determining the sampling distributions of the outlier detection statistics, critical regions were estimated through simulations. The developed outlier detection procedure was applied for testing the presence of LC outliers in the daily observations of the Index of Consumer Product Price (ICP) for the period 1990 to 2005. Over the period, the results indicate that LC outlier occurred in year 1998.

Keywords: LC outlier; GARCH; simulation; least squares method

1 Introduction

This paper reports the results of a study to detect outliers of the level change (LC) type in GARCH(1,1) processes. Generally an outlier is a data point located “far away” from the rest. Outliers can also be an ‘extreme value’, or an ‘extraordinary’ value relative to the variance of the data set. Issues in outliers have raised a lot of interest in time series literature (see, \textit{inter alia}, Hawkins, 1980; Gelman \textit{et al}, 1995; Mann, 2001, Pena, 2001; Balke and Fomby, 1994). The presence of these “strange” observations in a data set raises questions regarding its reliability. Further, their presence has been shown to cause estimation problems (Chang \textit{et al}., 1988; Chen and Liu, 1993; Tsay \textit{et al}., 2000; Franses and van Dijk, 2000; Berkoun \textit{et al}., 2003). Serious errors may occur if a time series contaminated by outliers is used to estimate a forecasting model. The estimated parameters in the model may be severely distorted.

Since the 1970s, there have been numerous researches on outlier handling procedures, mostly involving outliers in autoregressive (AR) and autoregressive moving average (ARMA) processes. Fox (1972) introduced formal definitions of types of outliers in AR processes. This later led to the growth of outlier studies in other processes, including, ARIMA, bilinear, ARCH, and GARCH.

1.1 Effects of Outlier on Data Analysis: A Brief Literature Review

The presence of outliers can have deleterious effects on statistical data analysis. Outlier could affect model identification, estimation, forecasting, diagnosing and testing. Outliers may distort estimates of residual variances (Hogg, 1979 and Martin, 1980), affect estimates of variance (Pena, 1990) and not only increases residuals but also distort model specification and parameter estimates (Abraham and Chuang, 1989).

Outlier may also increase error variance and reduce power of statistical tests (Franses \textit{et al}., 2004), decrease normality and bias or influence estimates (Schwager and Margolin, 1982;
Further, outliers have also been found to distort the estimated spectra when a time series spectrum is analyzed (Kleiner et al., 1979). Chernick et al. (1982) noted that outliers may cause severe effects on the estimation of correlation coefficient and spectral density. Outliers have also been found to affect estimates of intraclass correlation coefficient (Giraudeau and Chastang, 1999) and autocorrelation (ACF) and partial autocorrelation (PACF) estimations. Deutsch et al. (1990) studied outlier effects on ARMA model identification and found that at the identification stage, the effect of outliers was confusing in the way it affected the correct identification of the lag length of the model.

Outliers have been found to affect forecasts by increasing the estimated variance, making the forecast ambiguous. Trivez (1993) found that level change (LC) outliers and temporary change (TC) outliers affect forecasting estimates by increasing the width of the prediction intervals. Other studies of forecasting effects of outliers include those of Ledolter (1989), Hotta (1993) and Chen and Liu (1993b).

1.2 Outlier Representation
To algebraically model the presence of outliers in a time series, consider an ARMA process. An ARMA process “uncontaminated” by an outlier is represented as:

\[ \phi(L)Y_t = \theta(L)e_t ; \quad t = 1, 2, ..., n \]  

where \( n \) is the number of observations, \( \phi(L) \) and \( \theta(L) \) are lag polynomials with all roots outside the unit circle, \( L \) is the backshift operator, and \( \{e_t\} \) is a sequence of identically and independently distributed (IID) random variables with zero mean and constant variance.

In the presence of an outlier, the process is ‘contaminated’ and is represented by

\[ Y_t^* = Y_t + \xi(L)\omega I_t(\tau) \]  

where

\( Y_t^* \) is the observed contaminated series,
\( Y_t \) is the uncontaminated series,
\( \omega \) is the outlier effect,
\( \tau \) is the time of occurrence of the outlier,
\( \xi(L) \) is the dynamic pattern of the outlier,
\( I_t(\tau) \) is an indicator function for the occurrence of outlier effect and is defined as:

\[
I_t(\tau) = \begin{cases} 
0 & \text{if } t \neq \tau \\
1 & \text{if } t = \tau 
\end{cases}
\]

Box and Tiao (1965) introduced the LC type of outlier. It is called as level change since its behaviour is to change the level or mean of the observed series after the first effect and this outlier creates a sudden and permanent change in the observed series. In a simple equation, the changes can be denoted as a change from:

\[ Y_t = \phi(L)e_t \]

to

\[ Y_t = \frac{\phi(L)}{\theta(L)}e_t + \sigma_{LC} \]  

(1.3)

\( \sigma_{LC} \) can be positive or negative. The dynamic pattern \( \xi(L) \) for level change outlier is equal to \( \frac{1}{(1-L)} \). Thus, an LC model can be written as:

\[ Y_t^* = Y_t + \frac{1}{(1-L)}\omega I_t(\tau) \]  

(1.4)

2 Effect of LC in GARCH(1,1) Processes
Consider a random walk process \( y_t = y_{t-1} + \epsilon_t \). Let the GARCH process have the corresponding return series \( (1-L)y_t = \epsilon_t \) as its mean equation. The corresponding GARCH(1,1) process is then

\[(1-L)y_t = \epsilon_t = z_t\sqrt{h_t} , \quad \epsilon_t | \varphi_{t-1} \sim N(0,h_t) \]  

(2.1)

\[ h_t = \omega + \alpha \epsilon_{t-1}^2 + \beta h_{t-1} \]  

(2.2)

Equation (2.2) is the conditional variance equation of the GARCH model and \( \{z_t\} \) is a sequence of identically and independently distributed (IID) random variables with zero-mean and constant variance:

\[ E(z_t) = 0 \]
\[ E(z_t^2) = 1 \]
The method of Chen and Liu (1993) to derive outlier detection statistics for ARMA processes is now extended to GARCH(1,1) by taking the analogy of GARCH(1,1) as being equivalent to ARMA(1,1) for the $\varepsilon_i^2$, $\varepsilon_i$ being the residuals. This approach is motivated by Franses and Ghijsels (1999).

Following Bollerslev (1986) the variance component of the GARCH(1,1) process is assumed to have an ARMA(1,1) in $\varepsilon_t$ and thus, (2.3)

$$2\beta + \alpha - 1 = 0$$

Substituting equation (2.2) into (2.3) yields

$$v_t = \varepsilon_i^2 - h_i$$

Equation (2.3) is the outlier-free model where $\varepsilon_i$ is the outlier-free series and $v_t$ is the outlier-free residual, respectively and $t = 1, 2, \ldots$ When an outlier is present, the GARCH (1,1) representation is modelled as:

$$\varepsilon_i^{*2} = \varepsilon_i^2 + \xi(L)\sigma_{LC}I_i(\tau)$$

where, $\varepsilon_i^{*}$ represent the “contaminated” counterpart of $\varepsilon_i$, and $\sigma$ denotes the outlier effect. $I_i(\tau)$ is an indicator function for the occurrence of outlier and it takes a value of 1 at $t = \tau$ and zero otherwise with $\tau$ denoting the time point at which an outlier is present. In general, the point is unknown and in this study it is assumed that $\tau$ is unknown.

$$I_i(\tau) = \begin{cases} 1 & t = \tau \\ 0 & \text{otherwise} \end{cases}$$

The polynomial $\xi(L)$ represents the dynamics of the outlier effect. The dynamic pattern of outlier effect for LC was shown by Chen and Liu (1993) to be

$$\xi(L) = \frac{1}{1-L}$$

This type of outlier gives permanent change on the series at time $t = \tau$ and also on the subsequent observation $t > \tau$. It can be shown as:

$$\varepsilon_{i,LC}^{*2} = \varepsilon_i^2 + \xi(L)\sigma_{LC}I_i(\tau)$$

In general, an LC affects a GARCH(1,1) process in the following manner:

$$\varepsilon_{i,LC}^{*2} = \begin{cases} \varepsilon_i^2 + \sigma_{i,LC} & t < \tau \\ \varepsilon_i^2 + \sigma_{i,LC}I_i(\tau) & t = \tau \\ \varepsilon_i^2 + \sigma_{i,LC} & t = \tau + k \end{cases}$$

(2.7)

Assume that an LC occurs at $t = \tau$. Following equation (2.6) the relationship between the original and the contaminated residual then becomes

$$v_t^{*} = \pi(L)\sigma_{LC}\xi(L)I_i(\tau) + \varepsilon_t$$

In general, the effect of LC on the residuals of GARCH(1,1) variance equation can be written as:

$$v_{t,LC}^{*} = \begin{cases} v_t & t < \tau \\ v_t + \sigma & t = \tau \\ v_t + \left(1 - \sum_{i=1}^{k} \alpha_i \beta_i^{t-i}\right)\sigma \ \text{and,} \ t = \tau + k \end{cases}$$

(2.8)

where $k = 1, 2, \ldots, n - \tau$

### 3 The Outlier Effect and Test Statistic

The least squares estimate of the outlier effect for LC at $t = \tau$ is

$$\widetilde{\delta}_{LC} = \frac{\sum_{t=0}^{\tau-1} v_{t+k,LC}X_{k,LC}}{\sum_{k=0}^{n-\tau} X_{k,LC}^2}$$

(3.1)

where
The derived statistic for testing for the presence of an LC type outlier is
\[
\hat{t}_{LC}(\tau) = \left( \frac{\hat{\sigma}_{LC}(\tau)}{\hat{\sigma}_v} \right) \left( \sum_{k=0}^{n-\tau} \hat{x}_k^2 \right)^{1/2}
\]
\[
(3.2)
\]
where \( \hat{\sigma}_v = 1.483 \times \text{median} \| \hat{\epsilon}_i - \hat{\mu} \| \) is the residual standard deviation computed using the median absolute deviation (MAD) with \( \hat{\mu} \) being the median of the estimated residuals (Hampel et al., 1986). Because of the complicated nature of the distribution of the test statistic, their critical values for testing for the presence of LC were estimated by simulation.

### 4 Procedures for Testing for LC

The LC outlier is detected using the following steps:

(i) Estimate an initial GARCH(1,1) model

(ii) Obtain the estimate of the conditional variance, \( \hat{\gamma}_t \) and hence, \( \hat{\epsilon}_i = \hat{\epsilon}_i - \hat{\mu}_t \)

(iii) Compute the outlier effect and test statistic for all possible \( \tau = 1, 2, \ldots, n \)

(iv) If the value of the test statistic is less than the critical value, there is no outlier at the observation for which the t-statistic of \( \hat{\sigma}(\tau) \) is a maximum (in absolute value). An outlier is present at the time point where \( \hat{\sigma}(\tau) \) is maximum (in absolute value) if the value of the test statistic is greater than the estimated critical value.

### 5 Application: Detection of Presence of LC in Consumer Product Price Index (ICP)

Consumer product price index was collected from August 1\(^{st}\), 1996 to August 29\(^{th}\), 2003. The GARCH(1,1) process representing the returns of the ICP series was estimated as:
\[
y_i = 3.7263 \times 10^{-5} + \epsilon_i
\]
\[
h_i = 8.3533 \times 10^{-7} + 7.5945 \times 10^{-2} \epsilon_{i-1}^2
\]
\[+ 9.2406 \times 10^{-1} h_{i-1}
\]
\[8.1\]
Figure 5.1 is a plot of the returns series of ICP. The outlier detection procedure was applied. Three points were suspected to contain LC outliers. After the testing procedure was applied it was found that LC occur at location 546, with the value of test statistics of 101.84.

A summary of the test statistic of the LC detection on the returns of ICP is exhibited in Table 5.1. Results from Table 8.1 indicates the presence of a LC outlier during this period, which falls in September, 1998. This was due to the economic recession in the ASEAN region. The economy was severely affected between the year 1997 to 1998.

### Table 5.1: Summary of the LC detection

<table>
<thead>
<tr>
<th>Data</th>
<th>Date</th>
<th>Location</th>
<th>Test statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICP</td>
<td>3/9/1998</td>
<td>546</td>
<td>101.84</td>
</tr>
</tbody>
</table>

### References


