An Evaluation of Test Statistics for Detecting Level Change in BL(1,1,1,1) Models

[1]IBRAHIM MOHAMED, [2]AZAMI ZAHARIM, [3]SHAHRUM ABDULLAH AND [4]MOHD. SAHAR YAHYA
[1]Institute of Mathematical Sciences, University of Malaya,
50603 Kuala Lumpur
MALAYSIA
[2,3] Faculty of Engineering, National University of Malaysia,
43600 UKM Bangi, Selangor
MALAYSIA
[4]Centre for Foundation Studies in Sciences, University of Malaya,
50603 Kuala Lumpur
MALAYSIA

Abstract: - A study is carried out to investigate the sampling properties of the outlier test statistics of a procedure developed for detecting level change in BL(1,1,1,1) processes. It is done with respect to the sample size, the type of outlier and the size of the coefficients of the BL(1,1,1,1) process. The results show that, in general, the outlier detection procedure is capable of detecting level change, although the performance is affected if ω is large.

Key-Words: - Level change, bilinear process, outlier test statistics, outlier detection procedure, sampling properties.

1 Introduction

Observations that deviate from the rest of the observations exist frequently in time series data. These observations are known by different names, such as, "outliers", "contaminants", "discordant observations" and "extreme values". In 1983, Beckman and Cook [1] defined discordant observation as any observation that appears discrepant to the investigator, while a contaminant was defined as any observation that is not a realization of a target distribution. On the other hand, an outlier is just a collective name referring to either a contaminant or a discordant observation. The results of studying outliers can be used, among others, as diagnostic tools to test the strength and weakness of a model, to accommodate outliers in order to make inferences about a parameter, to improve the model and to examine their influence on a model. In general, studies on four types of outliers; additive outlier (AO), innovational outlier (IO), level change (LC) and temporary change (TC), had been mentioned in the literature. In this paper, only the LC case will be considered.

Outliers may happen in nonlinear time series data. In 2005, Mohamed [2] had looked into the occurrence of LC in time series data generated by BL(1,1,1,1) process and a test statistics had been derived. In the subsequent sections, the sampling properties of this test statistics are investigated.

2 Statistical Models for Computational Simulation

2.1 Bilinear Models

Biliner model was first documented by Granger and Andersen [3]. The general bilinear model, denoted by BL(p,q,r,s), is given by

$$Y_t = \sum_{i=1}^{p} a_i Y_{t-1} + \sum_{j=1}^{q} c_j e_{t-j} + \sum_{k=1}^{r}\sum_{\ell=1}^{s} b_{k\ell} Y_{t-k} e_{t-\ell} + e_t \tag{1}$$

where a_i, c_j and $b_{k\ell}$ are constant, and e_t's are assumed to follow normal distribution with mean zero and precision τ, $\tau > 0$. The first two components on the right-hand side of (1) are basically the ARMA model with parameters p and q. The second last component is nonlinear which helps to explain the nonlinearity characteristic of the data being modeled. Thus, ARMA (p,q) is a special case of the BL(p,q,r,s) when $r = s = 0$. The BL(1,1,1,1), can be deduced from equation (1) giving

$$Y_t = a\,Y_{t-1} + c\,e_{t-1} + b\,Y_{t-1}\,e_{t-1} + e_t \tag{2}$$

2.2 Measure of Outlier Effect

Let Y_t^* and e_t^* be the contaminated observed values and contaminated residuals obtained when an outlier exists in the data. In order to detect an outlier, the contaminated residuals are examined and given by

$$e_t^* = Y_t^* - \left(a + b\,e_{t-1}^*\right)Y_{t-1}^* - c\,e_{t-1}^* \tag{3}$$

The equation can be rewritten as

$$e_t^* = Y_t^* - \left(a + b\,e_{t-1}\right)Y_{t-1}^* - b\left(e_{t-1}^* - e_{t-1}\right)Y_{t-1}^* - c\,e_{t-1}^* \tag{4}$$

The formulation and symbols discussed here will be retained throughout the paper. Mohamed [2] had derived the measure of outlier effect using the least squares method, denoted by $\hat{\omega}$. The formulation is given by

$$\hat{\omega}_{LC} = \frac{\sum_{k=0}^{n-d} e_{d+k,LC}^*\,A_{k,LC}}{\sum_{k=0}^{n-d} A_{k,LC}^2} \tag{5}$$

where

$$A_{k,LC} = \begin{cases} 1 & k=0 \\ 1-\left(a+be_{d+(k-1)}\right)-\left(bY_{d+(k-1),LC}^*+c\right)A_{k-1,LC} & k=1,2,\ldots \end{cases} \tag{6}$$

and δ is the decaying factor.

2.3 The Outlier Detection Procedure

The mean and variance of the measures given by equations (5) can be obtained using the bootstrap method [4,5](Efron and Thibshirani [1986], Efron [1993]) as follows:

$$\tilde{\sigma}_{BS} = \left\{ \frac{\sum_{M=1}^{B}\left(\tilde{\omega}(M) - \overline{\tilde{\omega}}_{BS}\right)^2}{(B-1)} \right\}^{1/2} \tag{7}$$

where

$$\overline{\tilde{\omega}}_{BS} = B^{-1}\sum_{M=1}^{B}\tilde{\omega}(M). \tag{8}$$

Hence, the test criteria used for detecting individual outlier is

$$\hat{\tau}_t = \frac{\left(\hat{\omega}_t - \overline{\tilde{\omega}}_{BS,t}\right)}{\tilde{\sigma}_{BS,t}} \tag{9}$$

In general, the time point where an outlier occurs is unknown. Alternatively, the test statistics, $\hat{\tau}_t$, can be calculated at every time point t, t = 1, 2, ..., n. Hence, the determination of type of outlier is carried out using the following test criteria:

$$\hat{\eta} = \max_{t=1,\ldots,n}\left\{\left|\hat{\tau}_t\right|\right\}. \tag{10}$$

That is, the maximum values of the test statistics are examined in order to identify the a particular point t in BL(1,1,1,1) model where a LC occurs. The full steps are described below:

(a) Compute the least squares estimates of BL(1,1,1,1) model based on the original data. Hence, residuals of models can be obtained.
(b) Compute $\hat{\tau}_t$ for every *t, t* = 1, 2, ..., n using the residuals obtained in Stage one.
(c) Let $\hat{\eta} = \max_{t=1,\ldots,n}\left\{\left|\hat{\tau}_t\right|\right\}$. Given a pre-determined critical value C, if $\eta_t = \left|\tau_t\right| > C$, then a LC occurs at time point t.

Similar approach had been used by several authors such as Chen and Liu [6] on detecting outliers in the ARIMA model.

3 Results and Discussion

3.1 Sampling Behaviour of the Test Statistics

The simulation study in this section is carried out in order to investigate the sampling properties of the maxima of the outlier test statistics. It is associated with the sample size, and the coefficients chosen for BL(1,1,1,1)

Models 1 – 6, which is given in Table 1, are considered. They represent a broad choice of coefficients of BL(1,1,1,1) models. For instance, the coefficients of model 1 are all negative whereas the coefficients of model 2 are all positive. On the other hand, the coefficients of models 3 – 5 are the same in magnitude but with different set of signs. Model 6 has mixed signs with large coefficient of AR term.

For each model, three cases of sample size are examined, n = 60, 100 and 200. The random errors, e_t's, are assumed to follow standard normal distribution. For each model and each sample size, 500 series are generated. The test statistics for LC are calculated separately based on equations (10). The focus is to examine the sampling behaviour of $\eta = \max_{t=1,2,\ldots,n}\left\{\left|\tau_t\right|\right\}$. In particular, the percentiles of the test statistics at the 1%, 5% and 10% levels are estimated when no outlier is present in the series. There are three plots given in Figure 1 displaying percentiles for each of 6 models and 3 sample sizes.

Table 1. List of models

Model	Full model
1	$0.3Y_{t-1} - 0.3e_{t-1} - 0.3Y_{t-1}e_{t-1} + e_t$
2	$2Y_{t-1} + 0.2e_{t-1} + 0.2Y_{t-1}e_{t-1} + e_t$
3	$1Y_{t-1} + 0.1e_{t-1} - 0.2Y_{t-1}e_{t-1} + e_t$
4	$0.1Y_{t-1} + 0.1e_{t-1} + 0.2Y_{t-1}e_{t-1} + e_t$
5	$1Y_{t-1} - 0.1e_{t-1} + 0.2Y_{t-1}e_{t-1} + e_t$
6	$0.5Y_{t-1} + 0.1e_{t-1} - 0.2Y_{t-1}e_{t-1} + e_t$

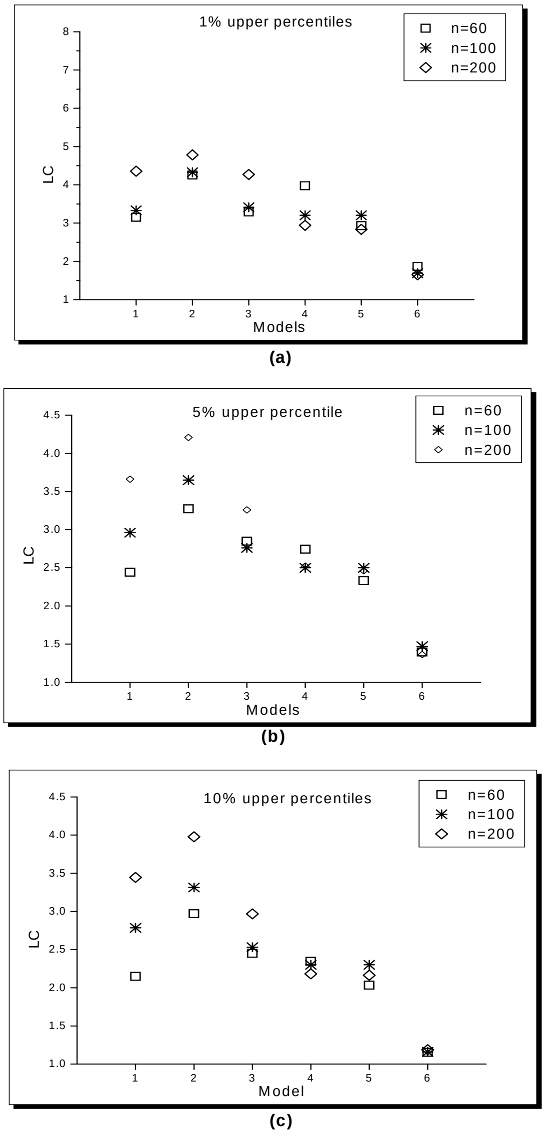

Figure 1 Plot of critical values of LC

The estimates for models 1 – 2 show a trend of an increasing function in sample size, n. However, the trend is not clear for models 3 – 5. The estimated percentiles for model 6 are smaller than that of the other models and not affected by sample size. Model 6 is different from the other models in its coefficients values. It has a large coefficient value of AR term (a = –0.5) if compared to other models. The estimated 5% percentiles for LC range from 1.36 to 4.21. In general, the increase of estimates is moderate with the exception of model 6. Based on the results, critical values of 2.5 to 4.0 seem to be appropriate for a series with length of 60 to 200. In practice, more than one critical value is recommended for the analysis.

3.2 Performance of the Outlier Detection Procedure

Interest here is to investigate the performance of the outlier detection, referred as outlier detection procedure herewith, through simulation work. The test criteria are applied to cases characterized by a combination of the following factors:

a) One underlying BL(1,1,1,1) models but with different combinations of coefficients as given in Table 1.
b) A single LC at t = 40 in samples of size 100.
c) Three different values of outlier effect, ω = 6, 8, 10.
d) Five different levels of critical values are chosen: 2, 2.5, 3, 3.5, 4.

Series are generated to contain a LC. The standard deviation of the noise process for each model is set to be unity. For a given model, 500 series of length 100 are generated using the rnorm procedure in S-Plus. The relative frequency or proportion of correct detection is reported. The correct detection is defined as a correct identification of both type and location of an outlier. For example, when a LC is included at time t = 40, outlier detection procedure is applied on the series to see whether a LC is correctly detected.

The results are shown in Table 2 (in the last page of this paper). It is observed that the proportion of correctly detecting LC is high if critical values of 2 or 2.5 are used. However, the proportions do not follow any general pattern of increasing or decreasing in ω. In several cases, the test criterion performs better when ω = 6 than when ω = 8 or 10; for instance, model 1 and 6. The estimation of BL(1,1,1,1) model is affected by LC especially when larger ω is used. This subsequently affects the performance of the outlier detection procedure. It is also observed that $\hat{\eta}$ tend to take smaller values than 4; for instance, for model 3 with $\omega = 6$, 60%

of $\hat{\eta}$ that detect LC correctly are greater then 3.0, 42% of $\hat{\eta}$ are greater than 3.5 and only 24% of $\hat{\eta}$ are greater than 4.

4 Conclusion

In this paper, the performance of the outlier detection procedure for LC is studied through simulation works. Results show that the performance of the procedure for detecting LC is better when the magnitude of outlier effect is not too large. In general, the procedure performs well.

References:

[1] RJ. Beckman, RD. Cook, Outlier .... s, *Technometrics*, Vol. 25, 1983, pp. 119-149.

[2] IB. Mohamed, Outliers in Bilinear Time Series Model, *PhD Thesis, Universiti Teknologi MARA (Malaysia)*, 2005.

[3] CWJ. Granger, AP. Andersen, Introduction to Bilinear Time Series Models. *Gottinge: Vandenhoeck and Ruprecht*, 1978.

[4] B. Efron, R. Tibshirani, Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, *Statistical Science 1*, Vol. 1, 1986, pp. 54-77.

[5] B. Efron, *An Introduction to the Bootstrap*. Chapman and Hall, 1993.

[6] C. Chen, LM. Liu, Joint estimation of model parameters and outlier effects in time series, *Journal of American Statistical Society*, Vol. 88, 1993, pp. 284-297.

Table 2. Proportion of correctly detecting LC using the outlier detection procedure

Model (ω)	Critical values				
	2	2.5	3	3.5	4
Model 1 $(\omega = 6)$	0.99	0.98	0.96	0.93	0.87
Model 1 $(\omega = 8)$	1.00	0.97	0.95	0.90	0.83
Model 1 $(\omega = 10)$	1.00	0.99	0.95	0.88	0.78
Model 2 $(\omega = 6)$	0.98	0.69	0.39	0.20	0.11
Model 2 $(\omega = 8)$	0.98	0.63	0.33	0.20	0.10
Model 2 $(\omega = 10)$	0.97	0.58	0.32	0.19	0.09
Model 3 $(\omega = 6)$	1.00	0.80	0.60	0.42	0.24
Model 3 $(\omega = 8)$	0.99	0.84	0.64	0.41	0.29
Model 3 $(\omega = 10)$	1.00	0.74	0.50	0.29	0.16
Model 4 $(\omega = 6)$	1.00	0.99	0.93	0.80	0.55
Model 4 $(\omega = 8)$	1.00	0.94	0.78	0.60	0.43
Model 4 $(\omega = 10)$	1.00	0.91	0.77	0.55	0.35
Model 5 $(\omega = 6)$	0.99	0.95	0.84	0.71	0.51
Model 5 $(\omega = 8)$	1.00	0.94	0.80	0.60	0.44
Model 5 $(\omega = 10)$	1.00	0.86	0.76	0.52	0.37
Model 6 $(\omega = 6)$	1.00	0.98	0.95	0.91	0.85
Model 6 $(\omega = 8)$	0.99	0.97	0.94	0.87	0.74
Model 6 $(\omega = 10)$	1.00	0.95	0.90	0.80	0.63