Detection Procedure for A Single Outlier in a Bilinear Model

[1]Azami Zaharim, [2]Mohammad Said Zainol, [3]Ibrahim Mohamed, [4]Mohd Sahar Yahaya

[1]Unit Asas Kejuruteraan, Fakulti Kejuruteraan, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor.
[2]Jabatan Sains Kuantitatif, Fakulti Teknologi Maklumat dan Sains Kuantitatif, Universiti Teknologi MARA, 40450 Shah Alam. [3]Institut Sains Matematik, Universiti Malaya, 50603 Kuala Lumpur
[4]Pusat Asasi Sains, Universiti Malaya, 50603 Kuala Lumpur

ABSTRACT

A single outlier detection procedure for data generated from BL(1,1,1,1) models is developed. It is carried out in three stages. Firstly, the measure of impact of an IO, AO, TC and LC, denoted by ω_{IO}, ω_{AO}, ω_{TC} and ω_{LC}, respectively are derived based on least squares method. Secondly, test statistics and test criteria are defined for classifying an observation as an outlier of its respective type. Finally, a general single outlier detection procedure is presented to distinguish a particular type of outlier at a time point t.
Keywords: Bilinear models, outlier, impact measure, test statistics.

1. Introduction

Granger and Anderson [1978] presented various types of bilinear processes with discussions on their invertibility and stationarity properties. Bilinear models were shown to perform well in comparison to linear model when applied to the Wőlfer sunspot data and the IBM daily common stock closing prices available in Box and Jenkins [1976]. An interesting feature of bilinear processes is the fact that they are merely an extension of the linear ARMA processes as well as being a simplified case of nonlinear Volterra series expansions (Weiner [1958]).

While most studies on detection of outliers are for cases in linear time-series processsess, Chen [1997] and Zaharim [1996] studied the detection in bilinear processes. The former used the Gibbs sampling method for general bilinear model, but only dealt with the additive outliers. On the other hand, the latter used the least squares method for simple bilinear model to detect four type of outliers, the additive outlier (AO), innovational outlier (IO), temporary change (TC) and level change (LC). This article reports a study in which the work of Zaharim [1996] is extended to the bilinear BL(1,1,1,1) case. The detection procedure employed is shown to perform well in detecting and identifying the most probable type of outliers.

2. Methodology

A BL(1,1,1,1) process for y_t is represented by:

$$y_t = \phi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \beta_1 y_{t-1} \varepsilon_{t-1} \tag{1}$$

where φ_1, θ_1 and β_1 are any real number satisfying the stationary condition of the process, whereas ε_t is the Gaussian white noise. The first three components on the right-hand side of (1) constitute an ARMA (1,1) process. The last component is the portion which represents the nonlinearity characteristic of the data being modeled.

2.1 Outlier Effect Measures

The nonlinear least squares estimation method as found in Priestly [1991] is extended for use in the estimation of the parameters of a BL(1,1,1,1) model. In order to detect outliers, the measures of outlier effect are derived by minimizing the sum of squares of the residuals $S = \sum_{t=1}^{n} e_t^2$. The following measures of outlier effect $\hat{\omega}$ were obtained:

Additive Outlier (AO):

$$\hat{\omega}_{AO} = \frac{\sum_{k=0}^{n-d} \left[\{-1\}^k \, e_{d+k} \, A_k \right]}{\sum_{k=0}^{n-d} A_k^2} \qquad \text{(outlier occurring at time t = d)} \qquad (2)$$

where $A_k = \begin{cases} 1 & k = 0 \\ (a + be_d) + (bY_d + c) & k = 1 \\ (bY_{d+(k-1)} + c)A_{k-1} & k \geq 2 \end{cases}$ with an outlier occurring at time t = d

Innovational Outlier (IO):

$$\hat{\omega}_{IO} = \frac{\sum_{k=0}^{n-d} \left[\{-1\}^k \, e_{d+k} \, A_k \right]}{\sum_{k=0}^{n-d} A_k^2} \qquad (3)$$

where $A_k = \begin{cases} 1 & k = 0 \\ (bY_d + c) & k = 1 \\ (bY_{d+(k-1)} + c)A_{k-1} & k \geq 2 \end{cases}$

Temporary Change (TC):

$$\hat{\omega}_{TC} = \frac{\sum_{k=0}^{n-d} e_{d+k} \, A_k}{\sum_{k=0}^{n-d} A_k^2} \qquad (4)$$

where $A_k = \begin{cases} 1 & k = 0 \\ \delta^k - \delta^{k-1}\left(a + be_{d+(k-1)}\right) - \left(bY_{d+(k-1)} + c\right)A_{k-1} & k \geq 1 \end{cases}$

Level Change (LC):

$$\hat{\omega}_{LC} = \frac{\sum_{k=0}^{n-d} e_{d+k} \, A_k}{\sum_{k=0}^{n-d} A_k^2} \qquad (5)$$

Where $A_k = \begin{cases} 1 & k = 0 \\ 1 - \left(a + be_{d+(k-1)}\right) - \left(bY_{d+(k-1)} + c\right)A_{k-1} & k \geq 1 \end{cases}$

The complexity of the above formulations makes the task of finding an exact expression for the variance of each $\hat{\omega}$ measure difficult. To circumvent the difficulty a bootstrap method is employed to obtain the bootstrap mean, $\overline{\tilde{\omega}}_{TP,BS,t}$, and standard deviation, $\tilde{\sigma}_{TP,BS,t}$, at time t where TP is either AO, IO, TC or LC. The test statistics, $\hat{\tau}_{TP,t}$, then are the standardized measure of each outlier effect, $\hat{\omega}_{TP}$, using the bootstrap mean and standard deviation.

In reality, the time point where an outlier occurs is unknown. The test statistics, $\hat{\tau}_{TP,t}$, for AO, IO, TC and LC are calculated at every time point t, $t = 1, 2, ..., n$. To determine the type of outlier a test is carried out using the following criteria:

$$\hat{\eta}_{TP} = \max_{t=1,...,n}\left\{\left|\hat{\tau}_{TP,t}\right|\right\}. \tag{7}$$

In order to identify the type of outlier the maximum value of the test statistic $\hat{\tau}_{TP,t}$ for each type of outliers are examined

2.2   Identification of Outlier

The procedure for identifying the type of outlier at a particular point t in BL(1,1,1,1) model begins with modeling the original time series by assuming that there are no outlier in the data. The steps are listed below:

(a) Estimate the least squares parameters of BL(1,1,1,1) model using the available data, hence compute the residuals.

(b) Compute $\hat{\tau}_{IO,t}$, $\hat{\tau}_{AO,t}$, $\hat{\tau}_{TC,t}$ and $\hat{\tau}_{LC,t}$ for $t = 1, 2, ..., n$.

(c) Let $\eta_t = \max\left\{\left|\hat{\tau}_{IO,t}\right|, \left|\hat{\tau}_{AO,t}\right|, \left|\hat{\tau}_{TC,t}\right|, \left|\hat{\tau}_{LC,t}\right|\right\}$. For any t, if $\eta_t = \left|\tau_{TP,t}\right| > C$, where C is a pre-determined critical value, then the null hypothesis that the data point at time t is not an outlier of type TP is rejected.

3.   Application to the Kampung Aring Data Set

The procedure was applied to the monthly volume of rainfall collected at Kampung Aring weather station in the state of Kelantan, Malaysia. Before the detection procedure was applied, the Keenan's test (see Keenan [1985]) and the F-test (see Tsay [1986]) were used to test whether non-linear models suit the data sets.

3.1. Detection of Outlier in the Kampong Aring Rainfall Data

The rainfall data is for the period of August 1995 to July 2002. The plot of monthly average in millimeter is given in Figure 1. It can be observed that the data is generally stationary in mean and variance except at time points t = 41 and t = 77, where heavy rainfalls were recorded..
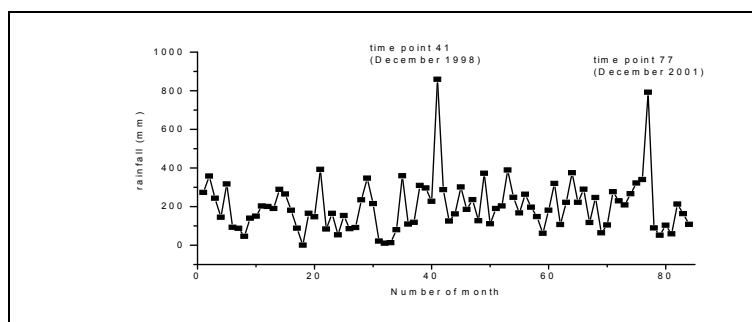

Figure 1: Plot of the Kampung Aring rainfall data

When nonlinearity tests are applied, the p-values of Keenan's test and F-test are 0.0293 and 0.2777 respectively. The Keenan's test strongly suggests that the data is nonlinear. The data was then modeled as a BL(1,1,1,1) process. Estimation yields parameter estimates; a = 0.6239, c = –0.2886 and b = –0.0007. The variance of the residuals is 17,761.01 representing a drop of 8.617% compared to the variance of the original data which equals to 19,442.36.

Table 1: Test statistics for outlier detection on the Kampung Aring rainfall
data

Point	Test Statistic for Outlier Type			
	IO	AO	TC	LC
41	−0.0319	**4.7475**	0.8530	0.0637
77	−0.0484	**4.4820**	−0.3204	0.2353

When the detection procedure is applied on the data, it is found out that an AO is detected at time point 41. The test statistics for AO at time point 77 is also large. The results are given in Table 1. Time points 41 and 77 correspond to December 1998 and December 2001 respectively. For December 1998, the average rainfall in the final 7 days of the month averages 66 mm per day. For December 2001, the reading of rainfall on 21st December 2001 is 110.5 mm and the reading is nearly double on the following day with 204 mm. This makes sense since December is within the rainy period in the east coast of Malaysia. In years other than 1998 and 2001, there was only a day or two with the rainfalls recorded at above 50 mm.

4. Conclusion

This article reports an outlier detection procedure for data modeled as BL(1,1,1,1) process. The procedure was then applied to detect outliers in a real data set. In the Kampung Aring rainfall data, two isolated high rainfall records are detected as AO-type outliers.

5. References

Box, G.E.P. and Jenkins, G.M. 1976. *Time Series Analysis Forecasting and Control,* Holden-Day, San Francisco.
Chen, C.W.S. 1997. Detection of additive outliers in bilinear time series. *Computational Statistics and Data Analysis*, **24**: 283-294.
Granger, C.W.J. and Andersen, A.P.1978. *Introduction to Bilinear Time Series Models*, Vandenhoeck and Ruprecht, Gottinge.
Keenan, D. M. 1985. A Tukey non-additivity type test for time series nonlinearity. *Biometrika* **72,** 1, pp. 39-44.
Mohler, R.R. 1973. *Bilinear Control Process,* Academic Press, New York.
Priestly, M.B. 1991. *Non-linear and Non-stationary Time Series Analysis*, Academic Press, San Diego.
Ruberti, A., Isidori, A. and D'allesandro, P. 1972. *Theory of Bilinear Dynamical System*, Springer-Verlag, Berlin.
Tsay, R.S. 1986. Nonlinearity test for time series. *Biometrika* 73, **2**, pp. 461-466.
Weiner, N. 1958. *Non-Linear Problems in Random Theory*, M.I.T. Press, Cambridge, Mass.
Azami Zaharim. 1996. Outliers and Change Points in Time Series Data. PhD Thesis, University of Newcastle Upon Tyne.