The Impact of Outliers on Parameter Estimation Method for Bilinear Model

Mohd Isfahani Ismail*, Hazlina Ali and Sharipah Soaad Syed Yahaya

School of Quantitative Sciences, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia

Nonlinear least squares (*NLS*) method along with Newton-Raphson (*NR*) iterative procedure is the best method to estimate parameters for bilinear model. However, the existence of outliers will affect the estimated value of the parameter and its validity can be doubtful. This statement was proven by conducting simulation analysis for the bilinear model, especially on bilinear (1,0,1,1) model without and with the existence of additive outlier (AO), innovational outlier (IO), temporary change (TC) and level change (LC) in the data. The performance of the *NLS* method is measured in terms of bias. Numerical results show that, in general, the *NLS* method performs better in estimating the parameters without the existence of AO, IO, TC or LC in the data.

Keywords: bilinear model; nonlinear least squares; Newton-Raphson; additive outlier; innovational outlier; temporary change; level change

I.   INTRODUCTION

Time series model can be divided to linear and nonlinear models. Linear models are more popular due to its simplicity. However, not all linear models are suitable for a time series data. As alternative, the nonlinear models may be more suitable. The simplest among the nonlinear models is bilinear model as it shows the most natural way to move from linear to nonlinear model (Ramakrishnan & Morgenthaler, 2010).

Parameter estimation is important phase in time series modeling since the estimate will be further incorporated into the subsequent phases. There are various types of methods to estimate time series parameters especially in bilinear model, among others is the nonlinear least squares (*NLS*) method. The main advantage of *NLS* method over other techniques is the flexibility in fitting various forms of functions and the efficient use of data even with relatively small data sets. Meanwhile, to calculate the estimated parameters, the *NLS* procedure needs to use iterative optimization to obtain a better estimate. In the meantime, before starting iterative procedure, an initial value of an unknown parameter is required for the software to run the

optimization process. Thus, *NR* iterative procedure is used to generate these initial values Therefore, the implementation of *NR* iterative procedure along with *NLS* method is much needed.

In general, there are four types of outliers known as additional outlier (AO), innovational outlier (IO), temporary change (TC) and level change (LC).AO is the type of outliers that affects a single observation at time point $t = d$ (Abuzaid, Mohamed & Hussin, 2014). Meanwhile, IO is characterized by a single strange observation at time point $t = d$ but in addition it also affects subsequent observations with the effect gradually dying out (Abuzaid *et. al.*, 2014). The third type of outlier is called as LC because its behaviour is to change the level or mean of the observed series. This type of outlier provides a sudden and permanent change to the observed series at the time point $t = d$ and continues until the end of the observed period, $t > d$.

Meanwhile, TC is a change which the effect is reduced exponentially. TC affects a series at a certain time, and its effect is weakened exponentially according to a dampening factor (δ).

*Corresponding author's e-mail: md_isfahani@yahoo.com

In this study we take $\delta = 0.7$ as recommended by Chen and Liu (1993).

The existence of outliers in the data will affect the estimated parameters, which consequently will jeopardize the validity of the model. Therefore, detecting and correcting outlier effects is an important task in the construction of a good predictor model. Hence, this paper will show the effect of outlier in nonlinear, specifically in building bilinear model. This study focusses on AO, IO, TC and LC in bilinear (1,0,1,1) model since this model is the simplest form of bilinear models. The investigation covers the performance of *NLS* method with *NR* iterative procedure in estimating the coefficients of bilinear (1,0,1,1) model without and with the existence of outlier in the data. Through simulation study, the performance of the *NLS* method will be measured in terms of bias using simulated data of several study specifications.

I. LITERATURE REVIEW

A. Bilinear Model

The general formulation of bilinear (p,q,r,s) model is represented by:

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

where Y_t and e_t each represents the observations and residuals at time t, where $t = 1,2,3,...$. The e_t's are assumed to follow normal distribution with mean zero and variance σ^2. Meanwhile, a_i, c_j and b_{kl} are the coefficients of the model. Based on (1), the first two components represent the autoregressive moving average (ARMA) linear model with order p and q, while the third component, which represents nonlinearity, helps to explain the nonlinearity characteristic of the data being modelled with order r and s. In this paper, the bilinear (1,0,1,1) model is considered and given by:

$$Y_t = a Y_{t-1} + b Y_{t-1} e_{t-1} + e_t \quad (2)$$

where a and b are the coefficients, while Y_t is outlier-free observation and e_t is outlier-free residual, such that $t = 1,2,3,...$. Meanwhile, the bilinear (1,0,1,1) model with existence of outlier is represented by:

$$Y_t^* = a Y_{t-1}^* + b Y_{t-1}^* e_{t-1}^* + e_t^* \quad (3)$$

where Y_t^* is the contaminated observations and e_t^* represents the contaminated residuals. The Y_t^* and e_t^* exist when there is outlier in the data at certain time point t, where $t = 1,2,3,...,n$.

B. AO Effects on Original Observations and Residuals

When there is no outlier existing in the data at time point t, the observations (Y_t) is known as the original observations. If AO exists in the data, the symbol $Y_{t,AO}^*$ is used to signify of the existence of the outlier and is known as "AO effect on observation". The effect of this outlier exists only at time point $t = d$ with ω as magnitude of outlier effect from bilinear (1,0,1,1) model. For time point $t \neq d$, clearly $Y_{t,AO}^* = Y_t$ and the full formulation of AO effects on Y_t is given by:

$$Y_{t,AO}^* = \begin{cases} Y_t & for \quad t \neq d \\ Y_t + \omega & for \quad t = d \end{cases} \quad (4)$$

where $t = 1,2,3,...,n$ and $d = 1,2,3,...,n$

From (4), it is indicated that the effect of AO on Y_t occurs only at one time point while the rest of the time points are unaffected.

Meanwhile, the original residual (e_t) are obtained when there is no outlier existing in the data at time point t. The "AO effect on residual" is denoted by $e_{t,AO}^*$. At time point $t < d$, the $e_{t,AO}^* = e_t$. While, at time point $t \geq d$ and $k \geq 0$, the formulation for $e_{d+k,AO}^*$ is:

$$e^*_{d+k,AO} = e_{d+k} - \omega A_{k,AO} \qquad (5)$$

where

$$A_{k,AO} = \begin{cases} -1 & for \quad k=0 \\ (a_k + b_{k1}e_{d+k-1}) - \sum_{j=1}^{k}(b_{1j}Y^*_{d+k-1,AO})A_{d+k-j,AO} & for \quad k \geq 1 \end{cases}$$

a and b are constant values. Based on (5), several residuals for time point $t \geq d$ should be affected.

C. IO Effects on Original Observations and Residuals

The IO effects on observations at time point $t < d$ is given by $Y^*_{t,IO} = Y_t$ while the IO effects on Y_t for $t \geq d$ is represent by:

$$Y^*_{d+k,IO} = Y_{d+k} + \omega A_{k,IO} \qquad (6)$$

where

$$A_{k,IO} = \begin{cases} 1 & for \quad k=0 \\ \sum_{m=1}^{k}(a_m + b_{m1}e_{d+k})A_{k-m,IO} & for \quad k \geq 1 \end{cases}$$

Based on (6), the existence of IO in bilinear (1,0,1,1) model effects Y_t not only at one time point but also at some of the subsequent Y_t.

The symbol $e^*_{t,IO}$ is used when there is IO effect on the residual in bilinear (1,0,1,1) model. At time point $t < d$, The $e^*_{t,IO} = e_t$ while at time point $t \geq d$ and $h \geq 0$, the equation for $e^*_{d+h,IO}$ is given by:

$$e^*_{d+h,IO} = e_{d+h} + \omega f_{d+h} \qquad (7)$$

where

$$f_{d+h} = \begin{cases} A_{0,IO} & for \quad h=0 \\ A_{h,IO} - (\sum_{m=1}^{h} b_{m1}f_{d+h-m})Y^*_{d+h-1,IO} - \sum_{k=1}^{h}(a_k + b_{k1}e_{d+h-k})A_{h-k,IO} & for \quad h \geq 1 \end{cases}$$

This equation indicates that the existence of IO not only changes the residual at $t = d$ but also changes some of the subsequent residuals.

D. TC Effects on Original Observations and Residuals

For TC effects on observations at time point $t < d$ is given by $Y^*_{t,tc} = Y_t$. Meanwhile, the equation of TC effects on Y_t for time point $t \geq d$ and $k \geq 0$ is given by:

$$Y^*_{d+k,tc} = Y_{d+k} + \delta^k \omega \qquad (8)$$

where δ is dampening factor for TC effects.

The symbol $e^*_{t,tc}$ denotes the residual which is affected by TC in bilinear (1,0,1,1) model. At time point $t < d$, $e^*_{t,tc} = e_t$. While, at time point $t \geq d$ and $k \geq 0$, the equation of TC effects is given by:

$$e^*_{d+k,tc} = e_{d+k} + \omega A_{k,tc} \qquad (9)$$

where

$$A_{k,tc} = \begin{cases} 1 & for \quad k=0 \\ \delta^k - \sum_{i=1}^{k}(a_i + b_{i1}e_{d+k-1})\delta^{k-i} + \sum_{j=1}^{k}(b_{1j}Y^*_{d+k-1,tc})A_{k-j,tc} & for \quad k \geq 1 \end{cases}$$

a and b are constant numbers. Based on (9), it can be seen that with the existence of TC, more than one observation and residuals are effected.

E. LC Effects on Original Observations and Residuals

The LC effect on observations at time point $t < d$ is $Y^*_{t,lc} = Y_t$, while, at time point $t \geq d$ is given by $Y^*_{t,lc} = Y_t + \omega$. The full formulation of AO effects on Y_t is given by:

$$Y^*_{t,lc} = \begin{cases} Y_t & for \quad t < d \\ Y_t + \omega & for \quad t \geq d \end{cases} \qquad (10)$$

The symbol $e^*_{t,lc}$ is used for LC effect on residual in bilinear (1,0,1,1) model. At time point $t < d$, $e^*_{t,lc} = e_t$. While at time point $t \geq d$ and $k \geq 0$, the formulation for $e^*_{d+k,lc}$ is:

$$e_{d+k,lc}^{*} = e_{d+k} + \omega A_{k,lc} \qquad (11)$$

where

$$A_{k,lc} = \begin{cases} 1 & for \quad k = 0 \\ 1 - (a_k + b_{k1}e_{d+k-1}) - \sum_{j=1}^{k}(b_{1j}Y_{d+k-1,lc}^{*})A_{k-j,lc} & for \quad k \geq 1 \end{cases}$$

a and b are constant numbers. Based on (11), several residuals for time point $t \geq d$ should be affected. Various methods of estimating the parameters of bilinear models are available. In this paper, the *NLS* method along with *NR* iterative procedure is used to estimate a and b.

II.METHODOLOGY

A. Nonlinear Least Squares (NLS) Method

The popular classical method of estimating parameters in the past for the bilinear model is known as nonlinear least squares (*NLS*) method. Some of the earliest researchers who worked on this method are Goldfeld and Quandt (1972). Meanwhile, Granger and Andersen (1978a) and Liu (1985), have studied and applied this method to bilinear $(1,0,1,1)$ and $(2,1,1,1)$ models on actual data. The general procedure of *NLS* method has been presented for BL $(p,0,r,s)$ model by Priestley (1991). The *NLS* estimation method for bilinear $(1,0,1,1)$ model is described as follows.

Let $\theta = (\theta_1 = a, \theta_2 = b)'$ denotes the complete set of parameters for bilinear $(1,0,1,1)$ model. The objective of the method is to minimize the equation of:

$$Q(\theta) = \sum_{t=\lambda+1}^{n} e_t^2 \qquad (12)$$

Where e_t is from equation (2). Then, the process of minimization is accomplished through Newton-Raphson (*NR*)iterative procedure which is given by:

$$\theta^{(i+1)} = \theta^{(i)} - H^{-1}(\theta^{(i)})G(\theta^{(i)}) \qquad (13)$$

where $\theta^{(i)}$ represents the vector of parameter estimate for i-th iteration, while the vector of gradient denoted by G and Hessian matrix denoted by H. The formulations of G vector and H matrix are

$$G(\theta) = \left\{ \frac{\partial Q}{d\theta_1}, \frac{\partial Q}{d\theta_2} \right\} \qquad (14)$$

$$H(\theta) = \left\{ \frac{\partial^2 Q}{\partial\theta_1\partial\theta_2} \right\} \qquad (15)$$

Based on (14) and (15), then the partial derivatives of equation (12) are given by:

$$\frac{\partial Q}{\partial\theta_1} = 2\sum_{t=2}^{n} e_t \frac{\partial e_t}{\partial\theta_1} \qquad (16)$$

$$\frac{\partial^2 Q}{\partial\theta_1\partial\theta_2} = 2\sum_{t=2}^{n} \frac{\partial e_t}{\partial\theta_1}\frac{\partial e_t}{\partial\theta_2} + 2\sum_{t=2}^{n} e_t \frac{\partial^2 e_t}{\partial\theta_1\partial\theta_2} \qquad (17)$$

Based on (2) and (14), the following partial derivatives can be obtained as shown below:

$$\frac{\partial e_t}{\partial a} = -Y_{t-1} - bY_{t-1}\frac{\partial e_{t-1}}{\partial a},$$

$$\frac{\partial e_t}{\partial b} = -Y_{t-1}e_{t-1} - bY_{t-1}\frac{\partial e_{t-1}}{\partial b},$$

$$\frac{\partial^2 e_t}{\partial a^2} = -bY_{t-1}\frac{\partial^2 e_{t-1}}{\partial a^2},$$

$$\frac{\partial^2 e_t}{\partial b^2} = -2Y_{t-1}\frac{\partial e_{t-1}}{\partial b} - bY_{t-1}\frac{\partial^2 e_{t-1}}{\partial b^2},$$

$$\frac{\partial^2 e_t}{\partial a\partial b} = -Y_{t-1}\frac{\partial e_{t-1}}{\partial a} - bY_{t-1}\frac{\partial^2 e_{t-1}}{\partial a\partial b},$$

For simplicity, the most common choice, if no prior information is available, is to choose the following conditions:

$$e_t = \frac{\partial e_t}{\partial\theta_1} = \frac{\partial^2 e_t}{\partial\theta_1\partial\theta_2} = 0, \text{ for all } t = 1 \text{ and } 2.$$

From (13), the iterative procedure can now be implemented. The iteration is stopped when the following conditions are met, $a^1, a^2, ..., a^N$ and $b^1, b^2, ..., b^N$ which are constructed for each parameter a and b until $\left|\theta_i^N - \theta_i^{N-1}\right| < \in$, for $i = 1, 2$, $\theta = (a,b)'$ and $\in$ is tolerance. In this study, we follow the same approach by Priestley (1991) in determining the initial

value where $\in$ was chosen to be 10^{-3}. The initial value selection is one of the most important elements to consider in the iteration procedure. According to Priestley (1991), in determining the initial value, he presented the given procedure:

"If bilinear $(p,0,1,1)$ model is to be fitted, then the parameter estimates of AR (p) model form initial estimates of $a_1,...,a_p$ while the initial estimate of b_{11} is taken to be zero. For bilinear $(p,0,2,1)$ and bilinear $(p,0,1,2)$ models, the estimates with the initial values of b_{21} and b_{12} respectively taken to be zero. For bilinear $(p,0,2,2)$ model, the parameter estimates of bilinear $(p,0,2,1)$ model or bilinear $(p,0,1,2)$ model are used as the initial values and the initial value for b_{22} is taken to be zero. The process then continues."

In the next section, we will discuss the complete steps in obtaining estimates of the parameters using the *NLS* method as well as the simulation studies as done by Priestley (1991).

B. Parameter Estimation Step

To start the iteration, the *NR* procedure requires initial values for the parameters. The process of getting the estimated parameters of bilinear (1,0,1,1) model are shown below. The first step explains how to obtain the initial value, while the second and third steps are the parameter estimation procedures.

1. From equation (2) and (3) of t data set (with and without outlier), where $t = 1,2,...,n$, then AR (1) model is used to estimate a. The formulation of the AR (1) model is $Y_t = \phi_0 + \phi_1 Y_{t-1} + e_t$, where $e_t =$ the term of error at time point t with mean equal to zero and constant variance equal to σ_e^2, and ϕ_0 is constant term.

2. Let the initial values for each a and b be a^* and 0 in the *NR* iteration procedure to estimate the bilinear (1,0,1,1) model parameters. Then, the parameter estimation procedure is complemented by the *NLS* method along with the implementation of the *NR* iteration procedure as described previously.

3. Finally, the estimates of a and b are obtained and symbolized by $\hat{a}$ and $\hat{b}$.

III. SIMULATION AND RESULTS

A simulation study was carried out to observe the performance of the procedure using S-Plus package. To assess and evaluate the performance of the procedure, the combination of the following factors is considered:

a) Generate five underlying bilinear (1,0,1,1) models for different combinations of known coefficients values (a,b).

b) Obtain the estimates of a and b using the aforementioned parameter estimation procedure (based on previous subsection).

c) Use number of simulations $(s = 100)$ and length $n = 100$.

d) Calculate bias by letting $\gamma = \{a,b\}$ where γ_i represents the estimation of a and b $(i = 1,..., s)$, and given mean γ is

$$\bar{\gamma} = \sum_{i=1}^{s} \bar{\gamma}_i$$ .Bias for the paramaters is $(|\gamma - \bar{\gamma}|)$.

We investigate the performance of parameter estimation procedure for bilinear (1,0,1,1) model based on bias. Tables 1 and 2show the results for bilinear (1,0,1,1) model without and with the existence of outlier. The first column of each table displays the different combinations of coefficients of the bilinear models (a,b). The combination of different coefficients is used to see how the combined effect of the estimates. For example, the combined coefficient (0.1,0.2) results is a smaller bias value compared to the coefficient of combinations (0.4,0.2). This shows the greater the combined value of the coefficient used the greater the bias value.

The estimates and biases are given in the subsequent row according to coefficients of each model. For example, in Table 1, the second column lists the estimators for a and b, i.e $\hat{a}$ and $\hat{b}$ respectively, while the fourth column shows the bias for the estimators, and the different approaches namely as *NLS* without outlier and *NLS* with outlier (AO,IO, LC and TC)are shown based on the given columns of tables.

Based on the bias values of the estimates in the tables, it is clearly shown that the *NLS* method without outlier is the best

approach based on the smallest bias values obtained as compared to the others. Bias values among the *NLS* method with AO, IO, TC and LC do not differ much from each other. Based on the models with different combination coefficients, we observe that by increasing the value of coefficient, the bias value increase for the *NLS* method without outlier and drastically increase for the *NLS* method with outlier.

Table 1. Parameter estimation for bilinear (1,0,1,1) model without and with AO and IO

Coefficient	Estimation	*NLS* without outlier	Bias	*NLS* with AO	Bias	*NLS* with IO	Bias
(0.1,0.2)	$\hat{a}$	0.088	0.012	0.062	0.038	-0.054	0.154
	$\hat{b}$	0.202	0.002	0.068	0.132	0.025	0.175
(0.1,0.3)	$\hat{a}$	0.089	0.011	0.072	0.028	58.726	58.626
	$\hat{b}$	0.299	0.001	0.100	0.200	0.086	0.214
(0.2,0.2)	$\hat{a}$	0.189	0.011	0.177	0.023	-14.802	15.002
	$\hat{b}$	0.199	0.001	-0.086	0.286	-0.037	0.237
(0.3,0.2)	$\hat{a}$	0.284	0.016	-4.648	4.948	-14.802	15.102
	$\hat{b}$	0.192	0.008	0.124	0.076	-0.037	0.237
(0.4,0.2)	$\hat{a}$	0.373	0.027	-6.911	7.311	25.599	25.199
	$\hat{b}$	0.187	0.013	0.085	0.115	-0.016	0.216

Table 2. Parameter estimation for bilinear (1,0,1,1) model without and with LC and TC

Coefficient	Estimation	*NLS* without outlier	Bias	*NLS* with LC	Bias	*NLS* with TC	Bias
(0.1,0.2)	$\hat{a}$	0.088	0.012	-11.058	11.597	5.527	5.427
	$\hat{b}$	0.202	0.002	0.094	0.106	0.122	0.078
(0.1,0.3)	$\hat{a}$	0.089	0.011	2.039	1.939	2.636	2.536
	$\hat{b}$	0.299	0.001	0.104	0.196	0.114	0.186
(0.2,0.2)	$\hat{a}$	0.189	0.011	23.741	23.541	-3.325	3.525
	$\hat{b}$	0.199	0.001	0.097	0.103	0.145	0.055
(0.3,0.2)	$\hat{a}$	0.284	0.016	-10.044	10.344	-12.738	13.038
	$\hat{b}$	0.192	0.008	0.034	0.166	0.052	0.148
(0.4,0.2)	$\hat{a}$	0.373	0.027	-38.868	39.268	-65.898	66.298
	$\hat{b}$	0.187	0.013	0.097	0.103	-0.138	0.338

IV.   CONCLUSION

In this paper, we discuss the method of parameters estimation known as nonlinear least square (*NLS*) method along with Newton-Raphson (*NR*) iterative procedure. The simulated values of parameter estimation are presented. The finding reveals that the *NLS* method without outlier performs well in estimating the coefficients of bilinear (1,0,1,1) model. In contrast, with the existence of outliers, AO, IO, TC and LC, the estimation of the coefficients of bilinear (1,0,1,1) model are adversely affected especially when the value of coefficients increases.

V.   ACKNOWLEDGEMENTS

VI.   REFERENCES

Abuzaid, A.H.M. Mohamed, I.B. & Hussin, A.G., 2014, Procedures for Outlier Detection in Circular Time Series Models. Environmental and Ecological Statistics, 21, pp. 793-809.

Chen & Liu, 1993, "Joint Estimation of Model Parameters and Outlier Effects in Time Series." Journal of the American Statistical Association 88: pp. 284-297.

Goldfeld & Quandt 1972, Non-linear Methods in Econometrics. Amsterdam: North-Holland.

Granger & Andersen 1978a, An Intro-duction to Bilinear Time Series Models. Vandenhoeck & Ruprecht: Göttingen.

Liu 1985, Theory of Bilinear Time Series Models. Comm. Statist. Theory Meth., 14, pp. 2549-2561.

Priestley 1991, Non-linear and Non-stationary Time Series Analysis. San Diego: Academic Press.

Ramakrishnan, R. & Morgenthaler, S. 2010, Robust Multivariate and Nonlinear Time Series Models. Lausanne, EPFL.10.5075/epfl-thesis-4688.