American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)
ISSN (Print) 2313-4410, ISSN (Online) 2313-4402

http://asrjetsjournal.org/

On the Efficiency of Outlier Generating Mechanisms in Multivariate Time Series

Olusesan Oluyomi Olufolabo[a*], Olanrewaju Ismail Shittu[b], Kazeem Adesola Adepoju[c]

[a]Department of Statistics, Yaba College of Technology, Lagos, Nigeria
[b,c]Department of Statistics, University of Ibadan, Ibadan, Nigeria
[a]Email: olufolabosesan@yahoo.com
[b]Email: oi.shittu@hotmail.com
[c]Email: ka.adepoju@.ui.edu.ng

Abstract

In this paper, two new outlier generating mechanisms for the detection of outliers in multivariate time series setting were derived. This is achieved by specifying two-variable vector autoregressive models and assuming additive and convolution effect of outliers on time series data. The magnitude and variance of outlier were derived for the generating models by method of least squares. Also a modified test statistics were developed to detect single outliers both in the response and explanatory variables. In order to establish the validity and efficiency of the derived models, the models were applied to both simulated and existing data. The results from the analysed data were also compared to some existing models and the result showed that the convolution model is best in terms of the number of outliers detected and the residual variance. This result confirms the finding in previous studies of outlier detection in univariate time series.

Keywords: Additive outlier; Convolution outlier; Innovative outlier; Multiplicative outlier; Vector auto regressive.

1. Introduction

The problem of outlier detection in time series has gained much attention in recent times and various methods of detection are available, but in most cases, it is limited to univariate time series.

----

* Corresponding author.

It is a known fact, that in time series analysis, outliers can cause biases in parameter estimation as well as inappropriate predictions, resulting in misleading conclusion [28]. The essence of outlier detection is to discover the unusual data; whose behaviour is very exceptional when compared to the rest of the data set. Examining the extraordinary behaviour of outliers

will surely help to uncover the valuable knowledge hidden behind them and to help the decision makers to improve on the quality of data.

Generally, detection methods are divided into two parts: univariate and multivariate methods. In univariate methods, observations are examined individually while in multivariate methods, associations between variables in the same dataset are taken into account.

Several outlier detection methods have been proposed for univariate time series including [11,8,9,26,1,14 and 32]. All the listed works were based on time domain and almost all make use of iterative procedure in the outlier detection process. However, [25] in their work considered the identification of outliers in frequency domain using the spectral method.

On the detection of outlier in multivariate time series, [12] made use of projection pursuit technique while [3] proposed the Independent Component Analysis (ICA) as a tool capable of identifying the locations of multiple outliers in multivariate time series. The authors [10] used meta-heuristic methods to detect additive outliers in multivariate time series. The work of [13] introduced the coefficient of vector autocorrelation, obtained its influence function together with its distribution and used it to test the hypothesis of presence of outliers. [31] in his paper used an efficient two-phase algorithm for detecting outlying samples in multivariate time series datasets. The Bounded Coordinate System metric was used to measure the similarity between two multivariate time series samples, and the outlierness of a sample is measured by average distance to its k nearest neighbours. Then a heuristic and two pruning rules were utilized to quickly remove multivariate time series samples that are not possible outlier candidates, reducing significantly the distance computation among objects.

As a result of outlier masking effect of both Additive and Innovative on the estimates of parameters and the multiplicative effect on parameters estimated, [26] introduced two other types of outliers which are Convolution Outlier (CO) and Multiplicative Outlier (MO) for univariate time series. The work of [26] was extended to multivariate time series by [19] whereby two generating mechanisms; Innovative and Multiplicative were considered. It was concluded that Multiplicative outlier model was more sensitive to outlier with minimum standard error of the estimate.

For this paper, two outliers generating mechanisms; Convolution and Additive will be extended to multivariate time series and their performances in terms of outlier detection will be compared to the existing ones.

2. Methodology

In this section, by assuming that outliers have either Additive or Convolution effect on a series for bivariate time series and specifying two-variable Vector Autoregressive (VAR) models, the estimate of the parameter for the two models will be derived and their corresponding test statistics developed.

2.1 Derivation of Outlier Generating Mechanisms for Additive Outlier (AO) Model

In this subsection outlier generating mechanism for additive outlier model will be derived.

Generally, an additive outlier represents an unexpected change in the value of one of the observations. It can appear as a result of a recording or measurement error or other single effect.

The additive outlier is defined as

$$X_t = Z_t + \omega \xi_t^{(T)} \tag{1}$$

where $X_t = (x_{1t,} x_{2t,} \dots\dots\dots.. x_{kt})$ is a k-dimensional time series, Z_t is an outlier free time series that is assumed to follow the Autoregressive Moving Average of Order (p,q) i.e. ARMA $_{(p,q)}$, $\xi_t^{(T)}$ is a time indictor such that $\xi_t^{(T)} = 1$ for all $t = T$ *and* $\xi_t^{(T)} = 0$ otherwise, and $\omega = \left(\omega_1, \dots, \omega_k\right)'$ is the size or the magnitude of outlier.

Now, given vector models X_{1t} and X_{2t} such that X_{1t} contains outlier and X_{2t} is outlier free, the magnitude of such outlier and its corresponding variance can be obtained by specifying the two variable VAR(2) as:

$$X_{1t} = \varphi_{11} X_{1t-1} + \varphi_{12} X_{2t-1} + a_{1t} \tag{2}$$

$$X_{2t} = \varphi_{21} X_{2t-1} + \varphi_{22} X_{1t-1} + a_{2t} \tag{3}$$

where, X_{1t} and X_{1t-1} are the current and lag values of the response variable respectively, X_{2t} and X_{2t-1} are current and lag values of the explanatory variable respectively.

Now considering equation (3)

$$X_{2t} = \emptyset_{21} X_{2t-1} + \emptyset_{22} X_{1t-1} + a_{1t}$$

When X_{2t-1} is contaminated and assumed additive model, we have

$$X_{2t} = \emptyset_{21}\left[Z_{t-1} + \omega\xi_{t-1}^{(T)}\right] + \emptyset_{22} X_{1t-1} + a_{1t} \tag{4}$$

$$= \emptyset_{21} Z_{t-1} + \phi_{21}\, \omega\xi_{t-1}^{(T)} + \emptyset_{22} X_{1t-1} + a_{1t}$$

$X_{2t} = \emptyset_{22} X_{1t-1} + \phi_{21} Z_{t-1} + \phi_{21} \omega \xi_{t-1}^{(T)}$

$X_{2t} = \emptyset_{22} X_{1t-1} + \phi_{21} \varphi(\beta) \xi_{t-1} + \phi_{21} \omega \xi_{t-1}$

$X_{2t} = \emptyset_{22} X_{1t-1} + \phi_{21} \xi_{t-1}\big(\omega + \varphi(\beta)\big)$

Therefore, the general the additive model is given as

$$Additivemodel: X_{Ait} = \phi_{ii} X_{jt-1} + \emptyset_{ij} \epsilon_{t-1} \big(\omega + \varphi(\beta)\big) \tag{5}$$

2.1.1 Derivation of the Magnitude of Outlier for AO

With $X_{1t} = \varphi_{11} X_{1t-1} + \varphi_{12} X_{2t-1} + a_{1t}$ as defined in equation (2)

Then,

$$Z_t + \omega \xi_t^{(T)} = \varphi_{11}\left(Z_t + \omega \xi_t^{(T)}\right) + \varphi_{12} X_{2t-1} + a_{1t} \tag{6}$$

$$\frac{\theta(\beta)}{\varphi(\beta)} a_t + \omega \xi_t^{(T)} = \varphi_{11}\left[\frac{\theta(\beta)}{\varphi(\beta)} + \omega \xi_t^{(T)}\right] + \varphi_{12} \phi(\beta) \mathrm{l}_t + a_{1t}$$

$$a_t\left[\frac{\theta(\beta)}{\phi(\beta)} - 1\right] = \phi_{11}\left[\frac{\theta(\beta)}{\phi(\beta)} + \omega \in_t^{(T)}\right] + \phi_{12} \varphi(\beta) \ell_t - \omega \xi_t^{(T)}$$

$$a_t = \frac{\phi_{11}\left[\frac{\phi(\beta)}{\theta(\beta)} + \omega \xi_t^{(T)}\right] + \phi_{12} \varphi(\beta) \ell_t - \omega \xi_t^{(T)}}{\frac{\phi(\beta)}{\theta(\beta)} - 1} \tag{7}$$

Summing the square of equation (1.9) over n we have

$$\sum_{t=1}^{n} a_t^2 = \sum_{t=1}^{n}\left[\frac{\phi_{11}\left(\frac{\phi(\beta)}{\theta(\beta)} + \omega \xi_t^{(T)}\right) + \phi_{12} \varphi(\beta) \ell_t - \omega \xi_t^{(T)}}{\frac{\phi(\beta)}{\theta(\beta)} - 1}\right] \tag{8}$$

Differentiating equation (1.7) with respect to ω and setting to zero, we obtain the magnitude of outlier in the model as

$$\frac{\partial a_t^2}{\partial \hat{\omega}_A} = \frac{2}{\left(\frac{\phi(\beta)}{\theta(\beta)} - 1\right)^2} \sum_{t=1}^{n} \phi_{11}\left(\frac{\phi(\beta)}{\theta(\beta)} + \omega\xi_t^{(T)}\right) + \phi_{12}\varphi(\beta)\mathsf{l}_t - \omega\xi_t^{(T)} = 0 \tag{9}$$

$$\sum_{t=1}^{n} \phi_{11}\left(\frac{\phi(\beta)}{\theta(\beta)} + \omega\right) + \sum_{t=1}^{n} \phi_{12}\varphi(\beta)\ell_t - \sum_{t=1}^{n} \omega = 0$$

$$\phi_{11}\left(\frac{\phi(\beta)}{\theta(\beta)}\right) + \phi_{11}\omega + \phi_{12}\varphi(\beta)\ell_t - \omega = 0$$

$$\omega(1-\phi_{11}) = \phi_{11}\frac{\theta(\beta)}{\phi(\beta)} + \phi_{12}\varphi(\beta)\ell_t$$

$$\hat{\omega}_A = \frac{\phi_{11}\frac{\theta(\beta)}{\phi(\beta)} + \phi_{12}\varphi(\beta)\ell_t}{1-\phi_{11}}$$

$$\hat{\omega}_A = \frac{\varphi(\beta)(1+\phi_{12}\,\ell_t)}{1-\phi_{11}} \tag{10}$$

Therefore, the estimate of the variance is

$$V(\hat{\omega}_A) = \frac{\phi\left(\beta^{2}\right)\varphi_{12}^2\,\sigma_a^2}{(1-\varphi_{11})^2}$$

$$= \left(\frac{\varphi_{12}}{1-\varphi_{11}}\right)^2 \phi\left(\beta^{2}\right)\sigma_a^2 \tag{11}$$

With the estimates of mean and variance of the magnitude of AO derived, the test statistic for testing the presence of outlier for additive model is constructed as follows:

$$\lambda_i = \frac{1+\varphi_{12}\mathsf{l}_t}{1-\varphi_{11}} \cdot \frac{1-\varphi_{11}}{\varphi_{12}\,\sigma_a}$$

$$\lambda_i = \frac{(1+\varphi_{12}\,\mathsf{l}_t)}{\varphi_{12}\,\sigma_a} \tag{12}$$

2.2 Derivation of Outlier Generating Mechanisms for Convolution Outlier (MO) Model

The outlier effect on a given series may be either additive or innovative and the effect may be a combination of the two [26].

By this, we propose the convolution of the additive and innovative outliers for the multivariate setting as follows:

The innovative and additive models are defined respectively as follows:

$$X_{tI} = Z_t + \varphi(\beta)\omega\xi_t^T \quad \text{for innovative model}$$

$$X_{tA} = Z_t + \omega\xi_t^{(T)} \quad \text{for additive model}$$

The convolution involved adding both innovative and additive models [26]. This gives

$$X_{tC} = 2\ Z_t + \omega\xi_t^T\left(1+\varphi(\beta)\right) \quad (13)$$

For the general case of CO, now considering,

$X_{2t} = \emptyset_{21}X_{2t-1} + \emptyset_{22}X_{1t-1} + a_{1t}$ as defined in equation (2)

Assuming X_{2t-1} is contaminated, we have

$$X_{2t} = \emptyset_{21}\left(2Z_{t-1} + \omega\xi_t^{(T)}(1+\varphi(\beta))\right) + \emptyset_{22}X_{1t-1} + a_{1t} \quad (14)$$

$X_{2t} = 2\emptyset_{21}Z_{t-1} + \emptyset_{21}\omega\xi_t^{(T)}(1+\varphi(\beta)) + \emptyset_{22}X_{1t-1} + a_{1t}$

where $Z_t = \varphi(\beta)\xi_t^{(T)}$and $Z_{t-1} = \varphi(\beta)\xi_t^{(T)}$

we then have

$$X_{2t} = 2\emptyset_{21}\varphi(\beta)\xi_t^{(T)} + \emptyset_{21}\omega\xi_t^{(T)}(1+\varphi(\beta)) + \emptyset_{22}X_{1t-1} + a_{1t} \quad (15)$$

$X_{2t} = 2\emptyset_{21}\varphi(\beta)\xi_t^{(T)} + \emptyset_{21}\omega\xi_{t-1} + \emptyset_{21}\omega\xi_t^{(T)}\varphi(\beta) + \emptyset_{22}X_{1t-1} + a_{1t}$

$X_{2t} = \emptyset_{22}X_{1\,t-1} + 2\emptyset_{21}\varphi(\beta)\xi_t^{(T)} + \emptyset_{21}\omega\xi_t^{(T)} + \emptyset_{21}\omega\xi_t^{(T)}\varphi(\beta)$

$X_{2t} = \emptyset_{22}X_{1\,t-1} + 2\emptyset_{21}\varphi(\beta)\xi_t^{(T)} + \emptyset_{21}\omega\xi_t^{(T)}\left(1+\varphi(\beta)\right)$

$$X_{2t} = \emptyset_{22}X_{1\,t-1} + \emptyset_{21}\xi_t^{(T)}[2\varphi(\beta) + \omega(1+\beta)] \quad (16)$$

Therefore, in general, the CO generating mechanism is

$$Convolutionmodel: X_{Cit} = \phi_{ii} X_{jt-1} + \emptyset_{ij} \xi_{t-1}^{T} [2\varphi(\beta) + \omega(1+\beta)] \tag{17}$$

2.1.2 Derivation of Magnitude of Outlier for CO

Now, specifying $X_{1t} = \phi_{11} X_{1t-1} + \phi_{12} X_{2t-1} + a_{1t}$ (18)

and substituting X_{tC} in equation (18) gives

$$2 Z_t + \omega \xi_t^{(T)} (1+\varphi(\beta)) = \phi_{11} (2 Z_{t-1} + \omega \xi_t^{(T)} (1+\varphi(\beta))) + \phi_{12} X_{2t-1} + a_{1t} \tag{19}$$

$$2\varphi(\beta) a_t + \omega \xi_t^{(T)} (1+\varphi(\beta)) = \phi_{11} [2\varphi(\beta) a_{t-1} + \omega \xi_t^{T} (1+\varphi(\beta))] + \phi_{12} \varphi(\beta) \ell_{t-1} + a_{1t}$$

$$2\varphi(\beta) a_t - \phi_{11} 2\varphi(\beta) a_{t-1} - a_{1t=} \phi_{12} \varphi(\beta) \ell_{t-1} - \omega \xi_t^{(T)} (1+\varphi(\beta)) + \phi_{11} \omega \xi_t^{(T)} (1+\varphi(\beta))$$

$$a_t [2\varphi(\beta) - 2\varphi(\beta)\phi_{11} - 1] = \phi_{12} \varphi(\beta) \ell_{t-1} - \omega \xi_t^{(T)} (1+\varphi(\beta)) + \phi_{11} \omega \xi_t^{(T)} (1+\varphi(\beta)) \tag{20}$$

Summing and squaring equation (20) gives

$$\sum_{t=1}^{n} a_t^2 = \frac{\sum_{t=1}^{n} [\phi_{12} \varphi(\beta) \ell_{t-1} - \omega \xi_t^{T} (1+\varphi(\beta)) + \phi_{11} \omega \xi_t^{T} (1-\varphi(\beta))]^2}{[2\varphi(\beta) - 2\varphi(\beta)\phi_{11} - 1]^2} \tag{21}$$

Differentiating equation (21) with respect to ω and equating to 0 we have

$$\frac{\partial \sum a_t^2}{\partial \omega} = \frac{2(1+\varphi(\beta)) + \phi_{11}(1+\varphi(\beta))}{[2\varphi(\beta) - 2\varphi(\beta)\phi_{11} - 1]^2} = 0 \tag{22}$$

$$\sum_{t=1}^{n} [\phi_{12} \varphi(\beta) \ell_{t-1} - \omega(1+\varphi(\beta)) + \phi_{11} \omega (1+\varphi(\beta))] = 0$$

$$\phi_{12} \varphi(\beta) \ell_{t-1} - \omega(1+\varphi(\beta)) + \phi_{11} \omega (1+\varphi(\beta)) = 0$$

$$\omega(1+\varphi(\beta)) - \phi_{11} \omega\omega (1+\varphi(\beta)) = \phi_{12} \varphi(\beta) \ell_{t-1}$$

$$\omega\left[\left(1+\varphi(\beta)\right)-\phi_{11}\left(1+\varphi(\beta)\right)\right]=\phi_{12}\,\varphi(\beta)\ell_{t-1}$$

$$\hat{\omega}_C=\frac{\phi_{12}\,\varphi(\beta)\ell_{t-1}}{\left(1-\phi_{11}\right)\left(1-\varphi(\beta)\right)} \qquad (23)$$

The corresponding variance is

$$\mathrm{V}(\hat{\omega}_C)=\frac{\varphi_{12}\,\phi(\beta)^2\,\sigma_a{}^2}{\left(1-\varphi_{11}\right)^2\left(1-\phi(\beta)\right)^2} \qquad (24)$$

Therefore, the test statistic is

$$\lambda_i=\frac{\hat{\omega}_C}{S.e(\hat{\omega}_C)}$$

$$\lambda_i=\frac{\phi_{11}\,\varphi(\beta)\ell_{t-1}/\left(1-\phi_{11}\right)\left(1+\varphi(\beta)\right)}{\sqrt{\phi_{12}^2\,\varphi\left(\overset{2}{\beta}\right)\sigma_a^2\left|\left(1-\phi_{11}\right)^2\left(1+\varphi(\beta)\right)^2\right.}} \qquad (25)$$

$$=\frac{\phi_{12}\;\varphi(\beta)\ell_{t-1}}{\left(1-\phi_{11}\right)\left(1+\varphi(\beta)\right)}*\frac{\left(1-\phi_{11}\right)\left(1+\varphi(\beta)\right)}{\phi_{12}\;\varphi(\beta)\sigma_a}$$

$$\lambda_i=\frac{\ell_{t-1}}{\sigma_a} \qquad (26)$$

Table 1: Summary of Estimates and Test Statistic for the two models when X_{1t} contains outlier

MODELS	MAGNITUDE	VARIANCE	TEST STATISTIC
Additive	$\frac{\varphi(\beta)\left(1+\phi_{12}\,\lambda_t\right)}{1-\phi_{11}}$	$\left(\frac{\phi_{12}}{1-\phi_{11}}\right)^2\varphi\left(\overset{2}{\beta}\right)\sigma_a^2$	$\frac{\left(1+\phi_{12}\,\ell_t\right)}{\phi_{12}\,\sigma_a}$
Convolution	$\frac{\phi_{12}\,\varphi(\beta)\ell_{t-1}}{\left(1-\phi_{11}\right)\left(1-\varphi(\beta)\right)}$	$\frac{\phi_{12}\,\varphi\left(\overset{2}{\beta}\right)\sigma_a^2}{\left(1-\phi_{11}\right)^2\left(1-\varphi(\beta)\right)^2}$	$\frac{\ell_{t-1}}{\sigma_a}$

3. Application

In this section, analysis of both simulated and real data sets will be used to test the validity and efficiency of the derived outliers generating mechanisms. In other to compare the performance of the two newly derived models with the existing ones, Data on Nigerian Bank Deposits and Loans from Annual Statistical Bulletin of the Central Bank of Nigeria, 2011 were made used of.

From the derived outlier generating mechanisms in section 2 and with the estimation of the magnitudes of outliers and their variances, the test statistics constructed will be used to detect the existence of outliers in both the generated series and real data.

For the simulated data, a uniform distribution is assumed with contaminated observation with varying sizes of 10, 50, and 100. The data were analysed with the R-package of version 3.0.1.

3.1 Analysis of Simulated Data when X_{1t} Contains an Outlier

The results of the models on simulated data assuming a uniform distribution in terms of their outlier detection performance are tabulated below.

The sample sizes considered are 10, 50 and 100.

Table2: Summary of Result on Detection Rate of the Models on Simulated Data when X_{1t} contains outlier

	N=10			N=50			N=100		
Model Type	No of outliers injected	No of outliers detected	% of outliers detected	No of outliers injected	No of outliers detected	% of outliers detected	No of outliers injected	No of outliers detected	% of outliers detected
Additive	2	1	50	5	4	80	8	6	75
Convolution	2	2	100	5	5	100	8	8	100
Innovative*	2	0	0	5	2	40	8	2	25
Multiplicative*	2	2	100	5	4	80	8	5	80

Source* [19]

The Convolution model from the summary in Table2 had 100% outlier detection compared to Additive model as the sample size increases.

When compared with existing models, the Convolution model is most sensitive to outlying observations.

3.2 Detection of Outlier in Real Data

In order to investigate the performance of the proposed models, a pair of data on Deposit and Loan was used. The data was extracted from the Annual Statistical Bulletin of the Central Bank of Nigeria, 2011.

3.2.1. Assumed Model of Deposits and Loans

Here two cases are considered. The first case is when loan is contaminated.

The vector autoregressive model is given as

$$X_{1t} = \emptyset_{11} X_{1t-1} + \emptyset_{12} X_{2t-1} + \ell_t \quad (27)$$

where X_{1t} is the current value of deposit, X_{1t-1} is the immediate past value of deposit, and X_{2t-1} is the immediate past value of loan.

The estimated VAR model via the use of statistical package R is as follows

$$X_{1t} = 0.4826\ X_{1t-1} — 0.1579\ X_{2t-1} \quad (28)$$

s.e (0.1836) (0.1561)

t (2.628) (–1.012)

P-value (0.0142) (0.3210)

When deposit is contaminated, the vector autoregressive model is:

$$X_{2t} = \emptyset_{21} X_{2t-1} + \emptyset_{22} X_{1t-1} + \ell_t \quad (29)$$

where X_{2t} is the current value of loan,X_{2t-1} is the immediate past value of loan and X_{1t-1} is the immediate past value of deposit.

The estimated VAR model via the use of statistical package R is as follows

$$X_{2t} = 0.9605\ X_{2t-1} — 0.3339\ X_{1t-1} \quad (30)$$

S.e (0.1712) (0.2015)

t (5.610) (–1.657)

P (6.78e.06) (0.1095)

The detection performance of both Additive and Convolution models on the real data are shown on tables 3 and 4 below.

Table 3: Detection Performance of Additive Model on Deposit and Loan Data

Deposit	**Loan**	**(W_a)**	**T**	**Remarks**
111.7	35.9	-7425.1230	-3.5562723	ND*
131.2	44.2	-7287.8225	-3.4905121	ND
276.6	58.2	-7320.9776	-3.5063918	ND
311.4	114.9	-6766.7184	-3.2409286	ND
873.5	373.6	-6641.4233	-3.1809184	ND
1229.2	492.8	-6645.0523	-3.1826565	ND
1378.4	659.9	-2347.0674	-1.1241310	ND
5722.0	3721.1	-1321.7090	-0.6330342	ND
8360.1	4730.8	-214.7493	-0.1028544	ND
10580.7	5962.1	-7060.4273	-3.3816009	ND
4612.2	1895.3	10107.6624	4.8410781	D**
19542.2	10910.4	-10360.6635	-4.9622533	D
4855.2	1602.2	-790.9809	-0.3788413	ND
8807.1	8659.3	2051.1317	0.9823922	ND
12442.0	4411.2	6231.8871	2.9847704	ND
19047.6	11158.6	3575.8049	1.7126364	ND
18513.8	11852.7	1289.6953	0.6177012	ND
15860.5	7498.1	6662.9285	3.1912183	ND
20640.9	11150.3	1167.7164	0.5592793	ND
16875.9	12341.0	1158.2959	0.5547673	ND
14861.6	8942.2	7283.8830	3.4886253	ND
20551.8	11251.9	48840.8554	23.3923914	D
64490.0	34118.5	-14779.9643	-7.0788832	D
18461.9	16105.5	-10755.6646	-5.1514396	D
3118.6	24274.6	-2097.8978	-1.0047909	ND
3082.3	27263.5	8721.0511	4.1769588	D
13411.8	46521.5	-3338.5137	-1.5989855	ND
3296.2	15590.5	-2683.9759	-1.2854938	ND
3953.1	63769.4	747.8194	0.3581691	ND

D** = Outlier detected

ND* = No outlier detected

The critical value (c) = 4

Table 4: Detection Performance of Convolution Model on Deposit and Loan Data

Deposit	Loan	(W_c)	T	Remarks
111.7	35.9	543.83931	3.4944194	ND*
131.2	44.2	533.78300	3.4298029	ND
276.6	58.2	536.21138	3.4454064	ND
311.4	114.9	495.61569	3.1845603	ND
873.5	373.6	486.43869	3.1255938	ND
1229.2	492.8	486.70449	3.1273017	ND
1378.4	659.9	171.90659	1.1045794	ND
5722.0	3721.1	96.80611	0.6220241	ND
8360.1	4730.8	15.72891	0.1010655	ND
10580.7	5962.1	517.12786	3.3227859	ND
4612.2	1895.3	-740.31692	-4.7568790	D**
19542.2	10910.4	758.84751	4.8759466	D
4855.2	1602.2	57.93392	0.3722523	ND
8807.1	8659.3	-150.23132	-0.9653058	ND
12442.0	4411.2	-456.44297	-2.9328574	ND
19047.6	11158.6	-261.90317	-1.6828492	ND
18513.8	11852.7	-94.46133	-0.6069578	ND
15860.5	7498.1	-488.01380	-3.1357146	ND
20640.9	11150.3	-85.52721	-0.5495519	ND
16875.9	12341.0	-84.83722	-0.5451185	ND
14861.6	8942.2	-533.49445	-3.4279489	ND
20551.8	11251.9	-3577.25755	-22.9855362	D
64490.0	34118.5	1082.53098	6.9557628	D
18461.9	16105.5	787.77864	5.0618425	D
3118.6	24274.6	153.65662	0.9873149	ND
3082.3	27263.5	-638.75715	-4.1043105	D
13411.8	46521.5	244.52322	1.5711749	ND
3296.2	15590.5	196.58282	1.2631357	ND
3953.1	63769.4	-54.77264	-0.3519396	ND

D** = Outlier detected

ND* = No outlier detected

The critical value (c) = 4

Table 5: Summary of Outlier Detection of the Two Models on Deposits and Loan Data

Model	No of outliers detected
Convolution	6
Additive	6
**Innovation	5
**Multiplicative	Nil

Source: [19]

4. Discussion of Results

From the analyzed simulated data with varying sample sizes of 10, 50, and100, the average percentage rates of outlier detection for AO and CO are 68% and 100% respectively of the injected outliers. From the result, CO was consistent in outlier detection as the sample size increases. Comparing the performance of these two newly derived models with the existing models, the CO outperformed both Multiplicative and Innovative models that have average detection rate of 86.7% and 21.7% respectively for the simulated data.

For the real data set of Deposit and Loan, 6 outliers were equally detected by the two models when we consider the case of deposit depending on loan. The two derived outlier-generating mechanisms were able to detect potential outlier independently in multivariate time series. However, comparing the performance of these models with the existing ones, AO and CO detected 6 outliers while Innovative model was able to detect 5 but Multiplicative model detected no outlier as a result of non-multiplicative nature of data. [19].

In summary, CO was found to be most sensitive to outliers for the simulated data sets as the sample increases and also for the real data. When compared also with the existing models, CO has been found to be most efficient with minimum standard error of the estimate and is therefore recommended for outlier detection in multivariate time series data.

5. Conclusion

This work was undertaken to develop test statistic for detecting outliers assuming two different outliers generating mechanisms in multivariate time series models. In line with the main objective of this paper, the test statistics were derived for each generating mechanism namely; the Additive and Convolution models. The model with greatest detective power in terms of their sensitivity to the number of outliers detected by applying the models to both simulated and a pair of real data were determined. All these were achieved using theoretical and analytical methods. The convolution model was found to be most sensitive to outlier detection when compared with existing models, it is therefore recommended for outlier detection in multivariate time series.

References

[1] Z. Azami, A. Ibrahim and S. Mohd. “Detection Procedure for a Single Additive Outlier in Bilinear Model.” Journal of Pak. Stat. Oper. Res. Vol. No 1 PP. 1-5, 2007.

[2] R. Baragona and F. Battaglia. “Outlier Detection in Multivariate Time Series by Independent Component Analysis.” Neutral Computation, 19:1962-1984.

[3] R. Baragona, F. Battaglia and C. Alzini. “Genetic Algorithms for the Identification of Additive and Innovational Outliers in Time Series” Computational Statistics and Data Analysis. 30,147,2001.

[4] V. Barnett. “The study of outliers: Purpose and Model.” Applied Statistics, 27(3), 242–250,1978.

[5] V. Barnett and T. Lewis. Outlier in Statistical Data. John Wiley & Sons U. K. 1994.

[6] G.E.P. Box, G.M. Jenkins and G. Reinsel. Time Series Analysis: Forecasting and Control,3rd Ed., New Jersey: Prentice-Hall,1994.

[7] K. Chaloner and R. Brant. “A Bayesian Approach to Outlier Detection and Residual Analysis.” Biometrika, 25, 651 – 660,1988.

[8] I. Chang, et. al. “Estimation of Time Series Parameters in the Presence of Outliers” Technometrics, 3, 193.204,1988.

[9] C. Chen and L.M. Liu. “Joint Estimation of Model Parameters and Outlier effects in Time Series.” Journal of the American Statistical Association, 88, 284 – 297,1993.

[10] D. Cucina, A.Di Salvatore and M. Protopapas. ‘’Meta-heuristic Methods for Outliers Detection in Multivariate Time Series.’’ Comisef working paper series, 003,270,2008

[11] A.J. Fox. “Outliers in Time Series.” Journal of the Royal Statistical Society. B34: 350 – 363,1972

[12] P. Galeano, D. Pena and R.S. Tsay. “Outlier Detection in Multivariate Time Series via Projection Pursuit.” Working paper 0-42. Statistics and Econometrics Series II, Dept. De Estadistica, Universidad Carlos III de Madrid,2004

[13] J. Helbling and R. Cleroux. “On Outlier Detection in Multivariate Time Series.” Mathematical Volume 34, Number 1, pp. 19-26,2009.

[14] A. Kaya. “Modelling Outlier Factors in Data Analysis.” Advances in Information Systems, LNCS 3261, 88 – 95,2010.

[15] A. Khattree and D.N. Naik. “Detection of Outliers in Bivariate Time Series Data.” Communications in

Statistics – Theory and Methods, 16(12): 3701 – 3714,1987.

[16] G.M. Ljung. “On Outlier Detection in Time Series.” J. R. Statist. Soc. B. 55 No. 2, 559 -567,1993.

[17] C.R. Nelson, and C.L. Plosser. “Trends and Random Walks in Macroeconomic Time Series.” Journal of Monetary Economics, 10, 139 – 162,1982.

[18] D. Olivier and C. Amelie. “The Impact of Outliers on Transitory and Permanent Components in macroeconomic Series.” Economic Bulleting, Vol. 3, No 60 PP 1 – 9,2008.

[19] O. Olufolabo, O.I. Shittu and K.A. Adepoju. “Performance of Two Generating Mechanisms in Detection of Outliers in Multivariate Time Series.” American Journal of Theoretical and Applied Statistics.5(3), 115-122,2016.

[20] A. Pankratz. Forecasting with Univariate Box-Jenkins Models: Concepts and Cases, New York: John Wiley and Sins,1983.

[21] A. Pankratz. “Detecting and Treating Outliers in Dynamic Regression Models.” Biometrika,80, 47-54,1993.

[22] D. Pena and G.E.P. Box. “Identifying a Simplifying Structure in Time Series.” Journal of the American Statistical Association, 82, 836-843,1987.

[23] S. Ruey and R.S. Tsay. “Outliers, Level Shifts, and Variance Changes in Time Series.” Journal of Forecasting, Vol. 7, I-20 Department of Statistics, Carnegie Mellon University, U. S.A,1988.

[24] D.K. Shangodoyin. “On the Specification of time series Models in the Presence of Aberrant Observations”. Ph.D. Thesis in the Dept. of Statistics, Univ. of Ibadan,1994.

[25] I.O. Shittu and D.K. Shangodoyin. ‘’Detection of Outliers in Time Series Data: A Frequency Domain Approach.’’ Asian Journal of Scientific Research 1, (2) 130-137,2008.

[26] I. O. Shittu. “On Performance of Some Generating Models in Detection of Outliers Under Classical Rule.’’ M.Phil. Thesis. Dept. of Statistics, Univ. of Ibadan,2000.

[27] C. Sims. “Macroeconomics and Reality.” Econometricsa 48 (1), 11-46, JSTOR 112017,1980.

[28] R.S. Tsay. “Time Series Model specification in the Presence of Outlier.” Jour. Amer. Stat. Asso. 81, 132 – 141,1986.

[29] R.S. Tsay. “Outliers, Level Shifts and Variance Changes in Time Series.” Journal of Forecasting, 7, 1-20,1988.

[30] R.S. Tsay, D. Pena and A.E. Pankratz. “Outliers in multivariate time series.” Biometrika, 87, 789-804,2000.

[31] X.Wang. “Two-Phase Outlier Detection in Multivariate Time series.” Fuzzy Systems and Knowledge Discovery, Eighth International Conference. Vol.3,2011.

[32] Ji. Yanjie et al. (2013). “Detection of Outliers in a Time Series of Available Parking Spaces.” Mathematical Problems in Engineering, Volume 2013: 1-12,2013.