

Performance of Two Generating Mechanisms in Detection of Outliers in Multivariate Time Series

Olufolabo Olusesan Oluyomi.¹, Shittu Olarenwaju Ismail.², Adepoju Kazeem Adesola.²

¹Department of Statistics, Yaba College of Technology, Yaba, Nigeria

²Department of Statistics, University of Ibadan, Ibadan, Nigeria

Email address:

olufolabosesan@yahoo.com (Olufolabo O. O.)

To cite this article:

Olufolabo Olusesan Oluyomi., Shittu Olarenwaju Ismail., Adepoju Kazeem Adesola. Performance of Two Generating Mechanisms in Detection of Outliers in Multivariate Time Series. *American Journal of Theoretical and Applied Statistics*. Vol. 5, No. 3, 2016, pp. 115-122. doi: 10.11648/j.ajtas.20160503.16

Received: April 5 2016; **Accepted:** April 25, 2016; **Published:** May 10, 2016

Abstract: This work is focused on developing two outlier generating mechanisms for the detection of outliers in the multivariate time series setting that is capable of ameliorating the swamping effect on regular observations in time series data. Specifying two-variable Vector Autoregressive (VAR) models and assuming innovative and multiplicative effect of outliers on time series data, the magnitude and variance of outlier were derived for the generating models by method of least squares. A modified test statistics were also developed to detect single outliers both in the response and explanatory variables. Real and simulated data were used to establish the validity of the models. The results show that the multiplicative is better than the additive model in terms of the number of outliers detected and the residual variance. This result is in line with previous studies in outlier detection in univariate time series.

Keywords: Innovative Outlier, Additive Outlier, Multiplicative Outlier, Vector Auto Regressive

1. Introduction

In time series or any classical data, it has been established that outliers do cause biases in parameter estimation as well as model misspecification, and poor forecast performance to misleading conclusion. For this reason, several outlier detection techniques and robust estimation procedures have been proposed in the literature for univariate time series analysis but however very limited for multivariate time series.

“An early and detailed examination of detection of outliers in stationary univariate time series was done by Fox [18]”. Ever since, a quiet number of literature have been dedicated to the study of impact of outliers in univariate time series. Some of the authors include; Denby and Martin [15], Pena [33], Tsay [45], Chang, Tiao and Chan [11] in which they all use iterative procedure for the detection of outliers. R. Baragona, F. Battaglia and D. Cucina [5] “proposed Identification and estimation of outliers in time series by using empirical likelihood methods.” Theory and applications are developed for stationary autoregressive models with outliers distinguished in the usual additive and

innovation types. Pena and Maravall [34] considered the case of when the model is known and when it is unknown alongside the effect of missing data linked with outlier. Chan and Liu [13], McCulloch and Tsay [29]. Le Martin and Raftery [25] and Luceno [28] used the method based on robust Bayes factors in the consideration of additive outliers.

However, Justel, Pena and Tsay [47], in their paper, “proposes a procedure to detect patches of outliers in an autoregressive process”. The procedure is an improvement over the existing detection methods via Gibbs sampling. It was shown that the standard outlier detection via Gibbs sampling may become extremely inefficient in the presence of sever outlier.

Shittu [41], in his work, considered two additional outlier generating models, which are Multiplicative and Convolution and concluded that Convolution model preforms more efficiently than all other single outlier generating models. Ji. Yanjie, D. Tang, A. Gou, P. T. Blythe and G. Reu [48] in their work “introduced outlier mining and nonparametric detection methods for detecting and analyzing outlier in available parking space data sets. The technique was able to detect Additive and Innovative outlier simultaneously”.

Shittu and Sangodoyin [40], considered the identification of outliers in frequency domain using the spectral method.

The above and other literature shows that not much work has been done on outlier detection in multivariate time series. Among available works on multivariate outlier detection in time series is the projection pursuit techniques used by Galeano, Pefia and Tsay [19] to find the linear combination of a multivariate time series that maximizes kurtosis with the purpose of best reproducing the outlying signal. Detection of time points of outliers and estimating its magnitudes were accomplished by employing univariate searching methods.

Baragona and Battaglia [4] proposed the Independent Component Analysis (ICA) as a tool for identifying the locations of multiple outliers in multivariate time series. The ICA was therefore used at identifying a set of independent unobservable variables that are supposed to generate the data set of interest. An unknown mixing matrix was postulated to linearly transform the unobservable variables to produce a set of observable mixed ones. Both unobservable variables and the mixing matrix have to be estimated from the data. ICA has been applied successfully to a variety of fields such as biomedicine, speech, and radar, signal processing and time series.

In their own work, Cucina, Salvatore and Protopapas [14], used meta-heuristic methods to detect additive outliers in multivariate time series. The implemented algorithms were; simulated annealing, threshold accepting and two different versions of genetic algorithm. They used the same objective function, the generalized AIC-like criterion, and in contrast with many of the existing methods, they do not require specifying a vector auto regressive moving averages model for the data and are able to detect any number of potential outliers simultaneously. They concluded that “almost all available methods for outlier detection are iterative, but the difference with respect to the meta-heuristic algorithms is that it seems to be able to provide more flexibility and adaptation to the outlier detection problem”.

Furthermore, Robert and Cleroux, [44] in their own work, introduced the coefficient of vector autocorrelation, obtained its influence function together with its distribution, and used it for testing the hypothesis of presence of outliers.

Barnett and Lewis, [7] and Shittu, [41] emphasis on the challenges in outlier analysis; namely smearing and masking. These concepts are related to the detection of outliers in statistical data and can even be intertwined to complicate the situation even further. Smearing (popularly known as swamping in the literature of outlier identification in statistical data is talked of when one outlier affects the series in a manner that makes the other observations appear to be outliers as well even when they are actually not. Conversely, masking occurs when one outlier tends to hide the others from being identified. It is generally believed that these notions are closely connected to specific outlier detection methods and not properties of data itself and smearing and masking are only deficiencies of certain methods, not types of outliers as such.

As a result of the effect of both Additive and Innovative

outliers on the estimates of parameters, Shittu [41] introduced the Convolution Outlier (CO) and Multiplicative Outlier (MO) models in univariate time series. To this effect Multiplicative and Innovative generating mechanisms outlier will be extended to multivariate time series in this paper with a view to comparing their performance in terms of parameter estimates and outlier detection capabilities.

2. Derivation of the Models

In this section, by assuming that outliers have either Innovative or Multiplicative effect on a series for bivariate time series. The estimate of the parameter shall be derived and the corresponding test statistics developed.

2.1. Innovative Outlier Model

An Innovative Outlier (IO) represents an unexpected change in the innovations that drive the vector time series. Suppose that the noise in a bivariate series consisting of oven temperature and a chemical concentration reading is mainly due to the random variation of the feed rate. Then a sudden change in the feed rate that happens at just a particular time point, due to some exogenous effect, will produce an IO in the series.

The innovative outlier-generating model for univariate series is defined as:

$$X_t = Z_t + \phi(B) \omega \xi_t^{\xi(t)} \quad (1)$$

with the unobservable free series given by

$$Z_t = \frac{\theta(B)}{\phi(B)} a_t \quad (2)$$

where $a_t \sim \mu(0, \sigma_{a_t}^2)$, $Var(e_t) = \sigma_{a_t}^2$ and

$$\ell_t = a_t + \phi(B) \omega \xi_t^{\xi(t)} \quad (3)$$

where $X_{it} = (x_{1t}, \dots, x_{kt})$ is a k-dimensional time series, Z_t is an outlier free time series that is assumed to be ARIMA(p, q), $\xi_t^{(T)}$ is a time indicator for outliers such that $\xi_t^{(T)} = 1$ for all $t = T$ and $\xi_t^{(T)} = 0$ otherwise, $\phi(B) = 1 - \Theta_1 B - \Theta_2 B^2 \dots - \Theta_p B^p$ are polynomials of order p and $\omega = (\omega_1, \dots, \omega_k)'$ represent the size of the magnitude of outliers.

Now, given a vector model 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 bivariate VAR(2, 2) as:

$$X_{1t} = \phi_{11} X_{1t-1} + \phi_{12} X_{2t-1} + a_{1t} \quad (4)$$

$$X_{2t} = \phi_{21} X_{2t-1} + \phi_{22} X_{1t-1} + a_{2t} \quad (5)$$

Where; X_{1t} is the current value of the response variable

X_{1t-1} is the lag value of the current variable

X_{2t} is the current value of the explanatory variable

X_{2t-1} is the lag value of the explanatory variable

Now, when X_{1t} contains an outlier,

Then

$$X_{1t} = Z_{1t} + \phi(B)\omega\xi_t^{(T)} \quad (6)$$

Substituting (6) into (4), to have:

$$Z_{1t} + \phi(B)\omega\xi_t^{(T)} = \varphi_{11}(Z_{1t-1} + \phi(B)\omega\xi_{t-1}^{(T)}) + \varphi_{12}X_{2t-1} + a_{1t} \quad (7)$$

recall that $e_t = \pi(B)X_t$ and $Z_t = \frac{\theta(B)}{\phi(B)}dt$

$$X_t = \frac{e_t}{\pi(B)} = \phi(B)e_t \quad (8)$$

We then have

$$\frac{\theta(B)}{\phi(B)}a_t + \phi(B)\omega\xi_t^{(T)} = \varphi_{11}\left(\frac{\theta(B)}{\phi(B)}dt + \phi(B)\omega\xi_{t-1}^{(T)}\right) + \varphi_{12}\phi(B)e_{t-1} + a_{1t} \quad (9)$$

$$\frac{\theta(B)}{\phi(B)}at - \varphi_{11}\frac{\theta(B)}{\phi(B)}a_t - a_{1t} = \varphi_{11}\phi(B)\omega\xi_t^{(T)} + \varphi_{12}\phi(B)e_{t-1} - \phi(B)\omega\xi_t^{(T)}$$

assuming $a_t = a_{t-1}$; when $\xi_t^T = 1$

$$a_t \left[\frac{\theta(B)}{\phi(B)} - \varphi_{11}\frac{\theta(B)}{\phi(B)} - 1 \right] = \varphi_{11}\phi(B)\omega_t^T - \phi(B)\omega_t^{(T)} + \varphi_{12}\phi(B)e_{t-1}$$

therefore

$$a_t = \frac{\varphi_{11}\phi(B)\omega_t^{(T)} - \phi(B)\omega_t^{(T)} + \varphi_{12}\phi(B)e_{t-1}}{\frac{\theta(B)}{\phi(B)}(1 - \varphi_{11}) - 1}$$

Using the least squares method to obtain

$$\sum_{t=1}^n a_t^2 = \sum_{t=1}^n \left[\frac{\varphi_{11}\phi(B)\omega_t^{(T)} - \phi(B)\omega_t^{(T)} + \varphi_{12}\phi(B)e_{t-1}}{\frac{\theta(B)}{\phi(B)}(1 - \varphi_{11}) - 1} \right]^2$$

$$\frac{\partial \sum a_t^2}{\partial \omega} = \frac{[2\varphi_{11}\phi(\beta)\omega_t^{(T)} - \phi(\beta)\omega_t^{(T)}]}{\left[\frac{\theta(\beta)}{\phi(\beta)}(1 - \varphi_{11}) - 1 \right]^2} \sum_{t=1}^n [\varphi_{11}\phi(\beta)\omega_t^{(T)} - \phi(\beta)\omega_t^T + \varphi_{12}\phi(\beta)e_{t-1}] = 0 \quad (10)$$

$$\sum_{t=1}^n [\varphi_{11}\phi(B)\omega_t^{(T)} - \phi(B)\omega_t^T + \varphi_{12}\phi(B)e_{t-1}] = 0 \quad t = T \text{ and } \xi_t^{(T)} = 0 \text{ otherwise, we have}$$

$$\text{Since } \xi_t^{(T)} \text{ is a time indicator where } \xi_t^{(T)} = 1 \text{ for all } \sum_{t=1}^n \omega(\phi(B) - \phi(B)) + \varphi_{12}\phi(B) \sum_{t=1}^n e_{t-1} = 0 \quad (11)$$

$$\sum_{t=1}^n \omega(\phi(B) - \phi_{11}\phi(B)) = \phi_{12}\phi(B) \sum_{t=1}^n e_{t-1} \quad \lambda_i = \frac{\phi_{12} e_{t-1} / (1 - \phi_{11})}{\frac{\phi_{12}}{1 - \phi_{11}} \sigma_a} = \frac{e_{t-1}}{\sigma_a} \quad (16)$$

$$\omega\phi(B) (1 - \phi_{11}) = \phi_{12}\phi(B)e_{t-1}$$

Therefore, the estimator of the magnitude outlier for IO is

$$\hat{\omega}_t = \frac{\phi_{12}}{1 - \phi_{11}} e_{t-1} \quad (12)$$

Its variance is

$$V(\hat{\omega}_t) = \left(\frac{\phi_{12}}{1 - \phi_{11}} \right)^2 V(e_{t-1}) \quad (13)$$

Therefore

$$V(\hat{\omega}_t) = \left(\frac{\phi_{12}}{1 - \phi_{11}} \right)^2 \sigma_a^2 \quad (14)$$

Having obtained the estimate and its corresponding variance, we then construct the test statistic for innovative model as

$$\lambda_i = \frac{\hat{\omega}_t}{S.e(\hat{\omega}_t)} \quad (15)$$

2.2. Multiplicative Outlier Model

Since outlier may have multiplicative interaction effect on a series (Shittu, 2003), there is need to develop the outlier generating model.

The multiplicative outlier model is defined as:

$$X_t = Z_t \omega \xi_t^{(T)} \quad (17)$$

again using $e_t = \pi(B)X_t$ and $Z_t = \frac{\theta(B)}{\phi(B)} a_t$

Specifying $X_{1t} = \phi_{11} X_{1,t-1} + \phi_{12} X_{2,t-1} + a_{1t}$ as defined in equation (4)

with the outlier free series $Z_t = \frac{\theta(B)}{\phi(B)} a_t = \phi(B) a_t$

linearize (17) by taking the logarithm to have

$$\log(\pi(B)^{-1}) = \log(\pi(B)^{-1}) a_t + \log \omega \in_t^{(T)} \quad (18)$$

$$(a_t - 1) \log \pi(B)^{-1} = \log w \in_t^{(T)}$$

Let $a_t = \log[\pi(B)^{-1}] \ell_t$, $\Omega_t^T = \log \omega \in_t^{(T)}$ and $a_t = \ell_t - \Omega_t^T$

Therefore, $Z_t \omega \in_t^{(T)} = \phi_{11} Z_{t-1} \omega \in_{t-1}^{(T)} + \phi_{12} X_{2,t-1} + a_t$

$$\log Z_t + \log \omega \xi_t^{(T)} = \phi_{11} [\log Z_{t-1} + \log \omega \xi_{t-1}^{(T)}] + \phi_{12} X_{2,t-1} + a_t \quad (19)$$

since, $Z_t = \frac{\theta(B)}{\phi(B)} a_t = \phi(B) a_t$

then $\log Z_t = \log \phi(B) a_t$

$$\log \phi(B) a_t \log \omega \xi_t^{(T)} = \phi_{11} [\log \phi(B)] a_{t-1} + \log \omega \xi_{t-1}^{(T)} + \phi_{12} \phi(B) e_t + a_t \quad (20)$$

If we let $\Omega_t^T = \log \omega \xi_t^{(T)}$, $e_t = \log \phi(B) a_t$ and $a_t = e_t - \Omega_t^T$

Then we have

$$e_t + \Omega_t^{(T)} = \phi_{11} (e_{t-1} + \Omega_{t-1}^{(T)}) + \phi_{12} \phi(B) e_t + a_t \quad (21)$$

$$e_t + \Omega_t^{(T)} = \phi_{11} e_{t-1} + \phi_{11} \Omega_{t-1}^{(T)} + \phi_{12} \phi(B) e_t + a_t$$

$$a_t = \ell_t + \Omega_t^{(T)} - \phi_{11} \ell_{t-1} - \phi_{11} \Omega_{t-1}^{(T)} - \phi_{12} \phi(B) \ell_t$$

By sum of squares of a_t we have:

$$\sum_{t=1}^n a_t^2 = \sum_{t=1}^n [\ell_t + \Omega_t^{(T)} - \phi_{11} \ell_{t-1} - \phi_{11} \Omega_{t-1}^{(T)} - \phi_{12} \phi(B) \ell_t]^2 \quad (22)$$

Differentiating equation (32) with respect to Ω_t and equating to zero, to get

$$\frac{\partial \sum a_i^2}{\partial e} = 2(e_i - \phi_{11}) \sum_{i=1}^n (e_i + \Omega_i^{(T)} - \phi_{11}e_{i-1} + \phi_{11}\Omega_i^{(T)} - \phi_{12}\phi(B)e_i) = 0 \quad (23)$$

$$(1 - \phi_{11})\Omega_i^{(T)} + e_i - \phi_{11}e_{i-1} - \phi_{12}\phi(B)e_i = 0$$

$$\Omega_i^{(T)} = \frac{\phi_{11}e_{i-1} + \phi_{12}\phi(B)e_i - e_i}{1 - \phi_{11}} \quad (24)$$

Setting $e_i = e_{i-1}$

$$\Omega_i^{(T)} \approx \frac{e_i \left[(\phi_{11} + \phi_{12}\phi(B) - 1) \right]}{1 - \phi_{11}} \quad (25)$$

recall that $\Omega_i^{(T)} = \log \omega \xi_i^{(T)}$ for $\xi_i^{(T)} = 1$ in the presence of outlier, we have

$$\log \omega_m \equiv \frac{e_i \left[\phi_{11} + \phi_{12}\phi(B) - 1 \right]}{1 - \phi_{11}}$$

$$\hat{\omega} \equiv \text{Anti log} \frac{e_i \left[\theta_{11} + \theta_{12}\phi(B) - 1 \right]}{1 - \phi_{11}} \quad (26)$$

$$V(\hat{\omega}_m) = \text{Anti log} \left[\frac{\phi_{11} + \phi_{12}\phi(B) - 1}{1 - \phi_{11}} \right]^2 \sigma_a^2 \quad (27)$$

Hence the test statistic is defined as:

$$\lambda_i = \frac{\hat{\omega}_m}{S.e(\hat{\omega}_m)} \quad (28)$$

$$= \frac{\text{Anti log} \left[\frac{e_i (\phi_{11} + \phi_{12}\phi(B) - 1)}{1 - \phi_{11}} \right]}{\left[\left(\frac{(\phi_{11} + \phi_{12}\phi(B) - 1)}{1 - \phi_{11}} \right) \sigma_a \right]}$$

The variance of $\hat{\omega}_m$ is

$$\lambda_i = \frac{e_i}{\sigma_a^2} \quad (29)$$

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

MODELS	MAGNITUDE	VARIANCE	TEST STATISTIC
Innovative	$\left(\frac{\phi_{12}e_{i-1}}{1 - \phi_{11}} \right)$	$\left(\frac{\phi_{12}}{1 - \phi_{11}} \right)^2 \sigma_a^2$	$\frac{e_{i-1}}{\sigma_a}$
Multiplicative	$\text{Anti log} \frac{e_i \left[\phi_{11} + \phi_{12}\phi(B) - 1 \right]}{1 - \phi_{11}}$	$\text{Anti log} \left[\frac{\phi_{11} + \phi_{12}\phi(B) - 1}{1 - \phi_{11}} \right]^2 \sigma_a^2$	$\text{Anti log} \left(\frac{e_i}{\sigma_a} \right)$

3. Analysis of Data

From the derived outlier generating mechanisms in section two 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.

Simulation data of varying sample sizes of 10, 50, and 100 will be used to evaluate the performance of the derived models, while data of commercial bank deposits and loans from Nigerian commercial banks extracted from the Annual

Statistical Bulletin of the Central bank of

Nigeria, 2011 will also be used to establish the validity of the developed models.

Statistical software R3.0.1 is used to analysing the data. The results and outcome for the two models i.e. Innovative and Multiplicative models are summarised below.

3.1. Analysis of Simulated Data When X_{it} Contains Outliers

The results of the two models in terms of their outlier detection performance from simulated data are tabulated below.

Table 2. Summary of Result on Detection Rate of the Models on Simulated Data when X_{it} contains outlier.

Model Type	n=10			n=50			n=100		
	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
Innovative	2	0	0%	5	2	40%	8	2	25%
Multiplicative	2	2	100%	5	4	80%	8	5	80%

As shown in Table 2 above, the multiplicative model is more sensitive to outlying observations than the innovative model for different sample sizes.

3.2. Detection of Outlier in Real Data

In order to investigate the performance of the derived models, a pair of real data on Deposit and Loan of banks in Nigeria obtained from the Annual statistical bulletin of the Central bank of Nigeria, 2011 were used.

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

is contaminated. The vector autoregressive model is given as

$$X_{1t} = \phi_{11}X_{1t-1} + \phi_{12}X_{2t-1} + \ell_t$$

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 granted.

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

$$X_{1t} = 0.4826 X_{1t-1} - 0.1579 X_{2t-1}$$

s.e (0.1836) (0.1561)

t (2.628) (-1.012)

P-value (0.0142) (0.3210)

The second case is when Deposit is contaminated.

Then, the vector autoregressive model is given as

$$X_{2t} = \phi_{21}X_{2t-1} + \phi_{22}X_{1t-1} + \ell_t$$

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

$$X_{2t} = 0.9605 X_{2t-1} - 0.3339 X_{1t-1}$$

s.e (0.1712) (0.2015)

t (5.610) (-1.657)

P (6.78e.06) (0.1095).

Table 3. Detection Performance of Innovative Model on Deposit and Loan Data when Loan when Loan is contaminated.

Deposit (x10 ⁶ Naira)	Loan (x10 ⁶ Naira)	ω_I	λ_t	Remark
111.7	35.9	3094.71	-3.49442	ND
131.2	44.2	-3037.48	-3.4298	ND
276.6	58.2	-3051.3	-3.44541	ND
311.4	114.9	-2820.29	-3.18456	ND
873.5	373.6	-2768.07	-3.12559	ND
1,229.20	492.8	-2769.58	-3.1273	ND
1,378.40	659.9	-978.232	-1.10458	ND
5,722.00	3,721.10	-550.874	-0.62202	ND
8,360.10	4,730.80	-89.5051	-0.10107	ND
10,580.70	5,962.10	-2942.71	-3.32279	ND
4,612.20	1,895.30	4212.763	4.756879	D
19,542.20	10,910.40	-4318.21	-4.87595	D
4,855.20	1,602.20	-329.672	-0.37225	ND
8,807.10	8,659.30	854.8892	0.965306	ND
12,442.00	4,411.20	2597.382	2.932857	ND
19,047.60	11,158.60	1490.356	1.682849	ND
18,513.80	11,852.70	537.5308	0.606958	ND
15,860.50	7,498.10	2777.035	3.135715	ND
20,640.90	11,150.30	486.6914	0.549552	ND
16,875.90	12,341.00	482.765	0.545119	ND
14,861.60	8,942.20	3035.842	3.427949	ND
20,551.80	11,251.90	20356.33	22.98554	D
64,490.00	34,118.50	-6160.13	-6.95576	D
18,461.90	16,105.50	-4482.84	-5.06184	D
3,118.60	2,474.60	-874.381	-0.98731	ND
3,082.30	2,763.50	3634.838	4.104311	ND
13,411.80	4,652.50	-1391.46	-1.57117	ND
3,296.20	1,590.50	-1118.65	-1.26314	ND
3,953.10	2,376.90	311.6829	0.35194	ND

D = Outlier detected

ND = No outlier detected

The critical value (c) = 4.

Table 4. Detection Performance of Multiplicative Model on Deposit and Loan Data when Loan is contaminated.

Deposit (Billion Naira)	Loan (Billion Naira)	ω_I	λ_t	Remark
111.7	35.9	Inf	NA	ND
131.2	44.2	Inf	NA	ND
276.6	58.2	Inf	NA	ND
311.4	114.9	Inf	NA	ND
873.5	373.6	Inf	NA	ND
1,229.20	492.8	Inf	NA	ND
1,378.40	659.9	Inf	NA	ND
5,722.00	3,721.10	Inf	NA	ND
8,360.10	4,730.80	Inf	NA	ND
10,580.70	5962.1	Inf	NA	ND
4,612.20	1,895.30	0	NA	ND
19,542.20	10,910.40	Inf	NA	ND
4,855.20	1,602.20	0	NA	ND
8,807.10	8,659.30	Inf	NA	ND
12,442.00	4,411.20	0	NA	ND
19,047.60	11,158.60	0	NA	ND
18,513.80	11,852.70	Inf	NA	ND
15,860.50	7,498.10	0	NA	ND
20,640.90	11,150.30	0	NA	ND
16,875.90	12,341.00	Inf	NA	ND
14861.6	8942.2	0	NA	ND
20,551.80	11,251.90	0	NA	ND
64,490.00	34,118.50	Inf	NA	ND
18,461.90	16,105.50	0	NA	ND
3,118.60	2,474.60	Inf	NA	ND
3,082.30	2,763.50	0	NA	ND
13,411.80	4,652.50	Inf	NA	ND
3,296.20	1,590.50	0	NA	ND
3,953.10	2,376.90	0	NA	ND

It could be deduced from the Table 4 above that no outlier was detected for multiplicative as a result of non-multiplicative nature of the data.

4. Discussion of Results

Results obtained from the simulated data with varying sample sizes (from small, moderate to large sample) of 10, 50, and 100 gave an average detection rates for Innovative Outlier model (IO) and Multiplicative Outlier model (MO) as (0% and 100%), (40% and 80%) and (25% and 80%) respectively for sample sizes 10, 50 and 100 for the injected outliers. However, as the sample size increases, MO was found to be most sensitive to outliers considering the simulated data sets.

For the real data set of Deposit and Loan, 5 pairs of observations were identified as outliers by IO, however, MO could not identify any outlier as a result of non-multiplicative nature of the data.

Considering the two-outlier detection models, MO 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 paper introduced outlier generating mechanism in multivariate time series using VAR. It also developed test statistic for detecting outliers assuming two different nature of outliers, the innovative and multivariate models. The test statistics were derived for each generating mechanism. Attempts were made also to unravel the model with greatest detective power in terms of relative efficiency and their sensitivity to outliers by applying the models to both simulated and real data. All these were achieved using theoretical and analytical means. The multiplicative model was found to be more sensitive to outlier detection but its ability to detect outliers in real data depends heavily on the nature of the series (whether the bound is multiplicative or not)

This work can be further extended to the frequency domain since this work is limited to time domain.

References

- [1] A. Arribas-Gil, and J. Romo, "Shape Outlier Detection and Visualization for Functional Data: the Outliergram," *Biostatistics*, 15, 603-619, 2014.
- [2] Z. Azami, A. Ibrahim, and S. Mohd, S, "Detection Procedure for a Single Additive Outlier in Bilinear Model" *Journal of Pak. Stat. Oper. Res.* 1, 1-5, 2007.
- [3] R. Baragona, and F. Battaglia, "Outlier Detection in Multivariate Time Series by Independent Component Analysis". *Neural Computation*, 19: 1962-1984, 2007.
- [4] R. Baragona, F. Battaglia and D. Cucina, "Empirical Likelihood for Outlier Detection and Estimation in Autoregressive Time Series" *Journal of Time Series Analysis* Vol 37, 3, 315-336, 2015.
- [5] R. Baragona, F. Battaglia, and T. Calzini, "Genetic Algorithms for the Identification of Additive and Innovational Outliers in Time Series." *Computational Statistics and Data Analysis* 30, 147, 2001.
- [6] V. Barnett, and T. Lewis, *Outlier in Statistical Data*, John Wiley & Sons U. K, 1995.
- [7] G. E. P. Box, G. M. Jenkins, and G. Reinsel, *Time Series Analysis: Forecasting and Control*, 3rd Ed., New Jersey: Prentice-Hall, 1994.
- [8] K. Chaloner and R. Brant, "A Bayesian Approach to Outlier Detection and Residual Analysis," *Biometrika*, 25, 651-660, 1988.
- [9] I. Chang, "Outlier in time series". Technical Report, Department of statistics, University of Wisconsin, 1982.
- [10] I. Chang, G. D. Tiao, and C. Chen, "Estimation of Time Series Parameters in the Presence of Outliers. *Technometrics*," 3, 193. 204, 1988.
- [11] S. Chattfield, *The Analysis of Time series an introduction*, New York, Charian and Hall, 1980.
- [12] 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.
- [13] 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.
- [14] L. Denby, and R. D. Martin, "Robust Estimation of the First Order Autoregressive Parameter," *Journal of the American Statistical Association*, 74, 140-146. 1979.
- [15] W. Enders, *Applied Econometric Time Series*, 2nd Edition, John Wiley & Sons, ISBN 0-471-23065-0, 2003.
- [16] M. Forni, and I. Reichlin, "Let's Get Real: A Dynamic Factor Analytical Approach to Disaggregated Business Cycle". *Review of Economic Studies*, 65, 453/474, 1998.
- [17] A. J. Fox, "Outliers in Time Series". *Journal of the Royal Statistical Society*. B34: 350 – 363, 1972.
- [18] 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 Estadística, Universidad Carlos III de Madrid, 2004.
- [19] I. Georgiev, "A Factor Model for Innovational Outliers in Multivariate Time Series", ICEE, First Italian Congress of Econometrics and Empirical Economics, Venice, 24-25, 2005.
- [20] A. Justel, D. Pena, and R. S. Tsay, "Detection of Outlier Patches in Autoregressive Time Series," *Statistica Sinica*, 11, 651-673. 2000.
- [21] A. Kaya, "An Investigation: The Analysis of Outliers in Time Series". PhD Thesis. Dokuz Eylül University, Izmir, Turkey, 1999.
- [22] A. Kaya, "Modelling Outlier Factors in Data Analysis", (advances in Information Systems), LNCS 3261, 88-95, 2010.
- [23] A. Khatree and D. N. Naik, "Detection of Outliers in Bivariate Time Series Data". *Communications in Statistics – Theory and Methods*, 16 (12): 3701-3714, 1987.
- [24] N. D. Le, A. E. Raftery, and R. D. Martin, R. D. "Robust Order Selection in Autoregressive Models Using Robust Bayes Factors". *Journal of the American Statistical Association*, 91, 123-131. 1996a.
- [25] G. M. Ljung, "On Outlier Detection in Time Series," *J. R. Statist. Soc. B*. 55 No. 2, 559-567, 1993.
- [26] A. Luceno. Detecting Possibly Non-Consecutive Outliers in Industrial Time Series.
- [27] *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 60 (2): 295-310, 1998.
- [28] H. Lutkepohl, *New Introduction to Multiple Time Series Analysis*, Springer, Berlin, 2005.
- [29] R. E McCulloch, and R. S. Tsay, "Bayesian Analysis of Autoregressive Time Series Via the Gibbs sampler", *Journal of Time Series Analysis* 15, 235-50, 1994.
- [30] C. R. Nelson, and C. I. Plosser, "Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics*, 10, 139-162, 1982.
- [31] D. Olivier, and C. Amelie, "The Impact of Outliers on Transitory and Permanent Components in Macroeconomic Series". *Economic Bulletin*, Vol. 3, No 60 PP 1-9, 2008.

- [32] A. Pankratz, "Detecting and Treating Outliers in Dynamic Regression Models," *Biometrika*, 80, 847-54, 1993.
- [33] D. Pena, and G. E. D. Box, "Identifying a Simplifying Structure in Time Series", *Journal of the American Statistical Association*, 82, 836-843, 1987.
- [34] D. Pena, and A. Maravall, "Interpolations, Outliers and Inverse Autocorrelations", *Communications in Statistics, Theory and Methods* 20, 3175-86. 1991.
- [35] B. Rolin, "Comparing Classical and Resistant Outlier Rules," *Journal of American Stat. Ass.* Vol. 412 pp. 1083–1090, 1990.
- [36] B. Rosner, "On the Detection of Many Outliers," *Technometrics*. 17, 221–227, 1995.
- [37] S. Ruey and R. S. Tsay, "Outliers, Level Shifts, and Variance Changes in Time Series. *Journal of Forecasting*, Vol. 7, 1-20 Department of Statistics, Carnegie Mellon University, U.S.A, 1988.
- [38] M. J. Sanchez, D and Pena, D. "The Identification of Multiple Outliers in ARIMA Models," *Communications in Statistics, Part A—Theory and Methods*, 32, 1265–1287. 2003.
- [39] 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.
- [40] 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.
- [41] I. O. Shittu, "On Performance of Some Generating Models in Detection of Outliers Under Classical Rule" Mphil Thesis Dept. of Statistics, Univ. of Ibadan, 2000.
- [42] C. Sims, "Macroeconomics and reality" *Econometrica* 48 (1), JSTOR 112017, 1980.
- [43] S. Sridevi, S. Abirami, and S. Rajaram, "Detecting and Revamping of X-Outliers in Time Series Database," *International Journal of Computer Applications* 60 (19): 28-33, 2012.
- [44] C. Robert and J. Helbling, "On Outlier Detection in Multivariate Time Series" *Acta Mathematica Vietnamica*, 34, 1, 19-26, 2009.
- [45] R. S. Tsay, "Time Series Model Specification in the Presence of Outlier". *Jour. Amer. Stat. Asso.* 81, 132–141, 1986.
- [46] R. S. Tsay, "Outliers, Level Shifts and Variance changes in Time Series," *Journal of Forecasting*, 7, 1-20, 1988.
- [47] R. S. Tsay, D. Pena, and A. E. Pankratz, "Outliers in Multivariate Time Series. *Biometrika*, 87, 789-804, 2000
- [48] Ji. Yanjie, D. Tang, A. Gou, P. T. Blythe and G. Reu, "Detection of Outliers in a Time Series of Available Parking Spaces," *Mathematical Problems in Engineering*. Volume 2013: 1-12, 2013.