ISSN 1684-8403
Journal of Statistics
Volume 18, 2011, pp. 67-83

Role of Structural Coefficients in the Study of the Influence of an Additive Outlier on the Residuals in ARMA (1,1), AR(1), and MA(1) Models

Jallikop Vaman Janhavi[1] *and Suresh Ramaiah*[2]

Abstract

Motivation for this paper comes from our perception that the Simulation study on nature of Additive Outlier (AO) in ARMA(1,1) model by Zaharim et al. (2009) ought to be more extensive for a fuller understanding of the influence of the Outlier. Stated briefly, Zaharim et al. (2009) based on Simulations of ARMA(1,1) Time Series model with one AO, have concluded that the AO affects the observations only at its time of occurrence, and affects, noticeably, the residuals at the time instance of the AO and the one immediately next to it. In this paper, we have demonstrated, both analytically and through Simulation, that in the case of ARMA(1,1) with an AO one or more residuals can get affected depending on the values of the model parameters. The number of residuals that get noticeably affected depends on both the magnitude of AO and also the values of the parameters of the underlying model. In order to gain a deeper insight into this behavior of residuals, we narrow down the study to AR(1) and MA(1). This type of investigation reveals that in the case of AR(1), an AO affects the residuals at its time point of occurrence, and at the subsequent point. In contrast, in MA(1) model, all the residuals at and after the incidence of AO can get affected (as in the case of ARMA(1,1) model), emphasizing the crucial role of moving average parameter. These deeper implications have escaped the attention of the studies made by Zaharim et al. (2009).

[1] Department of Statistics, Bangalore University, Bangalore 560 056, India.
Email: jvjanhavi@gmail.com
[2] Department of Statistics, Bangalore University, Bangalore 560 056, India.
Email: sureshramaiah@gmail.com

Keywords

ARMA(1,1) model, AR(1) model, MA(1) model, Additive outlier(AO), Simulation study.

1. Introduction

Observed Time Series are sometimes influenced by Outliers that may be the result of gross errors of measurement, collection and processing of the data, or to some unusual event influencing the phenomenon under study, such as wars, strikes, an economic crisis or a temporary change in experimental conditions such as special promotional schemes or system changes, and so forth. There are many descriptions of Outliers proposed by different authors in the literature, one commonly used informal definition (Barnett and Lewis, 1984) is; An Outlier in a set of data is an observation or a patch of observations which appears to be inconsistent with the remainder of that set of data. The inconsistency refers to the case where this outlying observation (or group of observations) is generated by some mechanism other than that of the rest (i.e., the majority) of the data. Fox (1972) discussed two characterizations of Outliers that are found in Time Series data. It appears that he is the first to consider Outliers in Time Series. He called them as Type I and Type II, which are now known as Additive Outlier (AO) and Innovational Outlier (IO). In the literature, we have come across other types of Outliers like, Level Change (LC) and Transient Change (TC). However, Additive Outliers are in practice more common than the other types of Outliers. These Outliers have adverse effects in many situations like model building and forecasting in Time Series analysis which have been theoretically shown. For details, see (Abraham and Chuang, 1989; Chen and Liu, 1993; Ledolter, 1989; Tsay, 1986 and 1988).

The main goal of the present work is to study exhaustively the effect of model parameters on the observations and specifically on the residuals in the presence of a single AO for different time series models like Autoregressive Moving Average of order (1,1) (ARMA(1,1)) and its particular cases Autoregressive of order one (AR(1)) and Moving Average of order one (MA(1)). We use some of the deductive steps in Zaharim et al. (2009) which are on the same lines as in Wei (2006) for the sake of readability and completeness. Further, we narrow the study to its particular cases; AR(1) and MA(1). In the study, we assume that the values of the parameters in the respective models, the time of occurrence ($t = T$) of AO and also its magnitude δ, are known.

Let, $\{x_t\}$ be an Autoregressive Moving Average process of order p and q, ARMA(p,q) defined as;

$$\emptyset(B)x_t = \theta(B)a_t \tag{5.1}$$

where, $\{a_t\}$ is a sequence of independent and identically distributed Gaussian variables with mean zero and variance σ_a^2 and the polynomial $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ are polynomials in B, B is the backward shift operator such that $Bx_t = x_{t-1}$ and $\emptyset(B) = 0$ and $\theta(B) = 0$ have all roots outside the unit circle. The Autoregressive part of order p, AR(p), is then stationary and the Moving Average part of order q, MA(q), is invertible, i.e., it can be written in terms of AR(∞) representation (Box et al., 1994). Equation (5.1) can be written as;

$$x_t = \frac{\theta(B)}{\phi(B)} a_t \tag{5.2}$$

Note that, if in the above model (5.2) the values of p and q are equal to one then we get ARMA(1,1) model. Now on, we shall focus our discussion on this model. An observed Time Series has an AO at time $t = T$ of size δ if it satisfies;

$$y_t = x_t + \delta I_t^{(T)} \tag{5.3}$$

where,

y_t is the observed and contaminated series, x_t is unobservable Outlier-free series in (5.1) with p and q equal to one, δ represents the magnitude of the AO, and $I_t^{(T)}$ is an Indicator variable such that $I_t^{(T)} = 1$ when an Outlier is spotted at $t = T$ and $I_t^{(T)} = 0$ otherwise. The model given in (5.3) is known an Additive Outlier ARMA(1,1) model.

Using (5.2) equation, (5.3) can be written as;

$$y_t = \frac{\theta(B)}{\phi(B)} a_t + \delta I_t^{(T)} \tag{5.4}$$

The remainder of this paper is organized as follows. Section 2 deals with the routine algebraic evaluation of residuals with respect to the ARMA(1,1), and its particular cases AR(1) and MA(1) in detail. The effect of AO is investigated graphically through a Simulation study in Section 3. A discussion of the results and conclusions are presented in Section 4.

2. Residuals of the Contaminated Series

Let us now define the residuals for ARMA(1,1) which is contaminated by an AO as;

$$e_t = \pi(B) y_t \tag{5.5}$$

where,

$$\pi(B) = \frac{\phi(B)}{\theta(B)} = 1 - \pi_1 B^1 - \pi_2 B^2 - \pi_3 B^3 - \cdots \tag{5.6}$$

Using (5.4) and (5.6) in (5.5), we get;

$$e_t = \frac{\phi(B)}{\theta(B)} \left\{ \frac{\theta(B)}{\phi(B)} a_t + \delta I_t^{(T)} \right\}$$

$$= a_t + \frac{\phi(B)}{\theta(B)} \delta I_t^{(T)}$$

$$e_t = \pi(B) \delta I_t^{(T)} + a_t \tag{5.7}$$

Thus, the residuals for $t < T$ and $t = T$ are as given below;

$$e_t = \begin{cases} a_t, & t < T \\ \delta + a_t, & t = T \end{cases} \tag{5.8}$$

By expanding (5.7), residuals for $t = T + j,\ j = 1, 2, 3, \ldots, n - T$ are obtained as follows;

$$e_t = \delta \pi(B) I_t^{(T)} + a_t$$

$$= \delta (1 - \pi_1 B^1 - \pi_2 B^2 - \cdots - \pi_j B^j - \cdots) I_t^{(T)} + a_t$$

$$= \delta (I_t^{(T)} - \pi_1 B^1 I_t^{(T)} - \pi_2 B^2 I_t^{(T)} - \cdots - \pi_j B^j I_t^{(T)} - \cdots) + a_t,$$

where,

$$B^j I_t^{(T)} = I_{t-j}^{(T)}$$

Consequently,

$$e_t = \delta (I_t^{(T)} - \pi_1 I_{t-1}^{(T)} - \pi_2 I_{t-2}^{(T)} - \cdots - \pi_j I_{t-j}^{(T)} - \cdots) + a_t$$

And, at $t = T + j$

$$e_{T+j} = \delta(I_{T+j}^{(T)} - \pi_1 I_{T+j-1}^{(T)} - \pi_2 I_{T+j-2}^{(T)} - \cdots - \pi_j I_{T+j-j}^{(T)} - \cdots) + a_{T+j}$$
$$= \delta(0 - \pi_1(0) - \pi_2(0) - \cdots - \pi_{j-1}(0) - \pi_j(1) - \pi_{j+1}(0) - \cdots) + a_{T+j}$$
$$= \delta(0 - 0 - 0 - \cdots - 0 - \pi_j - 0 - \cdots) + a_{T+j}$$
$$e_{T+j} = \delta(-\pi_j) + a_{T+j} \tag{5.9}$$

For 'n' number of observations, equations (5.7)-(5.9) can be summarized as follows;

$$e_t = \begin{cases} a_t, & t < T \\ \delta + a_t, & t = T \\ \delta(-\pi_j) + a_t, & t > T, i.e., t = T + j \ (j = 1, 2, 3, \ldots, n - T) \end{cases} \tag{5.10}$$

For an ARMA (1,1) model, we know that;

$$\pi_j = \theta^{j-1}(\phi - \theta), \forall \ j \geq 1 \tag{5.11}$$

From the above relation, it is clear that the value of π_j decreases at a 'slow' rate for $|\theta|$ very near to one and hence the number of residuals that are affected by the AO will be 'large' and the effect diminishes gradually, whereas when the value of θ is closer to zero, less number of residuals will be affected. Further, we notice that if ϕ and θ are equal, only the residual at the time of occurrence of AO is affected.

2.1 The case of contaminated AR(1) Model

When the value of the Moving Average parameter θ in the ARMA(1,1) model is equal to zero, the model turns out be AR(1);

$$\phi(B)x_t = a_t \tag{5.12}$$

where,

$$\phi(B) = 1 - \phi B$$

The AO contaminated AR(1) model is;

$$y_t = \frac{a_t}{\phi(B)} + \delta I_t^{(T)}$$

The residuals are given by;

$e_t = a_t + \delta\phi(B) I_t^{(T)}$

Residuals at time $t < T$ and $t \geq T$ are as follows;

$$e_t = \begin{cases} a_t , & t < T \\ \delta\phi(B) + a_t , & t \geq T \end{cases} \tag{5.13}$$

Note that;

$$e_t = \delta\phi(B) I_t^{(T)} + a_t$$

$$= \delta\{1 - \phi B\} I_t^{(T)} + a_t$$

$$= \delta\{I_t^{(T)} - \phi B I_t^{(T)}\} + a_t$$

$$e_t = \delta\{ I_t^{(T)} - \phi I_{t-1}^{(T)}\} + a_t \tag{5.14}$$

At, $t = T$, $I_T^{(T)} = 1$, therefore;

$$e_T = \delta\{1 - \phi(0)\} + a_T$$

$$e_T = \delta + a_T \tag{5.15}$$

Thus, the residuals for $t < T$ and $t = T$ are as given below;

$$e_t = \begin{cases} a_t , & t < T \\ \delta + a_t , & t = T \end{cases} \tag{5.16}$$

From equation (5.14) residuals for $t = T + j$ $(j = 1, 2, 3, \ldots, n - T)$ are obtained as;

$$e_{T+j} = \delta\{ I_{T+j}^{(T)} - \phi I_{T+j-1}^{(T)}\} + a_{T+j} \tag{5.17}$$

For $j = 1$ in equation (5.17), we get;

$$e_{T+1} = \delta\{I_{T+1}^{(T)} - \phi I_{T+1-1}^{(T)}\} + a_{T+1}$$

$$= \delta\{0 - \phi(1)\} + a_{T+1}$$

$$e_{T+1} = \delta(-\phi) + a_{T+1} \tag{5.18}$$

Similarly, for $j = 2$, we get;

$$e_{T+2} = \delta\{0 - \phi(0)\} + a_{T+2}$$

$$e_{T+2} = a_{T+2}$$

Hence, for $j = 3, \ldots, n - T$, we get;

$$e_{T+j} = a_{T+j} \tag{5.19}$$

For 'n' number of observations, equations (5.16)-(5.19) can be summarized as follows;

$$e_t = \begin{cases} a_t\,, & t < T \\ \delta + a_t\,, & t = T \\ \delta(-\phi) + a_t & t = T+1 \\ a_t & t > T+j, (j = 2,3\ldots, n-T) \end{cases} \quad (5.20)$$

From the above equation, it is clear that in an AR(1) model, the AO affects only the residuals at $t = T$ and $t = T+1$.

2.2 The case of contaminated MA(1) Model

When the value of the Autoregressive parameter ϕ in the ARMA(1,1) model is equal to zero, the model turns out be a MA(1) model. The model is;

$$x_t = \theta(B)a_t \quad (5.21)$$

where,

$$\theta(B) = 1 - \theta B$$

Furthermore, by proceeding on lines similar as we did for AO contaminated ARMA(1,1) model, we can get the AO contaminated MA(1) time series $\{y_t\}$ and the corresponding residuals $\{e_t\}$ as follows;

$$y_t = \theta(B)a_t + \delta I_t^{(T)}\ , \quad (5.22)$$

$$e_t = a_t + \frac{\delta}{\theta(B)} I_t^{(T)}. \quad (5.23)$$

Since, $\theta(B) = 1 - \theta B$, and for an invertible MA process;

$$\frac{1}{\theta(B)} = \frac{1}{(1-\theta B)} = 1 + \theta^1 B + \theta^2 B^2 + \theta^3 B^3 + \cdots + \theta^j B^j + \cdots. \quad (5.24)$$

Using equation (5.24) in (5.23), we get;

$$e_t = a_t + \delta\{1 + \theta^1 B + \theta^2 B^2 + \theta^3 B^3 + \cdots + \theta^j B^j + \cdots\} I_t^{(T)}$$

$$= a_t + \delta\{I_t^{(T)} + \theta^1 B I_t^{(T)} + \theta^2 B^2 I_t^{(T)} + \theta^3 B^3 I_t^{(T)} + \cdots + \theta^j B^j I_t^{(T)} + \cdots\}$$

$$= a_t + \delta\{I_t^{(T)} + \theta^1 I_{t-1}^{(T)} + \theta^2 I_{t-2}^{(T)} + \theta^3 I_{t-3}^{(T)} + \cdots + \theta^{j-1} I_{t-(j-1)}^{(T)} + \theta^j I_{t-j}^{(T)} + \theta^{j+1} I_{t-(j+1)}^{(T)} + \cdots\}$$

Residuals at time $t<T$ and $t=T$ can be got from the above expression and it is;

$$e_t = \begin{cases} a_t, & t<T \\ \delta + a_t & t=T \end{cases} \tag{5.25}$$

Now for residuals at time $t>T$, i.e., for $t=T+j,(j=1,2,\ldots,n-T)$

$$e_{T+j} = a_{t+j} + \delta\{I_{T+j}^{(T)} + \theta^1 I_{T+j-1}^{(T)} + \theta^2 I_{T+j-2}^{(T)} + \theta^3 I_{T+j-3}^{(T)} + \cdots + \theta^{j-1} I_{T+j-(j-1)}^{(T)} + \theta^j I_{T+j-j}^{(T)} + \theta^{j+1} I_{T+j-(j+1)}^{(T)} + \cdots\}$$

$$= a_{T+j} + \delta\{I_{T+j}^{(T)} + \theta^1 I_{T+j-1}^{(T)} + \theta^2 I_{T+j-2}^{(T)} + \theta^3 I_{T+j-3}^{(T)} + \cdots + \theta^{j-1} I_{T+1}^{(T)} + \theta^j I_T^{(T)} + \theta^{j+1} I_{T-1}^{(T)} + \cdots\}$$

$$= a_{T+j} + \delta\{0 + \theta^1(0) + \theta^2 I(0) + \theta^3 I(0) + \cdots + \theta^{j-1}(0) + \theta^j(1) + \theta^{j+1}(0) + \cdots\}$$

$$= a_{T+j} + \delta\{0+0+0+0+\cdots+0+\theta^j+0+\cdots\}$$

$$e_{T+j} = \delta\theta^j + a_{T+j} \tag{5.26}$$

Given ‘n’ observations, equations (5.25) and (5.26) can be summarized as follows;

$$e_t = \begin{cases} a_t & t<T \\ \delta + a_t & t=T \\ \delta(\theta)^j + a_t & t>T, i.e., t=T+j(j=1,2,\ldots,n-T) \end{cases} \tag{5.27}$$

From the above equation, it is clear that in a MA(1) model, the AO affects all the subsequent residuals from the time of its occurrence.

3. Illustration

The results of the Simulation study on the contaminated models discussed above are presented in this section by focusing on the effect of an AO on the observations and residuals.

3.1 Effect of an AO on observations and residuals

A Simulation study was carried out with different values of respective parameters in the three models which are considered in the present study along with three different magnitudes of the AO. Firstly, Outlier-free Time Series of $n=100$ observations are simulated with different values of the parameters (ϕ and θ).

These observations are plotted. Next, to illustrate the effects of an AO on the observations, δ of magnitudes 5, 10, and 15 are applied to create an artificial AO at $T = 50$. To examine the effects of AO on the residuals, we assume that the Time Series parameters are known and the series is observed from $t = 0$ to $t = n$.

3.1.1 The contaminated ARMA (1, 1) Model

The effect of AO with three different magnitudes and with different values of the parameters on the observations and the errors in the ARMA(1,1) model were investigated through a Simulation study and are plotted (for a clearer picture only the middle 20 i.e. from $t = 40$ to $t = 60$ observations and residuals are plotted) in the below figure 1.

3.1.2 The contaminated AR(1) Model

The effect of AO with three different magnitudes and with different values of the parameter ϕ on the observations and the errors in the AR(1) model were observed and are plotted in the figure 2.

3.1.3 The contaminated MA(1) Model

The effect of AO with three different magnitudes and with different values of the parameter θ on the observations and the errors in the MA(1) model were observed and are plotted in the figure 3.

4. Conclusion

Outliers are commonly encountered in Time Series data analysis. The presence of such extraordinary events could easily mislead the conventional Time Series analysis procedures resulting in erroneous conclusions. The impact of those events is often overlooked. So, here in the present work, we have tried to study the effect of an AO on the residuals of the underlying model in relation to the structural coefficients values. A cursory glance on the figures 1, 2, and 3 shows that, in the presence of an AO, while the observations appear to be similar; the residuals exhibit different behaviors across the values of the parameters.

From the limited Simulation studies on an ARMA(1,1) model, it is observed that, given the occurrence of an AO at $t = T$, the AO is seen to have an effect not only on the residual at $t = T$ but also on subsequent ones at $t = T + j, j \geq 1$.

However, the number of residuals noticeably affected depends on the parameters ϕ and θ and the magnitude of the AO δ. This observation has not been noticed in Zaharim et al. (2009). The Moving Average parameter θ plays a dominant role and values of $|\theta|$ nearer to one result in large number of residuals getting affected after the incidence of AO, which in turn affects the inference for such Time Series. Narrowing down to AR(1) model, the AO affects only at the time of occurrence i.e., at $t = T$ and the next subsequent residual $t = T + 1$, irrespective of the value of the parameter ϕ. This can also be observed from the equation (5.12) obtained when the value of $\theta = 0$, whereas, in the case of MA(1), all the residuals at the time of occurrence ($t = T$) and subsequent to the occurrence($t = T + j, j \geq 1$) of AO get affected. The extent of the effect depends on the value of θ and, the magnitude of the AO.

Holistically, figures 1, 2, and 3 reveal the following for the models under consideration.

4.1 ARMA(1,1)

- When both the AR and MA parameter values are almost equal (exactly) i.e., when $\theta \approx \phi$, the Additive Outlier has an effect on the residuals in the contaminated series only at the time of its incidence. (Refer to figure 1(a)-1st row, figure 1(b)-5th row, figure 1(c)-4th row).
- When $\theta \neq \phi$ and $|\theta| \approx 1$, the magnitude of a large number of residuals in the contaminated series from the time of occurrence of the Additive Outlier are noticeably more than that of the original series with positive sign for all the residuals when $\theta > 0$ (Refer to figure 1(a)-4th row, figure 1(b)-2nd and 6th rows) and with alternating signs when $\theta < 0$ (Refer to figure 1(a)-4th row, figure 1(b)-2nd row and figure 1(c)-1st row).
- When $\theta \neq \phi$ and $|\theta| \approx 0$, the magnitude of only a few number of residuals in the contaminated series from the time of occurrence of the Additive Outlier are slightly more than that of the original series with positive sign for all the residuals when $\theta > 0$ (Refer to figure 1(a)-3rd row, figure 1(b)-1st row and figure 1(c)-3rd and 5th rows) and with alternating signs when $\theta < 0$. (Refer to figure 1(a)-2nd and 6th rows, figure 1(b)-4th row and figure 1(c)-2nd).

4.2 AR(1)

- When $\phi \approx 1$, the residuals of the contaminated series at $t = T$ and $t = T+1$ are of different sign, as $\phi \downarrow 0$, the magnitude of the residuals at $t = T+1$ diminishes. Further, when $\phi \approx -1$, the residuals of the contaminated series at $t = T$ and $t = T+1$ are of same sign and with approximately the same magnitude, as $\phi \uparrow 0$, the magnitude of the residuals at $t = T+1$ diminishes. (Refer to figure 2).

4.3 MA(1)

- The effect of the AO on the residuals in this model for various values of parameter θ, is similar to what has been observed in the case of the ARMA (1,1) model. (Refer to figure 3).

These finer aspects of influence of an AO in ARMA(1,1), appears to be a refinement over the findings of Zaharim et al. (2009).

Acknowledgement
The authors sincerely thank the referees for their valuable comments, which have led to this version.

References

1. Abraham, B. and Chuang, A. (1989). Outlier detection and time series modeling. *Technometrics*, **31(2)**, 241-248.
2. Barnett, V. and Lewis, T. (1984). Outliers in Statistical Data, Second Edition, Wiley.
3. Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994). Time Series Analysis - Forecasting and Control, Third Edition, Pearson Education.
4. Chen, C. and Liu, L. M. (1993a). Forecasting time series with outliers. *Journal of Forecasting*, **12**, 13-35.
5. Chen, C. and Liu, L. M. (1993b). Joint estimation of model parameters and outlier effects in time series. *Journal of American Statistical Association*, **88**, 284-297.
6. Fox, A. J. (1972). Outliers in time series. *Journal of Royal Statistical Society (Series B)*, **34(3)**, 350-363.
7. Ledolter, L. (1987). The effect of additive outliers on the forecasts from ARIMA models. Technical Report No. 136, Department of Statistics and Actuarial Science, University of Iowa.

8. Tsay, R.S. (1986). Time series model specification in the presence of outliers. *Journal of American Statistical Association*, **81**, 119-122.
9. Tsay, R.S. (1988). Outliers, level shifts, and variance changes in time series. *Journal of Forecasting*, **7**, 1-20.
10. Wei, W.S. (2006). Time series analysis: Univariate and multivariate methods, Boston: Pearson Addison Wesely, 223-229
11. Zaharim, A., Rajali, R., Atok, R. M., Ibrahim, K. and Razali, A. M. (2009). A study on the nature of an additive outlier in ARMA(1,1) models. *European Journal of Scientific Research*, **32(3)**, 362-368.

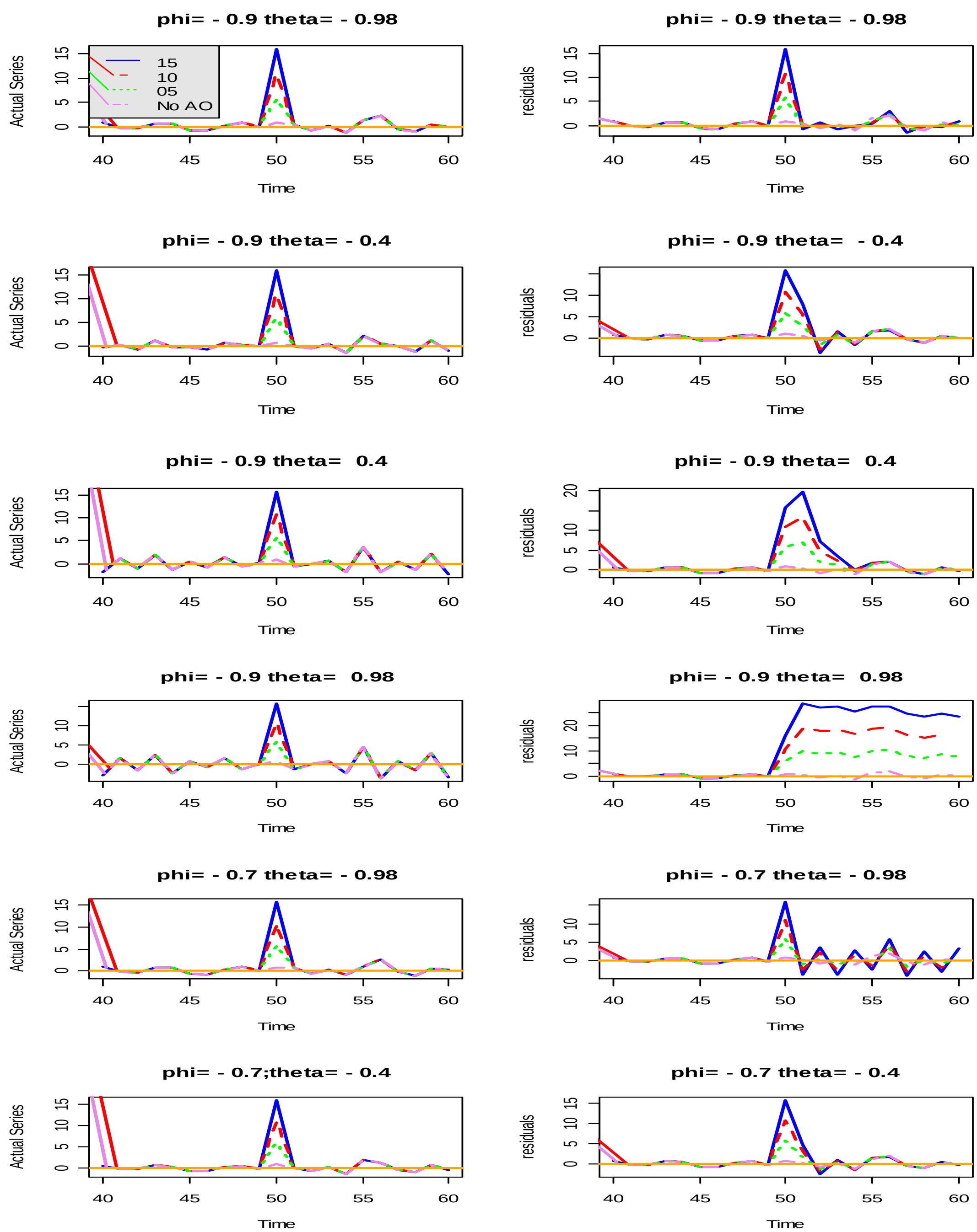

Figure 1(a): Time Series and Residual Plots of ARMA(1,1) Model

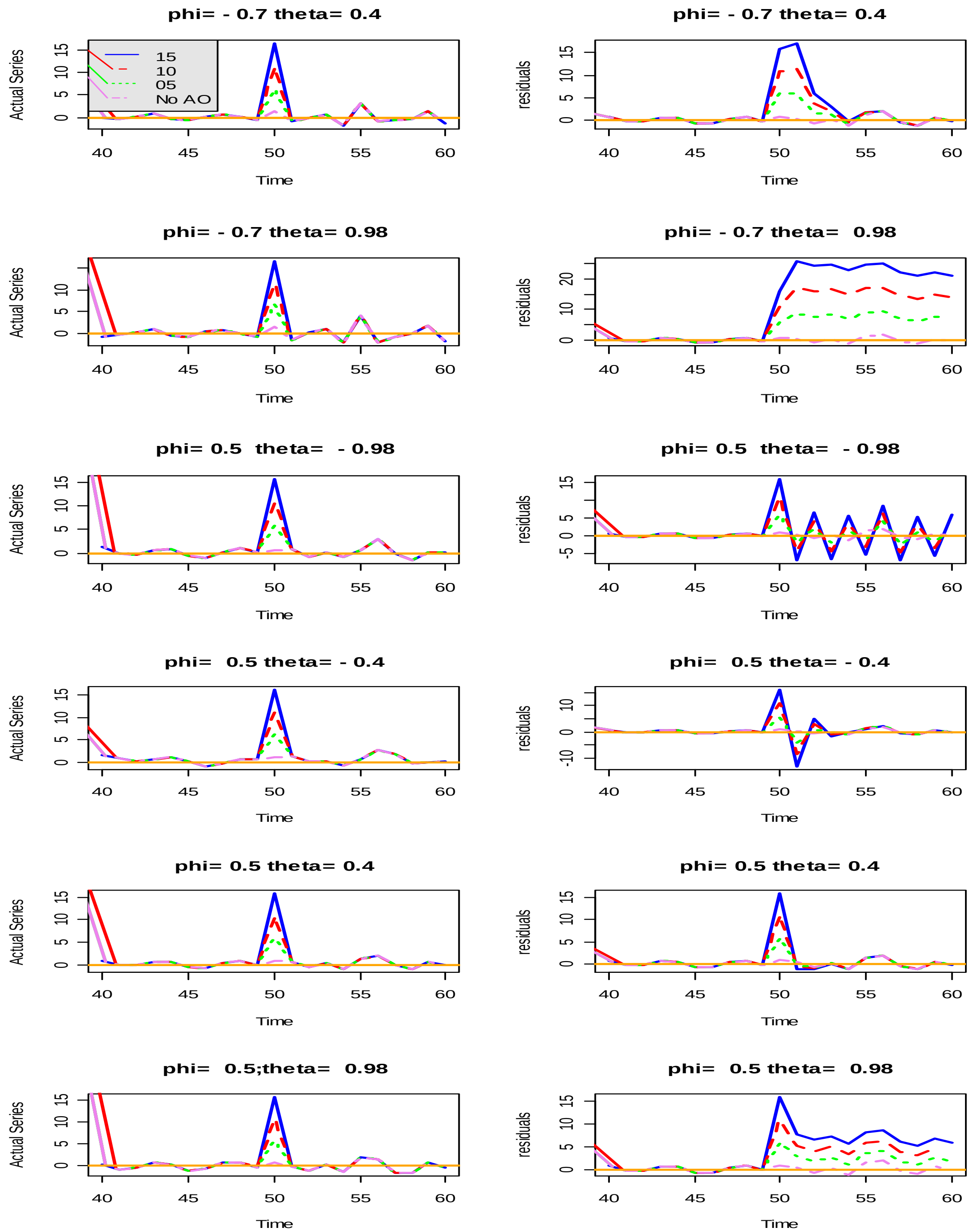

Figure 1(b): Time Series and Residual Plots of ARMA(1,1) Model

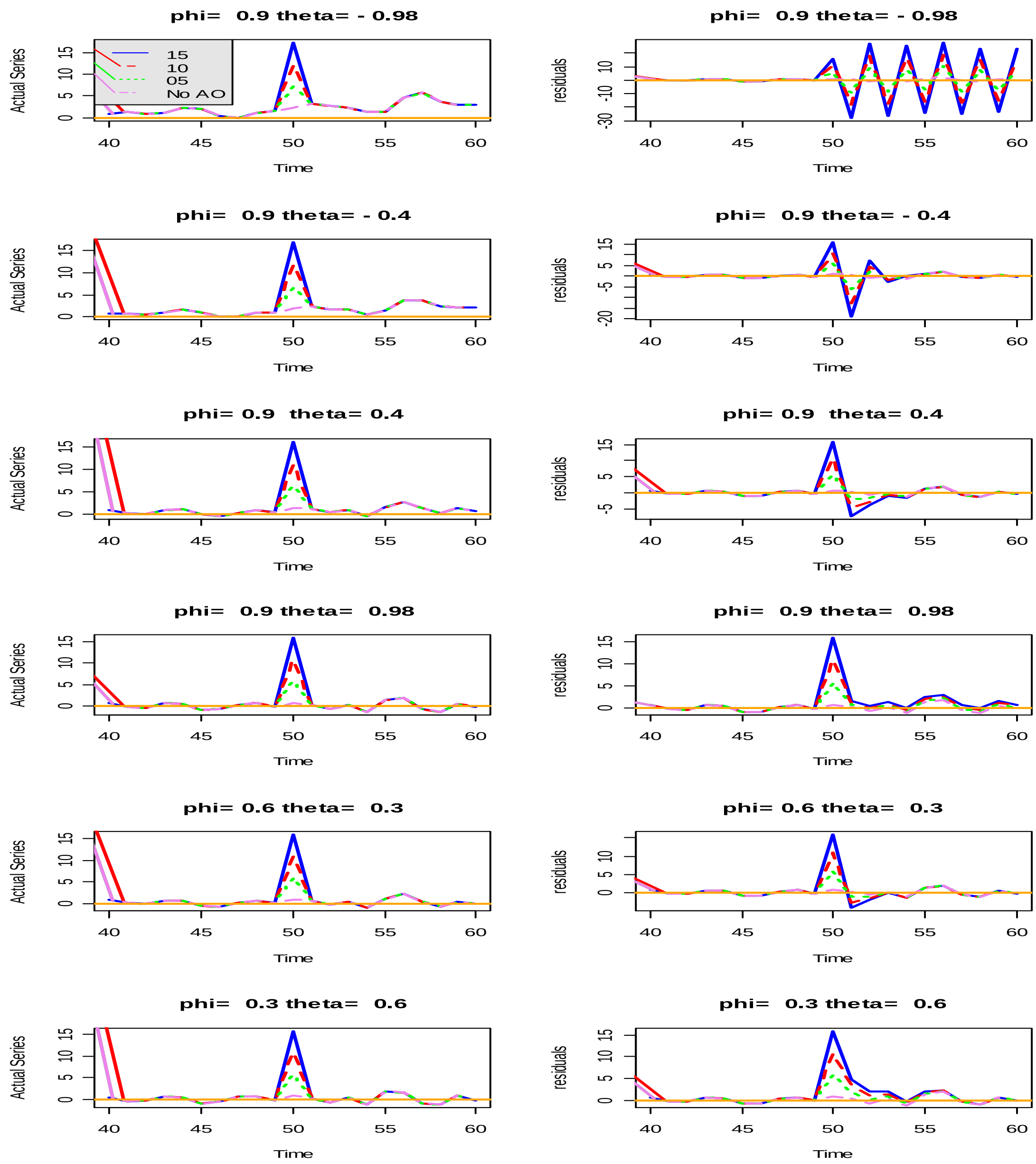

Figure 1 (c): Time Series and Residual Plots of ARMA(1,1) Model

Note: In figures 1(a, b, and c), the first column is the plot of actual Time Series, and the second column is the corresponding Residual Plots for various values of the parameters: Autoregressive parameter ϕ (phi) and Moving Average parameter θ (theta) of ARMA (1, 1) Model and for different magnitudes (5, 10, and 15) of the Additive Outlier.

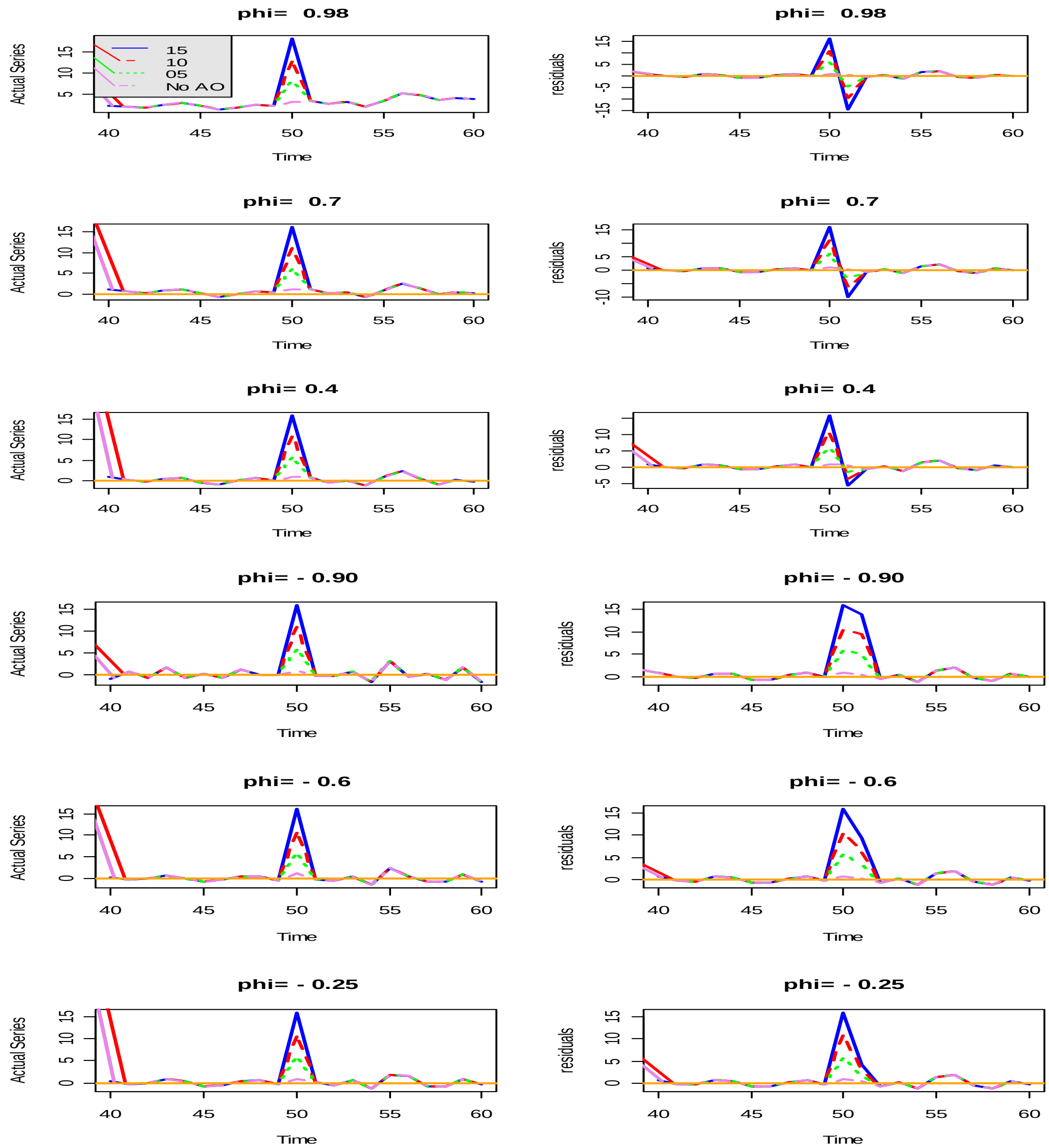

Figure 2: Time Series and Residual Plots of AR(1) Model

Note: In figure 2, the first column is the plot of actual Time Series, and the second column is the corresponding Residual Plots for various values of the Autoregressive parameter ϕ (phi) of AR (1) Model and for different magnitudes (5, 10, and 15) of the Additive Outlier.

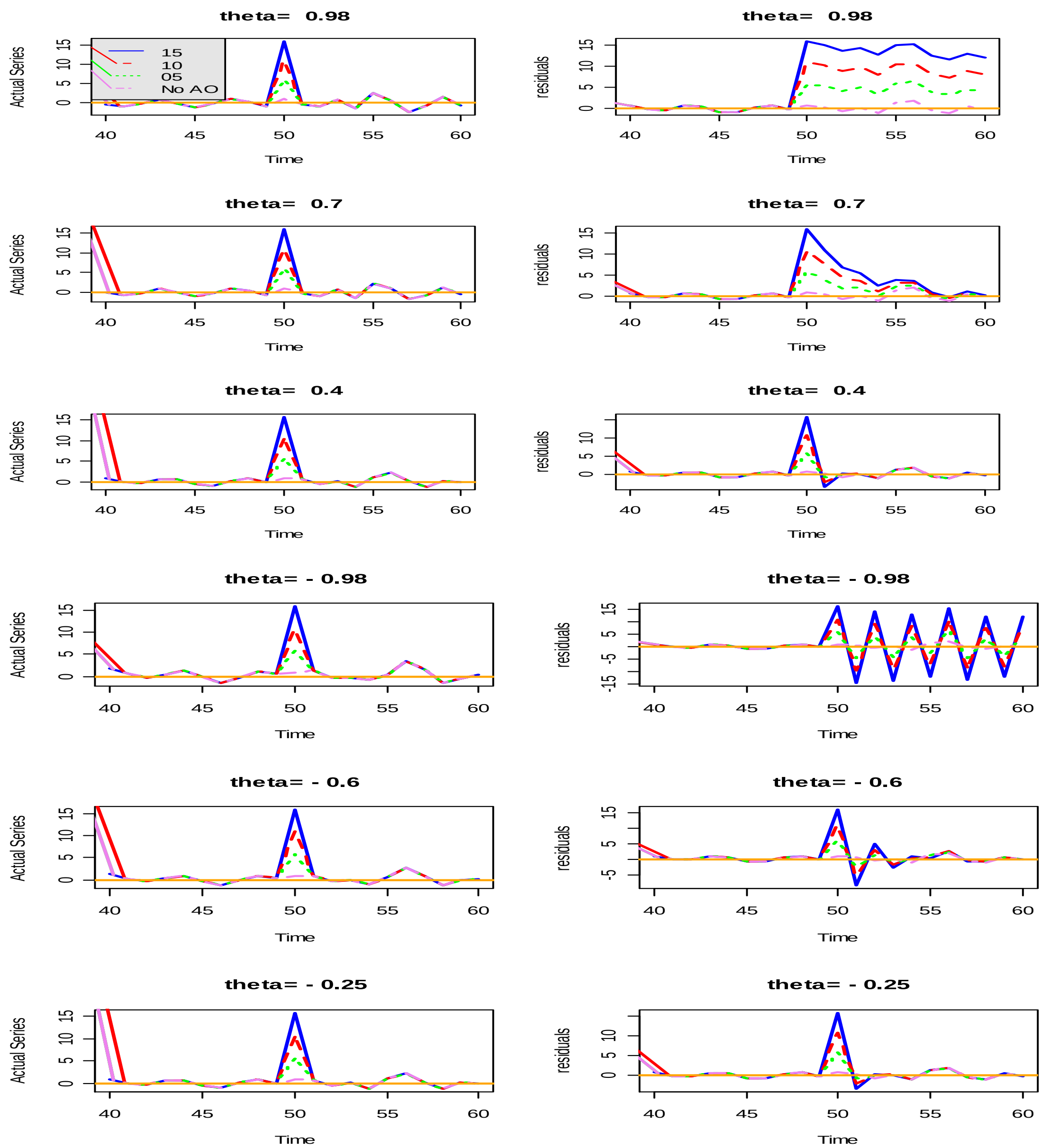

Figure 3: Time Series and Residual Plots of MA(1) Model

Note: In figure 3, the first column is the plot of actual Time Series, and the second column is the corresponding Residual Plots for various values of the Moving Average parameter θ (theta) of MA (1) Model and for different magnitudes (5, 10, and 15) of the Additive Outlier.