VILNIUS UNIVERSITY
FACULTY OF MATHEMATICS AND INFORMATICS
MODELLING AND DATA ANALYSIS
MASTER'S STUDY PROGRAMME

Outlier Search in ARMA Models

Master's thesis

Author: Abdellah Djaadi
VU email address: abdellah.djaadi@mif.stud.vu.lt
Supervisor: (Prof., Habil. dr. Vilijandas Bagdonavicius)

Vilnius

2022

Abstract

The purpose of this research is to study the problem of outliers in ARMA models and compare two outlier search techniques: Chang and BP methods. The aim of first method is based on an iterative procedure for parameters estimations then locating the additive and innovative outlier then distinguishing their types. The objective of second method is to generalize the BP outlier identification method for time series (ARMA). Our analysis shows that in most cases, the iterative procedure proposed by Chang is better than BP in detecting outliers. The latter method can not identify the type of declared outliers. However, there is some situation with adjacent outliers BP method slightly outperforms the existing techniques.

Keywords: ARMA; additive outlier; innovative outlier; BP; Chang.

Contents

List of Figures

List of Tables

Notation

- AO denotes additive outlier.
- IO denotes innovative outlier.
- LS denotes level shift.
- TC denotes time change.
- ω denotes outlier size.
- T denotes the moment that an outlier is intoduced.
- tsa denotes the method used in R package *TSA*
- tsout denotes the method used in R package *tsoutliers*
- robarim denotes the method used in R package *robustarima*
- i.i.d. means *independent identically distributed random variables.*

1 Introduction

In the last few years, the study of outliers received the attention of many authors and is now one of the main tasks in data analysis. An outlier is a data point that differs significantly from other observations. Outliers can arise from several different mechanisms and causes; the two major causes are those arising from errors in the data, and those arising from the inherent variability of the data. Moreover, outliers are often referred to as anomalies, discordant observations, discords, exceptions, aberrations, surprises or contaminants.
Nowadays, recent advances in technology allow us to collect a large amount of data over time. This leads the problem of multiple outliers to become a challenge for many analysts. Therefore, Outliers Identification become an interesting field for many researchers. It has various applications on time series data and is widely utilized in diverse research areas such as fraud detection [19], fault detection [18], medicine [9] and many more.

In the first study on this topic which was conducted by Fox (1972)[10], who defined two characterizations of outliers; *Type I* or *Additive Outlier* which affects a single observation; and *Type II* or *Innovative Outlier* which corresponds to the situation in which a single "innovation" is extreme. This will affect not only the particular observation but also subsequent observations. Also, Fox (1972) proposed the use of maximum likelihood ratio tests to detect them. In 1988, Chang et al. [6] extend the results of Fox (1972) to *ARIMA* models and present an iterative procedure for outlier identification and parameter estimation which were generalized by Tsay (1988) [17] for detecting level shifts and temporary changes which are other types of outliers. Chen and Liu (1993) [7] proposed an outlier detection and parameter estimation procedure that is widely used.

In this master thesis, our main goal was to compare the performance of some existing outliers identification methods in detecting the *AO* and *IO* in *ARMA* models with different outliers size. We considered two methods, the existing outlier identification method proposed by Chang et al. (1988)[6] and the generalization of the Bagdonavicius-Petkevicus (BP) method for outlier search in linear regression models (see [1]) to time series. For the practical usage of proposed procedures; the R packages *TSA*, *tsoutlier*s and *robustarima* and an original R script were used.

The procedures given in packages are based on iterative algorithms for detecting the location of outliers and then identifying the type of outliers (AO or IO) according to statistical tests. Therefore, we generate a sample 300 observations using the *AR(1)*, *MA(1)* and *ARMA(1,1)* models with $\phi = 0.6, \theta = 0.4$ the coefficients of *AR* and *MA* respectively; then in the first step, multiple additive outliers with different effects were introduced to data at different times and then the performance of the mentioned methods were compared based on masking and swamping values which are investigated in several simulations. We repeated the same steps with innovative outliers
The rest of this paper is organized as follows. Section 2 presents the literature review. In section 3, we illustrate the notions of additive and innovative outliers in the ARMA models. Next, we describe the the considered outlier detection methods. In section 5, we study the performance of the existing

identifications methods in detecting AO and IO outliers. Finally, the conclusion and some remarks are summarized in Section 6.

2 Literature Review

The study of outliers in ARIMA models has been a popular subject of research. Fox (1972)[10] describes additive and innovative outliers and advocates for their detection using maximum likelihood ratio tests. Chang and Tiao (1983) and Chang et al. (1988) generalize Fox's (1972) findings to ARIMA models and provide an iterative technique for outlier identification and parameter estimation. Tsay (1988)[17] expands on this technique for identifying level shifts and temporary changes.

However, these proposed outlier detection methods suffer from three significant problems. (b) Masking, the usual procedures based on the detecting outliers one by one, may also fail in the identification of outliers when they have similar effects. (a) When there is a level shift in the series, most of the techniques fail to distinguish between the level-shifts and innovative outliers.

Thus, many authors such as Balke (1993)[2] presents a new technique for solving the confusing problem between level-shift and innovative outliers. Chen and Liu (1993)[7] describe an outlier identification and parameter estimation approach that improves upon prior approaches and seems to be widely used. However, this technique has the potential to misinterpret level shift as innovative outliers, and certain outliers may be missed entirely because of masking effects. Outliers are not always significant observations, and Pena (1990, 1991)([15],[14]) gives statistics for determining the outliers' effect on model parameters.

These techniques we mentioned have one common factor; they all deal with outliers while estimating parameters of ARMA models. They begin by estimating the model parameters using maximum likelihood and then analyze the residuals using an iterative diagnostic technique to locate outliers.

3 Definition of AO and IO in ARMA models and Examples.

3.1 Models and Notations

Suppose that a stationary time series Y_t is outliers-free series, $t = 0, \pm 1, \pm 2, \ldots$, is represented by an ARMA process as follow:

$$\begin{aligned} Y_t &= a_1 Y_{t-1} + a_2 Y_{t-2} + ... + a_p Y_{t-p} + \varepsilon_t + b_1 \varepsilon_{t-1} + ... + b_q \varepsilon_{t-q} \\ &= \sum_{i=1}^{p} a_i Y_{t-i} + \sum_{j=1}^{q} b_j \varepsilon_{t-j} + \varepsilon_t, \quad t = 1, ..., n, \end{aligned} \tag{1}$$

The model in (1) can be written more concisely using the back-shift operator as it is described in Robert, S. et D book [16] as follows:

$$\phi(B) Y_t = \theta(B) \varepsilon_t \tag{2}$$

Where

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p, \quad \theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q. \tag{3}$$

Here $\phi(B)$ and $\theta(B)$ are polynomials in B of degrees p and q, respectively, B is the back-shift operator such that $B^k Y_t = Y_{t-k}, E(Y_t) = 0$, and $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variables of mean zero and variance σ^2. The process $\{Y_t\}$ is stationary and invertible because the root of the polynomial $\phi(B)$ are outside the unit circle (See prove [13]). Thus, model 2 can be defined as follow:

$$\pi(B)Y_t = \varepsilon_t \tag{4}$$

with

$$\pi(B) = \frac{\phi(B)}{\theta(B)} = 1 - \pi_1 B - \pi_2 B^2 - \ldots \tag{5}$$

When Y_t has outliers, it becomes a contaminated series, which contains either additive or innovational outliers. We note it as Z_t.

It is said that an AO is observed at the moment $t = T$ if:

$$Z_t = \begin{cases} Y_t, & t \neq T \\ Y_t + \omega, & t = T \end{cases}$$

Here ω defines outlier size and it is a constant or a random variable.

Moreover, the presence of the *additive outlier* (AO) does not influence values of Y_t, when $t \neq$T. So shortly Z_t can be written in the form

$$Z_t = Y_t + \omega \mathbf{1}_{\{t=T\}}. \tag{6}$$

where $\mathbf{1}_{\{t=T\}}$ is an indicator function that takes the value 1 if $t = T$, and 0 otherwise.

In the other hand, the presence of the *innovative outlier* (IO) does not effect only Y_T but also a sequence of observations $Y_{T+1}, Y_{T+2}, \ldots$. Thus, the contaminated series Z_t is given in the following manner:

$$\begin{aligned} Z_t &= \omega \pi^{-1}(B) \mathbf{1}_{\{t=T\}} + Y_t \\ &= \omega k_{t-T} \mathbf{1}_{\{t=T\}} + Y_t \end{aligned} \tag{7}$$

where k_{t-T} are the coefficients of $\pi^{-1}(B)$ which is defined previously in (5). Denoting $k_0 = 1$, we have

$$Z_t = \omega k_{t-T} \mathbf{1}_{\{t \geq T\}} + Y_t.$$

Note that

$$\pi^{-1}(B)\mathbf{1}_{\{t=T\}} = \sum_{j=0}^{\infty} k_j B^j \mathbf{1}_{\{t=T\}} = \sum_{j=0}^{\infty} k_j \mathbf{1}_{\{t-j=T\}}.$$

For $t < T$ all indicators $1_{\{t-j=T\}}$ are equal to zero. For $t \geq T$ only one indicator with $j = t - T$ is equal to one. So

$$\pi^{-1}(B)\mathbf{1}_{\{t=T\}} = k_{t-T}\mathbf{1}\{t \geq T\}.$$

Thus, the equation (7) is written as follows:

$$Y_t = \omega k_{t-T}\mathbf{1}_{\{t \geq T\}} + Y_t. \tag{8}$$

In the case of m outliers are presented in the series:
For AO:

$$Z_t = \sum_{i=1}^{m} \omega_i 1_{\{t=T_i\}} + Y_t,$$

so

$$Z_t = \omega_i + Y_t, \quad t = T_i, \quad i = 1, ..., m, \quad Z_t = Y_t, \quad t \neq T_1, ..., T_m.$$

For IO

$$Z_t = \sum_{i=1}^{m} \omega_i k_{t-T_i} 1_{\{t \geq T_i\}} + Y_t.$$

The last formula implies that for $t < T_{(1)}$

$$Z_t = Y_t.$$

For $T_{(1)} \leq t < T_{(2)}$

$$Z_t = \omega_{(1)} k_{t-T_{(1)}} + Y_t.$$

$T_{(j-1)} \leq t < T_{(j)}$, $j = 2, ..., m$,

$$Z_t = \sum_{i=1}^{j-1} \omega_{(i)} k_{t-T_{(i)}} + Y_t.$$

For $t \geq T_{(m)}$

$$Z_t = \sum_{i=1}^{m} \omega_{(i)} k_{t-T_{(i)}} + Y_t.$$

Furthermore, let us suppose that in our series m_1 AO, and m_2 IO are presented; where m_1, and $m_2 \in \mathbb{N}$. Then

$$Z_t = \sum_{i=1}^{m_1} \omega_A 1_{\{t=T_{iA}\}} + \sum_{i=1}^{m_3} \omega_I k_{t-T_{iI}} 1_{\{t \geq T_{iI}\}} + Y_t$$

3.2 Examples

1) In the case of MA(1) process $Y_t = \varepsilon_t - b_1\varepsilon_{t-1}$, $|b_1| < 1$, we have

$$\phi(B) = 1, \quad \theta(B) = 1 - b_1 B, \quad \theta^{-1}(B) = \sum_{j=0}^{\infty} b^j B^j,$$

$$\pi(B) = \theta^{-1}(B)\phi(B) = \theta^{-1}(B) = \sum_{j=0}^{\infty} b^j B^j.$$

$$\varepsilon_t = \sum_{j=0}^{\infty} b^j y_{t-j} = \sum_{j=0}^{t-1} b^j y_{t-j}.$$

For AO:

$$Z_t = \omega \mathbf{1}_{\{t=T\}} + Y_t.$$

For IO: Note that $1/\pi(B) = 1/\theta^{-1}(B) = \theta(B) = 1 - b_1 B$, so $k_1 = -b_1$, $k_j = 0$, $j > 1$,

$$Z_t = \omega + y_t, \quad t = T, \quad Z_t = -\omega b_1, \quad t = T+1, \quad Z_t = Y_t, \quad t \neq T, T+1.$$

Shortly:

$$Z_t = \omega(\mathbf{1}_{\{t=T\}} - b_1 \mathbf{1}_{\{t=T+1\}}) + Y_t.$$

2) In the case of AR(1) process $Y_t = a_1 Y_{t-1} + \varepsilon_t$ we have

$$\phi(B) = 1 - a_1 B, \quad \theta(B) = 1, \quad \theta^{-1}(B) = 1,$$

$$\pi(B) = \theta^{-1}(B)\phi(B) = \phi(B) = 1 - a_1 B.$$

$$\varepsilon_t = \phi(B) y_t = Y_t - a_1 Y_{t-1}.$$

For AO:

$$Z_t = \omega \mathbf{1}_{\{t=T\}} + Y_t.$$

For IO: Note that $1/\pi(B) = 1/(1 - a_1 B) = \sum_{i=0}^{\infty} a_1^i B^i$, so $k_i = a_1^i$,

$$Z_t = \omega + y_t, \quad t = T, \quad Z_t = \omega a_1^{t-T}, \quad t = T+1, \quad Z_t = Y_t, \quad t \neq T, T+1.$$

Shortly:

$$Z_t = \omega k_{t-T} \mathbf{1}_{\{t \geq T\}} + Y_t = \omega a_1^{t-T} \mathbf{1}_{\{t \geq T\}} + Y_t.$$

So presence of innovative outlier at point $t = T$ influences later observations in exponentially decreasing manner.

So presence of innovative outlier at point $t = T$ influences only the next observation.

3) In the case of ARMA(1,1) process we have

$$\phi(B) = 1 - a_1 B, \quad \theta(B) = 1 - b_1 B, \quad \theta^{-1}(B) = \sum_{i=0}^{\infty} b_1^i B^i,$$

we have,

$$\pi(B) = \theta^{-1}(B)\phi(B) = \sum_{i=0}^{\infty} b_1^i B^i (1 - a_1 B)$$
$$= \sum_{i=0}^{\infty} b_1^i B^i - a_1 \sum_{i=0}^{\infty} b_1^i B^{i+1}$$
$$= \sum_{i=0}^{\infty} b_1^i B^i - a_1 \sum_{i=1}^{\infty} b_1^{i-1} B^i 1$$
$$= 1 + \sum_{i=1}^{\infty} (b_1 - a_1) b_1^{i-1} B^i.$$

Similary

$$\pi^{-1}(B) = \phi^{-1}(B)\theta(B) = 1 + (a_1 - b_1) \sum_{i=1}^{\infty} a_1^{i-1} B^i.$$

For AO:

$$Z_t = \omega \mathbf{1}_{\{t=T\}} + Y_t.$$

For IO: Note that

$$k_{t-T} \mathbf{1}_{\{t \geq T\}} = 1_{\{t=T\}} + (a_1 - b_1) a_1^{t-T-1} \mathbf{1}_{\{t>T\}}.$$

It implies:

$$Z_t = \omega(\mathbf{1}_{\{t=T\}} + (a_1 - b_1) a_1^{t-T-1} \mathbf{1}_{\{t>T\}}) + Y_t.$$

So presence of innovative outlier at point $t = T$ influences later observations in exponentially decreasing manner.

4 Outlier detection techniques

4.1 Theoretic background for the Chang identification method

The *Chang* method is an interative procedure for outlier detection based on the one developed by Chang et al. (1988) [6]. This approach is mainly based on estimating the model parameters using maximum likelihood and then analysing the residuals with an iterative procedure to detect outliers. The iterative procedure has two stages. In the first stage, initial parameter estimation, a robust initial estimates of model parameters are computed. In the second stage, outlier detection, outliers are identified one by one using the likelihood ratio test.

4.1.1 ARMA Parameters and σ^2 Are Known

Supposing that the time series parameters ϕ_i's, and θ_j's are known. It is necessary first to estimate the effect parameters ω of an IO, and AO in equations (7), and(6) respectively.
For $t = 1, \ldots, n$, we define:

$$U_t = \pi(B) Z_t \tag{9}$$

Note that for AO:

$$U_t = \omega_A \pi(B)\mathbf{1}_{\{t=T\}} + \pi(B)Y_t = \omega_A \pi(B)\mathbf{1}_{\{t=T\}} + \varepsilon_t,$$

For IO

$$U_t = \omega_I \pi(B) k_{t-T}\mathbf{1}_{\{t\geq T\}} + \varepsilon_t = \omega \pi(B)\pi^{-1}(B)\mathbf{1}_{\{t=T\}} + \varepsilon_t = \omega_I \mathbf{1}_{\{t=T\}} + \varepsilon_t.$$

Set $i = I, A$, and $x_t = \pi(B)\mathbf{1}_{\{t=T\}}$ for AO, and $x_t = \mathbf{1}_{\{t=T\}}$ for IO.
Then U_t can be written in unified way:

$$U_t = \omega_i x_t + \varepsilon_t. \tag{10}$$

The effect ω_I of an IO at time T is estimated by the residual U_T at that specific point; where

$$\hat{\omega}_I = U_T. \tag{11}$$

However, the information about AO'impact is scattered over a string of "residuals". Thus, the best estimate for the ω_{AO} is a linear combination of $U_T, U_{T+1}, \cdots ,$. It can be computed using the technique of least squeares (see [7]). Thus, the estimate $\hat{\omega}_A$ is giving by:

$$\hat{\omega}_A = \rho^2 \pi(F) U_T. \tag{12}$$

$$= \rho^2 \pi(F)\pi(B) Z_t$$

where $\rho^2 = \left(1 - \pi_1^2 + \cdots - \pi_{n-T}^2\right)^{-1}$; F is the forward-shift operator such that $FU_t = U_{T+1}$ and $\pi(F) = 1 - \pi_1 F - \pi_2 F^2 - \cdots - \pi_{n-T} F^{n-T}$
The variances of these estimators are:

$$\operatorname{Var}(\hat{\omega}_I) = \sigma^2, \quad \operatorname{Var}(\hat{\omega}_A) = \rho^2\sigma^2. \tag{13}$$

In order to test for a single outlier, one may test the following hypotheses (i) $H_0 : \omega_I = \omega_A = 0$; (ii) $H_I : \omega_I \neq 0$, (iii) $H_A : \omega_A \neq 0$, and the likelihood ratio test statistics for testing H_0 vs. H_I, and H_A are respectively $\lambda_{i,T} = \widehat{\omega}_i/\sigma_i$ for $i = I, A$; where σ_i is the square root of the variance of the estimators which are defined previously in 13. More specificity,

$$\begin{aligned} H_0 \ \ vs. \ \ H_I : &\quad \lambda_{I,T} = \widehat{\omega}_I/\sigma, \\ H_0 \ \ vs. \ \ H_A : &\quad \lambda_{A,T} = \widehat{\omega}_A/\rho\sigma \end{aligned} \tag{14}$$

Under the null hypothesis of no outliers, these statistics are asymptotically distributed as $N(0, 1)$.

4.1.2 ARMA Parameters and σ^2 Are Unknown

In practice, the ARMA parameters and σ^2 are usually unknown. Then the parameters along with the effect ω can be obtained by maximizing the likelihood function of $(\phi_1, \cdots, \phi_p, \theta 1, \cdots, \theta q, \omega, \sigma^2)$ as described by Box and Jenkins (1976) [3]. Based on these estimates, the likelihood ratios for testing the hypotheses, one against another, in (10) can be computed. Fox (1972) ([10]) first discussed these ratio criteria in an AR context.

The criteria for detecting an outlier at an unknown position then follow. Fox discussed these ratio in a model. The maximum likelihood estimates for the parameters, as well as the likelihood ratios, do not reduce to explicit expressions in general because of the non-linear nature of the AO model and that of the general ARMA models. In this circumstance, the non-linear estimation algorithms are necessary to implement the likelihood ratio tests.

Suppose that the time series Z_t does not contain outliers.
Let $\hat{\phi}_i, \hat{\theta}_j$, and $\hat{\sigma}^2$ be the ML estimate of the parameters ϕ_i, θ_j, and σ^2 respectively. Moreover, let $\hat{U}_t$ be the residuals computed from such an estimated model and $\hat{\pi}(B) = \hat{\phi_B}/\hat{\theta_B}$.

$$\begin{aligned} \hat{\lambda}_{I,T} &= \frac{\hat{\omega}_I}{\hat{\sigma}} \widehat{\omega}_I \\ \hat{\lambda}_{A,T} &= \frac{\widehat{\omega}_A}{\hat{\rho}\hat{\sigma}}. \end{aligned} \tag{15}$$

where

$$\hat{\omega}_I = \hat{U}_T,$$

$$\hat{\omega}_A = \hat{\rho}^2(1 - \hat{\pi}_1 F - \hat{\pi}_2 F^2 - \cdots - \hat{\pi}_{n-T} F^{n-T})\hat{U}_T$$

and $\hat{\rho}^2 = \left(1 - \hat{\pi}_1^2 + \cdots - \hat{\pi}_{n-T}^2\right)^{-1}$.
One can see that $\hat{\lambda}_{I,T}$, and$\hat{\lambda}_{A,T}$ are asymptotically equivalent to the likelihood ratio criteria $\lambda_{I,T}$, and $\lambda_{A,T}$ in the previous testing hypotheses respectively.

In order to detect an IO or an AO at an unknown position, we can calculate η_t, $t = 1, \cdots, n$ where,

$$\eta_t = \max\left\{\left|\hat{\lambda}_{I,t}\right|, \quad \left|\hat{\lambda}_{A,t}\right|\right\}$$

Where in each iteration, the maximum of a given test statistic is selected as the candidate for that type of outlier, and the grand maximum across the tests is identified as the most likely outlier. This grand maximum is then compared with a predetermined positive critical value C so that the existence of an exogenous outlier can be judged.
At $t = T$ we assume an:

$$\text{IO}: \quad \text{if} \quad \max \eta_t = \left|\hat{\lambda}_{I,T}\right| \geq C \tag{16}$$

$$\text{AO}: \quad \text{if} \quad \max \eta_t = \left|\hat{\lambda}_{A,T}\right| \geq C \tag{17}$$

4.1.3 Distinguishing an AO From an IO

In practice, there is insufficient information on the discovered outlier's type; the detection test in 17, and 16 is unclear if it is appropriate for a given situation. When an improper type of test is utilized, the detecting power of the test may be dramatically decreased. Furthermore, even if it is known that

an outlier has occurred at a certain moment, the potentially negative impact of the outlier may be difficult to eliminate unless its nature is properly defined. When the position of the possible outlier is uncertain, we may need to repeat the test at several time periods, which might be time consuming problem. To simplify the task, we consider a basic rule proposed by Fox (1972)[10] as a possible approach to distinguish between an IO and an AO.
At any suspected point T, the possible outlier is categorized as an IO, and AO if $\left|\hat{\lambda}_{I,T}\right| > \left|\hat{\lambda}_{A,T}\right|$, and $\left|\hat{\lambda}_{I,T}\right| \leq \left|\hat{\lambda}_{A,T}\right|$ respectively.

4.1.4 Iterative Procedure For Outlier Detection

Based on the considerations in section 4.1, an iterative procedure has introduced to handle scenarios in which there may exist an unknown number of IO's or AO's. Simply, the approach begins with modelling the original series Z_t and assuling there is no outlier. Then the outlier-detection steps and the parameter-estimation processes will be performed alternately. The following is a full description of the procedure:

I) Outlier-Detection Stage:

1. We compute the residuals $\hat{U}_t$ from the estimated model, and let $\hat{\sigma}^2 = n^{-1}\sum_{t=1}^{n}\hat{U}_t^2$ be the estimate of σ^2. A possible robust alternative estimate of $\hat{\sigma}$ may be based on the median of the absolute values of residuals.

2. Compute $\hat{\lambda}_{I,T}$ and $\hat{\lambda}_{A,T}$ as in 15 and let $\eta_t = \max\left\{\left|\hat{\lambda}_{I,t}\right|, \left|\hat{\lambda}_{A,t}\right|\right\}$ for $t = 1, \ldots, n$. If $\max_t \eta_t = \left|\hat{\lambda}_I, T\right| > C$, where C is a predetermined positive constant, then there is the possibility of an IO at T. The impact ω of this possible IO is estimated by $\hat{\omega}_I$ in (2.2a). We then remove its impact by introducing a new residual $\hat{U}_T = \hat{e}_T - \hat{\omega}_I = 0$ at T. If $\max_t \eta_t = \left|\hat{\lambda}_{A,T}\right| > C$, then there is the possibility of an AO at T and its effect is estimated by $\hat{\omega}_A$ in 15. This AO's effect can be eliminated by defining new residuals $\hat{U}_t = \hat{U}_t - \hat{\omega}_A\hat{\pi}(B)\mathbf{1}_{\{t=T\}}$ for $t \geq T$. In each of the above situations, we obtain new estimate of $\hat{\sigma}^2$ based on the modified residuals.

3. If an IO or an AO is found in step 2, recompute $\hat{\lambda}_{I,T}$ and $\hat{\lambda}_{A,T}$ using the same initial estimates of the time series parameters, but based on the modified residuals $\hat{U}_t$'s and the estimate $\hat{\sigma}^2$, and repeat step 2 .

4. Steps 2 and 3 should be repeated until no further outlier candidates can be discovered.

II) Parameters Estimation Stage:

5. Suppose that IO's or AO's are suspected at different k time points $T_1, T_2, \ldots, T_k$. We consider these times to be known, and simultaneously we estimate the outlier parameters $\omega_1, \omega_2, \ldots, \omega_k$ and the time series parameters, as given by Box and Tiao (1975) [4], using models of the form

$$Z_t = \sum_{j=1}^{k}\omega_j L_j(B)\mathbf{1}_{\{t=T\}} + \frac{\theta(B)}{\phi(B)}\varepsilon_t, \tag{18}$$

Here $L_j(B) = 1$ for an AO and $L_j(B) = \theta(B)/\phi(B)$ for an IO at $t = T_j$.

Considering model 18, we perform the outlier detection stage again. The notations $\hat{\pi}_j$'s, $\hat{\omega}_j$'s and $\hat{U}_t$ represent the estimated values derived from the joint estimation of all model parameters in 18. If no further outliers are detected, we stop. Otherwise, the estimation step is repeated, with the newly detected outliers integrated into model 18, until no more outliers can be discovered and all of the outlier impacts have been simultaneously estimated with the time series parameters.

4.2 Theoretic background for the BP identification method

Suppose that the $ARMA(p,q)$ model as defined previously in 1

$$Y_t = a_1 Y_{t-1} + a_2 Y_{t-2} + ... + a_p Y_{t-p} + \varepsilon_t + b_1 \varepsilon_{t-1} + ... + b_q \varepsilon_{t-q} =$$

$$\sum_{i=1}^{p} a_i Y_{t-i} + \sum_{j=1}^{q} b_j \varepsilon_{t-j} + \varepsilon_t, \quad t = 1, ..., n,$$

is considered; here $\varepsilon_t \sim N(0, \sigma^2)$ are i.i.d. $N(0;1)$ random variables.

Denote by $|\varepsilon|_{(1)} \leq ... \leq |\varepsilon|_{(n)}$ the ordered absolute values $|\varepsilon_1|, ..., |\varepsilon_n|$.
Suppose that Φ is the c.d.f of the normal distribution. Set

$$b_n = \Phi^{-1}\Big(1 - \frac{1}{2n}\Big), \quad a_n = \frac{1}{b_n}. \tag{19}$$

Note that $b_n \to \infty$ as $n \to \infty$ and for any $\delta > 0$.

$$\lim_{n\to\infty} \frac{b_n}{n^\delta} = 2^\delta \lim_{x\to\infty} x[1 - \Phi(x)]^\delta = 2^\delta \lim_{x\to+\infty} (x^{\frac{1}{\delta}-1}\varphi(x))^\delta = 0, \tag{20}$$

because $x[1 - \Phi(x)] \sim \varphi(x)$ as $x \to +\infty$; here $\varphi(x)$ is the probability density function of the standard normal distribution.

We apply theorem 2.1.1 described in [8] to the random variables $|\varepsilon|_{(n)}/\sigma$, and we have for fixed s

$$\Big(\frac{|\varepsilon|_{(n)}/\sigma - b_n}{a_n}, \frac{|\varepsilon|_{(n-1)}/\sigma - b_n}{a_n}, ..., \frac{|\varepsilon|_{(n-s+1)}\sigma - b_n}{a_n}\Big) \xrightarrow{d} L_0 \tag{21}$$

as $n \to \infty$; here

$$L_0 = (-\ln E_1, -\ln(E_1 + E_2), ..., -\ln(E_1 + ... + E_s)) \tag{22}$$

and $E_1, ..., E_s$ are i.i.d. standard exponential random variables.
The parameters $\hat{a_i}$, $\hat{b_j}$ and $\hat{\sigma}$ are robust estimators of a_i, b_j, and σ respectively and they are computed using the R package robustarima. Set

$$r_{(i)} = \hat{|\varepsilon|}_{(i)}(\hat{a}_1, ..., \hat{a}_p, \hat{b}_1, ..., \hat{b}_q))/\hat{\sigma}.$$

If n is large, then for fixed s the distribution of the random vector

$$\Big(\frac{|r|_{(n)} - b_n}{a_n}, \frac{|r|_{(n-1)} - b_n}{a_n}, ..., \frac{|r|_{(n-s+1)} - b_n}{a_n}\Big)$$

is approximated by the distribution of the random vector L_0, where

$$L_0 = (-\ln E_1, -\ln(E_1 + E_2), ..., -\ln(E_1 + ... + E_s))$$

and $E_1, ..., E_s$ are i.i.d. standard exponential random variables.

4.2.1 Outlier identification

Begin by computing robust estimators of the parameters a_i, b_i and σ and the residuals $r_{(i)}$. Set

$$U^{(n)}_{(n-i+1)} = 1 - F_{\chi^2_{2i}}(2e^{-(|r|_{(n-i+1)} - b_n)/a_n}), \tag{23}$$

where $F_{\chi^2_{2i}}(x)$ is the c.d.f. of the chi-square distribution with $2i$ degrees of freedom. Set

$$U^{(n)}(s) = \max_{1 \le i \le s} U^{(n)}_{(n-i+1)}. \tag{24}$$

Under ARMA model and large n the distribution $U^{(n)}(s)$ is approximated by the distribution of the random variable

$$V(s) = \max_{1 \le i \le s} V_i,$$

where $V_i = 1 - F_{\chi^2_{2i}}(2(E_1 + ... + E_i))$, $i = 1, ..., s$, and $E_1, ..., E_s$ are i.i.d. standard exponential random variables. The random variables $V_1, ..., V_s$ are *dependent* identically distributed and the distribution of each V_i is uniform: $V_i \sim U(0, 1)$.

Denote by $v_\alpha(s)$ the α critical value of the random variable $V(s)$. They are easily found many times simulating i.i.d. s standard exponential random variables and computing the values of $V(s)$.

Our simulations showed that the below-proposed outlier identification methods based on exact and approximate critical values of the statistic $U^{(n)}(s)$ give practically the same results, so for samples of size $n \ge 20$ we recommend to approximate the α-critical level of the statistic $U^{(n)}(s)$ by the critical values $v_\alpha(s)$ which depend only on s. We shall see that for the outlier identification only the critical values $v_\alpha(5)$ are needed. We found that the critical values $v_\alpha(5)$ are: $v_{0.1}(5) = 0.9677$, $v_{0.05}(5) = 0.9853$, $v_{0.01}(5) = 0.9975$.

Outlier search procedure begins with an investigation of observations corresponding to the largest values of $|r|_{(i)}$. We recommend beginning with five largest. So take $s = 5$ and compute the value of the statistic $U^{(n)}(5) = \max_{1 \le i \le 5} U^{(n)}_{(n-i+1)}$.
If $U^{(n)}(5) \le v_\alpha(5)$, then we conclude that outliers do not existed and no further investigation is done. If $U^{(n)}(5) > v_\alpha(5)$, then it is concluded that outliers exist and the following classification scheme is done.

Note that (see the classification scheme below) if $U^{(n)}(5) > v_\alpha(5)$, then minimum one observation is declared as an outlier. Thus the probability to declare absence of outliers does not depend on the following classification scheme.
Step 1. Set $d_1 = \max\{i \in \{1, ..., 5\} : U^{(n)}_{(n-i+1)} > v_\alpha(5)\}$. If $d_1 < 5$, then classification is finished at this step: d_1 observations are declared as outliers, other observations are declared as non-outliers. If $d_1 = 5$, then it is possible that the number of outliers is higher than 5. Then the observation corresponding to

$i = 1$ (i.e corresponding to $|r|_{(n)}$)) is declared as an outlier and we proceed to the step 2.

Step 2. The above written procedure is repeated taking $\max_{1\leq i\leq 5} U^{(n-1)}_{(n-i)} = U^{(n-1)}(5)$ instead of $U^{(n)}(5)$; here

$$U^{(n-1)}_{(n-i)} = 1 - F_{\chi^2_{2i}}(2e^{-(|r|_{(n-i)}-b_{n-1})/a_{n-1}}), \quad i = 1, ..., 5,$$

Set $d_2 = \max\{i \in \{1, ..., 5\} : U^{(n-1)}_{(n-i)} > v_\alpha(5)\}$. If $d_2 < 5$, the classification is finished and $d_2 + 1$ observations are declared as outliers.

If $d_2 = 5$, then it is possible that the number of outliers is higher than 6. In such a case, the observation corresponding to the largest residual $|r|_{(n-1)}$ is declared as an outlier, in total 2 observations (i.e. corresponding to $|r|_{(n)}, |r|_{(n-1)}$) are declared as outliers at this step, but classification is not finished and we repeat the procedure.

Classification finishes at the lth step if $d_l < 5$. So we declare $(l - 1)$ outliers in the previous steps and d_l outliers in the last one. The total number of observations declared as outliers is $l-1+d_l$. These observations correspond to $|r|_{(n)}, ..., |r|_{(n-d_l-l+2)}$.

Note that for fixed α ($\alpha = 0.05$, for example) only one critical value $v_\alpha(5)$ ($v_{0.05}(5) = 0.9853$, for example) is needed.

5 Experiment and Results

5.1 Simulation Scheme

We compare the performance of the *Chang* and *BP* outlier identification methods by running several experiments using three different R libraries; tsa [5], tsoutliers [11], and robustarima [12] for Chang method, and original R script for generalized BP method. These libraries tsoutliers and robustarima modify the Chang procedure using different approach of robust estimation and including search of other than AO and IO types of outliers such as LS and TC. However, we are interested only on AO and IO. The performance of the methods is concluded according to masking values since the swamping values are not high in majority of the cases. The simulation scheme is described as follows.

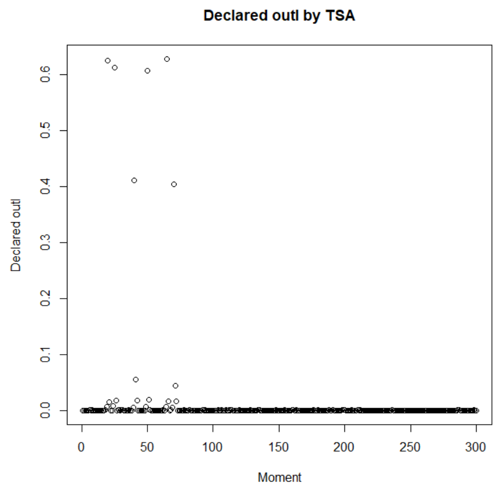
Assume that $n = 300, m = 5000$, and $\sigma = 1$ is the sample size, number of the simulations, and the value of standard deviation. To allow comparison between the two techniques. Three factors are considered: (a) type of outlier, AO, and IO; (b) time series structure, AR(1), MA(1) and ARMA(1,1); (c) outlier size, $\pm 4\sigma$, $\pm 5\sigma$; (d) number of outliers, 8, and 4 for AO and IO cases receptively.

5.2 Additive Outlier Case

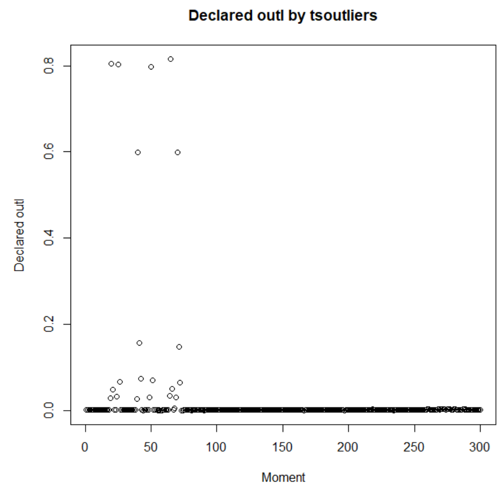
5.2.1 Comparative analysis for AR(1)

For fixed n, m, and σ. First, we generate an AR(1)according to the model, $y_t = 0.6y_{t-1} + \varepsilon_t$. Then, these series is perturbed with eight additive outliers at different time $T = (20, 25, 40, 41, 50, 65, 70, 71)$, with outlier's size $\pm 4\sigma$, then $\pm 5\sigma$.

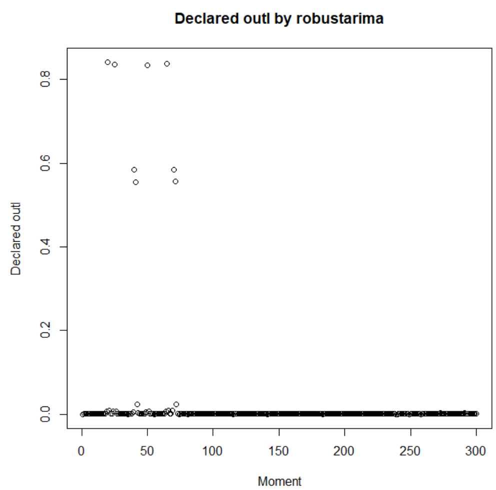
With $\omega = 4$ which is not large effect (size of outlier), the tsa and tsoutliers methods do not find adjacent (neighboring) outliers whereas robustarima and BP methods find them often as shown in Figure 1. Thus, we included into the table 1 only observations neighbouring with outliers because after many simulations other observations where declared as outliers very rarely.



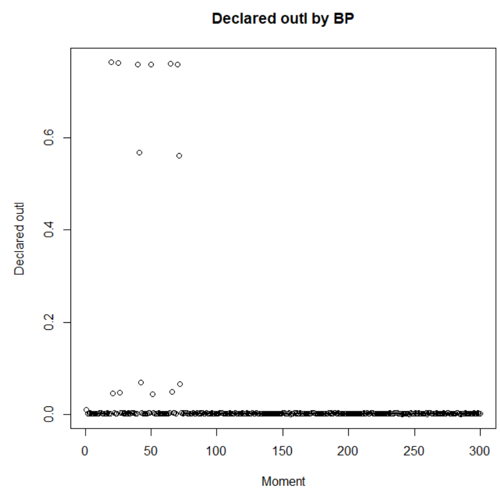
(a) Detected outliers by tsa



(b) Detected outliers by tsoutlier



(c) Detected outliers by robarim



(d) Detected outliers by BP

Figure 1: Detected additive outliers in AR(1) with $\omega = 4$

In Table (1), we can see that outliers at $T = 20, 25, 40, 50, 65$ were mostly detected by most of the methods but the highest detecting rate was for robarima. At $T = 40$ and $T = 70$ the BP method considerably outperforms robustarima (0.76 against 0.58 in both cases) . At the next neighboring points (41 and 71) the power is a little smaller for both BP and robustarima methods (0.57 and 0.56) whereas TSA and tsoutliers methods fail to find outliers.
In other hand, when ω increases to five. Table 2 shows that most of the methods detected the introduced outliers more often than the previous case with low outlier's impact ($\omega = 4$). From the results we can see that the correct detecting rate increase with higher $\omega = 5$. It is varying between 90% to 97% for all methods. The interesting results were at $T = 41$, and 71 and even though they are neighbouring points, the BP and robarim has succefully detect them as true outliers with rate of 87% this time much

	19	**20**	21	24	**25**	26
tsa	0.0062	**0.6246**	0.0152	0.0076	**0.6128**	0.0168
tsout	0.0284	**0.8050**	0.0488	0.0316	**0.8026**	0.0652
robarim	0.0058	**0.8408**	0.0080	0.0060	**0.8358**	0.0068
BP	0.0014	**0.7630**	0.0448	0.0020	**0.7614**	0.0474
	39	**40**	**41**	49	**50**	51
tsa	0.0046	**0.4110**	**0.0550**	0.0058	**0.6078**	0.0190
tsout	0.0264	**0.5994**	**0.1562**	0.0294	**0.7974**	0.0704
robarim	0.0054	**0.5834**	**0.5540**	0.0050	**0.8344**	0.0072
BP	0.0016	**0.7586**	**0.5670**	0.0030	**0.7576**	0.0434
	64	**65**	66	69	**70**	**71**
tsa	0.0066	**0.6276**	0.0158	0.0048	**0.4038**	**0.0442**
tsout	0.0342	**0.8150**	0.0492	0.0298	**0.5990**	**0.1474**
robarim	0.0062	**0.8382**	0.0092	0.0076	**0.5834**	**0.5554**
BP	0.0026	**0.7602**	0.0492	0.0020	**0.7586**	**0.5612**

Table 1: Declared AO in AR(1) ($\omega = 4$)

better than last experiment.

	19	**20**	21	24	**25**	26
tsa	0.0098	**0.8982**	0.0240	0.0106	**0.8950**	0.0280
tsout	0.0610	**0.9710**	0.0884	0.0588	**0.9708**	0.1194
robarim	0.0036	**0.9814**	0.0024	0.0054	**0.9808**	0.0022
BP	0.0010	**0.9534**	0.0236	0.0024	**0.9478**	0.0276
	39	**40**	**41**	49	**50**	51
tsa	0.0100	**0.7226**	**0.1458**	0.0100	**0.8944**	0.0298
tsout	0.0618	**0.8788**	**0.3688**	0.0570	**0.9728**	0.1264
robarim	0.0050	**0.9164**	**0.8750**	0.0052	**0.9784**	0.0030
BP	0.0024	**0.9502**	**0.8678**	0.0028	**0.9508**	0.0260
	64	**65**	66	69	**70**	**71**
tsa	0.0124	**0.9004**	0.0254	0.0082	**0.7210**	**0.1372**
tsout	0.0662	**0.9718**	0.0834	0.0618	**0.8742**	**0.3542**
robarim	0.0046	**0.9774**	0.0036	0.0054	**0.9130**	**0.8752**
BP	0.0020	**0.9468**	0.0278	0.0020	**0.9486**	**0.8648**

Table 2: Declared AO in AR(1) ($\omega = 5$)

Let us compare how BP-method, and Chang iterative procedure work with $r = 8$ of outliers and investigate swamping and masking effect to test their performance. For AR(1), masking values are presented in 3 Table 3 shows that the masking values of the BP and Chang methods is high for a

Method	Masking	Swamping
tsa	57.66%	0.04%
tsoutlier	40.97%	0.27%
robarim	29.68%	0.1%
BP	28.9%	0.32%

Table 3: The masking and swamping values for AR model ($\omega = 4$)

small outlier's effect $\omega = 4$. Chang-method has an enormous masking effect of 57.66% of outliers are not found; 29.68%, and 28.9% of outliers are not found for robarima and BP respectively. In this case, BP method has a heavier swamping effect 0.32%, but smaller masking effect. However, robarim has pretty small masking and swamping values which promote it to be the best method since it can identify different type of outliers such as (LS, and VC) whereas BP not. The masking values for all methods decrease when ω increase to five as we can see in table 4. Most of the methods have a dramatically decrease in the masking values. In this case, only 6.28%, and 7.12% of true outliers are not identified by robarima and BP respectively. Moreover, the masking values for tsa, and tsoutlier decreased to 33.56%, and 20.47% respectively. However, the swamping values has slightly increment. For BP method swamping effect is maximum among the others.

Conclusion: robustarima method (it appeared only in 2021) is the best method, second is the BP method, other are considerably worse. But in situations with adjacent outliers the BP method slightly outperforms the robustarima method. The drawback of the BP method is that it does not define the type of outliers.

Method	Masking	Swamping
tsa	33.56%	0.07%
tsoutlier	20.47%	0.45%
robarim	6.28%	0.1%
BP	7.12%	0.29%

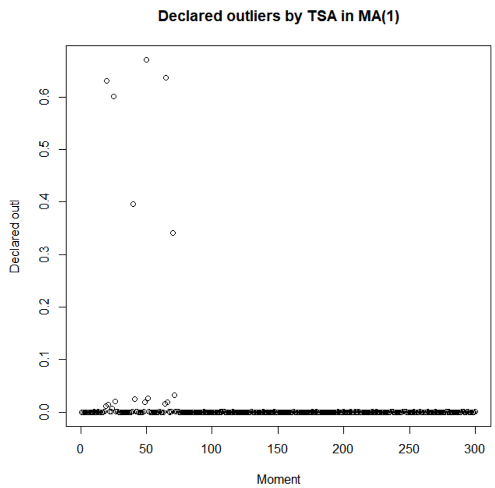
Table 4: The masking and swamping values for AR model ($\omega = 5$)

5.2.2 Comparative analysis for MA(1)

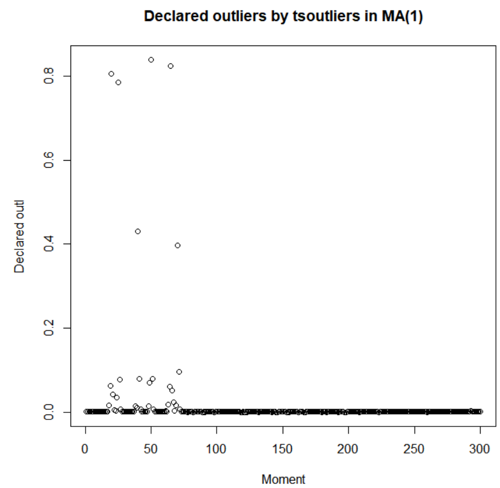
In this section was investigated the performance of the outlier identification methods in MA(1) using the following model

$$Y_t = \varepsilon_t - 0.4\varepsilon_{t-1}.$$

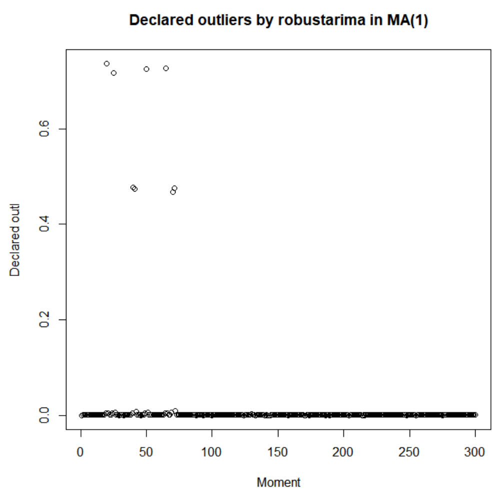
In this section we noticed almost same remarks as the previous ones (See figure:2). Most of the methods successfully detected the outliers in different time points. However, they fail to detect the neighbouring outliers at $T = 41$, and 71, where tsa and tsout have identified them correctly only 10% time over all simulations (See table: 5). Also, at same points BP method and robustarima results are almost same (0.47 in both cases). However, $T = 40$, and $T = 70$, BP method considerably outperform robustarima (0.69 against 0.47 in both cases). Moreover, at $T = 20$, 25, and 50 robarim has the highest detecting rates more than 80% but it decreases at some other outliers. One can notice that BP shows good results, it discovered most of the outliers with steady detection rate around 70% and does not change much.



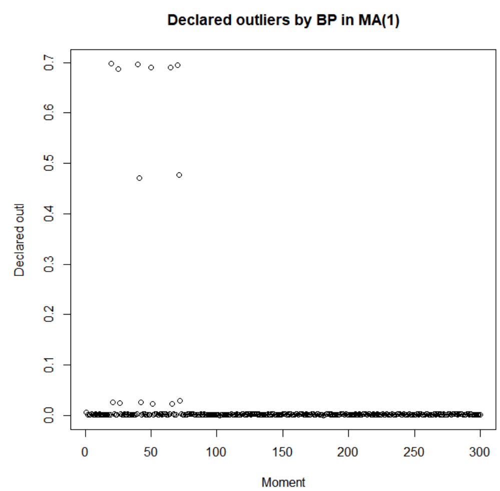
(a) Detected outliers by tsa



(b) Detected outliers by tsoutlier



(c) Detected outliers by robarim



(d) Detected outliers by BP

Figure 2: Detected additive outliers in MA(1) with $\omega = 4$

	19	**20**	21	24	**25**	26
tsa	0.0114	**0.6302**	0.0144	0.0068	**0.6006**	0.0198
tsout	0.0624	**0.8062**	0.0412	0.0342	**0.7854**	0.0778
robarim	0.0042	**0.9428**	0.0226	0.0066	**0.9424**	0.0196
BP	0.0012	**0.6976**	0.0254	0.0016	**0.6866**	0.0250
	39	**40**	41	49	**50**	51
tsa	0.0008	**0.3962**	**0.0242**	0.0190	**0.6704**	0.0256
tsout	0.0098	**0.4302**	**0.0794**	0.0704	**0.8394**	0.0790
robarim	0.0038	**0.4770**	**0.4736**	0.0036	**0.7252**	0.0060
BP	0.0020	**0.6966**	**0.4718**	0.0018	**0.6900**	0.0236
	64	**65**	66	69	**70**	**71**
tsa	0.0158	**0.6372**	0.0188	0.0002	**0.3406**	**0.0316**
tsout	0.0602	**0.8244**	0.0508	0.0158	**0.3976**	**0.0952**
robarim	0.0048	**0.7270**	0.0044	0.0052	**0.4674**	**0.4758**
BP	0.0022	**0.6906**	0.0236	0.0018	**0.6944**	**0.4772**

Table 5: Detected AO for MA(1) model ($\omega = 4$)

	19	**20**	21	24	**25**	26
tsa	0.0202	**0.8972**	0.0236	0.0116	**0.8862**	0.0320
tsout	0.0234	**0.9340**	0.0222	0.0216	**0.9272**	0.0388
robarim	0.0040	**0.9518**	0.0026	0.0044	**0.9538**	0.0018
BP	0.0010	**0.9216**	0.0164	0.0018	**0.9232**	0.0158
	39	**40**	**41**	49	**50**	51
tsa	0.0010	**0.7230**	**0.0732**	0.0276	**0.9192**	0.0358
tsoutl	0.0392	**0.8688**	**0.3886**	0.0192	**0.9272**	0.0318
robarim	0.0040	**0.8602**	**0.8428**	0.0040	**0.9522**	0.0018
BP	0.0016	**0.9234**	**0.8252**	0.0016	**0.9216**	0.0158
	64	**65**	66	69	**70**	**71**
tsa	0.0230	**0.9030**	0.0280	0.0002	**0.6794**	**0.0854**
tsoutl	0.0278	**0.9296**	0.0204	0.0416	**0.8762**	**0.3976**
robarim	0.0054	**0.9484**	0.0032	0.0046	**0.8578**	**0.8414**
BP	0.0024	**0.9234**	0.0142	0.0022	**0.9264**	**0.8284**

Table 6: Detected AO for MA(1) model ($\omega = 5$)

Table 6 shows that the method's performance improved when we increase the AO's size to 5. With this high outlier's impact, majority of outliers were detected correctly. BP and robarim has the highest detecting rate which it reaches 90% at $T = 20$, 25, 40, 65, and 70. Also, at the neighbouring points $T = 41$, and $T = 71$ robarim and BP methods performance is almost same. The latter method succeeded in detecting these critical outliers with rate of 82% against robarima with rate approximately to 85%. However, tsa fails to detect the outliers at the neighbouring points with rate of 0.07, and 0.08 at $T = 41$, and $T = 71$ respectively.

Let us compare BP, with Chang method when 8 outliers are introduced and check how both of these procedures work for different outlier's sizes.

1) For $\omega = 4$. Table: 7 shows that comparing masking values of BP and Chang methods the latter method's value are larger than BP method's. Masking values for tsa, and tsoutlier are 0.58, and 0.46 respectively. These values are big comparing to BP method which has the lowest value with 0.36 and it performs better than robarim which has 0.4. From the results, one can notice that swamping values for tsa and robarim are near to 0. However, the highest value is for tsoutl and this might be due to the confusing between the AO, and IO. Also, we can notice that swamping values for BP method is 0.22% and is bigger than robarim.

Method	Masking	Swamping
tsa	58.36%	0.05%
tsoutlier	46.77%	2.66%
robarim	40.01%	0.06%
BP	36.19%	0.22%

Table 7: The masking and swamping values for MA model ($\omega = 4$)

2) For $\omega = 5$. In table:8 data for all methods shows that all masking values have significant decrease when ω increases. This prove, that for large outlier's size all methods detect more outliers than previous case. Masking value for tsa is higher with 0.35. It is clear that BP-method and robarim have the lowest

masking values with 10.08%, and 9.8% receptively. This means that the identification methods perform much better when the size of the outlier is large. In other hand, swamping values on tsoutlier decreases to around zero but for BP-method remains the same and did not change.

Conclusion: Therefore, the results are very similar as AR model (see Section: 5.2.1). In most considered scenarios, the BP method is better than the iterative procedure of Chang since swamping values are not important as masking values.

Method	Masking	Swamping
tsa	35.41%	0.07%
tsoutlier	21.88%	0.17%
robarim	9.8%	0.07%
BP	10.08%	0.22%

Table 8: The masking and swamping values for MA model ($\omega = 5$)

5.2.3 Comparative Analysis For ARMA(1,1)

In this section, we analyze the performance of BP and Chang methods for ARMA. The following model is used

$$Y_t = 0.5Y_{t-1} + 0.6\varepsilon_{t-1} + \varepsilon_t.$$

At T = 20,25, 40, 41, 50, 65, 70, and 71 we introduce eight additive outliers with size four. Table 9 shows that majority of the added outliers were correctly declared as outliers by all methods. In this situation, tsa and tsout outperform the robarim and BP methods at T = 20, 25, 50, and 65 with an approximate correct detection rate of 85%, and 96% for tsa and tsout respectively. However, BP method has the highest identification rate with 61% at $T = 40$, and $T = 70$. One can notice that this rate is steady for other points except at the neighbouring points $T = 41$, $T = 71$ where its power is slightly better than robarim. Moreover, at these adjacent positions, tsa and tsoutl have detection's rate near to zero. Furthermore, the correct time detection rate dramatically increases when ω increases to five as it shown in table 10. At $T = 20, 25, 50$, and 65, tsa, tsoutl and robarim almost detect all the outliers with high rate around 0.99 while the results of BP method improved from 61% to 87% while it performs better than robarim at $T = 40$, and $T = 70$. However, the neighbouring outliers at $T = 41$, and 71 tsa and tsoutl fail to detect them whereas both of robarim and BP declared them as outliers with correct detection rate approximately around 64% for both cases.

	19	**20**	21	24	**25**	26
tsa	0.1098	**0.8510**	0.1682	0.0698	**0.8592**	0.1786
tsout	0.3234	**0.9568**	0.3546	0.2444	**0.9610**	0.4420
robarim	0.0030	**0.7282**	0.0266	0.0050	**0.7282**	0.0206
BP	0.0020	**0.6106**	0.1388	0.0026	**0.6140**	0.1374
	39	**40**	**41**	49	**50**	51
tsa	0.0034	**0.3268**	**0.0022**	0.1234	**0.8820**	0.1982
tsout	0.0220	**0.2754**	**0.0154**	0.3426	**0.9688**	0.4580
robarim	0.0048	**0.3084**	**0.2820**	0.0050	**0.7250**	0.0208
BP	0.0016	**0.6186**	**0.2920**	0.0030	**0.6148**	0.1370
	64	**65**	66	69	**70**	**71**
tsa	0.1230	**0.8644**	0.1858	0.0010	**0.2942**	**0.0028**
tsout	0.3390	**0.9630**	0.3852	0.0114	**0.2576**	**0.0202**
robarim	0.0040	**0.7198**	0.0222	0.0058	**0.3018**	**0.2768**
BP	0.0020	**0.6104**	0.1400	0.0024	**0.6144**	**0.2894**

Table 9: Declared AO in ARMA(1,1) ($\omega = 4$)

	19	**20**	21	24	**25**	26
tsa	0.1586	**0.9804**	0.2394	0.1084	**0.9838**	0.2546
tsout	0.4642	**0.9988**	0.4936	0.3904	**0.9974**	0.5778
robarim	0.0042	**0.9428**	0.0226	0.0066	**0.9424**	0.0196
BP	0.0024	**0.8722**	0.1104	0.0032	**0.8788**	0.1082
	39	**40**	41	49	**50**	51
tsa	0.0054	**0.6246**	0.0072	0.1652	**0.9838**	0.2626
tsout	0.0526	**0.6128**	0.0650	0.4738	**0.9974**	0.5920
robarim	0.0058	**0.7144**	0.6416	0.0060	**0.9372**	0.0194
BP	0.0012	**0.8812**	0.6550	0.0030	**0.8784**	0.1116
	64	**65**	66	69	**70**	**71**
tsa	0.1690	**0.9810**	0.2588	0.0052	**0.5894**	**0.0092**
tsout	0.4760	**0.9984**	0.5062	0.0330	**0.5864**	**0.0668**
robarim	0.0042	**0.9398**	0.0226	0.0070	**0.7000**	**0.6362**
BP	0.0020	**0.8790**	0.1084	0.0022	**0.8738**	**0.6412**

Table 10: Declared AO in ARMA(1,1) ($\omega = 5$)

Let us compare the performance of the methods according to the masking and swamping values. Table 11 data shows that masking values for all method are close. The values are approximately in range between 44.77% to 49.12%. The lowest value is for tsoutl and then BP which perfrom better than robarim. The latter method has the highest masking value, where 49.12% are not detected. Moreover, swamping value are heavy for tsa and BP methods but for robarim is small enough.

In other hand , table 4 data presents that masking values decrease if the outlier size ω increases to five. In this case, masking values of BP-method and robarima are 18%, and 19.32% respectively. These values are much smaller than tsa and tsoutl methods, where 35.5%, and 33.46% of true outliers are not discovered respectively using these methods. The obtained results confirm that outliers with large size are more likely to be detected by the majority of methods. Furthermore, swamping values are heavy for all methods. BP and robarim methods have less swamping values than tsa and tsoutl where 0.49%,

Method	Masking	Swamping
tsa	48.96%	0.4%
tsoutlier	44.77%	1.08%
robarim	49.12%	0.09%
BP	46.69%	0.49%

Table 11: The masking and swamping values for ARMA model ($\omega = 4$)

and 0.12% are not detected respectively. It is clear that tsoutl has the highest swamping value with 1.5%.

Method	Masking	Swamping
tsa	35.5%	0.57%
tsoutlier	33.46%	1.5%
robarim	19.32%	0.12%
BP	18%	0.49%

Table 12: The masking and swamping values for ARMA model ($\omega = 5$)

Conclusion: Considering all data it can be stated that robarim and BP methods work better than tsa and tsoutl with both outliers size $\omega = 4$, and $\omega = 5$.

5.3 Innovative Outlier Case

5.3.1 Comparative Analysis For AR(1)

In this section, we investigate the performance of previous methods in detecting the innovative outliers when they are introduced to the series at $T = 20, 90, 150$, and 200 with outlier size of four. Table 13 data shows that BP method is better in detecting the outliers in all points with correct identification rate of 76% then tsout comes next with an approximate rate of 66%. Moreover, robarim fails to detect the added IO with lowest rate around 45%. Furthermore, from results we see that robarim and BP methods detect some outliers at the neighbouring since the introduced outliers effect on the sequence of next observations.

	20	21	**90**	91
tsa	**0.5006**	0.0000	**0.5106**	0.0002
tsout	**0.6614**	0.0006	**0.6606**	0.0008
robarim	**0.4628**	0.0070	**0.4550**	0.0062
BP	**0.7630**	0.0032	**0.7690**	0.0032
	150	151	**200**	201
tsa	**0.5058**	0.0000	**0.5130**	0.0000
tsout	**0.6596**	0.0004	**0.6648**	0.0012
robarim	**0.4530**	0.0060	**0.4624**	0.0056
BP	**0.7664**	0.0032	**0.7720**	0.0046

Table 13: Declared IO for AR model ($\omega = 4$)

In the other hand, majority of the methods are more likely to detected the introduced outliers when we increase the outlier size to five. Table 14 data presents that detecting rate of tsa, tsoul and robarim improve where almost all the outliers were discovered whereas for BP method was steady with an approximate 88%.

	20	21	**90**	91
tsa	**0.9780**	0.2010	**0.9796**	0.2056
tsout	**0.9594**	0.1464	**0.9628**	0.1456
robarim	**0.9516**	0.0086	**0.9550**	0.0104
BP	**0.8802**	0.0816	**0.8808**	0.0756
	150	151	**200**	201
tsa	**0.9798**	0.1976	**0.9774**	0.2122
tsout	**0.9610**	0.1446	**0.9606**	0.1546
robarim	**0.9512**	0.0094	**0.9492**	0.0090
BP	**0.8872**	0.0778	**0.8784**	0.0796

Table 14: Declared IO for AR model ($\omega = 5$)

Let us investigate the power of methods and compare their masking, and swamping values. For small outlier size $\omega = 4$. Table 15 data shows that the masking value for BP method is 23.24% and is much smaller than tsoutl which comes second with 33.84% whereas, 54.17% of innovative outliers were not declared as true outliers with robarim and this may be due to the small size of outlier. Moreover, the swamping values is high for BP and tsout methods with rate of 0.22%, and 0.19% respectively and for tsa and robarim is around zero.

Method	Masking	Swamping
tsa	49.25%	0.04%
tsoutlier	33.84%	0.19%
robarim	54.17%	0.05%
BP	**23.24%**	0.22%

Table 15: The masking and swamping values for AR model ($\omega = 4$)

Method	Masking	Swamping
tsa	17.41%	0.09%
tsoutlier	8.72%	0.32%
robarim	23.44%	0.07%
BP	**4.76%**	0.23%

Table 16: The masking and swamping values for AR model ($\omega = 5$)

In other hand, for large outlier size Table 16 data shows that comparing masking values of the methods, BP method has lowest value where only 4.73% of true innovative outliers are not detected whereas tsout comes next with value 0.0872. In this situation, robarim shows worse performance with with highest masking value 0.2344. However, tsout and BP has the highest swamping values rate 0.23, and 0.09 receptively.

Conclusion: The data shows that the outlier size does effect on the method's performance. Moreover, the results lead to that BP method outperform the robarim method in both cases. The latter method has high masking values among the other and this might be due to its power to detect other outlier's types. Thus, it suffers from the confusion problem between IO and LS.

5.3.2 Comparative Analysis For MA(1)

In this section, the performance of the outlier detection methods was tested for MA(1) model with outlier size $\omega = 4$. Table 17 data shows that majority of the outliers were declared by all method. Outliers at $T = 20, 90, 150$, and 200 are more often declared correctly as outliers by tsoutl and robarim with high rate around 82% whereas BP method, and tsa have approximately same rate 70% most of the time. Furthemore, outliers at the neighbouring points $T = 21, 91, 151$, and 201 are more likely to be detected as outliers than previous section 5.2.2 due to the effect of innovative outliers on the next observations. However, table 18 data shows that the detecting power of all methods increase when innovational outlier size increase to five ($\omega = 5$). It can be seen that at $T = 20, 90, 150$, and 200 tsout and robarim have the highest correct identification rate reach 96% and BP method performance improved to reach 91% whereas tsa fails to detect some outliers where it has rate of 87%. Moreover, at the neighbouring points $T = 21, 91, 151$, and 201 tsout and BP perform better than robarim which detect the innovative outliers only 24% from all the simulations. They successfully detect these adjacent outliers with an approximate high rate of 67%, and 50% for tsout and BP respectively.

	20	21	**90**	91
tsa	**0.6858**	0.2762	**0.6930**	0.2822
tsout	**0.8342**	0.5242	**0.8388**	0.5314
robarim	**0.8216**	0.1658	**0.8206**	0.1710
BP	**0.6928**	0.3470	**0.7006**	0.3462
	150	151	**200**	201
tsa	**0.6956**	0.2790	**0.6960**	0.2820
tsout	**0.8350**	0.5264	**0.8446**	0.5242
robarim	**0.8228**	0.1704	**0.8188**	0.1696
BP	**0.6942**	0.3476	**0.7066**	0.3462

Table 17: Declared IO for MA model ($\omega = 4$)

	20	21	**90**	91
TSA	**0.8712**	0.3720	**0.8772**	0.3780
tsout	**0.9538**	0.6772	**0.9568**	0.6820
robarim	**0.9610**	0.2422	**0.9590**	0.2448
BP	**0.9196**	0.4976	**0.9236**	0.5014
	150	151	**200**	201
TSA	**0.8728**	0.3730	**0.8812**	0.3778
tsout	**0.9550**	0.6822	**0.9528**	0.6746
robarim	**0.9572**	0.2474	**0.9590**	0.2454
BP	**0.9170**	0.4996	**0.9236**	0.5030

Table 18: Declared IO for MA model ($\omega = 5$)

Table 19 data shows that masking values are high for all methods except the robarim that fails to detect only 8.89% of outliers which is the lowest among the other. Also, tsa and BP methods fail to detect some true outliers with masking value around 0.3 for both methods. However, swamping values are heavy for all methods and the heaviest one is for tsoutl with 0.88% then 0.67% for BP.

Method	Masking	Swamping
tsa	30.74%	0.38%
tsoutlier	16.18%	0.88%
robarim	8.89%	0.27%
BP	30.14%	0.67%

Table 19: The masking and swamping values for MA model ($\omega = 4$)

In the other hand, by increasing the outlier size the performance of BP and tsoutl methods significantly improved (see Table 20). The masking values of the latter method and BP are 4.54%, and 7.9% respectively whereas robarim remains the best detecting method with small masking value 4.09%.

Conclusion: robarim method identifies innovational outliers in MA model more successfully than BP method. This, maybe due to its power to solve the confusion between the IO and LS.

Method	Masking	Swamping
tsa	12.44%	0.51%
tsoutlier	4.54%	1.15%
robarim	4.09%	0.37%
BP	7.9%	0.89%

Table 20: The masking and swamping values for MA model ($\omega = 5$)

5.3.3 Comparative Analysis For ARMA(1,1)

In this section, we analyze the performance of BP and Chang methods for ARMA in the case of innovative outliers. For small outlier size ($\omega = 4$), data in Table 21 shows that most of the methods detect the IO at $T = 20, 90, 150,$ 200 but still some outliers are undeclared. The tsa has the highest detecting rate with 85% then tsout with 75% whereas the correct identification rate for robarim and BP methods is 73%, and 62% respectively. It is clear that majority of the methods fail to detect the adjacent outliers at $T = 21, 91, 151,$ and 201. The BP method outperform robarim (approximately 8% against 1%) and tsa has the highest detecting rate. In the other hand, Table 22 data shows that the performance of the methods improve. With large outlier size ($\omega = 5$) most of the outliers are detected by all methods but more with tsa. However, the methods fail to detect the adjacents outliers and their performance do not improve.

	20	21	**90**	91
TSA	**0.8534**	0.1494	**0.8534**	0.1484
tsout	**0.7514**	0.0736	**0.7538**	0.0754
robarim	**0.7338**	0.0106	**0.7344**	0.0116
BP	**0.6276**	0.0838	**0.6292**	0.0794
	150	151	**200**	201
TSA	**0.8502**	0.1420	**0.8544**	0.1536
tsout	**0.7530**	0.0656	**0.7464**	0.0784
robarim	**0.7362**	0.0098	**0.7332**	0.0124
BP	**0.6322**	0.0868	**0.6204**	0.0854

Table 21: Declared IO for ARMA model ($\omega = 4$)

	20	21	**90**	91
TSA	**0.9780**	0.2010	**0.9796**	0.2056
tsout	**0.9594**	0.1464	**0.9628**	0.1456
robarim	**0.9516**	0.0086	**0.9550**	0.0104
BP	**0.8802**	0.0816	**0.8808**	0.0756
	150	151	**200**	201
TSA	**0.9798**	0.1976	**0.9774**	0.2122
tsout	**0.9610**	0.1446	**0.9606**	0.1546
robarim	**0.9512**	0.0094	**0.9492**	0.0090
BP	**0.8872**	0.0778	**0.8784**	0.0796

Table 22: Declared IO for ARMA model ($\omega = 5$)

Let us compare the performance of the methods based on the masking and swamping values. Table 23 data shows that tsa has the lowest masking value 0.14 where tsout and robarim have close results (0.24 and 0.26 for each respectively) whereas BP method does not perform good in detecting the innovative outliers and its masking value is the highest with rate of 37.26%. Moreover, swamping values are high for all methods except for robarim which has only 0.04%.

Method	Masking	Swamping
tsa	14.14%	0.38%
tsoutlier	24.88%	0.19%
robarim	26.56%	0.04%
BP	37.26%	0.33%

Table 23: The masking and swamping values for ARMA model ($\omega = 4$)

However, with large outlier size ($\omega = 5$). One can see that masking values dramatically decrease. The tsa method has again the lowest masking value but highest swamping value. Despite the masking value of BP method decreases from 0.37 to 0.11, it remains the highest comparing to the others.

Conclusion: Our analysis shows that for ARMA model the tsa method which based on Chang procedure perform better in detecting the innovational outliers even though the swamping values are high. However, BP method does not success in most cases.

Method	Masking	Swamping
tsa	2.13%	0.54%
tsoutlier	3.9%	0.37%
robarim	4.82%	0.05%
BP	11.83%	0.36%

Table 24: The masking and swamping values for ARMA model ($\omega = 5$)

6 Conclusion

We compared the performance of two outliers detection methods. As a first remark, we were interested only in the case of AO and IO. Also, swamping values were not high for all methods, so more important were the masking values. Generally, the simulation results on the AR(1), MA(1) and ARMA(1,1) for both cases (AO and IO) showed that the robarim which is based on Chang iterative procedure but has some modifications to identify other types of outliers (LS and TC) has the lowest masking value while the BP method shows its power in detecting adjacent outliers and it gives slightly better results in some situations. However, the drawback of the latter method is the inability to define the type of outliers and distinguish between AO and IO.

For fixed numbers of additive and innovative outliers, our results showed that the masking value was affected by the size of the outlier. In the majority of cases, the masking values decreased for all methods when we use large outlier sizes whereas the swamping values increased.

As a result, our analysis showed that we should recommend using the package *robustarima* for outlier search in ARMA (and ARIMA) models.

References

[1] Vilijandas Bagdonavičius and Linas Petkevičius. Multiple outlier detection tests for parametric models. *Mathematics*, 8(12):2156, 2020.

[2] Nathan S Balke. Detecting level shifts in time series. *Journal of Business & Economic Statistics*, 11(1):81–92, 1993.

[3] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[4] George EP Box and George C Tiao. Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical association*, 70(349):70–79, 1975.

[5] Kung-Sik Chan, Brian Ripley, Maintainer Kung-Sik Chan, and Sik Chan. Package 'tsa'. *R package version*, 1, 2020.

[6] Ih Chang, George C Tiao, and Chung Chen. Estimation of time series parameters in the presence of outliers. *Technometrics*, 30(2):193–204, 1988.

[7] Chung Chen and Lon-Mu Liu. Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, 88(421):284–297, 1993.

[8] Laurens De Haan, Ana Ferreira, and Ana Ferreira. *Extreme value theory: an introduction*, volume 21. Springer, 2006.

[9] Elisa Ferrari, Paolo Bosco, Sara Calderoni, Piernicola Oliva, Letizia Palumbo, Giovanna Spera, Maria Evelina Fantacci, and Alessandra Retico. Dealing with confounders and outliers in classification medical studies: The autism spectrum disorders case study. *Artificial Intelligence in Medicine*, 108:101926, 2020.

[10] Anthony J Fox. Outliers in time series. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(3):350–363, 1972.

[11] Ruihong Huang, Zhiwei Chen, Zhicheng Liu, Shaoxu Song, and Jianmin Wang. Tsoutlier: Explaining outliers with uniform profiles over iot data. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 2024–2027. IEEE, 2019.

[12] Rob J Hyndman. Cran task view: Time series analysis. 2021.

[13] Henrik Madsen. *Time series analysis*. Chapman and Hall/CRC, 2007.

[14] Daniel Pefia. Measuring influence in dynamic regression models. *Technometrics*, 33(1):93–101, 1991.

[15] Daniel Peña. Influential observations in time series. *Journal of Business & Economic Statistics*, 8(2):235–241, 1990.

[16] Robert H Shumway, David S Stoffer, and David S Stoffer. *Time series analysis and its applications*, volume 3. Springer, 2000.

[17] Ruey S Tsay. Outliers, level shifts, and variance changes in time series. *Journal of forecasting*, 7(1):1–20, 1988.

[18] Keith Worden, Graeme Manson, and Nick RJ Fieller. Damage detection using outlier analysis. *Journal of sound and vibration*, 229(3):647–667, 2000.

[19] Conghai Zhang, Xinyao Xiao, and Chao Wu. Medical fraud and abuse detection system based on machine learning. *International Journal of Environmental Research and Public Health*, 17(19):7265, 2020.