ISSN: 2752-3829 Vol. 3 No.2, (December, 2023)

*Stochastic Modelling and Computational Sciences*

# EXPLORING THE MULTIPLICATIVE OUTLIER IN AUTOREGRESSIVE TIME SERIES MODELING

**Jitendra Kumar and Bhagchand Meena**
Department of Statistics, Central University of Rajasthan, Bandersindri, Ajmer, Rajasthan, India

## ABSTRACT
*The present paper investigates multiplicative outliers in time series modeling. Multiplicative outliers are data points that exhibit a significant departure from the expected behavior and can distort the model's estimation. The study proposes a Bayesian approach for multiplicative outliers that incorporates prior knowledge and beliefs about the model parameters. It takes into account the outlier of the data. The results demonstrate the effectiveness of the proposed method in multiplicative outliers and improving the accuracy of the time series model. The study highlights the importance of considering outliers in time series modeling and provides a practical approach for the estimation of multiplicative outliers in this context.*

*Keywords: Multiplicative outlier, Autoregressive model, Posterior probability, Bayesian inference.*

## INTRODUCTION
A time series is a set of observations that are recorded in time. This may also be described as "a time series is a sequence of observations of variables recorded at equally spaced time points'' [1]. Time-series observations are sometimes affected by interrupting events such as strikes, pandemics, natural disasters, and sudden political and economic crises, such values are usually referred to as outliers. Because Outliers have been known to wreak havoc on parameter estimation. it is therefore important to have procedures that will deal with such outlier effects.

In real life, a time series may include anomalous data for a variety of causes that don't match the majority of the observations. Outliers are these anomalous findings that have an impact on both order and parameter (s). In such cases, managing outliers is crucial before beginning analysis to have a deeper comprehension of the data-generating process and to be aware of its implications. Four kinds of outliers have been offered for univariate time series analysis. These are temporary changes, inventive outliers, level shifts, and additive outliers. These four categories of outliers affect an observed time series and associated residual process; see [8, 9, 18].

Various techniques exist for identifying outliers. For a Bayesian approach, look to the work of [14]. For non-Bayesian methods, see the research of [8] and the accompanying references. The autoregressive model adheres to the idea of dependence, where the current observation is linearly dependent on past observations see [6]. Outliers are categorized based on their impact on the model of the series. The first category includes outliers that affect the model via addition, known as additive outliers. The second category includes outliers that affect the model through multiplication, known as multiplicative outliers. Researchers handle the outlier(s) by pursuing two main approaches: first, identifying the outlier(s), and second, studying its consequences in Time Series.

The detection of outliers in the autoregressive model was first examined by [10] using the likelihood criterion when the number of outliers is known. In addition, he made a clear distinction between additive and multiplicative outliers and demonstrated that additive outliers should be given more consideration than multiplicative outliers. In their studies, [17, 9] examined the detection of outliers in an autoregressive integrated moving average (ARIMA) model. They used likelihood ratio criteria to identify both kinds of outliers and also evaluated the parameters associated with them. [4] suggested a test that may be used to discover outliers, but only under certain circumstances, such as when the outliers have a known distribution. In their paper, [5] proposed a technique for detecting the location of an outlier and then determining its size in the context of a non-linear time series.

[2] examined the identification of outliers in the autoregressive time series model using the Bayesian technique and derived the posterior probability. They investigated the responsiveness of models to variations in both previous adjustments and model misrepresentation. [11] created an expanded version of the linear dynamic model

that incorporates a model-based approach. They then used this model to identify instances of abrupt shifts in the series. [18, 3] examined the concepts of additive and multiplicative outliers within the Bayesian framework. They also established a testing process to determine the specific area of these outliers within a subset of a series. For more details, see [16]. [15] introduced a Bayesian method to identify additive outliers in the Poisson integer-valued AR (1) time series model. [13] suggested a method for detecting an anomalous value in a time series that follows a stationary AR (1) model with an intercept term. This method was then expanded to include a linear temporal trend by [12].

Researchers are drawn to the outliers due to their significant influence on many statistical theories and practical applications. The number of observations that may behave as outliers is often relatively small. In recent years, there has been an increasing trend towards the use of individual observations due to the abundance of data and the growing familiarity with advanced computational systems and software.

Thus, the focus of this paper is the Bayesian analysis of the time series model AR(p) impacted by a multiplicative outlier. There are five parts to the paper. The literature is discussed in this section. An overview of autoregressive time series models is provided in Section 1. Previous assumptions form the basis For a Bayesian approach in Sections 2 and 3. The suggested simulation and empirical analysis are shown in Sections 4 and 5. The significance of function and future growth is concluded in the concluding section.

## 2. Model

The topic of multiplicative outliers is covered in this subsection. An observation that occurs regularly as a result of inevitable circumstances and modifies the model's structure is called a multiplicative outlier.

Let us consider that $y_t$ follows an autoregressive model of order p (AR(p))

$$y_t = \theta + y_{t-1}\alpha_1 + y_{t-2}\alpha_2 \dots\dots\dots\dots\dots . y_{t-p}\alpha_p + u_t \qquad t=1, 2, 3\dots\dots\dots T \qquad (1)$$

If the time series model is contaminated by a multiplicative outlier at time point T, then error ($u_t$) is partitioned into two parts as follows: (i) error without outlier (ii) error with outlier. We rewrite the Model (1)

$$\boldsymbol{y_t} = \begin{cases} \theta + Y_{t-1}\alpha + \epsilon_t & t \notin T_1, T_2 \dots\dots\dots T_k \\ \theta + Y_{t-1}\alpha + \sigma\epsilon_t & t \epsilon T_1, T_2 \dots\dots\dots T_k \end{cases}$$

Where θ is the intercept term, $\alpha = (\alpha_1, \alpha_2, \dots\dots\dots . \alpha_p)$ are autoregressive coefficients, $u_t$ is the error term and $Y_{t-1} = y_0 + y_{-1} + y_{-2} \dots\dots\dots\dots\dots . y_{t-p}$ are initial observations.

## 3. Bayesian Estimation

In the Bayesian technique, prior to defining the probability distribution of model parameters based on past experience or some initial confidence, experimentation and data collecting may start. When analyzing time series, this method is often used to derive conclusions on the parameters of the model under consideration. [7] extends Fisher's information measure to cover prior distributions, providing a method for selecting a prior distribution based on the relative value of experimental data and prior information. We consider the following prior distribution

$\theta \sim N(\theta^0, \tau^{-1}V)$

$\alpha \sim MN(u, \tau^{-1}\Sigma^{-1})$

$\sigma^2 \sim IG(a_1, b_1)$

$\tau \sim Gamma(a, b)$

The Likelihood function for the model

$$L(\theta)=\frac{\tau^{\frac{T}{2}}}{(2\pi)^{\frac{T}{2}}}\exp\left[-\frac{\tau}{2}\left\{\sum_{(t\in T_1T_2)}^{T_k}\frac{1}{\sigma^2}(y_t-\theta-Y_{t-1}\alpha)^2+\sum_{(t\notin T_1T_2)}^{T-K}(y_t-\theta-Y_{t-1}\alpha)^2\right\}\right]$$

The posterior distribution is inherited by multiplying the likelihood function with joint prior distribution. The posterior distribution is

$$=\frac{\tau^{\frac{T+p-1}{2}+a}\quad b^a\quad b_1^{a_1}\ |\Sigma|^{\frac{1}{2}}\quad(\sigma^2)^{-a_1-1}}{(2\pi)^{\frac{T+P+1}{2}}V^{\frac{1}{2}}\,\Gamma a_1\,\Gamma a}\ exp\left(-\frac{b_1}{\sigma^2}\right)$$

$$\times\exp\left[-\frac{\tau}{2}\left\{\sum_{(t\in T_1T_2)}^{T_k}\frac{1}{\sigma^2}(y_t-\theta-Y_{t-1}\alpha)^2\right.\right.$$

$$+\sum_{(t\notin T_1T_2)}^{T-K}(y_t-\theta-Y_{t-1}\alpha)^2$$

$$\left.\left.+\left(\frac{1}{v}(\theta-\theta^0)^2+(\alpha-u)'\Sigma(\alpha-u)+2b\right)\right\}\right]\ldots\ldots\ldots\ldots(1)$$

Using a loss function, we were able to determine the best estimator from the posterior distribution under the Bayesian situation. The squared error loss function (SELF) and the absolute loss function (ALF) are two symmetric loss functions that need now be examined. Estimators often need many integrations under these loss functions and solving them analytically is challenging. Numerical and computational methods are therefore used to get around this problem. Two methods in MCMC are used to calculate Bayes estimators: the Gibbs sampler and the Metropolis-Hastings (M-H) algorithm. This is accomplished by using the conditional posterior distribution for each model parameter, which is as follows:

$$\hat{\theta}\sim N\left(B_1 l_1^{-1},\frac{1}{\tau l_1}\right)$$

$$\hat{\alpha}\sim MN\left(B_2 l_2^{-1},\frac{1}{\tau l_2}\right)$$

$$\hat{\tau}\sim Gamma\left(\frac{T+p}{2}+a,C\right)$$

$$\widehat{\sigma^2}\sim IG(a_1,D)$$

Where, $B_1=\sum_{t\in T_1,T_2}^{T_k}\frac{1}{\sigma^2}(y_t-Y_{t-1}\alpha)+\sum_{t\notin T_1,T_2}^{T-k}(y_t-Y_{t-1}\alpha)+\frac{\theta^\circ}{V}$,

$$l_1=\frac{k}{\sigma^2}+T-k+\frac{1}{V}$$

$$B_2=\left[\sum_{t\in T_1,T_2}^{T_k}\frac{1}{\sigma^2}(y_t-\theta)Y_{t-1}+\sum_{t\notin T_1,T_2}^{T-k}(y_t-\theta)Y_{t-1}+\Sigma\mu\right]$$

$$l_2 = \sum_{t \in T_1, T_2}^{T_k} \frac{1}{\sigma^2}(Y'_{t-1}.Y_{t-1}) + \sum_{t \notin T_1, T_2}^{T-k} (Y'_{t-1}.Y_{t-1}) + \Sigma$$

Where,

$$C = \frac{1}{2}\left[\sum_{(t\in T_1 T_2)}^{T_k} \frac{1}{\sigma^2}(y_t - \theta - Y_{t-1}\alpha)^2 + \sum_{(t\notin T_1 T_2)}^{T-K}(y_t - \theta - Y_{t-1}\alpha)^2 + \frac{1}{v}(\theta - \theta^0)^2 + (\alpha - u)'\Sigma(\alpha - u) + 2b\right]$$

Where, $D = \left[\sum_{t \in T_1}^{T_k} \tau(y_t - \theta - Y_{t-1}\alpha)^2 + 2b_1\right]$

Note that the conditional posterior distribution is coming in standard distribution form. Thus, Gibbs sampler is used for the model parameters to simulate the conditional posterior distribution because of explicitly known nature of distributions.

## 4. Simulation

Here we conducted a simulation study to examine the effectiveness of model. In this simulation, we sampled observations from the AR model with T=200, 300, 500 and used the minimum values of the parameters to estimate the parameters of the proposed model

We calculate the conditional posterior for each parameter because we know that the posterior distribution of the model is not closed forms. We then used the Gibbs sampler to generate sample from the conditional distribution of parameter estimates. We produced 10,000 realizations of the parameters of the Markov chain for this purpose using the conditional posterior distribution. Using the generated series, we calculate mean square error (MSE) absolute bias (AB), present in the table respectively.

| **Parameter (True value)** | **Estimated Value T=200** | | | **Estimated value T=300** | | | **Estimated value T=500** | | |
|---|---|---|---|---|---|---|---|---|---|
| | SELF | ALF | ELF | SELF | ALF | ELF | SELF | ALF | ELF |
| θ(4.77) | 4.489609 | 4.485761 | 4.49387 | 4.481445 | 4.482203 | 4.48573 | 4.489509 | 4.490031 | 4.493786 |
| α(0.4) | 0.421787 | 0.420745 | 0.422391 | 0.419542 | 0.418212 | 0.42013 | 0.420513 | 0.419551 | 0.421104 |
| τ(1) | 0.969873 | 0.971614 | 0.973743 | 0.967437 | 0.969294 | 0.971966 | 0.972168 | 0.967136 | 0.977158 |
| σ(1.20) | 3.701777 | 3.203903 | 4.325319 | 3.557217 | 3.110811 | 4.099938 | 3.510317 | 3.109227 | 3.999608 |

| | Estimated Value T=200 | | | | | |
|---|---|---|---|---|---|---|
| | MSE | | | AB | | |
| Parameter | SELF | ALF | ELF | SELF | ALF | ELF |
| θ(4.77) | 0.089095 | 0.092513 | 0.086427 | 0.285034 | 0.287267 | 0.280891 |
| α(0.4) | 0.000589 | 0.000551 | 0.00062 | 0.022221 | 0.021043 | 0.02281 |
| τ(1) | 0.007145 | 0.007758 | 0.006987 | 0.068262 | 0.065941 | 0.066977 |
| σ(1.20) | 8.945202 | 6.134988 | 14.26065 | 2.507786 | 2.012814 | 3.130804 |

| | Estimated Value T=300 | | | | | |
|---|---|---|---|---|---|---|
| | MSE | | | AB | | |
| Parameter | SELF | ALF | ELF | SELF | ALF | ELF |
| θ(4.77) | 0.090626 | 0.092527 | 0.087912 | 0.288555 | 0.287797 | 0.28427 |

| | | | | | | |
|---|---|---|---|---|---|---|
| α(0.4) | 0.000514 | 0.000472 | 0.000542 | 0.020215 | 0.018707 | 0.020757 |
| τ(1) | 0.009056 | 0.010022 | 0.008766 | 0.07408 | 0.074248 | 0.072974 |
| σ(1.20) | 7.787707 | 5.410962 | 12.09363 | 2.357217 | 1.912656 | 2.899938 |

| | Estimated Value T=500 | | | | | |
|---|---|---|---|---|---|---|
| | MSE | | | AB | | |
| Parameter | SELF | ALF | ELF | SELF | ALF | ELF |
| θ(4.77) | 0.087601 | 0.088485 | 0.084999 | 0.283862 | 0.282711 | 0.279609 |
| α(0.4) | 0.000539 | 0.000499 | 0.000566 | 0.021248 | 0.020121 | 0.02177 |
| τ(1) | 0.007928 | 0.008828 | 0.007751 | 0.071921 | 0.07326 | 0.070572 |
| σ(1.20) | 7.239183 | 5.04308 | 10.98477 | 2.310317 | 1.911991 | 2.799608 |

Squaring the difference eliminate negative values for the differences and ensure that the mean square error is always greater than or equal to zero. This property is necessary when we want our model to have small error. In statistics the concept of mean square error is an essential measure used to determine the performance of an estimator.

Bayes estimator outperforms OLS in terms of MSE and ABS. The MSEs and ABs of the estimators decrease as the size of the series is increase for all the parameters. The estimated values of the parameters approach to the actual values when we increase the series size. The absolute loss function (ALF) outperforms other Bayes estimators with respect to Bayesian inference.

## 5. DATA ANALYSIS

The cryptocurrency market has expanded quickly in the past few years, and more people are starting to see it as a useful instrument to supplement conventional stock and futures markets.

These currencies are totally digital and are mostly utilized online, in contrast to cash. Digital currencies may compete with existing online payment options like PayPal and credit/debit cards. Though they are still in their infancy, digital currencies like Solana, Ethereum, Bitcoin, and others might have a significant long-term impact on payment systems and currencies. We examine two popular digital currencies using our model because only two currencies are suitable according to our model and also provide sufficient results. The currency's description and analysis are given below.

### 5. 1 Solana

Solana, a high-performance blockchain, has been the subject of various studies. Launched in 2020, the goal of the public blockchain platform Solana is to improve scalability over existing blockchains without sacrificing security and decentralization. It facilitates the development of decentralized apps (DApps) and smart contracts. Cui (2022) developed VRust, an auto Smated vulnerability detection framework for Solana smart contracts, while Pierro (2022) proposed a tool to verify the ownership of these contracts. These contributions are crucial for ensuring the security and transparency of the Solana blockchain. Solana currently has 894963197 in supply and is trading on 1382 active data. Data from 10 April 2020 to 21 January 2024 were collected, and the following data analysis was done:

| **Parameter** | **OLS** | **SELF** | **ALF** | **ELF** |
|---|---|---|---|---|
| $\theta$ | 0.3519501 | 0.362934 | 0.36302412 | 0.363059 |
| $\alpha$ | 0.9889833 | 0.988866 | 0.98886771 | 0.988866 |
| $\tau$ | 0.0940395 | 0.094316 | 0.09430019 | 0.094385 |
| $\sigma^2$ | 5473 | 220.3066 | 188.095105 | 259.2982 |

### 5.2 Ethereum

Ethereum is a blockchain platform that supports smart contracts, with a key feature being a full-featured programming language for complex business logic (Tikhomirov, 2017). It is designed as a secure, decentralized, and generalized transaction ledger, providing a plurality of resources that can interact with each other (Wood, 2014). The platform's data can be explored using a systematic and high-fidelity framework called Data Ether, which allows for comprehensive and precise data analysis (Chen, 2019). Ethereum has a current supply of 16182147521 and is traded on 2267 active data. We have analyzed ETH-USD data from 9 November 2017 to 23 January 2024, as follows.

| Parameter | OLS | SELF | ALF | ELF |
|---|---|---|---|---|
| $\theta$ | 0.351227 | 6.308601 | 6.312257 | 6.736590 |
| $\alpha$ | 0.953485 | 0.994077 | 0.994078 | 0.994077 |
| $\tau$ | 0.000126 | 0.000251 | 0.000251 | 0.000251 |
| $\sigma^2$ | 250.000006 | 115.577006 | 99.311184 | 99.311184 |

## CONCLUSION

The paper proposes a bayesian approach to handle multiplicative outliers. It takes into consideration the outlier as well as past information and assumptions about the model parameters. Under bayesian set-up the model is justifies by simulation and then application on cryptocurrencies. There are two suitable cryptocurrencies according to our model among existing digital currencies. Using the help of simulation studies, various sizes t=200, 300, 500 used to test the suggested technique against other outlier identification techniques already in use. The results show that the suggestion's approach works with multiplicative outliers and increases the time series model's accuracy. The study provides a helpful approach for performing multiplicative outliers in this particular scenario to improve knowledge of the time series modeling data generation process.

## REFERENCES

[1] Anderson and Theodore W.: *The statistical analysis of time series*, (2011), John Wiley & Sons.

[2] Abraham, B. and Box, G. E.: Bayesian analysis of some outlier problems in time series, *Biometrika* **66(2)** (1979) 229-236.

[3] Barnett, G., Kohn, R. and Sheather, S.: Bayesian estimation of an autoregressive model using Markov chain Monte Carlo, *Journal of Econometrics* **74(2)** (1996) 237–254.

[4] Barnett, V. and Lewis, T.: Outliers in statistical data, **3(1)** (1994), New York: Wiley.

[5] Battaglia, F. and Orfei, L.: Outlier detection and estimation in nonlinear time series, *Journal of Time Series Analysis* **26(1)** (2005) 107-121.

[6] Box and Jenkins: *Engineering and statistics handbook-ARMA Models*, INSEAD (1970) 64–65.

[7] Canfield, R. V. and Teed, J. C.: Selecting the prior distribution in Bayesian estimation, *IEEE Transactions on Reliability* **26(4)** (1977) 283-285.

[8] Chen, C. and Liu, L. M.: Joint estimation of model parameters and outlier effects in time series, *Journal of the American Statistical Association* **88(421)** (1993) 284-297.

[9] Chang, I., Tiao, G.C. and Chen, C.: Estimation of time series parameters in the presence of outliers, *Technimetrics* **3** (1988) 193–204.

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

 

[11] Gordon, K. and Smith, A. F. M.: Modeling and monitoring biomedical time series, *Journal of the American Statistical Association* **85(410)** (1990) 328-337.

[12] Kumar, J. and Shukla, A.: Identification of Additive Outlier in Stationary AR (1): A Bayesian Approach, *International Journal of Intelligent Technologies and Applied Statistics* **8(3)** (2015) 275-290.

[13] Kumar, J., Shukla, A. and Tiwari, N.: Bayesian analysis of a stationary AR (1) model and outlier, *Electronic Journal of Applied Statistical Analysis* **7(1)** (2014) 81-93.

[14] McCulloch, R. E. and Tsay, R. S.: Bayesian analysis of autoregressive time series via the Gibbs sampler, *Journal of Time Series Analysis* **15(2)** (1994) 235-250.

[15] Pierro, G. A. and Amoordon, A.: A Tool to check the Ownership of Solana's Smart Contracts. *In 2022 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER)* (2022) 1197-1202.

[16] Siegmund, D., Yakir, B. and Zhang, N.R.: Detecting simultaneous variant intervals in aligned sequences, *The Annals of Applied Statistics* (2011) 645–668.

[17] Tsay, R. S.: Time series model specification in the presence of outliers, *Journal of the American Statistical Association* **81(393)** (1986) 132-141.

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