



vol. 17 / 2023



## **The 7th International Conference on Science Technology**

organized by  
Faculty of Social Science and  
Law Universitas Negeri Manado and  
Consortium of International Conference  
on Science and Technology

# **The Innovation Breakthrough in Digital and Disruptive Era**

# Comparison of Forecasting Accuracy Using the Short Moving Average (SMA) Method Using Boxplot Outlier Filtering and Not Using Outlier Filtering for Data that has a high level of variation

Akhmad Tajuddin Tholaby MS <sup>1\*</sup>

<sup>1</sup>Universitas KH. Bahaudin Mudhari Madura, Sumenep, East Java, Indonesia

<sup>2</sup>Universitas Trunojoyo Madura, Bangkalan, East Java, Indonesia

**Abstrak.** The Short Moving Average (SMA) forecasting method is one of the most widely used forecasting methods, especially for processing data with a high level of variation and is not linear with time. However, opportunities to develop and improve forecasting performance using the SMA method are still wide open. The performance of a forecasting method can be seen from the distribution of errors. SMA does not see and does not sort the type of input data that will be processed into a forecast value, whether the input data has small or large variations, or has outlier data. If the input data has an outlier, then that outlier can make the forecasting performance not good. One of the efforts to improve SMA forecasting performance is by filtering outlier data. In this study, a comparison was made of the forecasting results for SMA using outlier filtering with the forecasting results for SMA not using outlier filtering. The next step is to compare the error values, namely those that produce the smallest Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE) values. From the results of the study it can be seen that the performance of SMA using the Boxplot filtering method gives better forecasting results than those without using outlier filtering.

**Keywords.** Short Moving Average, outliers, forecasting, Boxplot

## 1 Introduction

Every company that produces products both in the form of goods and services definitely wants the company to run well and continue to progress. To achieve this condition, all companies without exception must carry out three main management cycles on an ongoing basis, namely planning, implementing and evaluating which later the resulting evaluation results will become the basic input for the next stage of planning.

One form of this evaluation is to calculate the forecasting demand for products produced by the company. There are many forecasting methods that can be applied. One of them is the Short Moving Average (SMA) forecasting method. The SMA forecasting method is one of the most widely used forecasting methods, especially for processing data with a high level of variation and is not linear with respect to time. However, opportunities to develop and improve forecasting performance using the SMA method are still wide open. The performance of a forecasting method can be seen from the distribution of errors. SMA does not see and does not sort out what type of input data will be processed into a forecast value, whether the input data has small or large variations, or has outlier data, all of which is processed to become forecast values. If the input data has an outlier, then

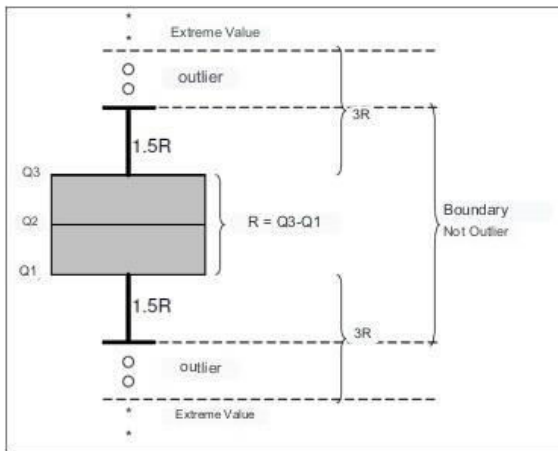
that outlier can damage the distribution of errors so that the forecasting performance is not good.

One of the efforts to improve SMA forecasting performance is by filtering outlier data. Outliers are data that do not describe the characteristics of the data as a whole, or in other words outliers are data whose value is far from the characteristics of the data as a whole.

One of the studies that did this was Chao Zhao and Jinyan Yang [1] who filtered and identified outliers using the boxplot method. However, it turns out that the resulting outlier deviates from the definition of the outlier itself. So it is very necessary to carry out deeper observations and calculations to improve these conditions.

In statistics, it is stated that there are several outlier identification methods, one of which is the Boxplot method [2]. This method can identify any outliers that can damage the distribution of errors in SMA forecasting performance. This method uses quartile values and ranges. Quartiles 1, 2, and 3 will divide a data sequence into four parts. The range (IQR, Interquartile Range) is defined as the difference between the 1st quartile and the 3rd quartile, or  $IQR = Q3 - Q1$ . Outlier data can be determined, namely values less than  $1.5 \cdot IQR$  for the 1st quartile and values more than  $1.5 \cdot IQR$  for the 3rd quartile. One of the studies using this method is R. Dawson [2] and

Soemartini [3]. - together perform filtering of outliers on the data to determine the impact of the presence and absence of outliers in data analysis (regression analysis).



**Fig. 1.** Outlier identification scheme using IQR or Boxplot

The data that is filtered using outlier filtering will be calculated for SMA forecasting. After that, the error distribution of MSE and MAPE will be calculated which will then be compared with the calculated values without using outlier filtering. The best forecasting value is seen from the smallest MSE and MAPE error values.

## 2 Literature Review

### 2.1 LQ45 index

The data used in this study is BEI stock price data. The BEI stock used in this thesis research is the LQ45 index, because the LQ45 index is the market capitalization value of the 45 most liquid stocks and has the largest capitalization value. LQ45 also has high data variation.

In summary there are three stages that must be passed in designing a forecasting method, namely:

- Perform analysis on past data. This step aims to get an overview of the pattern of the data in question.
- Select the method to be used. There are various methods available according to their needs. Different methods will produce different prediction systems for the same data. In general, it can be said that a successful method is one that produces the smallest error between the predicted results and the reality that occurs.

The process of transforming past data using the chosen method. If necessary, changes are made as needed.

### 2.2 Single Moving Average Method

Determining the forecast with the single moving average method is quite easy to do. If you apply the 4-month moving average, the forecast for May is calculated at the average of the previous 4 months, namely January, February, March, April. The mathematical equation of this technique is:

$$F_{t+1} = \frac{x_t + x_{t-1} + x_{t-2} + \dots + x_{t-n+1}}{N} \quad (1)$$

Information:

- $F_{t+1}$  : Forecast for period t+1
- $x_t$  : The real value of the t period
- $N$  : Moving average timeframe

The Single Moving Averages method has special characteristics.

- To determine forecasts for future periods requires historical data over a certain period of time.
- The longer the period of the moving average, the more visible the smoothing effect in the forecast or the smoother the moving average will be. This means that on a moving average with a longer timeframe, the difference between the smallest forecast and the largest forecast is smaller.

### 2.3 Forecasting Error Measures

Because demand is affected by many factors, and future value cannot be known with certainty, it makes no sense to get forecasts that are precise every time. The calculation of the average error made by the forecasting model over time is a measure of how precise the forecast is. Two commonly used error measurements are the Mean Absolute Percentage Error (MAPE) and the Mean Squared Error (MSE). The first thing is used because it is very useful to determine the tracking signal. MAD is the average error value in forecasting using the absolute value of the Forecasting Method, through the following stages: [5]

If  $X_t$  is the actual data for period t and  $F_t$  is the forecast (fitted value) for the same period, then the error is defined as:

$$e_t = D_{(t)} - F_{(t)} \quad (2)$$

Consideration of acceptance of a forecasting method is through the following criteria:

- Percentage error (PE)

$$PE = \left[ \frac{D_t - F_t}{D_t} \right] \times 100\% \quad (3)$$

- Mean Squared Error (MSE)

$$MSE = \frac{\sum_{t=1}^m |D_t - F_t|^2}{m} \quad (4)$$

- Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{\sum_{t=1}^m |PE_t|}{m} \quad (5)$$

Information:

- $D_t$  : Actual Price in Period t
- $F_{(t)}$  : Price Forecasting in Period t
- $m$  : Number of Forecasting Periods

### 3 Methodology

The steps in this research are described in the following flowchart.

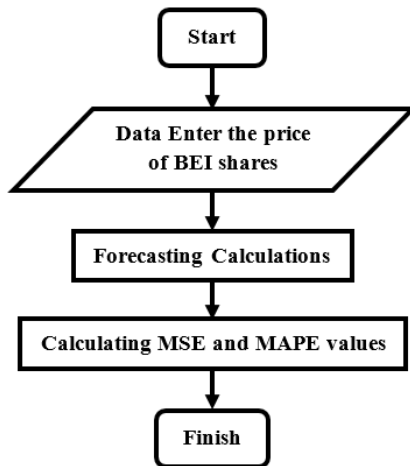


Fig. 2. Forecasting calculations without outlier filtering

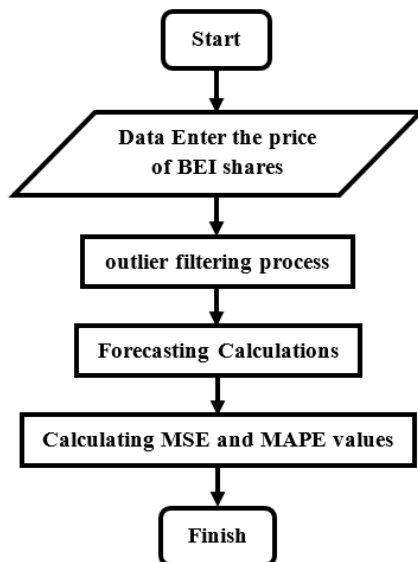


Fig. 3. Forecasting calculations with outlier filtering

After that, a comparison of the error values was carried out between those using outlier filtering and those not using outlier filtering.

### 4 Results and Discussion

From the tests that have been carried out in this study, the following results were obtained:

Table 1. Forecasting Value Without Filterization.

Period	Actual Value	Forecasting Value Without Filterization		
		Forecasting	MSE	MAPE
51	629,48	620,71788	6,488524	1,033498169
52	619,75	619,94844		
53	621,21	618,88258		
54	628,2	617,75248		
55	627,7	616,99488		

Table 2. Boxplot filtered data.

Period	Actual Value	Forecasting Value with Filterization		
		Forecasting	MSE	MAPE
51	629,48	616,3282174	8,595043528	1,370610622
52	619,75	616,6080213		
53	621,21	616,6735208		
54	628,2	616,7661429		
55	627,7	616,99488		

The experimental results show that the forecasting value without outlier filtering is better than forecasting using outlier filtering because it has better MSE and MAPE values.

### 5 Conclusion

The conclusion obtained in this study is that the boxplot outlier filter method is not good for improving forecasting performance for data that has high variation.

### 6 References

- [1] Zhao, Chao & Yang, Jinyan (2011). "A Robust Skewed Boxplot for Detecting Outliers in Rainfall Observations in Real-Time Flood Forecasting," *Advances in Meteorology* Volume 2019, Article ID 1795673, 7 pages
- [2] Dawson, R. (2011), "How Significant Is A Boxplot Outlier?". *Journal of Statistics Education*, Volume 19, Number 2(2011)
- [3] Soemartini (2007), "Pencilan (outlier)" Universitas Padjajaran, Jatinangor
- [4] Kr.Das, Mintu & Gogoi, Bipin (2015), "Usage of Graphical Displays to Detect Outlying Observations in

Linear Regression”. Indian Journal of Applied Research. Volume: 5, Issue: 5, May 2015, ISSN - 2249-555X

- [5] Atkinson A.C (1981). Two Graphical Display for Outlying and Influential Observation in Regression. *Technometrics*
- [6] Awallun, NK.; & Delyana, S (2012). ” Analisis Residual dengan Leverage Values, DFFITS, dan Cook’s Distance”. Jakarta: Sekolah Tinggi Ilmu Statistik Jakarta hal: 3-10.
- [7] Arumugam, P., & Anithakumari, V. (2013). “Fuzzy Time Series Method for Forecasting Taiwan Export Data”. *International Journal of Engineering Trends and Technology (IJETT) - Volume4 Issue8- August 2013*. India: Manonmaniam Sundaranar University.