Measuring the accuracy of Short Moving Average (SMA) forecasting performance with 3 medium data ranges

Akhmad Tajuddin Tholaby MS[1], V.Shyamala Susan[2]

[1]Sains dan Teknologi, Univeristas Bahaudin Mudhary, Sumenep, Indonesia.
[2]Computer Science Department in A.P.C. Mahalaxmi College, Thoothukudi, TamilNadu, India
(E-mail: akhmad.tajuddin@unibamadura.ac.id)

Correspondence:
Name, Department, University, City, Post Code, Country.
Email: xxx@company.ac.id
 (Times New Roman, size 10)

Abstract
The Short Moving Average (SMA) forecasting method is one of the most widely used forecasting methods, especially for processing data with low and high levels of variation and is not linear over time. However, the opportunity to develop and improve forecasting performance using the SMA method is still wide open. The performance of a forecasting method can be seen from the error distribution. One of the steps to calculate the SMA forecasting value is to determine the forecasting range. The length or shortness of the forecasting range can determine the accuracy of the SMA forecasting value. Therefore, in this study a comparison of SMA forecasting results was carried out using 3 different medium forecast ranges. 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 research results, it can be seen that the SMA forecasting performance using the shortest range has the smallest error value when compared to SMA forecasting using other forecasting ranges.

Keywords: SMA forecasting, Racial Medium Range, MSE, MAPE

Introduction

Every company that produces products in the form of goods or services definitely wants their company to run well and continue to progress. To achieve this condition, all companies without exception must carry out three main management cycles continuously, namely planning, implementation and evaluation, where the resulting evaluation results will become basic input for the next planning stage.

One form of evaluation is by calculating the estimated demand for the 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 low and high levels of variation and is not linear with time. However, the opportunity to develop and improve forecasting performance using the SMA method is still wide open. The performance of a forecasting method can be seen from the error distribution.

One of the steps to calculate the SMA forecasting value is to determine the forecasting range. The length or shortness of the forecasting range can determine the accuracy of the SMA forecasting value. Therefore, in this study a comparison of SMA forecasting results was carried out using 3 different medium forecast ranges. 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 research results, it can be seen that the SMA forecasting performance using the shortest range has the smallest error value when compared to SMA forecasting using other forecasting ranges.

Literature Review

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 low data variation. The data used in this test is LQ45 data from June 24 2011 to September 20 2011. As for the LQ45 data from June 24 2011 to August 12, 2011 are as follows:

Table 1. LQ45 data for the period June 24 2011 to August 12, 2011.

Date	Cloce
24/06/2011	681,495
27/06/2011	675,259
28/06/2011	678,082
30/06/2011	690,646
01/07/2011	698,858
04/07/2011	704,244

05/07/2011	696,897
06/07/2011	692,587
07/07/2011	696,514
08/07/2011	710,896
11/07/2011	708,959
12/07/2011	698,049
13/07/2011	704,851
14/07/2011	707,682
15/07/2011	713,178
18/07/2011	714,623
19/07/2011	710,207
20/07/2011	714,045
21/07/2011	718,035
22/07/2011	725,939
25/07/2011	721,087
26/07/2011	731,038
27/07/2011	740,992
28/07/2011	733,057
29/07/2011	729,836
01/08/2011	742,502
02/08/2011	740,636
03/08/2011	732,744
04/08/2011	730,104
05/08/2011	693,293
08/08/2011	681,946
09/08/2011	660,514
10/08/2011	685,247
11/08/2011	686,294
12/08/2011	689,603

The data ranges that will be used for SMA forecasting calculations in this experiment are 10, 15 and 20.

LQ45 data

Several selection criteria to determine whether an issuer can be included in the calculation of the LQ45 index are:

The first criterion is:
- Being in the TOP 95% of the total annual average transaction value of shares in the regular market.
- Be in the TOP 90% of the annual average market capitalization.

The second criterion is:
- Represents the highest order representing its sector in the JSX industry classification according to its market capitalization value.
- Is the highest order based on transaction frequency.

The LQ45 index only consists of 45 stocks that have been selected through various selection criteria, so that it will consist of stocks with high liquidity and market capitalization. Stocks on the LQ45 index must meet the following criteria and pass the main selection:
- Ranked in the top 60 of total stock transactions in the regular market (average transaction value over the last 12 months).
- Ranking by market cap (average market cap over the last 12 months)
- Has been listed on the BEJ for at least 3 months
- The financial situation of the company and its growth prospects, the frequency and number of trading days of regular market transactions.

The stocks included in the LQ45 are continuously monitored and a review will be held every six months (early February and August). If there are shares that do not meet the criteria, they will be replaced with other shares that meet the requirements. The selection of LQ45 shares must be reasonable, therefore the JSX has an advisory committee consisting of experts at BAPEPAM, universities and professionals in the capital market sector.

Factors that play a role in the movement of the LQ45 Index, namely:
- The SBI interest rate as a benchmark for investment portfolios in the Indonesian financial market,
- The level of investor tolerance for risk, and
- Index moving stocks (index mover stocks) which incidentally are large market cap stocks on the JSX.

The factors that influence the increase in the LQ45 Index are:
- Strengthening global and regional markets following the decline in world crude oil prices, and
- The strengthening of the rupiah exchange rate was able to lift the LQ 45 index to the positive zone.

The purpose of the LQ 45 index is to complement the JCI and in particular to provide an objective and reliable tool for financial analysts, investment managers, investors and other capital market observers in monitoring price movements of actively traded stocks.

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.

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.

Research Methods
Single Moving Average Method

Determining the forecast using the single moving average method is quite easy to do. If using a moving average with a range of 20, 25 and 30 periods, the forecasting value is calculated from the average value based on the number of each range. The mathematical equation of this technique is:

$$F_{t+1} = \frac{x_t + x_{t-1} + x_{t-2} \ldots x_{t-n+1}}{N} \tag{1}$$

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.

Sentiment Classification

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:

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

Consideration of acceptance of a forecasting method is through the following criteria:
- Percentage error (PE)

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

- Mean Squared Error (MSE)

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

- Mean Absolute Percentage Error (MAPE)

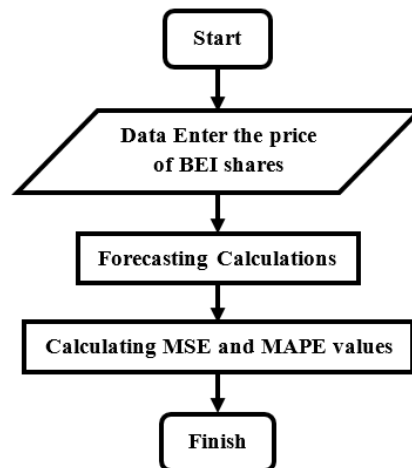$$MAPE = \frac{\sum_{t=1}^{m}|PE_t|}{m} \qquad (5)$$

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

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

Fig. 2. Forecasting calculations without outlier filtering

Results and Discussion

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

Table 2. Forecasting with a range of 20, 25 and 30.

Date	Cloce	Forecasting with		
		Range 20	Range 25	Range 30
24/06/2011	681,495			
27/06/2011	675,259			
28/06/2011	678,082			
30/06/2011	690,646			
01/07/2011	698,858			
04/07/2011	704,244			
05/07/2011	696,897			
06/07/2011	692,587			
07/07/2011	696,514			
08/07/2011	710,896			
11/07/2011	708,959			
12/07/2011	698,049			
13/07/2011	704,851			
14/07/2011	707,682			

Date	Close	R20	R25	R30
15/07/2011	713,178			
18/07/2011	714,623			
19/07/2011	710,207			
20/07/2011	714,045			
21/07/2011	718,035			
22/07/2011	725,939			
25/07/2011	721,087			
26/07/2011	731,038			
27/07/2011	740,992			
28/07/2011	733,057			
29/07/2011	729,836			
01/08/2011	742,502			
02/08/2011	740,636			
03/08/2011	732,744			
04/08/2011	730,104			
05/08/2011	693,293	**R20**	**R25**	**R30**
08/08/2011	681,946	720,54285	716,4798	711,21117
09/08/2011	660,514	719,1922	715,58788	711,2262
10/08/2011	685,247	717,31545	714,13256	710,7347
11/08/2011	686,294	716,33525	713,83896	710,97353
12/08/2011	689,603	715,26585	713,43016	710,82847

Table 3. MSE and MAPE values for SMA forecasting with a range of 20.

Date	Close	Forecasting	e	|e|	MSE	PE (%)	MAPE
08/08/2011	681,946	720,54285	-38,5969	38,59685	37,00952	5,65981	5,464415
09/08/2011	660,514	719,1922	-58,6782	58,6782		8,883718	
10/08/2011	685,247	717,31545	-32,0685	32,06845		4,679838	
11/08/2011	686,294	716,33525	-30,0413	30,04125		4,377315	
12/08/2011	689,603	715,26585	-25,6629	25,66285		3,721395	

Table 4. MSE and MAPE values for SMA forecasting with a range of 25.

Date	Close	Forecasting	e	|e|	MSE	PE (%)	MAPE
08/08/2011	681,946	716,4798	-34,5338	34,5338	33,97307	5,064008	5,017234
09/08/2011	660,514	715,58788	-55,0739	55,07388		8,338034	
10/08/2011	685,247	714,13256	-28,8856	28,88556		4,21535	
11/08/2011	686,294	713,83896	-27,545	27,54496		4,01358	
12/08/2011	689,603	713,43016	-23,8272	23,82716		3,4552	

Table 5. MSE and MAPE values for SMA forecasting with a range of 30.

Date	Close	Forecasting	e	|e|	MSE	PE (%)	MAPE
08/08/2011	681,946	711,211167	-29,2652	29,26517	30,27401	4,29142	4,472516
09/08/2011	660,514	711,2262	-50,7122	50,7122		7,677687	
10/08/2011	685,247	710,7347	-25,4877	25,4877		3,719491	
11/08/2011	686,294	710,973533	-24,6795	24,67953		3,596058	
12/08/2011	689,603	710,828467	-21,2255	21,22547		3,077926	

The experimental results show that the SMA forecasting value with range 5 has the best level of accuracy when compared to the others because it has the smallest MSE and MAPE error values.

Conclusion

The conclusion obtained in this research is that the accuracy value of SMA forecasting values will become more accurate if calculated with a smaller range.

References

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

Dawson, R. (2011), "How Significant Is A Boxplot Outlier?". Journal of Statistics Education, Volume 19, Number 2(2011)

Soemartini (2007), "Pencilan (outlier)" Universitas Padjajaran, Jatinangor

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

Atkinson A.C (1981). Two Graphical Display for Outlying and Influential Observation in Regression. Technometricss

Awallun, NK.; & Delyana, S (2012). " Analisis Residual dengan Leverage Values, DfFITS, dan Cook's Distance". Jakarta: Sekolah Tinggi Ilmu Statistik Jakarta hal: 3-10.

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.

Allahviranloo, T., Lotfi, F. M. and Firozja, A. (December 2007). Fuzzy Efficiency Measure with Fuzzy Production Possibility Set, AAM, Vol. 2, No. 2, pp. 152 – 166.

Chen, S. M. and Chung, N. Yi (2006). Forecasting Enrollments of Students by Using Fuzzy Time Series and Genetic Algorithms, Information and Management Sciences, No.3, Vol.17, pp. 1-17.

Rousseeuw, P.J. (1991), "Diagnostic Plot for Regression Outlier and Leverage Point," Statistical Software Newsletter, 127-129

Imon, Rahmatullah. 2007, "Robust Regression," Postgraduate Lecture Series 3. Institute of Mathematical Sciences University of Malaya.

Atkinson A.C (1981). Two Graphical Display for Outlying and Influential Observation in Regression. Technometricss

Dan Liao, Dan & Valliant, Richard (2012). "Variance inflation factors in the analysis of complex survey data". Component of Statistics Canada Catalogue no. 12-001-X Business Survey Methods Division

Yolcu, Ufuk (2012). *The Forecasting of Istanbul Stock Market with a High Order Multivariate Fuzzy Time Series Forecasting Model.* TJFS: Turkish Journal of Fuzzy Systems (eISSN: 1309–1190). An Official Journal of Turkish Fuzzy Systems Association. Vol.3, No.2, pp. 118-135, 2012. Turkey: University of Giresun, Faculty of Arts and Science, Department of Statistics

H Amjad, Usman.; A. Jilani, Tahseen., & Yasmeen, Farah. (2012). *A Two Phase Algorithm for Fuzzy Time Series Forecasting using Genetic Algorithm and Particle Swarm Optimization Techniques.* International Journal of Computer Applications (0975 – 8887) Volume 55– No.16, October 2012. University of Karachi

Alpaslan, Faruk.; Cagcag, Ozge.; Aladag, C.H.; Yolcu, U.; and Egrioglu, E.. (2012). *" A Novel Seasonal Fuzzy Time Series Method"* Hacettepe Journal of Mathematics and Statistics Volume 41 (3) (2012), 375 – 385. Turkey: Statistics Department, Ondokuz Mayis University, Samsun.

Suresh, S., & Senthamarai Kannan, K. (2011). *" Identifying Outliers in Fuzzy Time Series".* Journal of Modern Applied Statistical Methods. India:. ManonmaniamSundaranar University, Tirunelveli

Pratama, Jaka.; Awallun, NK.; & Delyana, S (2012). *" Analisis Residual dengan Leverage Values, DfFITS, dan Cook's Distance".* Jakarta: Sekolah Tinggi Ilmu Statistik Jakarta hal: 3-10.