Engineering and Applied Science Research October – December 2018;45(4):308-311 Research Article

EASR

Engineering and Applied Science Research

https://www.tci-thaijo.org/index.php/easr/index

Published by the Faculty of Engineering, KhonKaen University, Thailand

Forecasting techniques based on absolute difference for small dataset to predict the SET Index in Thailand

Watcharin Klongdee and Mathee Pongkitwitoon*

Department of Mathematics, Faculty of Science, KhonKaen University, Khon Kaen 40002, Thailand

Received 15 August 2017
Accepted 22 November 2017

Abstract

This research aims to use a simple statistical method as a forecasting model with a small dataset. Absolute difference methods, average absolute difference and minimum absolute difference, were used to adjust the dataset, i.e., the SET Index, before fitting using the following two forecasting models, an autoregressive forecasting model and a simple moving average forecasting model. Then we compared the quality of predictions using the mean square error and the mean absolute difference. These showed that the mean square error of the average absolute difference filtering method were 15.13%, 15.17% and 7.31% less than the original dataset for a one-period autoregressive forecasting model, a two-period autoregressive forecasting model and a three-period simple moving average forecasting model, respectively. The mean absolute differences were 8.36% , 8.39% and 4.10% less than the original dataset for a one-period autoregressive forecasting model, a two-period autoregressive forecasting model and a three-period simple moving average forecasting model, respectively. The mean square error of the minimum absolute difference filtering method were 66.02%, 58.94% and 16.33% less than the original dataset for a one-period autoregressive forecasting model, a two-period autoregressive forecasting model and a three-period simple moving average forecasting model, respectively. The mean absolute differences were 39.60% , 33.81% and 9.37% less than the original dataset for a one-period autoregressive forecasting model, a two-period autoregressive forecasting model and a three-period simple moving average forecasting model, respectively.

Keywords: Average absolute difference, Forecasting techniques, Minimum absolute difference, Simple statistical method, Small dataset

1. Introduction

Time series forecasting is one of many statistical techniques used to analyze and predict the behavior of a dynamic dataset. Researchers have developed new and improved tools to make accurate predictions, but they are too complicated for use by ordinary people and most require high computational effort. These include nonlinear forecasting, Bayesian techniques or stochastic approaches. Most forecasting techniques require large datasets. Often it is not possible to collect enough data to meet the minimum requirements to use specific forecasting methods [1-3]. We are interested in methods that can reduce noise within data to a minimum without the need for large datasets. So, our aim is to find a simple method that is easy and efficient for use with small datasets.

2. Data and methods

We used a simple statistical method, absolute difference (AD), which can easily calculate and simply manipulate small datasets. We introduced an average absolute difference (avgAD) and a minimum absolute difference (minAD) for this purpose, then fitted forecasting models using autoregressive models, AR(1) and AR(2), and a three-period simple moving average model for forecasting, MA(3). We compared the results using mean square error (MSE) and mean absolute deviation (MAD) [4-5]. These approaches were easy statistical tools to analyze and manipulate a small dataset for forecasting.

2.1 Dataset

We used a monthly SET index dataset (source: https://www.set.or.th) from December 2015 to February 2017 for this research and partition it into two parts. The first part was used for fitting the models, from December 2015 to November 2016.Thesecond section, from December 2016 to February 2017, was used to evaluate the predictions from time series analysis.

2.2 Purposed Algorithms

We computed the avgAD and minAD using equations

$$\mathrm{AD}(x_i, x_j) = |x_i - x_j|, \text{ for } i \neq j \tag{1}$$

*Corresponding author. Tel.: +66816610644
Email address: matpon@kku.ac.th
doi: 10.14456/easr.2018.41

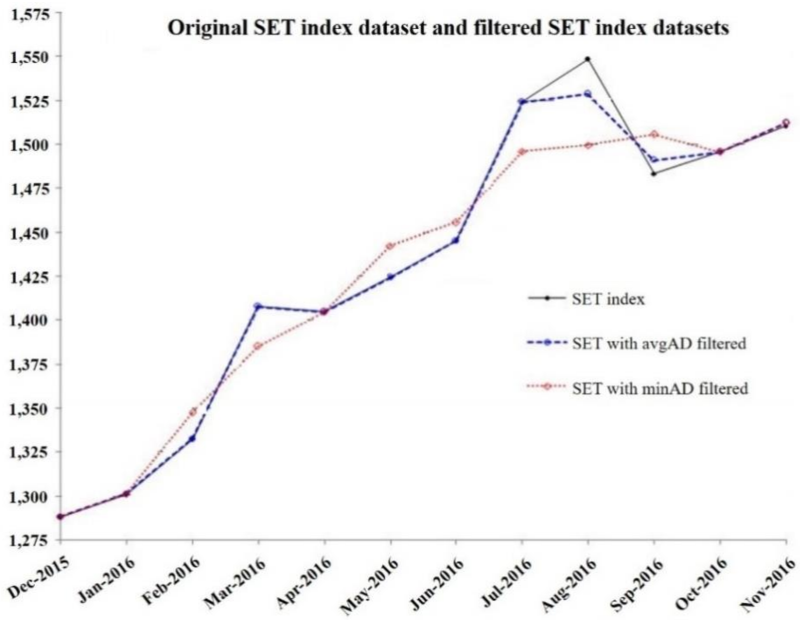

Figure 1 Original SET index and filtered SET index datasets

$$\text{avgAD} = \frac{1}{m}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\text{AD}(x_i, x_j), \qquad n > 1 \tag{2}$$

where

$$m = \sum_{i=1}^{n-1} i$$

and

$$\text{minAD} = \min_{\substack{i=1,2,\ldots,n-1\\ j=i+1,i+2,\ldots,n}}\{\text{AD}(x_i, x_j)\}\sum_{i=1}^{n-1} i \tag{3}$$

Let X be portion of the SET index dataset from December 2015 to November 2016, v_k be a value within a subset s_k,

$$X = \{x_1, x_2, \ldots, x_{12}\} \tag{4}$$

where x_1 be a SET index of December 2015, x_2 be a SET index of January 2016, and so on, and,

$$s_k = \{x_{k-2}, x_{k-1}, v_k, x_{k+1}, x_{k+2}\} \tag{5}$$

where $v_k = x_k$ and x_{k-2}, x_{k-1}, v_k, x_{k+1}, $x_{k+2} \in X$, if $x_w \notin X$, then x_w be null for $w=k-2$, $k-1$, $k+1$, $k+2$.

We introduce two algorithms that use avgAD and minAD as follows:

Algorithm I

1) Compute avgAD for $i \neq k$ and $j \neq k$.
2) Compute average(x_{k-2}, x_{k-1}, x_{k+1}, x_{k+2}) that is denoted by sM.
3) Compare v_k to these criteria
   a) If v_k is not satisfied
      $v_k <$ sM – avgAD
      then go to b)
      else set $v_k =$ sM – avgAD
   b) If v_k is satisfied
      $v_k >$ sM + avgAD
      then set $v_k =$ sM + avgAD.

Algorithm II

1) Compute minAD for $i \neq k$ and $j \neq k$.
2) Compute average(x_{k-2}, x_{k-1}, x_{k+1}, x_{k+2}) that is denoted by sM.
3) Compare v_k to these criteria
   a. If v_k is not satisfied
      $v_k <$ sM – minAD
      then go to b)
      else set $v_k =$ sM – minAD
   b. If v_k is satisfied
      $v_k >$ sM + minAD
      then set $v_k =$ sM + minAD.

Then, the algorithms were used with the original dataset, after which two new filtered datasets resulted. Figure 1 depicts the original dataset and the two new filtered datasets. We fitted the AR(1), AR(2) and MA(3) forecasting models using the original and two new datasets, filtered using equations 1 and 2. The AR(1), AR(2) and MA(3) forecasting models can be calculated using the respective equations:

$$\hat{x}_{t+1} = \beta_0 + \beta_1 x_t + \varepsilon_t \tag{6}$$

$$\hat{x}_{t+1} = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \varepsilon_t \tag{7}$$

and

$$\hat{x}_{t+1} = \frac{x_t + x_{t-1} + x_{t-2}}{3} \tag{8}$$

2.3 Test Results

We use the forecasting models to predict SET indices for three periods, the months of December 2016 to February 2017 and compared the quality of predictions using MSE and MAD comparisons between original dataset and two new datasets. The MSE and MAD were calculated using equations 9 and 10, respectively:

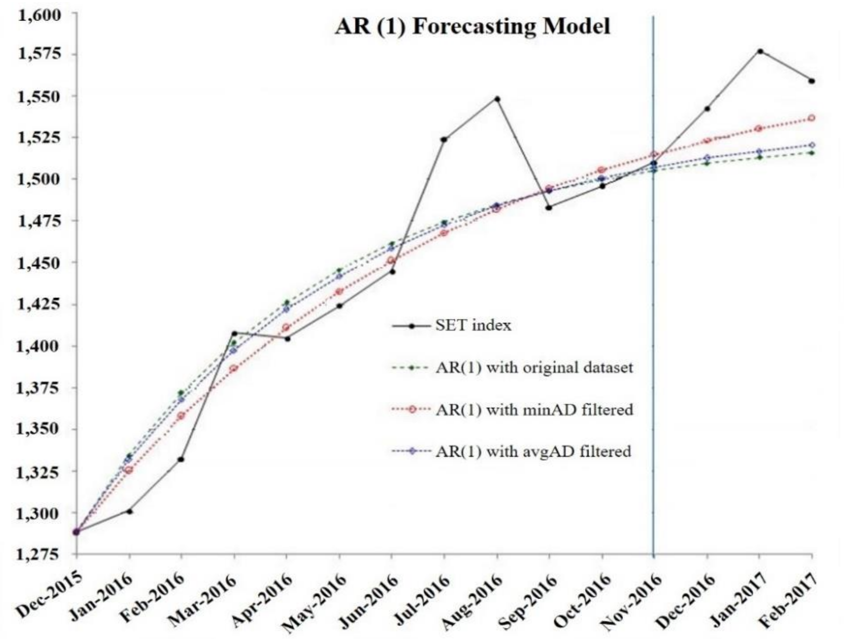

Figure 2 AR(1) forecasting model

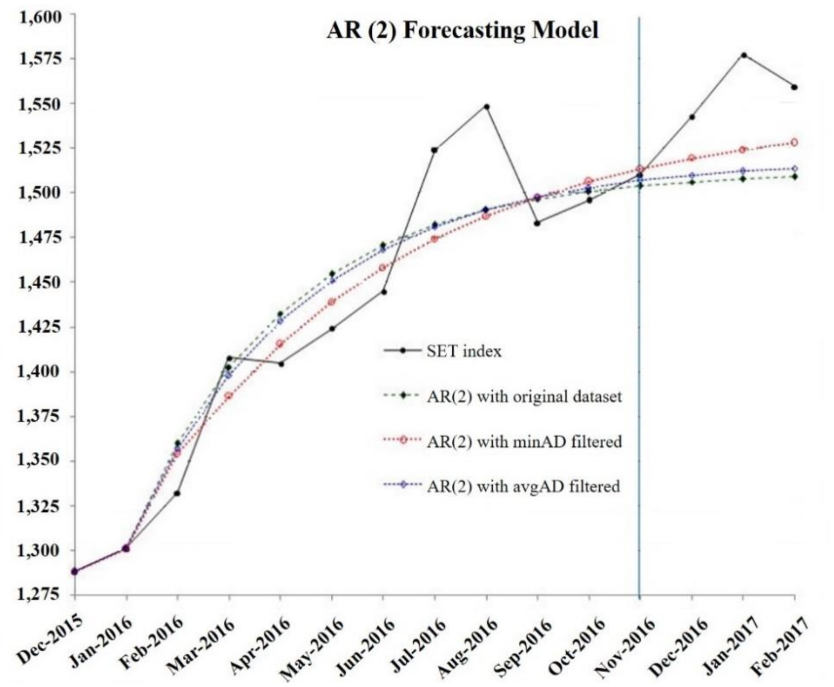

Figure 3 AR(2) forecasting model

$$\text{MSE} = \sum_{i=1}^{n} \frac{(x_i - \hat{x}_i)^2}{n} \tag{9}$$

where x_i is the real value and $\hat{x}_i$ is the predicted value, and

$$\text{MAD} = \sum_{i=1}^{n} \frac{|x_i - \hat{x}_i|}{n} \; . \tag{10}$$

3. Results and discussion

Figures 2, 3 and 4 show that the prediction from December 2016 to February 2017 of AR(1), AR(2) and MA(3) for the two new datasets were better than for the original dataset.

Table 1 shows that the original SET index dataset has larger MSE and MAD values than the SET index datasets using avgAD and minAD to adjust the data before fitting using the models. This research reveals that the quality of prediction for the dataset that used minAD to filter was better than for other datasets for all forecasting methods.

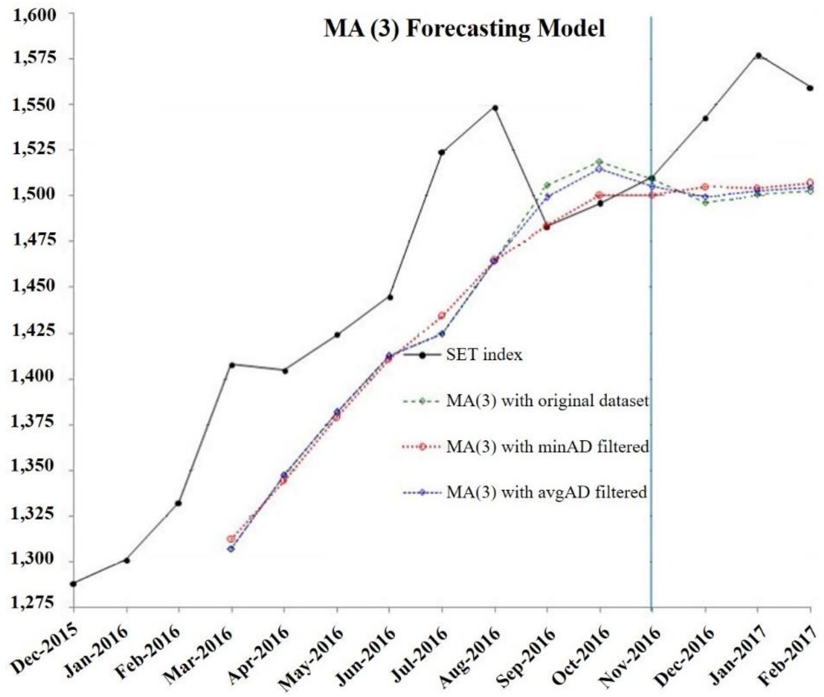

Figure 4 MA(3) forecasting model

Table 1 Quality of predictions

	A	B	C	D	E	F	G
	Original Dataset	**Dataset using avgAD filter**	**Dataset using minAD filter**	**Absolute (B–A)**	**Absolute (C–A)**	**(D/A) ×100%**	**(E/A) ×100%**
AR(1) forecasting model							
MSE	2380.8364	2020.6019	1046.7635	360.2345	1334.0729	15.13	66.02
MAD	47.0788	43.1448	29.9918	3.9340	17.0870	8.36	39.60
AR(2) forecasting model							
MSE	2928.4259	2481.3425	1465.8797	447.0834	1462.5462	15.27	58.94
MAD	52.4252	48.0270	36.1881	4.3982	16.2371	8.39	33.81
MA(3) forecasting model							
MSE	3760.7915	3485.9548	3191.5773	274.8367	569.2142	7.31	16.33
MAD	60.0552	57.5917	54.6613	2.4635	5.3939	4.10	9.37

The MSE values of the avgAD filtering method were 15.13%, 15.17% and 7.31% less than the original dataset for AR(1), AR(2) and MA(3) forecasting models, respectively. The MAD values were 8.36%, 8.39% and 4.10% less than the original dataset for AR(1), AR(2) and MA(3) forecasting models, respectively. Finally, the MSE of the minAD filtering method were 66.02%, 58.94% and 16.33% less than the original dataset for AR(1), AR(2) and MA(3) forecasting models, respectively.The MAD values were 39.60%, 33.81% and 9.37% less than the original dataset for AR(1), AR(2) and MA(3) forecasting models, respectively. The results revealed that simple statistical tools, avgAD and minAD, can be used to improve the accuracy of forecasting models.

4. Conclusions

The avgAD and minAD methods are very simple, easy to understand and use to simply manipulate small datasets and give the better results prediction. In addition to the better quality of prediction for small dataset, these methods incur lower computational costs and are efficient forecasting tools. In future research, a method or a procedure to optimize the coefficients of avgAD and minAD for improved accuracy may be developed.

5. References

[1] Abdol SS,Liangyue C. Modeling and forecasting financial data: techniques of nonlinear dynamics. New York, USA: Springer; 2002.

[2] George EPB, Gwilym MJ, Gregory CR. Time series analysis: forecasting and control. USA: John Wiley & Sons; 2008.

[3] Pedroza C. A Bayesian forecasting model: predicting U.S. male mortality. Biostatistics. 2006; 7(4):530-50.

[4] Dan RR, Nada RS. Operations Management. 4th ed. USA: John Wiley & Sons; 2010.

[5] Richard AJ, Gouri KB. Statistics: principles and methods. 6th ed. USA: John Wiley & Sons; 2010.