Simple Technique For Forecasting With Small Dataset Based On Minimum Absolute Difference: Thai Gold Jewelry Price

Mathee Pongkitwitoon[1*], Watcharin Klongdee[2*]
[1]*Deprtment of Statistics, Faculty of Science Khon Kaen University, Khon Kaen, 40002, matpon@kku.ac.th*
[2]*Department of Mathematics, Faculty of Science Khon Kaen University, Khon Kaen, 40002, kwatch@kku.ac.th*
(* corresponding authors)

Abstract

This article focuses on a simple statistical method which is easy to compute and understand, low computational cost for forecasting model fitting and suitable for a small dataset. The minimum absolute difference is used to deal with the data preprocessing stage for the monthly Thai gold jewelry price prediction by using three-period simple moving average forecasting model. The results indicate that the approach method is better than the traditional method in addition to having a lower computational cost.

Keywords: Forecasting, Minimum Absolute Difference, Small Dataset, Thai Gold Jewelry Price

1. Introduction

Many Thai people interest to invest in the 96.5% gold jewelry, its value is not only a precious one but also a financial security, but the gold price actually cannot accurately predict because it depends on many factors, not only demand and supply. The researchers try to develop many different forecasting models or improve forecasting methods in many aspects such as multiple linear regression methods, autoregressive regression methods, nonlinear forecasting approaches and Bayesian approaches. We interest in the data filtering method, a data preprocessing method, by using a simple statistical method which is easy to use and understand and has a low computational cost in addition to small dataset handling. We introduce a minimum absolute difference (minAD), a simple statistical method, to filter the small dataset before fitting the forecasting model against the original dataset and compare the quality of the models by using mean square error (MSE), mean absolute deviation (MAD) and relative absolute error (RAE) [1-3].

2. Research Methodology

2.1 Testing Small Dataset

The monthly Thai gold jewelry price dataset, from January 2016 to March 2017, is used in this experiment (source: Gold Traders Association, Thailand; web site: http://www.goldtraders.or.th).

2.2 Small Dataset Filtering based on Minimum Absolute Difference

The minAD can compute by using the following equations:

$$\text{minAD} = \min\{|x_i - x_j|:\ i = 1,2,\dots,n-1;\ j = i+1, i+2,\dots,n\} \tag{1}$$

Let X be a set of the monthly Thai gold jewelry price,

$$\text{X} = \{x_1, x_2, \dots, x_{15}\} \tag{2}$$

where x_1 be a gold price of January 2016, x_2 be a gold price of February 2016, and so on,

and w_k be a gold price that have to be examined which be in sub dataset s_k,

$$s_k = \{x_{k-2}, x_{k-1}, w_k = x_k, x_{k+1}, x_{k+2}\} \tag{3}$$

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

We use three datasets X_1, X_2 and X_3;

$$X_1 = \{x_1, x_2, \dots, x_{12}\},$$

$$X_2 = \{x_1, x_2, \dots, x_{13}\} \quad \text{and}$$

$$X_3 = \{x_1, x_2, \dots, x_{14}\},$$

to predict the next one period of gold price; $\hat{x}_{13}$, $\hat{x}_{14}$ and $\hat{x}_{15}$ respectively, and we have the sub datasets s_{ik} for $i=1,2,3$ and $k=1,2,3,\dots,n_i$ where $n_1=12$, $n_2=13$ and $n_3=14$.

The approach algorithm to filter the original dataset is simple steps as following:

Step 1 Calculate the sample mean of s_{ik} and $\bar{x}_{ik} = \frac{x_{k-2}+x_{k-1}+x_{k+1}+x_{k+2}}{4-(number\ of\ null\ values)}$,

Step 2 Find minAD_{ik},

Step 3 Compare the w_k as these criterions;

If $w_k < \bar{x}_{ik} - \text{t}_{lower} \cdot \text{minAD}_{ik}$ then

$$w_k = \bar{x}_{ik} - \text{t}_{lower} \cdot \text{minAD}_{ik}$$

or

If $w_k > \bar{x}_{ik} + \text{t}_{upper} \cdot \text{minAD}_{ik}$ then

$$w_k = \bar{x}_{ik} + t_{upper} \cdot \text{minAD}_{ik}$$

where

t_{lower} and t_{upper} are the correction factors, in this research, we use $\text{t}_{lower} = 1$ and $\text{t}_{upper} = 1$.

In addition to the approach algorithm we have introduced, we use a sample standard deviation (SD) to filter the original dataset in a similar way [4].

2.3 Forecasting Model: The Original Dataset Against The Filtered Datasets

The original dataset, as in Figure 1, has an odd shape so we consider using three-period simple moving average (MA3) to fit the forecasting models for the datasets.

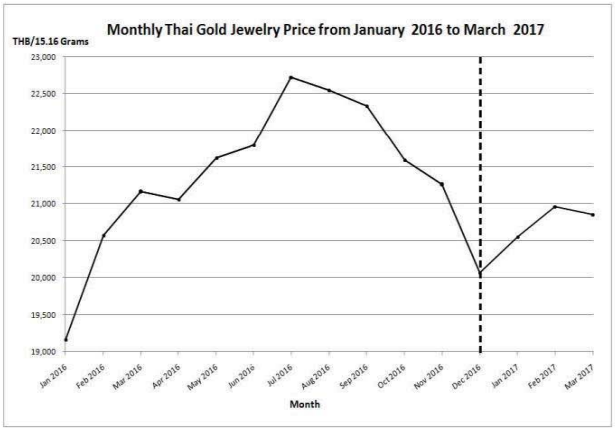

Figure 1 Monthly Thai gold jewelry price

The equation for calculate MA3 is

$$\hat{x}_{t+1} = \frac{x_t + x_{t-1} + x_{t-2}}{3} \tag{4}$$

2.4 Quality of Predictions

We use the forecasting models to predict the gold jewelry price for one period and measure the quality of predictions by using MSE, MAD and RAE that are calculated as the following equations [5]:

$$\text{MSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \hat{x}_i\right)^2}{n-1}} \tag{5}$$

$$\text{MAD} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \hat{x}_i| \tag{6}$$

and

$$\text{RAE} = \sum_{i=1}^{n}\frac{|x_i - \hat{x}_i|}{x_i} \tag{7}$$

3. Research Results and Discussions

The results, Figure 2 and Table 1, show that the prediction of original dataset of January 2017 is better than the predictions of both filtered datasets but the next two predictions, February and March 2017, the results of original dataset is worse than the filtered datasets.

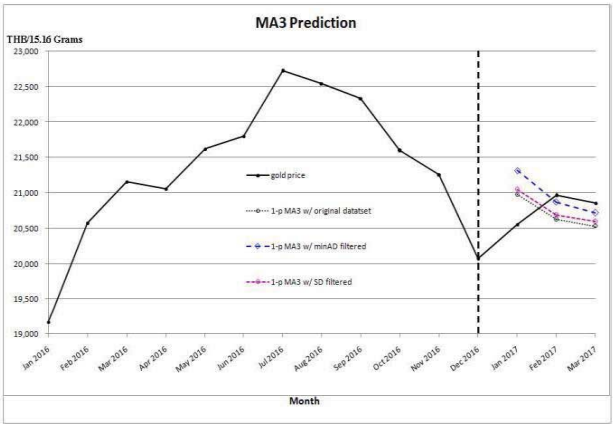

Figure 2 MA3 prediction

Table 1 Gold price predictions

Dataset	Jan 2017	Feb 2017	Mar 2017
Original	20,977.85	20,629.85	20,531.07
minAD filtered	21,317.95	20,872.75	20,717.79
SD filtered	21,052.23	20,691.42	20,597.51
Real prices	20,566.00	20,965.22	20,861.11

The MAD and RAE and the residuals plots of the predictions, Table 2 and Figure 3, show that the approach method, using minAD filtered, is better than traditional method for second and third predictions, although the MSE of the approach method is greater than the traditional method, and the results show that the SD filtered is also better than the traditional method.

Table 2 MSE, MAD and RAE of predictions

Dataset	MSE	MAD	RAE
Original	199,677.04	1,087.26	0.0523
minAD filtered	304,829.25	977.74	0.0483
SD filtered	195,345.90	1,033.62	0.0498

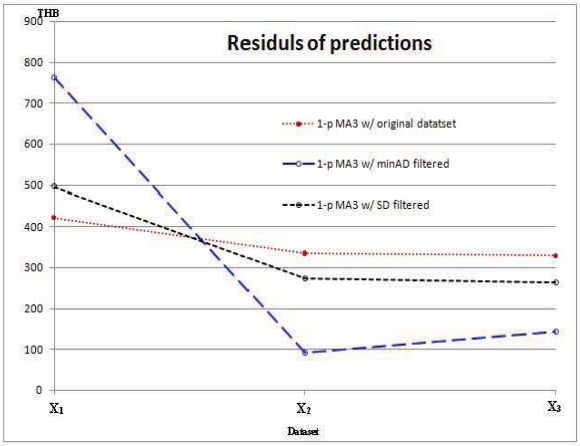

Figure 3. Residuals of predictions

These reveal that the minAD filtered and SD filtered methods are good for using to filter data before fitting the forecasting model and Figure 3 shows that if there are more data then the error of predictions will decrease.

We can conclude that the original dataset has to be filtered by using the appropriated method before fitting model because it usually is interfered with noise and the approach is the better method to filter because it is easy to calculate than the SD calculation.

4. Conclusions and Recommendations

We introduce the simple statistical method, minAD, for filtering dataset before fitting the forecasting model which is easy to calculate and understand and it gives a better forecasting model

than ordinary method in addition to consuming a low computational cost. In this experiment, we firstly assign the correction factors, t_{lower} and t_{upper}, to equal one, we hope that we may find the optimum values of the correction factors for better quality of forecasting in the next research.

5. References

[1] George E. P. Box, Gwilym M. Jenkins and Gregory C. Reinsel. Time series analysis: forecasting and control. John Wiley & Sons, Inc.; 2008.

[2] Abdol S. Soofi and Liangyue Cao. Modeling and forecasting financial data: techniques of nonlinear dynamics. New York, USA: Springer Science+Business Media; 2002.

[3] Jon Danielsson. Financial risk forecasting. UK. John Wiley&Sons Ltd.; 2011.

[4] A. Johnson and Gouri K. Bhattacharyya. Statistics: principles and methods. 6th ed. USA: John Wiley & Sons, Inc.; 2010.

[5] Ian T. Jolliffe and David B. Stephenson. Forecast verification: a pratitioner's guide in atmospheric science. UK. John Wiley&Sons Ltd.; 2003.