JIBC

Journal of Internet Banking and Commerce

An open access Internet journal (http://www.icommercecentral.com)

Journal of Internet Banking and Commerce, April 2017, vol. 22, no. 1

A HYBRID APPROACH EMD-MA FOR SHORT-TERM FORECASTING OF DAILY STOCK MARKET TIME SERIES DATA

AHMAD M AWAJAN

University Science Malaysia, 11800 Gelugor, Penang, Malaysia

Tel: 00960174528978;

Email: **awajanmath@yahoo.com**

MOHD TAHIR ISMAIL

University Science Malaysia, 11800 Gelugor, Penang, Malaysia

AL WADI S

University of Jordan, Queen Rania str., Amman, Jordan

Abstract

Recently, forecasting time series has attracted considerable attention in the field of analyzing financial time series data, specifically within the stock market index. Moreover, stock market forecasting is a challenging area of financial time-series forecasting. In this study, a hybrid methodology between Empirical Mode Decomposition with the Moving Average Model (EMD-MA) is used to improve forecasting performances in financial time series. The strength of this EMD-MA lies in its ability to forecast non-stationary and non-linear time series without a need to use any transformation method. Moreover, EMD-MA has a relatively high accuracy and offers a new forecasting method in time series. The daily stock market time series data of 10 countries is applied to show the forecasting performance of the proposed EMD-MA. Based on the five forecast accuracy measures, the results indicate that EMD-MA forecasting performance is superior to traditional Moving Average forecasting model.

Keywords: **Forecast Time Series; Empirical Mode Decomposition; Moving Average; Intrinsic Mode Function; Forecast Accuracy Measures**

INTRODUCTION

In financial time series analysis, one of the primary issues is modelling and forecasting financial time series data specifically stock market index. Usually, the transformation of a financial time series, rather than its original scale, is taken for describing its dynamics. Proper transformation is necessary to convert original non-stationary processes to stationary processes and subsequently to utilize mathematical and statistical properties for stationary processes.

The hybrid models combine strengths of few traditional models to get a better forecasting accuracy. Recently, several hybrid models were applied EMD in the literature for time series forecasting. That by using EMD to decompose the non-stationary and non-linear time series data into Intrinsic Mode Functions (IMFs) and residual components. And then use forecasting model to forecast each component. Then all these forecasted values were aggregated to produce the final forecasted value of the original time series.

Such as in Abadan et al. [1] used a hybrid EMD-ARIMA (Autoregressive integrated moving average) to forecasting the monthly prices of rice data [2]. Also, Li and Wang [3] also used the same methodology, but with wind speed data. A hybrid EMD-AR (Autoregressive) model was developed by coupling an AR model with the EMD technique in Duan et al. [4]. A hybrid EMD-LSSVR (least squares support vector regression) forecasting model has been applied on foreign exchange rate in Lin et al. [5]. While in Tatinati and Veluvolu [6] used a hybrid of EMD, LS-SVM (Least Squares-Support Vector Machines), and AR model with Kalman filter to predict wind speed data.

With regard to all those literature reviews, this study attempts to apply a hybrid of EMD-MA to forecast the daily stock market data of 10 countries. In order to assess the performance of forecasting, and the proposed method is compared with traditional Moving Average forecasting model. Experimental results show that the proposed method is superior to existing method in terms of three accuracy forecasting measure. Section 2 introduces methods is used in methodology in this paper which are EMD, IMF and Moving Average Model. Section 3 presented the proposed methodology with flowchart explain the steps. Section 4 analyzes the daily stock market time series data of 10 countries with a discussion the result showing the capability of EMD-MA. Finally, in Section 5 some concluding remarks are addressed.

METHODOLOGY

In this section, the various steps for the implementation of the EMD-MA forecasting method are described in detail. Which are Empirical Mode Decomposition and Moving Average Model?

Empirical Mode Decomposition [EMD]

EMD was described by Huang et al. [7], and this method has been modified by Jaber et al. [8] and Lu [9]. The main idea of EMD is the decomposing of nonlinear and non-stationary time series data into several of simple time series. And it analyzing time series with keeping the time domain of the signal. It supplies a strong and adaptive process to decompose a time series into a combination of time series that known as intrinsic mode functions (IMF) and residual. Later, the original signal can be constructed back as the following:

$$x(t) = \sum_{i=1}^{n} IMF_i(t) + r(t) \tag{1}$$

Where $x(t)$ represents the original time series, $r(t)$ represents the residue of the original time series data decomposition and IMF_i represent the i^{th} intrinsic mode function (IMF) series.

In order to estimate these IMFs, the following steps should be initiated and the process is called the sifting process of time series $x(t)$ are shown below:

1. Start the first step by taking the original time series $x(t)$ for sifting process and assuming the iteration index value is $i=1$
2. Then, evaluate all of local extreme values of the time series $x(t)$
3. After that, form the local maxima (local upper) envelope function $u(t)$ by connecting all local maxima values using a cubic spline line. In a similar way, form the local minimum (local lower) envelope function $l(t)$, and then form the mean function $m(t)$ by using this following;

$$m(t) = \frac{u(t) + l(t)}{2} \tag{2}$$

4. Next, define a new function $h(t)$ using the mean envelope $m(t)$ and the signal $x(t)$ on this formula:

$$h(t)=x(t)-m(t) \tag{3}$$

Check the function $h(t)$ is an IMF, according to IMF conditions (shown in the second part of this section). If the function $h(t)$ has satisfied IMF conditions, then go to step 5. If not, go back to step 2 and renew the value of $x(t)$ such that became $h(t)$, repeat steps 2 again until 4.

5. In step 5, firstly save the result of the IMF obtain from the last step. Secondly, renew the iteration index value such that became i=i+1. Thirdly attain the residue function $r(t)$ using the IMF and the signal $x(t)$ on the formula.

$$IMF_i(t)=h(t) \Rightarrow r_{i+1}(t)=x(t)-IMF_i(t) \tag{4}$$

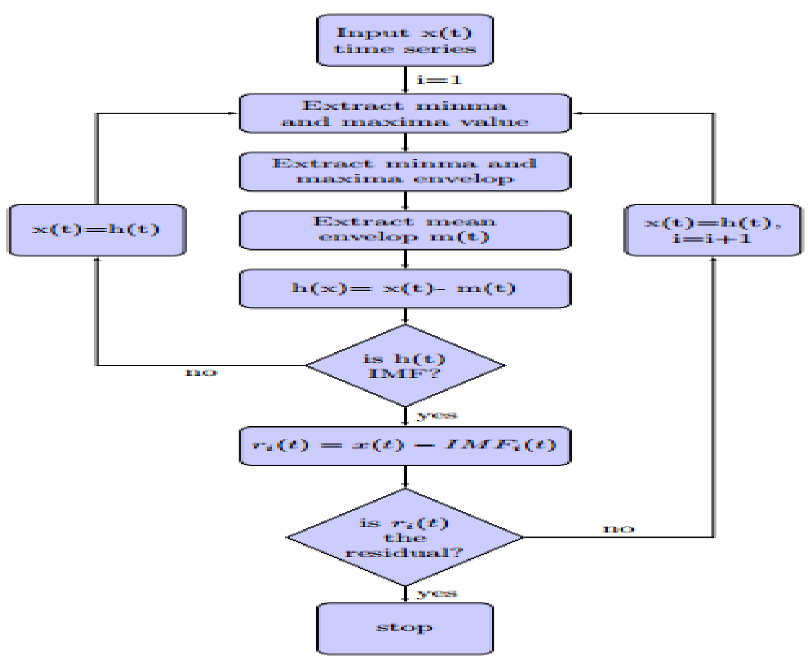
6. Finally, make a decision whether the residue function $r(t)$ that acquire from step 5 is a monotonic or constant function. Then, save the residue and all the IMFs obtained. If the residue is not monotonic or constant function, return to step 2.

The steps 1 to 6 which were discussed above allow the sifting process (EMD algorithm) to separate time-altering signal features. Figure 1 summarizes all the steps.

Intrinsic Mode Function (IMF)

Based on the EMD algorithm presented in the previous section, the IMF produces by the sifting process need to satisfy two conditions.

Figure 1: Flowchart of empirical mode decomposition estimation process.



1. $$|Num[extreme] - Num[cross - zero]| < 1 \quad (5)$$

Where Num.extreme represents the number of local extreme points (all local maxima and all local minima), also Num [cross-zero] represent the number of cross-zero points.

$$|m(t)| = \left|\frac{u(t) + l(t)}{2}\right| < \varepsilon \quad (6)$$

where $u(t)$ represents the envelope function generated by using cubic spline line on all local maxima, $l(t)$ represents the envelope function generated by using a cube spline line on all local minima, $m(t)$ represents the mean function that it was obtained by evaluating the mean of $u(t)$ and $l(t)$, and ε is a very small positive number that close to zero, sometime equal zero.

Moving Average Model (MA)

A moving-average of time series data is the result of calculating the average for a number of past data observation.

On other word, a moving-average model is a model of linear regression uses past error value of the time series.

The moving average model of order q is denoted by $MA(q)$, is formally given by [2]:

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \quad (7)$$

where μ is the mean of the series, and $\theta_1,\ldots, \theta_q$ are the parameters may be positive or negative. And the εt, $\varepsilon_{t-1},\ldots, \varepsilon_{t-q}$, are white noise error terms. The value of q is called the order of the MA model. This can be written using the backshift operator B as

$$X_t = \mu + (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t \quad (8)$$

The autocorrelation function is used to determine the order q.

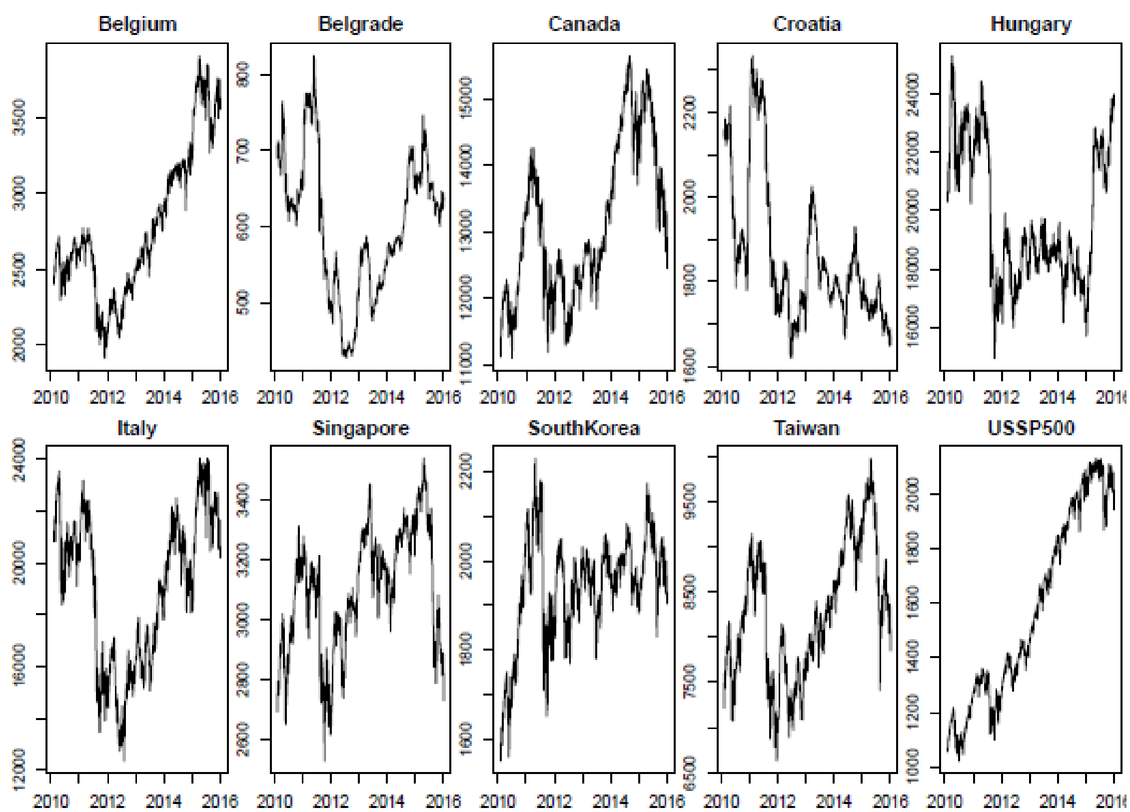
PROPOSE METHODOLOGY AND DATA

In this section, the various steps for the implementation of the EMD-MA forecasting method are described in detail. Which are Empirical Mode Decomposition and Moving Average Model?

Data

In this study, nonlinear and non-stationary time series data from the daily stock market of 10 countries are used. Table 1 presents the names of countries with the Basic statistics and the number of observations for each country, where SD is Standard Deviation, Sk is Skewness, Kts is Kurtosis and N is Number of observations. The data are extracted from the Yahoo finance website. Figure 2 shows the time series plot of these countries. The daily closing prices are used as a general measure of the stock market over the past six years. The whole data set for each country - covers the period from 9 February 2010 to 7 January 2016. The data set is divided into two parts. The first part (n observations) is used to determine the specifications of the models and parameters. The second part, on the other hand, (h observations) is reserved for out-of-sample evaluation. This part is used comparison of performances among various forecasting models. Singapore stock market data are taken as example. Where the number observation is N=1501, the first part is n=1500, 1499, 1498, 1497, 1496 and 1495 and the second part is h=1, 2, 3, 4, 5 and 6 respectively, are used.

Figure 2: Time Series Plots.



Propose Methodology

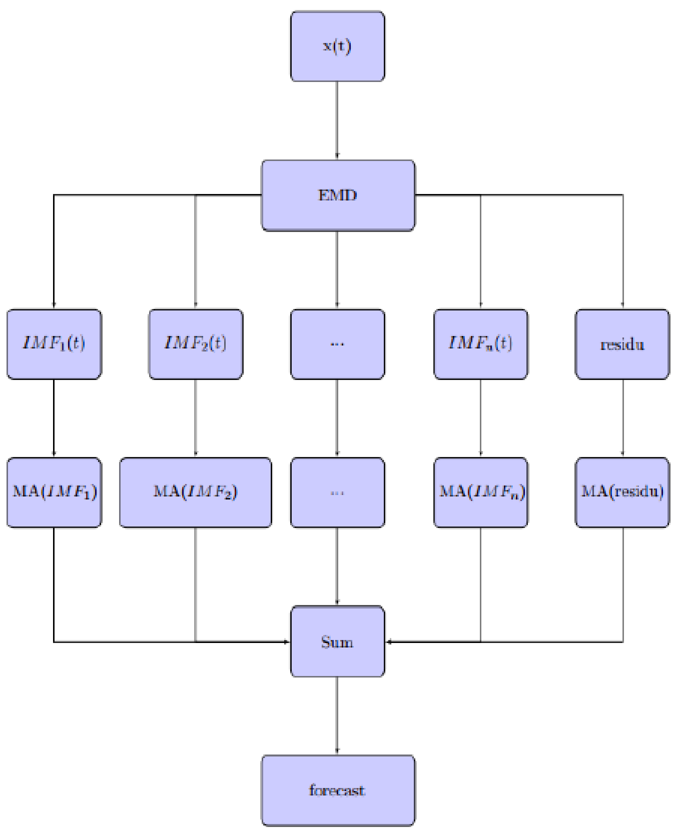
The proposed methodology consists of three stages.

Table 1: Basic statistics.

Country	Mean	Median	SD	Sk	Kts	N
Belgium	2785.21	2658.78	484.6	0.55	-0.64	1517
Belgrade	600.07	606.38	91.54	0.02	-0.76	1494
Canada	13157.33	12844.6	1160.69	0.4	-0.99	1500
Croatia	1861.88	1807.81	168.25	1.22	0.48	1477
Hungary	19773.29	18919.78	2345.84	0.4	-1.13	1474
Italy	19025.28	19671.55	2937.37	-0.3	-1.14	1520
Singapore	3099.76	3118.65	201.12	-0.25	-0.67	1501
SouthKorea	1943.25	1961.98	117.81	-0.74	0.58	1467
Taiwan	8280.26	8254.79	728.27	0.13	-0.87	1465
USSP500	1579.25	1493.69	344.31	0.2	-1.44	1490

This methodology is presented as a flowchart in Figure 3

Figure 3: Flowchart of a hybrid Empirical Mode Decomposition with Moving Average.



1. Firstly, the use of empirical mode decomposition (EMD) on the daily stock market time series data. In this stage, Intrinsic Mode Functions (IMFs) and residue are obtained.
2. Secondly, the Moving Average Model (MA) is applied on each IMFs and residue to forecast h days ahead.
3. Finally, in the last stage all the forecasted results for IMFs and residue are added up.

RESULT AND DISCUSSION

In this study, stock market time series data of 10 countries are used to present the forecasting accuracy of the EMD-MA method. Traditional Moving Average Model is used in order to validate the forecasting performance of EMD-MA. Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Absolute Scaled Error (MASE) and Theil's U-statistic (Theil) will be utilized to evaluate the forecasting accuracy for each method. Equations 9, 10, 11, 12 and 13 are showed the formula of RMSE, MAE, MAPE, MASE and Theil respectively. Where $\hat{y}_i$ is the forecast value of the variable y at time period i from knowledge of the actual series values (Table 2).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{9}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{10}$$

Table 2: Presented the different mean errors which include RMSE, MAE, MASE, MAPE and TheilU of EMD-MA and Moving Average forecasting method for forecasting at h=1, 2, 3, 4, 5 and 6 for each stock market time series data of 10 countries.

Error measure	Method	h=1	h=2	h=3	h=4	h=5	h=6
RMSE	MA	152.17	338.93	537.9	611.77	624.85	715.39
	EMD-MA	123.49	309.06	438.46	540.98	617.46	629.67
MAE	MA	152.17	313.85	497.03	546.65	556.39	666.34
	EMD-MA	123.48	284.79	386.76	471.68	540.28	577.98
MAPE	MA	2.634	4.428	6.174	7.675	8.536	8.996
	EMD-MA	1.534	3.641	5.479	6.854	8.282	8.657
MASE	MA	-	45.174	5.061	6.901	6.973	7.706
	EMD-MA	-	16.842	4.58	6.017	6.654	7.144
Theil	MA	-	65.581	6.161	7.403	7.279	7.798
	EMD-MA	-	27.656	5.941	6.66	7.247	7.414

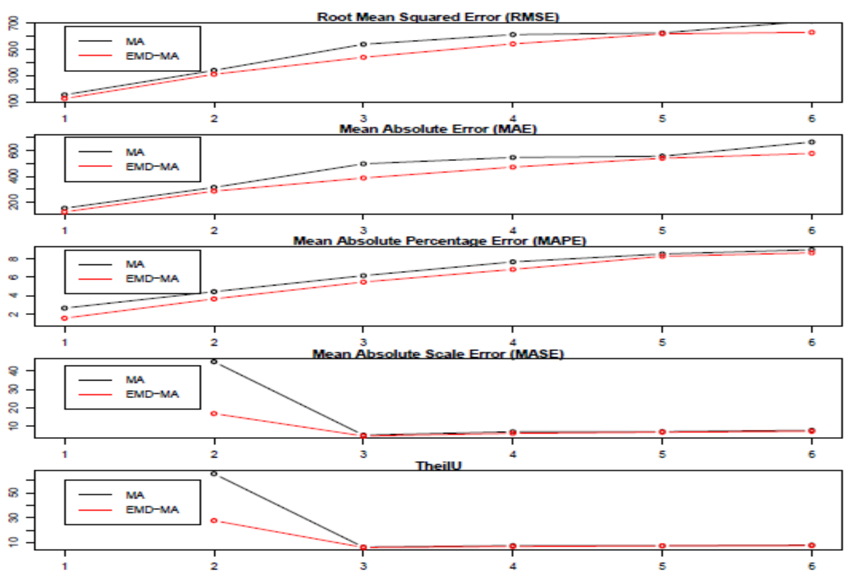
$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i}.100\% \tag{11}$$

$$MASE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{|y_i - \hat{y}_i|}{\frac{1}{n-1}\sum_{i=2}^{n}|y_i - \hat{y}_i - 1|}\right) \tag{12}$$

$$Theil = \frac{\sqrt{\sum_{i=1}^{n=1}\left(\frac{\hat{y}_i + 1 - y_i - 1}{y_i}\right)^2}}{\sqrt{\sum_{i=1}^{n=1}\left(\frac{y_i + 1 - y_i}{y_i}\right)^2}} \tag{13}$$

Table 2 and Figure 4 present the different mean errors which include RMSE, MAE, MAPE, MASE and Theil of EMD-MA and Moving Average Model for forecasting at h=1, 2, 3, 4, 5 and 6 for all stock market time series data of 10 countries. This indicates that the forecast accuracy for EMD-MA is better than the traditional Moving Average forecasting model.

Figure 4: Presented the different mean errors which include RMSE, MAE, MASE, MAPE and Theil U of EMD-MA and Moving Average forecasting method for forecasting at h = 1, 2, 3, 4, 5 and 6 for each stock market time series data of 10 countries.



CONCLUSION

Time series forecasting still remains as one of the most difficult area due to the non-stationary and non-linearity of financial time series data. In this study, we have presented a new hybrid method, a composite of empirical mode decomposition with Moving Average model (EMD-MA) for non-stationary and nonlinear time series forecasting. EMD-MA was tested on daily stock market time series data of 10 countries based on the comparison of five of forecast accuracy measurements. It was found that EMD-MA is able to outperform the traditional Moving Average model. Thus, this paper has strengthened the idea that EMD-MA forecasting method is suitable for non-stationary and nonlinear time series.

REFERENCES

1. Abadan S, Shabri A (2014) Hybrid empirical mode decomposition arima for forecasting price of rice. Applied Mathematical Sciences 8: 3133-3143.
2. Chris C (2016) The analysis of time series: an introduction. CRC press.
3. Li R, Wang Y (2008) Short-term wind speed forecasting for wind farm based on empirical mode decomposition. In Electrical Machines and Systems. International Conference. IEEE. pp: 2521-2525.
4. Duan WY, Huang LM, Han Y, Zhang YH, Huang S (2015) A hybrid AR-EMD-SVR model for the short-term prediction of nonlinear and non-stationary ship motion. Journal of Zhejiang University Science 16: 562-576.
5. Lin CS, Chiu SH, Lin TY (2012) Empirical mode decomposition based least squares support vector regression for foreign exchange rate forecasting. Economic Modelling 29: 2583-2590.
6. Tatinati S, Veluvolu KC (2013) A hybrid approach for short-term forecasting of wind speed. The Scientific World Journal.
7. Huang NE, SLong Z, Steven R, Wu, Manli C, et al. (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 454: 903-995.
8. Jaber AM, Ismail Tahir Md, Altaher AM (2014) Empirical mode decomposition combined with local linear quintile regression for automatic boundary correction, Abstract and Applied Analysis 2014: 8.
9. Lu Y, Oruklu E, Saniie J (2013) Chirplet signal and Empirical Mode Decompositions of ultrasonic signals for echo detection and estimation. Journal of Signal and Information Processing 4: 149-157.