International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 12, No.1 (Dec-2022)

https://dx.doi.org/10.12785/ijcds/1201108

Enhancing Stock Price Prediction Using Empirical Mode Decomposition, Rolling Forecast and Combining Statistical Methods

Mohammad Raquibul Hossain[1,2], Mohd Tahir Ismail[1] and Md. Jamal Hossain[1,2]

[1]*School of Mathematical Sciences, Universiti Sains Malaysia, Pulau Pinang, Malaysia*
[2]*Department of Applied mathematics, Noakhali Science and Technology University, Noakhali-3814, Bangladesh*

Received 14 Aug. 2021, Revised 27 Jul. 2022, Accepted 1 Dec. 2022, Published 8 Dec. 2022

Abstract: Data analytics especially predictive analytics is very important research domain which includes time series forecasting. Nonlinear nonstationary time series are challenging to predict. This paper presents the outcome of the research study in finding better forecasting methods for nonlinear nonstationary time series. Rolling forecast approach and locally adaptive empirical mode decomposition (EMD)-based hybridization were employed with autoregressive integrated moving average (ARIMA) and exponentially weighted moving average (EWMA). Thus, two methods were EMD-ARIMArolling and EMD-EWMArolling of which the later was found better in this study. Also, EMD-EWMArolling was combined with ARIMArolling and EWMArolling using affine combinations to develop affEEArolling and affEEErolling methods. Proposed affEEArolling and affEEErolling along with six other compared methods were employed on nine closing price stock datasets from NASDAQ Financial-100 companies and compared using accuracy measurements. From the results, it was found that proposed methods significantly improved forecast accuracy and outperformed the compared methods (e.g., in ACGL dataset, affEEArolling reduced RMSFE by 55.98% where rolling forecast, EMD-hybridization and affine combination improved 43.7%, 4.24% and 18.28% respectively and affEEErolling improved 56%). Hence, EMD-based hybridizations and forecast combinations can be useful tools for time series forecasting. In addition, such EMD-based advanced methods can be considered for inclusion in financial technologies.

Keywords: Empirical Mode Decomposition, Intrinsic Mode Functions, ARIMA, EWMA, Stock Price Prediction, Forecast Combination

1. Introduction

In many domains of human activities, time series data play important roles especially in decision making and future projection or prediction. Some time series may contain deterministic patterns and such series can be predicted comparatively easily by using traditional benchmark methods like autoregressive integrated moving average (ARIMA) and data smoothing methods (e.g., EWMA which stands for exponentially weighted moving average). However, many time series data especially human-involved data (namely financial and economic data) are very challenging to capture by traditional methods. Such reality is addressed in several research works including [1]. To face the challenges, there are many different research directions and approaches which started in the last century and currently its importance and practice is upward. Some of the works on time series forecasting are reviewed by [2] and [3]. Among many other works, some recent studies include [4] where they used ARIMA models and polynomial functions to predict COVID-19 per regions; [5] used ARIMA and Holt-Winters methods for forecasting potato price and [6] used ARIMA for stock return forecasting. Aside from single methods, many studies focused on hybridization and combination methods. In their forecast study of stock price, [7] used ARIMA and BPNN (i.e., Back Propagation Neural Network); [8] used ARIMA and SVM (i.e., Support Vector Machine) to predict daily price of rubber. For long-term pollution prediction, [9] used statistical and deep learning methods; [10] used padding-based Fourier transform denoising and deep learning methods to predict stock market indices.

Among different hybridized and combined forecast methods, Empirical mode decomposition (EMD)-based methods are very effective and useful. EMD is very efficient in decomposing a nonlinear and non-stationary time series. After the development of locally data adaptive algorithm EMD, many time series research studies focused on EMD-based forecasting. Even the original contributor of EMD, applied it on financial time series [11] which is an inspira-

E-mail address: raquib.math@gmail.com, m.tahir@usm.my, z_math_du@yahoo.com

tion and one of the influences. Later on, different authors made efforts in EMD-based forecasting research studies which include [12] where they used hybrid EMD-ARIMA in predicting short-term traffic speed. To forecast long-term streamflow, [13] used EMD-ARIMA and EEMD-ARIMA methods. In forecasting financial time series, [14] used EMD-LSTM method (where LSTM stands for Long Short-Term Memory). Recent studies on EMD-based hybrid methods also encompass [15] (which used EMD and ANN-based hybrid methods for stock price prediction), [16] (which used EMD-Neuro-Fuzzy hybrid method in analyzing and predicting financial time series), [17] (which used EMD-Theta method for stock price forecasting), [18] (which used CEEMDAN-LSTM method for stock index price prediction). Also, recently for predicting stock index [19] used combined model of IEME-EMD-SVR (where IEME stands for Improved Extreme Mirror Extension and SVR is for Support Vector Regression), in prediction of stock price [20] employed EMD-ARIMA-EWMA methods, to forecast stock price index [21] found high performing experimental results using MEMD-SVR-SVR (where MEMD stands for multivariate EMD), in predicting stock price [22] found outperforming results using CEEMD-CNN-LSTM (where CEEMD is short form of complete ensemble EMD) than other models. In forecasting of GDP data, [23] used EMD-LSTM which significantly improved the accuracy of only LSTM. These studies indicate that EMD-based prediction is active and potential research area of time series.

Many research studies both old and new encouraged and adopted forecast combinations for better accuracy. Some recent works include [24], [25], [26], [27], [28]. The work of [29] reported that majority of the accurate methods in the M-4 (i.e., Makridakis-4) competition were forecast combination methods. In addition, rolling forecast [30], [31] has additional advantage over traditional forecasting. Considering effectiveness of rolling forecast, this study adopted it in proposed methods.

This research study surveyed on existing literature from four points of view: first is general context, current trend and application of single forecasting methods; second is contemporary hybridization and combination methods; third is recent studies of EMD-based hybridization and combination; fourth is studies of financial time series prediction which particularly focused on EMD-based prediction methods. The research studies of all these types are refereed in this section and their context is briefly presented. Hence, non-financial time series studies contributed for potential methodologies and financial time series studies contributed for empirical results along with different methods and approaches. Methods and data were two main directions in this study in surveying existing literature of which EMD-based financial time series forecasting studies were closely relevant. Also, overall literature indicate that single popular conventional methods (ARIMA and EWMA) are generally useful and they are regarded benchmark methods. For comparison of this study with others is that although there are many works on EMD-based hybrid methods, combinations or especially affine combination-based studies of EMD are rare in existence. However, recent works on forecast combinations are highly worthy of mention although which are not EMD-based, they are feasibly used in time series forecasting practices. Thus, studies from surveyed literature influenced differently as directly (i.e., those with financial data), indirectly (i.e., those with non-financial data) and partially on this research study.

In pursuit of better forecasting methods in case of non-linear and non-stationary time series, there exists scope for research studies because existing methods are not always sufficient to serve the purpose of expected accuracy. To find better forecasting methods, this research study focused on the assumption that using EMD, statistical methods, rolling forecast and affine combination can collectively and significantly improve forecast accuracy. Searching through existing literature, such study is not yet done. On experimentation, this paper proposes two EMD-based affine combination methods with rolling forecasting approaches. One is affEEArolling (i.e., affine combination between EMD-EWMArolling and ARIMArolling) and another is affEEErolling (i.e., affine combination between EMD-EWMArolling and EWMArolling). From the empirical results, the proposed methods were found to produce better accuracy than compared methods. For both of the proposed methods, three improvement stages (i.e., rolling forecast, EMD-based hybridization and affine combination) significantly contributed on the forecast accuracy. These are essentially contributory elements of this research study and addition to the existing literature of EMD-based financial time series forecasting.

The following section conveys background studies (including benchmark methods and EMD-based hybridizations). Next sections contain proposed methods, datasets and experimental results, discussion with outcomes and lastly conclusion with future works.

2. Background Materials

This section includes essential details of benchmark methods ARIMA and EWMA along with their EMD-based hybrid methods EMD-ARIMA and EMD-EWMA and finally rolling forecast approach.

A. ARIMA Method

ARIMA is the short form for autoregressive integrated moving average [32], [33] which can fit and forecast a nonstationary time series. It methodically applies differencing process to transform a nonstationary time series (say $X = \{x_1, x_2, x_3, \ldots, x_t, \ldots\}$) into a stationary series (say $Y = \{y_1, y_2, y_3, \ldots, y_t, \ldots\}$). The lag difference generally taken once or twice to transform X into $Y$is known as order of differencing (d). After differencing process, ARMA method (i.e., Autoregressive Moving Average) is applicable to fit the underlying stationary time series Y. The mathematical formulation of a general ARMA(p,q) method is:

$$y_t = c + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + e_t + \beta_1 e_{t-1} + \beta_2 e_{t-2} + \cdots + \beta_q e_{t-q} \quad (1)$$

where c, y_i's and e_j's are respectively constant, past p autoregressive terms and past q error terms (also known as random shocks); α_i's are autoregressive parameters and β_j's are moving average parameters. Using order of differencing d (or equivalently known as order of integration), better written notation of ARIMA is $ARIMA(p, d, q)$. Although ARIMA was introduced long before, its usage is still popular as well as effective. Some recent ARIMA-based research studies include [34], [6], [4], [8], [35], [5].

B. EWMA Method

EWMA (i.e., exponentially weighted moving average) [36], [37], [38] is a data smoothing method. From a given time series $X = x_1, x_2, x_3, \dots, x_t, \dots$, EWMA (also known as single exponential smoothing or SES) creates the smoothing series $S = S_1, S_2, S_3, \dots, S_t, \dots$. A general EWMA method has the following mathematical formula:

$$S_2 = x_1 \quad (2)$$

$$S_t = x_{t-1} + (1-\gamma) S_{t-1},\ 0 < \gamma < 1,\ t \geq 2 \quad (3)$$

where γ, t and S_t are respectively smoothing parameter, time order and present smoothed term. Clearly, S_t is found from the convex combination of x_{t-1} (i.e., most recent original data) and S_{t-1}(i.e., most recent smoothing data). Another general form of EWMA can be presented as:

$$S_t = \gamma \sum_{i=1}^{t-2} (1-\gamma)^{i-1} x_{t-i} + (1-\gamma)^{i-1} S_2,\ t \geq 2 \quad (4)$$

The weights $\gamma(1-\gamma)^{i-1}$ of x_{t-i}'s (i.e., past data) exponentially diminishes as $(1-\gamma)^t$ is a decreasing function with increasing t.

C. EMD Method

In the seminal work on Hilbert-Huang transforms (HHT), [39] introduced empirical mode decomposition (EMD) as an essential part of HHT. Foundational or developmental purpose of EMD was the analysis of signal processing. Conceptually and practically, EMD being an effective decomposition procedure has locally adaptive nature. The best feature of EMD procedure is that it does not alter the original time domain. In an EMD process, a time series is decomposed into several subseries namely intrinsic mode functions (IMFs) and a residue. The defining properties of an IMF are (a) mean value has to be zero and (b) difference from number of extrema to number of zero crossings has to be zero or at most one. In the IMF generating process, EMD produces rapidly oscillating high frequency IMFs in the beginning and sequentially produces slowly oscillating low frequency IMFs. Finally, the process ends with a residue after extracting all possible IMFs. EMD process is presented through flowchart (Figure 1).

EMD as shown in flowchart of Figure 1 is basically an algorithm which a process named sifting process for extraction of decomposition components. EMD process or EMD sifting process for signal data or a time series dataset $x(t)$ essentially involves the following steps:

Step 1: Finding local minima and maxima (i.e., extrema) of $x(t)$ and temporary remainder (which is found immediately after extraction of each IMF).

Step 2: Fitting and formulating two sub-datasets of local minima and maxima with two envelops of cubic splines. Two envelopes are maxima-contained upper envelope (involving half of complete data above local mean) E_u and minima-contained lower envelope (with remaining half of dataset below local mean) E_l.

Step 3: Obtaining mean envelope E_m (i.e., arithmetic mean of E_u and E_l).

$$E_m = \frac{E_u + E_l}{2} \quad (5)$$

Step 4: Getting 1^{st} temporary remainder D_1 by subtracting E_m from $x(t)$.

$$x(t) - E_{m1} = D_1 \quad (6)$$

Step 5: Checking whether D_1 is an IMF by following IMF characteristics. If it is an IMF, then to follow step 6. If D_1 is not an IMF, next mean envelope and hence next temporary remainder is computed from steps 2 to 4 which are repeated for producing other subsequent mean envelopes (i.e., 2^{nd} to higher) and corresponding temporary remainders are also found.

$$D_1 - E_{m2} = D_2, D_2 - E_{m3} = D_3,\ \dots,\ D_{k-1} - E_{mk} = D_k \quad (7)$$

In EMD sifting process of IMFs extraction, stopping criterion (e.g., Cauchy convergence criterion involving standard deviation briefly CSD) is used which involve the following formula:

$$CSD_k = \sum_{t=0}^{T} \frac{(D_{k-1}(t) - D_{k-1}(t))^2}{D_{k-1}^2(t)} \quad (8)$$

If CSD_k generates smaller value than a pre-set minimum, the sifting process is completed.

Step 6: A D_k is considered as an IMF if it satisfies the pre-set stopping criterion. If $F_1 = D_k$is the 1^{st} IMF, first continual remainder $R_1$is the difference of F_1 from $x(t)$, i.e., $R_1 = x(t) - F_1$. Thus, during extraction of subsequent IMFs, current continual remainder R_i and other subsequent continual remainders R_{i+1}'s are computed by following steps 1 to 5.

$$R_2 = R_1 - F_2, R_3 = R_2 - F_3, \dots, R_n = R_{n-1} - F_n \quad (9)$$

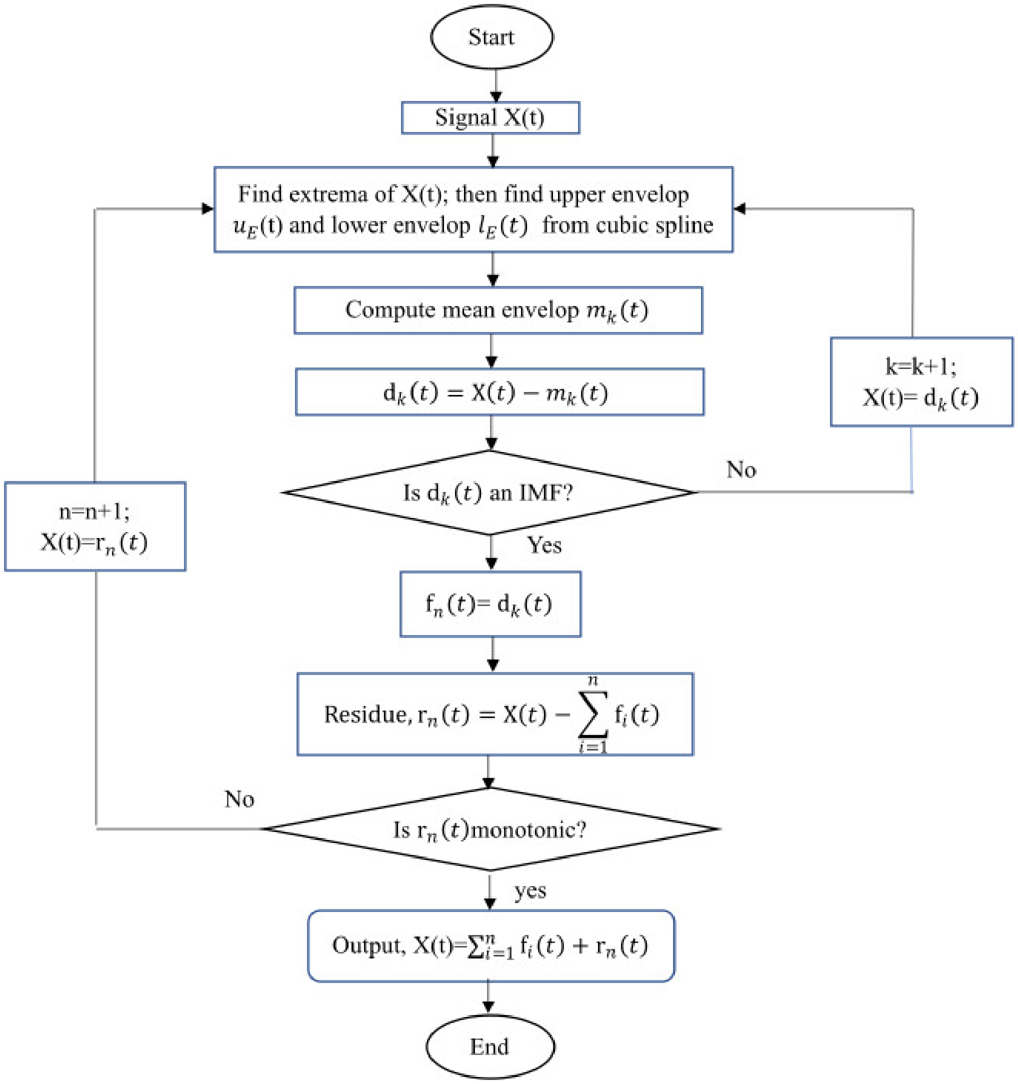

Figure 1. Flowchart of empirical mode decomposition (EMD)

Finally, the last continual remainder R_n (known as residual) is obtained.

D. EMD-ARIMA and EMD-EWMA Hybrid Methods

The EMD-ARIMA approach is the hybridization of EMD and ARIMA [12] where EMD is used for decomposition and ARIMA is used for data fitting and hence forecasting. First, using the locally adaptive EMD algorithm on a time series, decomposed components (i.e., IMFs and residue) are generated. Second, for each EMD component, a suitable and efficient ARIMA model is selected from the ARIMA method class. Third, all component forecasts are aggregated to obtain combined or hybridized forecast and accuracy measures are computed from the test data and forecast data. Also, the performance or accuracy of the method is presented against other competing methods. Definitely, developmental purpose of EMD or HHT was signal analysis. However, there exist a number of EMD-based research studies in other time series related domains encompassing finance and economics. The method or procedure of EMD-EWMA [40] is similar to EMD-ARIMA [17]. EMD is efficient in data decomposition with locally adaptive feature and EWMA is effective approach for data smoothing, fitting and forecasting. Therefore, suitable combination between these methods can be expectedly a good match for time series forecasting where data are nonlinear and non-stationary.

E. Rolling Forecast

Being updated with most recent available data is important as well as effective in decision making and forecasting process. Such approach or strategy adoption is getting attention, popularity and importance [30], [31]. In a rolling forecast approach, training data is updated or revised in a rolling basis to forecast data in the targeted future horizon [41]. A diagrammatic presentation of rolling forecast is shown in Figure 2. Rolling forecast is described more in datasets and results section.

3. Proposed Experimental Methods

This study firstly finds an EMD and rolling forecast based better method which is here EMD-EWMArolling. This method is expected to perform better than benchmark ARIMA, EWMA, ARIMArolling (i.e., rolling forecast using ARIMA), EWMArolling (i.e., rolling forecast using EWMA) and EMD-ARIMA rolling (i.e., rolling forecast using EMD-ARIMA) methods based on accuracy of experimental forecast results. Secondly, results of EMD-EWMArolling are improved further by combining with the results of ARIMArolling and EWMArolling using affine combination. Thus, two forecast combination methods are found. They are affEEArolling (i.e., affine combination of EMD-EWMArolling and ARIMArolling) and affEEErolling (i.e., affine combination of EMD-EWMArolling and EWMArolling).

Proposed methods are optimized in data driven way and using information criterion. Firstly, since EMD is data-driven locally adaptive decomposition algorithm, it disentangles local features in different intervals and frequencies. These EMD features are essentially used by ARIMA and EWMA for optimal model selection through information criterion (e.g., AIC, i.e., Akaike information criterion) and optimizing by reducing sum of square errors. Also, affine parameters are different for each data set and they are updated in each rolling as per data distribution. All these contributed for the overall optimization of the hybridization and combination of the proposed methods.

The two proposed methods (affEEArolling and affEEErolling) including EMD-ARIMArolling and EMD-EWMArolling are described below.

A. *EMD-ARIMArolling, EMD-EWMArolling and Selecting Better One of the Two*

To obtain better forecast results, rolling forecast approach was applied on EMD-based hybridizations EMD-ARIMA and EMD-EWMA. Rolling forecast approach was adopted in EMD-ARIMArolling and EMD-EWMArolling methods to improve the forecast accuracy of EMD-ARIMA and EMD-EWMA methods. In an EMD-ARIMArolling, firstly EMD algorithm is applied on a time series for decomposition. Secondly rolling forecast approach is employed on EMD components while using selected ARIMA models to fit and forecast. Thirdly, all the rolling forecasts of EMD components are summed up to get combined forecast for each rolling window, i.e., sequentially updated and forwarded training set. Fourthly, all the rolling window-based forecasts are aggregated for a complete forecast set; also, error measures (between test dataset and forecast dataset) are computed and performance is compared with other methods. EMD-EWMArolling is similar to EMD-ARIMArolling. On an average and from experimental results, EMD-EWMArolling was empirically expected to be better performer than EMD-ARIMArolling. Hence, EMD-EWMArolling was prioritized in affine combination.

B. *Combined Forecast Methods using Affine Combination*

Generally affine combination for k forecasted datasets $F_1, F_2, F_2, \ldots, F_k$ using k methods $M_{1,}\ M_{2,}\ M_{3,} \ldots, M_k$, can be presented as Equation 10:

$$\sum_i^k c_i F_{i,} \quad where \sum_i^k c_i = 1 \ and \ c_i \in \mathbb{R} \tag{10}$$

where, $\mathbb{R}$ is set of all real number.

For this research study, $k = 2$ was considered from Equation 10 to combine forecast values from two comparatively better methods. If any test dataset $T$is used to fit and forecast using selected model from a method, a forecasted dataset is found along with error. Considering two methods, ε_1 and F_1 are respectively error value and forecast set of method M_1, first method and similarly ε_2 and F_2 are respectively error value and forecast set of method $M_{2,}$second method. Thus, forecast affine combination aff-Comb of the two methods can be expressed as ($\theta \in \mathbb{R}$ is affine parameter):

$$aff - Comb = \theta * F_1 + (1-\theta) * F_2 \tag{11}$$

$$= \theta * (T + \varepsilon_1) + (1-\theta) * (T + \varepsilon_2) \tag{12}$$

[since $T - F_1 = \varepsilon_1$ *and* $T - F_2 = \varepsilon_2$]

$$= T - \theta * \varepsilon_1 - (1-\theta) * \varepsilon_2 \tag{13}$$

The error $-\theta * \varepsilon_1 - (1-\theta) * \varepsilon_2$ after combination as found from Equation 13 should be minimized, i.e., theoretically zero. Thus,

$$-\theta * \varepsilon_1 - (1-\theta) * \varepsilon_2 = 0 \tag{14}$$

From Equation 14 by solving for affine parameter, $\theta = \frac{\varepsilon_2}{\varepsilon_2 - \varepsilon_1}$.

C. *Formulation of affEEArolling and affEEErolling*

The forecast values of affEEArolling method are found from the affine combination of EMD-EWMArolling and ARIMArolling. Therefore,

$$affEEArolling = \theta_1 * EMD - EWMArolling + (1-\theta_1) * ARIMArolling \tag{15}$$

where $\theta_1 = \frac{\varepsilon_2}{\varepsilon_2 - \varepsilon_1}$, ε_1 is the error of EMD-EWMArolling and ε_2 is the error of ARIMArolling, $\theta_1, \varepsilon_1, \varepsilon_2 \in \mathbb{R}$. Similarly, affine forecast formula for affEEErolling method is:

$$affEEErolling = \theta_2 * EMD - EWMArolling + (1-\theta_2) * EWMArolling, \tag{16}$$

where $\theta_2 = \frac{\varepsilon_3}{\varepsilon_3 - \varepsilon_1}$, ε_1 is the error of EMD-EWMArolling and $\varepsilon_3$is the error of EWMArolling, $\theta_1, \varepsilon_1, \varepsilon_3 \in \mathbb{R}$.

Algorithms of both affEEArolling and affEEErolling

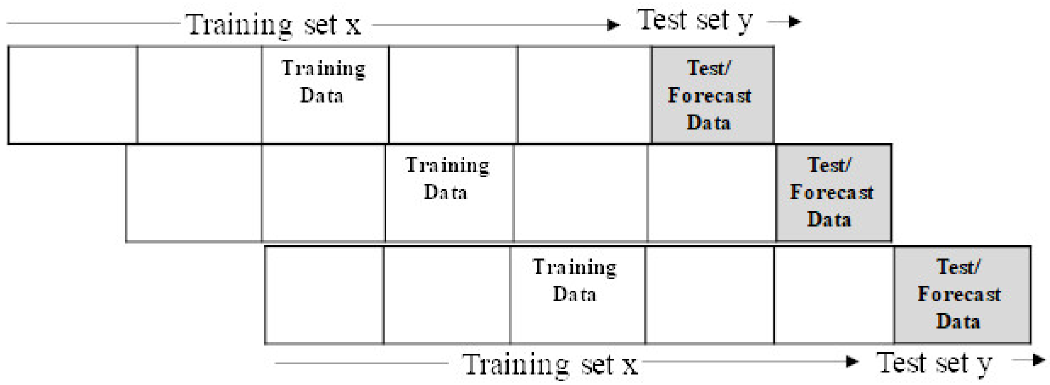

Figure 2. Diagram of rolling forecast

methods being similar, algorithm of affEEArolling is presented here.

- Algorithm: affEEArolling

Input: Time series data X (pre-processed)
Output: Prediction error results of affEEArolling
Step 1: Start Program
Step 2: Read the input series X
Step 3: Define $rollingFUN()$
Step 4: Split X using $rollingFUN(X)$ and store X_tr as training set and X_tt as test set, set h as forecast horizon
Step 5: Define $EMD_EWMA(G, H)$ and $ARIMA(G, H)$ as forecast functions
Step 6: Compute

$$fEE \leftarrow EMD_EWMA(X_tr, h)$$

and $fA \leftarrow ARIMA(X_tr, h)$
Step 7: Formulate $FCastErr(T, F)$
Step 8: Compute errors $\varepsilon_1 \leftarrow FCastErr(X_tt, fEE)$, $\varepsilon_2 \leftarrow FCastErr(X_tt, fA)$
Step 9: Compute $\theta_1 \leftarrow \frac{\varepsilon_2}{\varepsilon_2 - \varepsilon_1}$
Step 10: Compute $FCast \leftarrow \theta_1 * fEE + (1 - \theta_1) * fA$
Step 11: Compute error $\varepsilon \leftarrow FCastErr(X_tt, FCast)$
Step 12: Display error ε as error of affEEArolling
Step 13: End Program

D. Accuracy Measures to Evaluate Prediction Performance

Six error measures or equivalently accuracy measures (i.e., less error indicates better accuracy) are used in this study. These are Root Mean Squared Forecast Error (RMSFE), Mean Absolute Forecast Error (MAFE), Root Mean Squared Relative Forecast Error (RMSRFE) [42], Mean Absolute Percentage Forecast Error (MAPFE), Mean Absolute Scaled Forecast Error (MASFE) and Symmetric Mean Absolute Percentage Forecast Error (sMAPFE) [43]. These error measures or accuracy formulas are presented below using their mathematical forms:

$$RMSFE = \sqrt{\frac{\sum_1^n {e_t}^2}{n}} \tag{17}$$

$$MAFE = \frac{\sum_1^n |e_t|}{n} \tag{18}$$

$$RMSRFE = \sqrt{\frac{\sum_1^n \left(\frac{e_t}{x_t}\right)^2}{n}} \tag{19}$$

$$MAPFE = \frac{\sum_1^n \left|\frac{e_t}{x_t}\right|}{n} \tag{20}$$

$$MASFE = \frac{\sum_1^n \left|\frac{e_t}{\frac{1}{n-1}\sum_2^n |x_t - x_{t-1}|}\right|}{n} \tag{21}$$

$$sMAPFE = \frac{\sum_1^n 200\left|\frac{e_t}{x_t + f_t}\right|}{n} \tag{22}$$

where $e_t = x_t - f_t$ forecast error for a single point data, x_t is actual test value and f_t is forecast value, all at period t.

4. Datasets and Experimental Results

In this section, short details of datasets (i.e., brief description along with graphs and descriptive statistics) and empirical accuracy results of all methods (proposed as well as compared) are presented for all the nine datasets.

A. Datasets

Nine daily closing price time series datasets of nine NASDAQ Financial-100 companies are used in the experiment of this research study. The companies are Arch Capital Group Ltd. (ACGL as stock ticker), BOK Financial Corporation (BOKF), Capitol Federal Financial, Inc. (CFFN), FirstCash, Inc. (FCFS), Fifth Third Bancorp

(FITB), Fulton Financial Corporation (FULT), Investors Bancorp, Inc. (ISBC), Simmons First National Corporation (SFNC) and Washington Federal, Inc. (WAFD). All datasets include 1760 data in each set (from 20th February, 2014 to 16th February, 2021). These data are freely available at NASDAQ website (https://www.nasdaq.com/market-activity/quotes/nasdaq-financial-100-stocks) and also at yahoo finance website (https://finance.yahoo.com/). Selecting training and test datasets from a dataset $X = \{x_1,\ x_2, \ldots, x_1 760\}$ for rolling approach can be presented with the following mathematical expressions:

$$S_i \in \{x_{(i-1)h+1},\ x_{(i-1)h+2}, \ldots, x_{(i-1)h+d}\} \tag{23}$$

$$T_i \in \{x_{(i-1)h+d+1},\ x_{(i-1)h+d+2}, \ldots,\ x_{(i-1)h+d+k}\} \tag{24}$$

where $i$is index for a window, $S_i$is training set, $T_i$is test set, d is window size of training set, h is shifting or window moving size, k is forecast horizon or forecast size and $i, d, h,\ k \in \mathbb{N}$(where$\mathbb{N}$ presents set of all positive integers). This paper implemented window size d= 1600; shifting or moving size h=40 and rolling forecast horizon k=40. In the first rolling window, 1600 training data (S_1) are from 20th February, 2014 (x_1) to 26th June, 2020 (x_{1600}) and 40 test data (T_1) are from 29th June, 2020 (x_{1601}) to 24th August, 2020(x_{1640}). Training and test datasets in other rolling windows are distributed similarly.

Each time series has distinctive patterns and features. Quick tools to grab some features are line graph presentation and summary or descriptive statistics. Figure 3 contains the line graphs of nine stock price datasets. Also, Table I contains descriptive statistics of these datasets which involves 1600 training data for each of nine datasets. It presents different characteristics of involving datasets namely measure of statistical average or central tendency, measure of spread or dispersion, measures of normality (e.g., skewness and kurtosis). Table I includes mean and median as measures of central tendency; minimum (Min), maximum (Max) and coefficient of variation (COV) as measures of dispersion; skewness (Skew) for degree of symmetry and kurtosis (Kurt) for degree of peakedness both as shape characteristics or measures of normality. BOKF dataset has the highest mean (74.655) and median (75.03) values while ISBC has the lowest mean (11.985) and median (11.95) values among the nine datasets. CFFN has the least COV (0.08) while FCFS has highest COV (0.294). FULT, ISBC and SFNC are negatively skewed where degree or strength of skewness of ISBC (-0.471) is highest in the negative. Also, CFFN has highest degree of skewness (i.e., highly positively skewed) with value of 0.665. CFFN and ISBC have positive kurtosis where CFFN has the highest degree of peakedness (0.809) while WAFD (-1.562) has the flattest kurtosis among the remaining with negative kurtosis. The least skewed or highest symmetric data is WAFD (0.015) and the closest to mesokurtic data is ACGL (-0.03). These characteristics reflect the general and overall data feature. However, local as well as short term feature can be different and rolling with inclusion of most recent and exclusion of far distant data will also be more different yet more relevant. Also, since EMD disentangles local distinctive features reserving overall data patterns as a whole and accompanying hybridization component methods (here ARIMA and EWMA) use these features for suitable model selection, affine combinations characteristically minimize the error of proposed methods with the use of errors of accompanying methods. Thus, proposed methods along with affine combinations can feasibly reduce errors with the three improvement stages including data-adaptive feature of EMD for varied and sophisticated distributions of nonlinear and nonstationary data.

In the EMD process, each time series data is decomposed into components from high frequency stationary to gradually low frequency non-stationary. EMD components Graphs of all datasets being similar, the graphs of Arch Capital Group Ltd. (ACGL) are presented in Figure 4.

B. *Experimental Results*

Table II-VIII reflect the forecast accuracy of experimental results from eight methods employed in the study. These tables encompass results for nine stock price datasets (ACGL, BOKF, CFFN, FCFS, FITB, FULT, ISBC, SFNC and WAFD). Smaller values in forecast error are the indication of better performance or accuracy. The tables present results of forecast error measures RMSFE and MAFE as absolute error measures and RMSRFE, MAPFE, MASFE and sMAPFE as relative errors measures. Six out of the eight methods are used for comparison purpose. These methods include four benchmark methods (i.e., ARIMA, EWMA, ARIMArolling and EWMArolling), two hybridized EMD-based rolling forecast methods (i.e., EMD-ARIMArolling and EMD-EWMArolling) and two proposed methods (i.e., affEEArolling and affEEErolling).

5. Discussion and Study Outcome

In the performance comparison of methods, smaller errors in the prediction results are the reflections of better accuracy of a method. Considering both the absolute errors (RMSFE and MAFE) and relative errors (RMSRFE, MAPFE, MASFE and sMAPFE) presented in the tables (Table II-X) representing results, noticeably the proposed methods affEEArolling and affEEErolling are better performing methods than the six other compared methods. The nine datasets presented here have similar results and reflections on the methods. Inherent reasonable assumption is that performance of traditional ARIMA and EWMA methods are improved here in three improvement stages. These are adoption of rolling forecast, EMD-based hybridization and affine combination.

The experimental results show that rolling forecast approach can significantly improve forecast accuracy of traditional forecast approach. For ACGL dataset (Table II), this approach improved RMSFE of EWMA from 3.941 to 2.217 (i.e., error reduced by 43.7%); also, it improved MAFE of the EWMA from 3.351 to 1.733. Similar improvement also occurred in relative errors RMSRFE, MAPFE, MASFE

TABLE I. Descriptive statisitcs of nine daily stock price datasets (for first rolling window)

Dataset	Mean	Median	Min	Max	COV	Skew	Kurt	Count
ACGL	27.78	27.117	17.767	48.18	0.242	0.641	-0.03	1600
BOKF	74.655	75.03	35.4	105.26	0.192	0.08	-0.599	1600
CFFN	13.184	13.15	10.19	16.92	0.08	0.665	0.809	1600
FCFS	63.322	57.41	30.22	106.25	0.294	0.465	-1.083	1600
FITB	23.966	24.395	11.67	34.35	0.202	0.089	-1.087	1600
FULT	15.089	15.60	9.36	19.70	0.174	-0.085	-1.423	1600
ISBC	11.985	11.95	6.96	15.01	0.118	-0.471	0.610	1600
SFNC	24.558	24.675	14.51	33.35	0.16	-0.064	-0.748	1600
WAFD	28.422	27.915	19.67	38.14	0.194	0.015	-1.562	1600

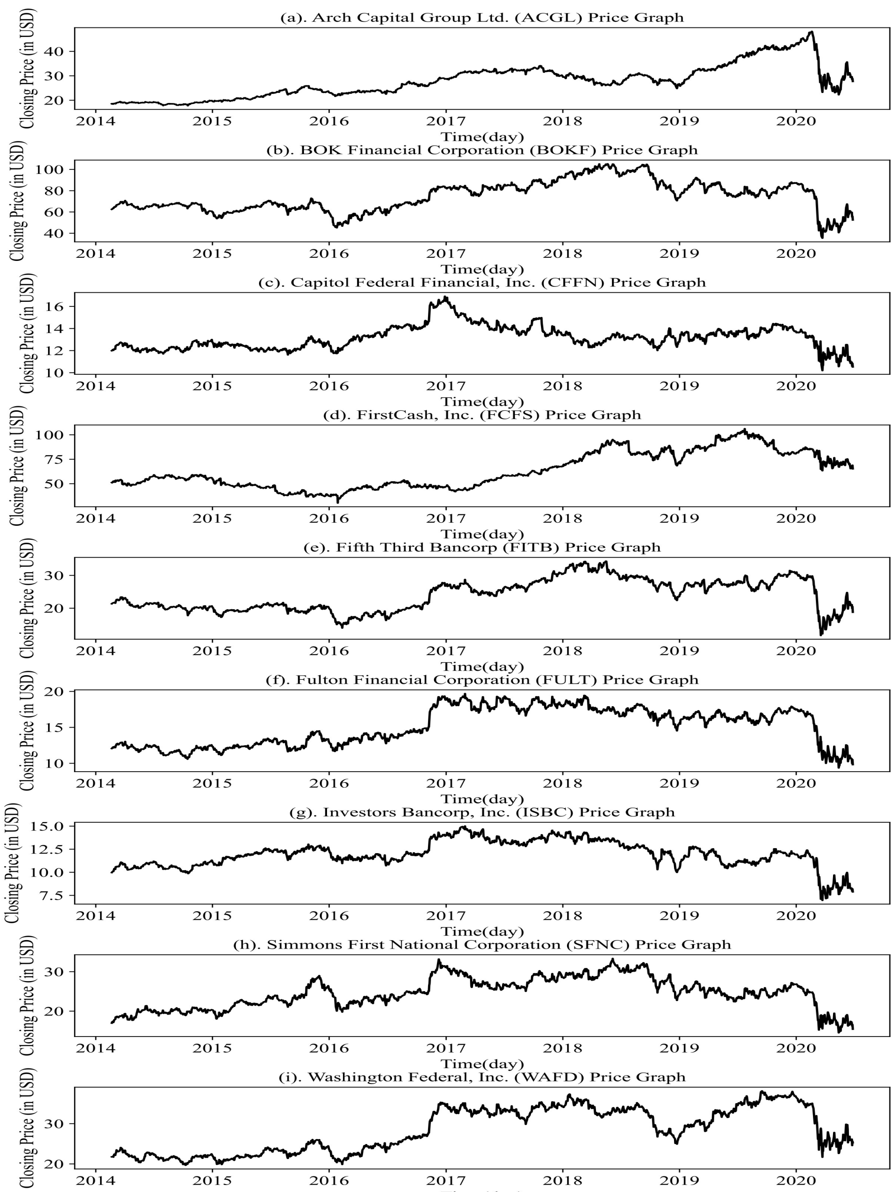

Figure 3. Daily closing price graphs of nine NASDAQ Financial-100 companies

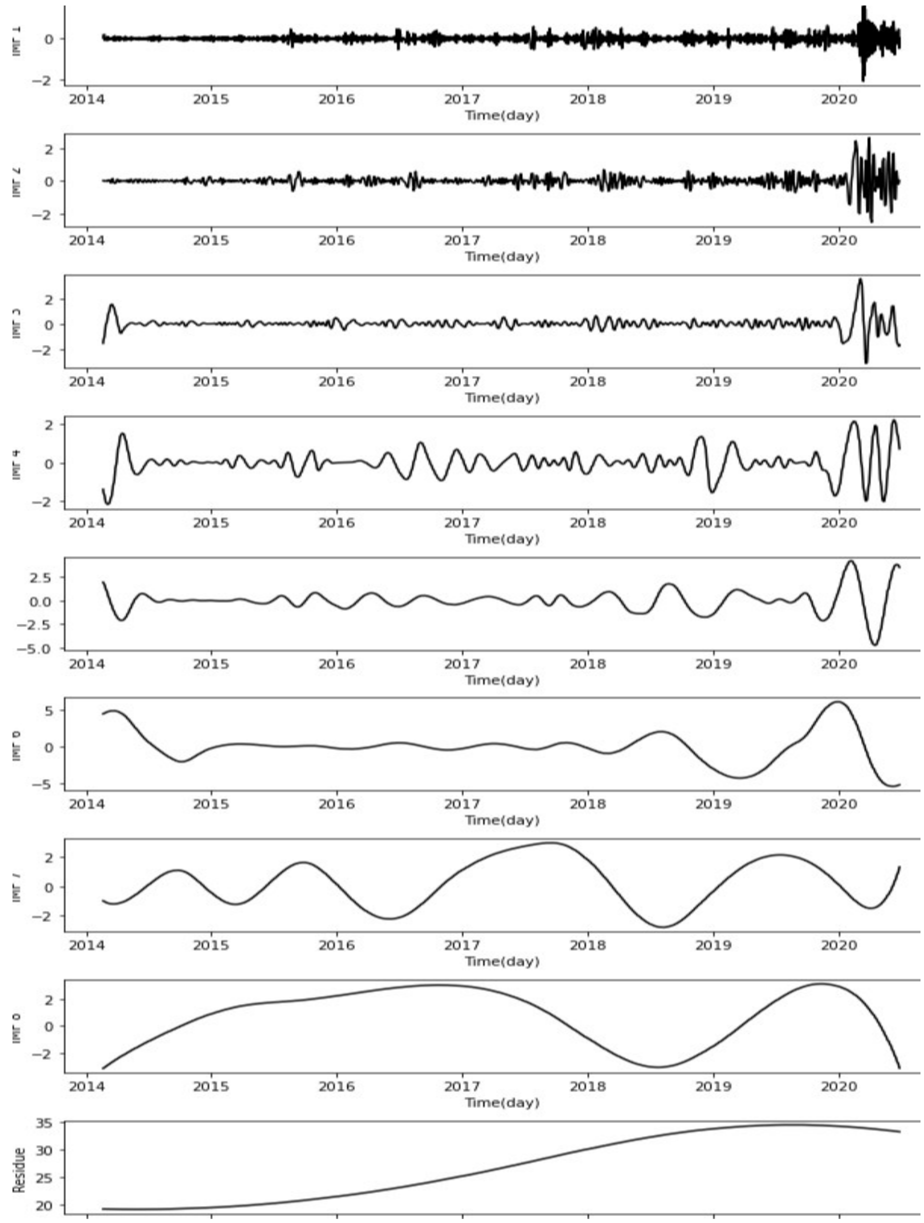

Figure 4. EMD components graphs of Arch Capital Group Ltd. (ACGL) closing price

TABLE II. ACGL stock price forecast accuracy results for 160 days

Methods	RMSFE	MAFE	RMSRFE	MAPFE	MASFE	sMAPFE
ARIMA	4.200	3.612	1.234	10.850	6.958	11.672
EWMA	3.941	3.351	1.155	10.055	6.455	10.768
ARIMArolling	2.375	1.844	0.722	5.696	3.552	5.889
EWMArolling	2.217	1.733	0.674	5.358	3.338	5.514
EMD-ARIMArolling	3.739	2.623	1.180	8.238	5.052	8.905
EMD-EWMArolling	2.123	1.683	0.647	5.201	3.242	5.343
affEEArolling	1.735	1.422	0.545	4.467	2.739	4.481
affEEErolling	1.731	1.415	0.545	4.453	2.726	4.465

TABLE III. BOKF stock price forecast accuracy results for 160 days

Methods	RMSFE	MAFE	RMSRFE	MAPFE	MASFE	sMAPFE
ARIMA	11.966	8.709	1.611	12.394	6.984	13.766
EWMA	11.793	8.583	1.589	12.232	6.882	13.544
ARIMArolling	7.249	5.373	1.04	8.01	4.309	8.511
EWMArolling	7.213	5.352	1.036	7.985	4.292	8.472
EMD-ARIMArolling	10.52	8.507	1.504	12.97	6.822	13.485
EMD-EWMArolling	7.18	5.308	1.029	7.895	4.256	8.388
affEEArolling	5.042	4.144	0.737	6.28	3.323	6.456
affEEErolling	5.047	4.135	0.737	6.265	3.316	6.441

TABLE IV. CFFN stock price forecast accuracy results for 160 days

Methods	RMSFE	MAFE	RMSRFE	MAPFE	MASFE	sMAPFE
ARIMA	1.442	1.283	1.253	11.316	7.828	11.531
EWMA	1.440	1.282	1.253	11.321	7.825	11.526
ARIMArolling	0.867	0.673	0.759	5.923	4.105	6.055
EWMArolling	0.867	0.672	0.76	5.912	4.098	6.045
EMD-ARIMArolling	1.911	1.755	1.783	16.124	10.707	15.765
EMD-EWMArolling	0.862	0.664	0.749	5.805	4.049	5.961
affEEArolling	0.685	0.505	0.603	4.455	3.083	4.565
affEEErolling	0.678	0.497	0.597	4.385	3.031	4.493

TABLE V. FCFS stock price forecast accuracy results for 160 days

Methods	RMSFE	MAFE	RMSRFE	MAPFE	MASFE	sMAPFE
ARIMA	8.158	7.149	1.406	12.045	6.421	11.156
EWMA	8.21	7.198	1.415	12.128	6.465	11.227
ARIMArolling	7.728	6.098	1.276	9.962	5.477	9.421
EWMArolling	7.704	6.094	1.272	9.959	5.474	9.420
EMD-ARIMArolling	9.707	8.112	1.625	13.41	7.286	13.588
EMD-EWMArolling	7.609	6.039	1.253	9.856	5.424	9.362
affEEArolling	5.885	4.655	0.928	7.459	4.181	7.451
affEEErolling	5.755	4.485	0.909	7.197	4.028	7.186

TABLE VI. FITB stock price forecast accuracy results for 160 days

Methods	RMSFE	MAFE	RMSRFE	MAPFE	MASFE	sMAPFE
ARIMA	5.607	4.228	1.941	15.636	8.548	17.701
EWMA	5.578	4.205	1.931	15.558	8.501	17.59
ARIMArolling	2.637	2.136	1	8.485	4.318	8.889
EWMArolling	2.639	2.138	1.001	8.495	4.321	8.892
EMD-ARIMArolling	4.967	4.037	1.808	15.689	8.161	17.25
EMD-EWMArolling	2.593	2.053	0.962	8.039	4.15	8.438
affEEArolling	1.752	1.422	0.66	5.593	2.876	5.762
affEEErolling	1.751	1.421	0.659	5.577	2.872	5.745

TABLE VII. FULT stock price forecast accuracy results for 160 days

Methods	RMSFE	MAFE	RMSRFE	MAPFE	MASFE	sMAPFE
ARIMA	1.989	1.568	1.503	12.631	6.290	13.530
EWMA	2.051	1.566	1.533	12.450	6.282	13.505
ARIMArolling	1.329	0.987	1.057	8.209	3.959	8.655
EWMArolling	1.299	0.945	1.024	7.794	3.793	8.245
EMD-ARIMArolling	2.119	1.795	1.835	15.529	7.203	15.253
EMD-EWMArolling	1.294	0.918	1.012	7.489	3.682	7.962
affEEArolling	0.857	0.666	0.673	5.463	2.671	5.601
affEEErolling	0.854	0.657	0.666	5.371	2.636	5.517

TABLE VIII. ISBC stock price forecast accuracy results for 160 days

Methods	RMSFE	MAFE	RMSRFE	MAPFE	MASFE	sMAPFE
ARIMA	1.812	1.26	1.629	12.149	7.181	13.492
EWMA	1.811	1.259	1.628	12.148	7.179	13.488
ARIMArolling	1.099	0.750	1.065	7.549	4.273	8.131
EWMArolling	1.098	0.747	1.065	7.526	4.261	8.104
EMD-ARIMArolling	1.771	1.521	1.855	16.291	8.673	16.302
EMD-EWMArolling	1.083	0.741	1.046	7.446	4.222	8.009
affEEArolling	0.813	0.577	0.786	5.857	3.291	6.072
affEEErolling	0.813	0.573	0.785	5.799	3.268	6.016

TABLE IX. SFNC stock price forecast accuracy results for 160 days

Methods	RMSFE	MAFE	RMSRFE	MAPFE	MASFE	sMAPFE
ARIMA	4.329	2.891	1.777	12.838	7.078	14.619
EWMA	4.438	2.987	1.824	13.269	7.312	15.181
ARIMArolling	2.578	1.825	1.129	8.494	4.467	9.116
EWMArolling	2.574	1.824	1.127	8.487	4.465	9.11
EMD-ARIMArolling	7.099	4.415	2.856	19.693	10.809	24.691
EMD-EWMArolling	2.573	1.821	1.123	8.48	4.459	9.074
affEEArolling	1.904	1.515	0.866	7.279	3.709	7.532
affEEErolling	1.884	1.502	0.859	7.221	3.676	7.472

TABLE X. WAFD stock price forecast accuracy results for 160 days

Methods	RMSFE	MAFE	RMSRFE	MAPFE	MASFE	sMAPFE
ARIMA	2.576	2.165	1.126	9.175	4.811	8.733
EWMA	2.449	2.018	1.059	8.511	4.485	8.166
ARIMArolling	1.964	1.611	0.792	6.538	3.581	6.574
EWMArolling	1.966	1.604	0.792	6.51	3.565	6.56
EMD-ARIMArolling	3.256	2.844	1.319	11.55	6.321	11.249
EMD-EWMArolling	1.936	1.559	0.777	6.312	3.465	6.374
affEEArolling	1.518	1.267	0.616	5.154	2.816	5.183
affEEErolling	1.504	1.248	0.61	5.071	2.774	5.102

and sMAFE for ACGL dataset in case of improvement from EWMA to EWMArolling. Such improvements in forecast accuracy or error reduction occurred for other datasets BOKF, CFFN, FCFS, FITB, FULT, ISBC, SFNC and WAFD (Table III-X). Also, in the similar way forecast errors of ARIMA was improved using ARIMArolling.

The second forecast approach or EMD-based hybridization further improved the forecast accuracy. Considering ACGL dataset (Table II), EMD-EWMArolling improved RMSFE of EWMArolling by 4.24% (from 2.217 to 2.123) and MAFE from 1.733 to 1.683. Accordingly, RMSRFE, MAPFE, MASFE and sMAPFE of of EWMArolling were significantly reduced by EMD-EWMArolling. Similarly, improved results are also found for other datasets (Table III-X). Therefore, application of divide-and-conquer principle using locally adaptive algorithm EMD was found effective for forecast improvement. Although accuracy of ARIMA was improved by ARIMArolling, in many cases EMD-ARIMArolling may not improve accuracy of ARIMArolling rather deteriorate accuracy which is also noticeable here (Table III-X) in this study. However, there are exceptions.

The accuracy of EMD-based rolling forecast was further improved by employing affine combination approach. From ACGL dataset (Table II), both affEEArolling and affEEErolling improved the accuracy of EMD-EWMArolling method. The affine combination approach affEEArolling improved RMSFE of EMD-EWMArolling by 18.28% (from 2.123 to 1.735) while affEEErolling improved the same from 2.123 to 1.73. Also, MAFE, RMSRFE, MASFE and sMAPFE of EMD-EWMArolling method were significantly reduced by both affEEArolling and affEEErolling. Such forecast accuracy improvements are found for all other datasets (Table III-X).

Overall, all the three improvement stages significantly improved the forecast accuracy. Considering ACGL dataset (Table II), both affEEArolling and affEEErolling significantly reduced the errors of ARIMA and EWMA methods. For absolute errors of EWMA, affEEArolling method reduced RMSFE from 3.941 to 1.735 (improved by 55.98%) while affEEErolling method reduced RMSFE from 3.941 to 1.731 (i.e., 56% reduction); affEEArolling method reduced MAFE from 3.351 to 1.422 (i.e., 57.56% reduction) while affEEErolling method reduced MAFE from 3.351 to 1.415 (improved by 57.77%). Similarly, affEEArolling method reduced RMSRFE from 1.155 to 0.545 while affEEErolling method reduced RMSRFE from 1.155 to 0.545; affEEArolling method reduced MAPFE from 10.055 to 4.467 while affEEErolling method reduced MAPFE from 10.055 to 4.453. Also, affEEArolling method reduced MASFE from 6.455 to 2.739 while affEEErolling method reduced MASFE from 6.455 to 2.726; affEEArolling method reduced sMAPFE from 10.768 to 4.481 while affEEErolling method reduced sMAPFE from 10.768 to 4.465. Among all the eight methods, affEEErolling method was found best and affEEArolling was second best. Such forecast error reductions or accuracy improvements are true for all other datasets (Table III-X).

Single benchmark methods are generally useful. However, they are limited by theoretical and foundational assumptions and developments, i.e., linearity and stationarity. In case of real-life data which are mostly complicated, nonlinear and nonstationary, these methods may not capture essential features of data. Therefore, in such cases any suitable hybrid or combination method, where component methods or approaches are best suited, can produce significantly improved forecast accuracy. Considering effectiveness and usefulness of a forecast method, EMD-based hybrid methods along with rolling forecast and affine combination can be very practicable. Based on the results of nine datasets and six error measures, proposed EMD-based affEEArolling and affEEErolling methods were better performing methods than ARIMA, EWMA, ARIMArolling, EWMArolling, EMD-ARIMArolling and EMD-EWMArolling. Presumably EMD being locally adaptive and efficient decomposition algorithm, such other not yet investigated EMD-based hybrid and combination approaches deserve further research attention.

From empirical forecast results along with complementary discussion, the outcome of the study is that due to improvement stages, EMD-based affEEArolling and affEEErolling methods were better than other compared methods. Moreover, our assumption is that such approaches can be applied on time series data of other fields related to forecasting. Savvy and sophisticated practitioners including stock, forex or other market traders, investors and others can include such methods in their forecast and analysis toolbox. Also, financial technologies and predictive analytics software can encompass such methods for improved forecast accuracy.

6. Concluding Remarks and Future Work

In many cases regarding datasets of different research domains, benchmark methods like ARIMA and smoothing methods performs satisfactorily. However, for sophisticated datasets involving non-stationarity and non-linearity properties, they fail to produce better or outperforming results. This research study employed EMD-based divide-and-conquer approach to improve results in such data characteristics. Rolling forecast approach was employed to use latest available information which improved accuracy. ARIMArolling and EWMArolling performed better than ARIMA and EWMA. Furthermore, EMD-EWMArolling even performed better although EMD-ARIMArolling did not. However, proposed affEEArolling and affEEErolling were best performing out of the eight methods. Hence, the hypothesis is that EMD-based hybridizations and forecast combinations with statistical methods can be quite useful and they deserve further study. Our future studies might include such research scopes with EMD-based forecasting methods which can be extended to machine learning methods and other predictive data analytics domain.

References

[1] S. Makridakis, "The art and science of forecasting an assessment and future directions," *International Journal of forecasting*, vol. 2, no. 1, pp. 15–39, 1986.

[2] J. G. De Gooijer and R. J. Hyndman, "25 years of time series forecasting," *International journal of forecasting*, vol. 22, no. 3, pp. 443–473, 2006.

[3] R. J. Hyndman, "A brief history of forecasting competitions," *International Journal of Forecasting*, vol. 36, no. 1, pp. 7–14, 2020.

[4] A. Hernandez-Matamoros, H. Fujita, T. Hayashi, and H. Perez-Meana, "Forecasting of covid19 per regions using arima models and polynomial functions," *Applied Soft Computing*, vol. 96, p. 106610, 2020.

[5] M. A. Şahinli, "Potato price forecasting with holt-winters and arima methods: A case study," *American Journal of Potato Research*, vol. 97, no. 4, pp. 336–346, 2020.

[6] M. L. Challa, V. Malepati, and S. N. R. Kolusu, "S&p bse sensex and s&p bse it return forecasting using arima," *Financial Innovation*, vol. 6, no. 1, pp. 1–19, 2020.

[7] J. Qin, Z. Tao, S. Huang, and G. Gupta, "Stock price forecast based on arima model and bp neural network model," in *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*. IEEE, 2021, pp. 426–430.

[8] L. J. Jong, S. Ismail, A. Mustapha, M. H. Abd Wahab, and S. Z. S. Idrus, "The combination of autoregressive integrated moving average (arima) and support vector machines (svm) for daily rubber price forecasting," in *IOP Conference Series: Materials Science and Engineering*, vol. 917, no. 1. IOP Publishing, 2020, p. 012044.

[9] P. Nath, P. Saha, A. I. Middya, and S. Roy, "Long-term time-series pollution forecast using statistical and deep learning methods," *Neural Computing and Applications*, pp. 1–20, 2021.

[10] D. Song, A. M. C. Baek, and N. Kim, "Forecasting stock market indices using padding-based fourier transform denoising and time series deep learning models," *IEEE Access*, 2021.

[11] N. E. Huang, M.-L. Wu, W. Qu, S. R. Long, and S. S. Shen, "Applications of hilbert–huang transform to non-stationary financial time series analysis," *Applied stochastic models in business and industry*, vol. 19, no. 3, pp. 245–268, 2003.

[12] H. Wang, L. Liu, S. Dong, Z. Qian, and H. Wei, "A novel work zone short-term vehicle-type specific traffic speed prediction model through the hybrid emd–arima framework," *Transportmetrica B: Transport Dynamics*, vol. 4, no. 3, pp. 159–186, 2016.

[13] Z.-Y. Wang, J. Qiu, and F.-F. Li, "Hybrid models combining emd/eemd and arima for long-term streamflow forecasting," *Water*, vol. 10, no. 7, p. 853, 2018.

[14] L. Chen, Y. Chi, Y. Guan, and J. Fan, "A hybrid attention-based emd-lstm model for financial time series prediction," in *2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD)*. IEEE, 2019, pp. 113–118.

[15] B. Alhnaity and M. Abbod, "A new hybrid financial time series prediction model," *Engineering Applications of Artificial Intelligence*, vol. 95, p. 103873, 2020.

[16] A. Vlasenko, Y. Rashkevych, N. Vlasenko, D. Peleshko, and O. Vynokurova, "A hybrid emd-neuro-fuzzy model for financial time series analysis," in *2020 IEEE Third International Conference on Data Stream Mining & Processing (DSMP)*. IEEE, 2020, pp. 112–115.

[17] M. R. Hossain and M. T. Ismail, "Empirical mode decomposition based on theta method for forecasting daily stock price," *Journal of Information and Communication Technology*, vol. 19, no. 4, pp. 533–558, 2020.

[18] Y. Lin, Y. Yan, J. Xu, Y. Liao, and F. Ma, "Forecasting stock index price using the ceemdan-lstm model," *The North American Journal of Economics and Finance*, vol. 57, p. 101421, 2021.

[19] Y. Pu and K. Wu, "Combined model of stock index prediction based on improved extreme mirror extension-empirical mode decomposition-support vector regression," in *2021 International Conference on Big Data Analysis and Computer Science (BDACS)*. IEEE, 2021, pp. 17–21.

[20] M. R. Hossain, M. T. Ismail, and S. A. B. A. Karim, "Improving stock price prediction using combining forecasts methods," *IEEE Access*, vol. 9, pp. 132 319–132 328, 2021.

[21] Y. Huang, C. Deng, X. Zhang, and Y. Bao, "Forecasting of stock price index using support vector regression with multivariate empirical mode decomposition," *Journal of Systems and Information Technology*, 2020.

[22] H. Rezaei, H. Faaljou, and G. Mansourfar, "Stock price prediction using deep learning and frequency decomposition," *Expert Systems with Applications*, vol. 169, p. 114332, 2021.

[23] S.-L. Lin, "Application of empirical mode decomposition to improve deep learning for us gdp data forecasting," *Heliyon*, p. e08748, 2022.

[24] A. F. Atiya, "Why does forecast combination work so well?" *International Journal of Forecasting*, vol. 36, no. 1, pp. 197–200, 2020.

[25] S. Jaganathan and P. Prakash, "A combination-based forecasting method for the m4-competition," *International Journal of Forecasting*, vol. 36, no. 1, pp. 98–104, 2020.

[26] M. Pawlikowski and A. Chorowska, "Weighted ensemble of statistical models," *International Journal of Forecasting*, vol. 36, no. 1, pp. 93–97, 2020.

[27] F. Petropoulos and I. Svetunkov, "A simple combination of univariate models," *International journal of forecasting*, vol. 36, no. 1, pp. 110–115, 2020.

[28] M. E. Thomson, A. C. Pollock, D. Önkal, and M. S. Gönül, "Combining forecasts: Performance and coherence," *International Journal of Forecasting*, vol. 35, no. 2, pp. 474–484, 2019.

[29] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The m4 competition: 100,000 time series and 61 forecasting methods," *International Journal of Forecasting*, vol. 36, no. 1, pp. 54–74, 2020.

[30] S. Yuan, C. Wang, B. Mu, F. Zhou, and W. Duan, "Typhoon intensity forecasting based on lstm using the rolling forecast method," *Algorithms*, vol. 14, no. 3, p. 83, 2021.

[31] T. L. Zeller, L. M. Metzger *et al.*, "Good bye traditional budgeting,

hello rolling forecast: has the time come?" *American Journal of Business Education (AJBE)*, vol. 6, no. 3, pp. 299–310, 2013.

[32] S. Jiang, C. Yang, J. Guo, and Z. Ding, "Arima forecasting of china's coal consumption, price and investment by 2030," *Energy Sources, Part B: Economics, Planning, and Policy*, vol. 13, no. 3, pp. 190–195, 2018.

[33] M. Ohyver and H. Pudjihastuti, "Arima model for forecasting the price of medium quality rice to anticipate price fluctuations," *Procedia Computer Science*, vol. 135, pp. 707–711, 2018.

[34] M. H. Alsharif, M. K. Younes, and J. Kim, "Time series arima model for prediction of daily and monthly average global solar radiation: The case study of seoul, south korea," *Symmetry*, vol. 11, no. 2, p. 240, 2019.

[35] J. W. Miller, "Arima time series models for full truckload transportation prices," *Forecasting*, vol. 1, no. 1, pp. 121–134, 2019.

[36] M. Akpinar and N. Yumusak, "Day-ahead natural gas forecasting using nonseasonal exponential smoothing methods," in *2017 IEEE International Conference on Environment and Electrical Engineering and 2017 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*. IEEE, 2017, pp. 1–4.

[37] H. Liu, C. Li, Y. Shao, X. Zhang, Z. Zhai, X. Wang, X. Qi, J. Wang, Y. Hao, Q. Wu *et al.*, "Forecast of the trend in incidence of acute hemorrhagic conjunctivitis in china from 2011–2019 using the seasonal autoregressive integrated moving average (sarima) and exponential smoothing (ets) models," *Journal of infection and public health*, vol. 13, no. 2, pp. 287–294, 2020.

[38] Q. T. Tran, L. Hao, and Q. K. Trinh, "A comprehensive research on exponential smoothing methods in modeling and forecasting cellular traffic," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 23, p. e5602, 2020.

[39] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[40] K. He, H. Wang, J. Du, and Y. Zou, "Forecasting electricity market risk using empirical mode decomposition (emd)—based multiscale methodology," *Energies*, vol. 9, no. 11, p. 931, 2016.

[41] I. Bou-Hamad and I. Jamali, "Forecasting financial time-series using data mining models: A simulation study," *Research in International Business and Finance*, vol. 51, p. 101072, 2020.

[42] S. I. Alzahrani, I. A. Aljamaan, and E. A. Al-Fakih, "Forecasting the spread of the covid-19 pandemic in saudi arabia using arima prediction model under current public health interventions," *Journal of infection and public health*, vol. 13, no. 7, pp. 914–919, 2020.

[43] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International journal of forecasting*, vol. 22, no. 4, pp. 679–688, 2006.

Mohammad Raquibul Hossain Mohammad Raquibul Hossain is an associate professor in the department of Applied Mathematics, Noakhali Science and Technology University, Bangladesh. He received his M.S. degree in Applied Mathematics from University of Dhaka, Bangladesh. He is currently pursuing the Ph.D. degree at School of Mathematical Sciences, Universiti Sains Malaysia, Malaysia. His research interest includes time series forecasting, predictive data analytics, Applied Statistics, Data Science and Machine Learning.

Mohd Tahir Ismail Mohd Tahir Ismail is an Associate Professor and researcher at the School of Mathematical Sciences, Universiti Sains Malaysia. His research interests are financial time series, econometrics, categorical data analysis, and applied statistics. He has published more than 150 publications in reviewed journals and proceedings (some of them are listed in ISI, Scopus, Zentralblatt, MathSciNet, and other indices). Currently, he is an Exco Member of the Malaysian Mathematical Sciences Society and an active member of other scientific professional bodies.

Md. Jamal Hossain Md. Jamal Hossain is an associate professor in the department of Applied Mathematics, Noakhali Science and Technology University, Bangladesh. He completed his M.S. degree in Applied Mathematics from University of Dhaka, Bangladesh. Currently he is doing his PhD at School of Mathematical Sciences, Universiti Sains Malaysia, Malaysia. His research interests include Applied Statistics, Financial Time Series and Economics.