Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v17n1p172

A new hybrid approach for forecasting of daily stock market time series data

By Awajan et al.

15 March 2024

Electronic Journal of Applied Statistical Analysis
Vol. 17, Issue 01, March 2024, 172-190
DOI: 10.1285/i20705948v17n1p172

A new hybrid approach for forecasting of daily stock market time series data

Ahmad M. Awajan*[a], Feras M. AL Faqih[a], Mohd Tahir Ismail[b], Bilal N. Al-Hasanat[a], Mohammed Z Swalmeh[c], and Sadam Al Wadi[d]

[a]*Department of Mathematics, Al Hussien bin Talal University, 71111 Ma'an, Jordan*
[b]*School of Mathematical Sciences, Universiti Sains Malaysia,11800 Penang, Malaysia*
[c]*Faculty of Arts and Sciences, Aqaba University of Technology, Aqaba 77110, Jordan*
[d]*Department of Finance, School of Business, The University of Jordan, Aqaba 77110, Jordan*

15 March 2024

In recent years, many researchers have focused on forecasting financial time series data, especially stock market data. Stock market data possesses so many features that forecasting may be very challenging. In the present study, a hybrid of two methodologies is proposed, which is the Empirical Mode Decomposition (EMD) and the Random Walk (RW) to enhance the stock market forecasting performance, denoted by (EMD-RW). The advantage of EMD-RW is its ability to forecast nonlinear and nonstationary stock market data without the need to use some transformation method or differencing a time series technique. Moreover, the new proposed EMD-RW produced high-accuracy results. Ten stock market time series for ten different countries are used in this study to demonstrate the forecasting accuracy of the EMD-RW. Results using four forecasting accuracy functions display that EMD-RW forecasting accuracy is better than the four compared methods.

keywords: Nonstationary time series, EMD, forecasting.

1 Introduction

As Castle et al. (2019) described, forecasting is a statement about the future. This future outcome is uncertain and needs an appropriate statistical approach to measure

*Corresponding authors: ahmad.m.awajan@ahu.edu.jo

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

it correctly. However, mistakes can happen as the data has many statistical characteristics so one statistical model cannot capture these attributes. Thus, many previous researchers introduced some hybrid forecasting methods to accommodate these changes to improve forecasting accuracy. Moreover, time series forecasting relies on recognizing the historical and current data values and studying the general trend of data. Besides, the next forecasting observations of the time series are usually determined by a specific hypothesis based on the experience, knowledge, and judgment of researchers.

One of the crucial issues in financial time series analysis is the forecasting of financial time data specifically in a stock market index Yang and Lin (2017). Lin et al. (2012) suggested, 50 years ago, that forecasting the financial time series data is concerned as one of the most difficult areas of research. Forecasting stock market data is difficult because of several reasons, including this time series data, is nonlinear and nonstationary Wang et al. (2015). Moreover, stock market data display high heteroscedasticity Kazem et al. (2013) and behave in a random walk manner Patel et al. (2015). Nonetheless, most forecasting techniques reviewed in the literature are based on the linear and stationary attributes of data Oh et al. (2009). Usually, the transformation of an economic time series data, instead of its original scale, is considered when describing its dynamics. A suitable transformation is requisite to transform the non-stationary process into a stationary process. When a process is stationary, most of its mathematical and statistical properties can be utilized for modeling or forecasting purposes Mills (2019).

There are many methods to handle or remove the nonlinear and non-stationary behavior with heteroscedasticity features in the data series. Several of these methods are applied to the transformation of data from the time domain into the frequency domain, the difference of time series, and decomposition techniques (Awajan et al., 2024). Unluckily, there are some issues in using the transformation and the different techniques, such as losing some statistical properties of the original data (Parsons et al., 2000). For this reason, it is preferable to use the decomposition data technique. In the literature, several decomposition techniques have been introduced. Based on the recent researchers' comparison, one of the superior techniques in this field is known as the Empirical Mode Decomposition (EMD) by Huang et al. (1998). EMD is a signal decomposition technique to break down non-stationary and nonlinear time series data, such as economic data. EMD is based on a simple set of mathematical processes and the local characteristic time scale for the time series. Also, EMD is adaptive and highly efficient in keeping the time domain. The EMD has been deeply employed in forecasting time series data in many fields (Awajan et al., 2019; Le An et al., 2005).

Commonly, hybrid forecasting methods gain strengths from single forecasting methods to obtain more accurate forecasting results (Ismail and Awajan, 2017). Lately, several studies have combined the EMD with a traditional forecasting model. Such studies are Nava et al. (2018); Awajan et al. (2018); Chowdhury et al. (2019); Wang et al. (2019); Büyükşahin and Ertekin (2019). In all the studies, EMD decomposes the time series data into Intrinsic Mode Functions (IMFs) and residual components. After that, each decomposed component is forecasted using a time series model. Then all these forecasted values were aggregated to generate the final forecasted value of the original time series. In Nava et al. (2018), the EMD-SVR (a hybrid of EMD with support vector regression

(SVR)) was applied to high-frequency data of the Standard & Poor's 500 Index. The EMD-SVR has gotten a significantly better result as compared to the benchmark methods. As contained in Awajan et al. (2018), a hybrid EMD with HW (Holt-Winter) and the moving block bootstrap (bagging) forecasting method was proposed and applied to the stock market in six countries. The results show that the EMD-HW bagging method is better than the fourteen methods being compared using five error measurements. A hybrid EMD with RF (Random Forest algorithm) forecasting model has been presented in Chowdhury et al. (2019). The EMD-RF model has been applied on the daily Dhaka Stock Exchange from January 2000 till December 2015. The results showed that the proposed EMD-RF method performs better in forecasting than the RF, SVR, EMD-SVR, PCA-SVR, ICASVR, and PCA-ICA-SVR. Meanwhile, Wang et al. (2019) introduced a hybrid of EMD and multilayer perceptron (MLP) of Artificial Neural Network (ANN), and Büyükşahin and Ertekin (2019) suggested EMD, ARIMA, and ANN hybrid method to improve the forecasting performance. Their results also showed that hybrid methods perform better than the traditional method.

Based on the previous literature, it seems that EMD combination with other methods has improved the forecasting performance. Thus, this study will attempt to employ a join of EMD-RW (random walk) to predict the stock market data. Then, to assess the performance of forecasting, the proposed method is compared with the forecast of ARIMA (Autoregressive Integrated Moving Average), STS (Structural Time Series), single HW (Holt-Winter), and single RW models. Results showed that the proposed method is better than these models in terms of RMSE, MAE, MAPE, and Theil U. Section 2 introduces methods that are used in this study, which are EMD and RW methods. Section 3 presented the proposed EMD-RW methodology with a flowchart. Section 4 examines the data used in this study with a discussion of the results present the strength of EMD-RW. In the last section, Section 5, some concluding comments are made.

2 Methodology

This section describes various steps to implement the EMD-RW forecasting method. The discussion focuses on the Empirical Mode Decomposition, Random Walk, and the Fourier transform.

2.1 Empirical mode decomposition (EMD)

The EMD method was introduced in Huang et al. (1998). The EMD effective method for processing and decomposing several types of non-stationary time series, the mode mixing problem is a limitation of the standard EMD algorithm (Trybek et al., 2023). One of our contributions to this study is that the EMD-RW methodology can overcome these limitations. The EMD technique was employed in several areas as shown in Awajan et al. (2019). The EMD method deals with the nonlinear and non-stationary time series (Razif and Shabri, 2023), to decompose this time series into various simple time series (Al-Jawarneh et al., 2022). Furthermore, the EMD decomposes the time series with a reserve for the time domain of the data. This attribute provides a powerful and adaptive process for decomposing a time series into a set of time series known as intrinsic mode functions (IMF) and residual (Al-Jawarneh et al., 2021). As a result, the original time series can be built from the IMFs by using Equation (1).

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

where $x(t)$ is the original time series, IMF_i and $r(t)$ are i^{th} intrinsic mode function and residue, respectively. Which is the result of decomposition for the original time series by EMD. To evaluate the IMFs, the sifting process algorithm is applied to the original data $x(t)$ Awajan et al. (2017). This algorithm is presented in detail as follows:

1. Input the original time series data $x(t)$ into the EMD algorithm, and let the iteration value i be equal to 1.

2. Determine the local extrema values for the $x(t)$.

3. Using the cubic spline line methods, the local maxima values are connected together, as a result, the local upper envelope function will be created, and it is denoted by $u(t)$.

4. Using the same last two processes, the local lower envelope function will be created from the local minimum values of $x(t)$, and it is denoted by $l(t)$.

5. Then the mean function will be determined from $u(t)$ and $l(t)$ using the following formula, and it is denoted by $m(t)$;

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

6. A new function $h(t)$ will be determined by using $m(t)$ and the signal $x(t)$ by Equation (3)

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

7. In this step, we check the following conditions on the function $h(t)$, which are called IMF conditions Rilling et al. (2003); these are

   - $$|N.Ex - N.C.Z| < 1 \tag{4}$$

   - $$|m(t)| = \left|\frac{u(t) + l(t)}{2}\right| < \varepsilon \tag{5}$$

   where $N.Ex$ means the number of local extreme points (local maxima and minima), while $N.C.Z$ means the number of cross-zero points, and ε is a too-small non-negative value that approaches 0, sometimes equal to 0.

8. If the function $h(t)$ satisfies the above conditions, then move to the next step. Otherwise, move to step 2 and we assume the function $h(t)$ is the original data, and apply steps 2 to 8 to the new original data.

9. As a result of the previous step; we have a new IMF; which is determined using Equation (6). And then, update the iteration value (by moving to the next integer) $i = i + 1$.

$$IMF_i(t) = h(t) \tag{6}$$

10. In this step; we will define a new function, which is called the residue function $r_i(t)$; this function is evaluated by Equation (7).

$$r_{i+1}(t) = x(t) - IMF_i(t) \tag{7}$$

11. Based on the result from the last step $r_{i+1}(t)$ we will tack one of the following decisions.

    a) If $r_i(t)$ is a monotonic or constant function, that concludes this is the last process. Hence; we save the residue of the sifting process which is $r(t) = r_i(t)$, and we save all the IMFs that have been obtained.

    b) If the residue component does not satisfy the monotonic (or constant) condition, go back to the second step.

Figure 1 presents the flowchart of the empirical mode decomposition estimation process.

The IMFs and the residual components of the original data (the Malaysian stock market data as an example of the original time series) are presented in Figure 2.

2.2 Random Walk with Drift (RW)

The Random Walk (RW) method is defined as an algorithm to build the current observation according to the last observation considering the error part (white noise)) (Tyree and Long, 1995). Mathematically, the RW is written as follows:

$$x_t = x_{t-1} + \varepsilon_t \tag{8}$$

where x_t and x_{t-1} represent the observation's value the observations value of the time series at time t, and ε_t represents the error part at time t. The error part is known as the white noise terms which are Independent and identically distributed has a mean equal to zero and a fixed variance σ^2. This technique assumes that the best time series prediction for the current value is the previous value. The RW technique is not a stationary method to ensure its variance is not fixed.

The drift in the random walk method is represented as a trend, and this model can be represented mathematically in the following equation:

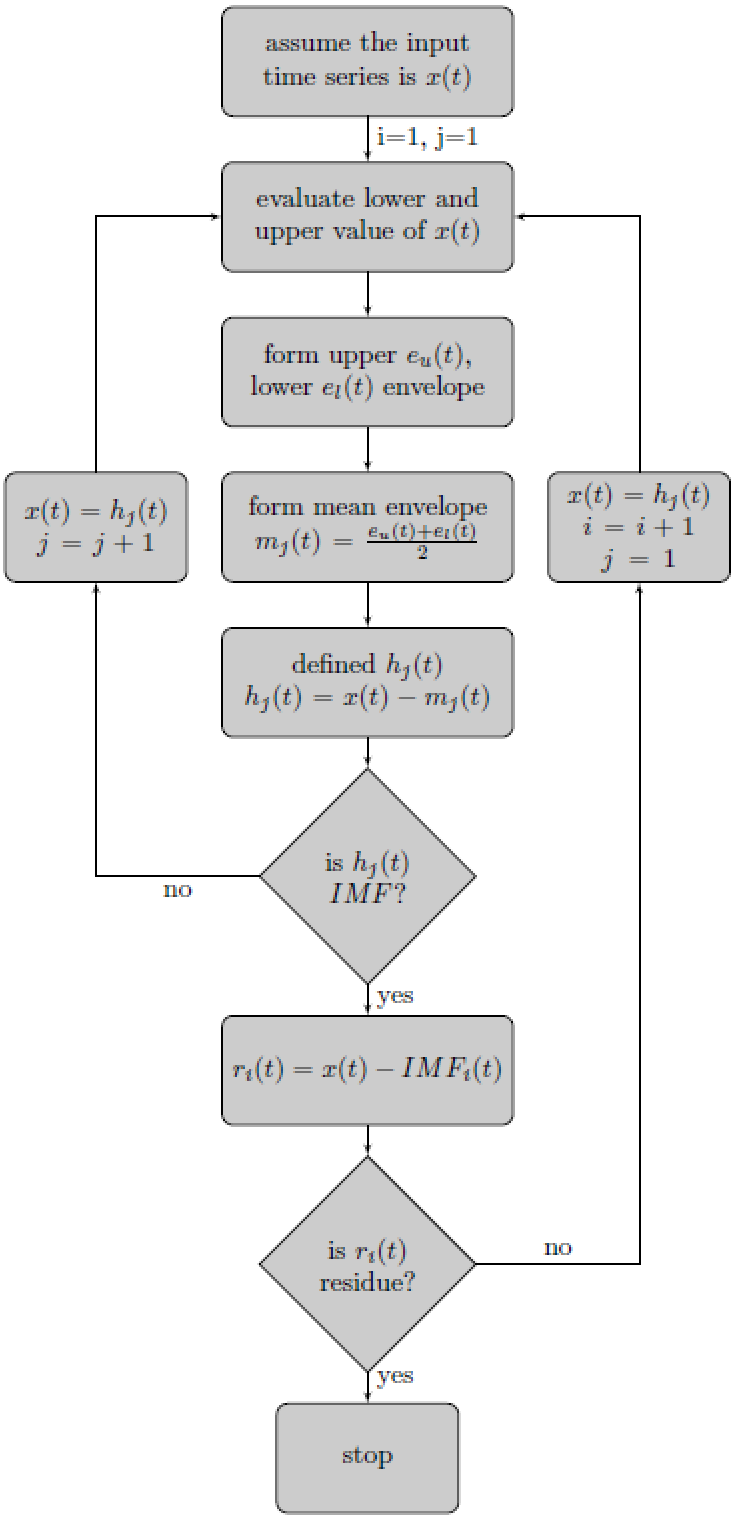

Figure 1: Flowchart of empirical mode decomposition estimation process

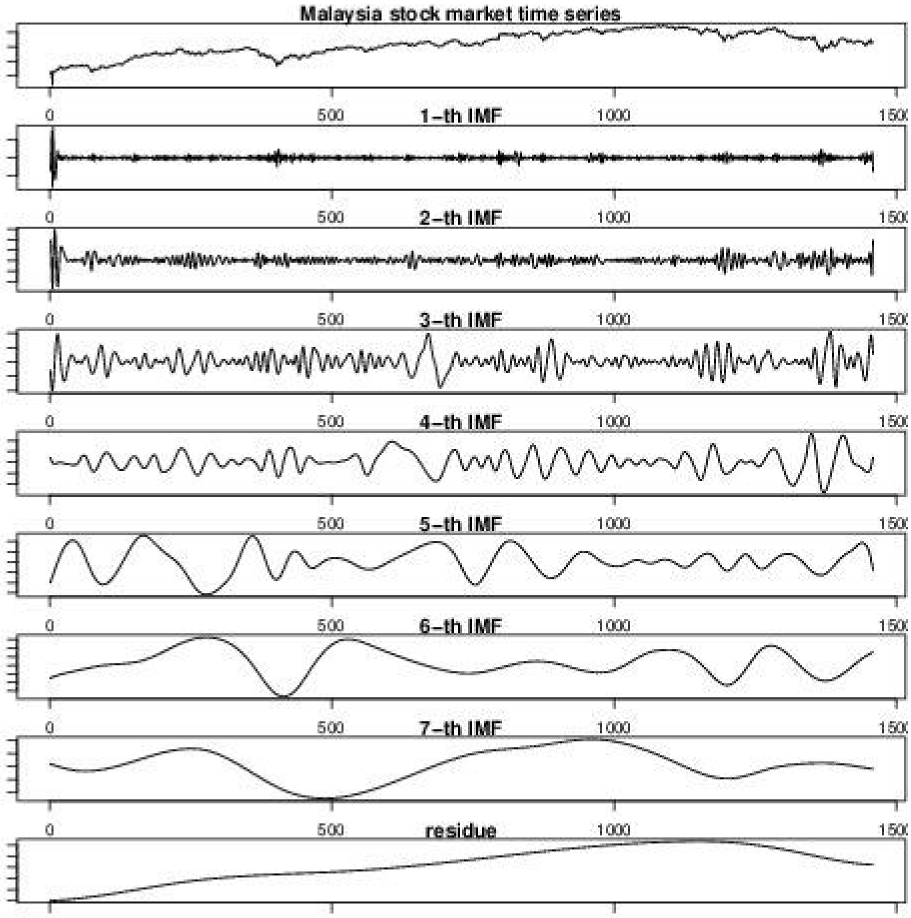

Figure 2: The IMFs and the residual components of the original data

$$x_t = x_{t-1} + \varepsilon_t + \alpha \tag{9}$$

for $\alpha > 0$, the technique will display a trend. This technique indicates both a deterministic and a stochastic trend, as presented in Rowland Rowland (2003). Forecasting values in the RW model are given by

$$x_{n+h} = x_n + \alpha.h \tag{10}$$

where h represents the number of the days-ahead forecast, and n represents the number of observations in the time series.

2.3 Fourier Transform

The Fourier transform (FT) is a conversion of the signal of data from the original domain to the frequency domain representation (Donoghue, 2014). The fast Fourier transform

(FFT) has been employed to estimate the spectral density of the IMFs time series components. The FFT is a mathematical technique for computing the finite discrete FT of signal data (Cooley et al., 1967; Bloomfield, 2004). The discrete FT of limited length can be written as the following mathematical formula (Van Loan, 1992).

Given that $Y_0, Y_1, ..., Y_{N-1}$ be the N data points of a time series. Then, this time series can be written by discrete FT as the following equation;

$$Y_k = \sum_{n=0}^{N-1} y_n e^{-i2\pi k \frac{n}{N}}; \qquad k = 0, \ldots, N-1. \tag{11}$$

3 Comparison between statistical techniques

In this study, the performance of the proposal forecasting EMD-RW technique is assessed with four forecasting models. The models are the Holt-Winter method by Holt (2004) and Winters (1960), the ARIMA models discussed by Peiris and Perera (1988), the structural time series model has been presented by Harvey and Koopman (2014); Harvey (1990); Turner and Witt (2001), and the random walk method. These models were selected according to their forecasting accuracy in the letterer. The structural time series is implemented by Kurihara et al. (2018) on the Japanese stock market by using the state space approach. They found that a state-space model is a useful tool for analyzing stock market prices. ARIMA (autoregressive integrated moving average) or Box-Jenkins models are commonly expressed as ARIMA(p, d, q). Here parameters p, d, and q are non-negative integers. The p represents the rank of the Autoregressive part, d denotes the number of differencing, and q means the rank of the Moving-average part. The value of d is obtained using repeated KPSS tests by Kwiatkowski et al. (1992). After finding the d, the p and q values are chosen from the minimum value of Akaike information criterion with bias correction (AIC_C). Equation (12) presents the formula of AIC_C, where n represents the sample size, k denotes the number of parameters and L denotes the maximum value of the likelihood function for the mathematical formula.

$$\text{AICc} = \text{2k} - 2\ln(L) + \frac{2k^2 + 2k}{n - k - 1} \tag{12}$$

The ARIMA was applied in a number of studies in forecasting financial data in the letterer. Such as Badmus et al. (2011) employed the ARIMA forecasting method for the cultivated area and production of maize in Nigeria. Then, Moroke (2014) used the ARIMA method to estimate the model used for forecasting household debt in South Africa. The Holt-Winters statistical technique is another form of exponential smoothing statistical technique, which is much simpler. Moreover, in this technique, the current values in time series affect more than elderly values in time series in forecasting future value. There are two types of the Holt-Winter technique, which are the Multiplicative and the Additive (Al-Gounmeein et al., 2023). Based on the seasonal part of this technique is determined the additive or multiplicative (Valakevicius and Brazenas, 2015). Mathematically, the additive HW is presented as the following formula:

$$\hat{y}_{t+h/t} = a_t + h * b_t + s_{t-p+1+(h-1)mod(p)}, \tag{13}$$

where a_t, b_t and s_t are given by

$$a_t = \alpha(y_t - s_{t-p}) + (1-\alpha)(a_{t-1} + b_{t-1}) \tag{14}$$

$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1} \tag{15}$$

$$s_t = \gamma(y_t - a_t) + (1-\gamma)s_{t-p} \tag{16}$$

And the multiplicative Holt-Winters forecasting function is defined by the following :

$$\hat{y}_{t+h/t} = (a_t + h * b_t) * s_{t-p+1+(h-1)mod(p)}, \tag{17}$$

where a_t, b_t and s_t are given by

$$a_t = \alpha(y_t / s_{t-p}) + (1-\alpha)(a_{t-1} + b_{t-1}) \tag{18}$$

$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1} \tag{19}$$

$$s_t = \gamma(y_t / a_t) + (1-\gamma)s_{t-p} \tag{20}$$

The at denoted the level of series at time t, b_t denoted the slope (growth) at time t, s_t denoted the seasonal component of the series at time t, and p denoted the number of seasons in a period. The constants α, β, and γ are smoothing parameters in the [0,1]-interval, h is the number of forecast values ahead. This algorithm employed the maximum likelihood function to determine the initial parameters, and then it may determine all the parameters iteratively to forecast future values of time series data. The data are required to be non-zero for a multiplicative model, but it makes sense if time series data are greater than zero.

Equations (21), (22), (23), and (24), show the formula of Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Theil's U-statistic (TheilU), respectively. These measurements will be used to evaluate the forecasting performance of the forecasting models. Where $\hat{y}$ is the forecast value of the variable y_i at step i from the actual series value.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2} \tag{21}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y}_i| \tag{22}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \widehat{y}_i|}{y_i}.100\% \tag{23}$$

$$TheilU = \frac{\sqrt{\sum_{i=1}^{h-1}\left(\frac{\widehat{y}_{i+1}-y_i}{y_i} - \frac{y_{i+1}-y_i}{y_i}\right)^2}}{\sqrt{\sum_{i=1}^{h-1}\left(\frac{y_{i+1}-y_i}{y_i}\right)^2}} \tag{24}$$

4 Data and the proposed method

This section starts by presenting the data used to implement the proposed methodology. Then details the discussion about the hybrid methodology of EMD-RW and finally displays the algorithm of the proposed methodology.

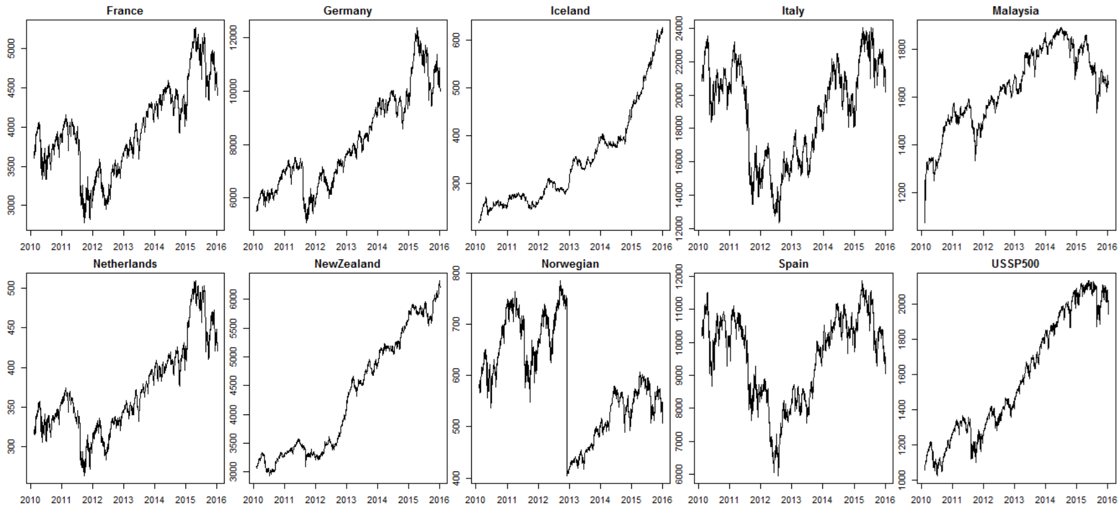

Figure 3: Time Series Plots

4.1 Data

Table 1 presents basic statistics for ten countries' stock market indexes and the number of observations for each stock market data. The ten countries are selected based on data availability without any specific preference. The stock market time series were obtained from the Yahoo Finance website (https://finance.yahoo.com/). Whereas, Figure 3 presents plots of the daily time series of these data. The close prices are employed as a general measure of the daily stock market data through the six years. These data cover the period from 9 February 2010 to 7 January 2016. Figure 3 presents the ten stock markets. The stationary and the linearity of these data are tested, and the KPSS(Kwiatkowski-Phillips-Schmidt-Shin (Banerjee et al., 1993)), RESET (Ramsey Regression Equation Specification Error Test (Ramsey, 1969)), BP (Breusch-Pagan test (Breusch and Pagan, 1979)) has been applied on the original data. Based on these tests, the time series is significantly non-linear and non-stationary, it has a high heteroscedasticity attribute also. Moreover, the basic statistics in Table 1 reveal that all the series are volatile, and display skewed and platykurtic distribution. To evaluate the proposed forecasting method in this study, the time series is partitioned into two sets. The observations from day-1 into day-m will be the first set. This data set is employed to select the suitable forecasting method. The second set is the observations from day-$(m + 1)$ into day-N (the last observation). This data set is reserved for out-of-sample evaluation,

Table 1: Basic statistics

Country	Mean	Median	Standard Deviation	Skewness	Kurtosis	Number of observations
France	3968.26	3939.82	557.54	0.21	-0.6	1516
Germany	8102.02	7637.87	1791.81	0.45	-0.91	1510
Iceland	348.33	326.35	101.33	1.07	0.23	1463
Italy	19025.28	19671.55	2937.37	-0.3	-1.14	1520
Malaysia	1638.2	1643.89	164.52	-0.4	-0.68	1459
Netherlands	370.77	355.92	56.19	0.65	-0.32	1516
New Zealand	4317.8	4340.73	1024.04	0.28	-1.47	1419
Norwegian	591.02	581.59	97.87	0.05	-0.99	1476
Spain	9537	9933.1	1297.48	-0.55	-0.67	1515
USSP500	1579.25	1493.69	344.31	0.2	-1.44	1490

which is employed to make a comparison of the accuracy with the selected forecasting methods. In this study, Malaysia stock market data are taken as an example to explain the partition of time series data. The number of observations is N = 1459, the first part is m = 1453, 1279, and 1094 and the second part is h = 6 days, 180 days, and 365 days respectively, are used.

4.2 Propose Methodology

The EMD-RW technical contains five steps as follows:

1. In the first step, the EMD method is used to decompose the original data. As a result, we obtained several IMFs (assumed equal to n) with one residue (denoted by r). These components can be written as the equation (1) Huang (2014).
2. The Fast Fourier Transformation (FFT) (Bloomfield, 2004) is employed to transform the IMF's components from the time domain into the frequency domain.
3. Based on the results from the previous step, the IMFs components are separated into two sets of data, which are high-frequency (HF) and low-frequency (LF) sets of data, with 0.02 threshold criteria (McAssey et al., 2013). Based on this threshold criteria, we define a new integer number M, such that M is between 1 and the number of IMFs (n). According to the M-value, the HF and LF are separated into two sets as the following:

$$HF = \{IMF_1, IMF_2, ..., IMF_M\} \tag{25}$$

$$LF = \{IMF_{M+1}, IMF_{M2}, ..., IMF_n\} \tag{26}$$

4. The Random Walk forecasting technique is employed for each IMF's component in HF set and the residue to forecast h-days in the future. The Random Walk forecasting technique was given as Equation (27).

$$x_t = x_{t-1} + \varepsilon_t + \alpha \tag{27}$$

   Click here to enter text. for $\alpha > 0$ the process will show an upward trend. Where x_t and x_{t-1} are the observations of the time series and ε_t is white noise. The white noise terms are $i.i.d.-$normal variables, having zero mean and a constant standard deviation. Now, for each IMFs in LF set and residue, we have h-forecasting values.

5. All the forecasted values from the last Step are aggregated in this step. As a result, a path of the forecasting value with length h for the original time series $x(t)$ is obtained. The symbol denotes the forecasting value of $\hat{x}_h$. Mathematically, $\hat{x}_h$ is written as a formula (28).

$$\hat{x}_h = RW_h(x(t)) + \sum_{i=M+1}^{n} RW_h(IMF_i(t)) \tag{28}$$

Figure 4 presents the methodology of the EMD-RW technique as a flowchart. While Figure 5 presents the scheme algorithm of the EMD-RW procedure.

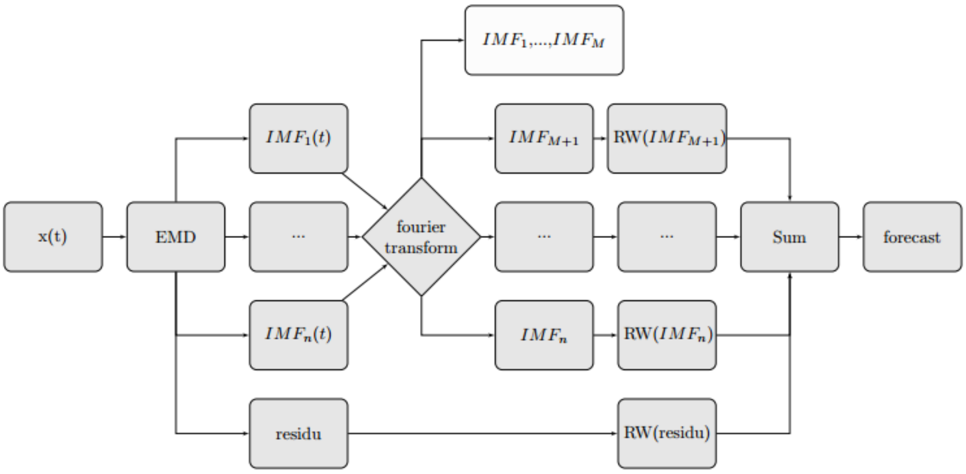

Figure 4: Flowchart of a hybrid Empirical Mode Decomposition with Random walk with drift model

5 Results and discussion

In this study, stock market data for ten countries are employed to display the forecasting accuracy performance for short-term (h=6 days), medium-term (h=180 days), and long-term (h=365 days) using the proposed method (EMD-RW). Besides, four traditional

procedure EMD-RW((time series,h))

[IMF_1,...,IMF_n,residue] ← EMD(time series)

[Low.IMF, High.IMFs] ← FFT(IMF_1,...,IMF_n,)

for ts←each component of (Low.IMFs and residue) **do**

h ahead forecast horizontal ← forecast(RW(ts), h)

end for

Add the forecasting results together.

end procedure

Figure 5: Summarized of the EMD-RW forecasting process

models (ARIMA, STS, single HW, and single RW are employed to present the forecasting accuracy of EMD -RW forecasting method. The RMSE, MAE, MAPE, and TheilU values for RW, ARIMA, HW, STS, and EMD-RW models, as shown in Table 2 and Table 3 for the ten countries.

From Table 2 and Table 3, nine of the ten countries shown the proposed method EMD-RW produced lower RMSE, MAE, MAPE, and TheilU for short-term (h=6 days) forecasting. As for the medium-term (h=180 days) forecasting performance, eight out of the ten countries indicate that EMD-RW is superior to the single RW, ARIMA, single HW, and STS. Meanwhile, for long-term (h=365 days) forecasting, out of the ten countries, only three countries displayed lesser forecasting performance measurement values for EMD-RW but chose another model to be better.

Moreover, results in Table 2 and Table 3 reveal that the HW and ARIMA models could be an alternative for long-term prediction due to the minimum measure of accuracy. Taylor (2003) indicates that ARIMA and HW are appropriate and adequate for a short-time forecast. However, in this study when long-term forecasting stock market, 365 days, the ARIMA model, compared to the remaining models, indicated minimum accuracy measures, (108.6, 90.8, 16.9, and 26.1) and (276.1, 236.8, 4.1, and 9.4) for Iceland and New Zealand respectively. At the same time, HW predicted France's, Germany's, Italy's, and Netherlands's index better. Therefore, in this case, the results indicate that ARIMA and HW are only adequate for long-term forecast, and this contradicts the conditional of ARIMA and HW modeling essence of short and medium forecasting (Al-Musaylh et al., 2018; Wong and Guo, 2010; Makridakis and Hibon, 1991; Makatjane and Moroke, 2016).

The proposed method, EMD-RW indicates the minimum accuracy measures, (154.4, 127.4, 2.8, and 2.5), (393.3, 351.1, 3.4, and 1.8), (606.4, 530.2, 2.6, and 1.5), and (13.8, 11.4, 2.7, and 2.2) when forecasting 6 and 180 days for the daily stock market index of France, Germany, Italy, and the Netherlands. Also, on average, the EMD-RW out-

Table 2: The RMSE, MAE, MAPE, and TheilU of EMD-RW, ARIMA, HW, STS, and RW for forecasting at h = 6, 180, and 365 for France, Germany, Iceland, Italy, and Malaysia countries.

		RMSE			MAE			MAPE			TheilU		
Country	Method\h	6	180	365	6	180	365	6	180	365	6	180	365
France	RW	183	493.8	559.5	158.3	443.9	466.8	3.5	9.4	9.6	2.9	6.9	8.3
	ARIMA	185.8	613.2	559.5	160.8	556.6	466.8	3.6	11.8	9.6	3	8.5	8.3
	HW	186.8	784.8	**400.6**	161.6	712.9	**317.4**	3.6	15.1	**6.5**	3	10.9	**5.9**
	STS	182.6	492.8	559.6	157.8	442.9	466.9	3.5	9.4	9.6	2.9	6.9	8.3
	EMD-RW	**154.4**	**451.8**	542.6	**127.4**	**396.7**	450.9	**2.8**	**8.4**	9.3	**2.5**	**6.3**	8
Germany	RW	398.7	1192.6	1262	353.9	1016.6	1022.8	3.5	9.7	9.3	1.8	7.2	7.8
	ARIMA	410.7	1652.5	1267.9	362.5	1450.2	1027.5	3.5	13.8	9.3	1.9	10	7.9
	HW	399.5	2605.1	**964.8**	354.5	2306.8	**756.3**	3.5	21.9	**6.9**	1.8	15.5	**6**
	STS	398.7	1192.6	1262	353.9	1016.6	1022.8	3.5	9.7	9.3	1.8	7.2	7.8
	EMD-RW	**393.3**	**1156.4**	1234.2	**351.1**	**975.3**	1001.5	**3.4**	**9.4**	9.1	**1.8**	**7**	7.7
Iceland	RW	9.3	86.5	139.5	8.8	76.2	117.6	1.4	13.2	22	2.3	21.6	33.6
	ARIMA	13.7	65	**108.6**	13.1	57.4	**90.8**	2.1	9.9	**16.9**	3.5	16.2	**26.1**
	HW	7.9	51.3	160.3	7.3	45.5	135.7	1.2	7.9	25.4	2	12.8	38.7
	STS	8.4	86.5	108.9	7.8	76.2	91.1	1.3	13.2	17	2.1	21.6	26.2
	EMD-RW	**5.2**	**86.1**	138.6	**4.3**	**75.7**	116.6	**0.7**	**13.1**	21.8	**1.4**	**21.5**	33.4
Italy	RW	698.6	1169.1	1676	581.5	947.3	1450	2.8	4.3	6.7	1.8	3.2	4.8
	ARIMA	699.7	1315.9	1675.9	582.2	1060.7	1449.9	2.8	4.9	6.7	1.8	3.6	4.8
	HW	681.9	3300.6	**1477**	569.4	2744.7	**1245.6**	2.8	12.5	**6**	1.7	9	**4.6**
	STS	700.4	1170.8	1675.7	582.4	948.5	1449.8	2.8	4.3	6.7	1.8	3.2	4.8
	EMD-RW	**606.4**	**1135.6**	1615.2	**530.2**	**924.5**	1398	**2.6**	**4.2**	6.5	**1.5**	**3.1**	4.7
Malaysia	RW	21.4	166.2	151.5	19	151.5	127	1.1	9.1	7.5	1.2	13	13.4
	ARIMA	20.8	212.7	273.8	18.8	193.9	236.3	1.1	11.6	13.8	1.2	16.6	24.1
	HW	22.1	274.7	15246.6	19.8	249.4	15185.3	1.2	15	869.7	1.3	21.5	1292
	STS	20.5	165.3	152	18.6	150.6	127.5	1.1	9.1	7.5	1.2	13	13.4
	EMD-RW	**21**	**163.4**	**135.4**	**18.9**	**148.7**	**108.5**	**1.1**	**8.9**	**6.4**	**1.2**	**12.8**	**12**

performed all four benchmark models when predicting the short, medium, and long-term forecast of Norway, Spain, and US stock market index, and this could be attributed to the fact that the new method improved the performance of the RW model.

In view of this, the result indicates that the EMD-RW is not only an improved excellent alternative to studying the stock markets index behavior but also a good hybrid model to achieve short- and medium-term forecasts with high accuracy. The proposed method is presented to be used for short-term and medium-term forecasting, more than long-term ones. This is due to short-term and medium-term forecasting being far more prevalent in commodities and stock markets where traders and investors focus on short-term and medium-term forecasting (Liu et al., 2022).

Despite this, the proposed method, EMD-RW indicates the minimum accuracy measures, (135.4, 108.5, 6.4, and 12), (27.9, 22.1, 4.1, and 4.4), (597,498.7, 4.7, and 4.1), and (86.7, 75.8, 3.7, and 4.5) when forecasting 365 days for the daily stock market index of Malaysia, Norwegian, Spain, and US S&P500. Also, in the stock markets of France, Germany, Italy, and New Zealand, the EMD-RW outperformed three benchmark models when predicting the long-term forecast.

Table 3: The RMSE, MAE, MAPE, and TheilU of EMD-RW, ARIMA, HW, STS, and RW for forecasting at h = 6, 180, and 365 for Netherlands, New Zealand, Norwegian, Spain, and US S&P 500 countries.

		RMSE			MAE			MAPE			TheilU		
Country	Method\h	6	180	365	6	180	365	6	180	365	6	180	365
Nether-lands	RW	16	52.2	64.3	13.7	45.5	54.8	3.2	10.1	11.7	2.6	8.1	10.6
	ARIMA	16.3	66.3	52.4	14	58.5	42.5	3.3	13	9	2.6	10.2	8.5
	HW	16.3	89.6	**48.8**	14	79.5	**38.7**	3.3	17.7	**8.2**	2.6	13.8	**7.9**
	STS	16	52.2	64.3	13.7	45.5	54.8	3.2	10.1	11.7	2.6	8.1	10.6
	EMD-RW	**13.8**	**48.2**	61.9	**11.4**	**40.8**	52.2	**2.7**	**9.1**	11.1	**2.2**	**7.5**	10.1
New Zealand	RW	44.2	199.6	640.3	38.9	149.1	576.4	0.6	2.5	9.9	1	5.9	21.6
	ARIMA	53	176.7	**276.1**	45.8	125.6	**236.8**	0.7	2.2	**4.1**	1.2	5.6	**9.4**
	HW	60.4	189.2	779.8	51.8	144	700.7	0.8	2.4	12.1	1.4	5.6	26.3
	STS	44.2	199.6	640.3	38.9	149.1	576.4	0.6	2.5	9.9	1	5.9	21.6
	EMD-RW	**40.6**	**188.2**	605.7	**34.1**	**145**	540.3	**0.5**	**2.4**	9.3	**0.8**	**5.6**	20.4
Nor-wegian	RW	17.9	43.7	29.2	14	36.3	23.4	2.7	6.6	4.3	2.1	6.5	4.6
	ARIMA	17.9	45.6	29	14	38.1	23.2	2.7	7	4.3	2.1	6.8	4.5
	HW	17.2	108.2	136.7	13.5	94.5	122	2.6	17.1	22	2	15.8	20.7
	STS	17.9	43.7	29.2	14	36.3	23.4	2.7	6.6	4.3	2.1	6.5	4.6
	EMD-RW	**17**	**41.3**	**27.9**	**13.5**	**33.4**	**22.1**	**2.6**	**6.1**	**4.1**	**1.9**	**6.1**	**4.4**
Spain	RW	377.2	1278.3	640.5	321.9	1088.2	514.2	3.5	10.8	4.8	2.9	9	4.3
	ARIMA	379.9	1260.1	637.3	327.2	1066.8	512.2	3.5	10.6	4.8	3	8.9	4.3
	HW	347	2248.7	669.4	294.4	1953.6	532	3.2	19.3	4.9	2.7	15.7	4.5
	STS	377.2	1278.3	640.5	321.9	1088.2	514.2	3.5	10.8	4.8	2.9	9	4.3
	EMD-RW	**340.4**	**1250.7**	**597**	**283**	**1055.9**	**498.7**	**3.1**	**10.5**	**4.7**	**2.7**	**8.8**	**4.1**
US S &P 500	RW	77	80.6	87.8	66.7	55.4	76.8	3.4	2.8	3.7	2.9	3.9	4.5
	ARIMA	78.4	146.6	122.8	67.9	119.3	87.9	3.4	5.9	4.4	2.9	7	6.6
	HW	79.2	167.4	220	68.6	138.8	176	3.4	6.9	8.7	3	8	11.7
	STS	76.7	151.4	87.8	66.3	124.2	76.8	3.3	6.2	3.7	2.9	7.2	4.5
	EMD-RW	**64.5**	**78.2**	**86.7**	**51.7**	**53.1**	**75.8**	**2.6**	**2.7**	**3.7**	**2.4**	**3.8**	**4.5**

6 Conclusions

In this study, a new hybrid EMD-RW forecasting method was presented for stock market forecasting. The EMD-RW was examined on ten different stock market data using four accuracy calculation functions. It was found that EMD-RW can outperform four selected forecasting methods for short and medium-term forecasting but the proposed was limited for long-term forecasting. Thus, this paper recommended using EMD-RW for forecasting stock market data in short and medium-term durations. In addition, the results of this study are aligned with previous literature that reveals the traditional forecasting method, when combined with EMD, will create a better forecasting model.

References

Al-Gounmeein, R. S., Ismail, M. T., Al-Hasanat, B. N., and Awajan, A. M. (2023). Improving models accuracy using kalman filter and holt-winters approaches based on arfima models. *IAENG International Journal of Applied Mathematics*, 53(3):98–107.

Al-Jawarneh, A. S., Ismail, M. T., and Awajan, A. M. (2021). Elastic net regression and empirical mode decomposition for enhancing the accuracy of the model selection. *International Journal of Mathematical, Engineering and Management Sciences*, 6(2):564.

Al-Jawarneh, A. S., Ismail, M. T., Awajan, A. M., and Alsayed, A. R. (2022). Improving accuracy models using elastic net regression approach based on empirical mode decomposition. *Communications in Statistics-Simulation and Computation*, 51(7):4006–4025.

Al-Musaylh, M. S., Deo, R. C., Adamowski, J. F., and Li, Y. (2018). Short-term electricity demand forecasting with mars, svr and arima models using aggregated demand data in queensland, australia. *Advanced Engineering Informatics*, 35:1–16.

Awajan, A. M., Al-Hasanat, B., Elkaroui, E., AL e'damat, A., Al-Gounmeein, R. S., Al-Jawarneh, A. S., Ayyoub, H. N., and AlFarajat, E. (2024). Time series forecasting of new cases for covid-19 pandemic in jordan using enhanced hybrid emd-arima. *Journal of Statistics Applications & Probability*, 13(1).

Awajan, A. M., Ismail, M. T., and Al Wadi, S. (2017). A hybrid emd-ma for forecasting stock market index. *Italian Journal of Pure and Applied Mathematics*, 38(1):1–20.

Awajan, A. M., Ismail, M. T., and Al Wadi, S. (2018). Improving forecasting accuracy for stock market data using emd-hw bagging. *PloS one*, 13(7):e0199582.

Awajan, A. M., Ismail, M. T., and Wadi, S. (2019). A review on empirical mode decomposition in forecasting time series. *Italian Journal of Pure and Applied Mathematics*, 43:301–323.

Badmus, M., Ariyo, O., et al. (2011). Forecasting cultivated areas and production of maize in nigerian using arima model. *Asian Journal of Agricultural Sciences*, 3(3):171–176.

Banerjee, A., Dolado, J. J., Galbraith, J. W., and Hendry, D. (1993). *Co-integration, error correction, and the econometric analysis of non-stationary data*. Oxford university press.

Bloomfield, P. (2004). *Fourier analysis of time series: an introduction*. John Wiley & Sons.

Breusch, T. S. and Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the econometric society*, pages 1287–1294.

Büyükşahin, Ü. Ç. and Ertekin, Ş. (2019). Improving forecasting accuracy of time series data using a new arima-ann hybrid method and empirical mode decomposition. *Neurocomputing*, 361:151–163.

Castle, J., Hendry, D. F., and Clements, M. P. (2019). *Forecasting*. Yale University Press.

Chowdhury, U. N., Chakravarty, S. K., Hossain, M. T., and Ahmad, S. (2019). Empirical mode decomposition based ensemble random forest model for financial time series forecasting. *International Journal of Engineering and Information Systems (IJEAIS)*, 3(1):1–13.

Cooley, J., Lewis, P., and Welch, P. (1967). Application of the fast fourier transform to computation of fourier integrals, fourier series, and convolution integrals. *IEEE Transactions on Audio and Electroacoustics*, 15(2):79–84.

Donoghue, W. F. (2014). *Distributions and Fourier transforms.* Academic Press.

Harvey, A. and Koopman, S. (2014). Structural time series models. *Wiley StatsRef: Statistics Reference Online.*

Harvey, A. C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press.

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International journal of forecasting*, 20(1):5–10.

Huang, N. E. (2014). *Hilbert-Huang transform and its applications*, volume 16. World Scientific.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H. (1998). 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*, 454(1971):903–995.

Ismail, M. and Awajan, A. M. (2017). A new hybrid approach emd-exp for short-term forecasting of daily stock market time series data. *Electronic Journal of Applied Statistical Analysis*, 10(2):307–327.

Kazem, A., Sharifi, E., Hussain, F. K., Saberi, M., and Hussain, O. K. (2013). Support vector regression with chaos-based firefly algorithm for stock market price forecasting. *Applied soft computing*, 13(2):947–958.

Kurihara, Y., Maeda, S., et al. (2018). The validity of applying state space model to japanese stock market. *International Journal of Economics and Financial Modelling*, 3(1):1–8.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1-3):159–178.

Le An, Y., Shou Yang, W., Lai, K., and Nakamori, Y. (2005). Time series forecasting with multiple candidate models: selecting or combining? *Journal of Systems Science and Complexity*, 18(1):1.

Lin, A., Shang, P., Feng, G., and Zhong, B. (2012). Application of empirical mode decomposition combined with k-nearest neighbors approach in financial time series forecasting. *Fluctuation and Noise Letters*, 11(02):1250018.

Liu, H., Qi, L., and Sun, M. (2022). Short-term stock price prediction based on cae-lstm method. *Wireless Communications and Mobile Computing*, 2022.

Makatjane, K. and Moroke, N. (2016). Comparative study of holt-winters triple exponential smoothing and seasonal arima: forecasting short term seasonal car sales in south africa. *Makatjane KD, Moroke ND.*

Makridakis, S. and Hibon, M. (1991). Exponential smoothing: The effect of initial values and loss functions on post-sample forecasting accuracy. *International Journal of Forecasting*, 7(3):317–330.

McAssey, M. P., Helm, J., Hsieh, F., Sbarra, D. A., and Ferrer, E. (2013). Methodological advances for detecting physiological synchrony during dyadic interactions. *Methodology*.

Mills, T. C. (2019). *Applied time series analysis: A practical guide to modeling and forecasting*. Academic press.

Moroke, N. D. (2014). The robustness and accuracy of box-jenkins arima in modeling and forecasting household debt in south africa. *Journal of Economics and Behavioral Studies*, 6(9):748–759.

Nava, N., Di Matteo, T., and Aste, T. (2018). Financial time series forecasting using empirical mode decomposition and support vector regression. *Risks*, 6(1):7.

Oh, H.-S., Suh, J.-H., and Kim, D.-H. (2009). A multi-resolution approach to non-stationary financial time series using the hilbert-huang transform. *The korean journal of applied statistics*, 22(3):499–513.

Parsons, S., Boonman, A. M., and Obrist, M. K. (2000). Advantages and disadvantages of techniques for transforming and analyzing chiropteran echolocation calls. *Journal of Mammalogy*, 81(4):927–938.

Patel, J., Shah, S., Thakkar, P., and Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications*, 42(4):2162–2172.

Peiris, M. and Perera, B. (1988). On prediction with fractionally differenced arima models. *Journal of Time Series Analysis*, 9(3):215–220.

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 31(2):350–371.

Razif, N. R. A. and Shabri, A. (2023). Application of empirical mode decomposition in improving group method of data handling. In *AIP Conference Proceedings*, volume 2500. AIP Publishing.

Rilling, G., Flandrin, P., Goncalves, P., et al. (2003). On empirical mode decomposition and its algorithms. In *IEEE-EURASIP workshop on nonlinear signal and image processing*, volume 3, pages 8–11. Grado: IEEE.

Rowland, P. (2003). Forecasting the usd/cop exchange rate: A random walk with a variable drift. *Borradores de Economía; No. 253.*

Taylor, J. W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, 54(8):799–805.

Trybek, P., Sobotnicka, E., Wawrzkiewicz-Jałowiecka, A., Machura, Ł., Feige, D., Sobotnicki, A., and Richter-Laskowska, M. (2023). A new method of identifying characteristic points in the impedance cardiography signal based on empirical mode decomposition. *Sensors*, 23(2):675.

Turner, L. W. and Witt, S. F. (2001). Forecasting tourism using univariate and multivariate structural time series models. *Tourism Economics*, 7(2):135–147.

Tyree, E. W. and Long, J. (1995). Forecasting currency exchange rates: Neural networks and the random walk model. In *City University Working Paper, Proceedings of the Third International Conference on Artificial Intelligence Applications*. Citeseer.

Valakevicius, E. and Brazenas, M. (2015). Application of the seasonal holt-winters model to study exchange rate volatility. *Inžinerinė ekonomika*, 26(4):384–390.

Van Loan, C. (1992). *Computational frameworks for the fast Fourier transform*. SIAM.

Wang, J., Zhang, D., and Zhang, J. (2015). Mean reversion in stock prices of seven asian stock markets: Unit root test and stationary test with fourier functions. *International Review of Economics & Finance*, 37:157–164.

Wang, J.-N., Du, J., Jiang, C., and Lai, K.-K. (2019). Chinese currency exchange rates forecasting with emd-based neural network. *Complexity*, 2019:1–15.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management science*, 6(3):324–342.

Wong, W. K. and Guo, Z. (2010). A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm. *International Journal of Production Economics*, 128(2):614–624.

Yang, H.-L. and Lin, H.-C. (2017). Applying the hybrid model of emd, psr, and elm to exchange rates forecasting. *Computational Economics*, 49(1):99–116.