Communications in Statistics - Simulation and Computation

ISSN: 0361-0918 (Print) 1532-4141 (Online) Journal homepage: https://www.tandfonline.com/loi/lssp20

Improving accuracy models using elastic net regression approach based on empirical mode decomposition

Abdullah S. Al-Jawarneh, Mohd Tahir Ismail, Ahmad M. Awajan & Ahmed R. M. Alsayed

To cite this article: Abdullah S. Al-Jawarneh, Mohd Tahir Ismail, Ahmad M. Awajan & Ahmed R. M. Alsayed (2020): Improving accuracy models using elastic net regression approach based on empirical mode decomposition, Communications in Statistics - Simulation and Computation

To link to this article: https://doi.org/10.1080/03610918.2020.1728319

Published online: 18 Feb 2020.

COMMUNICATIONS IN STATISTICS - SIMULATION AND COMPUTATION®
https://doi.org/10.1080/03610918.2020.1728319

Taylor & Francis
Taylor & Francis Group

Improving accuracy models using elastic net regression approach based on empirical mode decomposition

Abdullah S. Al-Jawarneh[a], Mohd Tahir Ismail[a], Ahmad M. Awajan[b], and Ahmed R. M. Alsayed[c]

[a]School of Mathematical Sciences, University Science Malaysia, Pulau Pinang, Malaysia; [b]Department of Mathematics, Al-Hussein Bin Talal University, Ma'an, Jordan; [c]Department of Economics, Management and Quantitative Methods, University of Milan, Milan, Italy

ABSTRACT

In this study, an elastic net (EN) regression model based on the empirical mode decomposition (EMD) algorithm is used in two applications, namely, numerical experiment and actual time series data. EMD is used to analyze a nonstationary and nonlinear signal dataset, which includes a set of orthogonal intrinsic mode functions (IMFs) and residual components. EN regression is used to select the most significant predictor variables influencing response variables and can address the multicollinearity problem between predictor variables. The main objective of this study is to apply the proposed method, EMD-EN, by using two variables for selecting important orthogonal IMFs and the residual components of predictor variables with significant effects on response variables. Moreover, this study uses the EMD-EN method in two different applications involving nonstationary and nonlinear problems. Results show that the proposed method outperforms other competitive methods in the numerical experiment and applications.

ARTICLE HISTORY
Received 13 July 2019
Accepted 6 February 2020

KEYWORDS
Elastic net regression; Empirical mode decomposition; LASSO regression; Multicollinearity; Ridge regression

1. Introduction

In various fields of science such as; medicine, physics, economics, and environmental science, the relationships of one or several predictor variables influencing response variables are interesting to be examined, but those variables often belong to nonstationary and nonlinear datasets, or a correlation may exist between two or more predictor variables and cause multicollinearity, which increases the variance of estimated coefficients and thus renders the estimated parameters inaccurate (Cho et al. 2010). In regression analysis, these problems lead to bias in selecting fitting models and can mislead the conclusions of studies. Empirical mode decomposition (EMD) and elastic net (EN) regression analysis are thus proposed to address these gaps (Zou and Zhang 2009).

The EMD algorithm was proposed in 1998 and was aimed at decomposing signals in nonstationary and nonlinear data (Huang 2014). EMD is different from traditional methods, namely, Fourier decomposition (Titchmarsh 1948) and wavelet decomposition (Chui 1995), which assume that signals are either stationary or linear. The main

CONTACT Abdullah S. Al-Jawarneh a_aljawarneh86@student.usm.my, abdssj2004@yahoo.com School of Mathematical Sciences, University Science Malaysia, Pulau Pinang, Malaysia.

advantage of the EMD method is that it does not require preconditions for signals (Crowley 2012). The practical principle of the EMD method is that the sifting process decomposes a signal into a finite set of orthogonal decomposition components, namely, intrinsic mode functions (IMFs) and residual components (Huang et al. 1998; Moore et al. 2018). These decomposition components via the EMD of the sifting process have varying wavelengths, amplitudes and frequencies, indicating that they are physically meaningful (Huang 2014).

The EN regression analysis proposed in 2005 is a penalized regularization method; it is a combination of the "ridge regression" and "least absolute shrinkage and selection operator (LASSO) regression." EN regression was proposed to analyze high-dimensional datasets and to tackle the limitations of the LASSO method. EN regression regularizes and selects important predictor variables to enhance the prediction accuracy of sparse modeling (Zou and Hastie 2005; Van der Kooij and Meulman 2008). It deals with the multicollinearity between predictor variables to produce models that are free from multicollinearity (Zou and Hastie 2005; Zou and Zhang 2009; Cho et al. 2010).

The EMD method combined with other statistical regression methods has been successfully applied in several scientific fields; it can address nonstationary and nonlinear signal datasets to discover the effects of predictor variables on response variables. The decomposition components of the original signal via EMD represent the predictors or response variables in the regression analysis method. For instance, Yang et al. (2011) and Yang, Tsai, and Huang (2011) combined ordinary least squares (OLS) regression with forward stepwise methods based on EMD (Yang, Tsai, and Huang 2011; Yang et al. 2011). Shen and Lee (2012) and Chu, Wei, and Qiu (2018) applied LASSO regression based on the decomposition components of ensemble empirical mode decomposition (EEMD) (Shen and Lee 2012; Chu, Wei, and Qiu 2018). Plakandaras et al. (2015) combined EN regression with a support vector regression (SVR) model based on the EEMD (Plakandaras et al. 2015). Qin et al. (2016) combined LASSO regression with decomposition components via EMD (Qin et al. 2016). Recently, Naik, Satapathy, and Dash (2018) combined the kernel ridge regression based on the decomposition components of EMD results (Naik, Satapathy, and Dash 2018). Masselot et al. (2018) used LASSO regression based on noise-assisted multivariate empirical mode decomposition to select the decomposition components, IMFs and residuals to identify their significant effects on response variables (Masselot et al. 2018).

The proposed EMD-EN method addresses the problem of nonstationary and nonlinear signal. EMD is used to decompose the original signals of datasets into a set of orthogonal IMF components and one residual component. In the next step, the decomposition components of EMD are used as orthogonal predictor variables in the EN regression analysis. Therefore, the proposed EMD-EN method combines EMD with EN regression to scan the relationships between predictor variables and response variables. The predictor variables of our study are IMF components and residual component, and their significant effects on the response variable are tested.

This article is organized as follows. Section 2 presents a brief description of EMD algorithms, EN regression, multicollinearity, the proposed EMD-EN method, and the test criteria. Section 3 provides the simulation data and the actual time series data used in this study. Section 4 provides the analysis and discussion. Finally, Sec. 5 concludes the study.

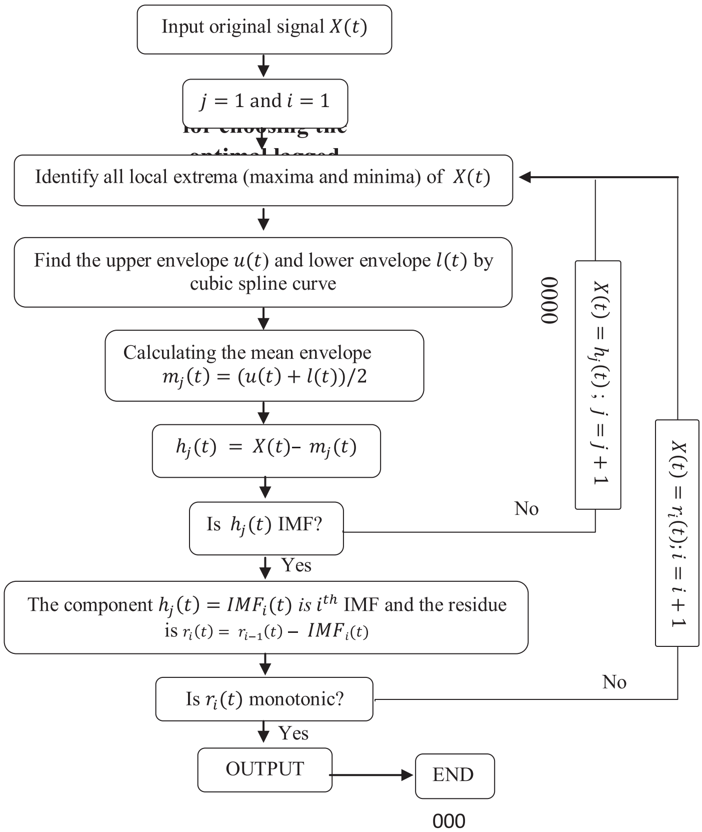

Figure 1. Empirical mode decomposition process.

2. Methodology

We explain the applied methods in detail. The first method deals with signals by using the EMD method and its implementation via the sifting process. The second method is the technical penalized least squares estimator method called EN regression. Finally, the proposed EMD-EN method is discussed.

2.1. Empirical mode decomposition

EMD was proposed by Huang et al. (1998) and then modified by several researchers (Rilling, Flandrin, and Goncalves 2003; Deering and Kaiser 2005; Jaber, Ismail, and Altaher Alsaidi 2013). EMD is the first part of the Hilbert–Huang transform (Huang 2014), and it improves the decomposition of a nonstationary and nonlinear original signal $X(t)$ into a finite set of orthogonal components called IMFs $\{\text{IMF}_i : \; i = 1, 2, \ldots, n\}$, and one component representing the trend of the signal is called the residual $r(t)$ component which keeps the time domain of the signal unchanged (Huang et al. 1998; Moore et al. 2018).

Signal analysis with the EMD algorithm is carried out by using the iterative algorithm method called the sifting process. The sifting process is used to extract the IMF and residual components from the original signal. The IMF components are orthogonal to each component as each of these IMF components satisfies two conditions:

- Over the entire length of a signal, the number of local extrema (LE) (i.e. local maximum and local minimum) and the number of zero crossings (ZCs) must be equal or differ at most by one.

$$(\#\mathrm{LE} = \#\,\mathrm{ZC}\ )\ \big|\ (|\#\mathrm{LE} - \#\,\mathrm{ZC}| = 1) \tag{1}$$

- At any point on a signal, the mean value between envelopes (i.e. upper $u(t)$ and lower $l(t)$ envelopes) is zero.

$$m(t) = \frac{u(t) + l(t)}{2} = 0 \tag{2}$$

The original signal $X(t)$ is a linear combination of a finite set of orthogonal IMF $\{\mathrm{IMF}_i(t) : i = 1, 2, \ldots, n\}$ components and one monotonic residual $r(t)$ component assumed as the $(n+1)^{\mathrm{th}}$ IMF component.

$$X(t) = \sum_{i=1}^{n} IMF_i(t) + r(t) \tag{3}$$

The steps of the sifting process to decompose the original signal $X(t)$ in the signal analysis into a set of IMFs and residual components are summarized in Figure 1 and outlined as follows:

Step 1: Insert the original signal $X(t)$ to start off the sifting process with the repetition index equal to 1 (i.e., $i,\ j = 1$).
Step 2: Identify all the local maximum and local minimum of the original signal $X(t)$.
Step 3: Connect all the local maximum and local minimum separately by using a cubic spline curve form to the upper $u(t)$ and lower $l(t)$ envelopes, respectively.
Step 4: Calculate the mean value for the upper and lower envelopes as follows:

$$m_j(t) = \frac{u(t) + l(t)}{2} \tag{4}$$

Step 5: Calculate the component $h_j(t)$, which is the difference between the original signal $X(t)$ and the mean value envelope $m_j(t)$, as follows:

$$h_j(t) = X(t) - m_j(t) \tag{5}$$

Step 6: Check whether the component $h_j(t)$ satisfies the conditions of the IMF.

NO: Replace $h_j(t)$ with $X(t)$, then, repeat the operation from step 1 on $h_j(t)$ and $j = j + 1$.
YES: $\boldsymbol{h_j(t) = IMF_i(t)}$, where $IMF_i(t)$ is the ith IMF $\{i = 1.2.3\ldots\ n\}$. Then, calculate the residual component in the following forms:

$$r_i(t) = r_{i-1}(t) - \mathrm{IMF}_i(t) \tag{6}$$

Step 7: Check whether the $r_i(t)$ component is a monotonic function or satisfies the stoppage criterion of the standard deviation SD_j for two consecutive successive siftings of the results, where $0.2 \leq SD_j \leq 0.3,$ as shown in the following formula:

$$SD_j = \sum_{t=0}^{T} \frac{\left(h_{j-1}(t) - h_j(t)\right)^2}{{h_{j-1}}^2(t)} \tag{7}$$

NO: Replace $r_i(t)$ with $X(t)$, and then, repeat the operations from step 1 on the residue $r_i(t)$ and $i = i + 1$ until $r_i(t)$ has a monotonic function or satisfies the stoppage criterion SD_j.

YES: Save all IMF and residual components, and then, end the sifting process.

2.2. Elastic net regression

EN regression is a technical penalized least squares estimator method proposed by (Zou and Hastie 2005) to analyze high-dimensional datasets and to overcome the limitations of the LASSO method. The structure of EN regression is a combination of two regression methods, namely, ridge regression (L_2 penalty) (Hoerl and Kennard 1970) and LASSO regression (L_1 penalty) (Tibshirani 1996). EN is a shrinkage and variable selection method that fits the general regression model that searches for a relationship between standardizing the predictor variables x_1, x_2, ..., x_p and response variable y as follows:

$$y = \sum_{j=1}^{p} x_j \beta_j + \varepsilon \tag{8}$$

where ε is the random observation error, $x_j = (x_{1j}, x_{2j}, ..., x_{mj})$ represents the predictors and β_j is the regression coefficient. For simplicity, the predictor and response variables are standardized (Yan and Su 2009; Januaviani et al. 2019). Equation (8) can be written in matrix form as follows:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{9}$$

where $[\boldsymbol{y}]_{m\times 1}$ is a vector of the response variable, $[\boldsymbol{X}]_{m\times \mathrm{p}}$ is a matrix of the predictor variables, $[\boldsymbol{\beta}]_{p\times 1}$ is a vector of the regression coefficient's predictor variables, and $[\boldsymbol{\varepsilon}]_{m\times 1}$ is a vector of the random observation errors.

EN regression aims to shrink the coefficient regression variable (β_j) selection toward zero or to be exactly equal to zero $\left\{\beta_j \ = 0 : \ j = 1, 2, ..., p\right\}$ and to eliminate the variables of this coefficient from the model by penalizing the coefficient regression variables, which are called the penalty term. The other coefficients still have nonzero $\left\{\beta_j \ \neq 0 : \ j = 1, 2, ..., p\right\}$. Thus, EN is called spare modeling (i.e. the number of coefficients is equal to zero). After the shrinking process, other regression nonzero coefficients are selected to build off the fitting regression model and thereby reduce prediction errors whilst dealing with the multicollinearity between predictor variables to produce models free from multicollinearity (Zou and Hastie 2005; Zou and Zhang 2009; Cho et al. 2010).

The vector of the coefficient variables $\boldsymbol{\beta}$ is unknown. Thus, it is estimated using the OLS estimator to obtain the vector regression coefficient $\boldsymbol{\beta}$ depending on the minimized sum of the squared error of prediction (SSE), which is known as the residual sum of squares (RSS) given that it sums up the differences between the actual $\boldsymbol{y}$ and the estimated $\hat{\boldsymbol{y}}$ as follows:

$$RSS\ (\boldsymbol{\beta}) = \sum_{l=1}^{m}\ (y_l - \hat{y}_l)^2 = ||\boldsymbol{y} - \hat{\boldsymbol{y}}||^2 \tag{10}$$

where $\boldsymbol{y}$ is the response vector variable and $\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ is the estimated vector model.

In Equation (10), the OLS estimation form depends on the minimum RSS.

$$\hat{\boldsymbol{\beta}}_{(\mathrm{OLS})} = \mathrm{argmin}_{\boldsymbol{\beta}}||\boldsymbol{y} - \hat{\boldsymbol{y}}||^2 \tag{11}$$

The OLS regression estimation has a number of drawbacks, such as the presence of multicollinearity between the predictor variables, low prediction accuracy, and difficulty in reducing the number of predictor variables. These problems are tackled with the ridge regression by adding the L_2 penalty (i.e., the sum of the squared coefficient variables) (Hoerl and Kennard 1970). However, ridge regression still cannot deal with the reduction of predictor numbers by shrinking the coefficients, that is, unnecessary predictor variables still exist in the model. Thus, LASSO regression is proposed by adding the L_1 penalty (i.e. the sum of absolute values of the coefficient variables) (Tibshirani 1996). However, the LASSO regression does not select variables when high multicollinearity exists (Zou and Hastie 2005; Zou and Zhang 2009). Hence, EN regression is proposed to deal with the limitations of LASSO and thereby improve the interpretability model and accuracy prediction by combining two penalties, namely, L_1 penalty (LASSO) and L_2 penalty (Ridge) (Zou and Hastie 2005; Zou and Zhang 2009; Friedman, Hastie, and Tibshirani 2010; Lee, Nguyen, and Wang 2016).

EN estimation is a penalized least squared estimator that includes two penalties, namely, L_1 penalty and L_2 penalty.

$$\hat{\boldsymbol{\beta}}_{(EN)} = (1 + \lambda_2)\left[\mathrm{argmin}_{\boldsymbol{\beta}}||\boldsymbol{y} - \hat{\boldsymbol{y}}||^2 + \lambda_1||\boldsymbol{\beta}|| + \lambda_2||\boldsymbol{\beta}||^2\right] \tag{12}$$

where $||\boldsymbol{\beta}|| = \sum_{j=1}^{p}|\beta_j\ |$ is the L_1_norm of β, $||\boldsymbol{\beta}||^2 = \sum_{j=1}^{p}\left(\beta_j\ \right)^2$ is the L_2_norm of $\boldsymbol{\beta}$ and the parameters $\lambda 1$ and λ_2 are tuning parameters and positive numeric values $(\lambda 1,\ \lambda_2\ > 0)$, respectively. They control the strength of shrinkage of the predictor variables. The values of the tuning parameters are dependent on the dataset, and they are automatically selected using cross-validation (Zou and Hastie 2005; Melkumova and Shatskikh 2017; Masselot et al. 2018). The best values of the tuning parameters λ_1 and λ_2 can be defined as the minimum mean square error (MSE) (Friedman, Hastie, and Tibshirani 2010; Lee, Nguyen, and Wang 2016). Equation (12) is equivalent to the optimization equation as follows:

$$\hat{\boldsymbol{\beta}}_{(EN)} = \mathrm{argmin}_{\boldsymbol{\beta}}\left[\frac{1}{2m}||\boldsymbol{y} - \hat{\boldsymbol{y}}||^2 + \lambda\ \left(\alpha||\boldsymbol{\beta}|| + \frac{1-\alpha}{2}||\boldsymbol{\beta}||^2\ \right)\right] \tag{13}$$

where $\alpha||\boldsymbol{\beta}|| + \frac{1-\alpha}{2}||\boldsymbol{\beta}||^2$ is the EN penalty, and alpha (α) is a regularization parameter between 0 and 1 (i.e. $0 \le \alpha\ \le 1$); when $\alpha = 0$, the EN estimation undergoes ridge regression, whereas when $\alpha = 1$, the EN estimation undergoes LASSO regression. Therefore, the range of EN penalty is between 0 and 1.

By using the numerical optimization method namely, coordinate descent (COD) method and soft-thresholding operator (Friedman, Hastie, and Tibshirani 2010) for solving the Equation (13) with the given α and λ values. The COD method will be used to optimize each predictor that solves exactly for one predictor $\boldsymbol{X}_j$ while the rest of predictors $\boldsymbol{X}_f$ except the jth predictor will be fixed (Gauraha 2018). Then the Equation (13) can be written as follows:

$$\hat{\boldsymbol{\beta}}_{(EN)} = \text{argmin}_{\boldsymbol{\beta}}\left[\frac{1}{2m}\left|\left|\boldsymbol{y} - \boldsymbol{X}_f\beta_f - \boldsymbol{X}_j\beta_j\right|\right|^2 + \lambda\left(\alpha||\boldsymbol{\beta}|| + \frac{1-\alpha}{2}||\boldsymbol{\beta}||^2\right)\right];\ f \neq j. \tag{14}$$

Hence, $\boldsymbol{y} - \boldsymbol{X}_f\beta_f$ is the partial residual r_f. Then, the Equation (14) becomes as follows:

$$\hat{\boldsymbol{\beta}}_{(EN)} = \text{argmin}_{\boldsymbol{\beta}}\left[\frac{1}{2m}\left|\left|r_f - \boldsymbol{X}_j\beta_j\right|\right|^2 + \lambda\left(\alpha||\boldsymbol{\beta}|| + \frac{1-\alpha}{2}||\boldsymbol{\beta}||^2\right)\right]. \tag{15}$$

Now, by using the COD to calculate Equation (15), the partial derivative of the Equation (15) with respect to β_j as follows:

$$\frac{\partial\hat{\boldsymbol{\beta}}_{(EN)}}{\partial\beta_j} = -\frac{1}{m}\boldsymbol{X}_j{}^t\left(r_f - \boldsymbol{X}_j\hat{\beta}_j\right) + \lambda\alpha\ \text{sign}\left(\hat{\beta}_j\right) + \lambda(1-\alpha)\hat{\beta}_j. \tag{16}$$

By solving Equation (16) in terms of $\hat{\beta}_j$:

$$\hat{\beta}_j = \frac{S\left(\frac{1}{m}\boldsymbol{X}_j{}^t r_f,\ \lambda\alpha\right)}{1 + \lambda(1-\alpha)} \tag{17}$$

where $\frac{1}{m}\boldsymbol{X}_j{}^t r_f$ is the simple OLS method to estimate the coefficient β_j of the predictor jth, and $S\left(\frac{1}{m}\boldsymbol{X}_j{}^t r_f, \lambda\alpha\right)$ is the soft-thresholding function with the value:

$$S\left(\frac{1}{m}\boldsymbol{X}_j{}^t r_f,\ \lambda\alpha\right) = \begin{cases} \frac{1}{m}\boldsymbol{X}_j{}^t r_f + \lambda\alpha & \textit{if}\ \frac{1}{m}\boldsymbol{X}_j{}^t r_f < -\lambda\alpha \\ 0 & \textit{if}\ -\lambda\alpha \leq \frac{1}{m}\boldsymbol{X}_j{}^t r_f \leq \lambda\alpha \\ \frac{1}{m}\boldsymbol{X}_j{}^t r_f - \lambda\alpha & \textit{if}\ \frac{1}{m}\boldsymbol{X}_j{}^t r_f > \lambda\alpha \end{cases} \tag{18}$$

2.3. Multicollinearity

Multicollinearity occurs when two or more predictor variables contain the same information. In mathematical terms, when the matrix $\boldsymbol{X}^t\mathbf{X}$ is close to singular, a high correlation exists between the predictor variables; such correlation tends to increase the variances of estimated regression coefficients and leads to an unstable estimation (Alin 2010; Cho et al. 2010; Jadhav, Kashid, and Kulkarni 2014).

The variance inflation factor (VIF) test is used to check the multicollinearity. A VIF value less than 10 (i.e. VIF $<$ 10) indicates no multicollinearity amongst predictor variables; the VIF form is as follows:

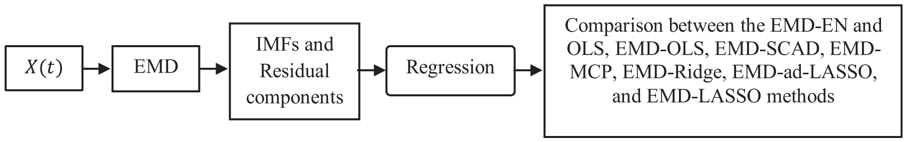

Figure 2. EMD-Regression method.

$$\mathrm{VIF}_j = \frac{1}{1 - R_j^2};\ j = 1, 2, \ldots, p \tag{19}$$

where R_j^2 is the coefficient of determination and p represents the predictor variables.

2.4. Proposed EMD-EN method

The EMD-EN method used for model building with two variables of the data is designed in two steps:

Step 1. Decompose the predictor variable (original signal) $X(t)$ by using the EMD algorithm into a finite set of orthogonal components IMF $\{\mathrm{IMF}_i(\mathrm{t}) : \ i = 1, 2, \ldots, n\}$ and one residual component $r(t)$, where $X(t)$ is equal to the summation of all decomposition components represented by Equation (3).

Step 2. Use all the decomposed components obtained from the predictor variable $X(t)$ via EMD in Step 1 as new predictor variables in EN regression with the response variable $y(t)$ to enhance the predication accuracy by selecting the subset of the decomposed components with the strongest effect on $y(t)$ using the following formula:

$$\hat{\boldsymbol{\beta}}_{(EN)} = \mathrm{argmin}_{\boldsymbol{\beta}} \left[\frac{1}{2m} \left\| y(t) - \beta_0 - \sum_{i=1}^{n} \beta_i \ \mathrm{IMF}_i(t) - \beta_{n+1} r(t) \right\|^2 \right] + \lambda H_\alpha(\beta) \tag{20}$$

where $H_\alpha(\beta) = \alpha \sum_{i=1}^{n+1} |\beta_i| + \left(\frac{1-\alpha}{2}\right) \sum_{i=1}^{n+1} (\beta_i\ )^2$.

Finally, we compare the proposed EMD-EN method with traditional methods, namely, OLS between the original predictor variable $X(t)$ and the response variable $y(t)$. Then, OLS, smoothly clipped absolute deviation (SCAD) (Fan and Li 2001), minimax concave penalty (MCP) (Zhang 2010), ridge regression, adaptive LASSO (ad-LASSO) (Zou 2006), and LASSO are compared in terms of the decomposed components of the predictor variable $X(t)$ via EMD and response variable $y(t)$ by using several criteria tests. The input data of the new predictor variables and response variables are standardized prior to regression. The analysis process is summarized in Figure 2.

2.5. Test criteria

Four test criteria are used to compare the performances of the estimated models of the seven methods with those of the EMD-EN method. The test criteria are the RSS (Equation (10)), root mean square error (RMSE; Equation (21)), mean absolute error (MAE; Equation (22)) and mean absolute scaled error (MASE; Equation (23)).

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{l=1}^{m}(y_l - \hat{y}_l)^2} \tag{21}$$

$$\text{MAE} = \frac{1}{m}\sum_{l=1}^{m}|y_l - \hat{y}_l| \tag{22}$$

$$\text{MASE} = \frac{1}{m}\sum_{l=1}^{m}\left(\frac{|y_l - \hat{y}_l|}{\frac{1}{m-1}\sum_{l=2}^{m}|y_l - y_{l-1}|}\right) \tag{23}$$

3. Data of study

In this section, we explain in detail the data used to apply the methods, namely, simulation data and actual time series data.

3.1. Numerical experiments

We technically implement the EMD-EN model and compare it with existing regression models (i.e. OLS, EMD-OLS, EMD-SCAD, EMD-MCP, EMD-Ridge, EMD-ad-LASSO, and EMD-LASSO). The sine wave function is used to generate the simulation dataset; here, predictor and response variables are simulated from actual signals selected from the work of Qin et al. (2016) with noise structure errors, namely, normal distribution with zero mean and unity variance $\varepsilon \sim iid\ N(0,1)$, added to predictor variable $X(t)$.

$$\begin{aligned}
X(t) &= 0.5t + \sin(\pi t) + \sin(2\pi t) + \sin(6\pi t) + \varepsilon; \quad t \in [0, 10] \\
y_1(t) &= 5\sin(6\pi t) + \sin(3\pi t) + \sin(8\pi t) + \sin(12\pi t) + \sin(13\pi t) \\
y_2(t) &= 5\sin(2\pi t) + \sin(3\pi t) + \sin(8\pi t) + \sin(12\pi t) + \sin(13\pi t)
\end{aligned}$$

The dataset is simulated with sample size $n = 250$. In each experiment, the 2000 replications of the sample size of 250 are modeled. The value of the tuning parameter λ will be automatically chosen by using 10-fold cross-validation. RSS, RMSE, MAE, and MASE are the numerical measures used to evaluate the quality of the estimation methods.

3.2. Application datasets

Two empirical applications are used to evaluate the performance of the proposed EMD-EN method relative to other methods. The first application has two variables, namely, the daily close price index (predictor variable) and the daily exchange rate (response variable) from March 1, 2011 to August 28, 2015 for Malaysia; the number of observations is 1150. The second application has two variables, namely, the daily closing prices of the Shanghai composite index (predictor variable) and the Shenzhen component index (response variable) from January 4, 2006 to January 12, 2011 of Chinese stock markets; the number of observations is 1222. All datasets are collected from the finance database (https://finance.yahoo.com/). Each dataset is divided into two parts: 70% is used for training, and the remaining 30% is used for testing. In general, the exchange

Table 1. Performance criteria in the numerical experiments.

	OLS	EMD-OLS	EMD-SCAD	EMD-MCP	EMD-Ridge	EMD-ad-LASSO	EMD-LASSO	EMD-EN
Experiment 1 (y_1(t))								
RSS	70.7160	39.15038	38.87427	38.88486	39.2541	38.59161	38.84474	**38.29632**
RMSE	0.96994	0.713548	0.711005	0.711072	0.714938	0.708604	0.711121	**0.706078**
MAE	0.84352	0.58282	**0.58010**	0.58022	0.59165	0.58094	0.58686	0.58217
MASE	1.1674	0.8019	0.7979	**0.7978**	0.8140	0.7989	0.8067	0.8005
Experiment 2 (y_2(t))								
	OLS	EMD-OLS	EMD-SCAD	EMD-MCP	EMD-Ridge	EMD-ad-LASSO	EMD-LASSO	EMD-EN
RSS	69.1147	44.2863	43.63225	43.57012	44.33144	43.35498	43.74513	**43.17313**
RMSE	0.95963	0.766336	0.760653	0.760138	0.767025	0.758428	0.762052	**0.757026**
MAE	0.81162	0.62050	0.61543	**0.61507**	0.62524	0.61549	0.62122	0.61646
MASE	0.7028	0.5349	0.5306	**0.5304**	0.5386	0.5306	0.5353	0.5315

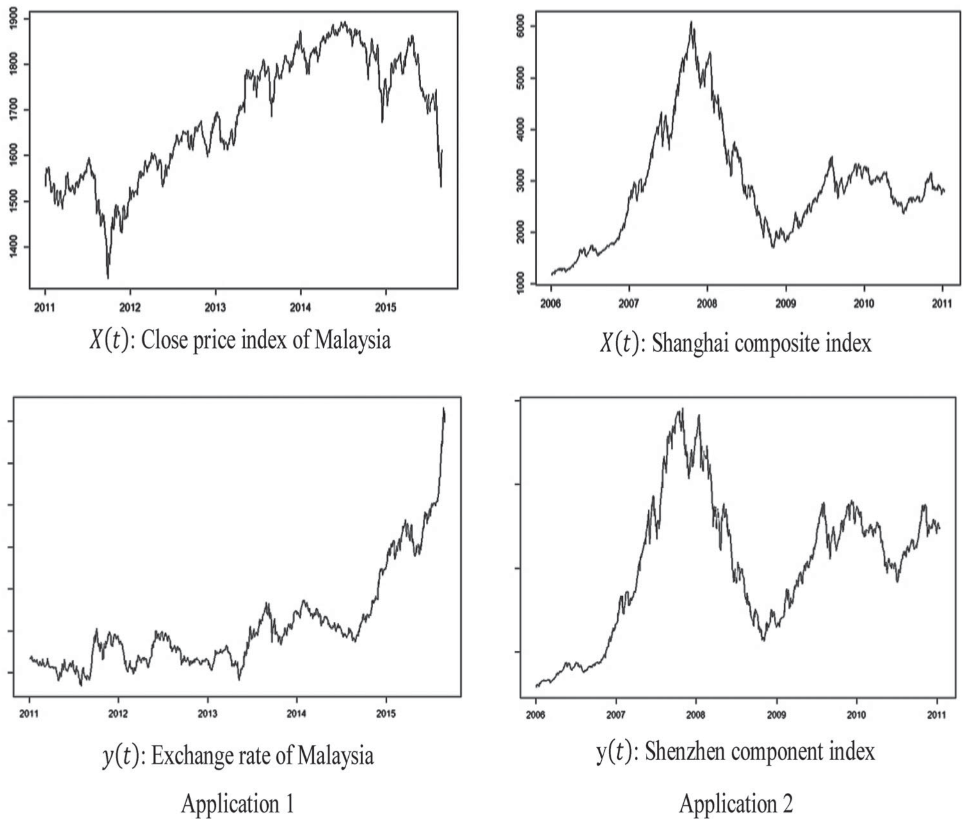

Figure 3. Plots of original signal $X(t)$ and $y(t)$ for two applications.

rate and stock market time series data often have autocorrelation, but the main purpose of the EMD-EN method is to understand a relationship between the decomposition components and response variable from a different point a view (multicollinearity point view).

The analyses are performed using the open-source R software (Team 2015) that uses an EMD package (Kim and Oh 2009) for signal analysis and the *glmnet* package (Friedman, Hastie, and Tibshirani 2010) for estimation.

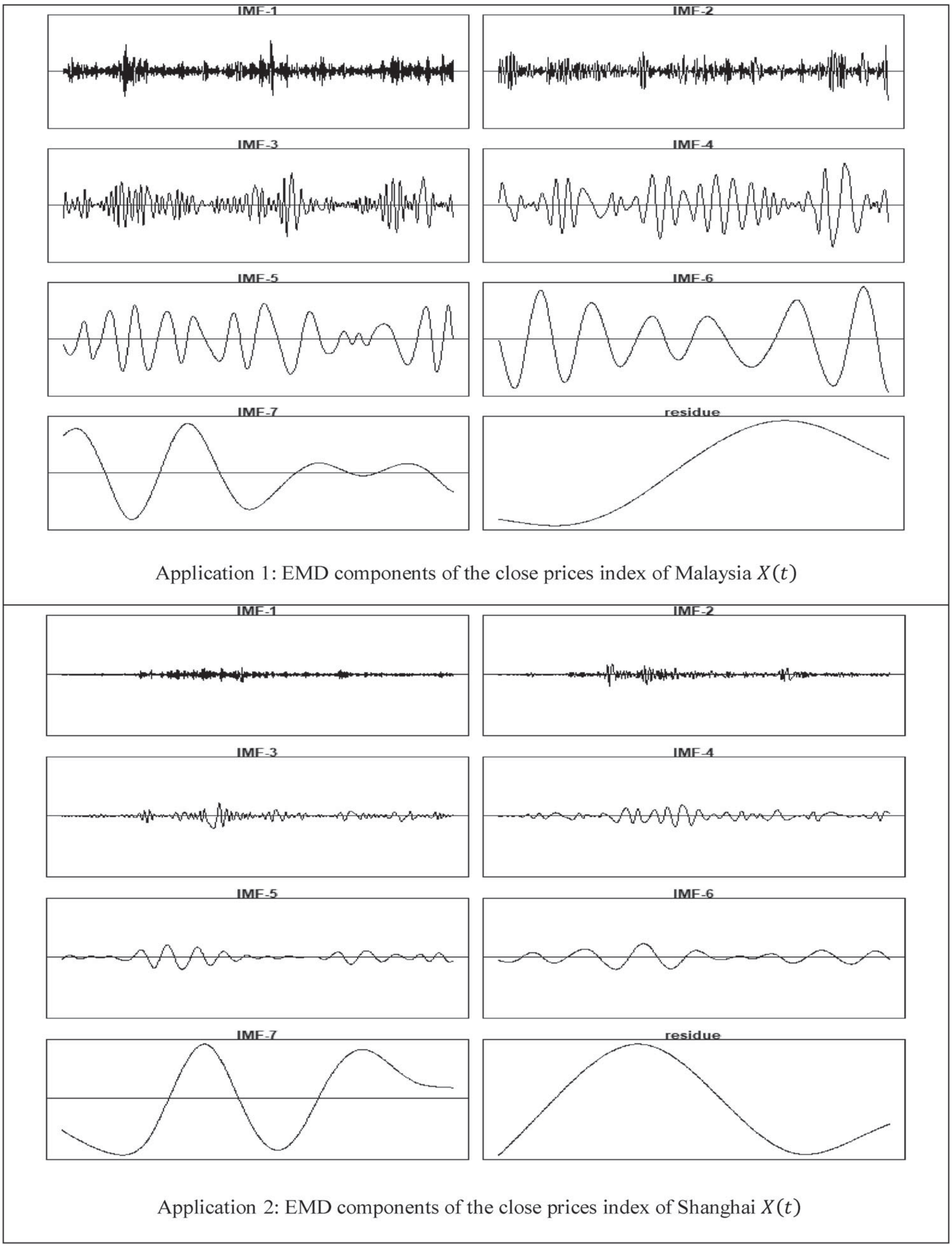

Figure 4. Decomposition of the original signal $X(t)$ via EMD in two applications.

Table 2. Variance inflation factors.

	VIF_1	VIF_2	VIF_3	VIF_4	VIF_5	VIF_6	VIF_7	VIF_8
Application 1	1.0158	1.0159	1.0185	1.0176	1.0194	1.0232	1.0455	1.0198
Application 2	1.0026	1.0197	1.0686	1.0459	1.0682	1.0352	1.0212	1.0436

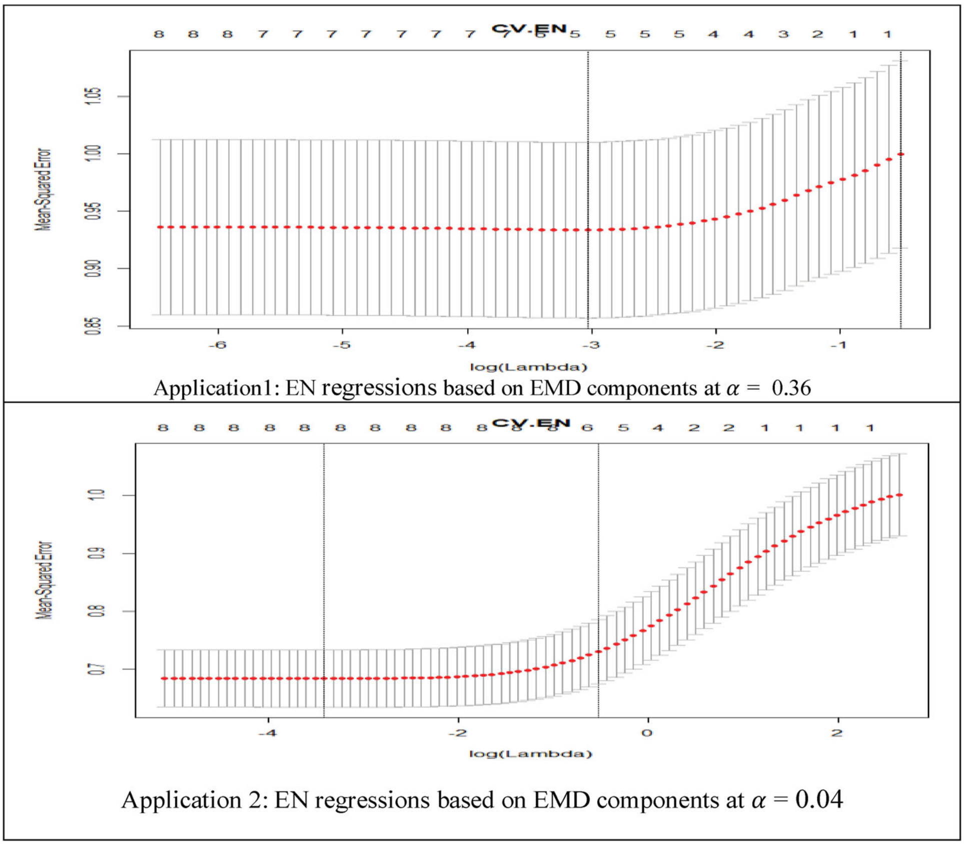

Figure 5. k-fold cross-validation estimation of the mean squared error as the Log (λ) for Elastic net regressions based on EMD at $k = 10$.

4. Results and discussion

The analysis and discussion results using the simulation and actual time series data are proposed.

4.1. Numerical results

Table 1 illustrates the performance criteria for all technical regression methods in our study. Specifically, we take the average of the performance criteria values (RSS, RMSE, MAE, and MASE) in the numerical experiments of the study. The results show that EMD-EN has the smallest error value in terms of RSS and RMSE. However, for MAE and MASE, a negligible difference is noted between the mean performance criteria of the proposed method and those of the EMD-SCAD, EMD-MCP, and EMD-ad-LASSO methods in the two experiments. The EMD-EN method has the highest reliability with high prediction accuracy.

4.2. Application results

Figure 3 presents the original signal of the predictor variables $X(t)$ and response variables $y(t)$ for the two applications of the study. In the first application, $X(t)$ represents

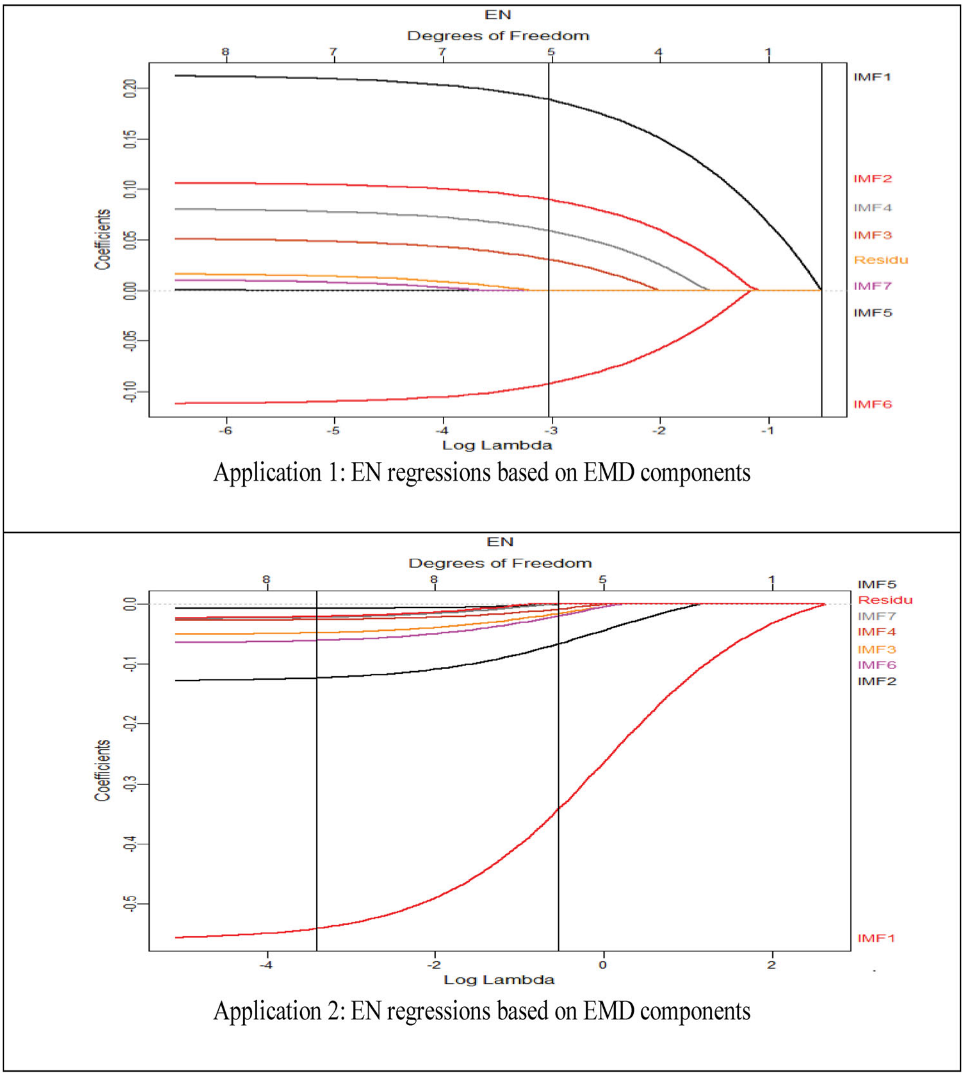

Figure 6. Coefficient estimation in two applications as Log (λ) function for EMD-EN regression methods by using 10-fold cross-validation estimation.

the daily close price index, and $y(t)$ represents the exchange rate of Malaysia. The second application $X(t)$ represents the daily close prices of the Shanghai composite index, and $y(t)$ represents the Shenzhen component index. The plots of the signals in the two applications show neither a straight line nor a constant value over time; such conditions are called nonlinear and nonstationary conditions, respectively.

Figure 4 shows the decomposition components of the sifting process for $X(t)$ in the first and second applications, in which the number of IMFs produced is equal to seven components and one residual component. All the extracted IMFs have different characteristics. Thus, IMF_1 has the shortest wavelength, maximum amplitude and highest frequency. This characteristic varies amongst the IMF plots in which the amplitude and frequency decrease whilst the wavelength increases. The last component is the

Table 3. RSS values in the train and test parts.

			RSS_{Train}	RSS_{Test}
		Application 1		
OLS			784.2697	354.1116
EMD-OLS			735.6008	333.0507
EMD-SCAD		$\lambda = 0.01621980$	735.9146	332.8378
EMD-MCP		$\lambda = 0.01999653$	735.9066	332.7999
EMD-Ridge	min$M$	$\lambda = 0.11442747$	736.2682	331.2642
	1se	$\lambda = 214.416181$	798.4079	338.8513
EMD-ad-LASSO	min$M$	$\lambda = 0.11740866$	736.0851	331.3869
	1se	$\lambda = 9.30443575$	799.0000	339.0000
EMD-LASSO	min$M$	$\lambda = 0.01739196$	737.1153	329.2542
	1se	$\lambda = 0.21441619$	799.0000	339.0000
EMD-EN	**min$M$**	$\boldsymbol{\lambda} = 0.0483109$	737.5599	**328.9841**
	1se	$\lambda = 0.59560052$	799.0000	339.0000
		Application 2		
OLS			829.9940	387.1897
EMD-OLS			565.0201	308.6346
EMD-SCAD		$\lambda = 0.07351708$	576.3729	312.2884
EMD-MCP		$\lambda = 0.07882995$	574.1943	309.3980
EMD-Ridge	min$M$	$\lambda = 0.05561297$	565.8339	305.4808
	1se	$\lambda = 0.68562272$	612.7504	302.9398
EMD-ad-LASSO	min$M$	$\lambda = 0.0333371$	567.6307	304.5973
	1se	$\lambda = 2.89948849$	614.3004	311.9243
EMD-LASSO	min$M$	$\lambda = 0.0235201$	568.2740	303.6796
	1se	$\lambda = 0.1821074$	614.3004	311.9243
EMD-EN	min$M$	$\lambda = 0.0328740$	565.3667	306.3329
	1se	$\boldsymbol{\lambda} = \mathbf{0.5880019}$	612.0503	**301.7346**

residual $r(t)$, which represents the trend of the original signal $X(t)$. Thus, the decomposition components have physical meaning.

The VIFs are used to test for multicollinearity amongst all the orthogonal IMF (from VIF_1 to VIF_7) components and VIF_8 of the residue component in the two applications. Table 2 shows the values of VIF for all components from VIF_1 to VIF_8 at less than 10. Therefore, no multicollinearity exists between all orthogonal IMFs and residue components.

Figure 5 shows the EN regression plot based on the EMD components in the two applications. The y-axis represents the estimation of the MSE, and the x-axis represents the $Log\ (\lambda)$ function. The upper horizontal line of the plot represents the numbers of nonzero regression coefficients for a given $Log\ (\lambda)$. The vertical dotted lines from left to right represent the first vertical line, which is the location of the minimum of the MSE (min$M$) rule. The second vertical line is the point selected by the one standard error (1se) rule. It shows the numbers of nonzero regression coefficient selected at min$M$ and 1se rules. When the value of λ is increased, the number of nonzero coefficients decreases in the model regression.

Figure 6 illustrates the relationship between $\log(\lambda)$ and the selected nonzero coefficient estimation in the EN regression based on EMD at current λ, which is the actual degrees of freedom. In the two applications, the nonzero coefficient estimation is different from the selected value of λ at min$M$ or 1se rules. The EN regression has regularization and selected variables. As the λ value increases, the estimation of coefficients shrinks toward zero and forces to become zero for unnecessary coefficient estimation (i.e. as $\lambda \to \infty$, then nonzero coefficients $\to 0$). In the first application, when the

Table 4. Coefficients estimation for the predictor variables in two applications.

Application 1: Regressions based on EMD components and response variable exchange rate of Malaysia							
	EMD-OLS	EMD-SCAD $\lambda = 0.0162198$	EMD-MCP $\lambda = 0.0199965$	EMD-Ridge min$M$ $\lambda = 0.1144275$	EMD-ad-LASSO min$M$ $\lambda = 0.1174087$	EMD-LASSO min$M$ $\lambda = 0.0173919$	EMD-EN min$M$ $\lambda = 0.0483109$
$\hat{\beta}_1$	0.21295721	0.2123702	0.2124173	0.1910800	0.2094858	0.1948450	0.1890288
$\hat{\beta}_2$	0.10686351	0.1084861	0.1084100	0.09726106	0.1035299	0.09238165	0.08994588
$\hat{\beta}_3$	0.05147557	0.04401691	0.04493453	0.04526517	0.03841548	0.03194924	0.03077258
$\hat{\beta}_4$	0.08104872	0.07988394	0.07990641	0.07169361	0.07235490	0.06138036	0.05930373
$\hat{\beta}_5$	0.00143060	0	0	0.00063227	0	0	0
$\hat{\beta}_6$	−0.1126013	−0.11319	−0.1132214	−0.1008141	−0.1078071	−0.09516301	−0.09.223514
$\hat{\beta}_7$	0.01080601	0	0	0.00929059	0	0	0
$\hat{\beta}_8$	0.01687016	0	0	0.01513029	0	0	0
Application 2: Regressions based on EMD components and response variables Shenzhen component index							
	OLS	EMD-SCAD $\lambda = 0.073517$	EMD-MCP $\lambda = 0.0788299$	EMD-Ridge 1se $\lambda = 0.6856227$	EMD-ad-LASSO min$M$ $\lambda = 0.0333371$	EMD-LASSO min$M$ $\lambda = 0.0235201$	EMD-EN 1se $\lambda = 0.5880019$
$\hat{\beta}_1$	−0.5601209	−0.55735	−0.557674	−0.331006	−0.557473	−0.5355650	−0.3414914
$\hat{\beta}_2$	−0.1285716	−0.05637375	−0.076594	−0.074221	−0.121019	−0.1063091	−0.0657540
$\hat{\beta}_3$	−0.0511991	0	0	−0.028306	−0.023523	−0.0245480	−0.0151887
$\hat{\beta}_4$	−0.0272796	0	0	−0.019708	0	−0.0070508	−0.0078805
$\hat{\beta}_5$	−0.0071676	0	0	−0.01192382	0	0	0
$\hat{\beta}_6$	−0.0649254	0	0	−0.03179910	−0.04132445	−0.03818629	−0.01915252
$\hat{\beta}_7$	−0.0245027	0	0	−0.01331322	0	0	0
$\hat{\beta}_8$	−0.0232208	0	0	−0.01089046	0	0	0

Table 5. Performance criteria.

	Application 1							
	OLS	EMD-OLS	EMD-SCAD	EMD-MCP	EMD-Ridge min*M*	EMD-ad-LASSO min*M*	EMD-LASSO min*M*	EMD-EN min*M*
RMSE	1.0205	0.9897	0.9894	0.9894	0.9871	0.9873	0.9841	**0.9837**
MAE	0.7401	0.7212	0.7211	0.7211	0.7179	0.7193	0.7160	**0.7153**
MASE	0.6795	0.6760	0.6759	0.6758	0.6729	0.6742	0.6711	**0.6705**
	Application 2							
	OLS	EMD-OLS	EMD-SCAD	EMD-MCP	EMD-Ridge 1*se*	EMD-ad-LASSO min*M*	EMD-LASSO min*M*	EMD-EN 1*se*
RMSE	1.0244	0.9145	0.9199	0.9157	0.9061	0.9086	0.9072	**0.9042**
MAE	0.7134	0.6231	0.6222	0.6200	0.6165	0.6168	0.6154	**0.6148**
MASE	0.7310	0.6528	0.6518	0.6495	0.6458	0.6462	0.6447	**0.6441**

exchange rate of Malaysia is the response variable, the selected nonzero coefficients of the decomposition components at min$M$ rule, including five nonzero coefficients, are selected with different significant strengths; the IMF_1, IMF_2, and IMF_6 components have significantly greater strength than the other components. However, at 1*se* rule, the IMF_1 component is only significant on the exchange rate. In the second application of the study, when the Shenzhen component index includes response variables, the selected nonzero coefficients of the components at min$M$ and 1*se* rules include those of five components with different strengths; the IMF_1 and IMF_2 components have considerable strength relative to the response variables.

Table 3 displays the RSS error values of the proposed EMD-EN method compared with those of the previous methods in the training and testing datasets. The smallest RSS value is noted in the training dataset and is achieved by EMD-OLS regression because EMD-OLS does not have any penalties on the coefficients and is thus different from the EMD-SCAD, EMD-MCP, EMD-Ridge, EMD-ad-LASSO, EMD-LASSO, and EMD-EN. In the testing dataset, the smallest value of RSS in the two applications is achieved by EMD-EN in the first application at the min$M$ rule $\lambda = 0.04831099$ and in the second application at 1*se* $\lambda = 0.5880019$. The RSS value in the testing dataset provides the best methods to select and support the fitting regression models. EMD-EN performs better than the other previous regression methods in the two applications.

Table 4 explains the estimation of the coefficients of the predictor variables (i.e., IMFs and residual components) for each of the regression techniques based on the RSS value. Most regression methods can reduce the number of predictor variables, except for OLS-EMD and EMD-Ridge; these methods have the same numbers of coefficient, that is, eight nonzero coefficients, and thus, all the predictor variables are entered into the final model. The numbers of nonzero coefficients of EMD-EN and other regression methods in the study are different from those of EMD-Ridge and EMD-OLS. For the EMD-EN in the first and second applications, the coefficient estimation equal to five nonzero coefficients of IMF_1, IMF_2, IMF_3, IMF_4, and IMF_6 components enters the EMD-EN regression model. Thus, the proposed EMD-EN method EMD-EN has relative great regularization and selected predictor variables with minimum RSS values.

Table 5 explains the prediction accuracy of all regression methods in the two applications by using RMSE, MAE, and MASE for the testing dataset depending on the minimum RSS values at min$M$ and 1*se* rules. The proposed EMD-EN provides the smallest error value in terms of RMSE, MAE, and MASE. The EMD-EN has lower

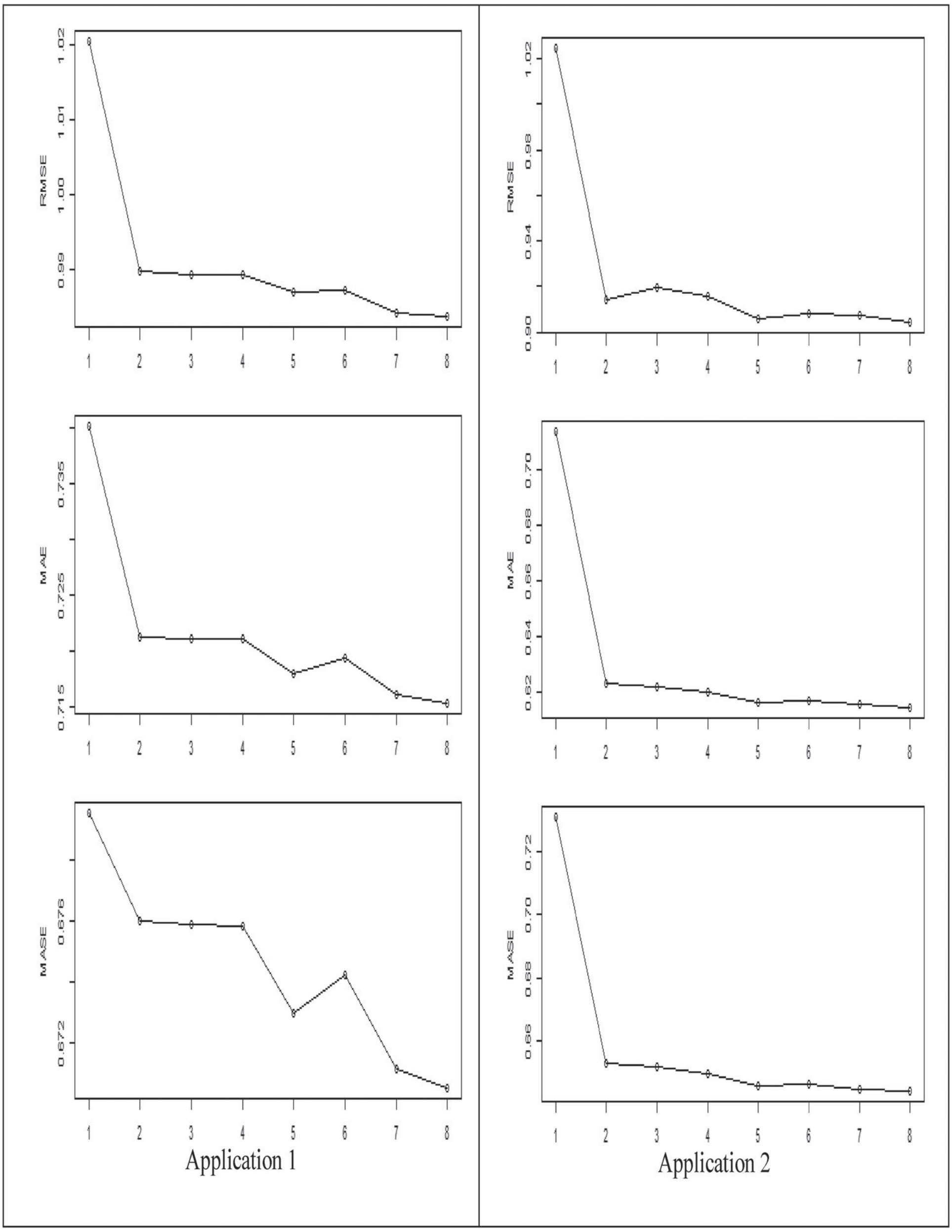

Figure 7. Plots of performance criteria in two applications.

prediction errors and produces better results than the OLS, EMD-OLS, EMD-SCAD, EMD-MCP, EMD-Ridge, EMD-ad-LASSO, and EMD-LASSO.

Figure 7 presents the performance criteria plots for all regression methods in the two applications. The plots of the performance criteria show that EMD-EN has low prediction error in terms of RMSE, MAE, and MASE values, indicating the best selected regression models. Therefore, the proposed EMD-EN method is a highly accurate regression model and produces the best estimates for fitting models.

5. Conclusions

We applied the proposed EMD-EN regression method that combines nonstationary and nonlinear datasets. The method was used to identify the relationship between the orthogonal decomposition components of the predictor variables and the response variable and to deal with multicollinearity between the decomposition components. The EMD-EN method identified the decomposition components of the predictor variable via the EMD algorithm. Then, the VIF test was used for the multicollinearity test between the IMFs and residual components. Lastly, we select the best orthogonal IMFs and residual components with the greatest strengths for the response variable, then, the EN regression was used.

Numerical experiments by simulation and actual time series data were carried out. The results showed that the EMD-EN method effectively selected the actual orthogonal IMFs and residual components which were most significant for the response variable and reduced the prediction error. The EMD-EN method more effective than the other methods. It selected the best fitting model with high prediction accuracy.

In the future, we will focus on several nonlinear and nonstationary predictor variables (i.e. two or more predictors) and the EMD-EN method to detect and deal with the internal relationships between decomposed components when two or more predictor variables are available. EN is a linear model, and the relationship between actual dataset is nonlinear in many situations. Thus, we will build other flexible nonlinear models for nonlinear and actual dataset. Moreover, we will also extend the proposed method for using in time series regression where the error term follows AR, MA, or ARIMA processes.

Acknowledgment

The authors thank the anonymous referees for their valuable recommendations that helped to improve the quality of the manuscript.

ORCID

Ahmed R. M. Alsayed http://orcid.org/0000-0002-6893-3144

References

Alin, A. 2010. Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics* 2 (3): 370–4. doi:10.1002/wics.84.

Cho, S., K. Kim, Y. J. Kim, J.-K. Lee, Y. S. Cho, J.-Y. Lee, B.-G. Han, H. Kim, J. Ott, and T. Park. 2010. Joint identification of multiple genetic variants via elastic-net variable selection in a genome-wide association analysis. *Annals of Human Genetics* 74 (5):416–28. doi:10.1111/j.1469-1809.2010.00597.x.

Chu, H., J. Wei, and J. Qiu. 2018. Monthly streamflow forecasting using EEMD-Lasso-DBN method based on multi-scale predictors selection. *Water* 10 (10):1486.

Chui, C. K. 1995. *Wavelet basics*. Boston, MA: Kulwer Academic Publishers.

Crowley, P. M. 2012. How do you make a time series sing like a choir? Extracting embedded frequencies from economic and financial time series using empirical mode decomposition. *Studies in Nonlinear Dynamics & Econometrics* 16:1–31.

Deering, R., and J. F. Kaiser. 2005. The use of a masking signal to improve empirical mode decomposition. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05) 2005*, iv/485-iv/88, Vol. 4, Philadelphia, PA: IEEE.

Fan, J., and R. Li. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96 (456):1348–60. doi:10.1198/016214501753382273.

Friedman, J., T. Hastie, and R. Tibshirani. 2010. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33 (1):1–22.

Gauraha, N. 2018. Introduction to the LASSO. *Resonance* 23 (4):439–64. doi:10.1007/s12045-018-0635-x.

Hoerl, A. E., and R. W. Kennard. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12 (1):55–67. doi:10.1080/00401706.1970.10488634.

Huang, N. E. 2014. Introduction to the Hilbert–Huang transform and its related mathematical problems. In *Hilbert–Huang transform and its applications*. Singapore: World Scientific.

Huang, N. E., S. Zheng, R. L. Steven, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 454 (1971):903–95. doi:10.1098/rspa.1998.0193.

Jaber, A. M., M. T. Ismail, and M. Altaher Alsaidi. 2013. Automatic boundary correction using empirical mode decomposition and local linear regression. *Far East Journal of Mathematical Sciences* 563:2013.

Jadhav, N. H., D. N. Kashid, and S. R. Kulkarni. 2014. Subset selection in multiple linear regression in the presence of outlier and multicollinearity. *Statistical Methodology* 19:44–59. doi:10.1016/j.stamet.2014.02.002.

Januaviani, T. M. A., N. Gusriani, K. Joebaedi, S. Supian, and S. Subiyanto. 2019. The best model of LASSO with the LARS (least angle regression and shrinkage) algorithm using Mallow's Cp. *World Scientific News* 116:245–52.

Kim, D., and H.-S. Oh. 2009. EMD: A package for empirical mode decomposition and Hilbert spectrum. *The R Journal* 1 (1):40–6. doi:10.32614/RJ-2009-002.

Lee, Y., V. Nguyen, and D. Wang. 2016. On variable and grouped selections of the elastic net. *Report CS532* 961:1–24.

Masselot, P., F. Chebana, D. Bélanger, A. St-Hilaire, B. Abdous, P. Gosselin, and T. B. Ouarda. 2018. EMD-regression for modelling multi-scale relationships, and application to weather-related cardiovascular mortality. *Science of the Total Environment* 612:1018–29. doi:10.1016/j.scitotenv.2017.08.276.

Melkumova, L. E., and S. Y. Shatskikh. 2017. Comparing Ridge and LASSO estimators for data analysis. *Procedia Engineering* 201:746–55. doi:10.1016/j.proeng.2017.09.615.

Moore, K. J., M. Kurt, M. Eriten, D. M. McFarland, L. A. Bergman, and A. F. Vakakis. 2018. Wavelet-bounded empirical mode decomposition for measured time series analysis. *Mechanical Systems and Signal Processing* 99:14–29. doi:10.1016/j.ymssp.2017.06.005.

Naik, J., P. Satapathy, and P. K. Dash. 2018. Short-term wind speed and wind power prediction using hybrid empirical mode decomposition and kernel ridge regression. *Applied Soft Computing* 70:1167–88. doi:10.1016/j.asoc.2017.12.010.

Plakandaras, V., R. Gupta, P. Gogas, and T. Papadimitriou. 2015. Forecasting the US real house price index. *Economic Modelling* 45:259–67. doi:10.1016/j.econmod.2014.10.050.

Qin, L., S. Ma, J.-C. Lin, and B.-C. Shia. 2016. Lasso regression based on empirical mode decomposition. *Communications in Statistics – Simulation and Computation* 45 (4):1281–94. doi:10.1080/03610918.2013.826361.

Rilling, G., P. Flandrin, and P. Gonçalvès. 2003. On empirical mode decomposition and its algorithms. *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing* 3:8–11.

Shen, Z., and C.-H. Lee. 2012. LASSO based ensemble empirical mode decomposition approach to designing adaptive clutter suppression filters. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 757–60.

Team, R. C. 2015. *A language and environment for statistical computing (version 3.2. 2)*. Vienna, Austria: R Foundation for Statistical Computing.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58:267–88. doi:10.1111/j.2517-6161.1996.tb02080.x.

Titchmarsh, E. C. 1948. *Introduction to the theory of Fourier integrals*. Second edition. Oxford: Clarendon Press.

Van der Kooij, A. J., and J. J. Meulman. 2008. Regularization with ridge penalties, the lasso, and the elastic net for regression with optimal scaling transformations. https://openaccess.leidenuniv.nl/bitstream/handle/1887/12096/04.pdf (accessed March 10, 2019).

Yan, X., and X. Su. 2009. *Linear regression analysis: Theory and computing.* Singapore: World Scientific.

Yang, A. C., J.-L. Fuh, N. E. Huang, B.-C. Shia, C.-K. Peng, and S.-J. Wang. 2011. Temporal associations between weather and headache: Analysis by empirical mode decomposition. *PLoS One.* 6 (1):e14612. doi:10.1371/journal.pone.0014612.

Yang, A. C., S.-J. Tsai, and N. E. Huang. 2011. Decomposing the association of completed suicide with air pollution, weather, and unemployment data at different time scales. *Journal of Affective Disorders* 129 (1–3):275–81. doi:10.1016/j.jad.2010.08.010.

Zhang, C.-H. 2010. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics* 38 (2):894–942. doi:10.1214/09-AOS729.

Zou, H. 2006. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* 101 (476):1418–29. doi:10.1198/016214506000000735.

Zou, H., and T. Hastie. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67 (2):301–20. doi:10.1111/j.1467-9868.2005.00503.x.

Zou, H., and H. H. Zhang. 2009. On the adaptive elastic-net with a diverging number of parameters. *Annals of Statistics* 37 (4):1733–51