Elastic Net Regression and Empirical Mode Decomposition for Enhancing the Accuracy of the Model Selection

3 authors:

Abdullah al Jawarneh
Jerash University
10 PUBLICATIONS   **24** CITATIONS

SEE PROFILE

Mohd Tahir Ismail
Universiti Sains Malaysia
238 PUBLICATIONS   **977** CITATIONS

SEE PROFILE

Ahmad M. Awajan
Al-Hussein Bin Talal University
23 PUBLICATIONS   **133** CITATIONS

SEE PROFILE

Elastic Net Regression and Empirical Mode Decomposition for Enhancing the Accuracy of the Model Selection

Abdullah S. Al-Jawarneh
School of Mathematical Sciences,
Universiti Sains Malaysia, Pulau Pinang, Malaysia.
Corresponding author: a_aljawarneh86@student.usm.my; abdssj2004@yahoo.com

Mohd. Tahir Ismail
School of Mathematical Sciences,
Universiti Sains Malaysia,Pulau Pinang, Malaysia.
E-mail: m.tahir@usm.my

Ahmad M. Awajan
Faculty of Science,
Al-Hussein Bin Talal University, Ma'an, Jordan.
E-mail: ahmad.m.awajan@ahu.edu.jo

Abstract
Elastic net (ELNET) regression is a hybrid statistical technique used for regularizing and selecting necessary predictor variables that have a strong effect on the response variable and deal with multicollinearity problem when it exists between the predictor variables. The empirical mode decomposition (EMD) algorithm is used to decompose the nonstationary and nonlinear dataset into a finite set of orthogonal intrinsic mode function components and one residual component. This study mainly aims to apply the proposed ELNET-EMD method to determine the effect of the decomposition components of multivariate time-series predictors on the response variable and tackle the multicollinearity between the decomposition components to enhance the prediction accuracy for building a fitting model. A numerical experiment and a real data application are applied. Results show that the proposed ELNET-EMD method outperforms other existing methods by capable of identifying the decomposition components that have the most significance on the response variable despite the high correlation between the decomposition components and by improving the prediction accuracy.

Keywords- Empirical mode decomposition, Elastic net regression, LASSO regression, Multicollinearity, Ridge regression.

1. Introduction
Several studies such as medicine, and economics interested in using time series datasets, where these datasets are often non-stationary and non-linear simultaneously. However, there is a lack of statistical methods that can effectively extract oscillatory patterns from the data because the traditional methods assumed that the dataset should be either stationary or linear. For example, Fourier decomposition (Titchmarsh, 1948), and wavelet decomposition (Chan, 1994). Recently, Huang et al. (1998) proposed the empirical mode decomposition (EMD) method, which aims to decompose non-stationary and non-linear data with keeping the time domain. The EMD method does not require any restrictions and pre-conditions on the nature of the data, such as stationary or linearity, in contrast to traditional methods.

EMD method divided the original signal into a finite set of decomposition components (i.e., intrinsic mode function (IMF) components and one residual component). Each of the

decomposition components has different information such as wavelength, amplitude, and frequency (Huang, 2014). Therefore, these components can be used as new predictor variables to predict the response variable and improve the prediction accuracy of regression analysis. Meanwhile, multicollinearity may exist among the decomposition components, especially in the case of dealing with original multivariate predictors. Multicollinearity defined as a relationship between the predictor variables, which increases the variance. This results in a wrong sign of coefficients and misleads the selection of a fitting model (Jadhav et al., 2014).

Elastic net (ELNET) regression analysis (Zou and Hastie, 2005) is a hybrid between LASSO (Tibshirani, 1996) and Ridge regression (Hoerl and Kennard, 1970). ELNET regression used to regularize and select the important predictor variables to obtain a simple model with the most significant predictor variables despite high multicollinearity between the predictors. ELNET can remove or select the predictor variables that have a high correlation in the final model and enhance the prediction accuracy (Liu and Li, 2017).

The behavior of the time series variables used in regression analysis, such as nonstationary and nonlinear, and multicollinearity problem may affect the prediction accuracy in model selection. This situation indicates several difficulties in the regression analysis with these existing issues. To address these gaps, the ELNET regression based on the EMD method (ELNET-EMD) is proposed in this study. The purpose of this study is to investigate the effect of the decomposition components of the original multivariate time-series predictors on the response variable and combat the correlation between the decomposition components to obtain accurate and reliable results. Where the EMD will be used to decompose the nonstationary and nonlinear multi-predictors signals of the data separately into a finite set of IMF components and one residual component. After that, all the decomposition components of the multi-predictors via EMD will be used as a new predictor variable in the ELNET regression to select the predictors which have the most effect on the response variable.

The rest of the paper is organized as follows. Section 2 presents the literature review. Section 3 describes the EMD method, ELNET regression, the ELNET-EMD method, and the test criteria. Section 4 discusses and applies the method through numerical experiments and gives an analysis. Section 5 illustrates the method by applying exchange rate data and provides an analysis and a discussion of the findings. Section 6 elaborates the conclusions of the study.

2. Literature Review
Recent studies have focused on using the EMD method combined with other established statistical regression or forecasting methods. This method has been successfully applied in several scientific fields. For instance, multiple linear regression models with forward stepwise regression (SR) methods based on EMD (Yang et al., 2011). The novel forecasting model based on the neural network and EMD (Chen et al., 2012). The least absolute shrinkage and selection operator (LASSO) regression based on ensemble EMD (EEMD) (Shen and Lee, 2012).

Moreover, LASSO regression and deep belief networks based on EEMD (Chu et al., 2018). The ELNET and the support vector regression (SVR) model based on EEMD (Plakandaras et al., 2015). LASSO regression based on decomposition components via EMD method (Qin et al., 2016). ELNET regression based on EMD method proposed to deal with nonlinear and non-stationary original univariate time-series predictor case (Al-Jawarneh et al., 2020). Kernel Ridge regression based on the decomposition components via EMD (Naik et al., 2018). Furthermore, LASSO

regression and noise-assisted multivariate EMD (NA-MEMD) (Masselot et al., 2018), SVR and particle swarm optimization based on EMD (Hong and Fan, 2019).

3. Methodology
This section briefly describes the applied methods. The first method is the EMD method by sifting process to decompose the original signal. The second method pertains to the technical penalized regularization method by ELNET regression. This section also discusses the proposed ELNET-EMD method.

3.1 EMD Method
Huang et al. proposed the EMD in 1998. The EMD focuses on the nonstationary and nonlinear time-series datasets by using the sifting process of the technical algorithm to decompose the original signal into a finite set of orthogonal decomposition components. Such components have different oscillatory patterns and by keeping the time domain of the signal unchanged. Their orthogonal decomposition components are called IMF components and one residual component (Huang, 2014; Moore et al., 2018).

3.1.1 Sifting Process
The iterative algorithm process of the EMD for extracting all the IMF components and one residual component is called the sifting process. It separates the original signal into orthogonal components of a non-overlapping time scale (Huang, 2014). The sifting process to decompose the original signal $x(t)$ is summarized as follows:

Step 1: The first step entails inserting the original signal $x(t)$ with repetition indicators that are equal to one (k , $j = 1$).

Step 2: All the local maximum and a local minimum of the original signal $x(t)$ are determined.

Step 3: By using the cubic spline curve, all local maximum and local minimum are connected separately to build an upper envelope $U(t)$ and a lower envelope $L(t)$, respectively, where all the original signal must be between these envelopes.

Step 4: The mean envelope value between the upper and lower envelopes is determined to build a new line curve, which represents the mean envelope by using the following equation:

$$M_j(t) = (U(t) + L(t)) / 2 \tag{1}$$

Step 5: Using the difference between the original signal $x(t)$ and the mean envelope value $M_j(t)$ as the new function $H_j(t)$:

$$H_j(t) = x(t) - M_j(t) \tag{2}$$

We verify whether the new function $H_j(t)$ satisfies the conditions of IMF. If $H_j(t)$ satisfies the conditions of IMF, then $H_j(t) = C_k(t)$, where $C_k(t)$ is the k-th IMF { $k = 1,2,...,K$} and we

continue to step 6. If not, then we replace $H_j(t)$ with $x(t)$ and repeat the operation from step 1 with repetition indicators $j = j + 1$ and $k = 1$.

Step 6: This step involves calculating the residual component $R_k(t)$ as in the following formula:

$$R_k(t) = R_{k-1}(t) - C_k(t); \ R_0(t) = x(t) \tag{3}$$

We check whether the residual component $R_k(t)$ is a monotonic function or satisfies the stoppage criterion of the standard deviation SD for two consecutive successive siftings of the results as follows:

$$SD_j = \sum_{t=0}^{T} \frac{(H_{j-1}(t) - H_j(t))^2}{H_{j-1}^2(t)}; \ 0.2 \leq SD_j \leq 0.3 \tag{4}$$

If not, then we replace $R_k(t)$ with $x(t)$ and repeat the operation from step 1 with repetition indicators $k = k + 1$. If yes, then we save all the IMF and residual components and end the sifting process.

The original signal $x(t)$ is the linear combination of the finite set of orthogonal IMF $\{C_k(t): k = 1,2, \dots, K\}$ and one monotonic residual $R(t)$ that are extracted via EMD, as in the following formula:

$$x(t) = \sum_{k=1}^{K} C_k(t) + R(t) \tag{5}$$

3.1.2 IMFs
The IMF's represent a new orthogonal design for the signals resulting from the division of the main original signal by using EMD. Each of IMF $\{C_k(t); k = 1,2, \dots, K\}$ components and residual $R(t)$ component has different and easy physical significant meanings. The IMF component function that satisfies two conditi (Huang et al., 1998; Huang, 2014) are as follows:

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

(2) At any point on a signal, the envelope mean value $M(t)$ between the upper envelope $U(t)$ defined by the local maximum and the lower envelope $L(t)$ defined by local minimum is equal to zero.

The first condition indicates that each IMF has only one local maximum or local minimum between two consecutive ZCs; meanwhile, the second condition explains that all the IMFs are stationary, which makes the analysis process highly flexible using these components (Raghuram et al., 2012).

3.2 ELNET Regression
In 2005, Zou and Hastie proposed ELNET regression. It is a hybrid technical penalized least square regression method that involves regularization and variable selection (Zou and Hastie, 2005). The

ELNET regression is a combination of two best techniques of shrinkage regression methods, namely, Ridge regression (L_2 penalty) for dealing with high-multicollinearity problems and the LASSO regression (L_1 penalty) for feature selection of regression coefficients (Wang et al., 2019).

The ELNET regression reduces the number of predictor variables by using the shrinkage of the coefficient regression toward zero or equal to zero for the less important variables by using the sum of the absolute values of the coefficient variables (L_1 norm) multiplied by the tuning parameter λ_1. The high correlation between the predictor variables is treated using the sum of the squared coefficient variables (L_2 norm) multiplied by the tuning parameter λ_2. This principle of the ELNET regression contributes to the production of an interpretable fitting model by limiting unnecessary variables that do not exist in the final model to increase the prediction accuracy. The ELNET deals with multicollinearity by keeping the predictor variables with high correlation into or out of the fitted model (Liu and Li, 2017).

The model structure of the multiple linear regression, which builds the relationship between the response variable and the predictor variables, is derived as follows:

$$y_i = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + ... + x_{ip}\beta_p + \varepsilon_i \tag{6}$$

$i = 1,2, ... , n$ and $j = 1,2, ... , p$, where y_i is the i-th response variable, β_0 is the intercept, x_{ij} is the j-th predictor variable of the i-th observation, β_j is the regression coefficient of the j-th predictor variable, which represents the average effect on y_i of per one unit change in the j-th predictor variable x_{ij}, and ε_i is the random error. For simplicity, we assume that the predictors and response variables are standardized by subtracting from the means and dividing by the standard deviations to obtain zero mean and unit variance (Melkumova and Shatskikh, 2017).

The traditional ordinary least squares (OLS) regression method is used to estimate the unknown regression coefficients by minimizing the residual sum of squared (*RSS*). Therefore, the *RSS* can be computed as follows:

$$RSS = \sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \tag{7}$$

Using Equation (7), the OLS regression coefficient estimation for the j-th element β (i.e., $\beta_j : j = 0,1,2, ... , p$) is calculated by minimizing the *RSS* form (Montgomery et al., 2012) as in the following formula:

$$\hat{\beta}^{OLS} = \arg\min_{\beta} \sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \tag{8}$$

The ELNET regression method is a penalized OLS estimator by adding the penalty terms (i.e., L_1 and L_2 penalties) to estimate the regression coefficients $\hat{\beta}$ (Zou and Hastie, 2005) as follows:

$$\hat{\beta}^{ELNET} = \arg\min_{\beta} \left[\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p}x_{ij}\beta_j)^2 + \lambda_1\sum_{j=1}^{p}|\beta_j| + \lambda_2\sum_{j=1}^{p}(\beta_j)^2 \right] \tag{9}$$

where, λ_1 and λ_2 are the tuning parameters and positive numeric values ($\lambda_1, \lambda_2 > 0$), respectively; they are automatically selected using cross-validation (CV). The tuning parameters control the strength of the regularization and selection of the predictor variables (Zou and Hastie, 2005). By denoting $\lambda_1 = 2n\lambda\alpha$ and $\lambda_2 = n\lambda(1-\alpha)$ (Haws et al., 2015), Equation (9) becomes equivalent to the following:

$$\hat{\beta}^{ELNET} = \arg\min_{\beta} \left[\frac{1}{2n}\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p}x_{ij}\beta_j)^2 + \lambda\left(\alpha\sum_{j=1}^{p}|\beta_j| + \frac{1-\alpha}{2}\sum_{j=1}^{p}(\beta_j)^2\right) \right] \tag{10}$$

where, α is a regularization parameter ($0 \le \alpha \le 1$), and λ is a tuning parameter ($\lambda > 0$).

The ELNET estimation represents the form of Ridge regression when $\alpha = 0$, while the ELNET estimation represents the form of LASSO regression when $\alpha = 1$. Therefore, the ELNET regression is setting the appropriate value for alpha between zero and one. The two penalty functions in ELNET regression are combined, where the L_2 penalty is used to stabilize the L_1 penalty regularization, while the L_1 penalty is used to generate a sparse model (Liu and Li, 2017). Therefore, Ridge and LASSO regressions are special cases of ELNET regression (Haws et al., 2015). The difference between the two penalties is the principle of the shrinkage of the regression coefficient and dealing with the correlation between the predictor variables.

3.3 Proposed ELNET-EMD Method

The ELNET-EMD is designed to explain the significance of the decomposition components of multivariate time-series predictors $x_j; j = 1,2,3,\dots p$ on the response variable y by selecting the most necessary components and tackle the multicollinearity using the following process:

(1) Decompose the original multi-predictors $x_j(t)$ separately into a finite set of orthogonal $C_{j,k}(t)$ components and one residual $R_j(t)$ component via EMD, where the original time-series predictor $x_j(t)$ is the summation of all decomposition components of the j-th predictor:

$$x_j(t) = \sum_{k=1}^{K}C_{j,k}(t) + R_j(t) \tag{11}$$

(2) Using Equation (6), all the decomposition components of the multi time-series predictors via EMD will be used as new predictor variables to explain the response variable $y(t)$:

$$y(t) = \sum_{j=1}^{p}(\sum_{k=1}^{K}C_{j,k}(t)\beta_{j,k}) + R_j(t)\beta_{j,K+1}) + \varepsilon(t) \tag{12}$$

(3) The ELNET regression will be used as in the following formula:

$$\hat{\boldsymbol{\beta}}^{ELNET} = \arg\min_{\beta} \left[\frac{1}{2n} \sum_{i=1}^{n} (y(t) - \sum_{j=1}^{p} \left(\sum_{k=1}^{K} C_{j,k}(t)\beta_{j,k} \right) - R_j(t)\beta_{j,K+1})^2 \right] + \lambda P_\alpha(\boldsymbol{\beta});$$

$$P_\alpha(\boldsymbol{\beta}) = \alpha \sum_{j=1}^{p} \sum_{k=1}^{K+1} |\beta_{j,k}| + \frac{1-\alpha}{2} \sum_{j=1}^{p} \sum_{k=1}^{K+1} (\beta_{j,k})^2 \tag{13}$$

The numerical experiments and real time-series data will be used to compare between the ELNET-EMD method against seven methods, namely, OLS-EMD, SR-EMD, Ridge-EMD, LASSO-EMD, The smoothly clipped absolute deviation (SCAD) (Fan and Li, 2001) based on EMD (SCAD-EMD), The minimax concave penalty (MCP) (Zhang, 2010) based on EMD (MCP-EMD), and Ad-LASSO (Zou, 2006) based on EMD (Ad-LASSO-EMD) methods.

To check the correlation among the decomposition components, the variance inflation factor (VIF) test will be used to assess the correlation value between the decomposition components. The decomposition components will be free from multicollinearity when the value of VIF is less than 10 (Jadhav et al., 2014) . The VIF form is presented as follows:

$$VIF_{j,k} = \frac{1}{1 - R_{j,k}^2} \tag{14}$$

Where, $R_{j,k}^2$ is the coefficient of determination of the decomposition component j, k on the remaining components in the model.

In this article, three criteria tests are used, that is, (a) root mean square error ($RMSE$), (b) mean absolute error (MAE), and (c) mean absolute percentage error ($MAPE$).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad ; \quad MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \quad ; \quad MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Where, y_i is the actual value, $\hat{y}_i$ is the estimated value, and n is the sample size.

4. Simulation Studies
We comprehensively explain the numerical experiment, which is simulated by the sine and cosine functions used to apply the methods, and the numerical results and discussion.

4.1 Numerical Experiment
The sine and cosine wave functions are used to generate simulated signals to demonstrate the ability of the ELNET-EMD method. These signals are numerical experiments that are employed in the application.

The numerical experiments are explained in two cases. In the first case, the predictor variables $x_1(t), x_2(t)$, and $x_3(t)$ are fixed variables without white noise error. In the second case, the predictor variables have white noise error with normal distribution of zero mean and unit variance (i.e. $x_j(t) = x_j(t) + \varepsilon$; $\varepsilon \sim iid$ N(0,1), $j = 1,2,3$). The response variable is simulated according to one or two components from the predictor variables. This notion is in line with the simulation study proposed by (Qin et al., 2016; Al-Jawarneh et al., 2020) for generating variables.

The datasets are simulated by predictors and response variables with the length of a sample size $n = 110$ and time sequence between zero and nine ($0 \leq t \leq 9$). For the second case, 2,000 replications of a sample size length $n = 110$ are made. The datasets are divided into two parts of 70% for training the model and 30% for testing and evaluating the performance criteria. The formula of the function test of the response and the predictor variables is presented as follows:

$$y(t) = \sin(\pi t) - \sin(2\pi t) - \cos(6\pi t) - \cos(13\pi t)$$

$$x_1(t) = \sin(2\pi t) + \cos(\pi t) + \sin(5\pi t) + \sin(9\pi t)$$

$$x_2(t) = 0.2t + \sin(\pi t) + \cos(6\pi t) + \cos(9\pi t)$$

$$x_3(t) = \sin(\pi t) + \sin(8\pi t) + \cos(7\pi t) + \cos(13\pi t)$$
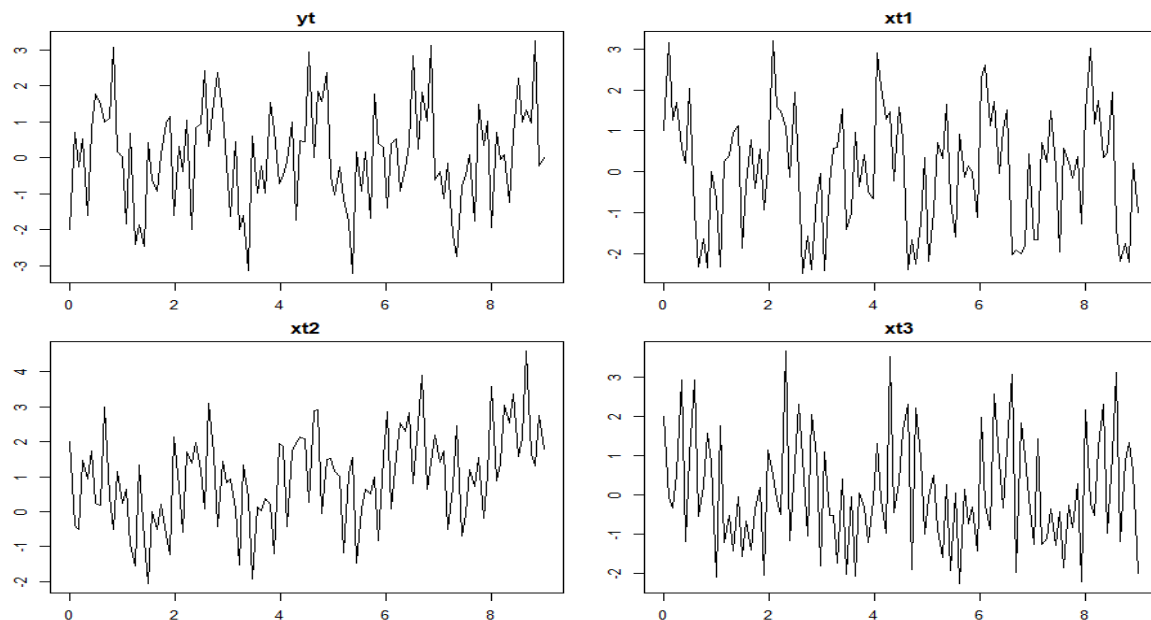
4.2 Numerical Results and Discussion

Figure 1. Plots of the original signals.

Figure 1 illustrates the shape of the original signals, where they neither show constant value over time. This property indicates that the predictor variables are nonstationary.
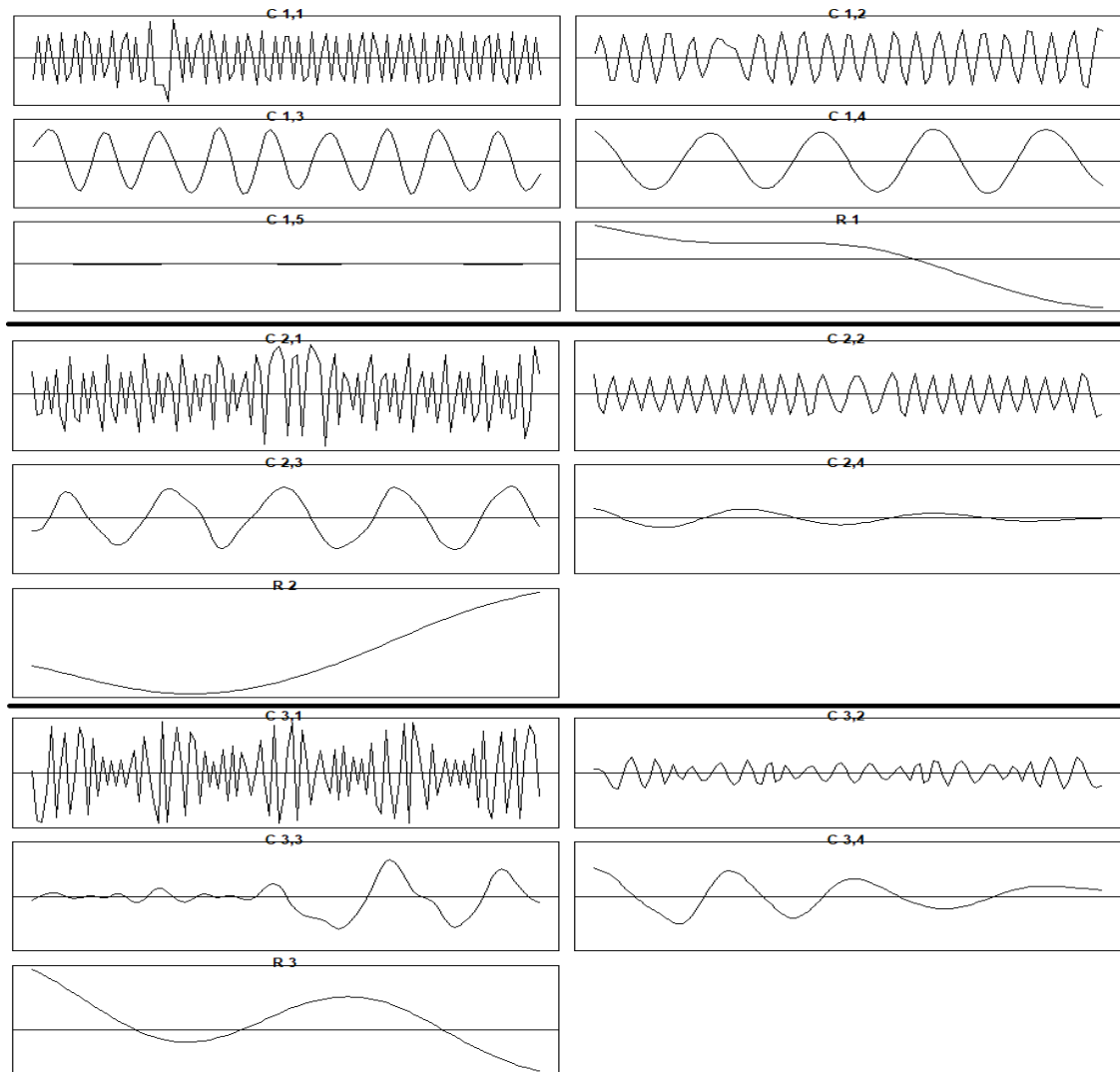
Figure 2. Decomposition of the main predictor signals $x_1(t)$ $x_2(t)$, and $x_3(t)$ via EMD.

Figure 2 shows the decomposition components via EMD method of the original multivariate predictors $x_1(t)$, $x_2(t)$ and $x_3(t)$. The first predictor $x_1(t)$ decomposed into five IMFs and one residual component. The second $x_2(t)$ and third $x_3(t)$ predictors decomposed into four IMFs and one residual component, separately. Each IMF component displays differences in physical meaning (i.e., wavelength, and frequency) compared with other IMF components; for example, the first IMF component, which is $\{C_{1,1}(t), C_{2,1}(t), C_{3.1}(t)\}$ has a high frequency and a short wavelength. This characteristic varies among the IMF components in which the frequency decrease, while the wavelength increases. The last component represents the trend of the original signal which is called the residual. Thus, each decomposition component has information concerning a particular frequency that is founded within the original signal. This finding contributes to the use of these components as new predictor variables.

Table 1 shows the values of the *VIF* ($VIF_{j,k}$) test of multicollinearity between the decomposition components for the original multivariate predictors via EMD. Based on the output of $VIF_{j,k}$, several decomposition components obtain values larger than 10 (i.e., $VIF_{1,6}, VIF_{2,3}, VIF_{2,5}$, and $VIF_{3,3} > 10$). This finding indicates that high multicollinearity (correlation) exists between the decomposition components.

Table 1. Variance inflation factors (*VIF*) (case 1).

$VIF_{1,1}$	$VIF_{1,2}$	$VIF_{1,3}$	$VIF_{1,4}$	$VIF_{1,5}$	$VIF_{1,6}$	$VIF_{2,1}$	$VIF_{2,2}$
1.025	1.019	2.943	1.696	2.618	49.40	1.071	1.037
$VIF_{2,3}$	$VIF_{2,4}$	$VIF_{2,5}$	$VIF_{3,1}$	$VIF_{3,2}$	$VIF_{3,3}$	$VIF_{3,4}$	$VIF_{3,5}$
13.68	3.33	41.36	1.115	1.071	14.39	9.676	5.852

*j, k is the *k*-th component of the original predictor *j*-th; $j = 1,2,3$

Figure 3 shows the CV and coefficient estimation plots of the ELNET-EMD. The first plot on the left represents CV at $K = 10$, where the y-axis represents the mean square error (*MSE*), and the x-axis represents the value of $\log(\lambda)$. The upper horizontal line represents the numbers of non-zero coefficients selected at the $\log(\lambda)$ value. The first vertical dotted line from the left is the location of the point selected at a minimum of the *MSE* (*minM*) rule, while the second vertical line denotes the location of the point selected at a minimum of *MSE* with the one-standard-error (*1se*) rule. Thus, the increase in λ leads to a decrease in the number of non-zero coefficients entering into the final model according to the CV plot. Therefore, the selection of the λ value at the *minM* or *1se* rule is based on the optimal minimum *MSE* value. The second plot represents the coefficient estimation and explains the order in which the number of non-zero regression coefficients of the decomposition components will be entered into the final model. It provides a sign on the response variable at current λ, which is the actual degree of freedom. Therefore, the number of non-zero coefficient estimation differs in relation to the selected value of λ at the *minM* or *1se* rule. For example, 14 components are identified under the *minM* rule, while seven components are observed under the *1se* rule.

Table 2 illustrates the results of the *RSS* values in the first case of the simulation for the regression methods at the values of lambda as selected by CV. The results show that the smallest *RSS* value is for the ELNET-EMD model at the *1se* rule ($\lambda = 0.25958$; $RSS = 16.21$). The *RSS* value provides the best method for selecting important variables and for supporting the fitting regression models using the ELNET-EMD model.
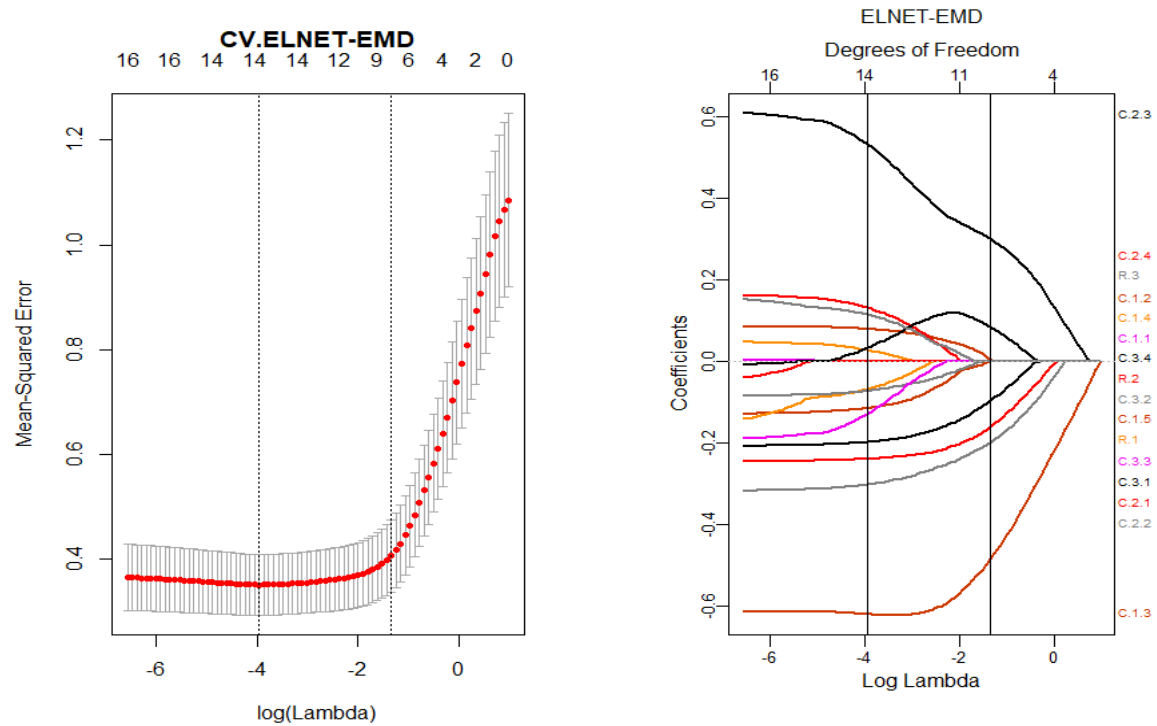
Figure 3. 10- Cross-Validation (10-CV) estimation and the coefficient estimation for ELNET-EMD.

Table 2. Residual sum of squares error (RSS) values (Case 1).

Method	Rule	λ	RSS
OLS-EMD			21.51 [12]
SR-EMD			19.32 [7]
SCAD-EMD		0.044985	18.08 [5]
MCP-EMD		0.048236	18.86 [6]
Ridge-EMD	$minM$	0.063766	19.57 [8]
	$1se$	0.340295	16.55 [2]
Ad-LASSO-EMD	$minM$	0.086895	19.58 [9]
	$1se$	1.705773	17.06 [4]
LASSO-EMD	$minM$	0.006681	20.18 [11]
	$1se$	0.099198	17.03 [3]
ELNET-EMD α = 0.24	$minM$	0.019185	20.11 [10]
	$1se$	0.259580	16.21 [1]

* 1-12 is the rank of the methods from minimum to maximum values

Table 3 explains the estimation of the coefficients of the decomposition components for each of the regression methods used in this study. Most regression methods can reduce the number of decomposition components, except for Ridge-EMD; these methods have the same numbers of coefficient, that is, all the decomposition components are entered into the final model. The numbers of nonzero coefficients of ELNET-EMD and other methods in the study are different from those of Ridge-EMD method. For the ELNET-EMD at $1se$ rule, the coefficient estimation equal to seven nonzero coefficients $C_{1,2}$, $C_{1,3}$, $C_{2,1}$, $C_{2,2}$, $C_{2,3}$, $C_{3,1}$, and $C_{3,4}$ components enter the ELNET-EMD regression model with varying levels of significance. That is, the $C_{1,3}$ and $C_{2,3}$ components are more significant than other selected components on the response variable.

Table 3. Coefficients estimation for the decomposition components.

	OLS-EMD	SR-EMD	SCAD-EMD	MCP-EMD	Ridge-EMD 1se	Ad-LASSO-EMD 1se	LASSO-EMD 1se	ELNET-EMD 1se
$B_{1,1}$	0.002	0	0	0	- 0.006	0	0	0
$B_{1,2}$	0.085	0.100	0.045	0.060	0.050	0	0	0.005
$B_{1,3}$	- 0.617	- 0.583	- 0.577	- 0.680	- 0.496	- 0.531	- 0.574	- 0.484
$B_{1,4}$	0.047	0	0.007	0	- 0.003	0	0	0
$B_{1,5}$	- 0.130	0	0	- 0.006	- 0.074	0	0	0
$B_{1,6}$	- 0.176	0	0	0	- 0.014	0	0	0
$B_{2,1}$	- 0.248	- 0.262	- 0.259	- 0.259	- 0.185	- 0.158	- 0.164	- 0.161
$B_{2,2}$	- 0.319	- 0.292	- 0.299	- 0.291	- 0.226	- 0.202	- 0.236	- 0.200
$B_{2,3}$	0.609	0.609	0.614	0.388	0.324	0.320	0.391	0.300
$B_{2,4}$	0.163	0	0.023	0	0.049	0	0	0
$B_{2,5}$	- 0.069	0	0	0	- 0.001	0	0	0
$B_{3,1}$	- 0.208	- 0.213	0.214	- 0.202	- 0.133	- 0.091	- 0.088	- 0.096
$B_{3,2}$	- 0.086	0	- 0.029	- 0.026	- 0.040	0	0	0
$B_{3,3}$	- 0.188	- 0.222	- 0.229	0	- 0.010	0	0	0
$B_{3,4}$	- 0.006	0	0	0.164	0.106	0.062	0	0.084
$B_{3,5}$	0.163	0	0.001	0.039	0.040	0	0	0

Table 4 illustrates the performance criteria of the prediction accuracy of the regression methods using *RMSE*, *MAE*, and *MAPE* in the first case of the numerical experiment. The results show that the ELNET-EMD has the smallest error values in terms of *RMSE*, *MAE,* and *MAPE*. Therefore, the ELNET-EMD method is highly reliable for the selected components with high prediction accuracy.

Table 4. Performance criteria (Case1).

Method	*RMSE*	*MAE*	*MAPE*
OLS-EMD	0.7953 [8]	0.6680 [8]	1.4188 [8]
SR-EMD	0.7539 [7]	0.6550 [7]	1.3498 [7]
SCAD-EMD	0.7292 [5]	0.6414 [5]	1.3221 [5]
MCP-EMD	0.7447 [6]	0.6475 [6]	1.3473 [6]
Ridge-EMD 1se	0.6978 [2]	0.6120 [2]	1.1732 [2]
Ad-LASSO-EMD 1se	0.7084 [4]	0.6227 [3]	1.2124 [4]
LASSO-EMD 1se	0.7077 [3]	0.6248 [4]	1.1884 [3]
ELNET-EMD 1se	0.6906 [1]	0.6063 [1]	1.1305 [1]

* 1-8 is the rank of the methods from minimum to maximum values

Moreover, in the second case of the numerical experiment, the white noise error $\varepsilon \sim iid$ N(0,1) is added to the main predictor variables. After performing the same process in the first case, the results show that components are similar to those in the first case for the selected predictor variables.

Table 5. Performance criteria (case 2).

Method	*RSS*	*RMSE*	*MAE*	*MAPE*
OLS-EMD	24.47 [8]	0.8512 [8]	0.6874 [8]	2.7763 [8]
SR-EMD	24.31 [7]	0.8481 [7]	0.6854 [7]	2.7193 [7]
SCAD-EMD	23.57 [5]	0.8358 [5]	0.6764 [5]	2.5216[5]
MCP-EMD	23.63 [6]	0.8368 [6]	0.6780 [6]	2.5299 [6]
Ridge-EMD	22.34 [4]	0.8150 [4]	0.6567 [4]	2.0626 [2]
Ad-LASSO-EMD	21.79 [2]	0.8048 [2]	0.6514 [2]	2.1924 [4]
LASSO-EMD	22.04 [3]	0.8093 [3]	0.6546 [3]	2.0705 [3]
ELNET-EMD	20.97 [1]	0.7894 [1]	0.6373 [1]	2.0573 [1]

* 1-8 is the rank of the methods from minimum to maximum values

Table 5 explains the mean performance criteria values (i.e., $RSS, RMSE, MAE$, and $MAPE$) for the ELNET-EMD method in the second case with the other regression methods. The results show that the ELNET-EMD has the smallest error value in these criteria tests. Therefore, the ELNET-EMD improves the prediction accuracy by producing the smallest error values in terms of RSS, $RMSE, MAE$, and $MAPE$.

Based on the numerical experiment results, the ELNET-EMD functions by selecting the necessary variables that have a significant effect on the response variable and improves the production accuracy despite the existence of white noise.

5. Application
In this section, we provide an empirical data analysis using the daily exchange rates of four countries against the US dollar (USD).

5.1 Exchange Rate
In this study, the proposed method contributes to the identification of the decomposed components via the EMD method of the multivariate original predictor variables, which exert a reflective effect on the response variable in case of multicollinearity among the decomposition components for enhancing the production accuracy. The daily close exchange rates of four countries against the USD, namely, Taiwan (TAW/USD), Malaysia (MYR/USD), Japan (JAP/USD), and China (CHN/USD), from 27/03/2015 to 25/10/2019, is used to evaluate the performance of the ELNET-EMD method relative to other technical selecting regression methods depending on the real time series dataset. The number of observations is 1,196, and all datasets are collected from the Wall Street Journal database (https://www.wsj.com/).

The response variable in this application is the daily close exchange rates of TAW/USD (Mean = 31.0993934, Standard Error = 0.02953328, Skewness = 0.33513459), and the three predictor variables are the daily close exchange rates of MYR/USD (Mean = 4.128473077, Standard Error = 0.005606923, Skewness = -0.555859832), JAP/USD (Mean = 112.0747492, Standard Error = 0.158407035, Skewness = 0.457279359), and CHN/USD (Mean = 6.648520569, Standard Error = 0.007161266, Skewness = -0.184059875). The datasets are divided into two parts, where the first part from 27/03/2015 to 11/06/2018 of the data sets (70% of the total dataset) are used for training the model, and the remaining datasets are used to evaluate the performance criteria and called the test dataset.

5.2 Application Results and Discussion
Figure 4 presents the daily close exchange rates of the main predictor variables MYR/USD ($x_1(t)$), JAP/USD ($x_2(t)$), and CHN/USD ($x_3(t)$), and the response variable y(t) is TAW/USD. The figure illustrates the plots of original predictors and response signals. The structure of the signals shows different values over time and excludes straight lines, which indicates that the signals are nonstationary and nonlinear.
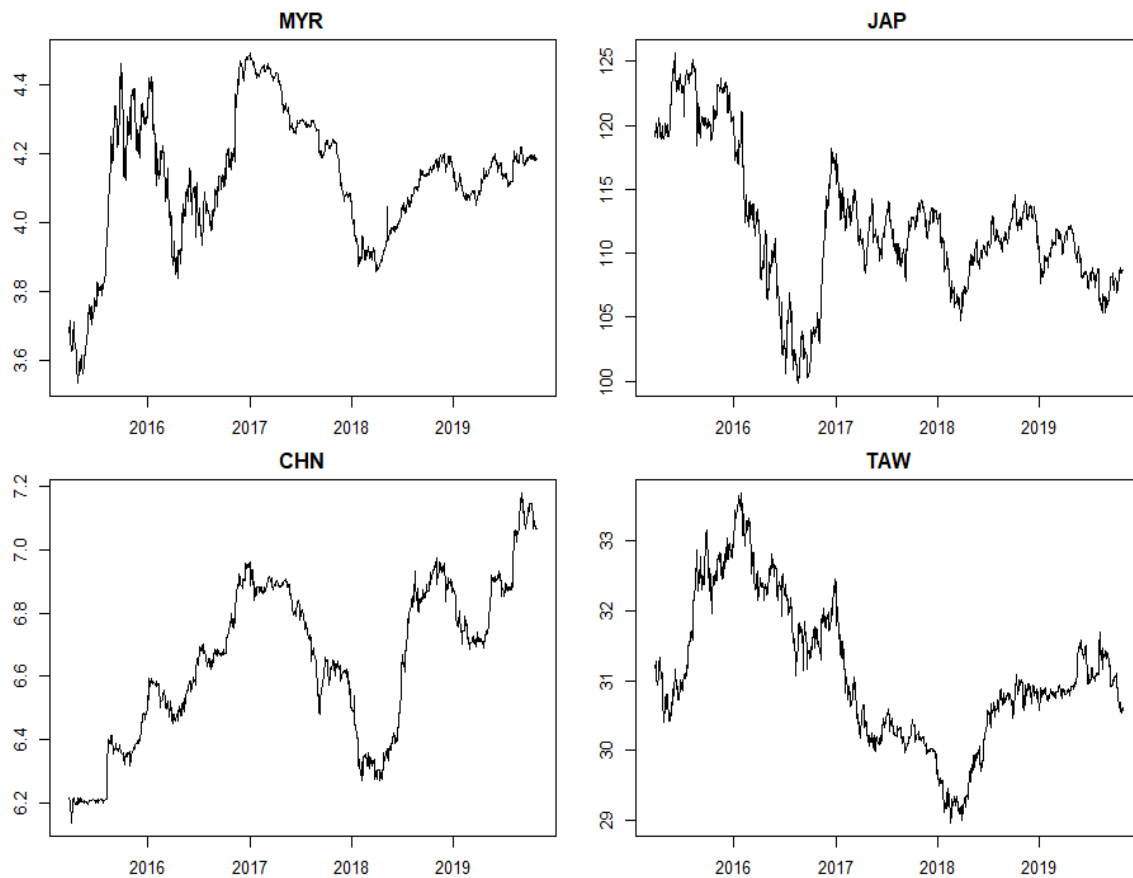
Figure 4. The daily exchange rates of MYR/USD, JAP/USD, CHN/USD, and TAW/USD.

Figure 5 shows the decomposing components using the EMD algorithm of the main predictor variables $x_1(t)$, $x_2(t)$, and $x_3(t)$. Each one of these variables is decomposed into seven IMFs and one residual component. Besides, the first three components have high frequency and short wavelength. These physical properties vary among the components, where the frequency decreases with the increase in the wavelength when the increase in the number of components.

Table 6 explains the values of the *VIF* test of multicollinearity between the decomposition components. The results show that several $VIF_{j,k}$ values are greater than ten ($VIF_{j,k} > 10$), such as $VIF_{1,7}, VIF_{2,7}, VIF_{2,8}$, and $VIF_{3,7}$. This finding indicates that a high correlation exists among the decomposition components of the MYR, JAP, and CHN variables, which subsequently indicates that multicollinearity exists.

Figure 6 shows the 10-CV and the coefficient estimation plots of the ELNET-EMD. The figures illustrate the number of decomposition components, which are selected in the final model by choosing the value of λ at the *minM* or 1se rule according to the CV plot. For the *minM* rule, twelve components from all decomposition components are selected into the final model, while for the 1*se* rule, only the $C_{1,1}$ and $C_{3,1}$ components have the strongest effect on the response variable and that is why it is entered into the final model.
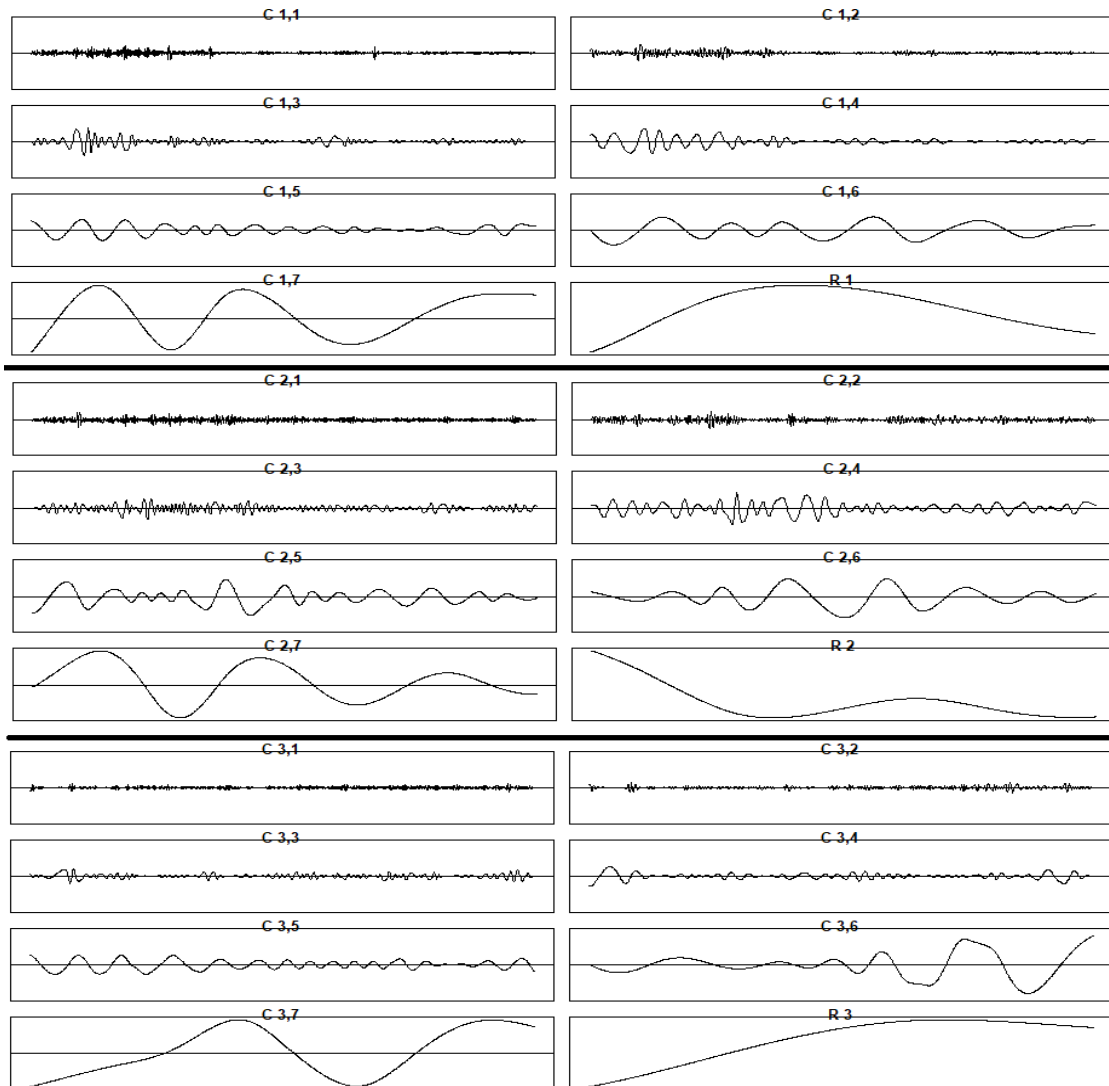
Figure 5. Decomposition of the main predictor variables $x_1(t)$, $x_2(t)$, and $x_3(t)$ via EMD.

Table 6. Variance inflation factors (*VIF*).

$VIF_{1,1}$	$VIF_{1,2}$	$VIF_{1,3}$	$VIF_{1,4}$	$VIF_{1,5}$	$VIF_{1,6}$	$VIF_{1,7}$	$VIF_{1,8}$	$VIF_{2,1}$	$VIF_{2,2}$	$VIF_{2,3}$	$VIF_{2,4}$
1.05	1.04	1.12	1.06	2.57	2.25	12.2	5.59	1.01	1.01	1.02	1.11
$VIF_{2,5}$	$VIF_{2,6}$	$VIF_{2,7}$	$VIF_{2,8}$	$VIF_{3,1}$	$VIF_{3,2}$	$VIF_{3,3}$	$VIF_{3,4}$	$VIF_{3,5}$	$VIF_{3,6}$	$VIF_{3,7}$	$VIF_{1,8}$
1.22	1.35	11.5	34.9	1.1	1.04	1.12	1.12	2.42	1.64	12.2	4.9

*j, k is the kth component of the original predictor j^{th}; $j = 1,2,3$
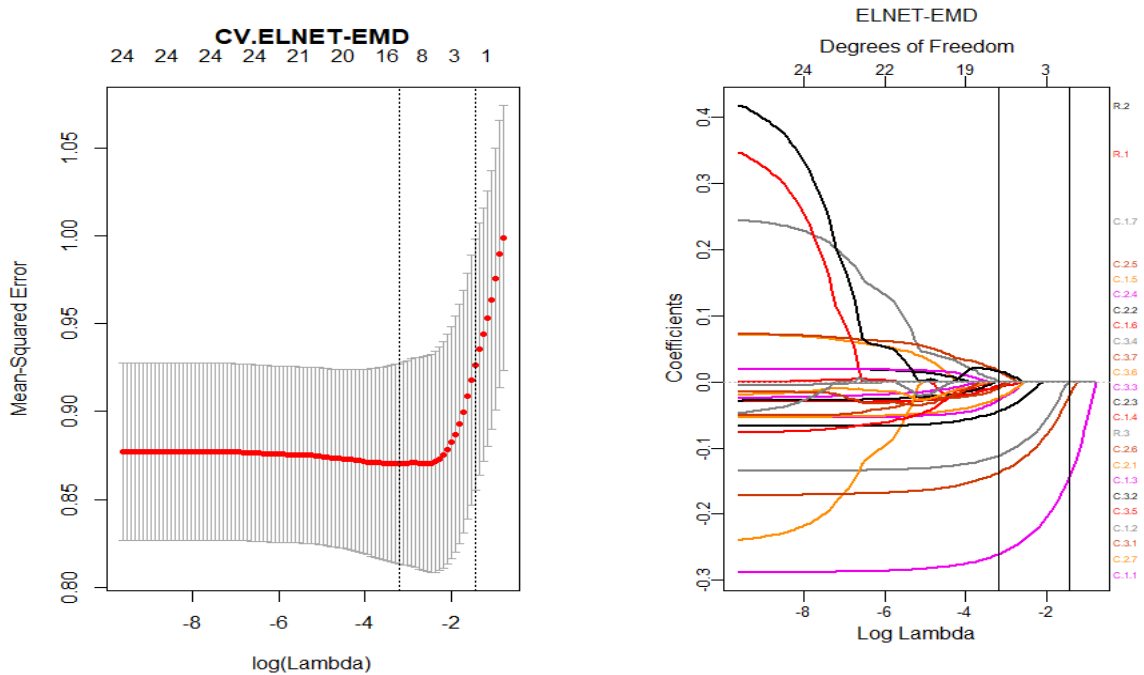
Figure 6. 10- Cross-validation (10-CV) estimation and the coefficient estimation for ELNET-EMD.

Table 7 illustrates the results of the RSS values for comparing the ELNET-EMD method with the other technical methods at the values of lambda that are selected by CV. The ELNET-EMD model has the smallest RSS value at $minM$ rule ($\lambda = 0.040645$ and $RSS = 299.444$). In this case, the ELNET-EMD model at $minM$ performs better than other methods in this study.

Table 7. Residual sum of squares error (RSS) values.

Method	Rule	λ	RSS
OLS-EMD			383.7779 [12]
SR-EMD			311.0713 [7]
SCAD-EMD		0.053576	303.3288 [4]
MCP-EMD		0.070824	305.5657 [6]
Ridge-EMD	$minM$	0.231917	301.5496 [2]
	$1se$	2.605181	324.7441 [8]
Ad-LASSO-EMD	$minM$	0.306543	302.3415 [3]
	$1se$	3.137560	339.5930 [11]
LASSO-EMD	$minM$	0.056127	305.3062 [5]
	$1se$	0.171404	336.0778 [9]
ELNET-EMD α = 0.72	$minM$	0.040645	299.4440 [1]
	$1se$	0.238060	336.5113 [10]

* 1-12 is the rank of the methods from minimum to maximum values

Table 8 explains the estimation of the coefficients of the decomposition components. For the ELNET-EMD at $minM$ rule compared with other regression methods used in the study, the coefficient estimation equal to twelve nonzero coefficients, where six components from Malaysia ($C_{1,1} \sim C_{1,4}$, $C_{1,7}$, and R_1), four components from Japan ($C_{2,1}, C_{2,5}, C_{2,6}$, and R_2), and two

components from China ($C_{3,1}$, and $C_{3,2}$) have varying levels of strong effects on the response variable. Thus, the ELNET-EMD method is more accurate in selecting non-zero coefficients than other regression methods when a high correlation exists among the decomposition components. Where the components that have high correlation like $C_{1,7}$ and R_2 components are selected into the final model.

Table 8. Coefficients estimation for the decomposition components.

	OLS-EMD	SR-EMD	SCAD-EMD	MCP-EMD	Ridge-EMD $minM$	Ad-LASSO-EMD $minM$	LASSO-EMD $minM$	ELNET-EMD $minM$
$B_{1,1}$	- 0.2884	- 0.2864	- 0.2937	- 0.2957	- 0.2375	- 0.2792	- 0.2432	- 0.2607
$B_{1,2}$	- 0.1353	- 0.1367	- 0.1124	- 0.1149	- 0.1139	- 0.1160	- 0.0916	- 0.1119
$B_{1,3}$	- 0.0544	- 0.0599	- 0.0047	0	- 0.0443	0	- 0.0005	- 0.0283
$B_{1,4}$	- 0.0287	0	0	0	- 0.0277	0	0	- 0.0055
$B_{1,5}$	0.0733	0	0	0	0.0226	0	0	0
$B_{1,6}$	- 0.0028	0	0	0	0.0032	0	0	0
$B_{1,7}$	0.2559	- 0.1718	0	0	0.0377	0	0	0.0010
$B_{1,8}$	0.4306	0	0	0	- 0.0105	0	0	- 0.0096
$B_{2,1}$	- 0.0538	- 0.0052	- 0.0015	0	- 0.0436	0	- 0.0003	- 0.0245
$B_{2,2}$	0.0181	0	0	0	0.0124	0	0	0
$B_{2,3}$	- 0.0288	0	0	0	- 0.0209	0	0	0
$B_{2,4}$	0.0195	0	0	0	0.0189	0	0	0
$B_{2,5}$	0.0737	0.0477	0	0	0.0439	0	0	0.0173
$B_{2,6}$	- 0.0491	- 0.0552	0	0	- 0.0273	0	0	- 0.0094
$B_{2,7}$	- 0.2544	- 0.1566	0	0	- 0.0050	0	0	0
$B_{2,8}$	0.4909	0.1029	0.0011	0	0.0128	0	0	0.0151
$B_{3,1}$	- 0.1718	- 0.1652	- 0.1399	- 0.1345	- 0.1440	- 0.1411	- 0.1159	- 0.1375
$B_{3,2}$	- 0.0663	- 0.0663	- 0.0189	- 0.0013	- 0.0613	- 0.0189	- 0.0229	- 0.0448
$B_{3,3}$	- 0.0247	0	0	0	- 0.0158	0	0	0
$B_{3,4}$	- 0.0066	0	0	0	- 0.0012	0	0	0
$B_{3,5}$	- 0.0782	0	0	0	- 0.0326	0	0	0
$B_{3,6}$	- 0.0237	0	0	0	- 0.0226	0	0	0
$B_{3,7}$	- 0.0146	0	0	0	- 0.0202	0	0	0
$B_{3,8}$	- 0.0659	0	0	0	- 0.0120	0	0	0

Table 9 illustrates the performance criteria of the prediction accuracy of the regression methods using $RMSE$, MAE, and $MAPE$ in the application using a real time-series dataset. The results show that the ELNET-EMD at the $minM$ rule has the smallest error value in the terms of $RMSE$ and MAE terms, while for the $MAPE$ achieved the fourth-order among the other methods. Therefore, the ELNET-EMD method selects decomposition components that have more significant effect on the response variables with high prediction accuracy.

Table 9. Performance criteria.

Method	$RMSE$	MAE	$MAPE$
OLS-EMD	1.0339 [8]	0.7942 [8]	4.2962 [8]
SR-EMD	0.9309 [7]	0.6906 [7]	2.5998 [3]
SCAD-EMD	0.9192 [4]	0.6771 [5]	2.7485 [7]
MCP-EMD	0.9226 [6]	0.6796 [6]	2.7453 [6]
Ridge-EMD $minM$	0.9165 [2]	0.6759 [4]	2.5771 [2]
Ad-LASSO-EMD $minM$	0.9177 [3]	0.6751 [2]	2.6828 [5]
LASSO-EMD $minM$	0.9222 [5]	0.6754 [3]	2.3970 [1]
ELNET-EMD $minM$	0.9133 [1]	0.6708 [1]	2.6308 [4]

* 1-8 is the rank of the methods from minimum to maximum values

6. Conclusions

This study applied the ELNET-EMD method by using nonstationary and nonlinear time-series data. The method is used to study the effect of the IMFs and the residual component of multivariate predictor variables on the response variable and tackle the high correlation among the IMFs and residual component for ensuring accuracy and reliability of the selected fitting model.

Numerical experiments and actual time-series dataset for the daily close exchange rates of MYR, JAP, CHN, and TAW are carried out. The ELNET-EMD method separately decomposes the multi-predictors into a finite set of IMFs and one residual component via EMD. Thereafter, the decomposition components are selected, which have the most significance on the response variable and address the multicollinearity between the decomposition components using the ELNET regression.

The results prove that the ELNET-EMD method is considerably more accurate than other regression methods. The ELNET-EMD method is highly capable of identifying the decomposition components that have the most significance on the response variable. Although the correlation between the decomposition components is high, the components have high correlation in/out the final model. The ELNET-EMD method selects the best fitting model that is free of multicollinearity and displays high prediction accuracy.

References

Al-Jawarneh, A.S., Ismail, M.T., Awajan, A.M., & Alsayed, A.R. (2020). Improving accuracy models using elastic net regression approach based on empirical mode decomposition. *Communications in Statistics-Simulation and Computation*, 1-20.

Chen, C.F., Lai, M.C., & Yeh, C.C. (2012). Forecasting tourism demand based on empirical mode decomposition and neural network. *Knowledge-Based Systems, 26*, 281-287.

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

Chan, Y.T. (1994). *Wavelet basics*. Springer Science & Business Media.

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association, 96*(456), 1348-1360.

Haws, D.C., Rish, I., Teyssedre, S., He, D., Lozano, A.C., Kambadur, P., Parida, L. (2015). Variable-selection emerges on top in empirical comparison of whole-genome complex-trait prediction methods. *PloS one, 10*(10), e0138903.

Hoerl, A.E., & Kennard, R.W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*(1), 55-67.

Hong, W.C., & Fan, G.F. (2019). Hybrid empirical mode decomposition with support vector regression model for short term load forecasting. *Energies, 12*(6), 1093.

Huang, N.E. (2014). Introduction to the Hilbert–Huang transform and its related mathematical problems. In *Hilbert–Huang transform and its applications* (pp. 1-26): World Scientific.

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., 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.

Jadhav, N.H., Kashid, D.N., & Kulkarni, S.R. (2014). Subset selection in multiple linear regression in the presence of outlier and multicollinearity. *Statistical Methodology, 19*, 44-59.

Liu, W., & Li, Q. (2017). An efficient elastic net with regression coefficients method for variable selection of spectrum data. *PloS one, 12*(2), e0171122.

Masselot, P., Chebana, F., Bélanger, D., St-Hilaire, A., Abdous, B., Gosselin, P., & Ouarda, T.B. (2018). EMD-regression for modelling multi-scale relationships, and application to weather-related cardiovascular mortality. *Science of The Total Environment, 612*, 1018-1029.

Melkumova, L., & Shatskikh, S.Y. (2017). Comparing ridge and LASSO estimators for data analysis. *Procedia Engineering, 201*, 746-755.

Montgomery, D.C., Peck, E.A., & Vining, G.G. (2012). *Introduction to linear regression analysis* (Vol. 821): John Wiley & Sons, New Jersey.

Moore, K.J., Kurt, M., Eriten, M., McFarland, D.M., Bergman, L.A., & Vakakis, A.F. (2018). Wavelet-bounded empirical mode decomposition for measured time series analysis. *Mechanical Systems and Signal Processing, 99*, 14-29.

Naik, J., Satapathy, P., & Dash, P. (2018). Short-term wind speed and wind power prediction using hybrid empirical mode decomposition and kernel ridge regression. *Applied Soft Computing, 70*(1), 1167-1188.

Plakandaras, V., Gupta, R., Gogas, P., & Papadimitriou, T. (2015). Forecasting the US real house price index. *Economic Modelling, 45*, 259-267.

Qin, L., Ma, S., Lin, J.C., & Shia, B.C. (2016). Lasso regression based on empirical mode decomposition. *Communications in Statistics-Simulation and Computation, 45*(4), 1281-1294.

Raghuram, M., Madhav, K.V., Krishna, E.H., Komalla, N.R., Sivani, K., & Reddy, K.A. (2012, May). HHT based signal decomposition for reduction of motion artifacts in photoplethysmographic signals. In *2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings* (pp. 1730-1734). IEEE. Graz, Austria.

Shen, Z., & Lee, C.H. (2012, March). A lasso based ensemble empirical mode decomposition approach to designing adaptive clutter suppression filters. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 757-760). IEEE. Kyoto, Japan.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(1), 267-288.

Titchmarsh, E.C. (1948). *Introduction to the theory of Fourier integrals* (Vol. 950): Clarendon Press Oxford, London.

Wang, Y., Kong, L., Jiang, B., Zhou, X., Yu, S., Zhang, L., & Heo, G. (2019). Wavelet-based LASSO in functional linear quantile regression. *Journal of Statistical Computation and Simulation, 89*(6), 1111-1130.

Yang, A.C., Fuh, J.L., Huang, N.E., Shia, B.C., Peng, C.K., & Wang, S.J. (2011). Temporal associations between weather and headache: analysis by empirical mode decomposition. *PloS one, 6*(1), e14612.

Zhang, C.H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of statistics, 38*(2), 894-942.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical Association, 101*(476), 1418-1429.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67*(2), 301-320.