Elastic-Net Regression based on Empirical Mode Decomposition for Multivariate Predictors

Abdullah Suleiman Al-Jawarneh* and Mohd. Tahir Ismail

School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia

ABSTRACT

The empirical mode decomposition (EMD) method is used to decompose the non-stationary and nonlinear signal into a finite set of orthogonal non-overlapping time scale components that include several intrinsic mode function components and one residual component. Elastic net (ELN) regression is a statistical penalized method used to address multicollinearity among predictor variables and identify the necessary variables that have the most effect on the response variable. This study proposed the use of the ELN method based on the EMD algorithm to identify the decomposition components of multivariate predictor variables with the most effect on the response variable under multicollinearity problems. The results of the numerical experiments and real data confirmed that the EMD-ELN method is highly capable of identifying the decomposition components with the presence or absence of multicollinearity among the components. The proposed method also achieved the best estimation and reached the optimal balance between the variance and bias. The EMD-ELN method also improved the accuracy of regression modeling compared with the traditional regression models.

Keywords: Elastic-net regression, empirical mode decomposition, LASSO, model selection, multicollinearity, ridge regression

INTRODUCTION

Regression analysis methods assume that the variables have stationary and linear properties and that the predictor variables are free from multicollinearity to achieve reliability and accuracy of results. Transformation (modification) methods can be used to convert the variables to become stationary and linearized; however, these

methods can lead to the loss of valuable information and features of the original dataset. Moreover, few statistical methods can handle variables selection when multicollinearity exists (Hamid et al., 2018; Hashibah & Mahat, 2013).

The empirical mode decomposition (EMD) method focuses on non-stationary and nonlinear variables and decomposes the variable into a set of decomposition components with different information (Huang, 2014). These components represent new predictor variables that can be used to study their effects on the response variable. Unlike traditional analysis methods, such as Fourier decomposition (Titchmarsh, 1948) and wavelet decomposition (Chui, 1995), the EMD method does not assume that the dataset is either stationary or linear.

Multicollinearity is one of the fundamental issues in variables selection. The presence of multicollinearity between predictor variables increases variance, highlights the wrong sign of coefficients, and misleads the selected model (Jadhav et al., 2014). To address this gap, Zou and Hastie (2005) proposed the technical penalized regularization method, which is a double shrinkage called elastic net regression (ELN). The ELN method is a combination of ridge regression (RR) (Hoerl & Kennard, 1970) and least absolute shrinkage and selection operator (LASSO) regression (Tibshirani, 1996). The ELN can achieve the best estimation of optimal balance between the variance and bias terms. It can also choose between predictors that exhibit a high correlation in the final model.

The EMD method and the penalized regularization regression method analysis have been used in several scientific fields to understand the significance of decomposition components on the response variable. Examples include the forward stepwise regression methods with EMD (Yang et al., 2011), the LASSO regression based on ensemble EMD (EEMD) (Shen & Lee, 2012), ridge regression with EEMD (Shen et al., 2012). LASSO regression based on EMD (Qin et al., 2016), LASSO regression, deep belief networks (DBN) with EEMD (Chu et al., 2018), kernel ridge regression and EMD (Naik et al., 2018), and LASSO regression based on noise-assisted multivariate EMD (NA-MEMD) (Masselot et al., 2018).

This article is an expansion of the study by Al-Jawarneh et al. (2020) that used the EMD-ELN method for univariate original predictor cases. Al-Jawarneh, et al. (2020) explored the relationship of the orthogonal decomposition components of the nonlinear and non-stationary original predictor with the response variable extracted through the EMD method. This article focused on the cases of the multivariate original predictors by applying the EMD method based on the ELN regression method. For the multivariate case, the effects of the decomposition components extracted through the EMD method of the nonlinear and non-stationary original multivariate predictor variables on the response variable are investigated. When multicollinearity among the decomposition components is determined, the proposed method removes the multicollinearity to arrive at a model free from multicollinearity, and to produce more accurate results

The rest of the paper is organized as follows. Section 2 describes the EMD algorithm, ELN regression, and EMD-ELN algorithm methods. Section 3 discusses and applies the method through numerical experiments using stock market data. Section 4 provides the conclusions of the study.

METHODS

This section describes the methods used in this study. The first method is the EMD algorithm method, which is employed to decompose the original signal (predictor variable) of the dataset. The second method is the regularization method by ELN regression. The proposed EMD-ELN algorithm is discussed in this section.

Empirical Mode Decomposition

Empirical mode decomposition (EMD) algorithm was presented by Huang et al. (1998). The EMD algorithm is a new technique analysis method that aims to decompose the non-stationary and nonlinear signal into a finite set of orthogonal decomposition components called intrinsic mode function (IMF) components. One component called the residual and represents the trend of the signal. The principle of the EMD algorithm is that it analyzes the original signal through an iterative process called the sifting process, which maintains the time domain of the signal (Huang, 2014). These decomposition components represent the temporal modes existing in the original signal, which have different physical significant meanings.

Each of the orthogonal decomposition components should satisfy the following IMF conditions: (1) Over the entire length of a signal, the number of local extrema (EX) (i.e. maximum and minimum) and the number of zero-crossing (Z) should be equal or differ at most by one ($|\#EX - \#Z| \leq 1$), where this condition indicates that each IMF has only one local extrema between two consecutive zero-crossing or vice versa. (2) At any point on a signal, the mean envelope (m) value between the upper (Ue) envelope defined by the local maximum and the lower (Le) envelope defined by the local minimum are zero ($m = \frac{Ue+Le}{2} = 0$), which explains that all IMF are stationary components (Huang, 2014).

The original signal $x(t)$ is the linear combination of the finite set of orthogonal IMF components and monotonic residual component through the EMD algorithm as indicated in the following Equation 1:

$$x(t) = \sum_{k=1}^{K} C_k(t) + r(t) \qquad [1]$$

where symbol t represent the sample index (time domain) $\{C_k(t)\}$, $k = 1, 2,..., K$ denotes the finite set of IMF components, and $r(t)$ is the residual component.

The iterative process of the EMD algorithm to decompose $x(t)$ into orthogonal $C_k(t)$ and $r(t)$ components of a non-overlapping time scale are summarized in Algorithm 1 and Figure 1.

Algorithm 1: EMD algorithm
Input: $x(t)$
Output: $\{C_k(t)\}; k = 1,2, \ldots, K$ and $r(t)$ components

    (1)   Initializations: $r_0(t) = x(t); j = 1$ and $k = 1$.

    (2)   Identify all local maximum and a local minimum of $x(t)$.

    (3)   Determine the upper $Ue_j(t)$ and lower $Le_j(t)$ envelopes through the cubic spline curve of the local maximum and minimum of the $x(t)$, respectively.

    (4)   Compute the mean envelope: $m_j(t) = [Ue_j(t) + Le_j(t)]/2$.

    (5)   Compute: $h_j(t) = x(t) - m_j(t)$.

    (6)   Check that $h_j(t)$ satisfies the conditions of IMF.

           Yes: $h_j(t) = C_k(t)$, save output $C_k(t)$ and go to the next step (7).

           No: $j = j + 1$ and repeat steps (2)–(6).

    (7)   Compute: $r_k(t) = r_{k-1}(t) - C_k(t)$.

    (8)   Check that $r_k(t)$ satisfies the stoppage criterion.

$$SD_j = \sum_{t=0}^{T}\left(h_{j-1}(t) - h_j(t)\right)^2 / h_{j-1}^2(t); \, 0.2 \leq SD_j \leq 0.3$$

           Yes: Go to the next step (9).

           No: $k = k + 1$ and repeat steps (2)–(8).

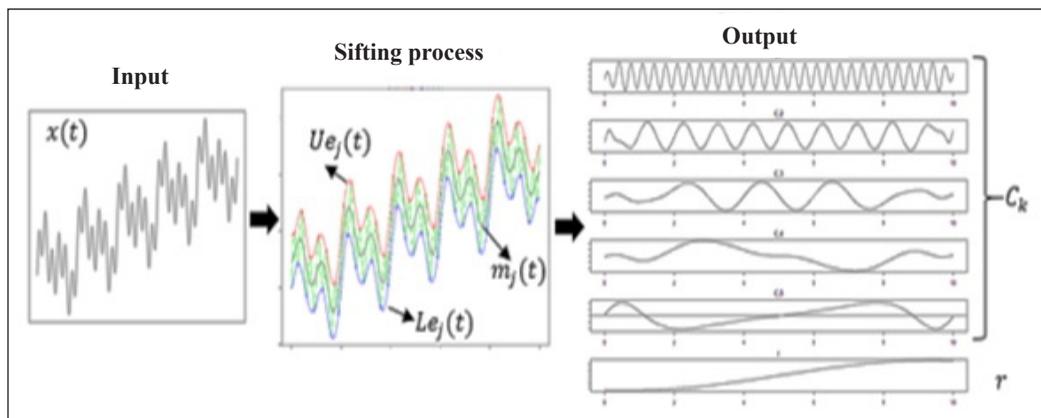   (9)   Save the output: $\sum_{k=1}^{K} C_k(t) + r_k(t)$.

Figure 1. EMD algorithm process

Elastic Net Regression

Elastic net regression (ELN) was introduced by Zou and Hastie (2005). The ELN is a hybrid technology of the penalized least squares regression method that involves regularization and variables selection and was developed to improve and overcome the limitations of the LASSO regression in some situations (i.e., when the number of observations (n) is larger than the number of predictors (p) and the predictor variables have a multicollinearity problem) (Zou & Hastie, 2005).

Penalized regressions methods, such as the ELN method introduce little increase to the bias and contribute to the decrease in the variance by adding penalties to the estimation regression, which leads to the treatment of the multicollinearity and the enhancement of the accuracy of the selection compared to the other methods, such as the ordinary least square (OLS). The ELN is a combination of the two more commonly used penalized regression methods: (1) LASSO method, which uses the sum of the absolute values of the coefficient variables (l_1-norm) with the tuning parameter and (2) the RR method, which uses the sum of the squared coefficient variables (l_2-norm) with the tuning parameter (Javaid et al., 2020). The practical principles of the ELN method are similar to LASSO and involve the shrinkage of the coefficient regression toward zero or should be equal to zero for unnecessary predictors, leading to a reduction in the number of predictor variables (Al-Jawarneh et al., 2020). In a multicollinearity case, the ELN method can remove or select the predictor variables that exhibit high correlation in the final model, unlike the LASSO, which lacks the tools to deal with multicollinearity.

The model structure of the multiple linear regression with n observations and p predictors is derived as Equation 2:

$$y_i = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{ip}\beta_p + \varepsilon_i,$$

$$y = \sum_{j=0}^{p} x_j\beta_j + \varepsilon \qquad [2]$$

where $i = 1,2,\ldots,n; j = 1,2,\ldots,p$, β_0 is the intercept, x_{ij} is the i-th observation of the j-th predictor variable x_j, and β_j is the unknown regression coefficient of the j-th predictor variable, which represents the average effect on y of per one unit change in the j-th predictor variable x_{ij}, while ε_i is the error term.

In the penalized regression methods, the predictors and response variables should be standardized to have zero mean and unit standard deviation (i.e. centered and normalized) before model fitting (Yan & Su, 2009; Zhou, 2013) as Equation 3:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \,, \; \bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \,\,, S_y = \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2} \,, \text{ and } S_{x_j} = \sqrt{\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2}$$

$$[3]$$

Then, the predictors and response variables are standardized as Equation 4:

$$\acute{y}_i = (y_i - \bar{y})/S_y \text{ , and } \acute{x}_{ij} = (x_{ij} - \bar{x}_j)/S_{x_j} \tag{4}$$

The classical OLS estimator is used to estimate the unknown regression coefficients by minimizing the residual sum of squares (RSS). The RSS is the sum of the squared differences between the actual value and the estimated value and is obtained as Equation 5:

$$RSS = \sum_{i=1}^{n}\left(\acute{y}_i - \beta_0 - \sum_{j=1}^{p}\acute{x}_{ij}\beta_j\right)^2 \tag{5}$$

Using Equation 5, the OLS regression for the j-th element $\widehat{\boldsymbol{\beta}}$ (i.e., $\beta_j: j = 0,1,2,\dots,p$) (Montgomery et al., 2012) is obtained as Equation 6:

$$\widehat{\boldsymbol{\beta}}^{OLS} = argmin_\beta \sum_{i=1}^{n}\left(\acute{y}_i - \beta_0 - \sum_{j=1}^{p}\acute{x}_{ij}\beta_j\right)^2 \tag{6}$$

The ELN regression method is the penalized version of the OLS estimator and produces the coefficients regression $\widehat{\boldsymbol{\beta}}$ (Zou & Hastie, 2005) given by Equation 7:

$$\widehat{\boldsymbol{\beta}}^{ELN} = argmin_\beta \sum_{i=1}^{n}\left(\acute{y}_i - \beta_0 - \sum_{j=1}^{p}\acute{x}_{ij}\beta_j\right)^2 + \lambda_1 \sum_{j=1}^{p}|\beta_j| + \lambda_2 \sum_{j=1}^{p}(\beta_j)^2 \tag{7}$$

where λ_1 and λ_2 are the tuning parameters $(\lambda_1, \lambda_2 > 0)$, which are automatically selected by using cross-validation (CV). Another way is to denote $\lambda_1 = 2n\lambda\alpha$ and $\lambda_2 = n\lambda(1-\alpha)$ (Haws et al., 2015), Equation 6 is equivalent to Equation 8:

$$\widehat{\boldsymbol{\beta}}^{ELN} = argmin_\beta \frac{1}{2n}\sum_{i=1}^{n}\left(\acute{y}_i - \beta_0 - \sum_{j=1}^{p}\acute{x}_{ij}\beta_j\right)^2 + \lambda\left(\alpha \sum_{j=1}^{p}|\beta_j| + \frac{1-\alpha}{2}\sum_{j=1}^{p}(\beta_j)^2\right) \tag{8}$$

where α is a regularization parameter $(0 \leq \alpha \leq 1)$ and λ is a tuning parameter $(\lambda \geq 0)$. When $\lambda = 0$, the ELN estimation follows the OLS method in Equation 6 while the ELN estimation follows RR regression shown in Equation 9 when $\alpha = 0$, but when $\alpha = 1$ the

ELN estimation performed using the LASSO regression shown in Equation 10. Thus, the RR and LASSO regressions are cases from the ELN regression. Therefore, the ELN regression sets the appropriate value for α between zero and one.

$$\widehat{\boldsymbol{\beta}}^{RR} = argmin_\beta \frac{1}{2n} \sum_{i=1}^{n} \left(\acute{y}_i - \beta_0 - \sum_{j=1}^{p} \acute{x}_{ij}\beta_j \right)^2 + \frac{\lambda}{2} \sum_{j=1}^{p} (\beta_j)^2 \qquad [9]$$

$$\widehat{\boldsymbol{\beta}}^{LASSO} = argmin_\beta \frac{1}{2n} \sum_{i=1}^{n} \left(\acute{y}_i - \beta_0 - \sum_{j=1}^{p} \acute{x}_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \quad [10]$$

The coordinate descent method and soft-thresholding operator (Friedman et al., 2010) are used to solve Equation 8 with given values of λ and α. The coordinate descent method is used to optimize each predictor separately and solves exactly for one predictor x_{ij} while the rest of the predictors x_{if} except the j-th predictor are fixed in each coordinate descent step. Equation 7 can be rearranged to isolate β_j (Equation 11):

$$\widehat{\boldsymbol{\beta}}^{ELN} = argmin_\beta \frac{1}{2n} \sum_{i=1}^{n} \left(\acute{y}_i - \beta_0 - \sum_{f \neq j}^{p} \acute{x}_{if}\beta_f - \acute{x}_{ij}\beta_j \right)^2 + \lambda\alpha \sum_{j=1}^{p} |\beta_j| + \lambda \frac{1-\alpha}{2} \sum_{j=1}^{p} (\beta_j)^2$$

$$[11]$$

where $\acute{y}_i - \beta_0 - \sum_{f \neq j}^{p} \acute{x}_{if}\beta_f$ is the partial residual $\boldsymbol{r}_f$ that represents the differences between the actual and estimated values that does not involve x_{ij}. Let $X_j = \sum_{i=1}^{n} \acute{x}_{ij}\beta_j$, then Equation 11 can be rearranged as Equation 12:

$$\widehat{\boldsymbol{\beta}}^{ELN} = argmin_\beta \frac{1}{2n} \left(\boldsymbol{r}_f - X_j\beta_j \right)^2 + \lambda\alpha \sum_{j=1}^{p} |\beta_j| + \lambda \frac{1-\alpha}{2} \sum_{j=1}^{p} (\beta_j)^2 \qquad [12]$$

Using the coordinate descent to compute Equation 12, the partial derivative for β_j is as Equation 13:

$$\frac{\partial \widehat{\boldsymbol{\beta}}^{ELN}}{\partial \beta_j} = \frac{1}{n} X_j^T \left(\boldsymbol{r}_f - X_j\hat{\beta}_j \right)^2 + \lambda\alpha \, sign(\hat{\beta}_j) + \lambda(1-\alpha)\hat{\beta}_j \qquad [13]$$

Solving Equation 13 in terms of $\widehat{\boldsymbol{\beta}}_j$ yields the following Equation 14:

$$\hat{\beta}_j = S\left(\frac{1}{n} X_j^T r_f, \lambda\alpha \right) / (1 + \lambda(1-\alpha)) \qquad [14]$$

where $\frac{1}{n}X_j{}^T r_f$ is the simple OLS method to estimate the coefficient $\widehat{\boldsymbol{\beta}}_j$ and $S\left(\frac{1}{n}X_j{}^T r_f, \lambda\,\alpha\right)$ is the soft-thresholding function with the value (Equation 15):

$$
S\left(\frac{1}{n}X_j{}^T r_f, \lambda\,\alpha\right) = \begin{cases} \frac{1}{n}X_j{}^T r_f + \lambda\,\alpha & , & \frac{1}{n}X_j{}^T r_f < -\lambda\,\alpha \\ 0 & , & \left|\frac{1}{m}X_j{}^t r_f\right| \le \lambda\,\alpha \\ \frac{1}{n}X_j{}^T r_f - \lambda\,\alpha & , & \frac{1}{n}X_j{}^T r_f > \lambda\,\alpha \end{cases} \quad [15]
$$

The mean square error (*MSE*) is a measure of the average squared difference between the actual and estimated values as Equation 16:

$$
MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\acute{y}_i - \sum_{j=1}^{p} x'_{ij}\beta_j\right)^2 \quad [16]
$$

The *MSE* is written as the sum of the variance and bias squared terms of the estimator $\widehat{\boldsymbol{\beta}}$ as Equation 17:

$$
MSE(\widehat{\boldsymbol{\beta}}) = E[\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}]
$$
$$
= E\left[\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{E}(\widehat{\boldsymbol{\beta}})\right)^2\right] + \left[\boldsymbol{E}(\widehat{\boldsymbol{\beta}}) - \boldsymbol{\beta}\right]^2
$$
$$
= var(\widehat{\boldsymbol{\beta}}) + Bias^2(\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta}) \quad [17]
$$

where the bias and variance have a trade-off, and thus, the ELN regression aims to achieve the best estimator through an optimal balance between the variance and bias terms by increasing the bias value, leading to a decrease in the variance value to investigate the minimum *MSE*.

Proposed EMD-ELN Method

The proposed EMD-ELN method is intended to understand the significance of the decomposition components, $\{C_{jk}(t)\}; k = 1,2,\ldots,K$ and $r_j(t)\ j = 1,2,\ldots p$, that have the most effect on response variable y in two case studies. In the first case, the components are free from multicollinearity while in the second case, high multicollinearity can be observed among the components. The EMD-ELN method is summarized in Algorithm 2 and Figure 2.

Algorithm 2: EMD-ELN algorithm

Input: $x_j(t)$ and $y(t)$; $j = 1,2,\ldots,p$, and $y(t)$.

Output: $\beta_{jk} \neq 0$; $j = 1,2,\ldots p$ and $k = 1,2,\ldots K, K+1$.

    (1)   Decompose $x_j(t)$ via EMD separately.

$$x_j(t) = \sum_{k=1}^{K} C_{jk}(t) + r_j(t)$$

    (2)   Use all $C_{jk}(t)$ and $r_j(t)$ components as new predictor variables to explain $y(t)$.

$$y(t) = \sum_{j=1}^{p} \left(\sum_{k=1}^{K} C_{jk}(t)\beta_{jk} + r_j(t)\beta_{jK+1} \right) + \varepsilon(t)$$

    (3)   Check the multicollinearity by using the variance inflation factor (VIF) test,

$$VIF_{jk} = \frac{1}{1 - R_{jk}^2},$$

where R_{jk}^2 is the coefficient of determination.

    (4)   Select the variables via the ELN:

$$\widehat{\boldsymbol{\beta}}^{ELN} = argmin_\beta \frac{1}{2n} \sum_{i=1}^{n} \left(y(t) - \beta_0 - \sum_{j=1}^{p} \left(\sum_{k=1}^{K} C_{jk}(t)\beta_{jk} \right) - r_j(t)\beta_{jK+1} \right)^2 + \lambda P_\alpha(\boldsymbol{\beta});$$

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

    (5)   Build the final model by $\beta_{jk} \neq 0$.

APPLICATION

We explain the numerical experiments comprehensively using the sine waves function and the daily close stock market data to apply the EMD-ELN method and discuss the results.

Numerical Experiments

This study uses the sine function to demonstrate the use of the EMD-ELN method. The datasets are generated for the predictors and response signals with the length of a sample size $n = 250$ and time-domain between zero and seven ($0 \leq d \leq 7$) in the first experiment
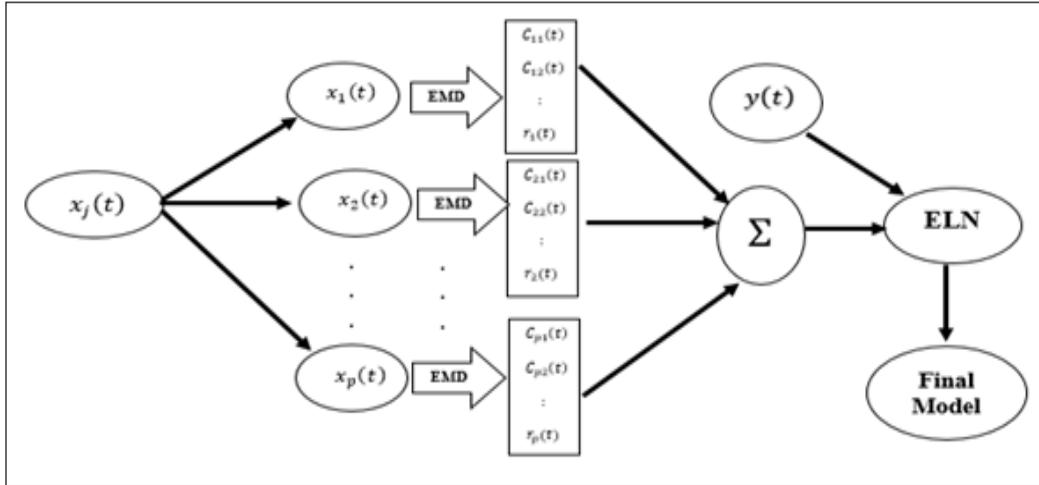
Figure 2. EMD-ELN modeling process

(Experiment 1) and $n = 450$ and $(0 \leq d \leq 9)$ second experiment (Experiment 2). The formulas of the function test in two experiments of the response and predictor variables are as follows.

Exeperiment 1:

$$x(d) = 0.5d + \sin(\pi d) + \sin(2\pi d) + \sin(6\pi d)$$

$$y(d) = \sin(2\pi d) + \sin(6\pi d) + \sin(8\pi d) + + \sin(13\pi d)$$

Experiment 2:

$$x_1(d) = 0.8d + \sin(0.3\pi d) + \sin(2\pi d) + \sin(7\pi d) + \sin(9\pi d)$$

$$x_2(d) = 0.4d + \sin(0.2\pi d) + \sin(6\pi d) + \sin(7\pi d) + \sin(9\pi d)$$

$$x_3(d) = 0.6d + \sin(\pi d) + \sin(7\pi d) + \sin(9\pi d)$$

$$y(d) = 0.5d + \sin(\pi d) + \sin(2\pi d) + \sin(6\pi d)$$

RESULTS AND DISCUSSION

Table 1 presents the number of decomposition components extracted through the EMD algorithm and the multicollinearity test among the components in two experiments. In the first experiment, the original $x(t)$ had three IMFs and one residual component. The VIF test also showed that the components were free from multicollinearity where all values were less than 10. In the second experiment, the results of the EMD algorithm for the original multiple predictors $x_1(t)$, $x_2(t)$, and $x_3(t)$ were as follows: each one of the $x_1(t)$ and $x_2(t)$ has four IMFs and one residual component, while the $x_3(t)$ has five IMFs and one residual component. Based on the VIF test values for this experiment, several decomposition

components had values of more than 10, indicating that high multicollinearity existed among the decomposition components in the second experiment.

Table 2 illustrates the results of the comparison of the methods that depended on the values of lambda, which were selected by 10-fold CV. In the two experiments, the results showed that EMD-ELN had the smallest values of RSS. The penalty value increase appeared to cause a negligible increase in the bias, which contributed to the decrease in the variance of the two experiments. This result indicates that the optimal balance between variance and bias was achieved. For the variable selection in the first experiment, which was free from multicollinearity, the END-ELN method similar to END-LASSO method selected variables C_1 and C_2 and had the most effect on the response variable. In the second experiment, the EMD-ELN was the best method for selecting the decomposition components that exhibited a high correlation and included C_{12}, C_{14}, C_{21}, C_{32}, and C_{33} in the final model, whereas the EMD-LASSO selected C_{21}, C_{32}, and C_{33}. Hence, the EMD-LASSO failed to select any decomposition components from the predictor $x_1(t)$ because of multicollinearity, while the EMD-ELN selected, two components C_{12} and C_{14}, from the $x_1(t)$.

Table 1
Output of EMD and VIF tests

	Expriment 1							
EMD	$C_1(t)$	$C_2(t)$	$C_3(t)$	$r(t)$				
VIF	1.00	1.01	1.00	1.01				
	Experiment 2							
EMD	$C_{11}(t)$	$C_{12}(t)$	$C_{13}(t)$	$C_{14}(t)$	$r_1(t)$	$C_{21}(t)$	$C_{22}(t)$	$C_{23}(t)$
VIF	2414	25.27	6.71	1.36	325.54	3.59	1.34	7.54
EMD	$C_{24}(t)$	$r_2(t)$	$C_{31}(t)$	$C_{32}(t)$	$C_{33}(t)$	$C_{34}(t)$	$C_{35}(t)$	$r_3(t)$
VIF	42.33	7.61	2380	24.28	1.47	38.75	3.38	384.26

Table 2
Comparison of different methods for numerical experiments

Method	Parameter(s)	RSS	Bias^2	Var	Variable Selection
		Experiment 1			
EMD-OLS	$\lambda = 0$	42.91298	5.89e-35	0.566789	All
EMD-RR	$\lambda = 0.05401, \propto = 0$	42.43922	1.05e-36	0.512713	All
EMD-LASSO	$\lambda = 0.05405, \propto = 1$	41.25816	2.59e-36	0.442239	C_1, C_2
EMD-ELN	$\lambda = 0.06434, \propto = 0.84$	41.24562	4.81e-35	0.432979	C_1, C_2
		Experiment 2			
EMD-OLS	$\lambda = 0$	124.2361	4.67 e-3	0.136539	All
EMD-RR	$\lambda = 0.083389, \propto = 0$	118.4152	3.97 e-5	0.084286	All
EMD-LASSO	$\lambda = 0.039912, \propto = 1$	115.2839	3.48 e-4	0.048209	C_{21}, C_{32}, C_{33}
EMD-ELN	$\lambda = 0.276647, \propto = 0.12$	114.9694	8.83 e-4	0.033481	$C_{12}, C_{14}, C_{21}, C_{32}, C_{33}$

Figure 3 shows the graphs of the decomposition components selected using the EMD-ELN method and the response variable in the two experiments. In each graph, the solid black line represents the response variable and other lines are the selected components. The response variable in the first experiment was stationary while the second experiment had modifications (i.e. transformed to be stationary). We showed similarities and matching lines between the components selected and the response variables for the two experiments.

Table 3 illustrates that the results of the compared methods with noise structure errors had zero mean and unity variance $\varepsilon \sim iid\ N(0,1)$ added to the predictor variables in the first and second experiments with 2000 replications modeled. The value of the tuning parameter was chosen automatically using a 10-fold CV. The results show that EMD-ELN had the smallest error value in terms of RSS and achieved the optimal balance between variance and bias in two experiments.
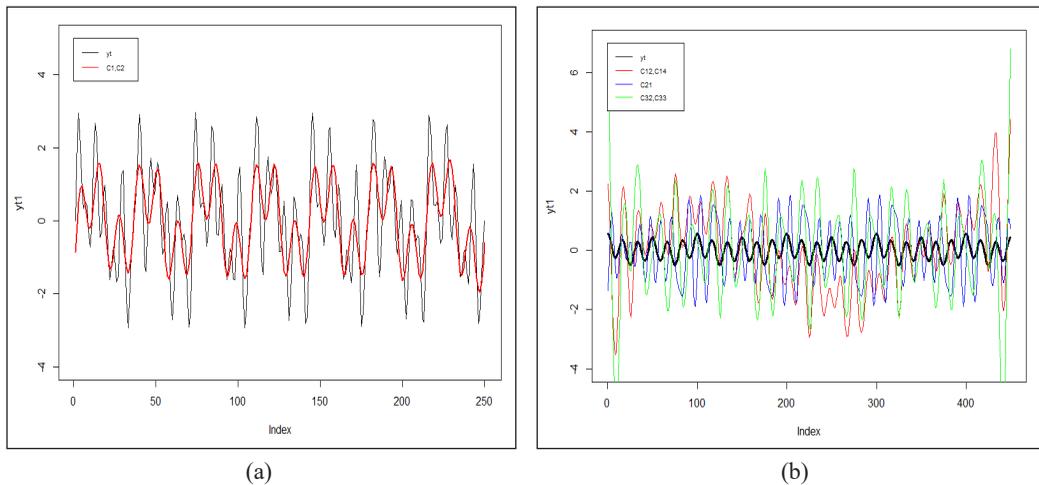

(a)   (b)

Figure 3. Variable selection and response variable: (a) experiment 1; and (b) experiment 2

Table 3
Comparison of different methods for simulation

Method	RSS	Bias^2	Var
Experiment 1			
EMD-OLS	52.03	4.99e-33	0.353
EMD-RR	51.77	9.06e-34	0.273
EMD-LASSO	51.51	8.36e-34	0.264
EMD-ELN	50.93	8.08e-34	0.263
Experiment 2			
EMD-OLS	131.22	0.010120	0.164156
EMD-RR	129.58	0.009838	0.060295
EMD-LASSO	128.40	0.009749	0.066662
EMD-ELN	126.71	0.009629	0.076161

Stock Market

In this study, the daily close stock market from March 26, 2010 to September 25, 2017 for three countries, namely, Japan (JAP), China (CH), and Singapore (SNG), were employed to evaluate the performance of the proposed EMD-ELN and traditional methods. All datasets were collected from the Yahoo finance database (https://finance.yahoo.com/). In this application, the response variable was the daily close stock market of SNG, while the predictor variables were the daily close stock markets of JAP and CH. Each dataset was divided into two parts: 70% for training (from March 26, 2010 until June 24, 2015) and the remaining 30% for testing.

Stock Market Results and Discussion

Figure 4 shows a graphical view of the original close daily stock market signals for CH, JAP, and SNG. Figure 4 shows that the signals did not show any constant value over time or straight lines, thereby indicating that the signals were non-stationary and nonlinear.

Figure 5 shows the plots of the decomposing components extracted by the EMD algorithm from the original predictor variables CH and JAP. The CH signal was decomposed into nine IMFs $\{C_{11}(t), C_{12}(t), C_{13}(t), C_{14}(t), C_{15}(t), C_{16}(t), C_{17}(t), C_{18}(t), C_{19}(t)\}$ and one residual $r_1(t)$ component, while the JAP signal was decomposed into seven IMFs $\{C_{21}(t), C_{22}(t), C_{23}(t), C_{24}(t), C_{25}(t), C_{26}(t), C_{27}(t)\}$ and one residual $r_2(t)$ component with different characteristics (i.e. frequency and wavelength).
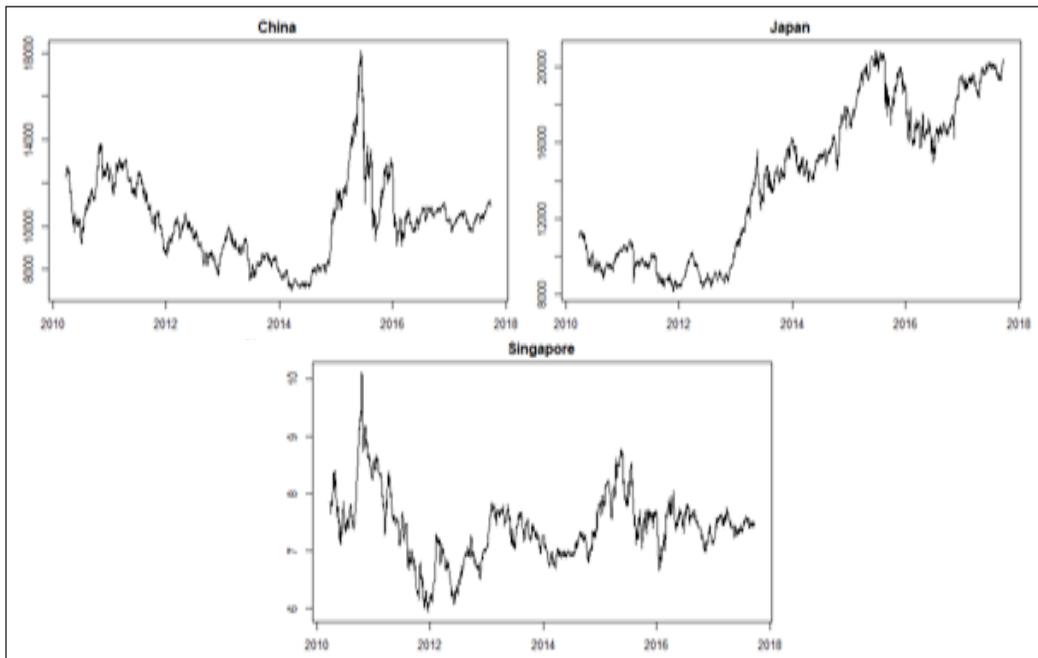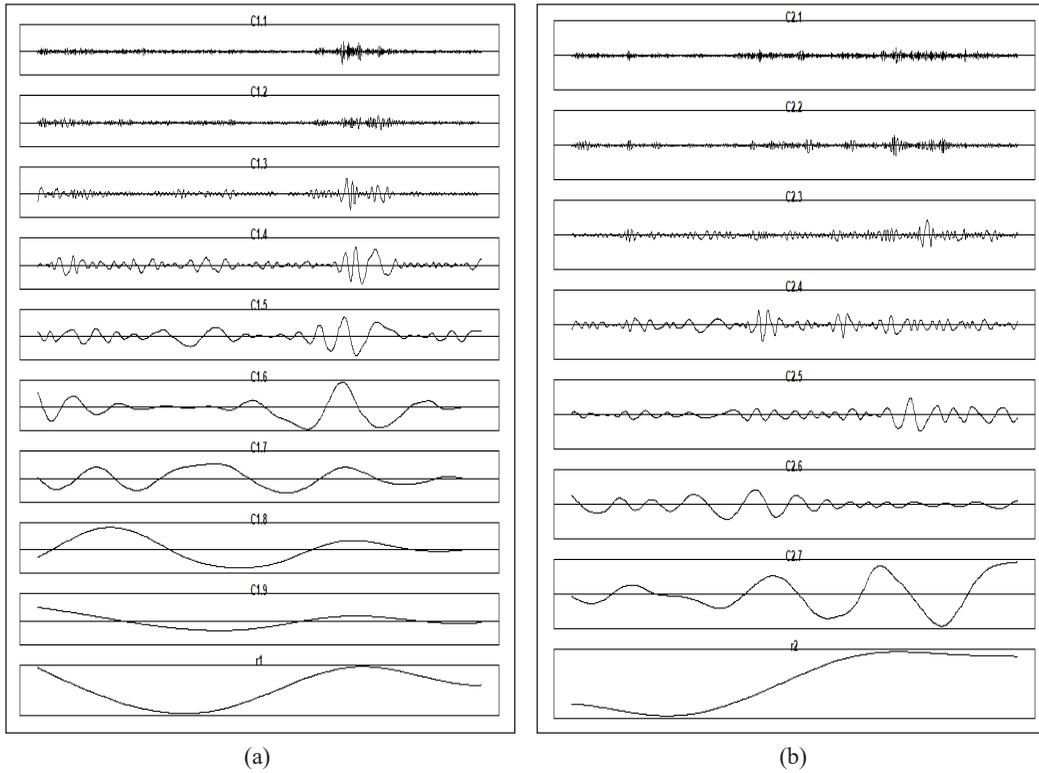

Figure 4. Plots of original signals

Figure 5. Decomposition of: (a) CH; and (b) JAP signals via EMD

Table 4 illustrates the values of the VIF test of multicollinearity among the decomposition components. The results of the VIF test shows that several decomposition components obtained values larger than 10 ($C_{19}(t)$, $r_1(t)$, and $r_2(t) > 10$), which indicates that high multicollinearity exists among the decomposition components.

Table 5 illustrates the results of the compared methods, which depend on the values of λ and were selected by 10-fold CV. The results show that the EMD-ELN outperformed its competitors. It had the smallest RSS ($RSS = 486.5276$ at that resulted in an optimal balance between variance and bias-square. For the variable selection, the EMD-ELN was the best method for selecting decomposition components that exhibit high multicollinearity. These chosen components included the eleven components in the

Table 4
VIF test

EMD	$C_{11}(t)$	$C_{12}(t)$	$C_{13}(t)$	$C_{14}(t)$	$C_{15}(t)$	$C_{16}(t)$	$C_{17}(t)$	$C_{18}(t)$	$C_{19}(t)$
VIF	1.04	1.02	1.05	1.23	1.12	1.52	1.52	4.19	172.0
EMD	$r_1(t)$	$C_{21}(t)$	$C_{22}(t)$	$C_{23}(t)$	$C_{24}(t)$	$C_{25}(t)$	$C_{26}(t)$	$C_{27}(t)$	$r_2(t)$
VIF	326.4	1.03	1.02	1.01	1.09	1.14	1.13	2.14	138.5

final model that had different signs on the response variable. The EMD-LASSO failed to select any decomposition components with high correlation, while the EMD-ELN could remove (like C_{19}) or chose the components that exhibited high correlation (like r_1 and r_2 components) in the final model.

Table 5
Comparison of different methods for stock market application

Method	Parameter(s)	RSS	Bias^2	Var	Variable Selection
EMD-OLS	$\lambda = 0$	759.3457	2.168e-29	0.5784	All
EMD-RR	$\lambda = 0.33320, \propto = 0$	487.0063	2.005e-32	0.0424	All
EMD-LASSO	$\lambda = 0.04205, \propto = 1$	489.6050	1.632e-34	0.0382	$\{C_{11}(t), C_{14}(t), C_{21}(t), C_{11}(t)\}$
EMD-ELN	$\lambda = 0.43233, \propto = 0.02$	486.5276	5.545e-33	0.0320	$\{C_{11}(t), C_{12}(t), C_{24}(t), C_{18}(t), r_1(t), C_{21}(t), C_{22}(t), C_{23}(t), C_{26}(t), C_{27}(t), r_2(t)\}$

CONCLUSION

In this paper, we developed a hybrid EMD-ELN method using non-stationary and nonlinear predictor variables. The proposed method can be used to determine which of decomposition components through the EMD of the original non-stationary and nonlinear predictors have the most effect on the response variable.

The results of the numerical experiments and stock market applications prove that the EMD-ELN method is highly capable of identifying the decomposition components with the most effect on the response variable in the presence or absence of multicollinearity and building a model free from multicollinearity. The proposed EMD-ELN also achieves a balance between bias and variance.

ACKNOWLEDGEMENT

REFERENCE

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, 2020*, 1-20. doi: https://doi.org/10.1080/03610918.2020.1728319

Chu, H., Wei, J., & Qiu, J. (2018). Monthly streamflow forecasting using EEMD-Lasso-DBN method based on multi-scale predictors selection. *Water, 10*(10), 1-15. doi: https://doi.org/10.3390/w10101486

Chui, C. (1995). *Wavelet basics*. Boston, Massachusetts: Kulwer Academic Publishers.

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

Hamid, H., Ngu, P., & Alipiah, F. (2018). New smoothed location models integrated with PCA and two types of MCA for handling large number of mixed continuous and binary variables. *Pertanika Journal of Science and Technology, 26*(1), 247-260.

Hashibah, H., & Mahat, N. I. (2013). Using principal component analysis to extract mixed variables for smoothed location model. *Far East Journal of Mathematical Sciences, 80*(1), 33-54.

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), 1-22. doi: 10.1371/journal.pone.0138903

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

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). Singapore: World Scientific. doi: https://doi.org/10.1142/9789814508247_0001

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. doi: https://doi.org/10.1098/rspa.1998.0193

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. doi: https://doi.org/10.1016/j.stamet.2014.02.002

Javaid, A., Ismail, M., & Ali, M. K. M. (2020). Efficient model selection of collector efficiency in solar dryer using hybrid of LASSO and robust regression. *Pertanika Journal of Science and Technology, 28*(1), 193-210.

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. doi: https://doi.org/10.1016/j.scitotenv.2017.08.276

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

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. doi: https://doi.org/10.1016/j.asoc.2017.12.010

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. doi: https://doi.org/10.1080/03610918.2013.826361

Shen, Z., Feng, N., & Shen, Y. (2012). Ridge regression model-based ensemble empirical mode decomposition for ultrasound clutter rejection. *Advances in Adaptive Data Analysis, 4*(1-2), 1-7. doi: https://doi.org/10.1142/S1793536912500136

Shen, Z., & Lee, C. H. (2012, March 25-30). 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). 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. doi: https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

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

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

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), 1-6. doi: https://doi.org/10.1371/journal.pone.0014612

Zhou, D. X. (2013). On grouping effect of elastic net. *Statistics and Probability Letters, 83*(9), 2108-2112. doi: https://doi.org/10.1016/j.spl.2013.05.014

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. doi: https://doi.org/10.1111/j.1467-9868.2005.00503.x