Afro-Asian Journal of Scientific Research (AAJSR)
المجلة الأفروآسيوية للبحث العلمي
E-ISSN: 2959-6505
Volume 4, Issue 1, 2026
Page No: 174-186
Website: https://aajsr.com/index.php/aajsr/index

SJIFactor 2024: 5.028 ISI 2025: 0.915 معامل التأثير العربي (AIF) 2025: 0.76

Regularization in Quantile Regression based on Empirical Mode Decomposition

Ali Ambark[1*], Mohd Tahir Ismail[2]
[1]Department of Statistics, Faculty of Science, Sebha University, Sebha, Libya
[2]School of Mathematical Sciences, University Science Malaysia, Pulau Pinang, Malaysia

التنظيم في الانحدار الكمي بناءً على تحلل الوضع التجريبي

علي صالح المصري امبارك[1*]، موهد طاهر إسماعيل[2]
[1]قسم الإحصاء، كلية العلوم، جامعة سبها، سبها، ليبيا
[2]مدرسة العلوم الرياضية، جامعة العلوم الماليزية، بولو بينانج، ماليزيا

*Corresponding author: ali.ambark@sebhau.edu.ly

Received: December 21, 2025	Accepted: January 28, 2026	Published: February 5, 2026

Abstract:
Quantile regression is considered a robust alternative to ordinary least squares (OLS) regression in the presence of outliers or heavy-tailed error distributions, providing insights into the conditional distribution of the response variable. The empirical mode decomposition (EMD) decomposes nonstationary and nonlinear signals into a finite set of orthogonal decomposition components, which are then used in several studies as new predictor variables in regression models. In this work, we develop the use of lasso, ridge, and elastic net regularizations in quantile regression utilising a modified percentile cross-validation based on empirical mode decomposition (EMD). These proposed methods aim to identify the decomposed components that exhibit the strongest effects and address the multicollinearity among decomposition components to improve the prediction accuracy. The simulation study and numerical examples utilising the stock market applications dataset from three countries are applied. The results showed that the proposed methods outperformed other existing methods at different quantiles by producing a model free from multicollinearity and effectively identifying the decomposition components that have a significant impact on the response variable, with high prediction accuracy.

Keywords: Empirical mode decomposition (EMD), Quantile regression, Ridge regularization, Lasso regularization, Elastic Net regularization, Multicollinearity.

الملخص:
يُعتبر الانحدار الكمي بديلاً فعالاً لانحدار المربعات الصغرى العادية (OLS) في حال وجود قيم متطرفة أو توزيعات خطأ ثقيلة الذيل، مما يوفر رؤىً شاملة للتوزيع الشرطي لمتغير الاستجابة. يُحلل تحليل الوضع التجريبي (EMD) الإشارات غير الثابتة وغير الخطية إلى مجموعة محدودة من مكونات التحليل المتعامدة، ثم تُستخدم المكونات الناتجة في العديد من الدراسات كمتغيرات تنبؤ جديدة في نماذج الانحدار. في هذا العمل، نُطور استخدام تنظيمات اللاسو والتلال والشبكة المرنة في الانحدار الكمي مع التحقق النسبي المُعدَّل بناءً على تحليل الوضع التجريبي (EMD). تهدف هذه الطرق المقترحة إلى تحديد المكونات المُحللة التي تُظهر أقوى التأثيرات، ومعالجة التعدد الخطي بين مكونات التحليل لتحسين دقة التنبؤ. تُطبق دراسة المحاكاة والأمثلة العددية باستخدام مجموعة بيانات تطبيقات سوق الأسهم لثلاث دول. أظهرت النتائج أن الطرق المقترحة تفوقت على الطرق الأخرى القائمة عند نسب مئوية مختلفة، وذلك بإنتاج نموذج خالٍ من التعدد الخطي، وتحديد مكونات التحليل التي لها تأثير كبير على متغير الاستجابة بدقة تنبؤ عالية.

الكلمات المفتاحية: تحلل الوضع التجريبي (EMD)، الانحدار الكمي، تنظيم التلال، تنظيم لاسو، تنظيم الشبكة المرنة، التعدد الخطي.

Introduction:

The mean regression analysis has been widely studied, and its procedures typically focus on the mean of the response. However, even though it has excellent properties, e.g., linearity and unbiasedness, it is unreliable if the error term has a heavy-tailed distribution or contains outliers. On the other hand, quantile regression, introduced by (Koenker & Bassett, 1978), has gained popularity as an alternative to least squares regression in recent years. Quantile regression has become a popular approach for studying the relationship between the response variable and the predictor variables at any quantile of the conditional distribution function, providing a more comprehensive view of the phenomenon under study. Quantile regression does not make any distributional assumption about the error term in the model. It can provide comprehensive information about the relationship between the response variable and predictors across the whole conditional distribution (Koenker & Bassett, 1978). It is robust against outliers and can handle heteroscedastic datasets (as opposed to linear regression). These properties have led to its widespread use in practical applications. A quantile estimator can describe the entire conditional distribution of the response variable given predictors and provide an overall assessment of the predictors' influence at various quantiles of the response variable (Koenker, 2005). In the regression model, it is often assumed that there is no dependence among the predictor variables, which might not be valid. If this assumption is violated, multicollinearity arises. Furthermore, the regression coefficients may have large sampling variance and misleading signs, which affect both inference and estimation. Thus, multicollinearity is a major issue in regression analysis (Ali et al., 2019).

The empirical mode decomposition (EMD) technique, proposed by (Huang et al., 1998), decomposes nonlinear and non-stationary time series data into different intrinsic mode functions (IMFs) and a residual component through a sifting process. Unlike earlier methods such as wavelet analysis (Chan, 1995) and Fourier analysis (Titchmarsh, 1948), EMD does not require a priori conditions on the data, such as linearity or stationarity, but instead allows the data to speak for itself. The sifting approach produces decomposition components with different wavelengths, amplitudes, and frequencies, indicating that they may be functionally important (Huang, 2014). EMD provides a fully data-driven and unsupervised signal decomposition, possessing the perfect reconstruction property: superimposing all extracted IMFs with the slow residual trend reconstructs the original signal without loss of information or distortion (Faltermeier et al., 2011). These decomposition components can be used as predictor variables to study their impact on a response variable (Al-Jawarneh et al., 2021).

Combining the EMD algorithm with regularization regression has been performed in several scientific fields. Examples include ridge regression with EEMD by (Shen et al., 2012). (Chu et al., 2018) applied LASSO regression based on Ensemble EMD (EEMD), (Qin et al., 2016) applied the LASSO regression based on EMD, and (Masselot et al., 2018) proposed the LASSO regression based on noise-assisted multivariate EMD (NA-MEMD). Recently, (Al-Jawarneh et al., 2021) proposed Elastic Net Regression based on Empirical Mode Decomposition. (Al-Jawarneh & Ismail, 2022) studied the adaptive LASSO regression with an empirical mode decomposition algorithm for enhancing modelling accuracy.

Variable selection is an essential tool for studying important predictors from a large quantity of predictors to create a sparse model with higher forecast accuracy. Regularized regression techniques are essential variable selection methods based on the concept of a penalized objective function that simultaneously conducts variable selection and coefficient estimation (Khan et al., 2019). There is a wide variety of penalized regression techniques, such as the ridge regression proposed by (Hoerl & Kennard, 1970), the least absolute shrinkage and selection operator (LASSO) proposed by (Tibshirani, 1996) and elastic net (EN) (Zou & Hastie, 2005), which combines both the ridge and Lasso penalties, and so on.

Furthermore, penalized approaches have also been applied successfully in quantile regression. For example, (Hu et al., 2021; Li & Zhu, 2008) proposed L1-penalized quantile regression model via the LASSO penalty. (Burgette et al., 2011) presented two approaches based on the LASSO and elastic net penalties for identifying potentially important predictors in quantile regression. To address multicollinearity, (Zaikarina et al., 2016) applied LASSO and ridge penalties in quantile regression with modified percentile cross-validation, and (Sadig & Bager, 2018) utilized ridge regression and quantile regression. Furthermore, (Yan & Song, 2019) studied penalized quantile regression with the elastic net (EnetQR) and adaptive elastic net penalty (AEnetQR). Under the Bayesian framework, (Alhamzawi et al., 2012) proposed Bayesian estimation with adaptive LASSO quantile regression (BALQR). (Tang et al., 2020) proposed the quantile regression with adaptive Lasso and Lasso penalty from a Bayesian point of view. Moreover, some researchers also have used the EMD method in quantile regression. For example, (Jaber et al., 2014) combined EMD, and local polynomial quantile regression (LLQ), and (Junior et al., 2020) employed the EEMD and Quantile-in-Quantile regression techniques. (Zhang et al.,

2020) used EMD to reduce noise in wind speed series for probability density forecasting based on quantile regression and kernel density estimation.

The significant contribution of this study is to deal with the multicollinearity, and to enhance the accuracy of model selection by identifying the decomposition components that have the most effect on the response variable. In this study, the methods, known as QR-Ridge, QR-LASSO, and QR-EN method based on the EMD algorithm is applied.

The remainder of this paper is structured as follows: Section 2 details the EMD method, quantile regression, LASSO, ridge, and elastic net penalties, culminating in a description of the proposed methods. Sections 3 and 4 present simulation studies and real-world applications using daily stock market closing data, respectively. Finally, conclusions appear in Section 5.

Methodology:

In this section, we discuss the methods used in this study. First, the EMD technique is selected, which deals with non-stationary and non-linear predictors by the sifting process. Second, quantile regression and the regularization methods including Ridge, Lasso, and Elastic Net regression will be applied. Finally, the proposed methods called EMD-QR-Ridge, EMD-QR-LASSO, and EMD-QR-EN will be presented.

Empirical Mode Decomposition (EMD):

The empirical mode decomposition (EMD) approach, introduced by (Huang et al., 1998), decomposes nonlinear and nonstationary time series data into several intrinsic mode functions (IMF's) components and one residual component via a sifting process. It is an unsupervised data-driven decomposition that does not require any prior information. Each IMF must satisfy two conditions: (1) the quantity of local extrema and the number of zero-crossings in the complete dataset must be equal or differ by one; and (2) at any given time point, the average value of the upper envelope, delineated by the local maxima, and the lower envelope, delineated by the local minima, must equal zero. The EMD process may be expressed simply as follows:

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

Where the $C_k(t)$ are intrinsic mode functions (IMFs), and $r(t)$ is the residue. The EMD algorithm analyzes signals using an iterative algorithm approach known as the sifting process. The sifting process separates the IMF and residual components from the original signal. It is worth noting that the IMF's have physical definitions for instantaneous frequency and amplitude. In other words, the IMFs are a physically meaningful time frequency energy representation of a time series (Huang & Wu, 2008). The flowchart for the sifting method is presented in Figure 1. The sifting procedure includes the following steps: The EMD implementation consists primarily of the following steps:

Empirical Mode Decomposition (EMD) Algorithm:

1. For given a time series $\boldsymbol{X(t)}$, determine all local maxima and minima.
2. Use cubic spline interpolation to connect all the minima and maxima, thereby forming the lower and upper envelopes.
3. Calculate the local mean of the upper and lower envelope.
$$\boldsymbol{m(t)} = \frac{\boldsymbol{u(t)+l(t)}}{\boldsymbol{2}}$$
4. Subtract the mean $\boldsymbol{m(t)}$ from the signal $\boldsymbol{X(t)}$ to obtain the first component IMF candidate
$$\boldsymbol{h(t)=X(t)-m(t)}$$
5. Repeat the sifting operation, which consists of step 1 to step 4 by considering $\boldsymbol{h(t)}$ as new X(t) until Step 4. the stopping criteria is reached:
   I. $\boldsymbol{m(t)}$ approaches zero
   II. the numbers of zero-crossings and extrema of $\boldsymbol{h(t)}$ differs at most by one, or
   III. maximum number of iterations is reached
6. Treat h(t) as new IMF and calculate the residual signal r(t) as: $\boldsymbol{r(t)=X(t)-h(t)}$
7. Use $\boldsymbol{r(t)}$ as new time series $\boldsymbol{X(t)}$ and repeat steps 1 to 6, until all IMFs are obtained.

The EMD method decomposes the complex signal into a finite, often tiny number of intrinsic mode functions (IMFs). IMF contains high to low frequencies, using the local characteristic 90 scales, defined as the distance between two successive local extrema in the signal. An IMF is a function of symmetric upper and lower envelopes. Furthermore, the number of zero-crossings and extremes are equal or vary by no more than one (Huang et al., 1998). Each computed IMF has oscillatory scales in a small spectrum and is typically regarded as a quasi-stationary variable. For instance, an IMF obtained from a three-month economic time series can be interpreted as the seasonal component. The specific decomposition process of EMD is shown in Figure 1.

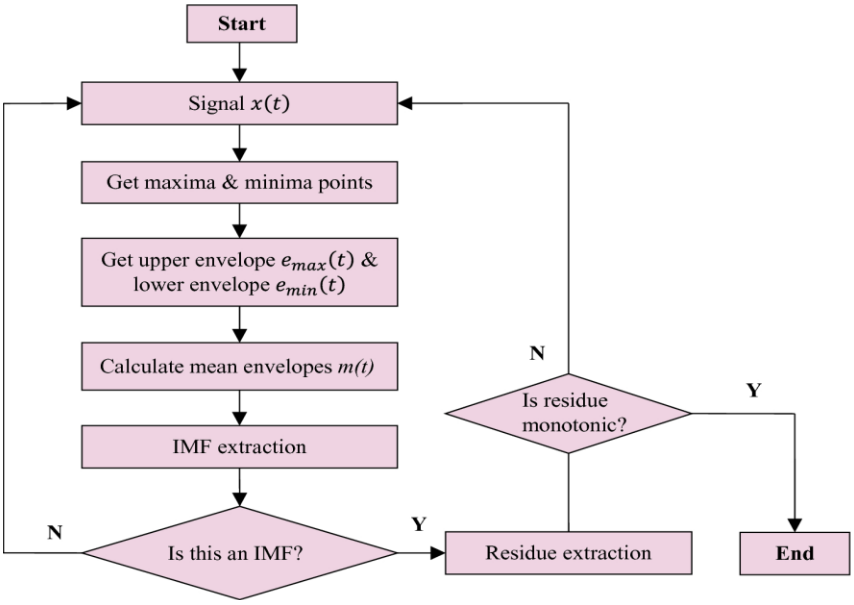

Figure (1): Flowchart for sifting process

Quantile Regression (QR):

Classical regression models provide a symmetric linear relationship between dependent and independent variables, with findings based on the mean-value relationship rather than analyzing the relationship at various degrees of conditional distribution of the response variable. The quantile regression classic model introduced by (Koenker & Bassett, 1978) extends the classical linear regression model that provides consistent and detailed covariate effects on the dependent variable by modelling their time-varying degree and dependency structure (Koenker, 2005). The QR method investigates the dependency of the tails of the dependent variable's distributions, allowing for a concise explanation of the relationship between variables.

Consider the linear regression model given by:

$$y_j = \beta_0 + x_j^T \beta + \varepsilon_j \quad for\ j = 1,2,3, \dots \dots , n \qquad (2)$$

Where y_i denotes the value of the response variable, $x_j^T = (x_{j1}, x_{j2}, \dots \dots , x_{jq})$ represents the q known predictor observations, β_0 is the intercept, indicates β a $q \times 1$ the vector of yet estimated unknown regression coefficients (parameters), and ε_j represents error terms.

In quantile regression, the parameters are estimated by:

$$\hat{\beta}_\tau = arg_\beta min \sum_{j=1}^{n} \rho_\tau \left(y_j - \beta_0 - x_j^T \beta\right) \qquad (3)$$

Where ρ is a quantile loss function, $\tau \in [0,1]$

LASSO Regularization:

The LASSO (Least Absolute Shrinkage and Selection Operator) is a regression method proposed by (Tibshirani, 1996). It is a standard statistical technique used with generalized linear models to select predictors by shrinking specific coefficients to zero. LASSO will constrain the regression coefficients by decreasing the residual number of squares under the condition that the sum of the absolute values of the coefficients is less than a constant.

The LASSO estimates are defined as:

$$\beta(\text{LASSO}) = argmin \left\{ \sum_{j=1}^{n} \left(y_j - \beta_0 + \sum_{i=1}^{q} x_{ij} \beta_i \right) s.t. \sum_{i=1}^{q} |\beta_i| \le s \right\} \qquad (4)$$

Where $s \ge 0$ is a tuning parameter. The LASSO penalty is often called an L1 penalty because of the first power in the penalty term.

$$\beta(\text{LASSO}) = argmin \left\{ \sum_{j=1}^{n} \left(y_j - \beta_0 + \sum_{i=1}^{q} x_{ji} \beta_i \right) + \lambda \sum_{i=1}^{q} |\beta_i| \right\} \qquad (5)$$

Since LASSO effectively overcomes the limitations of conventional variable-selection approaches, it has gained a lot of interest in the fields of regression and classification.

We consider quantile regression with the LASSO penalty based on (Li & Zhu, 2008) LASSO regularized quantile regression estimation given by:

$$\hat{\beta}_\tau = arg_\beta min \sum_{j=1}^{n} \rho_\tau \left(y_j - \beta_0 - x_j^T \beta\right) + \lambda \sum_{i=1}^{q} |\beta_i|, \qquad (6)$$

where λ is the penalty parameter (regularizer) that controls the amount of shrinkage.

Ridge Regularization:

(Hoerl & Kennard, 1970) proposed Ridge regression. It is the most common and has a wide range of applications. Ridge regression minimizes the residual sum of squares while keeping a limit on the coefficients' L2-norm. As a continuous shrinkage approach, Ridge regression achieves better prediction accuracy through a bias-variance trade-off. Ridge regression is usually used to overcome multicollinearity.

$$\beta(R) = argmin \left\{ \sum_{j=1}^{n} \left(y_j - \beta_0 + \sum_{i=1}^{q} x_{ji} \beta_i \right) + \lambda \sum_{i=1}^{q} \beta_i^2 \right\} \qquad (7)$$

Where λ is a positive ridge parameter $(0 < \lambda < 1)$.

The QR penalized with the ridge penalty (7) denoted by QR-Ridge. The QR-Ridge is given by the minimization problem of:

$$\hat{\beta}_\tau = arg_\beta min \sum_{j=1}^{n} \rho_\tau \left(y_j - \beta_0 - x_j^T \beta\right) + \lambda \sum_{i=1}^{q} \beta_i^2 \qquad (8)$$

Elastic Net (EN) Regularization:

(Zou & Hastie, 2005) proposed the Elastic Net regularization technique. It is a convex combination of the LASSO and Ridge penalty. Estimates of Elastic Net coefficients are obtained by decreasing the regression loss function using an Elastic Net penalty:

$$\sum_{i=1}^{q} [\alpha|\beta_i| + (1-\alpha)\beta_i^2] \le k \qquad (9)$$

The coefficient estimator in elastic-net regularized quantile regression is defined as:

$$\hat{\beta}_\tau = argmin \left\{ \sum_{j=1}^{n} \rho_\tau \left(y_j - \beta_0 - x_j^T \beta\right) + \lambda \sum_{i=1}^{q} [\alpha|\beta_i| + (1-\alpha)\beta_i^2] \right\} \qquad (10)$$

Which is the LASSO penalty for $\alpha = 1$ (Hoerl & Kennard, 1970), the Ridge penalty for $\alpha = 0$ (Hoerl & Kennard, 1970) and the Elastic Net penalty for $0 \le \alpha \le 1$ (Zou & Hastie, 2005).

The Proposed Methods:

In order to improve prediction accuracy, we proposed three hybrid prediction models, namely EMD-QR-Ridge, EMD-QR-LASSO and EMD-QR-EN. The flowchart of the proposed method is shown in Figure 2. It can be summarized as the following steps:

Step 1: Using the EMD method, each original signal $X(t)$ is decomposed into a finite set of IMF components and one residual component, expressed as follows.

$$X(t) = \sum_{k=1}^{K} C_k(t) + r(t) \qquad (11)$$

Step 2: Use all the decomposition components in Step 1 as predictor variables to explain the behavior of the response variable (Masselot et al., 2018).

$$y(t) = \sum_{i=1}^{q} \left[\sum_{k=1}^{K} C_{ik}\beta_{ik} + r_{ik}(t)\beta_{ik} \right] + \varepsilon(t) \qquad (12)$$

Step 3: Apply the proposed methods:

i. The EMD-QR-Ridge method:

$$\min_\beta \left[\frac{\rho_\tau}{n} \left(y(t) - \sum_{i=1}^{q} (\sum_{k=1}^{K} C_{ik}(t)\,\beta_{ik} - r_i(t)\beta_{ik+1}) \right)^2 \right] + \lambda P(\beta) \qquad (13)$$

$$\lambda P(\beta) = \lambda \sum_{i=1}^{q} \left[\sum_{k=1}^{K+1} \beta_{ik}^2 \right]$$

ii. The EMD-QR-LASSO method:

$$\min_\beta \left[\frac{\rho_\tau}{n} \left(y(t) - \sum_{i=1}^{q} (\sum_{k=1}^{K} C_{ik}(t)\,\beta_{ik} - r_i(t)\beta_{ik+1}) \right)^2 \right] + \lambda P(\beta) \qquad (14)$$

$$\lambda P(\beta) = \lambda \sum_{i=1}^{q} \left[\sum_{k=1}^{K+1} |\beta_{ik}| \right].$$

iii. The EMD-QR-EN method:

$$\min_{\beta}\left[\frac{\rho_\tau}{n}\left(y(t)-\sum_{i=1}^{q}(\sum_{k=1}^{K}C_{ik}(t)\,\beta_{ik}-r_i(t)\beta_{ik+1})\right)^2\right]+\lambda P(\beta) \qquad (15)$$

$$\lambda P(\beta)=\lambda\left(\alpha\sum_{i=1}^{q}\left[\sum_{k=1}^{K}\beta_{ik}^2\right]+\frac{(1-\alpha)}{2}\sum_{i=1}^{q}\left[\sum_{k=1}^{K+1}\beta_{ik}^2\right]\right)$$

Step 4: The performance of the proposed models is compared with the traditional methods using the following criteria: The Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Scaled Error (MASE), and correlation coefficient (R) are used to estimate the performance of hybrid models, as calculated by Equations (16) through (19), respectively.

$$RMSE=\sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(y_j-\hat{y}_j\right)^2} \qquad (16)$$

$$MAE=\frac{1}{n}\sum_{j=1}^{n}\left|y_j-\hat{y}_j\right| \qquad (17)$$

$$MASE=\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\left|y_j-\hat{y}_j\right|}{\frac{1}{n-1}\sum_{j=2}^{n}\left|y_j-y_{j-1}\right|}\right) \qquad (18)$$

$$R^2\frac{SSR}{SST}=\frac{\sum_{j=1}^{n}\left(\hat{y}_j-\bar{y}\right)^2}{\sum_{j=1}^{n}\left(y_j-\bar{y}\right)^2} \qquad (19)$$

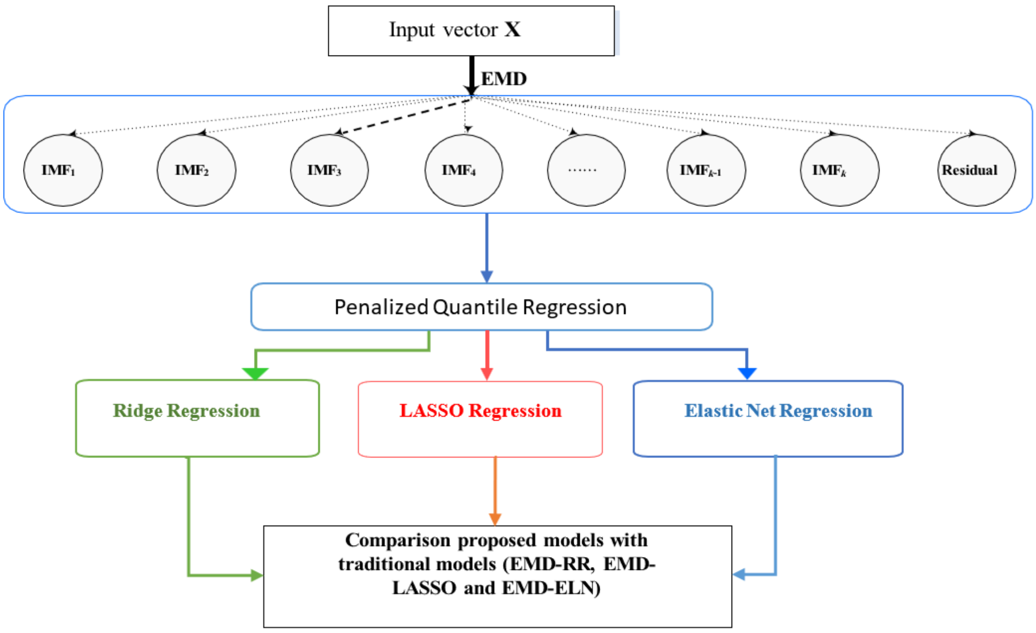

Figure 2: The full flowchart of the proposed models.

Where n represents the number of datasets; y_j denotes the observed data; $\hat{y}_j$ indicates the predicted value of variable y_j at the time period j.

Numerical study:

This section presents two experiments demonstrating that regularized quantile regression based on EMD is an effective variable selection technique and improves prediction accuracy. We implemented the proposed models EMD-QR-Ridge, EMD-QR-LASSO, and EMD-QR-EN in a simulation study to compare prediction accuracy between the proposed methods and traditional approaches. The sine function is used to illustrate the application of the proposed models. The predictor variables (X_j) and response variable Y were simulated from signals selected by the work of (Abdullah Suleiman Al-Jawarneh & Ismail, 2021). R programming was used for simulation study to evaluate and compare the methods. The simulation experiments are replicated 1000 times with a sample size of n = 300, and a time domain between $(0 \leq t \geq 9)$ and three quantile regression levels, $\tau = (0.25, 0.5, 0.75)$, are considered. The results are shown below.

$$x_1 = 0.8t + \sin(0.3\pi t) + \sin(2\pi t) + \sin(7\pi t) + \sin(9\pi t)$$
$$x_2 = 0.4t + \sin(0.2\pi t) + \sin(6\pi t) + \sin(7\pi t) + \sin(9\pi t)$$
$$x_3 = 0.6t + \sin(\pi t) + \sin(7\pi t) + \sin(9\pi t)$$
$$y = 0.5t + \sin(\pi t) + \sin(2\pi t) + \sin(6\pi t)$$

Numerical Results:

The VIFj values are given in Table 1 to test for multicollinearity among all orthogonal IMFs. Some of these values are greater than 10, such as VIF1, VIF6, VIF10, VIF11, VIF12, VIF14, and VIF15. The results indicate that high multicollinearity exists among the decomposition components (IMFs). Next, all the IMFs and the residual were applied to the proposed methods and other comparison methods. We compare the EMD-QR-Ridge, EMD-QR-LASSO, and EMD-QR-EN models to the QR-R, QR-LASSO, and QR-EN models and report the comparison results. Table 2 shows the MSE, MAE, RMSE, and MASE results of the simulation experiments.

The implementation of EMD significantly improves the estimation accuracy of the QR-R, QR-LASSO, and QR-EN methods. The experimental results show that the proposed methods obtain the lowest MSE, MAE, SSE, and RMSE values compared with other methods at the 0.25, 0.5, and 0.75 quantiles. Table 2 indicates that the results offered by EMD-QR-EN outperform those of EMD-QR-Ridge and EMD-QR-LASSO in MSE, MAE, RMSE, and MASE for each quantile. It has been found that EMD helps improve accuracy in most cases.

Table (1): Variance inflation factors

VIF1	VIF2	VIF3	VIF4	VIF5	VIF6	VIF7	VIF8	VIF9	VIF10	VIF11	VIF12	VIF13	VIF14	VIF15
1468.7	2.98	5.17	2.03	1.18	395.9	3.35	1.32	7.62	53.59	11.17	1462.9	1.94	35.91	448.8

Table (2): Comparison our methods with other methods for simulation data.

Quantile		EMD-QR-Ridge	EMD-QR-LASSO	EMD-QR- EN	QR- Ridge	QR- LASSO	QR-EN
0.25	MSE	0.13758205	0.0283528	0.03870044	2.22346946	2.0807123	2.1124556
	MAE	0.2986859	0.1301398	0.1503589	1.216709	1.181563	1.188918
	RMSE	0.3709205	0.1683829	0.1967243	1.49113	1.442467	1.453429
	MASE	0.7310645	0.3185307	0.3680188	2.97802	2.891998	2.91
	R^2	0.9811517	0.9944794	0.9925894	0.5406485	0.5488937	0.5489137
	λ_{min}	0.1265206	0.0126521	0.04217353	0.1174104	0.0117410	0.0391368
0.50	MSE	0.07760487	0.0206842	0.02648215	1.44364178	1.3840013	1.3904879
	MAE	0.2314241	0.0997259	0.121662	0.9661434	0.9413294	0.9452408
	RMSE	0.2785765	0.1438201	0.1627334	1.201516	1.176436	1.17919
	MASE	0.5664344	0.2440895	0.2977802	2.364736	2.304001	2.313574
	R^2	0.9818196	0.9932629	0.9920501	0.5406225	0.5487139	0.5481236
	λ_{min}	0.1646433	0.0164643	0.05488109	0.1572392	0.0157239	0.0524131
0.75	MSE	0.14261358	0.0285204	0.03716151	2.21744426	2.0499641	2.104961
	MAE	0.3079185	0.125885	0.1501264	1.203926	1.164461	1.178595
	RMSE	0.3776421	0.1688798	0.1927732	1.489109	1.43177	1.450848
	MASE	0.7536622	0.3081165	0.3674498	2.946733	2.850139	2.884732
	R^2	0.9792324	0.9938648	0.9920276	0.5399351	0.5480745	0.5488819
	λ_{min}	0.1268101	0.0126810	0.04227004	0.1305056	0.0130506	0.0435019

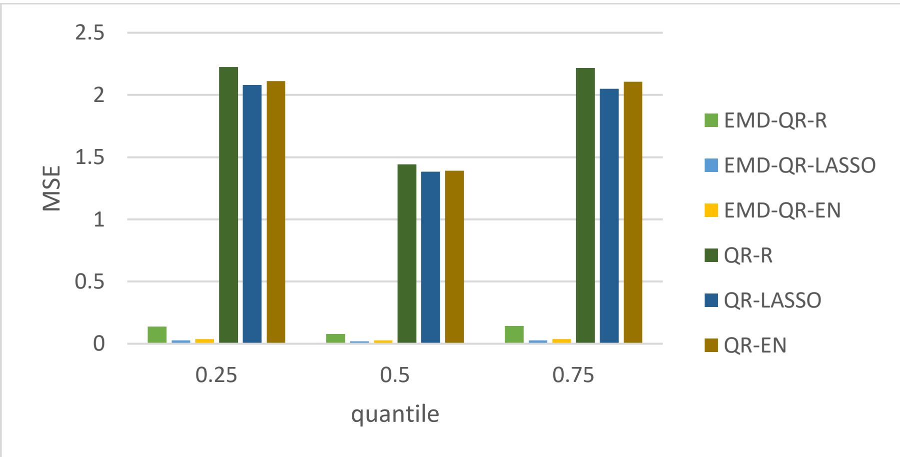

Figure 3: The result of MSE from simulation study

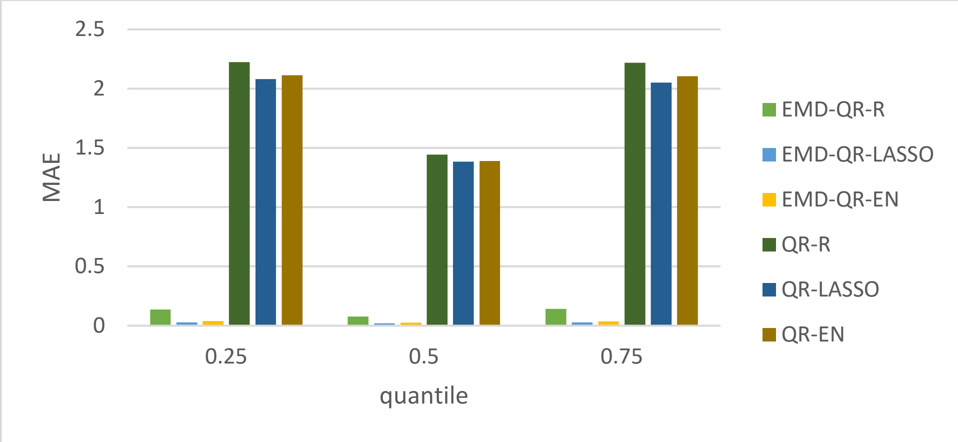

Figure 4: The result of MAE from simulation study

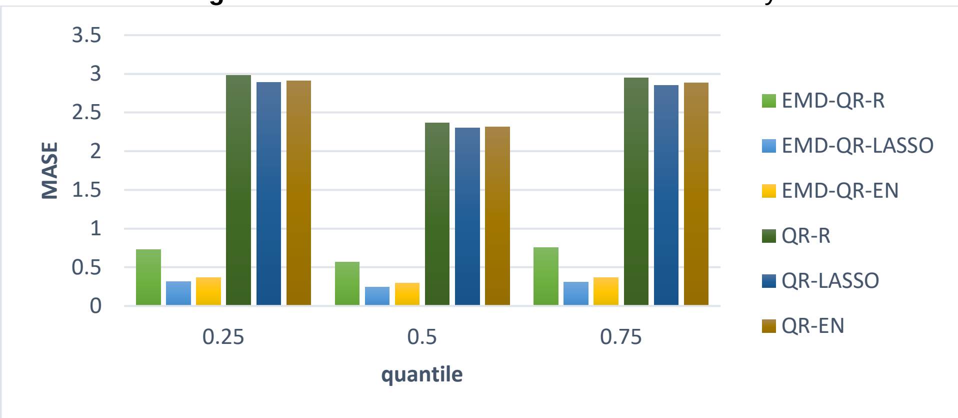

Figure 5: The result of MASE from simulation study

Figures 3, 4, and 5 present bar plots for the MSE, MAE and MASE, respectively. It can be seen that the results provided by EMD-QR-EN for each quantile outperform both EMD-QR-Ridge and EMD-QR-LASSO in terms of MSE, MAE, RMSE, MAPE and MASE. It has been found that EMD is helpful for accuracy improvement in most cases. Generally, it is indicated that the introduction of EMD can improve the accuracy of estimation. The proposed hybrid EMD-QR-Ridge, EMD-QR-LASSO, and EMD-QR-EN models outperform other estimation models in simulation data. On the contrary, in this study, the minimum MSE, MAE, RMSE, MAPE, and MASE values achieved by the other estimating models are higher than those of our models, which indicates that our proposed models are more accurate than those considered by others. Accordingly, the EMD-QR-Ridge, EMD-QR-LASSO, and EMD-QR-EN models are suitable and reasonable for estimation.

APPLICATION TO ACTUAL DATA:

In this study, to compare the suggested methods (EMD-QR-LASSO, EMD-QR-Ridge, EMD-QR-EN) performance to that of other techniques such as (EMD-LASSO, EMD-Ridge, EMD-EN), we use nonlinear and nonstationary time series from the daily stock market. The application has three variables: China's and Japan's daily closing stock market (predictor variables), and Singapore's daily closing stock market (response variable) for the period from March 1, 2011, to August 25, 2015, with 1,701 observations. All data are downloaded from the Yahoo financial database.

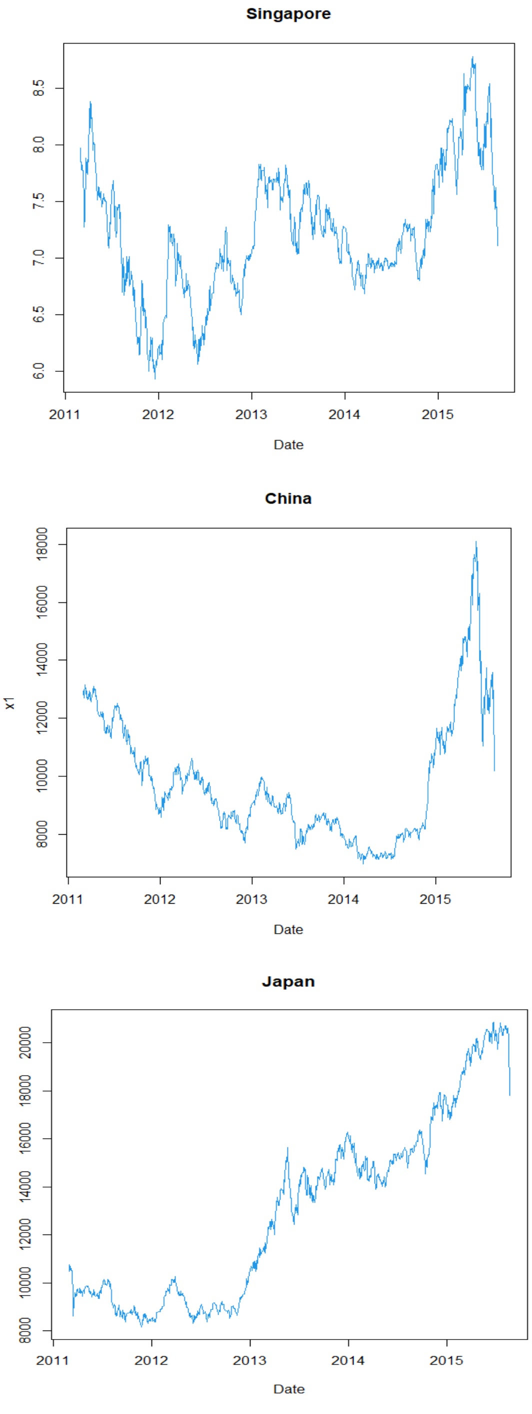

Figure (6): Figure daily stock market Index are plotted over time.

Application results:

Figure 6 displays a graphical representation of the original closing daily stock market signals for China, Japan, and Singapore. Figure 6 illustrates that the signals did not exhibit a constant value over time or follow straight lines, indicating that they were nonstationary and nonlinear. The EMD algorithm decomposes the original series predictor variables CH and JAP, as shown in Figure 7. The CHINA signal has been decomposed into six IMFs and one residue component, while the JAPAN signal has been decomposed into eight IMFs and one residue component.

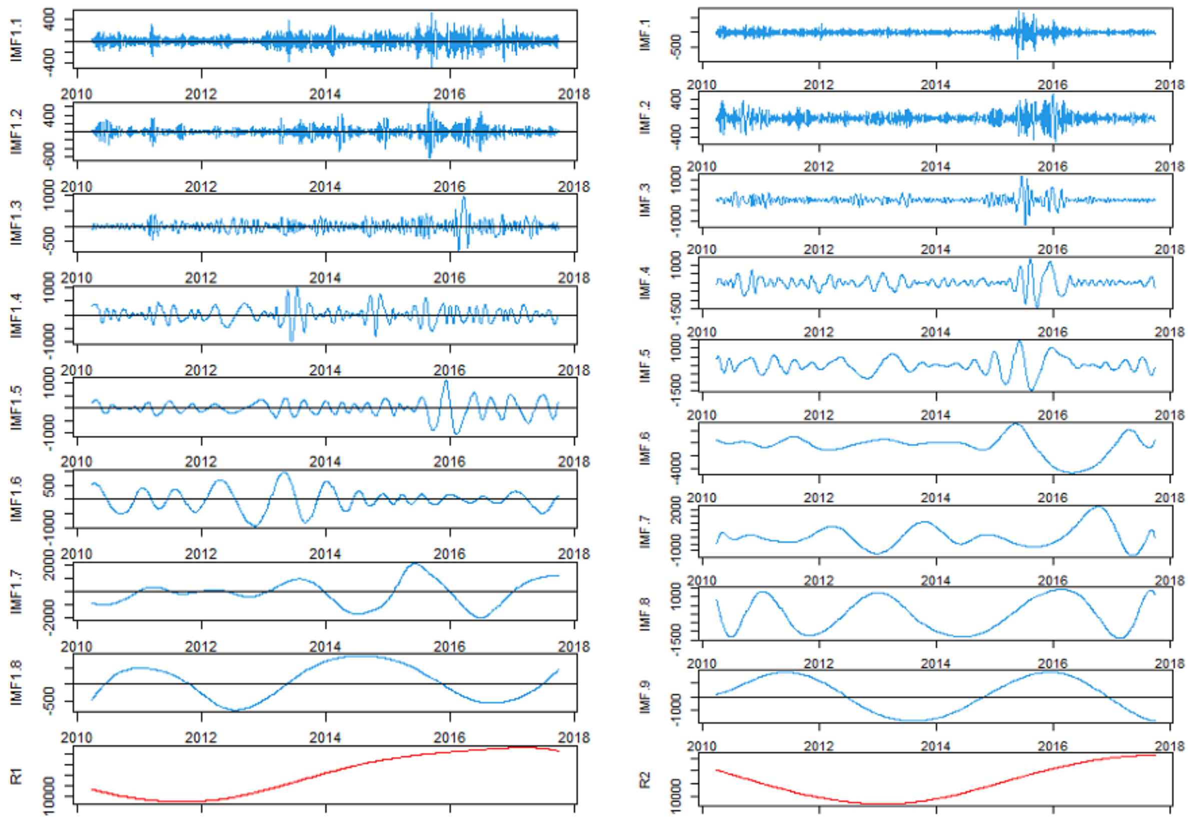

Figure (7): EMD decomposition results of China and Japan signals

Table (3): Variance inflation factors

VIF1	VIF2	VIF3	VIF4	VIF5	VIF6	VIF7	VIF8	VIF9	VIF10	VIF11	VIF12	VIF13	VIF14	VIF15	VIF16
1.052	1.017	1.022	1.276	1.087	3.577	19.07	1.048	1.027	1.048	1.24	1.035	1.312	3.269	1.366	18.79

Table (4): Comparison our methods with other methods for simulation data.

Quantile		EMD-QR-Ridge	EMD-QR-LASSO	EMD-QR- EN	QR- Ridge	QR- LASSO	QR-EN
0.25	MSE	0.2599072	0.2207811	0.2462280	0.3265920	0.3306543	0.3291636
	MAE	0.3594287	0.3328711	0.3506325	0.4053439	0.4072541	0.406479
	RMSE	0.509811	0.4698734	0.4962137	0.5714823	0.5750255	0.5737278
	MASE	5.758989	5.333467	5.618051	6.494672	6.525278	6.512858
	R^2	0.4386542	0.5270165	0.4776494	0.3021939	0.2960662	0.2980071
0.50	MSE	0.1801750	0.1625618	0.1736424	0.2067098	0.2050822	0.2057799
	MAE	0.2948102	0.2809755	0.2896979	0.3227566	0.3219075	0.3222266
	RMSE	0.4244702	0.4031895	0.4167042	0.4546535	0.45286	0.4536297
	MASE	4.723632	4.501963	4.641719	5.171406	5.157802	5.162915
	R^2	0.4977066	0.5384446	0.5139128	0.4052256	0.4052836	0.4052794
0.75	MSE	0.2732941	0.2776650	0.2925501	0.2851579	0.2847990	0.2819129
	MAE	0.3744988	0.3762371	0.38335	0.3887371	0.3890388	0.3877008
	RMSE	0.5227754	0.5269392	0.540879	0.5340018	0.5336656	0.5309547
	MASE	6.000453	6.028304	6.142271	6.228587	6.23342	6.211983
	R^2	0.4165107	0.42279	0.3829844	0.3634616	0.3712168	0.3731046

Table 3 illustrates that VIF results reveal that some decomposition components have values greater than 10, such as IMF7 and IMF16. These high values indicate that high multicollinearity exists among the decomposition components (IMFs). The prediction results for the close daily stock market are presented in Table 4, which compares the results obtained with the QR-Ridge, QR-LASSO and QR-EN models for the daily stock market. Table 4 shows that the MSE of the proposed (EMD-QR-Ridge, EMD-QR-LASSO, EMD-QR-EN) models are, respectively, (0.2599072, 0.2207811, 0.2462280). It can be observed that these values are smaller than those of other models. The numbers in bold indicate that our methods have the best performance for this close daily stock market under these performance measures. According to the prediction results, we can infer that the performance of our methods outperforms that of the other methods.

Table 5 illustrates how to estimate the coefficients of the predictor variables (IMFs and residual components) for our regression methods, indicating that our EMD-QR-LASSO and EMD-QR-EN methods behave similarly to LASSO quantile regression in terms of variable selection. Except for EMD-QR-Ridge, all predictor variables are included in the final model. Both regression techniques can reduce the number of predictor variables. For example, neither EMD-QR-LASSO regards IMF8, IMF9, IMF11, and IMF14 as important factors at 0.75.

In Figures 9 and 10, it can be seen that the EMD-QR-LASSO model exhibits the best prediction performance at 0.25 and 0.50, compared to the EMD-QR-Ridge and EMD-QR-EN models. Whereas the lowest MSE and MAE in 0.75 is EMD-QR-EN. Overall, it shows that (QR-Ridge, QR-LASSO and QR-EN) models will significantly improve when introducing EMD.

Table (5): Coefficients estimation for the predictor variables in the close daily stock market.

	0.25			**0.50**			**0.75**		
	EMD-QR-Ridge	**EMD-QR-LASSO**	**EMD-QR-EN**	**EMD-QR-Ridge**	**EMD-QR-LASSO**	**EMD-QR-EN**	**EMD-QR-Ridge**	**EMD-QR-LASSO**	**EMD-QR-EN**
λ_{min}	0.07492588	0.0074926	0.0249752	0.1169624	0.0116962	0.0389874 8	0.0954545 5	0.0095455	0.0318181
β_0	-97.870	-388.96	-18.582	-93.1849	-222.525	-135.581	-87.1851	-148.31	-61.387
β_1	0.00005	0	0	0.00009	0.00002	0.00005	0.00012	0.00007	0.00008
β_2	0.00017	0.00011	0.00012	0.00014	0.00012	0.00011	0.00016	0.00009	0.00008
β_3	0.00012	0.00012	0.00011	0.00012	0.00007	0.00008	0.0002	0.00016	0.00018
β_4	0.00022	0.00022	0.00021	0.00015	0.00016	0.00015	0.00008	0.00007	0.00006
β_5	-0.00003	-0.00006	-0.00004	0.00002	-0.00002	0	0.00007	0.00004	0.00005
β_6	0.00006	0.00012	0.00008	0.00009	0.00011	0.00011	0.00007	0.00011	0.00007
β_7	0.00008	0.00018	0.00011	0.00006	0.00011	0.00008	0.00002	0.00004	0
β_8	0.00005	0	0	0.00009	0	0.00002	0.00007	0	0
β_9	0.0001	0	0.00003	0.00003	0	0	0.00001	0	0
β_{10}	-0.00003	-0.00003	-0.00002	-0.00006	0	-0.00001	0.00009	0.00007	0.00005
β_{11}	-0.0001	0.00001	-0.00009	-0.00007	-0.00005	-0.00005	-0.00002	0	0
β_{12}	0.0001	0.00001	0.0001	0.00017	0.00018	0.00016	0.00021	0.0002	0.0002
β_{13}	0.0001	0.00001	0.0001	0.00016	0.00019	0.00016	0.00021	0.0002	0.0002
β_{14}	0.0001	0.00009	0.00009	0.00002	0.000000	0.000000	0.00007	0	0.00007
β_{15}	0.0001	0.00008	0.00009	0.00015	0.00015	0.00015	0.00019	0.00021	0.0002
β_{16}	0.01016	0.03848	0.0187	0.00975	0.02233	0.01387	0.00925	0.0152	0.00675

Conclusion:

This paper presents significant findings that extend the concept of using EMD multi-scale data decomposition with penalized quantile regression in the field of time series analysis. This approach utilizes the time and frequency domains to determine the relationship among variables, thereby improving model selection. This study concludes that providing more stationary time series data improves predictive performance. This conclusion motivates us to make the original time series data more stationary, to enhance the prediction accuracy of the model selection. Based on the above analysis, this study has developed three new hybrid methods by combining EMD and penalized quantile regression, namely EMD-QR-Ridge, EMD-QR-LASSO, and EMD-QR-EN, to enhance prediction accuracy. Subsequently, these methods are applied to simulation and real-world datasets and compared with the traditional methods, where the simulation and empirical results indicate that the proposed methods effectively address issues stemming from nonstationary and nonlinear signals. This is achieved by utilizing the EMD algorithm, which decomposes the original nonstationary and nonlinear signals into orthogonal IMF components and one residual component.

Moreover, the proposed methods have been effective in identifying the most important decomposition components, which have the most substantial effects on the response variable, thereby enhancing the study of the relationship between these components and the response variable using the EMD method. The simulation and empirical results also demonstrate that the proposed methods, based on EMD, yield better predictive accuracy compared to other approaches. This suggests that EMD-based techniques are particularly effective for improving the accuracy of predictive models in the research context. Finally, the proposed methods produce a model free from multicollinearity among the decomposition components and have high prediction precision with lower prediction error (PE).

References

1. Alhamzawi, R., Yu, K., & Benoit, D. F. (2012). Bayesian adaptive Lasso quantile regression. Statistical Modelling, 12(3), 279–297. https://doi.org/10.1177/1471082X1101200304.
2. Ali, M., Deo, R. C., Maraseni, T., & Downs, N. J. (2019). Improving SPI-derived drought forecasts incorporating synoptic-scale climate indices in multi-phase multivariate empirical mode decomposition model hybridized with simulated annealing and kernel ridge regression algorithms. Journal of Hydrology, 576, 164–184. https://doi.org/10.1016/j.jhydrol.2019.06.032.
3. Al-Jawarneh, A. S., & Ismail, M. T. (2022). The adaptive LASSO regression and empirical mode decomposition algorithm for enhancing modelling accuracy. Communications in Statistics: Simulation and Computation. https://doi.org/10.1080/03610918.2022.2032154.
4. Al-Jawarneh, A. S., Ismail, M. T., & Awajan, A. M. (2021). Elastic Net Regression and Empirical Mode Decomposition for Enhancing the Accuracy of the Model Selection. International Journal of Mathematical, Engineering and Management Sciences, 6(2), 564–583. https://doi.org/10.33889/IJMEMS.2021.6.2.034.
5. Al-Jawarneh, A. S., & Ismail, Mohd. T. (2021). Elastic-Net Regression based on Empirical Mode Decomposition for Multivariate Predictors. Pertanika Journal of Science and Technology, 29(1), 199–215. https://doi.org/10.47836/pjst.29.1.11.
6. Burgette, L. F., Reiter, J. P., & Miranda, M. L. (2011). Exploratory quantile regression with many covariates: An application to adverse birth outcomes. Epidemiology, 22(6), 859–866. https://doi.org/10.1097/EDE.0b013e31822908b3.
7. Chan, C. T. (1995). Wavelet Basics. In The Journal of the Acoustical Society of America (Vol. 99, Issue 2). https://doi.org/10.1121/1.414569.
8. Chu, H., Wei, J., & Qiu, J. (2018). Monthly streamflow forecasting using EEMD-Lasso-DBN method based on multi-scale predictors selection. Water (Switzerland), 10(10). https://doi.org/10.3390/w10101486.
9. Faltermeier, R., Zeiler, A., Tome', A. M., Brawanski, A., & Lang, E. W. (2011). Weighted Sliding Empirical Mode Decomposition. Advances in Adaptive Data Analysis, 03(04), 509–526. https://doi.org/10.1142/s1793536911000891.
10. Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Applications to Nonorthogonal Problems. Technometrics, 12(1), 69–82. https://doi.org/10.1080/00401706.1970.10488635.
11. Hu, J., Chen, Y., Zhang, W., & Guo, X. (2021). Penalized high-dimensional M-quantile regression: From L1 to Lp optimization. Canadian Journal of Statistics, 49(3), 875–905. https://doi.org/10.1002/cjs.11597.
12. Huang, N. E. (2014). INTRODUCTION TO THE HILBERT–HUANG TRANSFORM AND ITS RELATED MATHEMATICAL PROBLEMS. In Singapore: World Scientific (pp. 1–26). https://doi.org/10.1142/9789814508247_0001.
13. Huang, N. E., & Wu, Z. (2008). A REVIEW ON HILBERT-HUANG TRANSFORM : METHOD AND ITS APPLICATIONS. 2007, 1–23. https://doi.org/10.1029/2007RG000228.1.INTRODUCTION.
14. Huang, Shen, Z., Long, S. R., Wu, M. C., Snin, H. H., Zheng, Q., Yen, N. C., Tung, C. C., & Liu, H. H. (1998). The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 454(1971), 903–995. https://doi.org/10.1098/rspa.1998.0193.
15. Jaber, A. M., Ismail, M. T., & Altaher, A. M. (2014). Empirical mode decomposition combined with local linear quantile regression for automatic boundary correction. Abstract and Applied Analysis, 2014(March). https://doi.org/10.1155/2014/731827.
16. Junior, P. O., Tiwari, A. K., Padhan, H., & Alagidede, I. (2020). Analysis of EEMD-based quantile-in-quantile approach on spot- futures prices of energy and precious metals in India. Resources Policy, 68(May), 101731. https://doi.org/10.1016/j.resourpol.2020.101731.
17. Khan, D. M., Yaqoob, A., Iqbal, N., Wahid, A., Khalil, U., Khan, M., Rahman, M. A. A., Mustafa, M. S., & Khan, Z. (2019). Variable Selection via SCAD-Penalized Quantile Regression for High-

Dimensional Count Data. IEEE Access, 7, 153205–153216. https://doi.org/10.1109/ACCESS.2019.2948278.

18. Koenker, R. (2005). Quantile regression. In Nature Methods (Vol. 16, Issue 6). https://doi.org/10.1038/s41592-019-0406-y.
19. Koenker, R., & Bassett, G. (1978). Regression Quantiles. Econometrica, 46(1), 33. https://doi.org/10.2307/1913643.
20. Li, Y., & Zhu, J. (2008). L1-norm quantile regression. Journal of Computational and Graphical Statistics, 17(1), 163–185. https://doi.org/10.1198/106186008X289155.
21. Masselot, P., Chebana, F., Bélanger, D., St-Hilaire, A., Abdous, B., Gosselin, P., & Ouarda, T. B. M. J. (2018). EMD-regression for modelling multi-scale relationships, and application to weather-related cardiovascular mortality. Science of the Total Environment, 612, 1018–1029. https://doi.org/10.1016/j.scitotenv.2017.08.276.
22. 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. https://doi.org/10.1080/03610918.2013.826361.
23. Sadig, A., & Bager, M. (2018). Ridge Parameter in Quantile Regression Models . An Application in Biostatistics. 8(2), 72–78. https://doi.org/10.5923/j.statistics.20180802.06.
24. SHEN, Z., FENG, N., & SHEN, Y. (2012). Ridge Regression Model-Based Ensemble Empirical Mode Decomposition for Ultrasound Clutter Rejection. Advances in Adaptive Data Analysis, 04(01n02), 1250013. https://doi.org/10.1142/s1793536912500136.
25. Tang, Q., Zhang, H., & Gong, S. (2020). Bayesian Regularized Quantile Regression Analysis Based on Asymmetric Laplace Distribution. Journal of Applied Mathematics and Physics, 08(01), 70–84. https://doi.org/10.4236/jamp.2020.81006.
26. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. In Notes and Queries (Vols. s3-V, Issue 128, p. 490). https://doi.org/10.1093/nq/s3-V.128.490-a.
27. Titchmarsh, E. C. (1948). Introduction to the Theory of Fourier Integrals. In Nature (Vol. 141, Issue 3561). https://doi.org/10.1038/141183a0.
28. Yan, A., & Song, F. (2019). Adaptive elastic net-penalized quantile regression for variable selection. Communications in Statistics - Theory and Methods, 48(20), 5106–5120. https://doi.org/10.1080/03610926.2018.1508711.
29. Zaikarina, H., Djuraidah, A., & Wigena, A. H. (2016). Lasso and ridge quantile regression using cross validation to estimate extreme rainfall. Global Journal of Pure and Applied Mathematics, 12(4), 3305–3314.
30. Zhang, L., Xie, L., Han, Q., Wang, Z., & Huang, C. (2020). Probability density forecasting of wind speed based on quantile regression and kernel density estimation. Energies, 13(22), 1–24. https://doi.org/10.3390/en13226125.
31. 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. https://doi.org/10.1111/j.1467-9868.2005.00503.x.