Research Article

Empirical Mode Decomposition Combined with Local Linear Quantile Regression for Automatic Boundary Correction

Abobaker M. Jaber,[1] Mohd Tahir Ismail,[1] and Alssaidi M. Altaher[2]

[1] *School of Mathematical Sciences, Universiti Sains Malaysia, 11800 Minden, Penang, Malaysia*
[2] *Statistics Department, Sebha University, Sebha 00218, Libya*

Correspondence should be addressed to Abobaker M. Jaber; jaber3t@yahoo.co.uk

Empirical mode decomposition (EMD) is particularly useful in analyzing nonstationary and nonlinear time series. However, only partial data within boundaries are available because of the bounded support of the underlying time series. Consequently, the application of EMD to finite time series data results in large biases at the edges by increasing the bias and creating artificial wiggles. This study introduces a new two-stage method to automatically decrease the boundary effects present in EMD. At the first stage, local polynomial quantile regression (LLQ) is applied to provide an efficient description of the corrupted and noisy data. The remaining series is assumed to be hidden in the residuals. Hence, EMD is applied to the residuals at the second stage. The final estimate is the summation of the fitting estimates from LLQ and EMD. Simulation was conducted to assess the practical performance of the proposed method. Results show that the proposed method is superior to classical EMD.

1. Introduction

We consider the following general nonparametric regression model:

$$y = f(x) + \varepsilon_i \cdots ,\tag{1}$$

where Y is the response variable, x is a covariate, $f(x) = E(y \mid x)$ is assumed to be a smooth nonparametric function, and ε_i represents independent and identical random errors with mean 0 and variance σ^2.

Empirical mode decomposition (EMD) is a form of analysis based on nonparametric methods [1]. This technique is particularly useful for analyzing nonlinear and nonstationary time series. This method has been widely applied over the last few years to analyze data in different disciplines, such as biology, finance, engineering, and climatology. EMD can enhance estimation performance. Applying the capabilities of EMD as a fully adaptive method and its advantages of handling nonlinear and nonstationary signal behaviors leads to better results. However, EMD suffers from boundary extension, curve fitting, and stopping criteria [2]. Such problems may

corrupt the entire data and result in a misleading conclusion [3]. Given that finite data are involved, the algorithms must be adjusted to use certain boundary conditions. In EMD, the end points are also considered problems. The influence of the end points propagates into the data range during sifting. Data extension (or data prediction) is a risky procedure for linear and stationary processes and is more difficult for nonlinear and nonstationary processes. The work in [1] indicated that only the values and locations of the next several extrema, and not all extended data, need to be predicted for EMD. Widely used approaches, such as the characteristic wave extending method, mirror extending method [4], data extending method [5], data reconstruction method [6], and similarity searching method [7], were proposed to overcome the problem and generate a more reasonable solution. The work in [8] introduced quantile regression, a significant extension of traditional parametric and nonparametric regression methods. Quantile regression has been largely used in statistics since its introduction because of its ease of interpretation, robustness, and numerous applications in important areas, such as medicine, economics, environment

modeling, toxicology, and engineering [9, 10]. A robust version of classical local linear regression (LLR) known as local linear quantile regression (LLQ) by [11, 12] respectively, have increasingly drawn interest. With its robust behavior, LLQ exhibits excellent boundary adjustment. This characteristic can more efficiently distinguish systematic differences in dispersion, tail behavior, and other features with respect to covariates [12, 13].

The current study aims to use the advantages of LLQ to automatically reduce the boundary effects of EMD instead of using classical boundary solutions mentioned previously. The proposed method consists of two stages that automatically decrease the boundary effects of EMD. At the first stage, LLQ is applied to the corrupted and noisy data. The remaining series is then expected to be hidden in the residuals. At the second stage, EMD is applied to the residuals. The final estimate is the summation of the fitting estimates from LLQ and EMD. Compared with EMD, this combination obtains more accurate estimates.

The remainder of this study is organized as follows. In Section 2, we present a brief background of EMD and LLQ. Section 3 introduces the proposed method. Section 4 compares the results of the original EMD algorithm and the proposed new boundary adjustment by simulation experiments. Conclusions are drawn in Section 5.

2. Background

2.1. History of Boundary Treatment in Nonparametric Estimators. Most nonparametric techniques such as kernel regression, wavelet thresholding, and empirical mode decomposition show a sharp increase in variance and bias at points near the boundary. Lots of works have been reported in the literature in order to reduce the effects of boundary problem. For kernel regression solution, see [14, 15]. For wavelet thresholding, in addition to use of periodic or symmetric assumption, the authors in [16, 17] used polynomial regression to improve the boundary problem. For empirical mode decomposition the authors in [18] provided a new idea about the boundary extension instead of using the traditional mirror extension on the boundary, and they proposed a ratio extension on boundary. The authors in [19] applied neural network to each IMF to restrain the end effect. The work in [2] provided an algorithm based on the sigma-pi neural network which is used to extend signals before applying EMD. The authors in [20] proposed a new approach that couples the mirror expansion with the extrapolation prediction of regression function to solve boundary problem. The algorithm includes two steps: the extrapolation of the signal through support vector (SV) regression at both endpoints to form the primary expansion signal, and then the primary signal is further expanded through extrema mirror expansion and EMD is performed on the resulting signal to obtain reduced end limitations.

In this paper we have followed [16] and [17] strategies to handle end effects of boundary problem in EMD. Instead of using classical polynomial nonparametric regression we will replace it by using a more robust nonparametric estimator, called local linear quantile regression. Practical justifications for choosing such estimator will be explained in Section 2.4.

2.2. Empirical Mode Decomposition (EMD). EMD [1] has proven to be a natural extension and an alternative technique to traditional methods for analyzing nonlinear and nonstationary signals, such as wavelet methods, Fourier methods, and empirical orthogonal functions [21]. In this section, we briefly describe the EMD algorithm. The main objective of EMD is to decompose the data y_t into small signals called intrinsic mode functions (IMF). An IMF is a function in which the upper and the lower envelopes are symmetric; in addition, the number of zero-crossings and the number of extremes are equal or differ by at most one [22]. The algorithm for extracting IMFs for a given time series y_t is called shifting and consists of the following steps.

(I) Setting initial estimates for the residue as $r_0(t) = y_t$, $g_0(t) = r_{k-1}(t)$, $i = 1$, and the index of IMF as $k = 1$.

(II) Constructing the lower minima $I_{\min_{i-1}}$ and the upper $I_{\max_{i-1}}$ envelopes of the signal by the cubic spline method.

(III) Computing the mean values, m_i, by averaging the upper envelope and the lower envelope as $m_{i-1} = [I_{\max_{i-1}} + I_{\min_{i-1}}]/2$.

(IV) Subtracting the mean from the original signal, that is, $g_i = g_{i-1} - m_{i-1}$ and $i = i+1$. Steps II to IV are repeated until g_i becomes an IMF. If so, the kth IMF is given by $\text{IMF}_K = g_i$.

(V) Updating the residue as $r_k = (t) = r_{k-1}(n) - \text{IMF}_K$. This residual component is treated as new data and subjected to the process described above to calculate the next IMF_{K+1}.

(VI) Repeating the steps above until the final residual component $r(x)$ becomes a monotonic function and then considering the final estimation of residue $\hat{r}(x)$.

Many methods have been presented to extract trends from a time series. Freehand and least squares methods are the commonly used techniques; the former depends on the experience of users, and the latter is difficult to use when the original series are very irregular [23]. EMD is another effective method for extracting trends [19].

2.3. Local Linear Quantile Regression (LLQ). The seminal study by [8] introduced the parametric quantile regression, which can be considered an alternative to classical regression in both parametric and nonparametric fields. Many models for the nonparametric approach, including locally polynomial quantile regression by [11] and kernel methods by [24], have since been introduced into the statistical literature. In this paper we adopt local linear regression (LLQ) introduced by [12].

Let $\{(x_i, y_i),\ i = 1, \ldots, n\}$ be bivariate observations. To estimate the τth conditional quantile function of response y, the equation below is defined given $X = x$:

$$g(x) = Q_y(\tau \mid x). \tag{2}$$

Let K be a positive symmetric unimodal kernel function and consider the following weighted quantile regression problem:

$$\min_{\beta \in \mathbb{R}^2} \sum_{i=1}^{n} w_i(x) \rho_\tau \left(y_i - \beta_0 - \beta_1 (x_i - x) \right), \qquad (3)$$

where $w_i(x) = k((x_i - x)/h)/h$. Once the covariate observations are centered at point x, the estimate of $g(x)$ is simply $\hat{\beta}_0$, which is the first component of the minimizer of (2). $\hat{\beta}_1$ determines an estimate of the slope of the function g at point x.

The higher-order LLQ estimate is the minimizer of the following:

$$\min_{\beta \in \mathbb{R}^2} \sum_{i=1}^{n} w_i(x) \rho_\tau \left(y_i - \beta_0 - \beta_1 (x_i - x) - \cdots - \beta_p (x_i - x)^p \right). \qquad (4)$$

The choice of the bandwidth parameter h significantly influences all nonparametric estimations. An excessively large h obscures too much local structure by excessive smoothing. Conversely, an excessively small h introduces too much variability by relying on very few observations in the local polynomial fitting [13].

2.4. Bandwidth Selection.
The practical performance of $\hat{Q}_\alpha(x)$ depends strongly on selected of bandwidth parameter. In this study we adopt the strategy of [12]. In summary, we have the automatic bandwidth selection strategy for smoothing conditional quantiles as follows.

(1) Use ready-made and sophisticated methods to select h_{mean}; we use the technique of [25].

(2) Use $h_\tau = h_{\text{mean}} \{ \tau(1 - \tau)/\phi(\Phi^{-1}(\tau))^2 \}^{1/5}$ to obtain all other h_τs from h_{mean}.

Here, ϕ and Φ are standard normal density and distribution function and h_{mean} is a bandwidth parameter for regression mean estimation with various existing methods. As it can be seen, this procedure leads to identical bandwidth for τ and $(1 - \tau)$ quantiles.

2.5. The Behavior of Local Linear Quantile Estimator at Boundary Region.
To examine the asymptotic the asymptotic behavior of the local linear quantile estimators at the boundaries, we offer this theorem which has been discussed in detail; see [26]. Here we omitted the proofs and summarized the key points. Without loss of generality, one can consider only the left boundary point $u_0 = ch$, $0 < c < 1$, if U_t takes value only from $[0, 1]$. However, a similar result holds for the right boundary point $u_0 = 1 - ch$.

Define

$$u_{j,c} = \int_{-c}^{1} u^j K(u)\, du, \qquad v_{j,c} = \int_{-c}^{1} u^j K^2(u)\, du. \qquad (5)$$

Theorem 1 (see [26]). *Consider the following assumptions.*

(1) $a(u)$ is twice continuously differentiable in a neighborhood of u_0 for any u_0.

(2) $f_u(u)$ is continuous and $f_u(u_0) > 0$.

(3) $f_{y|u,x}(y)$ is bounded and satisfies the Lipschitz condition.

(4) The kernel function $K(\cdot)$ is symmetric and has a compact support, say $[-1, 1]$.

(5) $\{(X_t, Y_t, U_t)\}$ is a strictly α-mixing stationary process with mixing coefficient which satisfies $\sum_{t \geq 1}^{\infty} t^l \alpha^{(\delta-2)/\delta}(t) < \infty$ for some positive real number $\delta > 2$ and $l > (\delta - 2)/\delta$.

(6) $E\|X_t\|^{2\delta} < \infty$ with $\delta^* > \delta$.

(7) $\Omega(u_0)$ is positive-definite and continuous in a neighborhood of u_0.

(8) $\Omega^*(u_0)$ is continuous and positive-definite in a neighborhood of u_0.

(9) The bandwidth h satisfies $h \to 0$ and $h \to \infty$.

(10) $f(u, v \mid x_0, x_s; s) \leq M < \infty$ for $s \geq 1$ where $f(u, v \mid x_0, x_s; s)$ is the conditional density of (U_0, U_s) given $(X_0 = x_0, X_s = x_s)$.

(11) $n^{1/2-\delta/4} h^{\delta/\delta^*-1/2-\delta/4} = O(1)$.

The asymptotic normality of the local linear quantile estimator at the left boundary point is given by

$$\sqrt{nh} \left[\hat{a}(ch) - a(ch) - \frac{h^2 b_c}{2} a''(0+) + op\left(h^2\right) \right]$$

$$\longrightarrow N \left\{ o, \tau(1 - \tau) v_c \sum_\alpha 0+ \right\}, \qquad (6)$$

where

$$b_c = \frac{u_{2,c}^2 - u_{1,c} u_{3,c}}{u_{2,c} u_{0,c} - u_{1,c}^2},$$

$$v_c = \frac{u_{2,c}^2 v_{0,c} - 2u_{1,c} u_{2,c} v_{1,c} + u_{1,c}^2 v_{2,c}}{\left[u_{2,c} u_{0,c} - u_{1,c}^2 \right]^2}. \qquad (7)$$

Further, the asymptotic normality of the local constant quantile estimator at the left boundary point $u_0 = ch$ for $0 < c < 1$ is

$$\sqrt{nh} \left[\tilde{a}(ch) - a(ch) - \tilde{b}_c + op\left(h^2\right) \right]$$

$$\longrightarrow N \left\{ 0, \tau(1 - \tau) v_{0,c} \sum_\alpha \frac{(0+)}{u_{0,c}^2} \right\}, \qquad (8)$$

where

$$\tilde{b}_c = \left(\left[h u_{1,c} a'(0+) + \frac{h^2 u_{2,c}}{2} \right] \right.$$

$$\times \left\{ a''(0+) + \frac{2a'(0+) f_u'(0+)}{f_u(0+)} \right.$$

$$+ 2\Omega^{*-1}(0+) \Omega^{*'}(0+) a'(0+) \left. \right\} \right])$$

$$\times (u_{0,c})^{-1}.$$

TABLE 1: Formula of the test functions used in the simulation.

Test function	Formula
1	$f(x) = \sin(\pi x) - \sin(2\pi x) + 0.5x$
2	$f(x) = \begin{array}{ll} 10e^{-10x} + 2 & if\ x \le 0.5 \\ 3\cos(10\pi x) & 0.5 < x < 1 \end{array}$
3	$f(x) = 5e^{(-10(x-0.5)^2)} + e^{(-10x)}$

From the above theorem, one can deduce that, at the boundaries, the asymptotic bias term for the local constant quantile estimate is of the order h, compared to the order h^2 for the local linear quantile estimate. Hence, the local linear estimation possesses good behavior at boundaries and there is no need for any boundary correction. In other words, the local linear quantile estimate does not suffer from boundary effects but the local constant quantile estimate does. Therefore, local linear quantile is preferable in practice.

3. Proposed Method

This section elaborates on the proposed method. This technique combines EMD and LLQ (EMD-LLQ). Since local linear quantile regression produces excellent boundary treatment [27], it is expected that the addition of this component to empirical mode decomposition will result in equally well-boundary properties. Results from numerical experiments extremely support this claim.

The basic idea behind the proposed method is to estimate the underlining function f with the sum of a set of EMD functions, f_{EMD}, and an LLQ function, f_{LLQ}. That is,

$$\widehat{f}_{LLQ\cdot EMD} = \widehat{f}_{EMD} + \widehat{f}_{LLQ}. \tag{10}$$

We need to estimate the two components f_{EMD} and f_{LLQ} to obtain our proposed estimate, $\widehat{f}_{EMD\cdot LLQ}$, by the following steps.

(1) Applying LLQ to the corrupted and noisy data, y_i and obtaining the trend estimate $\widehat{f}_{LLQ}$.

(2) Determining the residuals e_i from LLQ; that is, $e_i = y_i - \widehat{f}_{LLQ}$.

(3) Applying EMD to e_i, given that the remaining series is expected to be hidden in the residuals. This step is accomplished by performing the following substeps.

   (I) Setting initial estimates for the residue as $r_0(t) = e$, $g_o(t) = r_{k-1}(t)$, $i = 1$, and the index of IMF as $k = 1$.

   (II) Constructing the lower minima $I_{\min_{i-1}}$ and $I_{\max_{i-1}}$ envelopes of the signal by the cubic spline method.

   (III) Calculating the mean values by averaging the upper envelope and the lower envelope. Setting $m_{i-1} = [I_{\max_{i-1}} + I_{\min_{i-1}}]/2$.

   (IV) Subtracting the mean from the original signal as $g_i = g_{i-1} - m_{i-1}$ and $i = i + 1$. Steps I to IV are repeated until g_i becomes an IMF. The kth IMF is then given as $IMF_K = g_i$.
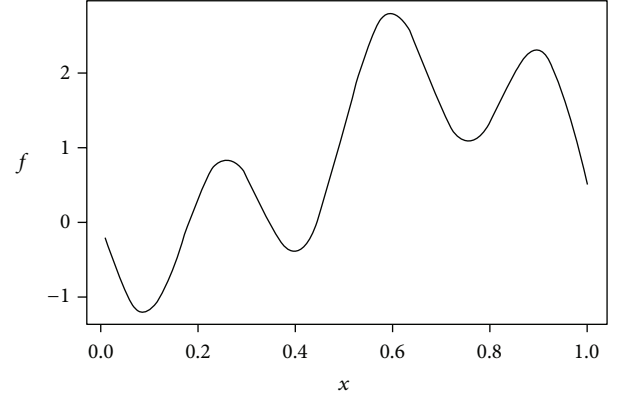
(a)
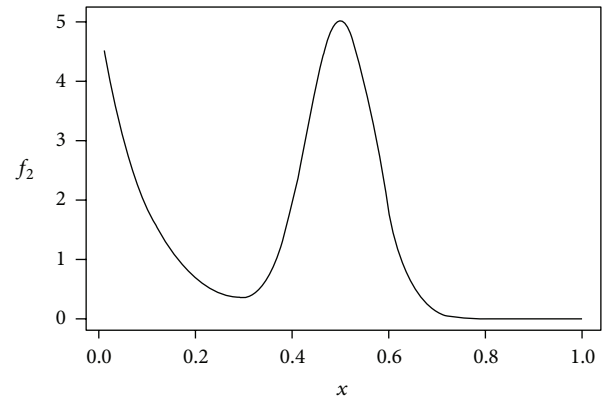
(b)
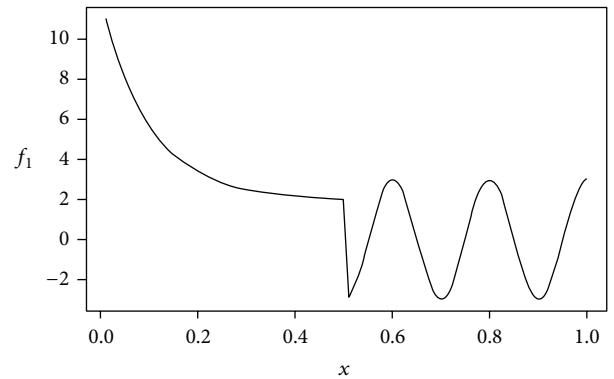
(c)

FIGURE 1: Three test functions used in the simulation.

   (V) Updating the residue $r_{k(x)} = r_{k-1}(n) - IMF_K$. This residual component is regarded as a new datum and is subjected to the process described above to calculate the next IMF_{K+1}.

   (VI) The steps above are repeated until the final residual component $r(x)$ becomes a monotonic function. The final estimation of the residue $\widehat{r}(x)$ is then considered.

TABLE 2: The MSE of the classical EMD and proposed method under variety of boundary solution noise structure, different values of quantile τ (0.25, 0.50, and 0.75), and sample size 100.

Test function 1								
Quantiles	$\tau = 0.25$			$\tau = 0.50$			$\tau = 0.75$		
Errors	$N{\sim}(0,1)$	$T(100,3)$	AR(0.5)	$N{\sim}(0,1)$	$T(100,3)$	AR(0.5)	$N{\sim}(0,1)$	$T(100,3)$	AR(0.5)
EMD (none)	0.269175	0.310785	0.32917	0.261690	0.3027736	0.3190445	0.2552796	0.306369	0.3223
EMD-LLQ	**0.11712**	**0.122119**	**0.145831**	**0.06343**	**0.058571**	**0.010130**	**0.070048**	**0.067575**	**0.10510**
Wilcoxon test $V =$	451927	459306	449112	49589	497928	480711	493926	494605	479475

Test function 1								
Quantiles	$\tau = 0.25$			$\tau = 0.50$			$\tau = 0.75$		
Errors	$N{\sim}(0,1)$	$T(100,3)$	AR(0.5)	$N{\sim}(0,1)$	$T(100,3)$	AR(0.5)	$N{\sim}(0,1)$	$T(100,3)$	AR(0.5)
EMD (periodic)	1.38123	1.40508	1.45777	1.42545	1.43629	1.47529	1.40223	1.41718	1.44426
EMD-LLQ	**0.11287**	**0.12451**	**0.14150**	**0.06375**	**0.05967**	**0.09826**	**0.06509**	**0.06284**	**0.10475**
Wilcoxon test $V =$	500496	50076	500427	500500	500076	500477	500493	500490	500476

P-value $< 2.2e^{-16}$

Test function 1								
Quantiles	$\tau = 0.25$			$\tau = 0.50$			$\tau = 0.75$		
Errors	$N{\sim}(0,1)$	$T(100,3)$	AR(0.5)	$N{\sim}(0,1)$	$T(100,3)$	AR(0.5)	$N{\sim}(0,1)$	$T(100,3)$	AR(0.5)
EMD (symmetric)	0.86598	0.91897	1.005607	0.87603	0.916301	1.005788	0.8820793	0.92558	0.96602
EMD-LLQ	**0.11656**	**0.12304**	**0.14698**	**0.8760**	**0.05816**	**0.09760**	**0.070185**	**0.06459**	**0.10498**
Wilcoxon test $V =$	500500	499701	500456	500498	499540	500497	500500	500497	500492

P-value $< 2.2e^{-16}$

Test function 1								
Quantiles	$\tau = 0.25$			$\tau = 0.50$			$\tau = 0.75$		
Errors	$N{\sim}(0,1)$	$T(100,3)$	AR(0.5)	errors	$N{\sim}(0,1)$	$T(100,3)$	AR(0.5)	Errors	$N{\sim}(0,1)$
EMD (wave)	1.20811	1.18084	1.250195	1.19274	1.20855	1.226365	1.199136	1.20274	1.22829
EMD-LLQ	**0.11318**	**0.11967**	**0.14845**	**0.06456**	**0.06173**	**0.098486**	**0.066864**	**0.06339**	**0.10583**
Wilcoxon test $V =$	500500	500500	500500	500500	500500	500500	500500	500500	500500

(4) The final estimate is the summation of the fitting estimates from LLQ and EMD, as follows:

$$\widehat{f}_{\text{LLQ·EMD}} = \widehat{f}_{\text{EMD}} + \widehat{r}(t). \tag{11}$$

4. Simulation Study

In this simulation, the software package R was employed to evaluate classical EMD by [1] and the proposed combined method, EMD-LLQ. The following conditions were set.

(1) Three different test functions (Table 1).

(2) Three different values of quantile τ (0.25, 0.50, and 0.75).

(3) Three different kinds of noise structure errors, namely:

    (a) normal distribution with zero mean and unity variance,

    (b) correlated noise from the first-order autoregressive model AR (1) with parameter (0.5),

    (c) heavy-tailed noise from t distribution with three degrees of freedom.

Datasets were simulated from each of the three test functions with a sample size of $n = 100$ (Figure 1). For each simulated dataset, the above two methods were applied to estimate the test function. In each case, 1,000 replications of the sample size $n = 100$ were made. The mean squared error (MSE) was used as the numerical measure to assess the quality of the estimate. The MSE was calculated for those observations that were at most 10 sample points away from the boundaries of the test functions:

$$\text{MSE}_{\Delta}\left(\widehat{f}\right) = \frac{1}{2\Delta} \sum_{i \in N(\Delta)} \left\{ f\left(x_i\right) - \widehat{f}\left(x_i\right)\right\}^2$$

$$\left(\Delta = 1, 2, \ldots, \left[\frac{n}{2}\right]; \ x_i = \frac{i}{n}\right), \tag{12}$$

where $N(\Delta) = \{1, \ldots, \Delta, n - \Delta + 1, \ldots, n\}$.

TABLE 3: The MSE of the classical EMD and proposed method under variety of boundary solution noise structure, different values of quantile τ (0.25, 0.50, and 0.75), and sample size = 100.

Test function 2								
Quantiles	$\tau = 0.25$			$\tau = 0.50$			$\tau = 0.75$		
Errors	$N\sim(0,1)$	$T(100,3)$	AR(0.5)	$N\sim(0,1)$	$T(100,3)$	AR(0.5)	$N\sim(0,1)$	$T(100,3)$	AR(0.5)
EMD (none)	13.71622	8.517261	7.290762	7.818736	8.094024	6.974746	7.168348	8.445187	7.364002
EMD-LLQ	**2.200657**	**2.148026**	**2.063819**	**2.147824**	**0.8982332**	**0.847309**	**1.68474**	**1.69877**	**1.57035**
Wilcoxon test $V =$	467459	463117	500260	499774	498875	489150	482567	483765	489150
P-value $< 2.2e^{-16}$									
Test function 2								
Quantiles	$\tau = 0.25$			$\tau = 0.50$			$\tau = 0.75$		
Errors	$N\sim(0,1)$	$T(100,3)$	AR(0.5)	$N\sim(0,1)$	$T(100,3)$	AR(0.5)	$N\sim(0,1)$	$T(100,3)$	AR(0.5)
EMD (periodic)	7.430924	6.976339	6.77826	7.135594	6.982454	6.827251	7.226472	6.865929	6.725184
EMD-LLQ	**2.118629**	**2.145495**	**2.052055**	**0.904618**	**0.9031342**	**0.8702984**	**1.672161**	**1.691775**	**1.598264**
Wilcoxon test $V =$	498562	497328	495251	500441	500350	500203	498068	496900	496766
P-value $< 2.2e^{-16}$									
Test function 2								
Quantiles	$\tau = 0.25$			$\tau = 0.50$			$\tau = 0.75$		
Errors	$N\sim(0,1)$	$T(100,3)$	AR(0.5)	$N\sim(0,1)$	$T(100,3)$	AR(0.5)	$N\sim(0,1)$	$T(100,3)$	AR(0.5)
EMD (symmetric)	8.693953	8.800117	8.678484	8.679718	8.902223	8.76185	8.623087	8.756977	8.718655
EMD-LLQ	**2.121942**	**2.142339**	**2.059669**	**0.918506**	**0.8823274**	**0.8637991**	**1.655741**	**1.670038**	**1.571081**
Wilcoxon test $V =$	500500	500500	500500	500500	500500	500500	500500	500500	500500
P-value $< 2.2e^{-16}$									
Test function 2								
Quantiles	$\tau = 0.25$			$\tau = 0.50$			$\tau = 0.75$		
Errors	$N\sim(0,1)$	$T(100,3)$	AR(0.5)	$N\sim(0,1)$	$T(100,3)$	AR(0.5)	$N\sim(0,1)$	$T(100,3)$	AR(0.5)
EMD (wave)	7.243865	7.493569	7.341059	7.138806	7.430882	7.449	7.124459	7.374468	7.302789
EMD-LLQ	**2.12519**	**2.149727**	**2.053073**	**0.908747**	**0.9056965**	**0.8560205**	**1.653102**	**1.704716**	**1.552656**
Wilcoxon test $V =$	500500	500500	500500	500500	500500	500500	500500	500497	500500
P-value $< 2.2e^{-16}$									

To compare the methods, Tables 2, 3, and 4 present the numerical results of the classical EMD with respect to the proposed method.

4.1. Results. From the simulation results, reported in Tables 2, 3, and 4, we have observed the following. Regardless of the boundary assumptions, test functions, noise structures, and different values of quantile, the proposed method is constantly superior to the classical EMD under periodic, symmetric (Mirror) and wave conditions. Tables 2, 3, and 4 summarize the results.

To ensure that the improvement in mean squared error is due to our proposed treatment, not to something else, we evaluated the classical method and our proposed one when no boundary treatment has been set up at all. From simulation result, we observed that even though the classical solutions help improve the mean squared error, our improvement is much better. Then, at the end, to get rid of some suspicions that the differences might not be significant, we used rank Wilcoxon test. This provided us evidence that our proposed method still achieves a better performance near the boundaries than EMD. All P value for Wilcoxon test are less than 0.05.

5. Conclusions

In this study, a new two-stage method is introduced to decrease the effects of the boundary problem in EMD. The proposed method is based on a coupling of LLQ at the first stage and classical EMD at the second stage. The empirical

TABLE 4: The MSE of the classical EMD and proposed method under variety of boundary solution noise structure, different values of quantile τ (0.25, 0.50, and 0.75), and sample size = 100.

	Test function 3								
Quantiles	$\tau = 0.25$			$\tau = 0.50$			$\tau = 0.75$		
Errors	$N\sim(0,1)$	$T(100,3)$	AR(0.5)	$N\sim(0,1)$	$T(100,3)$	AR(0.5)	$N\sim(0,1)$	$T(100,3)$	AR(0.5)
EMD (none)	0.034744	0.037490	0.06867	0.035888	0.035473	0.0708678	0.03540	0.0375	0.07260
EMD-LLQ	**0.01494**	**0.01827**	**0.04637**	**0.0142052**	**0.0168427**	**0.04288602**	**0.014913**	**0.0197**	**0.048020**
Wilcoxon test $V =$	453610	426890	371062	740664	435835	415746	455217	420619	393951
	P-value $< 2.2e^{-16}$								
	Test function 3								
Quantiles	$\tau = 0.25$			$\tau = 0.50$			$\tau = 0.75$		
Errors	$N\sim(0,1)$	$T(100,3)$	AR(0.5)	$N\sim(0,1)$	$T(100,3)$	AR(0.5)	$N\sim(0,1)$	$T(100,3)$	AR(0.5)
EMD (periodic)	0.01495	0.01601	1.46184	0.014931	0.017672	1.475389	0.01513	0.0173	1.44361
EMD-LLQ	**0.01201**	**0.01425**	**0.14555**	**0.010031**	**0.016859**	**0.0967748**	**0.01176**	**0.0157**	**0.10338**
Wilcoxon test $V =$	142798	167295	500426	159630	193475	500491	133946	170435	500474
	P-value $< 2.2e^{-16}$								
	Test function 3								
Quantiles	$\tau = 0.25$			$\tau = 0.50$			$\tau=0.75$		
Errors	$N\sim(0,1)$	$T(100,3)$	AR(0.5)	$N\sim(0,1)$	$T(100,3)$	AR(0.5)	$N\sim(0,1)$	$T(100,3)$	AR(0.5)
EMD (symmetric)	0.856264	0.938382	1.01019	0.853808	0.928291	0.9873542	0.86854	0.9322	0.96783
EMD-LLQ	**0.11334**	**0.12452**	**0.14649**	**0.065446**	**0.056100**	**0.09569597**	**0.06553**	**0.06180**	**0.104910**
Wilcoxon-test $V =$	500489	499304	500420	500500	500498	500474	500500	500485	500463
	P-value $< 2.2e^{-16}$								
	Test function 3								
Quantiles	$\tau = 0.25$			$\tau = 0.50$			$\tau = 0.75$		
Errors	$N\sim(0,1)$	$T(100,3)$	AR(0.5)	$N\sim(0,1)$	$T(100,3)$	AR(0.5)	$N\sim(0,1)$	$T(100,3)$	AR(0.5)
EMD (wave)	1.19549	1.19853	1.22884	1.180275	1.200136	1.217799	1.193706	1.20841	1.236941
EMD-LLQ	**0.11508**	**0.12732**	**0.14624**	**0.063552**	**0.055484**	**0.094623**	**0.06860**	**0.20362**	**0.110907**
Wilcoxon test $V =$	500500	500500	500500	500500	500500	500500	500500	499500	500500
	P-value $< 2.2e^{-16}$								

performance of the proposed method was tested on different numerical experiments by simulation and real data application. The results of these experiments illustrate the improvement of the EMD estimation in terms of MSE.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

References

[1] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society A*, vol. 454, no. 1971, pp. 903–995, 1998.

[2] Z. Liu, "A novel boundary extension approach for empirical mode decomposition," in *Intelligent Computing*, vol. 4113 of *Lecture Notes in Computer Science*, pp. 299–304, Springer, Berlin, Germany, 2006.

[3] W. Wang, X. Li, and R. Zhang, "Boundary processing of HHT using support vector regression machines," in *Computational Science—ICCS 2007*, vol. 4489 of *Lecture Notes in Computer Science*, pp. 174–177, Springer, Berlin, Germany, 2007.

[4] J. Zhao and D. Huang, "Mirror extending and circular spline function for empirical mode decomposition method," *Journal of Zhejiang University Science*, vol. 2, no. 3, pp. 247–252, 2001.

[5] K. Zeng and M.-X. He, "A simple boundary process technique for empirical mode decomposition," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS '04)*, pp. 4258–4261, September 2004.

[6] Z. Zhao and Y. Wang, "A new method for processing end effect in empirical mode decomposition," in *Proceedings of the International Conference on Communications, Circuits and Systems (ICCCAS '07)*, pp. 841–845, July 2007.

[7] J. Wang, Y. Peng, and X. Peng, "Similarity searching based boundary effect processing method for empirical mode decomposition," *Electronics Letters*, vol. 43, no. 1, pp. 58–59, 2007.

[8] R. Koenker and G. Bassett Jr., "Regression quantiles," *Econometrica*, vol. 46, no. 1, pp. 33–50, 1978.

[9] M. Buchinsky, "Quantile regression, Box-Cox transformation model, and the U.S. wage structure, 1963–1987," *Journal of Econometrics*, vol. 65, no. 1, pp. 109–154, 1995.

[10] Y. Wei, A. Pere, R. Koenker, and X. He, "Quantile regression methods for reference growth charts," *Statistics in Medicine*, vol. 25, no. 8, pp. 1369–1382, 2006.

[11] P. Chaudhuri, "Nonparametric estimates of regression quantiles and their local Bahadur representation," *The Annals of Statistics*, vol. 19, no. 2, pp. 760–777, 1991.

[12] K. Yu and M. C. Jones, "Local linear quantile regression," *Journal of the American Statistical Association*, vol. 93, no. 441, pp. 228–237, 1998.

[13] R. Koenker, *Quantile Regression*, John Wiley & Sons, New York, NY, USA, 2005.

[14] H.-G. Müller and J.-L. Wang, "Hazard rate estimation under random censoring with varying kernels and bandwidths," *Biometrics*, vol. 50, no. 1, pp. 61–76, 1994.

[15] H.-G. Müller, "Smooth optimum kernel estimators near endpoints," *Biometrika*, vol. 78, no. 3, pp. 521–530, 1991.

[16] H.-S. Oh, P. Naveau, and G. Lee, "Polynomial boundary treatment for wavelet regression," *Biometrika*, vol. 88, no. 1, pp. 291–298, 2001.

[17] H.-S. Oh and T. C. M. Lee, "Hybrid local polynomial wavelet shrinkage: wavelet regression with automatic boundary adjustment," *Computational Statistics & Data Analysis*, vol. 48, no. 4, pp. 809–819, 2005.

[18] Q. Wu and S. D. Riemenschneider, "Boundary extension and stop criteria for empirical mode decomposition," *Advances in Adaptive Data Analysis*, vol. 2, no. 2, pp. 157–169, 2010.

[19] Y. Deng, W. Wang, C. Qian, Z. Wang, and D. Dai, "Boundary-processing-technique in EMD method and Hilbert transform," *Chinese Science Bulletin*, vol. 46, no. 11, pp. 954–961, 2001.

[20] D.-C. Lin, Z.-L. Guo, F.-P. An, and F.-L. Zeng, "Elimination of end effects in empirical mode decomposition by mirror image coupled with support vector regression," *Mechanical Systems and Signal Processing*, vol. 31, pp. 13–28, 2012.

[21] C. D. Blakely, *A Fast Empirical Mode Decomposition Technique for Nonstationary Nonlinear Time Series*, vol. 3, Elsevier Science, New York, NY, USA, 2005.

[22] A. Amar and Z. El abidine Guennoun, "Contribution of wavelet transformation and empirical mode decomposition to measurement of US core inflation," *Applied Mathematical Sciences*, vol. 6, no. 135, pp. 6739–6752, 2012.

[23] Y. Fan, J. W. Zhi, and S. L. Yuan, "Improvement in time-series trend analysis," *Computer Technology and Development*, vol. 16, pp. 82–84, 2006.

[24] R. Koenker, P. Ng, and S. Portnoy, "Quantile smoothing splines," *Biometrika*, vol. 81, no. 4, pp. 673–680, 1994.

[25] D. Ruppert, S. J. Sheather, and M. P. Wand, "An effective bandwidth selector for local least squares regression," *Journal of the American Statistical Association*, vol. 90, pp. 1257–1270, 1995.

[26] X. Xu, *Semiparametric quantile dynamic time series models and their applications [Ph.D. thesis]*, University of North Carolina at Charlotte, Charlotte, NC, USA, 2005.

[27] Z. Cai and X. Xu, "Nonparametric quantile estimations for dynamic smooth coefficient models," *Journal of the American Statistical Association*, vol. 103, no. 484, pp. 1595–1608, 2008.