Wavelet regression Combined with Local Linear Quantile Regression for Automatic Boundary Correction

M. A. Ghazal[1], W. Alabeid[2], Gh. Alshreef[3]

[1]Professor, Dept. of Mathematics, Faculty of Science, Damietta University, Damietta, Egypt
[2] Students, Dept. of Mathematics, Faculty of Science, Damietta University, Damietta, Egypt
[3]Dr, Dept. of Mathematics, Faculty of Science, Damietta University, Damietta, Egypt

---

Abstract - *Classical wavelet thresholding methods (WTM) is used to analyze both the non-stationary and nonlinear time series. Yet, the bounded support of underlain time series limited the availability of the partial data within its boundaries. In addition, large biases at the edges was occurred by increasing the bias when a time series data is defined in (WTM) which also result in creating artificial wggles. This study suggests a new two-stage method to concurrently minimize the effect of the boundaries found in (WTM). Local Linear Quantile Regression (LLQ) is applied in early stage in order to provide more accurate description of damaged or noisy data. However, it is assumed that there will be a remaining series hidden in the residuals. At second stage (WTM) has been applied to the residuals. The final stage is the summation of the fitting estimated from both LLQ and (WTM). To assess the practical performance of the proposed method a simulation was run which shows that the optimized WTM could overcome the classical method used in non-stationary and nonlinear time series analysis.*

***Key Words*: Local Linear Quantile Regression, Wavelet Thresholding Methods, Bandwidth Selection, Non-stationary and Nonlinear Time Series Analysis, Wavelet Thresholding Methods - Local Linear Quantile Regression (WTM-LLQ).**

1. INTRODUCTION

Kernel smoothers, trigonometric regressions are among those most nonparametric smoothing approaches to overcome boundaries problems when dealing with smooth functions as suggested by [1]. They believe that polynomial-trigonometric regression where ƒ is the estimator of a sum of trigonometric functions in a low-order Polynomial. The latter is expected to account for the boundary problem. They also gave an examples were provided to manifest the approach and that showed the convergence rates of the estimators over a specific smoothness class of functions were optimal irrespective of the regression function is being periodic or not. [2] Have used a de-noising approach of two-stage robust based on wavelet thresholding with median filter. They study the corrupted data passed the median filter at the earliest stage, where the outliers are restrained. Then, they obtained an ultimate restructured signal when the data is recoiled. The boundary is a key issue in classical wavelet thresholding method which is occurred during the transformation of the wavelet to a finite signal. Recently, [3] introduced a new method based on the concept of pseudo data and later developed by them in 2009[4] to include the robust thresholding within robust estimation. The practical performance of this method can be improved by considering the automatic boundary treatment suggested by [5] and [6, 7].

2. Wavelet thresholding methods (WTM)

The term wavelets is used to refer to a set of orthonormal basis functions generated by dilation and translation of a compactly supported scaling function (or father wavelet), ∅, and a mother wavelet, φ, associated with an r-regular multiresolution analysis of $L^2(R)$. A variety of different wavelet families now exist that combine compact support with various degrees of smoothness and numbers of vanishing moments[8] and these are now the most intensively used wavelet families in practical applications in statistics. Any square integrable function ƒ admits the following expansion:

$$f_w(x) = \sum_{k=-\infty}^{\infty} C_{0,K}\,\phi_K(x) + \sum_{j=0}^{\infty}\sum_{k=-\infty}^{\infty} d_{j,k}\varphi_{j,k}(x) \qquad (1)$$

Where

$$\phi_k(x) = 2^{1/2}\phi(2x-k)$$
$$\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x-k)$$

The scaling and detail coefficients

$$C_{0,K} = \int_{-\infty}^{\infty} f(x)\phi_k(x)dx \quad \text{and}$$

$$d_{j,k} = \int_{-\infty}^{\infty} f(x)\varphi_{j,k}(x)dx \quad .$$

Equation (1) suggests the following classical nonlinear wavelet regression estimator:

$$\hat{f}_w(x) = \sum_{k=-\infty}^{\infty} \hat{C}_{0,K}\,\phi_K(x) + \sum_{j=0}^{\infty}\sum_{k=-\infty}^{\infty} \hat{d}^{S}_{j,k}\varphi_{j,k}(x) \qquad (2),$$

Where

$$\hat{c}_{0,k} = \sum_i y_i \phi_k (i/n), \hat{d}_{j,k} = \sum_i y_i \varphi_{j,k} (i/n)$$

are respectively the empirical scaling and detail coefficients and $\hat{d}^s_{j,k} = \text{sgn}(\hat{d}_{j,k}) \max(0, |\hat{d}_{j,k}| - \lambda)$.

Sometimes the soft-thresholded coefficients $\hat{d}^s_{j,k}$ are replaced by the hard-thresholded coefficients

$\hat{d}^H_{J,K} = \hat{d}_{j,k} I(|\hat{d}_{j,k}| > \lambda)$.

3. Local Linear Quantile (LLQ) Regression

The seminal study of [9] introduced parametric quantile regression, which is considered an alternative to the classical regression in both parametric and nonparametric fields. Numerous models for the nonparametric approach have been introduced in statistical literature, such as the locally polynomial quantile regression by [10] and the kernel methods by [11]. In this paper, we adopt the LLQ regression employed by [12]. Let $\{(x_i, y_i), i = 1....., n\}$ be bivariate observations.

To estimate the τth conditional quantile function of response y and $\tau \in [0,1]$, the equation below is defined given $X = x$:

$$g(x) = Q_Y(\tau / x) \qquad (1)$$

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

$$\frac{min}{\beta \in i^2} \sum_{i=1}^{n} w_i(x) \rho_\tau (yi - \beta_0 - \beta_1 (x_i - x)), \qquad (2)$$
$$w(x) = k((x_i - x)/h)/h$$

where $w(x) = k((x_i - x)/h)/h$. Once the covariate observations are centered at point, the estimate of $g(x)$ is simply β_0 , which is the first component of the minimizer of (1), and determines the estimate of the slope of the function g at point x .

3.1 Bandwidth Selection

The practical performance of $\hat{Q}_\tau(\text{x})$ 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) - We use ready-made and sophisticated methods in selecting h_{mean} ; we employ [13] which explored a "direct plugin" bandwidth selection procedure which relies on asymptotically optimal bandwidth:

$$h_{mean} = \left[\sigma^2 R(k)(b-a)/n\mu_2^2 \int m''^{2}_{(x)} p_{(x)} d(x)^{1/5} \right]$$

$$= c_1(k)\{\frac{\sigma^2(b-a)}{n\theta_{22}}\}^{1/5}$$

$$h_\tau = h_{mean} \left\{ \tau(1-\tau)(\Big/ \varnothing(\Phi^{-1}(\tau))^2 \right\}^{1/5}$$

(2) - We use to obtain all of the other h_τ from h_{mean} . ϕ and Φ are standard normal density and distribution functions, and h_{mean} is a bandwidth parameter for regression mean estimation with various existing methods. This procedure obtains identical bandwidths for the τ and (1 −τ) quantiles.

3. Proposed Method

This section elaborates on the proposed method. This technique combines WTM and LLQ (WTM-LLQ). Since local linear quantile regression produces excellent boundary treatment [14], it is expected that the addition of this component to Wavelet thresholding methods 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 WTM functions, f_{WTM}, and an LLQ function, f_{LLQ}. That is,

$$\hat{f}_{W.LLQ} = \hat{f}_{\text{WTM}} + \hat{f}_{LLQ}$$

To obtain the wavelet regression - Local Linear Quantile Regression estimate $\hat{f}_{W.LLQ}$ we need to estimate the two components: $\hat{f}_w$ and $\hat{f}_{LLQ}$. Inspired by the back-fitting algorithm of [15], we propose the following iterative algorithm for computing $\hat{f}_w$, $\hat{f}_{LLQ}$ and hence $\hat{f}_{W.LLQ}$.

1- Obtain an initial estimate $\hat{f}^0$ or f , and set $\hat{f}^0_{LLQ} = \hat{f}^0$

2- For j=1,....., iterate the following steps:

(a) Apply wavelet thresholding to $y_i - \hat{f}^{J-1}_{LLQ}$ and obtain $\hat{f}^J_{\text{WTM}}$.

(b) Estimate $\hat{f}^J_{LLQ}$ by fitting local quantile regression to $y_i - \hat{f}^J_{\text{WTM}}$

3- Stop if $\hat{f}_{W.LLQ} = \hat{f}_{\text{WTM}} + \hat{f}_{LLQ}$ converges.

To use the above algorithm, one needs to choose the initial curve estimate $\hat{f}^{0}$ in Step 1 and the smoothing parameter for the local quantile fit $\hat{f}^{J}_{LLQ}$ in Step 2(b). For computing $\hat{f}^{0}$, we use high-level statistical or mathematical software packages (R).

4. Simulation Study

In this simulation, the software package **R** was employed to evaluate classical WTM, and the proposed combined method, WTM-LLQ. The following conditions were set.

(1) Eight different test functions.
(2) Three different values of quantile τ (0.25, 0.50, and 0.75).
(3) Three different samples size (64,128, and 512).

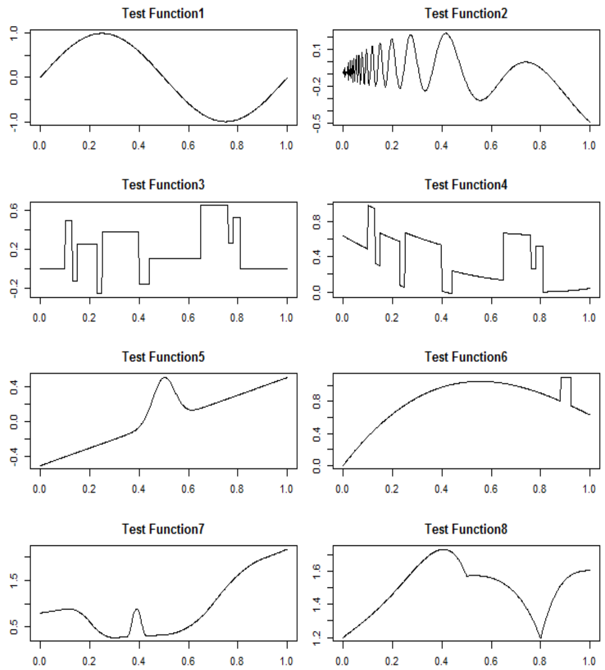

Fig-1: The eight test functions used in the simulation.

From fig1 we observed Each function has some abrupt changing features such as discontinuities or sharp bumps, Note that it is reasonable to assume a periodic boundary condition for Test Function 1,and 2, while for Test Functions 3,4,5,6,7 and 8, some boundary adjustment is strongly preferred, Datasets were simulated from each of the test functions with a Three samples size of n = 64,128 and 512. And level of signal-to-noise ratio (snr) were chosen: snr = 2, where snr is defined as $snr = f / \sigma$, For each simulated dataset, the above two methods were applied to estimate the test function. In each case, 500 replications of the Three samples size n = 64,128 and 512 were made. Throughout the whole simulation and for all the above two methods, the empirical Bayes procedure EBayes Thresh of [16] was used as the thresholding rule. 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:

$$MSE_{\ell}\left(\hat{f}\right)=\frac{1}{2\ell}\sum_{i\in N(\ell)}\left\{f\left(x_i\right)-\hat{f}\left(x_i\right)\right\}^2$$

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

Where

$$N\left(\ell\right)=\left\{1,\ldots,\ell,n-\ell+1,\ldots,n\right\}.$$ (See [17])

To gain an idea of how the two methods perform near the boundaries, we calculated the following MSE values for observations near the boundaries. (See table 1, 2, and 3).

Table -1: The MSE of the classical (WTM) and proposed method under different values of quantile τ (0.25, 0.50, and 0.75), and sample size 64

N=64						
Test functions	$\tau=0.75$		$\tau=0.50$		$\tau=0.25$	
	Mse (WTM-LLQ)	Mse (WTM)	Mse (WTM-LLQ)	Mse (WTM)	Mse (WTM-LLQ)	Mse (WTM)
Test function1	0.008062	0.041797	0.008536	0.055567	0.008927	0.038068
Test function2	0.004113	0.011419	0.005595	0.013284	0.004483	0.01865
Test function3	0.007669	0.004318	0.004826	0.004321	0.010852	0.002951
Test function4	0.056252	0.188786	0.030474	0.296407	0.058852	0.16709
Test function5	0.01291	0.04047	0.005344	0.080712	0.007155	0.05906
Test function6	0.011173	0.043078	0.006224	0.035673	0.020049	0.044433
Test function7	0.071131	0.137736	0.06412	0.132807	0.043316	0.195941
Test function8	0.002528	0.025158	0.001327	0.017944	0.00254	0.027108

Table -2: The MSE of the classical (WTM) and proposed method under different values of quantile τ (0.25, 0.50, and 0.75), and sample size 128

N=128						
Test functions	$\tau=0.75$		$\tau=0.50$		$\tau=0.25$	
	Mse (WTM-LLQ)	Mse (WTM)	Mse (WTM-LLQ)	Mse (WTM)	Mse (WTM-LLQ)	Mse (WTM)
Test function1	0.001319	0.045048	0.006085	0.022437	0.005927	0.027282
Test function2	0.003127	0.014359	0.004223	0.01768	0.004024	0.017169
Test function3	0.003644	0.001353	0.004495	0.00256	0.007799	0.004392
Test function4	0.019152	0.139156	0.025233	0.151319	0.005915	0.161533
Test function5	0.006106	0.029358	0.003535	0.034583	0.004296	0.060219
Test function6	0.004154	0.027534	0.003111	0.025203	0.007202	0.023229

Test function7	0.030169	0.122704	0.025159	0.142712	0.0412	0.144393
Test function8	0.000563	0.012015	0.000709	0.01128	0.001354	0.008262

Table -3: The MSE of the classical (WTM) and proposed method under different values of quantile τ (0.25, 0.50, and 0.75), and sample size 512

N=512						
Test functions	$\tau = 0.75$		$\tau = 0.50$		$\tau = 0.25$	
	Mse (WTM-LLQ)	Mse (WTM)	Mse (WTM-LLQ)	Mse (WTM)	Mse (WTM-LLQ)	Mse (WTM)
Test function1	0.000754	0.018025	0.000762	0.013343	0.000708	0.013169
Test function2	0.002472	0.007252	0.002364	0.013253	0.002368	0.014321
Test function3	0.002043	0.002233	0.001499	0.001452	0.001719	0.001636
Test function4	0.005368	0.049901	0.002768	0.054339	0.000798	0.046979
Test function5	0.001743	0.022847	0.001568	0.011604	0.00190	0.019046
Test function6	0.001523	0.017428	0.001791	0.013981	0.001904	0.00943
Test function7	0.00888	0.066047	0.005955	0.05849	0.009245	0.042798
Test function8	0.000685	0.008978	0.000342	0.011871	0.000538	0.013628

To compare the methods, Tables 1, 2, and 3 present the numerical results of the classical (WTM) with respect to the proposed method.

Results

Tables 1, 2, and 3indicated that proposed method performs better in terms of boundary assumptions, noise structures and test functions, under different samples size and values of quantile than the classical WTM.

It is therefore understood that Wavelet Thresholding Methods - local linear quantile regression WTM-LLQ is highly recommended in handling issues of boundaries in wavelet regressions.

5. Conclusions

This paper suggested a two-stage method to minimize issues in classical WTM boundaries. Coupling LLQ at early stages followed by classical WTM combination is the key design of the proposed method. A simulation was carried out to test the empirical performance of the method through different numerical experiments and real data application. Findings from this research indicated that WTM can be improved and resolve issues of boundaries and enhance MSE values.

REFERENCES

[1] Eubank & Speckman, "Convergence rates for trigonometric and polynomial-trigonometric regression estimators", Statist & Probability Letters, Vol. 11, Issue 2,1991a, pp.119-124.

[2] G. Wei, T. Jian, and R. Qiu, "Signal denoising via median filter & wavelet shrinkage based on heavytailed modeling," in *Proceedings of the 8th International Conference on Signal Processing*, vol. 1, 2006.

[3] H.-S. Oh, D. W. Nychka, and T. C. M. Lee,"The role of pseudo data for robust smoothing with application to wavelet regression, " *Biometrika*, vol. 94, no. 4, 2007, pp. 893–904.

[4] H.-S. Oh, D. Kim, and Y. Kim, "Robust wavelet shrinkage using robust selection of thresholds, " *Statistics and Computing*, vol. 19, no. 1, 2009, pp. 27–34.

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

[6] 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, 2005, pp. 809–819.

[7] A. M. Altaher and M. T. Ismail,"Robust estimation for boundary correction in wavelet regression, " *Journal of Statistical Computation and Simulation*, vol. 82, 2012, pp. 1531–1544.

[8] I.Daubechies,"*Ten Lectures on Wavelets*, "Philadelphia: SIAM, 1992.

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

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

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

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

[13] 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, 1995, pp. 1257–1270.

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

[15] T.J.Hastie and R.J. Tibshirani, "Generalized Additive Models. Generalized Additive Models, " Chapman & Hall, London, 1990.

[16] I.M .Johnstone, B.W.Silverman," Empirical Bayes selection of wavelet thresholds, " unpublished manuscript, 2003.

[17] A. M. Jaber, M. T. Ismail, and A. M. Altaher , "Empirical mode decomposition combined with local linear quantile regression for automatic boundary correction," Hindawi Publishing Corporation Abstract and Applied Analysis, vol. 2014, Article ID 731827, 8 pages, 2014.