Hindawi Publishing Corporation
The Scientific World Journal
Volume 2014, Article ID 708918, 5 pages
http://dx.doi.org/10.1155/2014/708918

Research Article

Application of Empirical Mode Decomposition with Local Linear Quantile Regression in Financial Time Series Forecasting

Abobaker M. Jaber,[1] Mohd Tahir Ismail,[1] and Alsaidi 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

Received 15 January 2014; Revised 23 June 2014; Accepted 25 June 2014; Published 22 July 2014

Academic Editor: Mohamed Hanafi

Copyright © 2014 Abobaker M. Jaber et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper mainly forecasts the daily closing price of stock markets. We propose a two-stage technique that combines the empirical mode decomposition (EMD) with nonparametric methods of local linear quantile (LLQ). We use the proposed technique, EMD-LLQ, to forecast two stock index time series. Detailed experiments are implemented for the proposed method, in which EMD-LPQ, EMD, and Holt-Winter methods are compared. The proposed EMD-LPQ model is determined to be superior to the EMD and Holt-Winter methods in predicting the stock closing prices.

1. Introduction

Recent studies have indicated that financial markets typically follow nonlinear and nonstationary behavior. Therefore, forecasting finance using classical techniques is quite difficult. The empirical mode decomposition (EMD) explored by Huang et al. [1] is a very powerful tool in modern quantitative finance and has emerged as a powerful statistical modeling technique [2, 3]. The capacity of the EMD to handle nonlinear and nonstationary behaviors has provided both researchers and practitioners with an attractive alternative tool. The EMD can explain the generation of time series data from an alternative perspective by breaking up time series signals into smaller numbers of independent and concretely implicational intrinsic modes based on scale separation. This distinguishing feature makes the EMD a valuable and desirable tool for forecasting financial time series signals [4]. The current study aims to extract and forecast the trend of two stock markets, namely, the Kuala Lumpur Bursa (KLSE) index and the New Zealand stock market index (NZX50), using the advantages of local linear quantile (LLQ) regression. The proposed method consists of two stages. In the first stage, LLQ is applied to corrupt and noisy data. The remaining series is subsequently expected to be hidden in the residuals. In the second stage, EMD is applied to the residuals. The final estimate is the summation of the fitting estimates from the LLQ and EMD. To extract and forecast the trend using EMD-LLQ and EMD, we summarize the steps as follows. (1) A signal is decomposed by the EMD-LLQ and EMD. (2) Meaningful intrinsic mode functions (IMFs) (components) are selected using the fast Fourier transform (FFT) (see [5]). (3) Selected components are added to the residue to obtain the trend. (4) The Holt-Winter method is based on the selected components and provides the forecasting results.

The remainder of this paper is organized as follows. Section 2 presents a brief background of the EMD and LLQ. Section 3 introduces the proposed method. Section 4 compares the results of the original EMD algorithm and the new proposed method by forecasting the daily closing prices of two stock markets, namely, the KLSE and the NZX50. Section 5 concludes.

2. Why EMD-LPQ (Empirical Mode Decomposition Combined with Local Linear Quantile Regression)?

Due to the edged effects, nonparametric techniques such as empirical mode decomposition show a sharp increase in

variance and bias at points near the boundary. The presence of such problem has dramatic effects on results. Varieties of works have been reported in literature in order to reduce the effects of boundary problem in traditional EMD. Two-stage methods or coupling methods nowadays have been used extensively for solving such problem; for instance [6], applied neural network to each IMF to restrain the end effect [7] provided an algorithm based on the sigma-pi neural network which is used to extend signals before applying EMD. Reference [4] proposed a new two-stage algorithm. 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. All previous methods have been shown to have good solutions to the end point and achieved a higher precision in an application part as well. In this paper we have followed [8]. The proposed method EMD-LPQ (empirical mode decomposition combined with local linear quantile regression) is designed to be a robust version of classical empirical mode decomposition especially in presence of edge effect problems.

3. Empirical Mode Decomposition

The EMD [1] has proven to be a natural extension of and an alternative technique to traditional methods for analyzing nonlinear and nonstationary signals, such as wavelet methods, Fourier methods, and empirical orthogonal functions [9]. In this section, we briefly describe the EMD algorithm. The EMD mainly decomposes the data y_t into smaller signals called IMFs. An IMF is a function in which the upper and the lower envelopes are symmetric. Moreover, the number of zero-crossings and the numbers of extremes are equal or differ by one, at the most [10]. The algorithm for extracting IMFs for a given time series y_t is called shifting, and it consists of the following steps.

(i) The initial estimates for the residue are set at $r_0(t) = y_t$, $g_0(t) = r_{k-1}(t)$, and $i = 1$, and the IMF index is set at $k = 1$.

(ii) The lower minima $I_{\min_{i-1}}$ and the upper $I_{\max_{i-1}}$ envelopes of the signal are constructedusing the cubic spline method.

(iii) The mean values m_i are computed by averaging the upper and lower envelopes as $m_{i-1} = [I_{\max_{i-1}} + I_{\min_{i-1}}]/2$.

(iv) The mean is subtracted from the original signal; that is, $g_i = g_{i-1} - m_{i-1}$ and $i = i + 1$. Steps (i) to (iv) are repeated until g_i becomes an IMF. Hence, the kth IMF is given by $\mathrm{IMF}_K = g_i$.

(v) The residue is updated via $r_k(t) = r_{k-1}(n) - \mathrm{IMF}_K$. This residual component is treatedas new data and subjected to the previously described process to calculate the next IMF_{K+1}.

(vi) The previous steps are repeated until the final residual component $r(x)$ becomes amonotonic function. The final estimation of residue $r(x)$ is subsequently considered.

Several 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, whereas the latter is difficult to use when the original series is very irregular [11]. The EMD is another effective method for extracting trends [6].

4. Local Linear Quantile (LLQ) Regression

The seminal study of Koenker and Bassett [12] 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 Chaudhuri [13] and the kernel methods by Koenker et al. [14]. In this paper, we adopt the LLQ regression employed by Yu and Jones [15].

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 $X = x$:

$$g(x) = Q_y(\tau \mid x). \tag{1}$$

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

$$\min_{\beta \in i^2} \sum_{i=1}^{n} w_i(x) \rho_\tau (y_i - \beta_0 - \beta_1 (x_i - x)), \tag{2}$$

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.

5. Bandwidth Selection

The practical performance of $\widehat{Q}_\alpha(x)$ strongly depends on the selected bandwidth parameter. We adopt the strategy of Yu and Jones [15]. In sum, we employ 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 [16] which explored

a "direct plugin" bandwidth selection procedure which relies on asymptotically optimal bandwidth:

$$h_{\text{mean}} = \left\{ \frac{\sigma^2 R(K)(b-a)}{n\mu_2^2 \int \{m''(x)\}^2 p(x)\,dx} \right\}^{1/5} = C_1(K) \left\{ \frac{\sigma^2 (b-a)}{n\theta_{22}} \right\}^{1/5}, \quad (3)$$

where and for later use we have introduced the array

$$\theta_{rs} = \int m^r(x) m^s(x) p(x)\,dx, \quad r, s \geq 0,\ r+s \text{ even}. \quad (4)$$

Again for later use we write down the following estimator with a bandwidth g for σ^2:

$$\sigma_g^2 = \frac{1}{\nu} \sum_{i=1}^{n} \left\{ Y_i - \widehat{m}_g(X_i) \right\}^2. \quad (5)$$

This is simply a normalized residual sum of squares and normalizing quantity ν, sometimes known as the degrees of freedom, which is given by $\nu = n - 2\sum_i w x_{i,h}(X_i) + \sum_{i,h} w^2 x_{j,h}(X_i)$.

Its presented guarantees $E(\sigma_g^2 \mid X_1, \ldots, X_n) = \sigma^2$, where $m(x)$ is either constant or linear [17].

(2) We use $h_\tau = h_{\text{mean}}\{\tau(1-\tau)/\phi(\Phi^{-1}(\tau))^2\}^{1/5}$ to obtain all of the other $h_\tau s$ 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 $(1-\tau)$ quantiles.

6. Proposed Method

The proposed method consists of two stages that automatically decrease the boundary effects of EMD [8]. At the first stage, LLQ which is considered as an excellent boundary treatment [18] 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.

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

$$\widehat{f}_{\text{LLQR.EMD}} = \widehat{f}_{\text{EMD}} + \widehat{f}_{\text{LPQR}}. \quad (6)$$

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

(1) The LLQ is applied to the corrupt and noisy data y_i, and the trend estimate $\widehat{f}_{\text{LLQ}}$ is subsequently obtained.

(2) The residuals of e_i from LLQ, that is, $e_i = y_i - \widehat{f}_{\text{LLQ}}$, are determined.

(3) The EMD is applied 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: this substep is accomplished by performing algorithms (i) to (vi).

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

$$\widehat{f}_{\text{LLQ.EMD}} = \widehat{f}_{\text{EMD}} + \widehat{r}(t). \quad (7)$$

7. Experiment Analysis and Results

In this section, we consider the daily closing prices of two stock markets, namely, KLSE and NZX50, from December 3, 2007, to December 6, 2013, see Figure 1. The last 10, 30, and 50 days of the KLSE and NZX50 stock indices are forecasted, respectively, based on the past sequences. The selection of these two indices aims to qualitatively and culturally compare the time series of two different markets. The data used in this study are collected from the website: http://in.finance.yahoo.com/. We analyze the two indices based on the EMD-LLQ and EMD, in combination with the FFT and the Holt-Winter methods. The approach consists of several steps. First, we decompose the daily closing prices of the stock markets into a finite number of components called IMFs and one residue. Second, we select significant components by applying the FFT to each IMF. Third, we add the significant component obtained from step two to the residue to acquire the trend. Finally, we employ the Holt-Winter method for forecasting the trend.

8. Comparison of Forecasted Data

Two criteria are used to evaluate the forecasting performance of the different models in empirical studies. The forecasting accuracy measures employed in this study are root mean square error (RMSE), mean error (MAE), and mean absolute square error (MASE). The RMS, MA, and MASE values obtained through the EMD-LLQ, EMD, and Holt-Winter methods in each test set for the two index series are summarized in Tables 1 and 2. The results demonstrate that the proposed EMD-LLQ method is more successful in all cases in forecasting the stock closing prices than the EMD and the Holt-Winter methods.

9. Conclusion and Future Research

We propose an EMD-LLQ model for forecasting future prices by considering the past sequences of daily stock prices. The EMD-LLQ method is a new two-stage forecasting method that combines the EMD and LLQ algorithm. The effectiveness

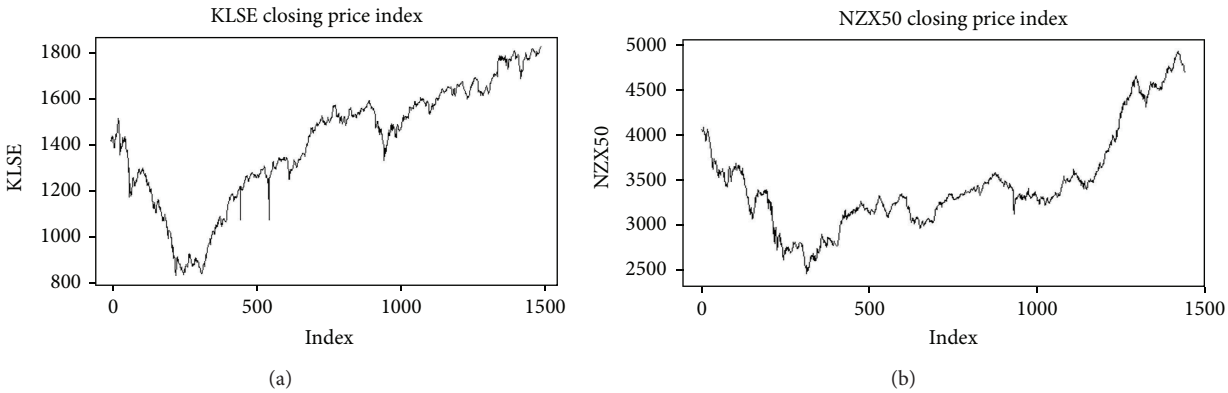

Figure 1: (a) and (b) KLSE price and NZX50 closing price index, respectively.

Table 1: Comparison of RMSE, MA, and MASE values for KLSE using the Holt-Winter, LLQ, EMD, and EMD-LLQ methods.

n.head = 10	RMSE	MAE	MASE
EMD-LLQ$_{\tau(0.25)}$	822.4843	468.8303	133.0985
EMD-LLQ$_{\tau(0.50)}$	822.2308	468.7086	129.4356
EMD-LLQ$_{\tau(0.75)}$	821.6983	468.3845	135.1209
EMD-LLQ$_{\tau(0.95)}$	822.3858	468.7594	135.0173
EMD	824.3217	470.1197	135.2341
Holt-Winters	824.0651	473.557	218.0672

Table 2: Comparison of RMSE, MA, and MASE values for NZX50 using the Holt-Winter, LLQ, EMD, and EMD-LLQ methods.

n.head = 10	RMSE	MAE	MASE
EMD-LLQ$_{\tau(0.25)}$	2002.658	1146.388	141.743
EMD-LLQ$_{\tau(0.50)}$	2003.291	1146.721	140.9419
EMD-LLQ$_{\tau(0.75)}$	2003.833	1146.989	146.0694
EMD-LLQ$_{\tau(0.95)}$	2002.921	1146.404	143.3273
EMD	2005.733	1147.129	154.8274
Holt-Winters	2010.189	1159.261	225.7123

of the new model is analyzed by performing experiments on KLSE and NZX50 data. The capability of the EMD-LLQ for forecasting the future daily stock closing prices is better than that of LLQ, EMD, and the Holt-Winter methods. The results demonstrate that EMD-LLQ is an effective method for forecasting financial time series. Future research should improve the performance of the proposed method. We can consider several forecasting strategies by analyzing the composition of daily stock closing prices. Another possible strategy is to apply the Hilbert-Huang transform for selecting meaningful IMFs (components) instead of using the FFT.

Conflict of Interests

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

Acknowledgment

The authors would like to thank Universiti Sains Malaysia (USM Short Term Grant 304/PMATHS/6313045) for financial support.

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 of London A: Mathematical*, vol. 454, pp. 903–995, 1971.

[2] K. Drakakis, "Empirical mode decomposition of financial data," *International Mathematical Forum*, vol. 3, no. 25, pp. 1191–1202, 2008.

[3] N. E. Huang, M. Wu, W. Qu, S. R. Long, and S. S. P. Shen, "Applications of Hilbert-Huang transform to non-stationary financial time series analysis," *Applied Stochastic Models in Business and Industry*, vol. 19, no. 3, pp. 245–268, 2003.

[4] A. Lin, P. Shang, G. Feng, and B. Zhong, "Application of empirical mode decomposition combined with K-nearest neighbors approach in financial time series forecasting," *Fluctuation and Noise Letters*, vol. 11, no. 2, Article ID 1250018, 2012.

[5] S. Rezaei, *Physiological Synchrony as Manifested in Dyadic Interactions*, University of Toronto, 2013.

[6] 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. 954–960, 2001.

[7] Z. Liu, "A novel boundary extension approach for empirical mode decomposition," in *Intelligent Computing*, pp. 299–304, Springer, 2006.

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

[9] C. D. Blakely, *A Fast Empirical Mode Decomposition Technique for Nonstationary Nonlinear Time Series*, Elsevier Science, Amsterdam, The Netherlands, 2005.

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

[11] J. Fan and R. Li, "Statistical challenges with high dimensionality: feature selection in knowledge discovery," http://arxiv.org/abs/math/0602133.

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

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

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

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

[16] 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, no. 432, pp. 1257–1270, 1995.

[17] T. C. M. Lee and V. Solo, "Bandwidth selection for local linear regression: a simulation study," *Computational Statistics*, vol. 14, no. 4, pp. 515–532, 1999.

[18] 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.