

Forecasting Outlier Occurrence in Stock Market Time Series Based on Wavelet Transform and Adaptive ELM Algorithm

Nargess Hosseinioun

Statistics Department, Payame Noor University, Tehran, Iran
Email: mails.students@gmail.com

Received 30 December 2015; accepted 23 February 2016; published 26 February 2016

Copyright © 2016 by author and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In financial field, outliers represent volatility of stock market, which plays an important role in management, portfolio selection and derivative pricing. Therefore, forecasting outliers of stock market is of the great importance in theory and application. In this paper, the problem of predicting outliers based on adaptive ensemble models of Extreme Learning Machines (ELMs) is considered. We found out that the proposed model is applicable for outlier forecasting and outperforms the methods based on autoregression (AR) and extreme learning machine (ELM) models.

Keywords

Component, Extreme Learning Machine, Outliers Forecasting, Wavelet Transform

1. Introduction

Outliers can have deleterious effects on statistical analyses. They can result in parameter estimation biases, invalid inferences and weak volatility forecasts in financial data. As a result when modeling financial data, their detection and correction should be considered seriously. Time-series data are often messed up with outliers due to the influence of unusual and non-repetitive events. Forecast accuracy in such situations is decreased dramatically due to a carry-over effect of the outliers on the point forecast and a bias in the estimate of parameters. The effect of additive outliers on forecasts is studied by Ledolter [1]. It was shown that forecast intervals are quite sensitive to additive outliers, but that point forecasts are largely unaffected unless the outlier occurs near the forecast origin. In such a situation the carry-over effect of the outlier can be quite substantial.

Considerable research has been devoted to the subject of forecasting and various methods have been suggested which have been divided into two main groups: classical methods mainly exponential smoothing, regres-

sion, Box-Jenkins autoregressive integrated moving average (ARIMA), generalized autoregressive conditionally heteroskedastic (GARCH) methods, and modern methods applying artificial intelligence techniques including artificial neural networks (ANN) and evolutionary computation (for more discussed details see [2]-[4]). Extreme learning machine (ELM) has been proposed as a class of learning algorithm for single hidden layer feedforward neural networks (SLFNs). In ELM algorithm, the connections between the input layer and the hidden neurons are randomly assigned and remain unchanged during the learning process. Thus by minimizing the cost function through a linear system the output connections are tuned. The computational burden of ELM has been significantly reduced as the only cost is solving a linear system. The low computational complexity attracted a great deal of attention from the research community, especially for high dimensional and large data applications. While considerable research has been devoted to detecting and removing outliers, few focused on forecasting them.

Outliers forecasting model has been discussed in [5] for the two market indexes and six individual stocks based on multi-feature extreme learning machine (ELM) algorithm. The purpose of this paper is to present adaptive ensemble model of Extreme Learning Machines (ELMs) for prediction which can lead to smaller predicting errors and more accuracy than some other forecasting methods. This paper is structured as follows: In Section 2, the theories of wavelet transform and ELM are presented, as well as how we combine both of them in the adaptive ensemble method. Section 3 describes the numerical studies, while Section 4 discusses the results.

2. Methodology

In this section, we present the methodology employed for forecasting outliers applying a wavelet decomposition technique and ELM algorithm.

2.1. Wavelet Transforms

This section contains some facts about wavelets, used throughout this paper. A thorough review of the wavelet transform is discussed in Mallat [6]-[8]. The wavelet analysis is a mathematical tool that offers decomposition of signal $s(t)$ into many frequency bands at many scales. In particular, the signal $s(t)$ is decomposed into smooth coefficients α and detail coefficients d , which are given by

$$\begin{aligned}\alpha_{j,k} &= \int s(t) \phi_{j,k}(t) dt, \\ d_{j,k} &= \int s(t) \psi_{j,k}(t) dt,\end{aligned}$$

where Φ is the father and Ψ is the mother wavelets, and j and k are, respectively, the scaling and translation parameters. The father wavelet (function) keeps the frequency domain properties (low-frequency) of the signal, while the mother wavelet keeps the time domain properties (high-frequency). The father wavelet Φ and the mother wavelet Ψ are defined as follows:

$$\begin{aligned}\phi_{j,k} &= 2^{-\frac{j}{2}} \phi(2^{-j}t - k), \\ \psi_{j,k} &= 2^{-\frac{j}{2}} \psi(2^{-j}t - k).\end{aligned}$$

The two wavelets Φ and Ψ satisfy the condition $\int \phi = 0$ and $\int \psi = 1$. Consequently, the orthogonal wavelet representation of the signal $s(t)$ is given by

$$s(t) = \sum_j \alpha_{j,k} \phi_{j,k} + \sum_j d_{j,k} \psi_{j,k} + \sum_j d_{j-1,k} \psi_{j-1,k} + \dots + \sum_j d_{1,k} \psi_{1,k}.$$

Using the above decomposition, the original signal $s(t)$ is represented with approximation coefficients $\alpha(t)$ and detail coefficients $d(t)$, by convolving the signal $s(t)$ with a low-pass filter (LP) and a high-pass filter (HP), respectively. The low-pass filtered signal is the input for the next iteration step and so on. The approximation coefficients $\alpha(t)$ contain the general trend (the low-frequency components) of the signal $s(t)$, and the detail coefficients $d(t)$ contain its local variations (the high-frequency components).

2.2. Extreme Learning Machine (ELM) Algorithm

The purpose of this paper is to discuss the mythology behind the Extreme learning machine (ELM). ELM is an

improved learning algorithm for the single feed-forward neural network structure. It notably differs from the traditional neural network methodology, since it is not essential to tune all the parameters of the feed-forward networks (input weights and hidden layer biases). For more information on efficiency of SLFNs with randomly chosen input weights, hidden layer biases and a nonzero activation function to approximate any continuous functions on any input set, one can refer to [9] and [10].

The proposed extreme learning machine (ELM) has shown its efficiency in training feedforward neural networks and overcoming the limitations faced by other conventional algorithms [11] [12]. The essences of ELM lie in two aspects, that is, random neurons and the tuning-free strategy. The learning phase of ELM generally includes two steps, namely, constructing the hidden layer output matrix with random hidden neurons and finding the output connections. Thanks to using random hidden neuron parameters which remain unchanged during the learning phase, ELM enjoys a very low computational complexity. The computational burden has been greatly reduced as the only cost is solving a linear system. At the same time, numerous applications have shown that ELM can provide a comparable or better generalization performance than the popular support vector machine (SVM) [13] [14] and the BP method in most cases [15]-[17].

ELM is a single-hidden layer feedforward network (SLFN) with a special learning mechanism which consists of three layers: input layer, hidden layer and output layer. Suppose the SLFN has n hidden nodes and non-linear activation function $g(x)$. For N training samples (x_i, t_i) , where x_i is the i th input vector and t_i is the i th desired output, the SLFN can be modeled by

$$\sum_{j=1}^n \beta_j g(w_j \cdot x_i + b_j) = o_i, \quad i = 1, \dots, N$$

where w_j is the input weight vector linking the j th hidden node and the input nodes, b_j is the bias of the j th hidden node, β_j is the output weight vector linking the j th hidden node and the output nodes, o_i is the actual network output. If ELM can approximate all the training samples (x_i, t_i) with zero error, then we claim that there exist β_j , w_j and b_j such that

$$\sum_{j=1}^n \beta_j g(w_j \cdot x_i + b_j) = t_i, \quad i = 1, \dots, N.$$

The above matrix can be expressed as $H\beta = T$, where H is called the hidden layer output matrix. As mentioned earlier, the input weights and hidden biases are randomly constructed and do not need tuning as in the case of traditional SLFN methodology. The evaluation of the output weights linking the hidden layer to the output layer is equivalent to determining the least-square solution to the given linear system. The minimum norm least-square (LS) solution to the linear system is

$$\hat{\beta} = H^+T.$$

The H in the above equation is the Moore-Penrose (MP) generalized inverse of matrix H , see [18] for more discussion. The minimum norm LS solution is unique and leads to smallest norm along all the LS solutions. The MP inverse method based on ELM algorithm is found to obtain a good generalization performance with a radically increased learning speed. One can present a general Algorithm for ELM as follows. For a given training set, activation function $g(x)$ and hidden neuron number L :

Step 1: Assign random input weight w_j and bias b_j , $j = 1, \dots, L$.

Step 2: Calculate the hidden layer output matrix H .

Step 3: Calculate the output weight.

Theoretical discussions and a more thorough presentation of the ELM algorithm are detailed in the original papers [19] [20].

2.3. Adaptive ELM

Comparable to other flexible nonlinear estimation methods, the ELM may suffer either under-fitting or over-fitting [19]. Over-fitting is particularly inaccurate since it can cause wild prediction far beyond the range of the training data even with the noise-free data. It may lead to poor predictive performance, as it may cause minor fluctuations in the data. In this work, the output of the network is only one value that is the predicted outliers.

The ensemble model is made up of a number of randomly initialized ELMs, which each have their own para-

meters. The model ELM_i has an associated weight w_i which determines its contribution to the prediction of the ensemble. Hence, we present our model only for one output. Let us define the input data as

$$Q_i = [q_i^1, q_i^2, \dots, q_i^n].$$

Comparing to the learned input patterns which is presented as

$$X_i = [x_i^1, x_i^2, \dots, x_i^n], \quad i = 1, \dots, N.$$

The determination of the closeness measure is the major factor in prediction accuracy, for which adaptive metrics are introduced to solve this problem and the arithmetic is defined by:

$$L_M(Q_i, X_i) = \left(|q_i^1 - x_i^1|^d + |q_i^2 - x_i^2|^d + \dots + |q_i^n - x_i^n|^d \right)^{\frac{1}{d}}.$$

Studying time-series forecasting, the information on trends and amplitudes plays an effective role. Adaptive metrics are introduced to solve this problem, while the arithmetic is presented as:

$$L_A(Q_i, X_i) = \min_{\lambda_r, u_r} f_r(\lambda_r, u_r), \quad (1)$$

where the parameter of minimization, λ_r equilibrates the amplitude difference between Q_i and X_i and

$$f_r(\lambda_r, u_r) = \left(|q_i^1 - \lambda_r x_i^1 - u_r|^d + |q_i^2 - \lambda_r x_i^2 - u_r|^d + \dots + |q_i^n - \lambda_r x_i^n - u_r|^d \right)^{\frac{1}{d}},$$

where h_r and l_r are the largest and smallest elements of vector correspondingly,

$\lambda_r \in \left[1, \frac{h_r}{l_r} \right]$, $u_r \in [0, h_r - l_r]$. The optimization problem (1) can be solved using the algorithm of Levenberg-Marquardt optimization or other gradient methods for $d \geq 1$. For $d = 2$, two equations may be presented as blow:

$$\begin{aligned} \frac{\partial f_r(\lambda_r, u_r)}{\partial \lambda_r} &= 0, \\ \frac{\partial f_r(\lambda_r, u_r)}{\partial u_r} &= 0. \end{aligned}$$

Then the solution of the minimization problem can be obtained analytically:

$$u_r = \frac{z_1 z_2 - z_3 z_4}{n z_2 - z_3^2},$$

where $z_j = \sum_{i=1}^n q_i^j$, $j = 1, 2$, $z_3 = \sum_{i=1}^n x_i$, $z_4 = \sum_{i=1}^n q_i x_i$. The adaptive k -nearest neighbors are chosen and the input vector of the first network can be defined as:

$$\text{Input}_g = (\hat{q}_i^{1g}, \dots, \hat{q}_i^{ng}) = \left(\frac{q_i^1 u_{r_g}}{\lambda_{r_g}}, \frac{q_i^2 u_{r_g}}{\lambda_{r_g}}, \dots, \frac{q_i^n u_{r_g}}{\lambda_{r_g}} \right).$$

The forecasting error increases considerably because of the big difference between training data and input data. In order to get more accurate results for time series q_i^1, \dots, q_i^n , k sets of inputs are used and the output vector are $\text{Output}_g = b_g$, $g = 1, \dots, k$. The mechanism for admixture of outputs is presented as follows:

$$\text{Result} = \frac{1}{k-1} \sum_{g=1}^k (\lambda_{r_g} b_g + u_{r_g}) \left(1 - \frac{d_g}{\sum_{g=1}^k d_g} \right),$$

where d_g is the distance between Q_i 's v th nearest pattern and Q_i . The model has been tested on both stationary and nonstationary time series, and the experiments show that in both cases the adaptive ensemble method leads to a prediction accuracy comparable to the best methods. For more detailed information see [16] [17].

3. Numerical Studies

The data used in the paper is the daily value of Petroleum sector Index, obtained from the DataStream database services of Tehran Over-the-Counter Market (OTC)¹. Since 2009, Iran has been developing an over-the-counter market for bonds and equities. OTC provides a complete available achieve of data, based on different sectors and dates. Our sample ranges from 28 Sep 2009 to 27 Dec 2015, with 1510 observations. *Petroleum*, the prime reason for the economic growth of the country, has been the primary *industry* in *Iran* since the 1920s. In 2012, Iran was the second-largest exporter among the Organization of Petroleum Exporting Countries², which exports around 1.5 million barrels of crude oil a day. Through primary wavelet decomposition, sequence V 's low frequency V_L and high frequency V_D are computed. In order to eliminate stochastic diffusion we set high frequency V_D equal zero. To get the main trend \hat{V} , inverse wavelet transform is used for low frequency V_L . Then we compute the absolute residual of V as sequence

$$\varepsilon_v = |V - \hat{V}|.$$

Based on sequences obtained from Matlab, we then construct an AD-ELM abnormal predicting model which can predict whether abnormal fluctuation will appear today or not. Since an ELM is essentially a linear model of the responses of the hidden layer, we apply PRESS statistics in R to retrain the ELM in an incremental way. The number of input nodes for ELM, and AD-ELM are set as 10, and the number of hidden is set to be 5. A detailed discussion of inputs and hidden nodes of ELMs with PRESS can be found in [21]. **Figure 1** shows the outliers in green color, while the red plus signs (115 points) represent abnormal points.

In order to analyze outlier detection accuracy of AD-ELM method with other methods, an adequate error measure method must be selected. In this paper we apply mean squared error (NMSE) and Mean Absolute Percentage Error (MAPE). The first is used as the error criterion, which is the ratio of the mean squared error to the variance of the time series, while the second one is regarded as one of the standard statistical performance measures. For a time series ε_v we have

$$\text{NMSE} = \frac{\sum_{i=1}^N (\varepsilon_v - \tilde{\varepsilon}_v)^2}{\sum_{i=1}^N (\varepsilon_v - \tilde{\varepsilon}_v)^2}$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \left(\frac{\varepsilon_v - \tilde{\varepsilon}_v}{\varepsilon_v} \right)^2 \cdot 100\%$$

where $\tilde{\varepsilon}_v$ is the predicted point and N is the number of predicted points.

Different prediction models on the data is summarized in **Table 1**. In our work, the AR method using $\text{AR}(m)$

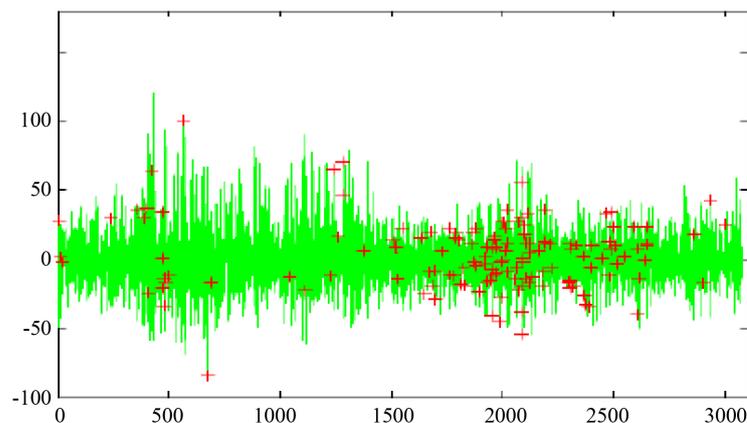


Figure 1. Outlier detection of daily value.

¹<http://www.ifb.ir>.

²Energy and the Iranian economy. United States Congress. July 25, 2006. Retrieved June 11, 2014.

Table 1. Comparisons of monthly forecasting.

	Error measures	
	NMSE	MAPE
AR	1.5678	42.65%
ELM	0.6345	12.54%
AD-ELM	0.08436	9.54%

where m is the number of input nodes of AD-ELM. In the simulation, the NMSE are 1.5678, 0.6345, and 0.08436 for AR, ELM, AD-ELM respectively, and the MAPE are 42.65%, 12.54%, 9.54% for AR, ELM, AD-ELM respectively. It is undeniable that the AD-ELM method improves upon the two other models.

4. Results and Conclusion

In this paper, forecasting models mostly have been used to forecast the stock market index value outliers. The proposed AD-ELM method is successfully used for market indexes of Tehran Over-the-Counter Market (OTC) for Petroleum sector for 1510 observations. Outliers of time series are firstly calculated through wavelet decomposition and then prediction is constructed using AD-ELM method. We plot outlier detection and evaluate forecast accuracy by mean squared error and Mean Absolute percentage error. The results reveal successfully that the accuracy of the proposed method can lead to smaller NMSE (0.08436) and MAPE (9.45%); comparing to autoregression (AR) and extreme learning machine (ELM) models, thus the AD-ELM method is a superior method for the practical forecasting of time series.

Acknowledgements

This research is supported by Payame Noor University, 19395-4697, Tehran, Iran. The author gratefully acknowledges the constructive comments, offered by anonymous referee which help to improve the quality of the paper significantly.

References

- [1] Ledolter, J. (1989) The Effect of Additive Outliers on the Forecasts from ARIMA Models. *International Journal of Forecasting*, **5**, 231-240. [http://dx.doi.org/10.1016/0169-2070\(89\)90090-3](http://dx.doi.org/10.1016/0169-2070(89)90090-3)
- [2] Araujo, R.A. (2011) A Class of Hybrid Morphological Perceptrons with Application in Time Series Forecasting. *Knowledge-Based Systems*, **24**, 513-529. <http://dx.doi.org/10.1016/j.knosys.2011.01.001>
- [3] Bodyanskiy, Y. and Popov, S. (2006) Neural Network Approach to Forecasting of Quasiperiodic Financial Time Series. *European Journal of Operational Research*, **175**, 1357-1366. <http://dx.doi.org/10.1016/j.ejor.2005.02.012>
- [4] Cao, J., Lin, Z., Huang, B. and Liu, N. (2012) Voting Based Extreme Learning Machine. *Information Sciences*, **185**, 66-77. <http://dx.doi.org/10.1016/j.ins.2011.09.015>
- [5] Fang, Z.J., Zhao, J., Fei, F.C., Wang, Q.Y. and He, X. (2013) An Approach Based on Multi-Features Wavelet and ELM Algorithm for Forecasting Outlier Occurrence in Chinese Stock Market. *Journal of Theoretical and Applied Information Technology*, **49**, 369-377.
- [6] Mallat, S. (1989) A Theory for Multiresolution Signal Decomposition the Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **31**, 679-693. <http://dx.doi.org/10.1109/34.192463>
- [7] Daubechies, I. (1992) Ten Lectures on Wavelets, CBMS-NSF Regional Conferences Series in Applies Mathematics. SIAM, Philadelphia.
- [8] Daubechies, I. (1988) Orthogonal Bases of Compactly Supported Wavelets. *Communication in Pure and Applied Mathematics*, **41**, 909-996. <http://dx.doi.org/10.1002/cpa.3160410705>
- [9] Huang, G.B., Chen, L. and Chee-Kheong, S. (2006) Universal Approximation Using Incremental Constructive Feed-forward Networks with Random Hidden. *IEEE Transactions on Neural Network*, **17**, 879-892. <http://dx.doi.org/10.1109/TNN.2006.875977>
- [10] Huang, G.B. and Slew, C.K. (2004) Extreme Learning Machine: RBF Network Case. *Proceedings of the 8th International Conference on Control, Automation, Robotics and Vision*, **2**, 1029-1036.

-
- [11] Guo, Z., Wu, J., Lu, H. and Wang, J. (2011) A Case Study on a Hybrid Wind Speed Forecasting Method Using BP Neural Network. *Knowledge-Based Systems*, **24**, 1048-1056. <http://dx.doi.org/10.1016/j.knosys.2011.04.019>
- [12] Li, M.-B., Huang, G.-B., Saratchandran, P. and Sundararajan, N. (2005) Fully Complex Extreme Learning Machine. *Neurocomputing*, **68**, 306-314. <http://dx.doi.org/10.1016/j.neucom.2005.03.002>
- [13] Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. *Machine Learning*, **20**, 273-297. <http://dx.doi.org/10.1007/BF00994018>
- [14] Hsu, C.W. and Lin, C.J. (2002) A Comparison of Methods for Multiclass Support Vector Machines. *IEEE Transactions on Neural Networks*, **13**, 415-425. <http://dx.doi.org/10.1109/72.991427>
- [15] Huang, G.B., Zhou, H., Ding, X. and Zhang, R. (2012) Extreme Learning Machine for Regression and Multiclass Classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, **42**, 513-529. <http://dx.doi.org/10.1109/TSMCB.2011.2168604>
- [16] Lin, C.T. and Lee, I.F. (2009) Artificial Intelligence Diagnosis Algorithm for Expanding a Precision Expert Forecasting System. *Expert Systems with Applications*, **36**, 8385-8390. <http://dx.doi.org/10.1016/j.eswa.2008.10.057>
- [17] Liu, N. and Wang, H. (2010) Ensemble Based Extreme Learning Machine. *IEEE Signal Processing Letters*, **17**, 754-757. <http://dx.doi.org/10.1109/LSP.2010.2053356>
- [18] Zhang, R., Lan, Y., Huang, G.B., Xu, Z.B. and Soh, Y.C. (2013) Dynamic Extreme Learning Machine and Its Approximation Capability. *IEEE Transactions on Cybernetics*, **43**, 2054-2065. <http://dx.doi.org/10.1109/TCYB.2013.2239987>
- [19] Huang, G.B., Zhu, Q.Y. and Siew, C.K. (2006) Extreme Learning Machine: Theory and Applications. *Neurocomputing*, **70**, 489-501. <http://dx.doi.org/10.1016/j.neucom.2005.12.126>
- [20] Xia, M., Zhang, Y., Weng, L. and Ye, X. (2012) Fashion Retailing Forecasting Based on Extreme Learning Machine with Adaptive Metrics of Inputs. *Knowledge-Based Systems*, **36**, 253-259. <http://dx.doi.org/10.1016/j.knosys.2012.07.002>
- [21] Myers, R.H. (1990) Classical and Modern Regression with Applications. 2nd Edition, Pacific Grove, Duxbury.