Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.4, No.12, 2014

Stock Market Returns and Direction Prediction: An Empirical Study on Karachi Stock Exchange

M. Khalid*, M. Sultana , Faheem Zaidi
Department of Mathematical Sciences,
Federal Urdu University of Arts, Science & Technology, Karachi-75300
*E-mail of corresponding author: khalidsiddiqui@fuuast.edu.pk

Abstract There has been much research in the recent past on the predictability of stock return, mainly due to its significance in managing economic gains on a high scale. Our research initiates the forecasting of the Karachi stock return with the help of the Wavelet analysis and Empirical mode decomposition method. This paper attends in large part to investors and traders to deduce a method for predicting the stock market. The collected data ranges from Jan 2009 to Dec 2012. Every training set is selected from January through October and the sets left over are used for testing. What we have discovered is that Empirical Mode decomposition (EMD) method supersedes all other models on the Mean square error and Mean Absolute error criteria. We may also evaluate the performance of these models by changing strategy direction and comparing payoffs to understand which framework performs as a better forecasting model. It is establishes by the results of the study that the same model serves better for forecasting in trading strategy and could rule over other possible models for most periods under consideration. It is our belief that this study will help stock investors to come to quick decisions about optimal buying or selling time in Karachi Stock Exchange

Key Words: Forecasting, KSE (Karachi Stock Exchange) 100 Index, Empirical Mode Decomposition, Wavelet transform ,Trading Strategy, Mean square error and Mean Absolute error

1. Introduction

Forecasting the stock market is an endeavor as old as the markets themselves. In Alice Louise Slotsky's words, "Records from ancient Mesopotamia are full of references to omens that were believed to predict commodity prices". In the world of today, a significant role is played by the financial markets in determining the economy of a country. Traditional statistical forecasting methods have been around for decades. The foundation of such traditional statistical forecasting has been linear models, mostly. These traditional methods, however, did not result in being helpful due to noise and non-linearity in the time series. In recent times, the hopes of financial researchers have been kindled because of the successful use of non-linear methods in other areas of research. The major advantage of nonlinear dynamics is that one can estimate at what point in time one state of nature(predictable) or the other (unpredictable) is more probable to come into being. The best advice one can hope to get from such a prediction is that it is best not to be in a particular market at all. Consequently, an effort has been made in this study to better understand the use of Wavelet analysis and Empirical Mode Decomposition on Karachi Stock Exchange (KSE-100 index), which is the biggest and most important stock exchange of Pakistan. In fact, in the year 2002, KSE was declared the world's best stock market by Bloomberg and it was repeatedly ranked first by "Business Weekly" for more than six years in a row as the best performing market globally.

The rest of the paper is designed as described; in the next section, the literature review is discussed. In Section 3 the data and methodology is talked about. Section 4 presents the results with discussion; and Section 5 concludes the entire paper.

2. Literature Review

The Stock market is an exciting and challenging monetary activity which makes the dry topic of economy interesting for the general public. The market climates change dramatically in one second and gains or losses come around in a twinkling. As of now, much work has been conducted in forecasting stock market returns.

Multiple studies have been conducted over the past two decades that deal with the prediction of stock prices. In a vast number of the cases, the researchers have attempted to create a linear relationship between the stock returns and input macroeconomic variables. Along with the discovery of nonlinearity in the stock market index returns [Abhyankar et al. (1997)], however, there has been a great diversion in the attention of the researchers towards the nonlinear prediction of the stock returns. Although since then a number of papers have appeared for the nonlinear statistical modeling of the stock returns, most of them needed the nonlinear model to be specified before the estimation takes place

Various sets of input variables are used to predict stock returns in literature of the kind that we are discussing. In effect, for predicting the same set of stock return date, a number of different input variables are employed. Input data from a single time series is used by some researchers where others considered the inclusion of heterogeneous market information and macro economic variables. Some researchers even processes these input data sets beforehand when feeding it to the ANN for predicting the market. Refined probabilistic NN (PNN) was used by Kim & Chun (1998) to forecast the index of a stock market. A neuro-fuzzy approach for predicting the prices of IBM stock were presented by Pantazopoulos et al. (1998). Wong put into practice a neural network model in lieu with the technical analysis variables for listed companies in Shanghai Stock Market.

A specialized neural network was used by Chenoweth et al. (1995) as the preprocessing component and a decision rule base. The job of the preprocessing component is to determine the most relevant features for stock market prediction, eliminate the noise, and split the remaining patterns into a couple of disjoint sets. Subsequently, these neural networks foretell the market's rate of return, with one network specifically trained to recognize positive and the other for the case of negative returns. The accuracy of ANN in predicting the stock market index return has been tested by researchers in the case of most developed economies worldwide. Literature is available for forecasting index returns of U.S markets like NYSE [Leigh et al. (2002)], FTSE [Brownstone (1996)], DJIA [Brown et al. (1998)], S&P 500 [Austin et al. (1997)] , Desai & Bharati (1998)). Studies in European context are available for markets like Euronext Paris Stock Exchange (Refenes et al. (1994), German Stock Exchange [Siekmann (2001)], and Madrid Stock Exchange (Fernandez-Rodriguez et al. (2000). Few papers are also available in context to Asian stock markets like Hang Seng Stock Exchange, Korea Stock Exchange Tokyo Stock Exchange and Taiwan Stock Exchange.

Kim and Han (2000) put their heads together and came up with a genetic algorithm for transforming continuous input values into discrete ones. The genetic algorithm played its part for reduction in the complexity of the feature space. Later, Kishikawa & Tokinaga (2000) incorporated a wavelet transform to extract the temporary features of stock trends.

Additionally, prominent literatures include those of of Siekmann et al. (2001) who used fuzzy rules to divide inputs into groups of increasing, stable, and decreasing trend variables. Siekmann et al. (2001) in his work employed a network structure that includes the adaptable fuzzy parameters in the weights of the connections between the first and second hidden layers. Roh (2007) changed the usual methods a little by integrating neural network and time-series model for the forecasting of the volatility of stock price index. Thawornwong & Enke (2004) then used redeveloped neural network models for foretelling which way the future stock return would go. Kim (2003) applied SVM (support vector machine) to predict the stock price index.

3. Data and Research Methodology

In this study we have only analysed data that ranges from Jan 2009 to Dec 2012 yielding a total of 1181 observations. Stock prices have been adopted from Bloomberg. For each year, every training set is taken from January to October and the reminder data set is used as a testing set.

The daily return series will be generated as

$$Returns = \frac{p_{t+1}}{p_t} - 1 = r_{t+1} \tag{1}$$

where p_t is the KSE-100 index value at time t .

3.1. Wavelet Analysis

Wavelet transform is the name given to a process which converts a signal into a series of wavelets. These converted wavelets are used to further simplify financial time series into bands such that each band has a set of near stationary signals. Furthermore, the low frequency band should be smooth enough to be converted to stationary series. The wavelet transform function (the mother wavelet) is represented as

$$\psi_{(a,b)}(t)=\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \qquad ;a>0\ ,\ b\in R \tag{2}$$

In this formula "a" is the scaling parameter and "b" is the translation parameter. The coefficient of this expression can be determined through the usual projection

$$\psi_{(a,b)}=\int_{-\infty}^{+\infty} f(t)\psi_{(a,b)}(t)\,dt \tag{3}$$

This coefficient calculates the variation of the field $f(t)$ about the point "b", with the scale given by "a". It helps to indicate high and low frequency features in a time series when decomposition take places at consecutive levels. Detailed incidents are captured when the wavelet transform is high in frequency. This is perfect for the analysis of non-stationary time series, as its job is to record detailed incidents at high frequency and long, in-time incidents when the frequency is low. For decomposing the return series, we use Daubechies wavelet of order 2. The Daubechies wavelet transforms are defined in the same way as the Haar wavelet transform by computing the running averages and differences via scalar products with scaling signals and wavelets. These scaling signals and wavelets are described very differently and that is where their uniqueness lies. This wavelet produces balanced frequency responses, but also non-linear phase responses. Daubechies wavelets use overlapping windows, so all high frequency changes are reflected by the high frequency coefficient spectrum. It looks like Daubechies of order 2 visually matches with KSE-100 index returns considerably better than the others. The To Shannon entropy criterion decides levels of decomposition. Then, they are forecasted with the help of the recursive ARIMA.

The crux of wavelet analysis lies in inverse transformation; it rebuilds the original suspected signal. After finding the approximation and details they are all independently forecasted using an approximation series. For each detail series, we use the ARIMA model. These values are then added up and we finally attain the actual forecasted values of the original series.

3.2. Empirical Mode Decomposition Model

Haung et al. (1998) proposed the Empirical Mode Decomposition algorithm, which helped to decompose a non-stationary time series into a sum of intrinsic mode functions (IMF). The basis of this algorithm is construction of smooth envelopes described by local maxima and minima of a sequence, and the successive subtraction of the mean of these envelopes from the primary sequence. This method takes into account all local extrema that are then attached by cubic spline lines, hence producing the upper and the lower envelopes.

The respective average of both envelopes are then subtracted from the initial sequence. This complete procedure assists in giving the empirical function in the first approximation. IMF extraction from EMD will satisfy the following requirements.

1) Number of IMF extrema should be equal to the number of zero-crossings or the difference should not be more than one;
2) At any point of an IMF, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima shall be zero.

To obtain the second IMF, subtract the previously obtained IMF from the original signal. The method above explained can be repeated until the required IMF is gained. The sifting procedure ends when the residue contains no more than two extrema. The similar operations are applied to the residue signals until the IMF properties are fulfilled.

4. Discussion of Results

This section provides accurate evaluations and comparisons, and the Mean Square Error (MSE) and Mean Absolute Error (MAE) are taken as the evaluation criterion. To verify the proposed model, the KSE-100 index return datasets from 2009 to 2012 are used as the experiment datasets, each year KSE dataset is named as sub-dataset. Each sub-dataset for the previous 10 months is used for training, and those from November to December are selected for testing. Furthermore, this paper compares the performances of the proposed model with the time-series model, AR (1) [Engle (1982)] model. Mean Square Error (MSE) and Mean Absolute Error (MAE) can be calculated by

$$\varepsilon_{MSE} = \frac{1}{N}\sum_{1}^{N} \left(y_{real} - y_{forecast}\right)^2$$

$$\varepsilon_{MAE} = \frac{1}{N}\sum_{1}^{N} \left|y_{real} - y_{forecast}\right| \qquad (4)$$

where y_{real} shows the real stock price return, $y_{forecast}$ is the predicted stock price return, $\bar{y}$ is the mean of y_{real}, and N is the number of data points.

It will also be interesting to compare the various methods on their one-step-ahead prediction errors (shown in table 1). With the assistance of the accuracy statistics, conclusive results are drawn; i.e. the Empirical Mode Decomposition model surpassed the other two models discussed earlier on. It is also observed that MAE and MSE calculates using Empirical Mode Decomposition (EMD) model and demonstrates a small deviation; proposing a slight difference in the actual and calculated values. Consequently, the use of EMD offers fast convergence, high accuracy and strong forecasting ability of real data.

‘Trend Following’ is a trading scheme that analysis and avails itself an advantage of long-term moves in different markets. It is capable of taking profits from ups and downs of the stock market. This strategy can help the trader know the general direction of the market. Traders like these who aim for trend following strategy are not concerned with the forecasting, but they grab the opportunity and utilize it. Several empirical studies have documented that the signs of stock returns are, to some extent, predictable. In this paper, we consider the predictive ability of the wavelet, through Autoregressive and Empirical mode decomposition methods, in predicting the course of stock returns

Trend following strategies were used to display the performance of the three used models. In this case trading simulation was of much help as it allowed the virtual investors to trade into Karachi Stock market by studying the next day's return predicted by the three models. The motive was to form a single opinion as to when an investor should buy or sell. Hence, with the help of his study, an investor should buy when the forecast return is positive, and sell if negative. Table 2 below presents a competitive comparison of the three models.

Studying the above brings us to the conclusion that the outcomes of the Empirical mode decomposition model not only beat the other model, but, more over, outdo even the benchmark model. Deductions can be ultimately made on the three employed models. In Empirical mode decomposition, training and testing sets are more consistence than Wavelet method and Autoregressive (AR) model (See table 2).

5. Conclusion

This research is one of a kind for the reason that it will help to establish wavelet analysis and Empirical Mode Decomposition as more forecasting tools for Karachi Stock market. During the course of this paper, we have attempted to deduce an optimal architecture of the wavelet analysis and Empirical Mode Decomposition to foretell the direction of the KSE-100 Index extremely accurately. According to researchers trend following strategies may be used for profitable trading in stock market. The goal is finding a way to automate the buying or selling process so as to take into account the price relative to a long time moving average value. We attempted, in this paper, to verify this trend-following strategies phenomenon by sticking to Autoregressive, Wavelet and Empirical mode decomposition models and put forward the idea that the Empirical mode decomposition model is a better model for forecasting.

Acknowledgements

The authors would like to thank reviewers for their valuable comments and suggestions for the improvement the quality of the paper. Thanks to Ms Wishaal Khalid for proof reading

References

Abhyankar,, A., Copeland, L.S. & Wong, W. (1997), “Uncovering nonlinear structure in real-time stock market indexes: The S&P 500, the DAX, the Nikkei 225, and the FTSE-100.” *Journal of Business & Economic Statistics*, 15:1–14.

Austin, M., Looney, C. & Zhuo, J. (1997), “Security market timing using neural network models.”, *New Review of Applied Expert Systems*, 3:3–14.

Brown, S.J., Goetzmann, W.N. & Kumar, A. (1998), “The Dow Theory: William Peter Hamilton′s track record reconsidered”, *Journal of Finance*, 53:1311–1333.

Brownstone, D. (1996), “Using percentage accuracy to measure neural network predictions in stock market movements”, *Neurocomputing*, 10:237–250.

Chenoweth, T., Obradovic, Z. & Lee, S. (1995), “Technical Trading Rules as a Prior Knowledge to a Neural Networks Prediction System for the S&P 500 Index,” *Proc. The IEEE Technical Applications Conference at Northcon 95*, Portland, 111-115.

Desai, V.S. & Bharati, R. (1998), “A comparison of linear regression and neural network methods for predicting excess returns on large stocks.”, *Annals of Operations Research*, 78:127–163.

Engle, R.F. (1982), “Auto regression Conditional Heteroscedasticity Estimates of the variance of UK Inflation”, *Econometrica*, 50:987-1008.

Fernandez-Rodriguez, F., Gonzalez-Martel, C. & Sosvilla-Rivero, S. (2000), “On the profitability of technical trading rules based on artificial neural networks: Evidence from the Madrid stock market”, *Economic Letters*, 69:89–94.

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C. & Liu, H.H. (1998), “The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Nonstationary Time Series Analysis”, *Proceedings of the Royal Society of London A*, 454:903–995.

Kim, K. J. & Han, I. (2000), “Genetic algorithm approach to feature discretization in artificial neuralnetwork for the prediction of stock price index”, *Experts systems with application*, 19:125-132.

Kim, K.J. (2003), “Financial time series forecasting using support vector machines”, *Neurocomputing*, 55:307–319.

Kim, S.K. & Chun, S.H. (1998), “Graded Forecasting Using An Array Of Bipolar Predictions: Application Of Probabilistic Neural Network To A Stock Market Index.” *International Journal of Forecasting*, 14:323-3378.

Leigh, W., Paz, M. & Purvis, R. (2002), “An analysis of a hybrid neural network and pattern recognition technique for predicting short-term increases in the NYSE composite index”, *Omega-International Journal of Management Science*, 30:69–76.

Kishikawa, J, & Tokinaga, S. (2000), “Realization of Feature Descriptive Systems for Clusters by using Rule Generations based on the Genetic Programming and its Applications”, *IEICE Trans. Fuadamentals*, 89(12): 2627-2635.

Pantazopoulos, K. N. (1998), “Numerical methods and software for the pricing of american financial derivatives”, *Doctoral Dissertation*, Purdue Univerity Wst Lafyett,USA.

Refenes, A.N., Zapranis, A. & Francis, G.. (1994), “Stock Performance Modeling Using Neural Networks: A Comparative Study with Regression Models”, *Journal of Neural Networks*, 7(2):375-388.

Roh, T.H. (2007), “Forecasting the volatility of stock price index”, *Expert Syst Appl*., 33:916-922.

Siekmann, S., Kruse, R. & Gebhardt, J. (2001) “Information fusion in the context of stock index prediction”, *International Journal of Intelligent Systems*, 16:1285–1298.

Thawornwong, S. & Enke, D. (2004), "The adaptive selection of financial and economic variables for use with artificial neural networks", *Neurocomputing* , 56:205–232.

Error Analysis		Mean Square Error			Mean Absolute Error		
Years	Data type	AR Returns	EMD Returns	Wavelet Returns	AR Returns	EMD Returns	Wavelet Returns
2009	Training	0.00032	0.00013	0.00010	0.01357	0.00886	0.00768
	Testing	0.00015	0.00005	0.00008	0.00940	0.00569	0.00733
2010	Training	0.00009	0.00005	0.00002	0.00684	0.00532	0.00351
	Testing	0.00005	0.00004	0.00002	0.00508	0.00502	0.00311
2011	Training	0.00010	0.00004	0.00003	0.00719	0.00481	0.00408
	Testing	0.00006	0.00004	0.00002	0.00607	0.00501	0.00357
2012	Training	0.00005	0.00003	0.00002	0.00560	0.00419	0.00310
	Testing	0.00002	0.00001	0.00001	0.00320	0.00233	0.00180

Table 1: Error analysis of daily stock returns; prediction by different forecasted method

Years	Data Type	Methods Used		
		AR Returns	EMD Returns	Wavelet Returns
2009	Training	121(56.54%)	162(75.70%)	177(82.71%)
	Testing	20(45.54%)	36(81.82%)	32(72.73%)
2010	Training	103(48.36%)	155(72.77%)	176(82.63%)
	Testing	22(50%)	28(63.64%)	36(81.82%)
2011	Training	121(56.81%)	167(78.40%)	182(85.45%)
	Testing	22(51.16%)	31(72.09%)	38(38.37%)
2012	Training	124(57.94%)	159(74.30%)	172(80.37%)
	Testing	21(48.84%)	33(76.74%)	34(79.07%)

Table 2: Forecasting returns by trend strategy using three above mentioned models