Business Excellence and Management
Volume 5 Issue 1 / March 2015

HYBRID NEURAL NETWORK MODEL TO PREDICT STOCK MARKET INDEX: EVIDENCE FOR THE TUNINDEX STOCK MARKET

Nadhem SELMI
University of Sfax, Sfax, Tunisia
nadhem.selmi@yahoo.fr

Rabia ACHAIRI
University of Sfax, Sfax, Tunisia
rabia.achairi@yahoo.fr

Abstract

In this paper we utilized as input variables the historical data and analyze the neural networks structure for financial time series forecasting. After chosen the better structure we compared the performance of the neural networks method and generalized autoregressive conditional heteroscedasticity GARCH model. In this paper we proved a combination of neural networks and statistical model is presented for forecasting the return TUNINDEX. This study motivated from the hybrid models (ARCH(2)-M-ANN, ARMA-ANN, GARCH-ANN) is better them the statistical process (ARCH, GARCH,…).
Keywords: Efficiency, Neural network, Stock market index.

1. INTRODUCTION

The study of the dynamics of stock market series and review of their future developments of current econometric models becomes a topic of interest to a large number of researchers. Motivated by the possible existence of nonlinear serial dependence in financial time series the present study examines the weak-form efficiency of Tunisian stock markets index. Among the goals of any exploration of a nonlinear forecasting method is to demonstrate improvements over linear forecasts. Using a battery of non linearity tests, the statistical results reveal that all the returns series still contain predictable non linearity even after removing linear serial correlation from the data.

Moreover, the issue of market efficiency as introduced by Fama (1965, 1970) remains the most important from resource allocation and portfolio investment point of view. Efficient mature markets are generally found to be weak-form efficient. With the hypothesis of efficiency relevance of a financial market is the ability of prices fully to reflect completely all the available information relative to the past events, present, and future. The efficient market hypothesis claims that prices fully reflect all information. This theory of investment informational efficiency dates back to the dissertation of Bachelier (1900).

Many models of financial time series assume a linear correlation structure among the time series data while there nonlinear patterns in such data that cannot be captured by GARCH models. The family of GARCH models is found to better fit empirical data of stock returns and accommodate for nonlinear and infrequent trading caused by thinness, lack of liquidity and regulatory changes and became commonplace in empirical finance.

Thus, an approximation of such complex real world problems by linear models may not always be satisfactory. Same, the school of statistics includes the class "autoregressive conditional heteroscedastic ARCH" in 1982 by Engle. The ARCH model was used to predict the volatility because of their ability to model financial series. This model provides good prediction for specific applications, but the problem is that there is a pre-specification model.

The use of this method is also due to the inability of linear models autoregressive integrated moving average ARIMA, to consider the past information contained in the time series. These models depend on the assumption of linearity between the variables and the normal distribution. However, the assumption of linearity and normal distribution cannot stand, although it has been demonstrated successfully in the treatment of movement of the stock price over the past decades. In addition, the series of stock prices are generally noisy, dynamic, non-linear, complex, non-parametric, and chaotic by nature.

Indeed, neural networks are able to describe the dynamics of non stationary of time series for the nonparametric adaptation and tolerance of ownership noise. Neural networks have become a tool increasingly used in various fields. However, they remain a subject of great interest to researchers who want to improve the performance of these networks and expand their applications.

Such, the essential element of neural networks is that they can capture the non-linear dependencies between the explanatory variables, which is possible in presence of non-linear transformation in the calculating the predicted value. Because neural networks are universal approximations tools that capture the fundamental relationships of data while avoiding model statistical fluctuations specific to set a particular training.

Another, the major advantages of Neural Networks is that, theoretically, they are capable of approximating any continuous function, and thus the researcher does not need to have any hypothesis

Business Excellence and Management
Volume 5 Issue 1 / March 2015

about the underlying model, or even to some extent, which variables matter. Rather, the model has a capacity for adaption based on the features extracted from the data.

The remaining sections of this paper are organized as follows: Section 2 gives the background of the related studies; Section 3 introduces the methodology used in this study and Section 4 provides results of each model using daily index of the Tunisian stock market. Final section gives the conclusion and recommendations for future researches.

2. LITERATURE REVIEW

2.1 ARCH model and extensions

The volatility of the financial time series is exhibit in several models. An early attempt to explain this phenomenon is provided by Engle (1982), who develops the autoregressive conditional heteroskedasticity (ARCH) framework to model the temporal dependence of conditional variance. Bollerslev (1986) extends the ARCH model to generalized ARCH (GARCH) to model financial time series. The TARCH model brought by Zakoian (1990) and Glosten et. al. (1993) and the EGARCH (exponential GARCH) model proposed by Nelson (1991).

The GARCH model, loosely speaking, can be thought of heteroscedastic time-varying variance. It is conditionally dependent on the observations of the immediate past. Autoregressive describes a feedback mechanism that in corporate past observations into the present. GARCH then is a mechanism that includes past variances in the explanation of future variances. More specifically, GARCH is a time series model that uses past variances to forecast future variances. GARCH models are generalized ARCH models, where the conditional variance at time t, σ_t, depends on earlier variances. Under GARCH (p,q), the conditional volatility is assumed to be a function of p lagged variances and q squared lagged errors where:

$$\sigma_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i} \tag{1}$$

If $p = 0$ GARCH becomes ARCH (q) and if $p = q = 0$, ε_t is simply white noise. GARCH model is one of the most popular methods used in modeling financial time series. Various recent studies show successful applications of the GARCH model in capturing the dynamics of stock indexes.

2.2 Neural Network

The artificial neural networks were developed in 1943 with the work of McCulloch and Pitts in order to impart intelligence to computers by simulating the behavior of the mammalian brain; it is inspired by those of real neurons. McCulloch and Pitts showed that a network of discrete neurons, unconstrained topology can represent any function. However, the development of efficient learning algorithms for adjusting the parameters of neuron was proposed in 1958 by Rosenblatt. Then in 1986, it appears the back propagation algorithm of the error of Rumelhart et al. with the model of the neural network with multi layers. RN then migrated from the field of artificial intelligence than statistical modeling then gradually become powerful computational tools specially adapted to situations involving non-linearity.

Neural Networks are an artificial intelligence method for modeling complex target functions. Artificial Neural Networks are a very powerful tool in modern quantitative finance and have emerged as a powerful statistical modeling technique. Neural network provide an attractive alternative tool for both researches and practitioners. Some articles have reviewed journal articles on how neural network can be applied to finance and economic (Avci, 2007; Khals and Bijari, 2010; Chen, et al, 2003; Kim, 2006; Jasemi, et al, 2011; Vanstone and Finnie, 2009)

The relationship between the output y_t and the inputs ($y_{t-1},\dots,y_{t-p}$) has the following mathematical representation:

$$y_t = w_0 + \sum_{j=1}^{q} w_j g(w_{0,j} + \sum_{i=1}^{p} w_{i,j} y_{t-i}) + \varepsilon_t \tag{2}$$

Where, $w_{i,j}$ $(i = 0,1,2,3,\dots,p;\ j = 1,2,3,\dots,q)$ and w_j $(\ j = 1,2,3,\dots,q)$ are model parameters often called connection weights; p is the number of input nodes; and q is the number of hidden nodes. Activation functions can take several forms. The type of activation function is indicated by the situation of the neuron within the network. In the majority of cases input layer neurons do not have an activation function, as their role is to transfer the inputs to the hidden layer. The most widely used activation function for the output layer is the linear function as non-linear activation function may introduce distortion to the predicated output.

2.3 GARCH-MLP models

Autoregressive conditional heteroscedasticity (ARCH) model considers the variance of the current error term to be a function of the variances of the previous time period's error terms. ARCH relates the error variance to the square of a previous period's error. If an autoregressive moving average model (ARMA

Business Excellence and Management
Volume 5 Issue 1 / March 2015

model) is assumed for the error variance, the model is a generalized autoregressive conditional heteroskedasticity (GARCH). In that case, the GARCH (p,q) model (where p is the order of the GARCH terms σ_t and q is the order of the ARCH terms ε_t) is given by:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \qquad (3)$$

Most of the financial series models are known to be easily modeled by GARCH (1, 1), so this research uses the extracted variables from GARCH (1,1) as Roh suggests (Roh, 2007). The GARCH (1,1) has the following formula:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \alpha_2 \varepsilon_{t-1}^2 \qquad (4)$$

Where σ_t is volatility at t, α_0 is the non conditional volatility coefficient, ε_{t-1}^2 residual at t-1, σ_{t-1}^2 is the variance at t-1. The newly extracted variables are as follows (Roh, 2007):

$$\sigma_t^{2'} = \beta_1 \sigma_{t-1}^2 \qquad (5)$$

$$\varepsilon_t^2 = \alpha_1 \varepsilon_{t-1}^2 \qquad (6)$$

We use these new variables as additional inputs for every type of ANN given above.

3. METHODOLOGY

The purpose of this research is to find a system that can predict exactly the price levels of actions by taking into account that the financial and economic factors such as historical price of stock indexes, transaction volume, T-bill rates, level of money supply, the exchange rate euro / dinar and dollar / dinar. Forecasting of efficiency stock price is a very complex and difficult task because there are too many factors such as political events, economic conditions, expectations of traders and other environmental factors that may influence the stock prices.

3.1 Data and preliminary tests

In this paper, we consider the Stock market index of Tunisia; we consider the daily closing prices of the TUNINDEX covering the period of 02/01/2008 to 12/09/2012, amounting to a total of 1162 observations.

The returns of the Tunisian stock market index are computed as:

$$R_t = 100 \times \log(\frac{P_t}{P_{t-1}}) \qquad (7)$$

Where, P_t is the closing price of TUNINDEX on day t and P_{t-1} is the TUNINDEX in the previous trading day.

We assume that return is stationary and if it is not stationary, then we need to convert it into stationary and only then we can estimate ARCH and GARCH model.

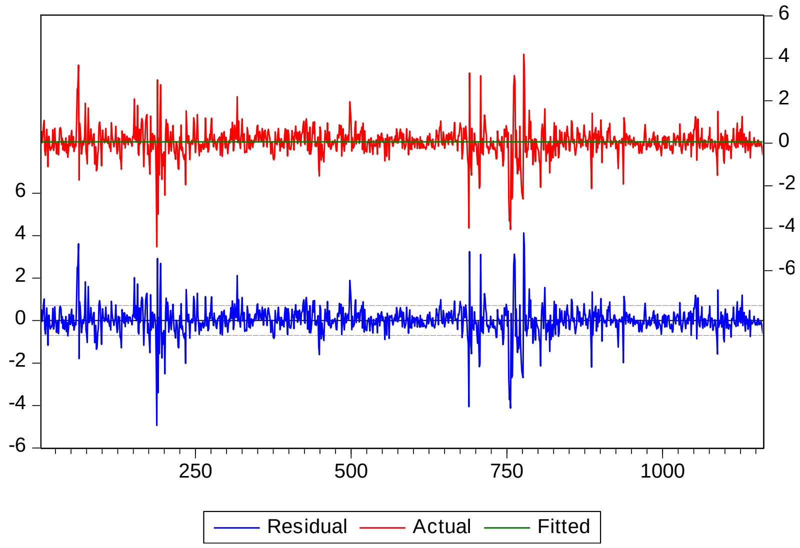

FIGURE 1 – GRAPHIC OF RETURN TUNINDEX

The graphic of return TUNINDEX indicate that is a processes homoscedastic further there is of high volatility of return.

TABLE1 - DESCRIPTION VARIABLE

	Description
Y	Historical price of stock market indexes
X	Transaction volume
K	T-bill rates
M	Level of money supply
E	The exchange rate euro/dinar
Z	The exchange rate dollar/dinar

The financial data are often affected by financial crises, mergers of companies, repurchases, etc. Which perturbs the characteristics of normality of series, including tests submitted by skewness, kurtosis and Jarque-Bera where the non-linearity of the temporal series.

- Kurtosis examines the centricity of the data around the mean. Three types of kurtosis may be utilized to explain a data set: Leptokurtic (sharper than a normal distribution around the mean with outliers on the wings, kurtosis>3), Platykurtic (fatter around the mean with outliers on the wings, kurtosis<3), Mesokurtosis (fitting the normal distribution, kurtosis=3).

- Skewness examines the curve to tell if it is centered on the mean. Three types of skewness may be utilized to explain a data set: Left skewed: the data set is shifted to the left, skewness>0), Right skewed: the data set is shifted to the right, skewness<0), Centered: the data set is centered on the mean).

TABLE 2 - STATISTICS DESCRIPTION DATA

	Y	X	K	M	E	Z
Sample Mean	4149,78	67989,34	415,81	54361,26	151483,05	69266,77
Standard Error	856,89	323858,82	317,17	140508,68	148215,28	129015,03
Skewness	-0,30	18,77	0,19	2,37	0,56	1,67
Kurtosis	-1,29	490,22	-1,29	4,02	-0,72	1,56
Jarque-Bera	98,92	117039,25	88,63	1874,29	87,05	659,49

Table 2 shows the basic statistical characteristics of the series. The kurtosis and the skewness in these data suggest that their series are not normal distribution. This is why it is necessary to use the non-linearity processes.

4. EMPIRICAL RESULTS

To evaluate forecast accuracy, this study compares the volatility forecasts of proposed hybrid models with the realized volatility. The realized volatility (RV) on day t is calculated by:

$$RV_t = \sqrt{\frac{1}{n}\sum_{i=1}^{t-n}(R_i - \bar{R})^2} \tag{8}$$

We shall be developing ARMA, ARCH, GARCH, EGARCH, ARCH-M, GARCH-M, ANN model and we shall compare all these models to find out which one is the best.

Three measures are used to evaluate the performance of models in forecasting volatility as follows: RMES, MAPE, and Theil IC. And lower the value of this criterion better fitted the model.

RMSE: Root mean square error this measure is defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\sigma_i - RV_i)^2} \tag{9}$$

MAPE: Mean absolute percent error was used to evaluate the performance model is presented by the following formula:

$$MAPE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\sigma_i - RV_i)^2} \tag{10}$$

$$Theil = \sqrt{\frac{\sum_{i=1}^{n}(R_i - \widehat{R})^2}{\sum_{i=1}^{n}(R_i - R_{i-1})^2}} \tag{11}$$

Guideline is lower the value of RMSE, MAPE and Theil IC, better the model fitted. The results are reported in table 3.

TABLE 3 - FORECAST RESULT OF ARCH FAMILY MODELS

	RMSE	Rank	MAPE	Rank	Theil IC	Rank
ARMA (1,1)	7.315	6	1.325	2	0.0081	7
ARCH(2)	7.214	4	1.287	6	0.0078	6
GARCH(1,1)	7.112	3	1.258	5	0.0068	5
EGARCH(1,1)	7.110	2	1.245	1	0.0065	2
ARCH(1)-M	7.358	4	1.471	3	0.0066	3
GARCH(1,1)-M	7.310	5	1.365	4	0.0067	4
ANN	6.878	1	1.214	4	0.00014	1

If we assume the MAPE criteria the EGARCH is the best model. The advantage of EGARCH model is the logarithmic form of conditional variance; therefore, it is not necessary to impose restrictions in the coefficients because it is a positive sign.

The best model turned out to be ANN according to RMSE and Theil IC criteria. The advantage of ANN is it can capture the non-linear high-level dependencies between variables. Further, the ANN model does not suppose a specification of the forms of relationship between the variables studied. The intelligent artificial is present a powerful tool for forecasting a return of TUNINDEX.

In spite of the advantages of ANN compared with the statistical models shelf space for the forecast of the financial data the reliability of these results in summer criticized because of a number of problem such as the sensibility of network in cut her of data and the size of the sample. Therefore, it is recommended that these models be combined with other models when applied to stock markets.

TABLE 4 - FORECAST RESULT OF HYBRID MODELS

	RMSE	Rank	MAPE	Rank	Theil IC	Rank
ARCH(2)-M-ANN	6.891	1	1.234	3	0.0064	1
ARMA-ANN	7.109	2	1.312	2	0.0067	2
GARCH-ANN	6.451	3	1.214	1	0.0002	3

Table 4 present the result of the application of the proposed hybrid models for forecasting volatilities.

Business Excellence and Management
Volume 5 Issue 1 / March 2015

Business Excellence and Management
Volume 5 Issue 1 / March 2015

The best model turned out to be ARCH(2)-M-ANN according to RMSE and Theil IC criteria and if we assume the MAPE criteria the GARCH-ANN is the best model. The hybrid model gives results better than the ANN and the statistical model.

5. CONCLUSIONS

In this paper a combination of neural networks and statistical model is presented for forecasting the return TUNINDEX. This study motivated from the hybrid models (ARCH(2)-M-ANN, ARMA-ANN, GARCH-ANN) is better them the statistical process (ARCH, GARCH,…).
As a possible extension of models of artificial neural networks hybrid models can be formalized. This in order to further improve the forecasting time series finance.

REFERENCES

Avcı, E. (2007). Forecasting daily and sessional returns of the ISE-100 index with neural network models, *Doğuş Üniversitesi Dergisi*, 8, 128-142.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 31, 307–327.

Chen, A.S., Leung M.T. and Daouk, H. (2003). Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index, *Computers and Operations Research* 30, 901-923.

Engle, R.F., (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* 50, 987–1007.

Fama, E. (1965). The behavior of stock market prices, *Journal of Business*, 38, 34−105.

Fama, E. (1970). Efficient capital markets: A review of theory and empirical work, *Journal of Finance*, 25, 383−417.

Fama, E. (1991). Efficient Capital Markets: II, *The Journal of Finance,* 46(5), 1575-1617.

Guresen, E., Kayakutlu G., and Daim, T.U. (2011). Using artificial neural network models in stock market index prediction, *Expert Systems with Applications* 38, 10389–10397.

Jasemi, M., Kimiagari A.M., and Memariani, A. (2011). A modern neural network model to do stock market timing on the basis of the ancient investment technique of Japanese Candlestick, *Expert Systems with Applications* 38, 3884–3890.

Khasli, M. and Bijari, M. (2010). An artificial neural network (p,d,q) model for time series forecasting, *Expert Systems with Applications* 37, 479-489.

Kyoung-Jae, K., (2006). Artificial neural networks with evolutionary instance selection for financial forecasting, *Expert Systems with Applications* 30, 519-526.

Neslon, D.B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica* 59, 347-370.

Principe, J. C., Euliano, N. R., and Lefebvre, W. C. (1999*). Neural and adaptive systems: Fundamentals through simulations*. New York, USA: John Wiley & Sons.

Roh, T. H, (2007). Forecasting the volatility of stock price index, *Expert Systems with Applications*, 33, 916–922.

Vanstone B. and Finnie, G. (2009). An empirical methodology for developing stockmarket trading systems using artificial neural networks, *Expert Systems with Applications* 36, 6668–6680.