IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-1, Jan 2019 E-ISSN: 2347-2693

Forecasting Financial Time Series using a Hybrid Non-stationary Model with ANN

D.K. Shetty[1*], B. Ismail[2]

[1]Department of Statistics ,Mangalore University,Mangalagangothri,India
[2]Department of Statistics ,Mangalore University,Mangalagangothri,India

Corresponding Author: dileepshetty1991@gmail.com, Tel: +919900885735

Available online at: www.ijcseonline.org

Accepted: 22/Jan/2019, Published: 31/Jan/2019

Abstract- Forecasting financial time series have been regarded as one of the most challenging applications of modern time series forecasting. Thus, numerous models have been depicted to provide the investors with more precise predictions. In recent years, financial market dynamics forecasting has been a focus of economic research. In this paper, we propose a hybrid non-stationary time series model with artificial neural network (ANN) for forecasting financial time series. The proposed model is non-stationary in trend component with regressor, lagged variable and non-linear component. The proposed model can capture both linear and non-linear structures in the time series. Non-linear structure is capture by Fed-Forward Neural Networks (FNN). The working of the proposed model is examined for SPY and VOO stock prices. Forecast based on the proposed model performs better than existing models in terms of Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percent Error (MAPE) criterion.

Keywords- ARIMA-ANN, ARIMA-GARCH, Trend, Hybrid and Accuracy.

I. INTRODUCTION

Financial time series analysis and forecasting is an active research area over the last few decades. Forecasting is essential for planning and operation in a variety of areas such as financial planning, production management and investment analysis. The accuracy of forecasting is important to many decision processes and hence the research for improving the accuracy of forecasting has never stopped. Autoregressive integrated moving average (ARIMA) and Artificial Neural Network (ANN) models are have been used for forecasting the time series. ARIMA model is used to handle the linear effect in the time series and ANN is used to model both linear and non-linear structures in the data even though they don't handle both structures equally well. Both theoretical and empirical findings in the literature show that combining different models is an efficient way to improve the accuracy in forecasting.

There are situations where the time series is non-stationary in trend, depends on exogenous factors (regressors), lagged variables and non-linear component. For this type of situation, an appropriate model needs to be developed. We propose a hybrid model by combining the time series model with ANN which takes care of the above factors.

In this paper, we modify Zhang's hybrid approach mentioned above. To estimate linear part of the problem, we propose to use $X_t = \sum_{i=0}^{k} a_i t^i + \phi X_{t-1} + \gamma u_t + \varepsilon_t$ model instead of ARIMA model. We follow a three steps procedure: In step step-I, we estimate and forecast X_t using $X_t = m(t;a) + \phi X_{t-1} + \gamma u_t + \varepsilon_t$ model, in step-II, we estimate and forecast residuals of step-I estimation using Artificial Neural Network approach and in the last step, we combine the forecast based on step-I and step-II.

Organization of this paper is as follows. Related work is given in the next section. In section III, we present the proposed model with the theoretical discussion. The working of the proposed model for forecasting is demonstrated with an application of SPY and VOO stock prices in section IV and concluding remarks are given in the last section. All computations of this work were carried out using R (3.4.4) software.

II. RELATED WORK

Pai[6] proposed hybrid ARIMA and support vector machines (SVM) model for forecasting stock price. Tseng [2] proposed a hybrid SARIMA and the neural network back propagation (SARIMABP) model to forecast two seasonal time series data.Zhang [5] proposed a hybrid ARIMA and Fed-Forward Neural Networks (FNN) model. Aladag [1] proposed a hybrid approach combining Elman's Recurrent Neural Networks (ERNN) and ARIMA model.

III. METHODOLOGY AND MAIN RESULT

Non-stationary time series model

The proposed hybrid non-stationary model has the representation

$$X_t = L_t + N_t \ , \quad (1)$$

where X_t denotes original time series, L_t denotes the linear component and N_t denotes the nonlinear component. The linear component may include a trend component, lagged variable and exogenous variable.

Linear component L_t can be represented as

$$L_t = m(t;a) + \phi X_{t-1} + \gamma u_t \ , \quad (2)$$

where $m(t;a)$ is the k[th] order polynomial trend with (k+1)- dimensional parameters $a \in R^{k+1}; X_{t-1}$ is lagged variable and u_t is exogenous variable.

We estimate the linear component L_t by modeling X_t as

$$X_t = \sum_{i=0}^{k} a_i t^i + \phi X_{t-1} + \gamma u_t + \varepsilon_t , \quad (3)$$

where the error term ε_t is independent random variable such that $E(\varepsilon_t | X_{t-1}, u_t) = 0$ and $Var(\varepsilon_t | X_{t-1}, u_t) = \sigma^2$.

We assume that

a) ε_t is a sequence of an independent identically distributed normal random variable with $E(\varepsilon_t)=0$ and $Var(\varepsilon_t)=\sigma^2$: $\varepsilon_t \sim N(0,\sigma^2)$

b) Non-autocorrelation : $E(\varepsilon_t \varepsilon_s) = 0$ for $t \neq s$

c) Uncorrelated with explanatory variables X_{t-1} & u_t : $E(X_{t-1}, \varepsilon_t) = 0$ and $E(u_t, \varepsilon_t) = 0$

Based on these assumptions we can estimate the parameters of the model (3) by conditional maximum likelihood estimation approach.

Linear component L_t estimated by the model (3) and residuals obtained from the estimated model is

$$Z_t = X_t - \hat{L}_t \ , \quad (4)$$

where $\hat{L}_t$ is the estimated value of the linear component for time t of the time series X_t by the model(3). Z_t can be estimated by FNN. With p input nodes, the ANN model for the residuals can be written as

$$Z_t = f(z_{t-1}, z_{t-2}, ..., z_{t-p}) + e_t \ , \quad (5)$$

where f (.) is a nonlinear function determined by the neural network and e_t is the random error. The estimation of Z_t in (5) will yield the estimation of non-linear component N_t in time series X_t . By this way, estimated values of the time series are obtained as follows:

$$\hat{X}_t = \hat{L}_t + \hat{N}_t \quad (6)$$

Forecasting

We derive minimum mean square error (MMSE) forecast for the time series represented by the hybrid model.

Let $\hat{X}_n(l)$ denote a forecast of X_{n+l} based on $I_n = \{X_n, X_{n-1}, ...\}$. We use quadratic loss function for forecasting X_{n+l}. Accordingly choosing the forecast $\hat{X}_n(l)$ so as to minimize

$$MSE(\hat{X}_n(l)) = E\left(X_{n+l} - \hat{X}_n(l)\right)^2 \quad (7)$$

The forecast with the smallest mean square error turns out to the conditional expectation of X_{n+l} given I_n:

$$\hat{X}_n(l) = E(X_{n+l} | I_n) \quad (8)$$

Let us consider the linear component L_t

$$L_t = a_0 + a_1 t + a_2 t^2 + ... + a_k t^k + \phi X_{t-1} + \gamma u_t \ , \quad (9)$$

Let $L_n(h)$ denote a forecast of L_{n+h} based on $I_n = \{X_1, X_2, . . . , X_n\}$

For h=1, one step ahead forecast value of L_t at origin $t = n$,

$$\begin{aligned}\hat{L}_n(1) &= E(L_{n+1} | I_n) \\ &= E((a_0 + a_1(n+1) + a_2(n+1)^2 + ... \\ &\quad + a_k(n+1)^k + \phi X_n + \gamma u_{n+1} + \varepsilon_{n+1}) | I_n)\end{aligned}$$

$$\begin{aligned}\hat{L}_n(1) &= a_0 + a_1(n+1) + a_2(n+1)^2 + ... \\ &\quad + a_k(n+1)^k + \phi X_n + \gamma \mu_u\end{aligned} \quad (10)$$

For h=2,

$$\begin{aligned}\hat{L}_n(2) &= E(X_{n+2} | \mathrm{I_n}) \\ &= E((a_0 + a_1(n+2) + a_2(n+2)^2 + ... \\ &\quad + a_k(n+2)^k + \phi X_{n+1} + \gamma u_{n+2} + \varepsilon_{n+2}) | I_n)\end{aligned}$$

$$\begin{aligned}\hat{L}_n(2) &= a_0 + a_1(n+2) + a_2(n+2)^2 + ... \\ &\quad + a_k(n+2)^k + \phi \hat{L}_n(1) + \gamma \mu_u\end{aligned} \quad (11)$$

Equation (10) in (11),

$$\begin{aligned}\hat{L}_n(2) &= a_0(1+\phi) + a_1[(n+2) + \phi(n+1)] \\ &\quad + a_2[(n+2)^2 + \phi(n+1)^2] + ... + a_k[(n+2)^k \\ &\quad + \phi(n+1)^k] + \phi^2 X + \gamma(1+\phi)\mu_u\end{aligned} \quad (12)$$

For h=3, three step ahead forecast value $\hat{L}_n(3)$ of L_{n+h}

$$\begin{aligned}\hat{L}_n(3) &= E(L_{n+3}|I_n)\\ &= a_0(1+\phi+\phi^2)+a_1[(n+3)+\phi(n+2)+\phi^2(n+1)]\\ &+a_2[(n+3)^2+\phi(n+2)^2+\phi^2(n+1)]+....\\ &+a_k[(n+3)^k+\phi(n+2)^k+\phi^3(n+1)^k]\\ &+\phi^3 X_n+\gamma(1+\phi+\phi^2)\mu_u\end{aligned} \quad (13)$$

Similarly, h-step ahead forecast of L_{n+h} is

$$\hat{L}_n(h) = a_0\sum_{j=0}^{h-1}\phi^j + a_1\sum_{j=0}^{h-1}\phi^j(n+h-j)$$

$$+a_2\sum_{j=0}^{h-1}\phi^j(n+h-j)^2+....$$

$$+a_k\sum_{j=0}^{h-1}\phi^j(n+h-j)^k+\phi^h X_n+\gamma\sum_{j=0}^{h-1}\phi^j\mu_u$$

In generally, we have

$$\hat{L}_t(h) = \sum_{i=0}^{k}a_i\sum_{j=0}^{h-1}\phi^j(t+h-j)^i+\phi^h X_t+\gamma\sum_{j=0}^{h-1}\phi^j\mu_u \quad (14)$$

Equation (14) represents forecast of linear component L_t for future h-periods.

As we discussed earlier, non-linear relationships can be discovered by modeling residual series $\{Z_t\}$ using ANN. Single hidden layer FFN is the most widely used model for time series modeling and forecasting [5]. The model is characterized by a network of three layers of simple processing units connected by a cyclic link. The relationship between the output (Z_t) and the inputs (Z_{t-1} , Z_{t-2} , . . ., Z_{t-p}) has the following representation:

$$Z_t = \alpha_0+\sum_{j=1}^{q}\alpha_j g(\beta_{0j}+\sum_{i=1}^{p}\beta_{ij}Z_{t-i})+e_t \quad (15)$$

where α_j (j=0,1,2,...,q) and β_{ij} (i=0,1,2,...,p; j=1,2,...,q) are the model parameters often called the connection weights; p is the number of input nodes and q is the number of hidden nodes. The logistic function is often used as the hidden layer transfer function, that is,

$$g(x)=\frac{1}{1+e^{-x}} \quad (16)$$

Hence, the ANN model (15) in fact performs a nonlinear functional mapping from the past observations (Z_{t-1} , Z_{t-2} , . . ., Z_{t-p}) to the present value Z_t .

The prediction of Z_t by (15) will yield the forecasting of nonlinear component (N_t) of time series X_t . With this forecasted values of the time series X_t are given by

$$\hat{X}_t(h)=\hat{L}_t(h)+\hat{N}_t(h) \quad (17)$$

where $\hat{L}_t(h)$ and $\hat{N}_t(h)$ are forecast of linear and non-linear components of X_t for h periods ahead respectively.

IV. DATA ANALYSIS

Two data sets—the SPY and the VOO stock price are used in this study to demonstrate the working of the hybrid method. We consider the daily closing values of SPY and VOO stock on trading days from 1st January 2015 to 2nd February 2018. To assess the forecasting performance of different models, each data set is divided into two sets of training and testing. The training data set is used exclusively for model development and then the test sample is used to evaluate the performance of the proposed model. The first 735 observations (1st January 2015 to 1st December 2017) are used for parameters estimation, while the next 41 observations (1st December 2017 to 2nd February 2018) are used for out-sample forecast evaluation.

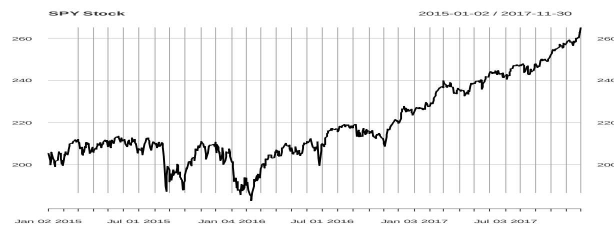

Figure 1. Daily closing values of the SPY stock price

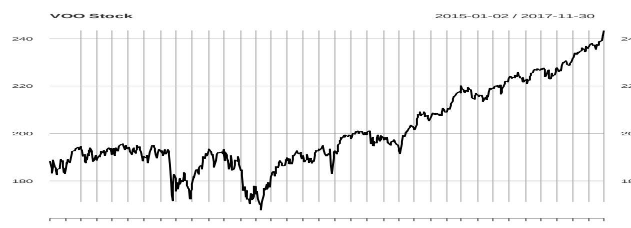

Figure 2. Daily closing values of the VOO stock price

Figure 1 and Figure 2 represent the graph of the SPY and VOO closing price series for the period 1st January 2015 to 1st December 2017 respectively. Both stock prices are non-stationary in terms of trend component. Both linear and nonlinear models have been applied to these data sets, although more or less non-linearities have been found in these series. First, model (3) is used for estimating the linear part of the problem. Analysis indicates that the estimated linear component ($\hat{L}_t$) is $\hat{a}_0+\hat{a}_1 t+\hat{a}_2 t^2+\hat{\phi}X_{t-1}+\hat{\gamma}u_t$ (Table.1) is presence in both the data sets, here u_t is drawn from the uniform distribution between 0 and 10. The residual series Z_t is obtained after eliminating $\hat{L}_t$ from the actual observations.

Table 1. Summary about fitted linear component ($\hat{L}_t$)

Coefficients	Estimates (For SPY Stocks)	Test for Significance of estimates (p-value)	Estimates (For VOO Stocks)	Test for Significance of estimates (p-value)
a_0	8.8477	0.00013	8.0048	0.000143
a_1	-0.00333	0.01788	-0.00299	0.01967
a_2	0.0000093	0.00049	0.0000085	0.000542
ϕ	0.9567	0.0000	0.9573	0.0000
γ	0.04826	0.0208	0.04212	0.0272

Next, the residual Z_t is estimated by the ANN model and we found that the most appropriate neural model is $1\times4\times1$ networks. The Error Statistic values for the last 41 forecasted observations of the proposed model, Zhang [5], ARIMA-GARCH and ANN models are summarized in Table.3 and Table.4. It is observed that RMSE, MAE and MAPE of the proposed method are the smallest compared to the other three approaches. Thus, it is concluded that the proposed method improves the accuracy of forecasting.

Table 3. Forecasting results of SPY stock

Error statistic	Proposed Approach	Zhang[5]	ARIMA-GARCH	ANN
MAPE	**1.4384**	2.2158	1.9532	2.6876
RMSE	**5.4852**	8.4504	7.4042	10.013
MAE	**4.0121**	6.1866	5.4470	7.4965

Table 4. Forecasting results of VOO stock

Error statistic	Proposed Approach	Zhang[5]	ARIMA-GARCH	ANN
MAPE	**1.4584**	2.3542	2.0090	2.9282
RMSE	**5.0676**	7.9618	6.9601	9.6937
MAE	**3.7375**	6.0314	5.1484	7.4952

V. CONCLUSION

The accuracy of forecasting is a topic of current interest and research. Statisticians have studied to obtain better forecasts for long terms and by these studies, hybrid models have been improved in the literature. Therefore, we propose a hybrid non-stationary model with ANN and compare the forecasting error between the proposed model, Zhang [5], ANN, ARIMA-GARCH models based on RMSE, MAE and MAPE criterion.

It is observed that the proposed model yields the least RMSE, MAE and MAPE than those produced by the other two models (Table.3 and Table.4). Hence the proposed model is superior to other models under error statistic criteria. Thus in the presence of trend and non-linear effect the hybrid non-stationary model with ANN gives better forecast than the existing models.

REFERENCES

[1] C.H.Aladag, E.Egrioglu, C.Kadil, *"Forecasting nonlinear time series with a hybrid methodology"*, Applied Mathematics Letters,Vol.**22**, Issue.**9,** pp.**1467-1470, 2009.**

[2] F.M. Tseng, H.C. Yu, G.H. Tzeng, *"Combining neural network model with seasonal time series ARIMA model"*, Technological Forecasting & Social Change,Vol. **69**, Issue.**1** , pp.**71-87, 2002.**

[3] G. Zhang, E.B. Patuwo, M.Y. Hu, *"Forecasting with artificial neural networks: the state of the art"*, Journal of Forecasting,Vol. **1,** Issue**.1,** pp.**35-62, 1998**.

[4] G.E.P. Box, G.M. Jenkins, *"Time Series Analysis: Forecasting and Control"*, Holden-Day, San Francisco, **1970.**

[5] G.Zhang, *"Time series forecasting using a hybrid ARIMA and neural network model"*,Neurocomputing,Vol. **50 ,** Issue. **17,** pp.**159-175, 2003.**

[6] M.H. Ahmad, P.Y. Ping, S.R Yaziz, N.A Miswan, *"A Hybrid Model for Improving Malaysian Gold Forecast Accuracy*", International Journal of Mathematical Analysis, Vol. **8,** Issue. **28,** pp. **1377-1387, 2014**.

[7] P.F. Pai, C.S. Lin, *"A hybrid ARIMA and support vector machines model in stock price forecasting"*, The International journal of Management Science, Vol. **33,** Issue. **6**, pp. **497-505, 2005**.

[8] R.S. Tsay, *"Analysis of Financial Time Series"*, 2nd ed., John Wiley& Sons, U.K, **2005**.

[9] T.Nakatsuma, *"Bayesian analysis of ARMA-GARCH models: A Markov chain sampling approach"*, Journal of Econometrics, Vol. **95**, Issue. **1**, pp. **57-69 2000**.

[10] W.W.S Wei, *"Time Series Analysis: Univariate and Multivariate methods"*, Pearson, San Francisco, **2006**.

AUTHOR PROFILE

Mr. Dileep Kumar Shetty He pursued his Master of Science in Statistics from Mangalore University, Karnataka INDIA in 2014. He is currently pursuing Ph.D in Statistics and published one research paper in Journal of Indian Society for Probability and Statistics.

Dr. Ismail B, He is professor of Statistics and Chairman in Department of Statistics Mangalore University, Karnataka India .He completed his Ph.D in Statistics in Mangalore University .Title of the thesis is some Contributions to the Statistical Analysis of Stochastic Models in Econometrics. Also Honorary Senior Research Fellow, Department of Statistics, University of Glasgow, Scotland U.K., 2000-2001. "Franca e Diego de Castro" Award (jointly with C T Jose) for the Research Paper "Estimation of Error Correlation in Nonparametric Regression" in honor of Diego de Castro, Torino, Italy, 2004.He has more than 27 years research experience in the field of Statistics and published many research papers in reputed journals.