Improving time series forecasting using a hybrid SARIMA and neural network model

Amaal El Sayed Abd El Ghany Mubarak
Department of applied statistics and insurance,
faculty of commerce , Damietta University, Egypt

Abstract

A hybrid forecasting model was proposed in this article; seasonal time series ARIMA and neural network back propagation(BP) models were combined together in which is known as SARIMABP. This model was used to improve forecasting of high frequency data with application on exchange rate (Egyptian pound / US dollar). the aim is to combine models to build a complete picture especially if a time series exhibits different patterns. The forecasting performance was compared among SARIMABP and SARIMA models and showed that The mean square error (MSE) and the mean absolute error (MAE) of the SARIMABP model were the lowest. The turning point evaluations also show that the proposed model has the ability to capture the actual direction of turning points of the time series.

Keywords: ARIMA; back propagation; foreign exchange; high frequency; neural network; SARIMA; SARIMABP; time series

1. Introduction

Usually, linear models of financial time series like exchange rates perform poorly and linear univariate models consistently give evidence for a random walk. therefore, an often followed strategy is to try to use nonlinear models to improve the fit and thus the prediction. Neural Networks are flexible functional forms that allow to approximate any continuous, also nonlinear, function. So, they can be expected to provide effective nonlinear models for financial time series and thus to allow for better predictions. Many authors had reported mixed evidence as to forecasting results of Neural Networks. although Neural Networks performed occasionally better than many statistical methods, also reported as prediction tools for different currency exchange rates.

However, despite all advantages cited for artificial neural networks, their performance for some real time series is not satisfactory. Improving forecasting especially time series forecasting accuracy is an important yet often difficult task. One of the major developments in neural networks over the last decade is the model combining or ensemble modeling. The basic idea of this multi-model approach is the use of each component model's unique capability to better capture different patterns in the data. Both theoretical and empirical results have indicated that integration of different models can be an effective way of improving upon their predictive performance, especially when the models in the ensemble are quite different. In a competitive architecture, the aim is to build appropriate modules to represent different parts of the time series, and to be able to switch control to the most appropriate. In a cooperative modular combination, the aim is to combine models to build a complete picture from a number of partial solutions. The assumption is that a model may not be sufficient to represent the complete behavior of a time series, for example, if a time series exhibits both linear and nonlinear patterns during the same time interval, neither linear models nor nonlinear models alone are able to model both components simultaneously. A good exemplar is models that fuse auto-regressive integrated moving average with artificial neural networks. An auto-regressive integrated moving average (ARIMA) process combines three different processes comprising an auto-regressive (AR) function regressed on past values of the process, moving average (MA) function regressed on a purely process, and an integrated (I) part to make the data series stationary by differencing. In such hybrids, whilst the neural network model deals with nonlinearity, the auto-regressive integrated moving average model deals with the non-stationary linear component.

Several studies (Ansuj, Camargo, Radharamanan, & Petry, 1996; Kohzadi, Boyd, Kermanshahi, & Kaastra, 1996; Hill, O'Connor, &Remus, 1996; Caire, Hatabian, & Muller, 1992) have compared between Neural Networks and time series techniques and all of them has proved that neural network was a superior method for forecasting in their test cases. Maier and Dandy (1996) suggested that the ARIMA model is better suited for short-term forecasts and that neural networks are better suited for longer-term forecasts and concluded that time series methods are preferable to explanatory approaches, at least for short-term horizons. There have been few studies suggesting hybrid models; combining time series models and neural networks (Voort, Dougherty & Watson, 1996; Wang &Leu, 1996; Su, Tong & Leou,1997).Their results showed that the hybrid model produced better forecasts than either the ARIMA model or the neural network by itself. In recent years, more hybrid forecasting models have been proposed, using auto-regressive integrated moving average and artificial neural networks and applied to time series forecasting with good prediction performance. Tseng, Yu, and Tzeng (2002), Khashei, and Bijari (2010) proposed new hybrid models in order to overcome the data limitation of neural networks and yield more accurate forecasting models, especially in incomplete data situations. the hybrid model has produced promising results in these studies. In this paper, we combine the seasonal time series ARIMA (SARIMA) model and the neural network back propagation (BP) model to forecast time series with seasonality. In this proposed model, the future value of a time series is considered as a function of several past observations and random errors, such SARIMA models. Therefore, in the first phase, an auto-regressive integrated moving average model is used in order to generate the necessary data from under study time series. Then, in the second phase, a neural network is used to model the generated data by SARIMA model, and to predict the future value of time series. Because previous studies have shown the quality of predictions obtained by the proposed method, that is considered a motivation for application and use of this method to improve the prediction of significant changes in our national economy; such as exchange rate, where the practical importance of this search is. Beside that, there is a scientific importance when we introduce the steps of the application of this method to users and beneficiaries.

2. Neural network with seasonal time series ARIMA model

2.1. SARIMA model

A time series $\{Z_t / t = 1,2,\ldots,\kappa\}$ is generated by SARIMA (p, d, q) (P, D,Q)$_s$ process with mean μ of Box and Jenkins time series model if

$$\varphi(B)\Phi(B^s)(1-B)^d(1-B^s)^D(Z_t-\mu)=\theta(B)\Theta(B^s)a_t \qquad (١)$$

Where $p,\ d,\ q,\ P,\ D,\ Q$ are integers; s is the periodicity of season; $\varphi(B)=1-\varphi_1 B-\varphi_2 B^2-\ldots-\varphi_p B^p$, $\theta(B)=1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q$, $\Phi(B^s)=1-\Phi_1 B^s-\Phi_2 B^{2s}-\ldots-\Phi_p B^{ps}$ and $\Theta(B^s)=1-\Theta_1 B^s-\Theta_2 B^{2s}-\ldots-\Theta_Q B^{Qs}$ are polynomials in B of degree $p,\ q,\ P,$ and Q ; B is the backward shift operator; d is the number of regular differences, D is the number of seasonal differences; Z_t denotes the observed value at time t, $t = 1, 2, \ldots, k$; and a_t is the estimated residual at time t. The a_t should be independently and identically distributed as normal random variables with mean equal 0 and variance σ^2.

The Box and Jenkins (1976) methodology includes three iterative steps of model identification, parameter estimation, and diagnostic checking. They proposed to use the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the sample data as the basic tools to identify the order of the SARIMA model. The basic idea of model identification is that if a time series is generated from an SARIMA process, it should have some theoretical autocorrelation properties. By matching the empirical autocorrelation patterns with the theoretical ones, it is often possible to identify one or several potential models for the given time series. Some other order selection methods have been proposed such as the Akaike's information criterion (AIC) and Schwarz Bayesian criterion (SBC). In the identification step, data transformation is often required to make the time series stationary, which means that the mean and the autocorrelation structure being constant over time. Differencing and power transformation are applied to remove the trend and to stabilize the variance before an SARIMA model can be fitted. In the parameters estimation step, the parameters are estimated such that an overall measure of errors is minimized. This can be accomplished using a nonlinear optimization procedure. The last step in model building is the diagnostic checking of model adequacy witch is basically checked if the model assumptions about the errors, a_t , are satisfied.

2.2. Neural network back propagation (BP) model

One of the most significant advantages of the neural network models over other classes of nonlinear models is that neural network models are universal approximations for a large class of functions with a high degree of accuracy. Their power comes from the parallel processing of the information from the data. No prior assumption of the model form is required in the model building process. Instead, the network model is largely determined by the characteristics of the data.

The neural network consists of an input layer, an output layer and one or more hidden layers. The hidden layers can capture the nonlinear relationship between variables. Each layer consists of multiple neurons that are connected to neurons in adjacent layers. The model parameters (connection weights) will be adjusted iteratively by a process of minimizing the forecast errors. For each training iteration, an input vector, randomly selected from the training set, was submitted to the input layer of the network being trained. The output of each processing unit (or neuron) was propagated through each layer of the network using the weights on the connections from unit to other and a sigmoid transformation(the activation function) was then applied for each unit in a hidden layer.

Although many different approaches exist in order to find the optimal architecture of a neural network, these methods are usually quite complex in nature and are difficult to implement. There is no simple clear-cut method for determination of these parameters and the usual procedure is to test numerous networks with varying numbers of input and hidden units (m,L), estimate generalization error for each and select the network with the lowest generalization error. Once a network structure (m,L) is specified, the network is ready for training a process of parameters estimation. The parameters are estimated such that the cost function of neural network is minimized. Cost function is an overall accuracy criterion such as the following mean squared error:

$$\begin{aligned} E &= \frac{1}{N}\sum_{t=1}^{N}(e_t)^2 \\ &= \frac{1}{N}\sum_{t=1}^{N}\Big(y_t - \big(w_0 + \sum_{j=1}^{L} w_j g(w_{0j} + \sum_{i=1}^{m} w_{ij} x_{it})\big)\Big)^2 \qquad (2) \end{aligned}$$

Where, y_t the output, x_{ti} the inputs, $w_{ij} (i = 0,1,2,\ldots,m, j = 1,2,\ldots,L)$ and $w_j (j = 0,1,2,\ldots,L)$ are model parameters often called connection weights; m is the number of input nodes; L is the number of hidden nodes and N is the number of error terms. This minimization is done with some efficient nonlinear optimization algorithms in which the parameters of the neural network, w_{ij}, are changed by an amount Δw_{ij}, according to the following formula:

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} \qquad (3)$$

Where, the parameter η is the learning rate and $\frac{\partial E}{\partial w_{ij}}$ is the partial derivative of the function E with respect to the weight w_{ij}. This derivative is commonly computed in two passes. In the forward pass, an input vector from the training set is applied to the input units of the network and is propagated through the network, layer by layer, producing the final output. During the backward pass, the output of the network is compared with the desired output and the resulting error is then propagated backward through the network, adjusting the weights accordingly. To speed up the learning process while avoiding the instability of the algorithm introduced a momentum term δ in the above equation, thus obtaining the following learning rule,:

$$\Delta w_{ij}(t+1) = -\eta \frac{\partial E}{\partial w_{ij}} + \delta \Delta w_{ij}(t) \qquad (4)$$

The momentum term may also be helpful to prevent the learning process from being trapped into poor local minima, and is usually chosen in the interval [0; 1]. Finally, the estimated model is evaluated using a separate hold-out sample that is not exposed to the training process.

2.3 The combined SARIMA with neural network model

Maier and Dandy (1996) suggested that the SARIMA model is suited for short-term forecasts, although the neural network model is better suited for long-term forecasts. However, when sample size is small, Liu, Kuo, and Sastri (1995) showed that the neural network model performs better than probability model for the clustering problem. The current paper proposes to combine these two models into a hybrid model to forecast seasonal time series data of foreign exchange. In the previous studies about the BP model (Zhang, 2003; Wang & Meng,2012) their forecasting errors provide feedback to revise the weights, but they do not provide feedback to modify the input variables. Thus, the information lost by the BP model can be very important. Very often, quarterly or monthly seasonal cycles exist in real-world time series. The SARIMA model has been shown to make good forecasts for seasonal time series especially for short-term periods, but it is limitation by the large amount of historical data that is required. The SARIMABP model combines the advantages of the SARIMA and the BP models that input the forecasts and residuals generated by a SARIMA model to the input layer of a BP model and so the latter attempt after the training process will minimize the residuals and improve the forecasts. There exists a problem when a neural network model is used to forecast future outcomes, because future residuals are not yet known. We proposed to use the weighted average of past residuals from the same period in the past as a first guess of the residuals in the forecast period.

3. Forecast evaluation methods

Four criteria of forecasting accuracy can be used to make comparisons of the forecasting capabilities among the SARIMA model and the SARIMABP model:

1. The mean square error (MSE):

$$MSE = \frac{1}{T}\sum_{t=1}^{T}(P_t - Z_t)^2 \quad (5)$$

Where P_t the predicted value is at time t; Z_t is the actual value at time t; and T is the number of predictions.

2. The mean absolute error (MAE):

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|P_t - Z_t| \quad (6)$$

3. The mean absolute percentage error (MAPE):

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}|P_t - Z_t| / Z_t \quad (7)$$

4. Turning point evaluations:

While the above criteria are good measures of the deviations of the predicted values from the actual values, they cannot reflect a model's ability to predict turning points which is critical for traders and analysts to forecast the change of monetary market direction.

There is another evaluation method which involve a model defines a forecast direction variable F_t and an actual direction variable A_t such that

$$A_t = 1 \ \text{if} \ \Delta Z_t \rangle 0 \ \text{and} \ A_t = 0 \ \ \text{if} \ \Delta Z_t \le 0$$

$$F_t = 1 \ \text{if} \ \Delta P_t \rangle 0 \ \text{and} \ F_t = 0 \ \ \text{if} \ \Delta P_t \le 0$$

Where Δ refers to the amount of change in a variable between time t-1 and t. The following regression equation can be established:

$$F_t = \alpha_0 + \alpha_1 A_t + \varepsilon_t \quad (8)$$

Where ε_t is the error term, α_1 is the slope of this linear equation and α_1 should be positive and significantly different from 0 in order to demonstrate those F_t and A_t have a linear relationship. This reflects the ability of a forecasting model to capture the turning points of a time series.

4- Experimental results

In the following, the performance of the SARIMABP model is compared with SARIMA model using the seasonal time series of exchange rate (Egyptian pound / US dollar). The monthly time series data of the exchange rate for the period from January1998 to December 2012 showed strong seasonality and growth trends. The author used a 160-observation data set from January 1998 to April 2011 as the training data, and forecasted for the remained 20 months ; in order to make comparisons, using SPSS program to formulate our models.

4.1 SARIMA model

The first-order regular difference and the first seasonal difference were taken in order to remove the growth trend and the seasonality. The sample (ACF), (PACF), and Akaike Information Criterion (AIC) were used to determine the best model. The model generated is ARIMA (1, 1, 1) $(0, 1, 1)_{12}$ as the following equation:

$$(1-0.774B)(1-B)(1-B^{12})Z_t = (1-0.604B)(1-0.965B^{12})a_t \tag{9}$$

The estimated parameters is presented in table(1)

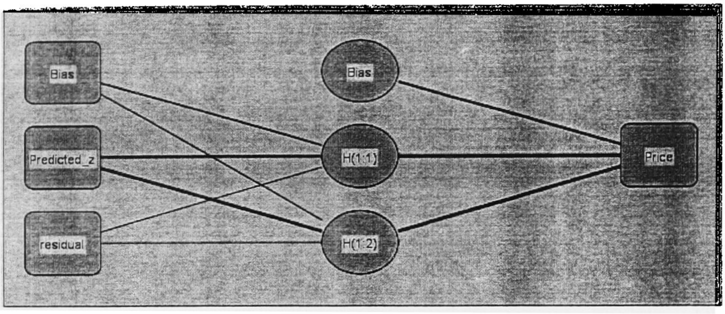
ARIMA Model Parameters		Estimate	SE	t	Sig.
AR	Lag 1	.774	.170	4.558	.000
Difference		1			
MA	Lag 1	.604	.212	2.850	.005
Seasonal Difference		1			
MA, Seasonal	Lag 1	.965	.447	2.156	.033

Table(1): The estimated parameters of ARIMA (1, 1, 1) $(0, 1, 1)_{12}$ model.

4.2 SARIMABP model

The neural network was trained using SARIMA outputs ($\hat{Z}_t, a_t$). First, we choose the best architecture for a neural network model. One hidden layer was set up, and then the input and the hidden neurons of the test time series were examined. The learning rate was initially set at 0.2, and the momentum was set at 0.8. The output layer had one neuron, which was the forecast value. There were two neurons in the input layer of the hybrid model; $\hat{Z}_t$, and residual value a_t (the results from the SARIMA model). The neural network structure is characterized by a network of three layers of simple processing units connected by acyclic links as shown in (Fig. 1).

Fig.(1): Structure of the best fitted network



The results showed that SARIMABP model is superior to the SARIMA model for the test cases of exchange rate time series. The MSE and MAE were the lowest for the SARIMABP model. The SARIMABP model also outperformed the other model in terms of turning points forecasts. The t-ratio of the slope coefficient α_1 of the SARIMABP model shows that it is positive and statistically different from zero, this implies that the SARIMABP model has good turning point forecasting ability. On the other hand, for the SARIMA model, the α_1 parameter is negative and not significantly different from zero. This reflects the fact that SARIMA model could not forecast turning points well. The results of these criteria are reported in Table 2.

model	M S E	M A E	MAPE	Turning point evaluation
SARIMA	111.39	9.683	0.0015	F_t = 1.0 - 0.059 A_t (t-stat.) (6.01) (-0.334)
SARIMABP	1.78	1.012	0.016	F_t = 0.5 + 0.5 A_t (t-stat.) (4.123) (3.9)

Table 2: The results of forecast evaluation methods

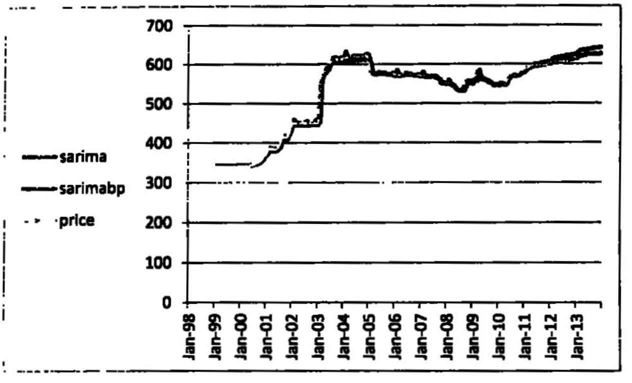
SARIMA model and the hybrid model are used to forecast the exchange rate for the period from January 2013 to December 2013, Table (3) shows these forecasts.

SARIMABP	SARIMA	DATE
613.3	623.41	Jan-13
619.29	631.84	Feb-13
620.48	634.68	Mar-13
621.56	634.83	Apr-13
621.93	635.46	May-13
626.68	637.11	Jun-13
623.56	637.8	Jul-13
623.88	639.29	Aug-13
624.91	640.36	Sep-13
624.68	640.66	Oct-13
624.2	640.59	Nov-13
626.37	643.01	Dec-13

Table (3): The forecasts outside the sample obtained from SARIMA and SARIMABP models.

Finally, Fig.(2) shows the actual time series over the all sample (January 1998- December 2012) and the fitted values obtained from SARIMA and SARIMABP models as well as the forecasts outside the sample (January 2013- December 2013).

Fig. (2):the actual values and the fitted values obtained from SARIMA and the hybrid models.



5- Conclusions

This article introduces a hybrid model for improving the performance of seasonal auto-regressive integrated moving average models. The proposed model based on the Box-Jenkins methodology in its first stage. An SARIMA model is used in order to generate the necessary data; the inputs for the neural network, and then a neural network is used to determine a model in order to capture the underlying data generating process and predict the future using pre-processed data. Empirical results indicate that the proposed (SARIMABP) model can be an effective way in order to yield more accurate model than traditional Box-Jenkins (SARIMA) approach. Thus, it can be used as an appropriate alternative for SARIMA model when higher forecasting accuracy is needed specially for financial data.

References

1- Ansuj, A.P., Camargo, M.E., Radharamanan, R., Petry, D.G. (1996):" Sales forecasting using time series and neural networks." Comput. Ind. Eng. 31 (1), 421-424.

2- Box , G.P., Jenkins, G.M. (1976), Time Series Analysis: Forecasting and Control, Holden-Day, New Jersey.

3- Cabarkapa , S., Kojic, N., Savic, A., Zivkoviv, B.(2010):" use of artificial neural networks in financial time series prediction and financial risk prediction." Infoteh-Jahorina vol.q,Ref.E-I I I-6,P.584-587.

4- Caire, D., Hatabian, G., Muller, C. (1992):" Progress in forecasting by neural networks." Int. Conf. Neural Networks II, 540-545.

5- Dhamija, A., Bhalla, V. (2010):" Financial time series forecasting : comparison of neural networks and ARCH Models." International research journal of finance and Economics- Issue 49.

6- Diaz, L.A., chow, J.C., Watson, j.g. , Moncada, J.A. (2008):" A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco-chile." Atmospheric Environment 42, 8331-8340.

7- Hill, T., O'Connor, M., Remus, W. (1996):" Neural network models for time series forecasts." Manage. Sci. 42 (7), 1082-1092.

8- Jiuran, H., Bingfeng, S.(2013):" The application of ARIMA-RBF model in urban rail traffic volume forecast." The 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE).

9- Khashei, M., Bijari, M. (2010):" An artificial neural network (p, d, q) model for time series forecasting." Expert Systems with Applications 37, 479-489.

10- Khashei, M., Bijari, M. (2011):" Exchange rate forecasting better with hybrid artificial neural networks models." J. Math. Comput. Sci. 1, No.1, 103-125.

11- Kohazdi, N., Boyd, M.E., Kermanshahi, B., Kaastra, I. (1996):" A comparison of artificial neural network and time series model for forecasting commodity prices." Neurocomputing 10 (2), 169-181.

12- Liu, M.C., Kuo, W., Sastri, T., An exploratory study of a neural network approach for reliability data analysis, Qual. Reliab. Eng. Int. 11 (1995) 107-112.

13- Maier, H.R., Dandy, G.C. (1996):" Neural network models for forecasting univariate time series." Neural Networks World 6 (5), 747-772.

14- Pao, H.T. (2008) :" A comparison of neural network and multiple regression analysis in modeling capital structure." Expert systems with Applications 35, 720-727.

15- Su, C.T., Tong, L.I., Leou, C.M. (1997):" Combination of time series and neural network for reliability forecasting modeling." I. Chin. Inst. Ind. Eng. 14 (40), 419-429.

16- Tseng, F.M., Yu, H.C., Tzeng, G.H. (2002):" Combining neural network model with seasonal time series ARIMA model." Technological Forecasting and Social Change 69, 71-87.

17- Voort, V.D., Dougherty, M., Watson, M. (1996):" Combinig Kohonen maps with ARIMA time series models to forecast traffic flow." Transp. Res. Circ. (Emerge Technol.) 4C.(5) ,307-318.

18- Wang, J.h., Leu, J.Y. (1996):" Stock market trend prediction using ARIMA-based neural networks." IEEE Int. Conf. Neural Networks, 4 (6), 2160-2165.

19- Wang, X., Meng, M.(2012):" A hybrid neural network and ARIMA model for energy consumption forecasting." Journal of computers, Vol. 7, No. 5.

20- Zhang, G.P. (2003):" Time series forecasting using a hybrid ARIMA and neural network model." Neurocomputing 50, 159-175.

21- Zhang, G.P., Qi, G. M. (2005):" Neural network forecasting for seasonal and trend time series." European Journal of Operational Research 160, 501-514.

22- Zou, H.F., Xia, G.P., Yang, F.T., Wang, H.Y.(2007)"An investigation and comparison of artificial neural network and time series models for Chines food grain price forecasting." Neurocomputing 70, 2913-2923.