مجلة التجارة والتمويل

https://caf.journals.ekb.eg/

كلية التجارة – جامعة طنطا

العدد : الثالث

سبتمر 2023

الجزء الثانى

جامعة طنطا

كلية التجارة

قسم الإحصاء والرياضة والتأمين

Hybrid Model "ARIMA -ANN" Using for forecasting Stock Index EGX30

إعداد

محمد سامي حلمي المالكي

Mohamed.elmalky@commerce.tanta.edu.eg

تحت إشراف

دكتور

مي محمد كامل

maie.kamel@commerce.tanta.edu.eg

دكتور

وجدي محمد كامل

Wagdi.abdelnabi@commerce.tanta.edu.eg

2023

Abstract

This research aims to evaluate the efficiency of Hybrid model ARIMA-ANN in forecasting by Stock market Index EGX30 since 3/1/2022 to 9/1/2022. In this research discussion three models which are Autoregressive Integrated Moving Average (ARIMA) , Artificial Neural Network (ANN) and the Hybrid model (ARIMA-ANN) , while ARIMA(0,1,1) has been used to estimate the linear Part of Model ,then Estimating the Non- linear Part of Model by the difference between Actual data and Estimated data of series, so the model of ANN (2,5,1) has been used to estimate the non-linear part of the model and by collecting the two Parts for getting finally the hybrid model for forecasting Processing , After comparing three models and based on standard group such as Mean Square Error (MSE) , Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) ,We achieved that " the Hybrid model (ARIMA-ANN) was the best model in forecasting by stock index EGX30 and it is better than ARIMA (0,1,1) and ANN (2,5,1) which did singularly ,that is because Hybrid Model has the minimum accurately values of forecasting standards .

Keywords: Forecasting, Stock Index EGX30, Hybrid Model, ARIMA, ANN

1-Introduction

There are several models which used in time series forecasting including traditional models such as Box and Jenkins models (ARIMA) which depend on essential assumption is time series must be linear , while the most of time series such as economic and financial series consist two parts (The linear and the non- linear parts) so these traditional models became unable to forecast so it is important to use some models able to do with non-linear part of the series such as artificial Neural Network (ANN).

From other hand, the merge between the traditional models (ARIMA) model and the modern model (ANN) by hybrid model (ARIMA- ANN) is so important to be able to work with two parts and to obtain the best results comparing with the results of an individual model.

While The Egyptian stock index (EGX30) is one of the best pricing models which measure the performance of the thirty highest companies of liquidity in The Egyptian stock market. (Al-Wasifi2012)

Egyptian stock market is continuing in improvement to EGX30 index to diversity index sectors , the stock market committee depended on adding and removing new standards for companies index to be identical with involving rules , some of these rules is determining the minimum of Company capital for putting it in companies index and also determine the maximum companies in the same sector , that is for a real reason which keeping with its attractive for investment so it is so important to arrange the companies within their liquidity .In fact, applied (Buffer rule)which consider one of the most followed methodologies which increase a stability of index ingredients , In the most major institutions make two reviews for the index by the Egyptian stock market , The first review index finishes in January and start to apply it from 1/7 to 31/12 and the second review finishes in July and start to apply it from 1/1 to 30/6 so the forecasting for EGX30 became an important tool for registered companies in this index specially and in the Egyptian stock market generally so different strategies were used to forecast by stock market index for helping the decision makers and achieving to the best results .

2-Review literature:

El-Waseefy(2012) This study interested in achieving to the best model in forecasting for stock market index EGX30 so applied to (ANN) and (ARIMA) then the merge between them by using the residuals from (ARIMA) and was applied to (ANN) and made a comparison between three models (ARIMA) ,(ANN) and (ARIMA-ANN) , After forecasting by three models and using forecasting quality standards such as MSE , MAE and MAPE , Finally achieved to the best model from three models is (ARIMA-ANN) for forecasting by EGX30 .**Babu As, Reddy Sk., (2015)** This study shared with explanation of daily exchange rate behavior for Indian Rupee against Dollar ,Sterling Pound , Euro and Japanese Yen , a study analyzing by (ARIMA) and (ANN) for forecasting in the traded currencies so was used daily exchange rate from January 2010 to April 2015 , Finally achieved that (ANN) model was better than ARIMA in forecasting .**Hill and Remus (2018)** This study helped in a comparison between artificial Neural Network (ANN) and Box and Jenkins (ARIMA) in forecasting and applied to Annual , quarterly and monthly date , then it was measured averages and standard deviation and achieved to (ANN) model was better than (ARIMA) because (ANN)model got averages were better than ARIMA within a significance. **El-Fadel and et al (2019)** This research consists the comparison between frequency domain Analysis by the Fourier series analysis and time analysis by ARIMA models , it was applied using Gas sales data in the United States of America from 1993-2014.The aim from analysis is checking the cycle of variability and estimate the best model of data . it was achieved the cycle of variability was achieved at 135 months and forecasted for 6 months. it obtained by time series analysis that the suitable model for these data is ARMA (1,0,1) and achieved the Fourier series analysis is near Time series analysis and both have a prominent level of accuracy by group of models estimating standards such as MSE, MAE and MAPE. **El-Sadye (2019)** Contributed this research in a comparison between Box and Jenkins model and the ANN model in forecasting by (Iraq trade balance) from 2003 to 2016 and targeted forecasting it because of consideration (Iraq trade balance) is one of the political tools used in effect on the economic

activity and covering state expenses, The problem lies in the forecasting by traditional methodologies may give inaccurate results in the small samples , while the between ARIMA as a example of traditional methodologies and (ANN) ,This research achieved to ANN is better than ARIMA in forecasting with available data. **El-Zayan(2019)** This study targeted forecasting for the future values of Euro Price series against shekels so depended on three models ARIMA (2,1,1), GAR (1,1) and ANN (2,6,1) so the comparison between three models with forecasting quality and efficiency by some standards such as MSE , MAPE and the statistical results obtained the (ANN) is the best Model so depends on it for forecasting by the forecasting by the future values of Euro price series against shekels .**Ghattfan and Eskander (2019)** This research contributed in a comparison between individual models ARIMA , ANN and the merge model (ARIMA-ANN) assume the series of data consist (linear and non- linear) parts , which the linear part was measured by ARIMA model and non- linear part was measured by the ANN model so measured ARIMA (1,1,1) and ANN (1,26,1) for the minimum temperature and measured ARIMA (0,1,2) and ANN (1,17,1) for the maximum temperature , this research achieved in the best model for applied data was (ARIMA-ANN) because had the smallest values of MSE, MAE and MAPE. **Mohamed and Mousa(2020)**This research was used the hybrid models for time series has two variables ,so it was used ARMAX model to analyze the linear part of data ,as well as it was used (ARCH , GARCH , GARCHX) to analyze the non-linear part of data ،It was applied time series of External variables (Exchange Rate) x_t and essentials time series (Unemployment Rate) y_t ، From jan,2000 to Dec,2017 by 216 observations ، so the aim of this study is forecasting for (Unemployment Rate) for 2018 means 12 periods. It was obtained by analyzing the ability and sufficiency of ARMAX (2,1,1,0) GARCHX (1,1,1) comparison with ARMAX (2,1,1,0) and ARMAX (2,1,1,0) GARCH (1,1) based on evaluating models standards MAE, MAPE and (Q-Like). **Abdul Qader (2022)** analyzed the behavior of the time series of Gross fixed capital formation in the Kingdom of Saudi Arabia in the time from 1969 to 2019.The study showed that the time series of Gross fixed capital formation can be represented through

ARIMA(2,1,4)through the evaluation criteria of MSE , MAE and MAPE models and used to predict the data for the years from 2020 to 2023, and the results showed that the fixed capital index in the Kngdom of Saudi Arabia is constantly increasing for the predicted years from 2020 to 2022compared the actual values of the variable of interest . **Fawzy and Alrweili (2022)** study for oil prices forecasting by using the Hybrid models and applied for the Kingdom Saudi Arabia from the period January 2001 to May 2021 for 239 observations. The first 215 observations was used to training for test the next 24 observations in the series , finally achieved ARIMA(1,1,0) and ANN(1:10:1) , while the Hybrid model ARIMA-ANN(1,1,0)(1:18:1) was the best model among three models based on MSE, MAE and MAPE . **Othman and Ismail (2022)** aimed at evaluating the ability of multi-layered neural network models to improve the accuracy of forecasting credit risks facing business establishments compared to logistical analysis and the observations credit 824. This study concluded that the multi –layered neural network models outperformed the logistic regression using financial indicators in predicting credit risks, as the highest level of prediction accuracy for the neural networks reached 86.67% compared to 85.19% for the logistic regression models. Acceptance of the first main hypothesis that the use of multi-layered neural networks leads to an improvement in the ability of financial indicators to predict the credit risks of business enterprises, as well as the acceptance of the second main hypothesis that there is an effect of the optimal combination of ratios and financial indicators on Accuracy of multi-layered neural networks for predicting. **Alam et al (2023)** A study on prediction of oil , coal and natural gas Prices for the period (2022 -2025) compared to prices compared with prices before and after the Covid 19 by applying to the data of the country of India by extending the data for the period from January 2020 to May 2022 , using different methodologies in the analysis and they are the style Box and Jenkins (ARIMA) , SEC methodology and ANN using the nearest neighbor learning methodology . This study was achieved the best model for forecasting for these data was ARIMA (0,1,1) based on criteria AIC and BIC.

3-Research Problem

The most of time series such as economic and financial series consist of two parts (the linear and the non- linear parts) so these traditional models such as Box and Jenkins models became unable to forecast so it is important to use some models able to do with non-linear part of the series such as artificial Neural Network (ANN) , So the merge between the traditional models (ARIMA) model and the modern model (ANN) by hybrid model (ARIMA- ANN) is so important to be able to work with two parts and to obtain the best results comparing with the results of an individual model**.**

4-Research Objectives

Research has an essential goal which is using a hybrid model (ARIMA-ANN) in forecasting by stock market index **EGX30** so has branch goals**,**

1- Making ARIMA model to forecast **EGX30** index**.**

2- Making ANN model to forecast **EGX30** index**.**

3- Making (ARIMA-ANN) model to forecast **EGX30** index**.**

4- A comparison between three models **(ARIMA), (ANN) and (ARIMA-ANN)**

for forecasting by three models and using forecasting quality standardssuch

as **MSE, MAE MAPE, AIC and BIC.**

5- **Research limits**

This research consists of time and places limits,

- Places limits, Egyptian stock market index EGX30.

-Time limits, based on EGX30 data from 2/1/2014 to 2/1/2022.

- Data sources,

Central Agency for Public Mobilization and Statistics (CAPMAS) (capmas.gov.eg)

The Egyptian Exchange (egx.com.eg)

6 -Methodologies

This research consists of three methods to forecast time series which are ARIMA model, ANN model and (ARIMA-ANN) model.

6-1- ARIMA Models: (Box and Jenkins)

The models were published by Box and Jenkins in 1970 and there are the most methods are used to analyze the time series, when we talk about Box and Jenkins, we mean models family which is called **A**uto-**R**egressive and **I**ntegrated **M**oving **A**verages (ARIMA (p, d, q)) and can be expressed in this model by this function. (Box G., and Jenkins G. 1970)

$$X_t = \zeta + \sum_{i=1}^{p} \varphi_i X_{t-i} + a_t - \sum_{j=1}^{q} \theta_j a_{t-j}$$

$$or \qquad (1)$$

$$Y_t = \sum_{i=1}^{p} \varphi_i Y_{t-i} + a_t - \sum_{j=1}^{q} \theta_j a_{t-j}$$

Where: X_t: refers to the actual stable observations of the time series of interest.

ζ: a fixed quantity equal to $u(1 - \sum_{i=1}^{p} \emptyset_i)$ for stationary serieswith mean equal u.

Y_t: shows the deviations of the observation from the mean**.**

ϕ_i: Shows Auto Regressive parameter.

θ_i: Shows moving averages parameter.

P: refers to Auto Regressive rank.

Q: refers to moving averages rank.

d: refers to the order of differences necessary to stabilize the series.

a_t: refers to a random variable follows a normal distribution with expected value equal zero and variance equal σ_a^2 . $a_t \sim$ i.i.d.N(0, σ_a^2 **)**

ARIMA Models are based on four stages which are Identification stage, Estimation Stage, checking model Stage and Forecasting model Stage.

1- Identification Stage. 2- Estimation Stage.
3- Diagnostic Checking Stage. 4-Forecasting Stage.

1- Identification stage :
At this stage must be confirmed that the data is stable , there are no unit roots and this is done by relying on the Autocorrelation function (ACF) and Partial Autocorrelation function (PACF) ,where unstable data leads to an increase forecasting error, After that the parameters determined by Autocorrelation and Partial Autocorrelation functions, graphs, Estimation of model parameters and follow ordinary least square method or Maximum Likelihood method and for the sake of comparison depends on the model that has the less values of mean square error MSE,MAE,MAPE,AIC,BIC and for model fit the residual series must be follow the normal distribution by expected value equal zero and constant variance for forecasting.
It can summary the behavior of Autocorrelation and Partial Autocorrelation functions in non-seasonal models as shown in table (1).

Table (1) Vandal-Walter (1992)
Determine the nonseasonal model by using ACF and PACF

Model	ACF	PACF
AR(P)	Dies down	Cuts off after lag
MA(q)	Cuts off after lag	Dies down
ARMA (p,	Dies down	Dies down

And it can summary the behavior of Autocorrelation and Partial Autocorrelation functions in seasonal models as shown in table (2).

Table (2) Vandal-Walter (1992)
Determine the seasonal model by using ACF and PACF

Model	ACF	PACF
AR ($p+ S_p$)	Dies down	Cut off after lag $p+S_p$
MA ($q+ S_q$)	Cut off after lag $q+S_q$	Dies down
ARMA ($p+ S_p, q+$ S	Dies down	Dies down

2-Estimation Stage:

After initial recognition models of the time series of interest must be estimated vector parameters of the proposed models. The method of Least Squares and Maximum Likelihood method are two basic appreciations. Estimate the model and check its parameters by the significance test for each parameter using T test and be sure about stability and reflections conditions for the estimated parameters.

3- Diagnostic Checking Stage:

Box and Jenkins suggested groups to look at and test the relevance of the models and these groups are:

1-analysis of the stationary state of control of the stationarity and reflection in the capabilities of the models proposed, proof of the inadequacy of the data models.

2- residuals analysis.

By analyzing residuals model for sure about are white noise or no by checking for the residuals Autocorrelation and Partial Autocorrelation Functions , Then calculate the first differences for the residuals and checking its ACF and PACF for Exponential and the Partial Autocorrelation coefficient equal -0.5 or near from it . There are two tests proposed for this purpose:

(1) statistical Box-Pierce $Q(k) = n\sum_{i=1}^{n} r_i^2\ (\hat{a})$

(2) Statistics Ljung-Box $Q(k) = n(n+2)\sum_{i=1}^{n} \frac{1}{n-i} r_i^2\ (\hat{a})$

Where: n is the number of hits of the time series after taking the necessary differences housed.

$r_i^2(\hat{a})$ Autocorrelation coefficient box overstocks at period i.

This follows the statistical distribution of Chi square degrees of freedom equal to the number of autocorrelation coefficients minus the number of estimated parameters of the form,

4- Forecasting Stage:

The prediction using information available at the time sequences associated with the predicted time string worked to improve the quality of those predictions.

6-2 ANN Model:

using of neural networks goes back to Me Culloch and Pitts in 1943, when the human brain consists of a group of neurons called neurons, and these cells or neurons spread in groups called networks, where each

network is composed of several thousand neurons connected or like each other. These cells enable the brain to perform its functions such as thinking, remembering and others.

In addition to neuron which is a processing unit that has a nucleus in the middle has some neural endings and it manages the inputs of the cell. This cellular also has a conductor responsible for the outputs, and these endings are fused with the neural endings of the second cell in what is known as synaptic transmission, and the signal is transmitted from one neuron to another. This signal through electrochemical interactions, and the brain through these connections and interactions, processes information in parallel, this is at the same moment.

Artificial neural networks consist of neurons or processing units connected to form a network of nodes. Each connection between these nodes has a set value called weights that contribute to decide the values resulting from each processing element which is based on the entered values. For this element, the input units are under a layer called the input layer, and the processing units are form the processing layer, which outputs the outputs of the network, and hidden layers are between each of these layers that connect each layer to the next layer, and the network contains only one layer of input units but it may contain more than one layer of processing layers, and mathematical functions are applied to calculate and process the values, then output the outputs and pass it to the other nodes .

The following figure (1) shows the comparison between the natural neural network and artificial neural network. (Basiouny 2008)

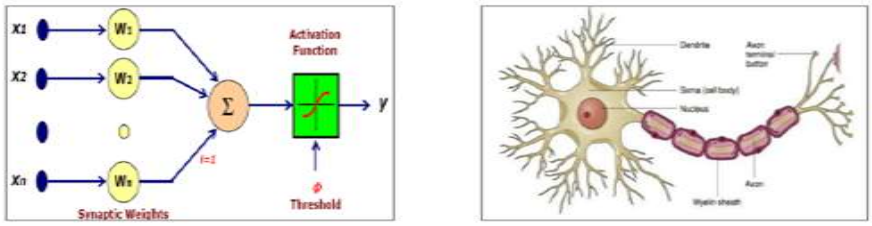

Figure (1) Comparison between natural neural network and artificial neural network

Components of an artificial neural network, A neural network consists of three main parts consider them as follows,

1- The inputs , we assume that the values of the outputs of neurons previous to the neuron (the current neuron are the values (X_1 , X_2 ,X_3,…….,X_n)where the synapses transfer a percentage of these values to the input neural fiber through gaps so that these ratios differ from gap to gap others, depending on the nature of each gap , and by representing that mathematically , assuming that they are weights , which take the form (W_1,W_2,W_3,……,W_n) and thus the part transferred to the entrance in each branch is (WX) **.(Issa 2000)**

2- Cell body or processing element, the cell body does the following equation,

A- Weighted sum,

$$\textbf{Total input} = W_1 X_1 + W_2 X_2 + W_3 X_3 + \ldots\ldots + W_n X_n \quad (2)$$

B-Comparing the sum with a boundary value known as the Heaviside function , if the weighted inputs are more than or equal to the boundary value of the function , The outputs will be equal to **one** and if the weighted inputs are less than the boundary value of the function , The outputs will be equal to **zero.**

3-The outputs, it can be either one or zero and can be used as one of the input branches of other neurons and can be written in the following form.

$$y = F\left(\sum_{i=1}^{n} W_i X_i\right) \quad (3)$$

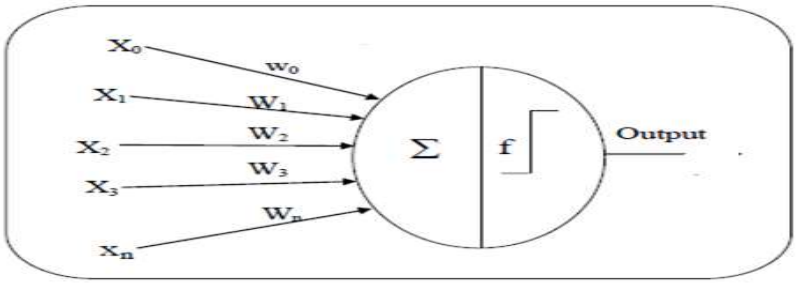

Figure (2) Mathematical model of neural cell
(Source, Mohamed Al-Sharqawy 1995)

-Hybrid model (ARIMA-ANN): (Zhang 2003)

The idea of merging between the time series method and the neural networks method has become important, as there is no single model alone capable of predicting with high accuracy. And this comes from the shortcomings that exist in each model, from here comes the idea of merging with its justifications, which is that each model tries to eliminate the shortcomings in the order model, which leads to a decrease in the negative impact resulting from using a model that may not be the best and is in the following function,

$$y_t = L_t + N_t \qquad (4)$$

Where: y_t , the time series.
L_t , the linear part of time series from ARIMA Model.
N_t , the non- linear part of time series from ANN Model

6-3 Hybrid model (ARIMA-ANN) can be built According to the following steps, (Zhang 2003)

The residuals of ARIMA model are used to build the neural network According to the following steps,

1- Building an ARIMA Model for the series (y_t) under study using Box and Jenkins method.
2- Obtaining the predicted values from the ARIMA Model to be the linear part of the series.
3- Building ANN Model for residual modeling based on the residuals from ARIMA.

$$e_t = y_t - \hat{L}_t \quad (5)$$

4- Calculate its future predictions.

$$e_t = f(e_{t-1}, e_{t-2}, \dots, e_n) + \varepsilon_t \qquad (6)$$

Where is,

f, the non-linear function that is determined by the neural network.

ε_t, a random error and neglected it we get the estimated values.

$$\hat{e}_t = f(e_{t-1}, e_{t-2}, \dots, e_n) \qquad (7)$$

5- Obtaining $\hat{y}_t$ by combining predictions from ARIMA Model with predictions from ANN Model,

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \qquad (8)$$

7 -Results and Discussion:

In this research, the data of the Egyptian stock index EGX30 were used, which can be summarized as in the following table.

Table (3): Descriptive Statistics for the time series data

Data Interval	**From 2/1/2014 to 2/1/2022**
Data Size	**1951 Observations**
Mean of Data	**11238.3**
Standard deviation	**2883.35**
Maximum value	**18363.3**
Minimum value	**5713.3**
Skweness Coefficient	**0.122**
Std of Skweness Coefficient	**0.055**
Kurtosis Coefficient	**-0.971**
Std of Kurtosis Coefficient	**0.111**

Analysis using different analytical methods, the research included application using three different methods in time series analysis as follows:

1- **Analysis by ARIMA (p, d, q) model,** this way is used by 4 steps as follows,

1-1- Identification of Model:

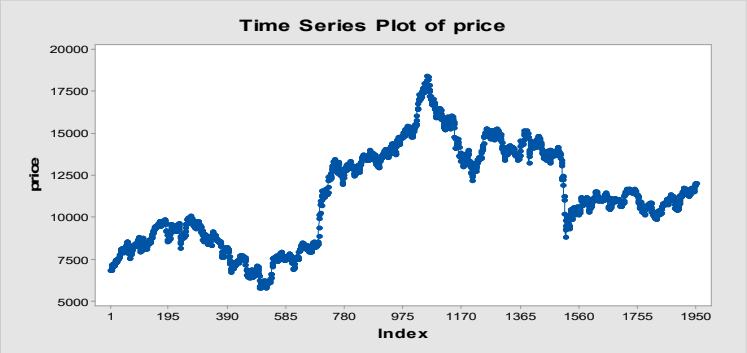

Figure (3) The graphic of time series

From the previous chart, the variance and the mean of the time series are not stable, which requires taking proper procedures to install both of them.

1- variance stabilization, many transformations were tried to stabilize the variance, and the researchers settled on taking LAG transformation for the data, as this transformation achieved the stability of the time series.

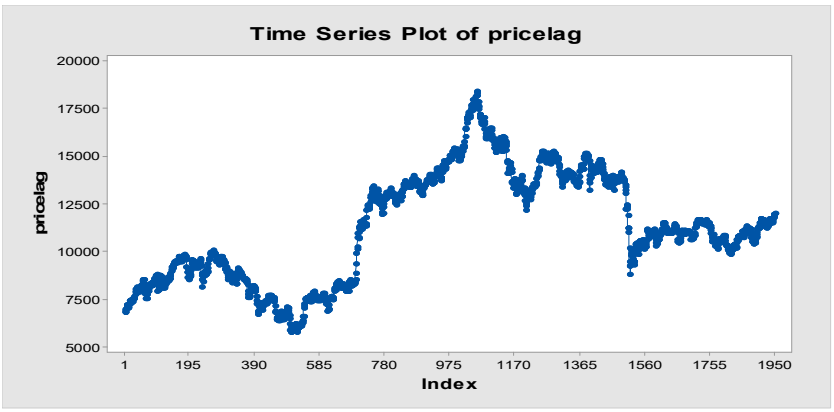

Figure (4) The graphic of of data after taking LAG transformation

By drawing the time series after taking LAG transformation, there is instability in the meaning of the series which is necessary taking the appropriate to install it.

2- Mean stabilization: After taking the transformation of the series data, the first difference of the time series was taken, and there is no benefit from taking the second or the third differences, so it is enough to take only the first difference.

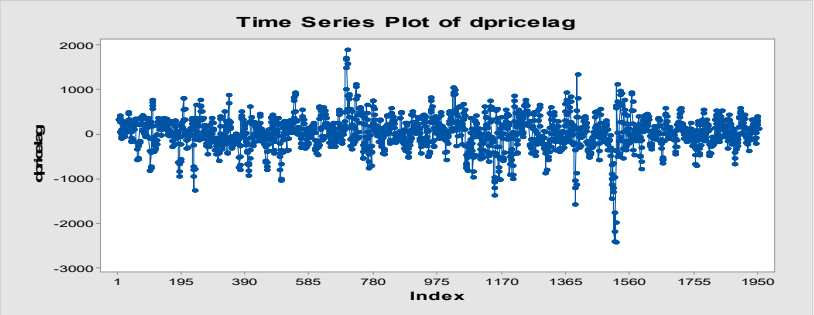

Figure (5) The graphic of the first difference of data after taking LAG transformation

The second and third differences of the time series were evaluated, but they did not lead to mean stabilization, so the first difference was sufficient.

3- Autocorrelation and partial autocorrelation function Graphis:

3-1- Autocorrelation function:

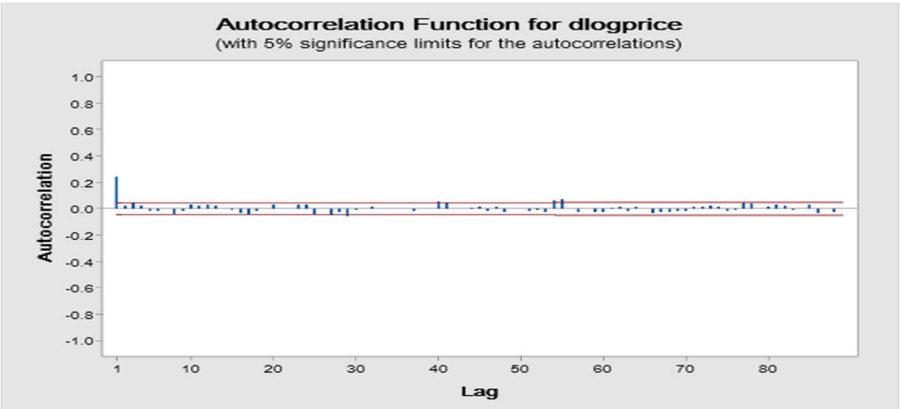

Figure (6) The graphic of the Autocorrelation function of the first difference of data after taking LAG transformation

Through the graphic of Autocorrelation function, the totality is decreasing and gradually approaching zero and taking an exponential form.

3-2- Partial Autocorrelation function:

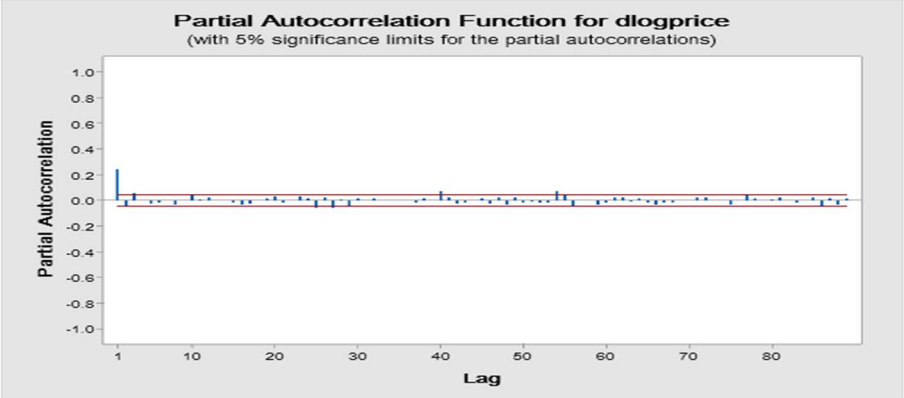

Figure (7) The graphic of the partial Autocorrelation function of the first difference of data after taking LAG transformation

From the graphic of the partial Autocorrelation function, the totality is decreasing and gradually approaching zero and taking an exponential form. So, it can be conclusion that the time series data may describe by ARIMA (0,1,1).

1-2- Model Estimation: after identification of model, can estimate its parameters, and the method of least square is one of the most reliable methods to estimate the parameters of the model.

Table (4): Estimated Model of data

Model	Estimate	S.E Error	t –ratio	P. v
ARIMA (0,1,1)	a_t = 0.0001	0.0002	0.80	0.422
	Θ_1= -0. 2535	0.0219	-11.57	0.00

The estimated model has been reached so that its equation is as follows,

$$y_t = 0.0001 + 0.2535\ e_{t-1} \qquad (9)$$

From previous results, it is so clear that P-value of the moving average coefficient is significant and the changes that occur in the model only purely random changes, and this clear from the results of the following table.

Table (5): the Values of $\mathcal{K}^2$

Lag	12	24	36	48
$\mathcal{K}^2$	10.78	20.59	39.38	53.67
DF	10	22	34	46
P. V	0.375	0.546	0.242	0.204

From $\mathcal{K}^2$ statistic to evaluate that the residuals of estimated model **a**re only purely random changes, where the value of P.V is more than Significant level 5%, so the residual of estimated model is only purely random changes.

1-3- Model checking: The third stage is checking the estimated model and must be sure about some characteristics such as,

1-3-1-Stability and Reflection Checking:

From Schedule (1), it obtains that the parameter of estimated model investigates Stability and reflection conditions because its value less than one.

1-3-2- Residuals Analysis:

The residuals were evaluated by residuals function graph and graph Autocorrelation and partial Autocorrelation functions for the residuals for sure it is only purely random changes.

A- **The graphic of the residuals**:

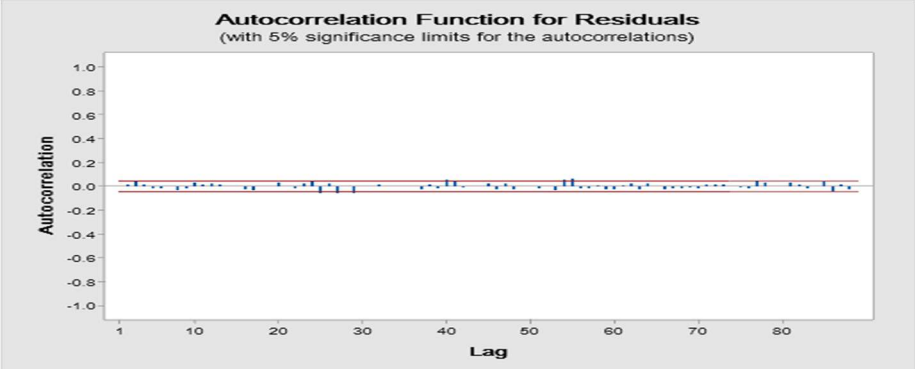

Figure (8)The graphic of the residuals

B- **The graphic of the autocorrelation function of residuals:**

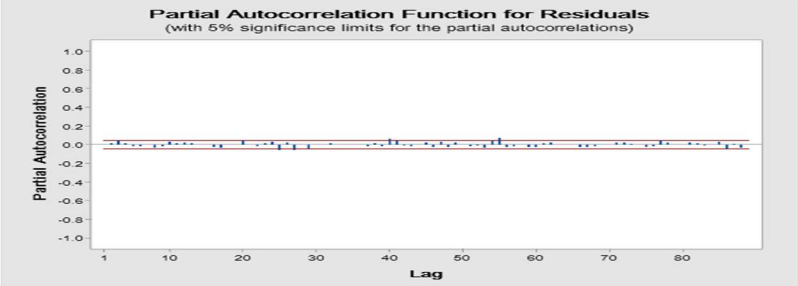

figure (9)The graphic of the autocorrelation function of residuals

C- **The graphic of the partial autocorrelation function of residuals**:

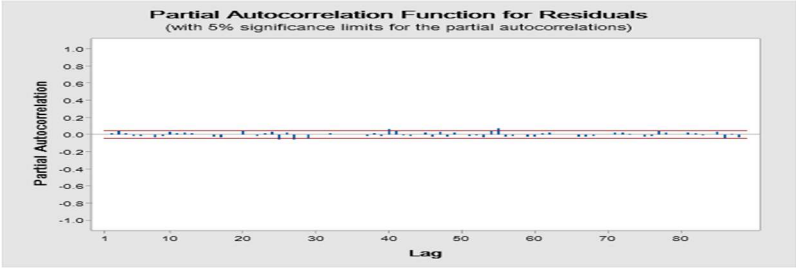

figure (10) The graphic of the partial autocorrelation function of residuals

It obtains by the graphic of Autocorrelation and partial autocorrelation functions that the residual of estimated model is only Purely random changes, then calculate and graph residuals differences.

D- The Residuals differences:

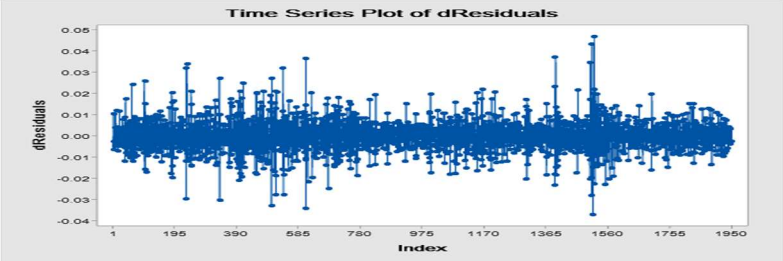

figure (11) The graphic of the first difference of residuals

E- Autocorrelation function of the Residuals differences**:**

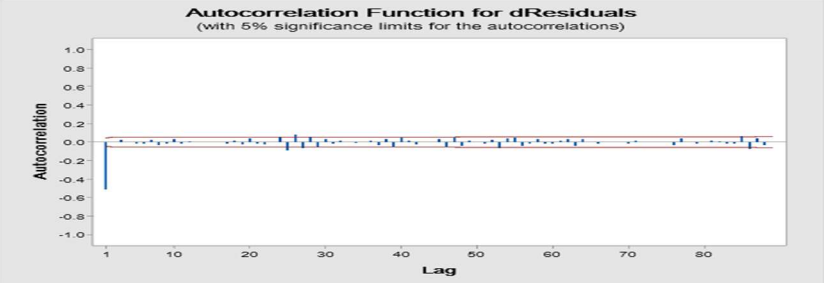

figure (12) The graphic of Autocorrelation function of the first difference of residuals

F- Partial Autocorrelation function of the Residuals differences**:**

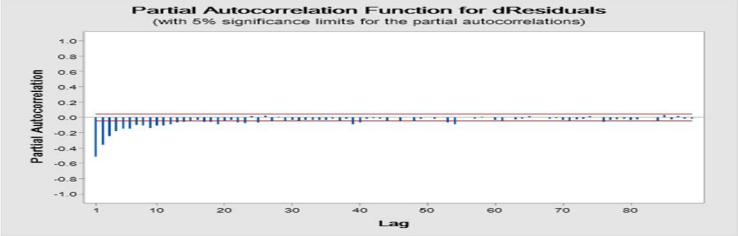

figure (13) The graphic of partial Autocorrelation function of the first difference of residuals

And it obtains by the graphic of Autocorrelation and partial autocorrelation functions of residuals differences, there are.

1- Autocorrelation coefficients are near -0.5

2- Partial Autocorrelation coefficient decreasing and gradually approaching zero and taking an exponential form.

3-The model for the residuals of model follows MA (1) by parameter equal 0.99.

From the Previous results obtained the ability of estimated model ARIMA (0,1,1) for forecasting.

1-4- forecasting, after estimation of the model and checking it for sure about adaptability for time series of EGX30 so the model function is in this form,

$$y_t = 0.0001 + 0.2535\ e_{t-1} \qquad (10)$$

Table (6): comparison between Actual and forecasted values by using ARIMA (0,1,1)

Date	Observed Values	Forecasted Values	Difference Percent
3/1/2022	11897.4	11954.3	0.478%
4/1/2022	11982.6	11957.8	%-0.207
5/1/2022	12043.4	11961.3	-0.682%
6/1/2022	11978.6	11964.7	-0.116%
9/1/2022	12022.7	11968.2	-0.453%

Table (7): Standards of evaluation accuracy of model

	MSE	MAE	MAPE	AIC	BIC
ARIMA (0,1,1)	52.4529	46.44	0.3872	1654.26	1656.84

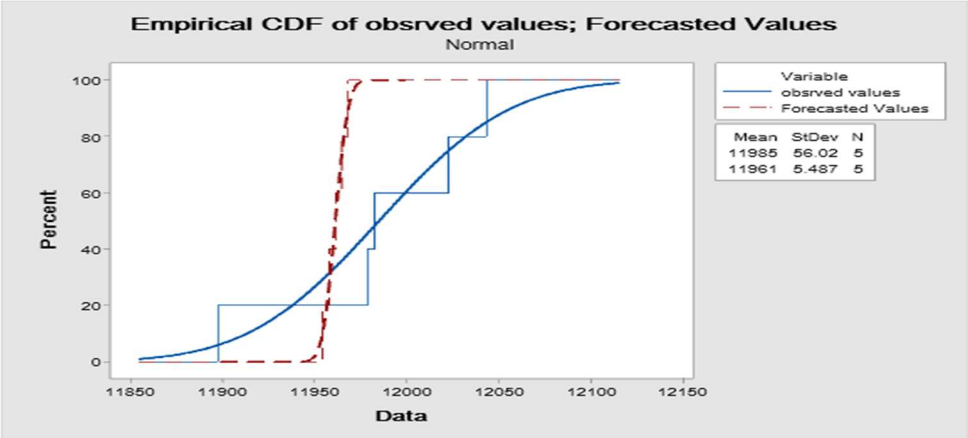

Figure (14) The graphic of actual and forecasted Values of EGX30 by using ARIMA (0,1,1)

2- Analysis by Artificial Neural Networks:

On the other hand, there is another method for forecast which is called Artificial Neural Networks while it can explain it by these steps,

1- Input layer, the data represented in the past week and the previous EGX30stock index values.

2- Hidden layer consists of operating elements.

3- Output layer, it is represented in the current indicator values.

Cross validation:

This method was used to learn the network by using 90% of the network size for training and 10% for testing.

Back Propagation learning method:

This method is considered one of the most famous learning methods in multi- layered and supervised neural networks. This method was discovered by (Paul **and Robbers**) in 1974, and the network relies on defining it with the required input and output patterns, the work of this method is to reduce the sum of the squares of the output errors which calculated through the network gradually.

It can be represented the functions are used in the back propagation learning method as follows,

$$X_i = \sum Y_i W_{ij} \qquad (11)$$

$$F(y)_i = \frac{1}{1 + e^{-X_j}} \quad (12)$$

Where is,
X_i, the sum of outputs which is the result of the collection process.
$\mathbf{F(y)}$, is the active function in the past layer.
$\mathbf{W_{ij}}$, link weight between the two units i, j.

And the neural network measures the error by calculate the outputs of all units as follows,

$$E = \frac{1}{2} \sum (y_i - d_j)^2 \quad (13)$$

Where is,
Yi , unit output of layer i
d_i , the real output

Adam Optimizer: It was used for increasing level of accuracy for much of data since 2014 and gave more accuracy for Back **Propagation learning method, while** input layer represented in the past week and the previous EGX30 stock index values, Hidden layer which consists of operating elements and Output layer represented in price of EGX30.
Finally, it has been concluded that the model followed by the neural network's method is ANN (2,5,1) model, which means there are two input layers, five hidden layers and one output layer represented in the predicted values of EGX30.

Table (8): comparison between Actual and forecasted values by using ANN (2:5:1)

Date	Observed values	Forecast values	Difference Percent
3-1-2022	11897.4	11913.84	0.138%
4-1-2022	11982.6	11921.94	-0.506%
5-1-2022	12043.4	11953.55	-0.746%
6-1-2022	11978.6	12009.51	0.258%
9-1-2022	12022.7	11981.97	-0.339%

Table (9): Standards of evaluation accuracy of model:

	MSE	MAE	MAPE	AIC	BIC
ANN (2,5 ,1)	237.915	178.484	0.0178	1711.52	1714.10

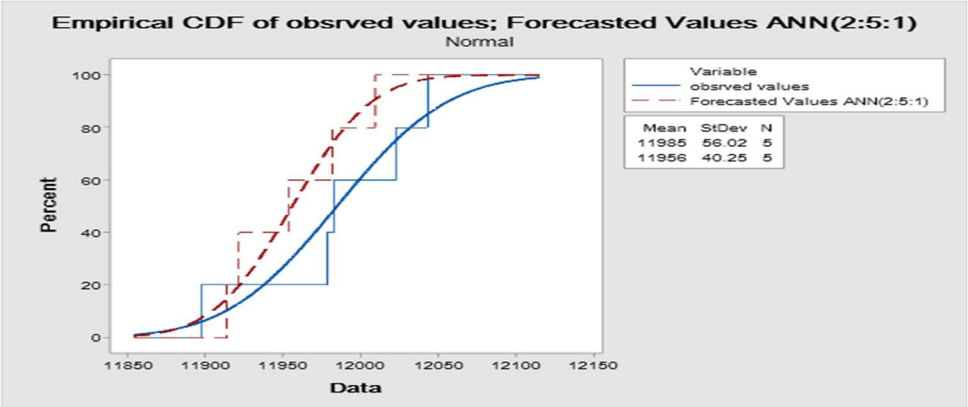

Figure (15) The graphic of actual and forecasted Values of EGX30 by using ANN (2:5:1)

3-Analysis by Hybrid model:

After completing the analysis of data using Box and Jenkins models ,as well as the analysis through the method of Artificial neural networks separately, The hybrid models method is based on merging the two previous methods through one of the merging methods ,through ARIMA-ANN model where it is reached the best fit model is through the time series model to obtain the linear part , while the neural networks method is applied to the non-linear part of the model and this is shown through the following equations,
Residuals from ARIMA model are used to build the neural network according to the following steps,

1- Building ARIMA model for the series by using Box and Jenkins method.

2- Obtaining predicted values$\hat{L}_t$ from ARIMA model to be the linear part of the series.

3- Building ANN model for modeling the residuals

$$e_t = y_t - \hat{L}_t \quad (14)$$

4- Calculate the future predictions $(\hat{N}_t)$

$$e_t = f(e_{t-1}, e_{t-2}, \dots, e_n) + \varepsilon_t \quad (15)$$

Where is,

$\boldsymbol{f}$, non- linear function is decided by ANN.

ε_t, random error

5- Obtaining $\hat{y}_t$ by combining the predictions from ARIMA model with predictions from ANN model

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \quad (16)$$

Table (10): comparison between Actual and forecasted values by using ARIMA –ANN (0,1,1) (2:5:1)

Date	Observed values	Residuals	Forecast values	Difference Percent
3-1-2022	11897.4	-0.0010334	11897.401	0.000008%
4-1-2022	11982.6	0.0024626	11982.598	-0.0000167%
5-1-2022	12043.4	-0.001298	12043.401	0.00000830%
6-1-2022	11978.6	0.002052	11978.598	0.00001713%
9-1-2022	12022.7	-0.0004667	12022.701	0.00000388%

Table (11): Standards of evaluation accuracy of model

	MSE	MAE	MAPE	AIC	BIC
ARIMA-ANN (0,1,1) (2:5:1)	0.01157	0.01109	0.1036	25.65	28.25

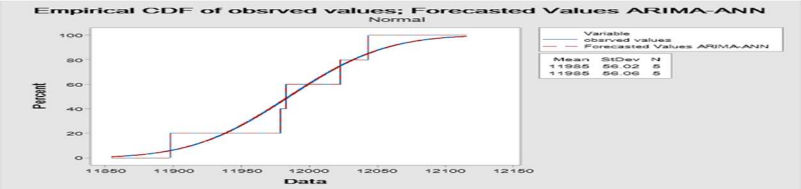

Figure (16) The graphic of actual and forecasted Values of EGX30 by using ARIMA-ANN (0,1,1) (2:5:1)

Table (12) Comparing between ARIMA (0,1,1), ANN (2:5:1) and ARIMA-ANN (0,1,1) (2:5:1) results.

Time	ARIMA	ANN	ARIMA-ANN
3/1/2022	11954.3	11913.84	11897.401
4/1/2022	11957.8	11921.94	11982.598
5/1/2022	11961.3	11953.55	12043.401
6/1/2022	11964.7	12009.51	11978.598
9/1/2022	11968.2	11981.97	12022.701
MSE	52.45	54.105	0.012
MAE	46.44	47.718	0.011
MAPE	0.387	0.397	0.104
AIC	1654.26	1711.52	25.65
BIC	1656.84	1714.10	28.25

From the earlier table, the hybrid model ARIMA-ANN is the best model for analysis among three models because it recorded the lowest values for the five criteria MSE, MAE, MAPE, AIC and BIC.

8- **Conclusion**, Tested the ability and efficiency of ARIMA, ANN and ARIMA-ANN models in forecasting by EGX30 index.This research proposes three models , where is ARIMA (0,1,1) model was used to estimate the linear part of the model , after that a series of model residuals was extracted by the difference between original and estimated series , then estimate the non- linear part by ANN(2,5,1) model and combined between linear part and non-linear part to achieve the hybrid model to make the final prediction , After comparing three models ,it was concluded that the ARIMA-ANN(0,1,1)(2:5:1) model was better and more accurate in predicting EGX30 index than ARIMA model and ANN model that operate separately .

References

[1] Abdul Qader, Qreeb (2022) " Using the Box and Jenkins methodology to predict the time series of Gross fixed capital formation –an applied study on the Kingdom of Saudi Arabia – Journal of Economic, Administrative and Legal sciences – Vol 6 No 14.

[2] Aisa, Alam Zaki (2000) "Engineering brown neural networks, algorithms, applications," beam for publishing and science.

[3] Alam, Md., Murshed Mu., Manikandan Pa., Duraisamy Pa and Shamansurova Zi. (2023), " Forecasting oil, coal, and natural gas prices in the pre-and post-COVID scenarios: Contextual evidence from India using time series forecasting tools. -Elsevier-Vol 81-March.

[4] Al-Saadi, (2019) "Comparative study between Box and Jenkins models and synthetic neural networks to predict Iraq's trade balance" - Economic Journal – Iraq

[5] Al-Wasifi, Al-Shaima (2012) " Prediction of the use of the integration of synthetic neural networks and models of Box and Jankins" (applied study) - Al-Mansoura University, Egypt.

[6] Babu AS, Reddy SK (2015), Exchange Rate Forecasting using ARIMA, Neural Network and Fuzzy Neuron, J Stock Forex Trad, P.P 1-5.

[7] Basiouny, Abdul Hamid (2008) "Synthetic neural network technology and sixth generation computing, Scientific Books Publishing and Distribution House.

[8] Bates J., and Granger C. 1969. " The combination of forecasts. Operational Research Quarterly Vol 20(4), P.P451–468.

[9] Box G., and Jenkins G. 1970. Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco, CA.

[10] El fadel S, El khedr A and El zawbaay A (2019)"Modeling the Gas price data using periodogram and Fourier Series Analysis" Arab journal of sciences and research publishing Vol 1 -3 March2019.

[11] Fawzy Ha .and Alrweili Hl. (2022)" Forecasting Crude Oil Prices Using an ARIMA-ANN Hybrid Model "- Faculty of Graduate Studies for Statistical Research, Cairo, Egypt Published 1Sep.

[12]Ghatfan and Eskandar 2019 "Time series analysis using ARIMA model, ANN model and Hybrid model (applied study).

[13] Hill and Remus (2018) Neural Network Models for Time series Forecast – Management science, INFORMS.

[14] Mohamed F, Mousa M (2020) "Applying some hybrid models for modeling bivariate time series assuming different distributions for random error with a particular application" – journal of economics and administrative sciences 2020, Vol .26, No 117 Pages 442-479.

[15]Othman, Hossam and Ismail, Tarek (2022) " the use of multi-layered neural networks in predicting credit risks for business establishments " an applied study journal of Financial and commercial Studies – First issue.

[16]Shabri, and Ani bin 1996. Applications of Artificial Neural Networks to theNonlinear Combination of Forecasts. Expert Systems 13, 195–201.

[17] Sharawy, Samir Mustafa. (2004) "Introduction to Modern Analysis of Time Series," Scientific Publishing Center - King Abdulaziz University.

[18] Vandal-Walter 1992) Time chains of applied orientation and Box and Jenkins models. Arabization and review of Dr. Abdelmarahi Hamid Azam, Dr. Ahmed Hussein Harun - Mars Publishing House.

[19] Zayan, Wasim Noah. (2019) "Comparative Study between ARIMA, GARCH and ANN Models in Time Chain Prediction", Faculty of Economics and Administrative Sciences _ Azhar University _ Gaza.

[20] Zhang, G.P. (2003), Time series forecasting using hybrid, Arima and ANN, Neurocomputing, Vol50, P159-175. .