International Academic Institute for Science and Technology

International Academic Journal of Business Management
Vol. 10, No. 1, 2023, pp. 01-11.

ISSN 2454-2768

International Academic Journal of Business Management

www.iaiest.com

Using Time Series Models and Neural Networks to Predict the Sales of Motor Oils in Iraq

Hayder Sabah Tuama [a], Ruqayah Yasir Abdulameer [b]

[a] *College of Administration and Economics, University of Maysan, Iraq. E-mail: haidar_1100@yahoo.com*
[b] *Finance and Banking Department, Imam Al-Kadhum College (IKC), Iraq.*
E-mail: rukaya.yasser@alkadhum-col.edu.iq

Abstract

Time series are the basis for statistical forecasting through building models that depend as a basis in economic analysis, and the most important of these models are Box-Jenkins models, which depends on the interpretation of the relationship between autocorrelation and partial autocorrelation, and when compared with the Artificial Neural Network (ANN) methodology which is considered as one of the new important means in building models, analyzing, evaluating data, forecasting and controlling it without referring to a common statistical model or method to pre-diagnose the behavior of the phenomenon, where the data is processed to reach the best model. This leads to the least possible number of errors and is close to reality and can be used in most cases. One of the goals of time series analysis is to obtain an accurate description of the time series, build a model to explain its behavior, and use the results for the purpose of predicting the series' behavior in the future, and to achieve this, it requires an analytical study of time series models based on statistical and computer methods.

The research aims to compare some time series models by studying and analyzing some time series for motor oils (Diesel S3 oil, hydraulic oil, and other types of oils) through Box-Jenkins, and neural networks models.

Keywords: Time Series, Neural Series, Motor Oil, Iraq.

DOI:10.9756/IAJBM/V10I1/IAJBM1001
Received:18 November 2022/Accepted:12 December 2022/Published:04 January 2023

Introduction

The time series represents a set of accidental observations of a random variable originating from its relationship to time according to the laws of probability, statistics and mathematics. Particularly, the time series is a set of observations of a studied phenomenon according to successive and often regular periods of time. Prediction using time series, including the Box-Jenkins methodology, is the traditional method for building and predicting models, as well as prediction are made through neural networks, where multi-layer neural networks are used for the purposes of the model design, and then the model is used for the purpose of prediction.

One of the advantages of neural networks is that the neural network can represent nonlinear systems in addition to its ability to reduce data instability and treat noise that associated with time series data and anomalies in behavior, on the basis of this principle, time series models were built through the Box-Jenkins method and some associated models a neural network model with a number of layers was also built for some oil sales variables according to the Perceptrons method.

In recent years, neural networks have gained increasing importance in processing and analyzing time series and calculating future predictions, due to their great flexibility compared to the common traditional methods and that were approved in this field, in addition to their ability to learn and adapt themselves to any model.

The Research Problem

The production of motor oils is the mainstay for the production of the oil refineries in Iraq, and planning for them and predicting their needs, whether at local or global levels, is of the utmost importance for this industry, this is because this industry achieved a return that based on converting the oil product into materials that has commercial value with high financial reward. Therefore, the research problem is limited to the extent of the possibility of building efficient models that help the economic planner in developing these products and maximizing their financial return.

The Research Importance

The purpose lies in building multiple models using different methods, including Box-Jenkins models, the models associated with them, and the neural networks model, then comparing them and choosing the best one according to statistical criteria. Any model or time series, whether it is linear, semi-linear, or non-linear, and whether the error follows a normal distribution or any other unknown distribution and adopting a modern methodology with intelligent capabilities for processing, time series analysis and forecasting, and providing a quantitative numerical standard to the senior management in order to make an effective decision through the use of the model for economic planning.

Research Objective

The main goal is to build time series models for motor oil sales data in Iraq according to the actual data and to understand the behavior of these series in terms of time series components such as the general trend, seasonality and other vehicles and in different cases and comparing the methods that applied, as well as the prediction methods, including the method of neural networks.

Theoretical Part

The most important principles of time series analysis will be reviewed through the Box-Jenkins method with some models that associated with them in the first part, then the principles of prediction using the neural network method will be reviewed with the possibility of applying them to the study data.

1. Box-Jenkins Models

Time Series Stability

The first step of analyzing time series is by drawing them to identify some of their components, such as the general trend, cyclical changes, seasonal changes, and finally episodic changes, then the stability of the time series is checked if it is strong, and by this we mean that the time series has a stable statistical distribution that is not related to the time period or is weak stability, what is meant here is stability in the mean and stability in the variance, where the unstable time series in variance turns into a stable series by performing transformations of the series data, including logarithmic transformations and power transformations. As for the stability in the average, it can be

DOI:10.9756/IAJBM/V10I1/IAJBM1001
Received:18 November 2022/Accepted:12 December 2022/Published:04 January 2023

achieved through regular or seasonal differences or both, according to the behavior of the time series. The stability of the time series is often achieved after the second deference.

1.1. P-order Autoregressive Model

The model is known as AR(p) and statistically the model is described as follows:

$$y_i = \sum_{j=1}^{p} \emptyset_j \, y_{j-1} + e_i$$

The errors of the model (e_i) are independent and symmetrically distributed (iid). Among its characteristics is that the autocorrelation function decreases exponentially and approaches zero as the displacement increases, while the partial autocorrelation function breaks off after the displacement p. The model can be described by several formulas, including the regression factor formula (back or forward) according to $B^k = y_{j-k}$.

According to the statistical theory, if the model errors are not independent and this phenomenon often occurs in time series and follows an autoregressive model of the first order, then the simple linear regression model is described as follows:

$$y_i = \alpha + \beta \, x_i + e_i$$
$$e_i = \rho \, e_{i-1} + u_i$$

It is noted that the errors of the regression model (e_i) are related to the first order autoregressive model AR(1), and thus four parameters of the model must be estimated.

As for the moving (Mobil) averages model, known as MA(q), it is described in terms of model errors and by the following relationship:

$$y_i = e_i + \sum_{j=1}^{q} \theta_j \, e_{i-j}$$

The behavior of the autocorrelation coefficients is equal to zero after the displacement q, and the partial autocorrelation coefficients decrease exponentially and turn to zero, and the errors of the regression model can also follow this model.

As for the autoregressive model and the moving averages ARMA(p,q) is described in terms of the previous models as follows:

$$\emptyset(L)(y_i) = \theta(L) \, e_i$$

Which combines in its behavior the specifications of the autoregressive models and the moving averages, and from the model we find that ∅(L), θ(L) represent the polynomials of the autoregressive model and the moving averages model in the model parameters.

1.2. Autoregressive Distributed Time Gaps Model [(1)]

The ARDL models are mix between the autoregressive AR model and the finite distributed lag models, which are of several orders that combine the variables and their displacements. It is considered as one of the methods of building dynamic response models by including the lagging variables as independent explanatory variables in the model, if we have the following functional relationship:

$$y = g(x) + e_i$$

The model can be described in terms of the deceleration variables for the independent variable (X_I) if it is of the following formula:

$$y_i = \emptyset_0 \, x_i + \emptyset_1 \, x_{i-1} + \cdots \ldots + \emptyset_p \, x_{i-p} + e_i$$

The dynamic behavior can be expressed by relying on the previous value of the internal variable, that is, it can be represented by the AR(P) (Autoregressive Model), meaning that the time series variable (y_i) depends on the previous values of the variable and is expressed by the following formula:

$$y_i = \emptyset_0 \, y_{i-1} + \emptyset_1 \, y_{i-2} + \cdots \ldots + \emptyset_p \, y_{i-p} + e_i$$

In other words, the additional or alternative way to contain the dynamic component in economic behavior is through the inclusion of decelerating endogenous variables along with exogenous variables as explanatory variables.

2. Neural Network Model

Artificial neural networks are one of the learning methods based on signal transmission within human brain cells. The scientist Nilsson defined neural networks as a computational system consisting of a number of processing units

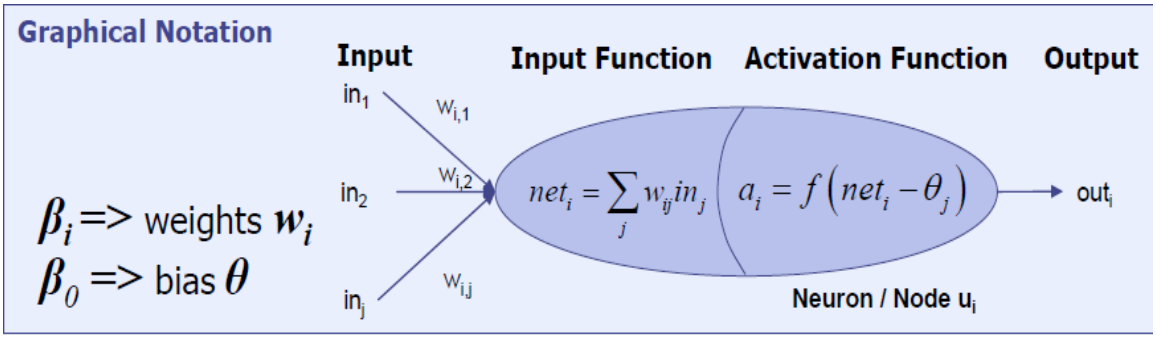
DOI:10.9756/IAJBM/V10I1/IAJBM1001
Received:18 November 2022/Accepted:12 December 2022/Published:04 January 2023

interconnected with each other and characterized by its dynamic and parallel nature in processing the data entering it. Zurada defined artificial neural networks as cellular physical systems in which it is possible to acquire, store and use empirical knowledge. Among other definitions, they are mathematical models concerned with data processing and learning from them, and thus mimic the characteristics that characterize biological systems [(2)]. From all of the above, "N***eural networks can be described as a data processing system based on statistical and mathematical models that have advantages in signal transmission similar to the biological neural network in humans through a set of non-linear models"*** [(3)].

2.1. Model and Effectiveness Functions

It is a function that defines the output by a weighted sum of the inputs from the neuron. The transmission functions of neurons in the hidden layer are often nonlinear and they provide nonlinearity to the network, after collecting and weighing the input from neuron 1 is mapped to i by the transfer function f, it can be shown as:

The input layer standardizes the explanatory variables vector values (x1... xp) so that the range of each variable is -1 to 1 [(4)]. The input layer distributes the values to each of the neurons in the hidden layer as well as the prediction variables, besides having a fixed limit called bias which is fed to each of the hidden layers the bias is multiplied by the weight and added to the sum going to the neuron.



The most important of these functions are as the following:

1. **Step Function:** This function determines the output from it between 0 and 1 so that the output becomes equal to one if the input is greater or equal to zero and the output becomes equal to zero if the input is greater than zero and according to the following formula [(5)]:

$$f(s) = \begin{cases} 0 & if \quad s < 0 \\ 1 & if \quad s \geq 0 \end{cases}$$

It is also called the Threshold function, and the following figure shows this function.

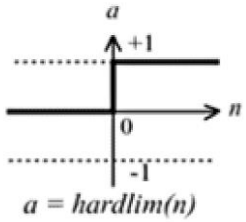


Figure 1: Shows the Threshold Function (Step Function)

3. Sigmoid Function [(7)]

It is based on the logistic distribution, and this function takes the input values between ∞ - and ∞ +, and the outputs are between 0 and 1 and it is the most commonly used function because of its ease of derivation, which takes the following formula within the field (1, 0).

$$F(x) = \frac{1}{1 + \exp\ (-\mu x)}$$

DOI:10.9756/IAJBM/V10I1/IAJBM1001
Received:18 November 2022/Accepted:12 December 2022/Published:04 January 2023

Where: X: represents the product of the input value with the corresponding weight; μ: represents the slope of the function and its value is equal to one, and the derivative of the function F(x) is:

$$f'(x)= f(x)[1-f(x)]$$

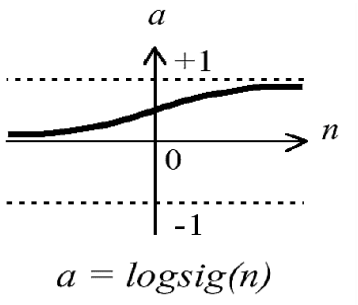


Figure 2: The Sigmoid Function

It is one of the most important functions used in artificial neural networks due to its ability to spread the values (1-0) in a good manner and in proportion to the logistic distribution and it is characterized by ease of calculation and implementation. The way neurons are connected to each other to form the network is illustrated in Figure 3 in which each component of the input ray is connected to each neuron through a weight matrix W.

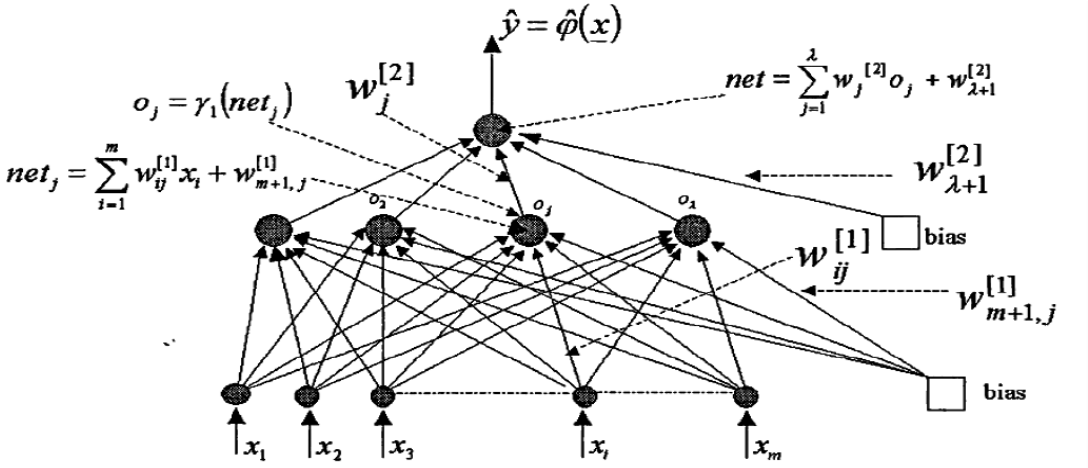


Figure 3: Represents the Architecture of the Artificial Neural Network

Every neuron contains a junction or node that combines the weighted input with the displacement to form the numerical output of the neuron. As a result, the output components of the neuron layer form the output ray (a one-column matrix) and the components of the input ray enter the network through the following weights matrix, as in the figure below [(8)]:

$$\mathbf{W} = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,R} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,R} \\ \\ w_{S,1} & w_{S,2} & \cdots & w_{S,R} \end{bmatrix}$$

The line pointers of the elements of this matrix are indicative of the target neuron; the column indicators show the input components source, which means, the indicators in element W12 point to that this weight is related to the first neuron, and the input component of this neuron is the second component. A neural network can consist of several

DOI:10.9756/IAJBM/V10I1/IAJBM1001
Received:18 November 2022/Accepted:12 December 2022/Published:04 January 2023

layers; each layer has a weight matrix W, an offset ray, and an output ray, and for distinguish, the layer number is added as an upper guide to each of the variables used through the network that shows the combined input, a neuron in the first layer, a neuron in the second layer, and so on in the same manner. It is also noted that the output of each middle layer is the income of the next layer. Thus, each layer in this network is considered as a single layer network. The layer that gives the output is called the output layer, while the input does not consider as a layer, and the rest of the layers are called the hidden layers. The triple network can be represented by the following figure [9]:

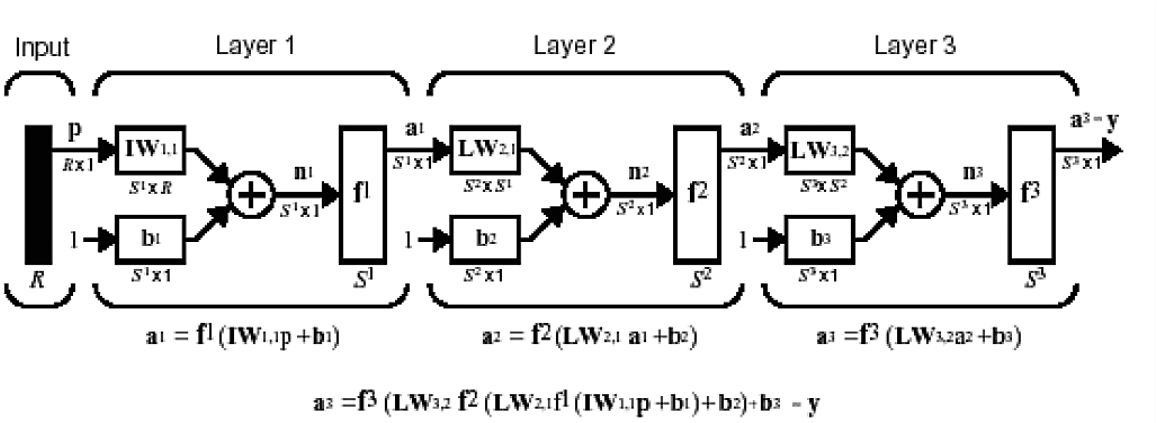
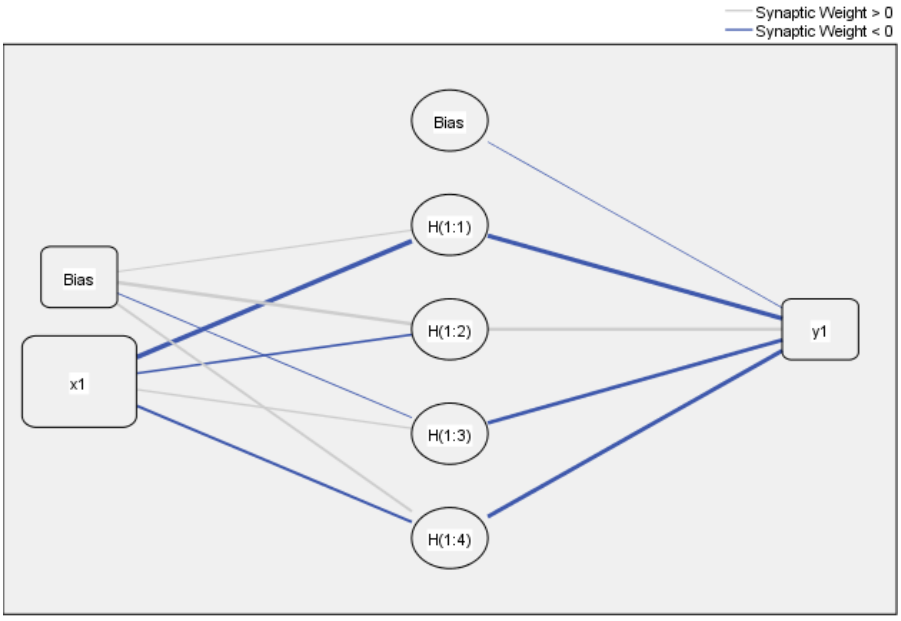


Figure 4: The Triangular Neural Network

Multilayer networks are highly effective networks, especially two-layer networks, which are widely used, these networks can solve many complex problems, but it takes longer time. The process of estimating the matrix of weights is through iterative methods and non-linear optimization methods to reach the best estimator for the matrix w, and it passes in three stages: the training stage, the testing stage, and then the model evaluation stage.

The Practical Side

A neural network model was built for the diesel oil sales variable S3 with a training stage of (70%) of the data and a testing stage, (30%), and with one hidden layer consisting of four nodes after converting the data to standard form and a hyperbdic tangent for nodes in the hidden layer and an output function (identity) for the results of the final output. The results were described in the following form, which represents the best estimated network after leaving the room for the computer to control the number of layers and nodes at each layer:



DOI:10.9756/IAJBM/V10I1/IAJBM1001
Received:18 November 2022/Accepted:12 December 2022/Published:04 January 2023

The following weights matrix between the input layer and the hidden layer with its four nodes

Parameter Estimates						
Predictor		**Predicted**				
		Hidden Layer 1				**Output Layer**
		H(1:1)	**H(1:2)**	**H(1:3)**	**H(1:4)**	**y1**
Input Layer	(Bias)	.091	.263	-.073-	.228	
	x1	-1.268-	-.193-	.097	-.234-	
Hidden Layer 1	(Bias)					-.035-
	H(1:1)					-.581-
	H(1:2)					.257
	H(1:3)					-.293-
	H(1:4)					-.444-

And with the following competency measures for the training and testing stages:

Model Summary		
Training	Sum of Squares Error	17.073
	Relative Error	.899
	Training Time	0:00:00.00
Testing	Sum of Squares Error	6.404
	Relative Error	.950
Dependent Variable: y1		
a. Error computations are based on the testing sample.		

And for estimation through time series models, especially Box-Jenkins models, the time series was first drawn as in Figure (5) in order to detect changes from the general trend and the rest of the compounds, and It is noted that the series fluctuates around an average that is almost fixed in relation to time with no general trend in it and here it does not need differences or changes in the stability of the variance.

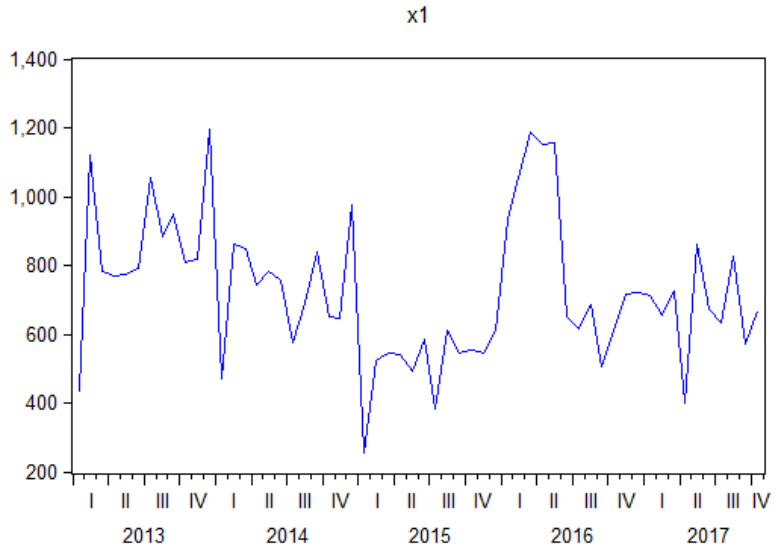


Figure 5: The Time Series

The revised Dickey–Fuller test was used to test the unit root, where the null hypothesis was rejected, it indicates that the time series is stable and the t-test value was (-5.48206) with a probability of (0.000), it is less than the specific level of significance (0.05), and from it is infer that the time series is integrated of degree I (0), which confirms the possibility of dealing with it directly.

DOI:10.9756/IAJBM/V10I1/IAJBM1001
Received:18 November 2022/Accepted:12 December 2022/Published:04 January 2023

Null Hypothesis: X1 has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.482063	0.0000
Test critical values:	1% level	-3.550396	
	5% level	-2.913549	
	10% level	-2.594521	

*MacKinnon (1996) one-sided p-values.

The autocorrelation function behavior and the partial autocorrelation.

Date: 03/23/18 Time: 21:45
Sample: 2013M01 2017M10
Included observations: 58

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.310	0.310	5.8697	0.015
		2	0.355	0.287	13.720	0.001
		3	0.294	0.153	19.196	0.000
		4	0.019	-0.210	19.220	0.001
		5	0.029	-0.087	19.276	0.002
		6	-0.022	-0.003	19.308	0.004
		7	-0.066	0.016	19.605	0.006
		8	-0.121	-0.104	20.623	0.008
		9	-0.022	0.061	20.656	0.014
		10	-0.064	0.010	20.949	0.021
		11	-0.323	-0.374	28.691	0.003
		12	-0.027	0.110	28.747	0.004
		13	-0.228	-0.037	32.769	0.002
		14	-0.167	0.007	34.962	0.001
		15	-0.085	-0.085	35.544	0.002
		16	-0.093	0.055	36.257	0.003
		17	-0.041	0.005	36.399	0.004
		18	-0.093	-0.175	37.148	0.005
		19	-0.112	-0.175	38.272	0.005
		20	-0.088	0.112	38.978	0.007

After calculating the function of autocorrelation coefficients and partial autocorrelation, the data did not show the nature and rank of the model, so it would be better to choose a group of models and do the comparison between them in terms of efficiency measures for these models, and the best model was a model and according to the following table.

DOI:10.9756/IAJBM/V10I1/IAJBM1001
Received:18 November 2022/Accepted:12 December 2022/Published:04 January 2023

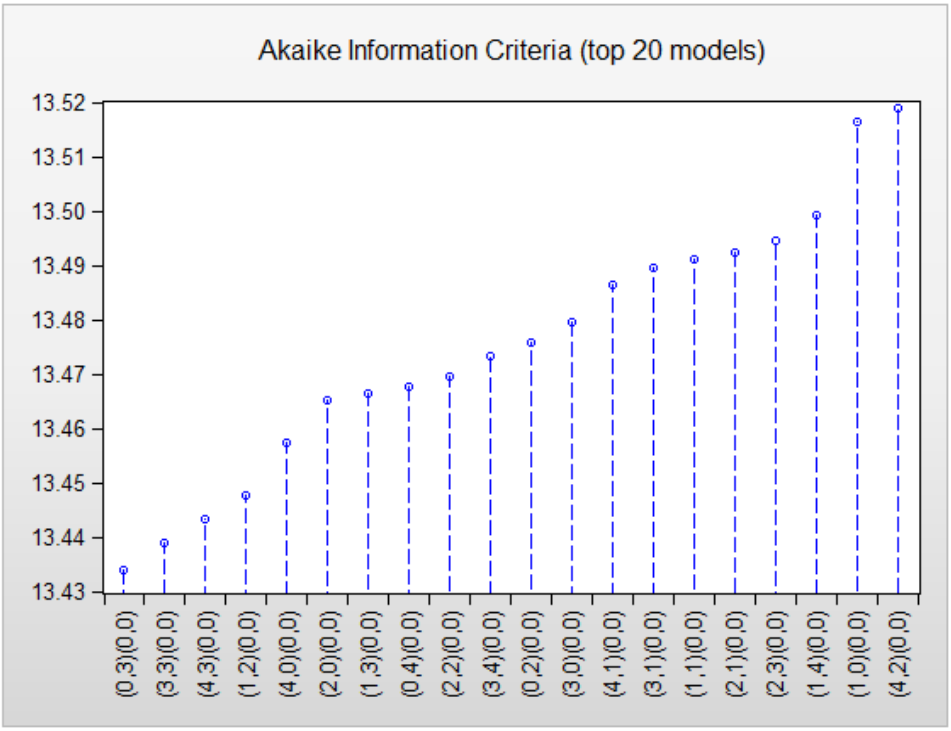
Dependent Variable: X1
Method: ARMA Maximum Likelihood (BFGS)
Date: 03/23/18 Time: 20:18
Sample: 2013M01 2017M10
Included observations: 58
Convergence achieved after 9 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	724.3774	46.51155	15.57414	0.0000
MA(1)	0.164014	0.123225	1.331008	0.1889
MA(2)	0.328079	0.143498	2.286300	0.0263
MA(3)	0.329243	0.188921	1.742757	0.0872
SIGMASQ	33332.73	6203.702	5.373039	0.0000

R-squared	0.232010	Mean dependent var	726.2931
Adjusted R-squared	0.174048	S.D. dependent var	210.1523
S.E. of regression	190.9904	Akaike info criterion	13.43398
Sum squared resid	1933298.	Schwarz criterion	13.61161
Log likelihood	-384.5855	Hannan-Quinn criter.	13.50317
F-statistic	4.002826	Durbin-Watson stat	1.897135
Prob(F-statistic)	0.006548		

Inverted MA Roots	.21-.72i	.21+.72i	-.58

The following table also shows the decrease in the Akaike information standard with the change of different models to reach the best model.



DOI:10.9756/IAJBM/V10I1/IAJBM1001
Received:18 November 2022/Accepted:12 December 2022/Published:04 January 2023

Conclusions and Recommendations

Conclusions

1. Box Jenkins models suffer from multiple problems, including the inability to determine the rank of the model, whether it is related to the autoregressive model or to the moving averages, because the reality of the data does not match the theoretical behavior of these models, and this is what was faced during the building of the model.
2. The results demonstrated the efficiency of the artificial neural network method in most of the models, especially in the seasonal models that include more than number of nodes in the input layer of the artificial neural network.
3. The superiority of the Box-Jenkins method to the traditional neural networks method in the moving medium of the models, this is due to the imposition of the Box-Jenkins methodology on the artificial neural networks method, which affects the input and then the output results of the neural network.
4. The results demonstrated the efficiency of the artificial neural networks method in processing semi-linear models, and the failure of the Box-Jenkins method in treating these "chains" states, despite the imposition of the Box-Jenkins methodology on the traditional neural networks method. This indicates the ability of artificial neural networks in processing semi-linear and non-linear "chain" models.
5. The lack of accuracy in determining or diagnosing the number and type of nodes of the input layer of the neural network, which affects the process of training the network and then the output of the artificial neural network.
6. Artificial neural network models are mostly solid models because they are well trained on data.

Recommendations

The most important recommendations that have been reached in the light of the conclusions can be summarized as follows:

1. Adopting the application of the artificial neural networks method in research and studies related to time series.
2. Using the wavelet analysis method to determine the inputs of the modified artificial neural network.
3. Adopting artificial neural networks in processing time series models with different and varied patterns, in which the error is not distributed according to the normal distribution, this where not stipulated, and the error can be distributed in it, while the Box-Jenkins models and the traditional methods assume that the error is normally distributed with a mean equal to zero and a variance equal to one, this will opens wide horizons for researchers in dealing with these models, while this condition was overlooked due to its complexity and the difficulty of the method of treating it.
4. Adoption of artificial neural networks in the treatment of multivariate time series (multiplicity of inputs and outputs) and models of the linear and non-linear transformation function and the canonical form.
5. The methodology of artificial neural networks can solve the problems of nonlinearity with the presence of external variables without resorting to data entry transformation functions for the purpose of making it stable.

References

Arabic References

Al-Zoba'i, O.M. (2007). *A proposed method for diagnosing time series models.* The first Arab Statistical Conference - Jordan, p. 45.

Gebara, M.J.M.A. (2012). *Predicting the Time Series for the Level of the Nile at Waddani Station*. A doctoral thesis submitted to Sudan University of Science and Technology.

Al-Bakri, A., & Al-Sayed. G. (2010). *Designing a system to distinguish patterns of geometric shapes using neural networks*. College of Science - University of Babylon, p. 119.

English References

Palit, A.K., & Popovic, D. (2005). *Computational Intelligence in Time Series Forecasting*. Springer-Verlag, London.

Wegman, E.J. (2000). Time Series Analysis: Theory, Data Analysis and Computation, Addison – Wesley publishing company, Inc., New York.

Priestley, M.B. (1981). *Spectral Analysis and Time Series*. Academic press Harcort Brace Jovanovich, Publishers, London, San Diego, New York, Tokyo Toronto.

DOI:10.9756/IAJBM/V10I1/IAJBM1001
Received:18 November 2022/Accepted:12 December 2022/Published:04 January 2023

Amir, F.A. (1999). Comparison Between Neural-Network Forecasting Techniques-Case Study: River Flow orecasting. *Journal of Transactions on Neural Networks*, *10*(2), 457-471.

Cadenas, E., Jaramillo, O.A., & Rivera, W. (2010). Analysis and forecasting of wind velocity in chetumal, quintana roo, using the single exponential smoothing method. *Renewable Energy*, *35*(5), 925-930.

Catalão, J.D.S., Pousinho, H.M.I., & Mendes, V.M.F. (2011). Short-term wind power forecasting in Portugal by neural networks and wavelet transform. *Renewable energy*, *36*(4), 1245-1251.

DOI:10.9756/IAJBM/V10I1/IAJBM1001
Received:18 November 2022/Accepted:12 December 2022/Published:04 January 2023