

Republic of Iraq

Ministry of Higher Education and Scientific Research

University of AL-Qadisiya

College of Computer of sciences & Mathematics

Department of Mathematics



# **A Comparison Between Spectral Analysis and Artificial Neural Networks to Determine the Best Prediction Model with Application**

A thesis

Submitted to the Council of the College of Computer Sciences & Mathematics of  
Al -Qadisiyah University as a Partial Fulfillment of the Requirements for the  
Degree of Master of Sciences in Mathematics

By

**Emaan Yousif Abdoon Al-Shabbani**

Supervised By

**Prof. Dr. Mohammed Habib Al- Sharoot**

1439 A.H

2017 A.D

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

يَرْفَعِ اللَّهُ الَّذِينَ آمَنُوا مِنْكُمْ وَالَّذِينَ أُوتُوا

الْعِلْمَ دَرَجَاتٍ

صَدَقَ اللَّهُ الْعَظِيمُ

سورة المجادلة / آية 11

# إهداء

الى رمز الرجولة والتضحية..

الى من علمني الصمود مها تبدلت الظروف..

الى من حصد الاشواك عن دربي ليمهد لي طريق العلم..

الى الشمعة التي تحترق كل يوم لتضيء دربي وتبدد ظلمتي..

الى من أحمل أسمه بكل فخر .. الى أبي

أهدي ثمرة جهدي و صبري..

ايمان يوسف

# ACKNOWLEDGEMENTS

First, thankfulness and glorification to Almighty Allah Who helped and inspired me to write this piece of work.

I would like to express my sincere gratitude to my supervisor Prof. Dr Mohammed Habib Al-Sharoot for his valuable suggestions, extensive discussions, encouragement and continuous guidance.

Also, I am deeply indebted to Professor Tahir Reisan for his help and guidance.

Finally, I would like to thank my family for their generosity and support at a challenging time.

Emaan, 2017.

# Supervisor Certification

I certify that the preparation of this thesis entitled " **A Comparison Between Spectral Analysis and Artificial Neural Networks to Determine the Best Prediction Model with Application** " was made under my supervision at the University of Al-Qadisiyah, College of Computer Sciences and Mathematics as a partial fulfillment of the requirements for the degree of Master of Science in Mathematics.

Signature: 

Name: Dr. Mohammed Habib Al-Sharoot

Scientific Grade: prof.

Date: / / 2017.

In view of the available recommendations, I forward this thesis for debate by the Examining Committee.

Head of the Mathematics Department

Signature: 

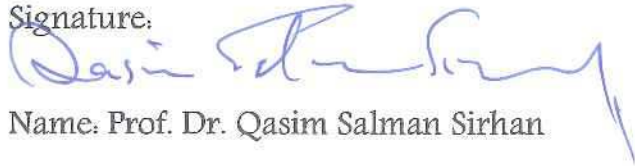
Name: Dr. Qusuay Hatim Eghaar

Date: / / 2017.

# Language Expert Certification

This is to certify that I have read the thesis entitled " **A Comparison Between Spectral Analysis and Artificial Neural Networks to Determine the Best Prediction Model with Application** " and corrected every grammatical mistake I found; therefore, it is now fully qualified for debate.

Signature:



Name: Prof. Dr. Qasim Salman Sirhan

Date: / / 2017.

## Examination Committee Certification

We certify that we have read this thesis " **A Comparison Between Spectral Analysis and Artificial Neural Networks to Determine the Best Prediction Model with Application** " and as an examining committee, we examined the student in its content, and that in our opinion it meets the standards of a thesis for the Degree of Master of Science in Mathematics.

Signature: 

Name: Prof. Dr. Fadhil Abdul Abbas

Date: 15 / 11 / 2017

Chairman

Signature: 

Name: Dr. Gorgees Shaheed Mohammad

Date: 13 / 11 / 2017


Member

Signature: 

Name: Asst. Prof. Ameera Jabber Mohaisen

Date: 14 / 11 / 2017

Member

Signature: 

Name: Prof. Dr. Mohammed Habib Al-Sharoot

Date: 13 / 11 / 2017

Member & Supervisor

In view of the available of the recommendations I forward this thesis for debate by examining committee .

Signature: 

Name: Asst. Prof. Dr. Hisham Mohammed ali Hasan

Date: 29 / 11 / 2017

# List of Abbreviations and Symbols

Symbols	Definitions
SACF	Sample Autocorrelation Function
SPACF	Sample Partial Autocorrelation Function
AR	Autoregressive Model
MA	Moving Average Model
ARMA	Autoregressive Moving Average Model
ARIMA	Autoregressive Integrated Moving Average Model
ADF	Augmented Dickey – Fuller Test
P.P	Phillips – Perron Test
KPSS	Kwiatkowski – Phillips – Schmidt – Shin Test
LM	Multiple Lagrange Test
EML	Exact Maximum Likelihood
$Q_{BP}$	Box – Pierce Test
$Q_{LB}$	Ljung–Box (Modified Box – Pierce) Test
AIC	Akaike Criterion
BIC	Bayesian Information Criterion
H-Q	Hannan – Quinn Criterion
ANN	Artificial Neural Network
b	Bias
$\eta$	Learning Rate
BP	Backpropagation Algorithm
$\gamma$	Momentum
MSE	Mean Square Error
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error



# Table of Contents

Chapter	Title	Section	Title	Page Number
First	Introduction and Literature Review	1-1	Introduction	1
		1-2	The Research Objective	2
		1-3	Literature Review	2
Second	The Theoretical Side: -Time Series Analysis -Spectral Analysis -Artificial Neural Networks	2-1-1	Time Series	9
		2-1-2	Autocovariance Function	9
		2-1-3	Autocorrelation Function	10
		2-1-4	Partial Autocorrelation Function	10
		2-1-5	Stationarity	11
		2-1-6	Time Series Models	12
		2-1-7	Autoregressive Integrated Moving Average Model	14
		2-1-8	Box – Jenkins Methodology	15
		2-1-9	Model Selection Criteria	20
		2-2-1	Spectral Analysis	21
		2-2-2	Harmonic Analysis	21
		2-2-3	Analysis of Harmonic Model	22
		2-2-4	Periodogram Analysis	24
		2-2-5	The Search for Hidden Periodicities	25
		2-2-6	Selection of Harmonic Model	27
		2-2-7	Spectrum	28
		2-2-8	Spectrum of ARMA Models	30
		2-2-9	The Relationship Between the Spectrum and the Autocorrelation Function	31

		2-2-10	Methods of Spectrum Estimation	32
		2-3-1	Artificial Neural Network	39
		2-3-2	Artificial Neural Network Structure	39
		2-3-3	Artificial Neural Networks Architecture	43
		2-3-4	Training Algorithms	47
		2-3-5	Hebb Learning Rule	48
		2-3-6	Least Mean Square (LMS) Algorithm	51
		2-3-7	Backpropagation Algorithm	54
		2-3-8	Criteria of Model Selection	58
Third	The Application Side -Artificial Neural Networks -Spectral Analysis	3-1	Data Set	59
		3-2	Artificial Neural Network Methodology	61
		3-3	Spectral Analysis	83
		3-4	Conclusion and Recommendations	94

# List of Tables

Table Number	Title	Page Number
2-1	Behaviour of Theoretical ACF and PACF for Stationary Process	17
2-2	Coefficients for Calculating Approximate Probability Limits for Cumulative Periodogram Test	26
3-1	Exchange Rate of Iraqi Dinar vs US Dollar	60
3-2	Tests Results of The Original Time Series $Y_t$	64
3-3	Tests Results of The Transformed Time Series $W_t$	66
3-4	Estimated Parameters of ARIMA(p,d,q) Models	67
3-5	Estimated Criteria of ARIMA(p,d,q) Models	68
3-6	Autocorrelation Coefficients of the Residuals	70
3-7	Results of Ljung-Box Test	71
3-8	First Model: ANN(1)	73
3-9	Second Model: AAN(2)	73
3-10	Third Model: ANN(3)	74
3-11	Forth Model: ANN(4)	74
3-12	Fifth Model: ANN(5)	75
3-13	Minimum Values of the Criteria	75
3-14	Values of the Criteria During Momentum Change and Fixed Learning Rate	79
3-15	Values of the Criteria When Data Partitioning	80
3-16	Values of Criteria for Final Model Without Processing	80
3-17	Final Values of Weights and Biases	81
3-18	Periodogram Analysis of $W_t$	84
3-19	Cumulative Periodogram of $W_t$	87
3-20	$F_k$ Statistic Application Results	89

3-21	Autocorrelation Coefficients of Residuals	90
3-22	Box-Ljung Test	91
3-23	Cumulative Periodogram of Residuals	92
3-24	Comparison between Artificial Neural Networks and Spectral Analysis	93

# List of Figures

Figure Number	Title	Page Number
2-1	Spectrum of AR(1)	37
2-2	Spectrum of MA(1)	38
2-3	Artificial Neuron	40
2-4	Linear Function	41
2-5	Sigmoid Function	42
2-6	Hyperbolic Function	42
2-7	Single-Layer Artificial Neural Network	44
2-8	Multi-layer Artificial Neural Network	46
2-9	Linear Associator	48
2-10	Adaptive Linear Neural Network	52
3-1	Exchange Rate of Iraqi Dinar vs US Dollar	62
3-2	Correlogram of the Original Time Series $Y_t$	63
3-3	Correlogram of The Transformed Time Series $W_t$	65
3-4	Correlogram of Residuals	69
3-5	The Final Artificial Neural Network Model	82
3-6	Periodogram	85
3-7	Cumulative Periodogram of $W_t$	88
3-8	Autocorrelation Function for Residuals	91
3-9	Cumulative Periodogram of Residuals	93

# Abstract

The variability of exchange rate, especially the unexpected rapid increase and decrease, has a significant effect on the national economy of any country. Iraq is no exception; therefore, the accurate prediction of Iraqi dinar exchange rate with respect to the US dollar is a crucial matter that affects the planning processes to secure economic stability in Iraq. This thesis aims to compare the spectral analysis methodology with the artificial neural networks methodology in terms of the prediction of Iraqi dinar exchange rate with respect to the US dollar within a ten-year time period that extends from 30/01/2004 to 30/12/2014. In this comparison, mean square error MSE, mean absolute error MAE and mean absolute percentage error MAPE were used as statistical criteria to compare the two methodologies. The results show that the neural network methodology is better than the spectral analysis methodology to predict the exchange rate of Iraqi dinar with respect to the US dollar.

# **Chapter One**

**Introduction and Literature  
Review**

# 1. Introduction and Literature Review

## (1-1) Introduction

Recently, the interest has increased in analyzing the time series for a certain phenomenon and predicting its future values based on the past observations of the phenomenon. The fluctuation that occurs in time series is usually analyzed in the time domain. However, there are some variations that occur in some phenomena due to hidden periodicities that cannot be detected in the time domain. Therefore, the analysis of the phenomenon is performed in the frequency domain using the spectral analysis methodology. In addition, the analysis in the time domain is normally performed using Box - Jenkins approach, which provides good predictions of linear time series only. However, some time series consists of linear and nonlinear components or only nonlinear component, which makes the prediction process more difficult and complicated. Therefore, the use of neural network methodology is considered a proper choice to model this type of series since this methodology offers the ability to analyze nonlinear relationship in the data.

The exchange rate is a phenomenon that is characterized by rapid, sometimes unexpected, variations. These variations have significant effects on the Iraqi national economy, which makes the accurate prediction of such variations play an important role in securing the economic stability of Iraq. Therefore, the analysis of this phenomenon using the neural network methodology in the time domain and the spectral analysis methodology in the frequency domain is a crucial and necessary matter to obtain the best model to represent the data and provide accurate predictions.

This thesis consists of three chapters. The first chapter consists of four sections that include the introduction, research problem, research objective, and a review for the previous literature related to the thesis topic. The second chapter consists of three sections. The basic concepts of time series and Box-Jenkins methodology are explored in the first section of this chapter. In the second section, the basic principles of spectral analysis methodology are explored, which include periodogram, spectrum, and spectrum estimation methods. The third section includes definition of neural network and its components as well as the training algorithms of the network, the backpropagation algorithm in particular.



The last chapter of the thesis consists of three sections. The first section presents an introduction about Iraqi Dinar exchange rate against the US Dollar in addition to the data table. In the second section, neural network methodology has been applied on the exchange rate data to find the best prediction model while the spectral analysis methodology has been applied on the same data in the third section to find harmonic model that represents the significant harmonic components in the data. The third section includes conclusions and recommendations of the researcher.

### **(1-2) The Research Objective**

This research aims at building two models using two different methodologies, the first methodology is the artificial neural networks to build the first model whereas the second one is the spectral analysis to build the second model in order to predict the Iraqi Dinar exchange rate with respect to the US Dollar and comparing the two models relying on MSE, MAE and MAPE as statistical criteria to perform the comparison.

### **(1-3) Literature Review**

In (2001) [3], the author conducted a comparison between two of the available spectral methods and three generalized methods in the case of univariate, in addition to two suggested methods. The application was performed on bivariate autoregressive model by applying simulation experiments of different sample sizes. The author concluded that the suggested method is better compared with the other estimation methods in the context of multi spectrum smoothing where the absolute value of eigen values is more than (0.5) for all sample sizes.

In (2001) [22], the authors used the ANN methodology to renewable energy problems especially the modeling of the solar energy field (solar steam generation plant). The backpropagation algorithm was used to train the network which consisted of more than one hidden layer. The data was processed and scaled down to the fit in the period [-1,+1] which allows for better training results. This approach proved that the ANN provides competitive results in this field comparing to the other techniques.

In (2003) [63], a comparison was conducted between the Autoregressive Integrated Moving Average (ARIMA) model and the Artificial Neural Networks (ANN) model in prediction. The data that was used in this work consists of three sets: sunspot, Canadian lynx and British Pound/US Dollar exchange rate. A hybrid methodology, combine ARIMA and ANN, was

used by using single hidden layer feedforward network with logistic activation function in the hidden layer. There was no rule to choose the number of the nodes in the hidden layer in this work. The MAD and MSE criteria was used in the comparison between the ARIMA, ANN, and the hybrid models. The results showed that the hybrid model provided more precise results.

In (2003) [29] the authors used the Elman Jordan Recurrent Network to predict the exchange rate of the US dollar and four other currencies: Swiss Franc, Europeans Money (EURO), Japanese Yen, and the Great Britain Pound (GBP). The data was reprocessed by using the normalization formula in order to remove the data correlation before applying the data to the network. They used 100 nodes in the hidden layer and the logistic activation function in both the hidden layer and the output layer, and the results showed that the neural network was successful in prediction.

In (2004) [4], the author studied the problem of determining the best estimation of spectral density that is consistent with the actual spectral density function to describe the behavior of nucleic acid chain of eight living beings. Based on simulation error square mean and simulation absolute errors mean as criteria, the author concluded that Parzen weight function is better than Priestley weight function, Kaiser weight function, and the rest of the other functions that were used to determine the best consistent estimation of the spectral density function.

In (2004) [23], a comparison study of three training algorithms was conducted to predict the exchange rate of the Australian Dollar. These algorithms were: Standard Backpropagation SBP, Scaled Conjugated Gradient SCG, and Backpropagation with Bayesian Regularization BPR. The data were collected from the Reserve Bank of Australia between 1991-2002. The data were divided into two set; training set (500 observations) and testing set (65 observations). The stopping criteria were set between 5000-10000 iterations. The network was designed in 30 different designs with different weights and learning parameters based on the MAE and the normalized means square error NMSE. The results showed that the SCG algorithm provided better prediction results than the other algorithms, and the retraining of the network by using new data is a good approach to obtain precise prediction. The results also showed that the convergence of the SBP algorithm was highly dependent on the learning rate and the momentum. In addition, the number of the nodes in the hidden layer had a high effect on the behaviour of the network and it was calculated by trial and error.

In (2004) [27] conducted a comparison study between the Seasonal Autoregressive Integrated Moving Average (SARIMA) model and the Artificial Neural Networks (ANN) model in the prediction of three sets of data: Airline, Tourist and Notten Data. The Automatic Relevance Determination (ARD) approach was used to identify the input variables of the neural networks. The neural network was trained by using the backpropagation with Levenberg-Marquard optimization technique. A linear activation function was used in the output layer, while the hyperbolic tangent activation function was used in the hidden layer. Based on the mean euclidean distance criteria MED, the results showed that the ARD-based method was better than the SARIMA-based method in prediction.

In (2005) [5], the author studied the two stages of identification and diagnostic checking of Box-Jenkins methodology. He compared the identification tools with the rank selection criteria and the diagnostic checking tests based on the concept of simulation. The author concluded that each of ACF, PACF, and inverse autocorrelation function IACF works properly in identifying and determining the rank of model MA(1) and model AR(1) while it cannot determine the rank of model ARMA(1,1). In addition, the author found that IACF can give more accurate results than PACF, and that the suggested extended sample inverse autocorrelation function ESIACF method is better than extended sample autocorrelation function ESACF in identifying the ARMA (1,1) model.

A comparison study [64] was conducted in (2005) to compare the SARIMA model and the ANN model in the prediction of the simulation and real data. The feed forward neural network was used with the backpropagation with Levenberg-Marquard optimization technique to train the network, the logistic activation function in the hidden layer and identity activation function in the output layer. Four models were build based on the type of the data that was applied to the network. The first model (O) was built by using the original time series, the second model (DT) was built by using the series after removing the trend, the third model (DS) was built by the series after removing the seasonal component, the fourth model (DTDS) was built by the series after removing both the trend and the seasonal component. Based on the root mean square error RMSE, MAPE and MAE, the results showed that the network model with DTDS data provided better results and did better than the SARIMA model in prediction. The results also showed that the network model was affected by the type of the data whether it has trend component or seasonal component or both. The results also showed that it is necessary to remove both trend component and seasonal

component from the data before applying it to the network because these two components can lead to bad results if they still exist in the data.

In (2007) [52], the authors conducted a comparison between an approximate conditional-mean ACM-type filter algorithm and robust least square RLS filter algorithm in terms of cleaning time series by robust way to obtain a robust spectral density estimator. Based on the simulation results, they concluded that the ACM- filter algorithm is better. Then, they applied the competitive method on actual heart rate variability measurement of diabetes patients.

A comparison between four methods was conducted in (2008) [31], which included curve fitting, Box-Jenkins, ANN, and extrapolation with periodic function to predict the wind speed in the atmosphere. The data was collected between 1992 and (2001) from the Indian costal station. In the ANN approach, the feed forward network was used with error backpropagation algorithm to train the network in addition to the data preprocessing by using the normalized formula. The logistic activation function was used in the hidden layer and the error trial method was used to find the number of nodes in the hidden layer. Based on the root mean square error RMSE criterion, both ANN and Box-Jenkins filed to predict the wind speed. However, the extrapolation using periodic curve fitting was better and showed a better precision in prediction.

In (2009) [16] , the authors introduced a test called periodogram coefficient of variation test (PCOV) that is based on spectral analysis technique to check the consistency of periodogram ordinates throughout different sections resulted from broken sample record. The purpose of the check was to assess the weak stationarity, and the application was performed using simulated and experimental data. The results showed that the suggested test can be considered as a useful tool for exploratory analysis of time series.

In (2010) [55], four types of networks were considered in a comparison study to predict the daily closing price of the IBM stock. The networks were: Backpropagation Network BPN, Layer Recurrent Network LRN, Radial Basis Network RBN, and Generalized Regression Network GRN. The data that was used covered the period from (1980) to (1992) and it was obtained from <http://robjhdman.com>. The data was divided into training set (70%), and the 30% was considered for testing. In the BPN, the backpropagation algorithm was used to train the network. The training iterations number is set

to 3000 and the number of nodes in the hidden layer was increased to ensure that the RMSE decreased. When the optimal number of nodes in the hidden layer was identified, the model was further enhanced by changing the value of the learning rate and the momentum. The authors found that when increasing the number of nodes in the hidden layer, the RMSE started to decrease and then back to increase again, and the same happened when increasing the number of the input variables, and this can be generalized for all the sets of data.

A hybrid model of the ANN was suggested in (2010) [26]. This model included Box-Jenkins model which is known as the ARIMA model for prediction. Real data was used in the training consisting of three groups: Sunspot, Canadian Lynx and British Pound/US Dollar exchange rate. The one-step-ahead forecasting was used. The authors compared the following four models: ARIMA, ANN, Zhang's Hybrid and proposed model. In this hybrid model, the model consisted of the application of ARIMA model for a group of data by using Box-Jenkins methodology and then the residuals were applied to the network. The ARIMA model was implemented by using EVIEWS software and the network was implemented by using MATLAB. The Sunspot data was represented by the AR(9), and the Canadian Lynx was represented by AR(12), and the British Pound data was represented by the Random Walk Model. Based on the MSE values for the testing set of each group of data, the results showed that the suggested hybrid model is better than the other models in prediction.

A comparison was conducted by [44] in (2011) between neural network model and the Hidden Markov model to predict the Nigerian foreign exchange rate vs three different currencies: EURO, Yen and the GBP. The data were set of 800 observations that was downloaded from the famous currency exchange website [www.oanda.com](http://www.oanda.com). The Multi-Layer Perceptron Network was used with three hidden layers and learning rate of 0.10. The Backpropagation algorithm was used to train the neural network. The data set was divided into three categories: 500 observations as a training set, 200 observations as a validation set, and 100 observations was used as a testing set. Based on the MSE criterion for the forecast error, the results showed that the prediction efficiency was 81% while the Hidden Markov was 69%. MATLAB software was used in this work.

In (2012) [19], the author developed a statistical methodology to analyze non-stationary categorical time series in frequency domain through a spectral envelope tool that relies on summarizing information from spectrum

matrix and easily displaying the understood information. In addition, the author suggested Tree-Based Adaptive Segmentation (TBAS) method to estimate the spectral envelope based on Piecewise stationary process model where the application was performed on simulated and real data. The results showed that this methodology can work quite well in the segmentation and correctly divide the DNA sequence into coding and noncoding sequence. Moreover, this methodology provides the solution of the spectral analysis of nonstationary time series.

In (2012) [20], the authors investigated testing problems for spectral densities of time series with unequal sample sizes. They concluded that the natural approach to use periodogram-based distances for testing hypothesis about different spectras has not turned out to be very promising due to the inconsistency of the periodogram.

In (2013) [46], the authors studied the stationary processes using unequal sample sizes where they introduced theoretical detailed framework to test equality of spectral densities in bivariate case. Then, they generalized their approach to the m-dimensional case and apply it on cluster financial time series data with different sample lengths.

In (2014) [36], the authors made a comparison between three different approaches of prediction that included ANN, Ramirez-Verduzco and Knothe-Steidley method. Biodiesel Kinematic Viscosity data, consisting of 150 samples, was used. The ANN was used, and the backpropagation with Levenberg-Marquard optimization technique was used to train the network. The hyperbolic tangent activation function was used in the hidden layer and the linear activation function was used in the output layer. One hidden layer with four nodes was also used. The data, set of 73 samples, was divided into three main sets: training set 70%, testing set 15%, and 15% for the validation. Based on the MSE as a performance function, the results proved that the ANN is superior to the other approaches in prediction.

In (2015) [38], the author studied the problem of determining the number of differences required to stationary time series. He conducted a comparison between unit root tests, ADF and P.P, and stationary tests, KPSS and LM. The application was performed through the simulation using different sample sizes. The results showed similarity in the results between ADF and P.P, and between LM and KPSS tests for all sample sizes. In addition, the results showed that the increase in sample size leads to stationary time series

for ADF and P.P test while the increase of sample size has lower effect on stationarity of the series for KPSS and LM tests.

In (2016) [12], the authors optimized the averaged periodogram methodology, which relies on Fast Fourier transform, to estimate power spectral density of electrical activity signals in the atmosphere by suggesting a satisfactory method to choose the optimal number of segments to be averaged.

# **Chapter Two**

## **The Theoretical Side:**

- Time Series Analysis**
- Spectral Analysis**
- Artificial Neural  
Networks**



# The Theoretical Side

## 2.1 Time Series Analysis

### (2-1-1) Time Series [7, 37, 59]

Time series is sequential independent observations for a certain phenomenon during a specific time interval. Time series can be classified into two main categories as follow:

#### 1- Continuous Time Series

In this type of time series, the observations are continuous with time and can be represented in the following formula:

$$\{Y_t : t \in T \text{ and } T \text{ is continuous set}\}$$

#### 2- Discrete Time Series

In this type of time series, the observations are recorded at fixed time intervals that can be either minutes, hours, days, months, seasons or even years. This type of time series, the discrete time series, can be represented in the following formula:

$$\{Y_t : t \in T \text{ and } T \text{ is discrete set}\}$$

It is important to note that the observations of the discrete time series are recorded at equal intervals.

### (2-1-2) Autocovariance Function [7, 10, 37, 59]

This function is represented by the symbol  $\gamma_k$  and defined as follows:

$$\gamma_k = COV(W_t, W_{t+k}) \quad : k = \text{number of lags} \quad (2 - 1)$$

$$= E[(W_t - \mu)(W_{t+k} - \mu)] \quad (2 - 2)$$

Where  $\mu = E(W_t)$  for all  $t$

To estimate the autocovariance function of a time series  $W_t$  consisted of  $N$  observations, the following formula is used:

$$c_k = \frac{1}{N} \sum_{t=1}^N (W_t - \bar{W})(W_{t+k} - \bar{W}) \quad (2 - 3)$$

where  $\bar{W} = \frac{1}{N} \sum_{t=1}^N W_t$

### (2-1-3) Autocorrelation Function (ACF) [7, 10, 37, 59, 47]

This function is represented by the symbol  $\rho_k$  and defined as follows:

$$\rho_k = \frac{COV(W_t, W_{t+k})}{\sqrt{Var(W_t)}\sqrt{Var(W_{t+k})}} \quad : k = 0, 1, 2, \dots \quad (2-4)$$

$$= \frac{\gamma_k}{\gamma_0} \quad (2-5)$$

where  $\gamma_0 = Var(W_t)$  for all  $t$

To estimate the autocorrelation function for a time series  $W_t$  consisted of  $N$  observations, the following formula is used:

$$r_k = \frac{c_k}{c_0} \quad (2-6)$$

$$= \frac{\sum_{t=1}^{N-k} (W_t - \bar{W})(W_{t+k} - \bar{W})}{\sum_{t=1}^N (W_t - \bar{W})^2} \quad (2-7)$$

where  $r_k = \hat{\rho}_k$

The equation (2-6) is called the sample autocorrelation function SACF. The properties of the Autocorrelation Function are:

- a)  $\rho_k$  is an even function and symmetric about the lag zero.  
i.e.  $\rho_k = \rho_{-k}$
- b)  $\rho_0 = 1$
- c)  $|\rho_k| \leq 1$

### (2-1-4) Partial Autocorrelation Function (PACF) [7, 10, 37, 47, 59]

This function is an indicator that shows the relation between two variables  $W_t$  and  $W_{t+k}$  after removing the effect of all the variables between them ( $W_{t+1}, W_{t+2}, \dots, W_{t+k-1}$ ).

By using Durbin method, the partial autocorrelation function can be represented as follow:

$$\varphi_{k+1, k+1} = \frac{\rho_{k+1} - \sum_{j=1}^k \varphi_{kj} \rho_{k+1-j}}{1 - \sum_{j=1}^k \varphi_{kj} \rho_j} \quad (2-8)$$

and

$$\varphi_{k+1,j} = \varphi_{kj} - \varphi_{k+1,k+1}\varphi_{k,k+1-j} : k \geq 0 , j = 1,2, \dots, k \quad (2-9)$$

In order to estimate the partial autocorrelation function for a time series consisted of N observations, the following formula is used:

$$\hat{\varphi}_{k+1,k+1} = \frac{r_{k+1} - \sum_{j=1}^k \hat{\varphi}_{kj} r_{k+1-j}}{1 - \sum_{j=1}^k \hat{\varphi}_{kj} r_j} \quad (2-10)$$

and

$$\hat{\varphi}_{k+1,j} = \hat{\varphi}_{kj} - \hat{\varphi}_{k+1,k+1}\hat{\varphi}_{k,k+1-j} : k \geq 0 , j = 1,2, \dots, k \quad (2-11)$$

It is important to know that the plot of  $r_k$  against the lag  $k$ , and  $\hat{\varphi}_k$  against the lag  $k$  is called correlogram.

### (2-1-5) Stationarity [7, 10, 37, 47, 59, 60]

The time series is called stationary if the data is oscillating (fluctuating) around a certain level without any increasing or decreasing trends. The stationarity can be classified into two types:

#### 1- Strictly Stationary

The time series can be called as strictly stationary if the joint probability distribution for any set of observations is not affected by the shift of the time interval either forward or backward, which means:

$$P_r(w_{t_1}, w_{t_2}, \dots, w_{t_m}) = P_r(w_{t_1+k}, w_{t_2+k}, \dots, w_{t_m+k})$$

Where

$$t_m = \text{time}$$

#### 2- Weakly Stationary

The time series is called weakly stationary if meet the following conditions:

$$E(W_t) = \mu \quad (2-12)$$

$$Var(W_t) = E(W_t - \mu)^2 = \gamma_0 \quad (2-13)$$

$$COV(W_t, W_{t+k}) = E[(W_t - \mu)(W_{t+k} - \mu)] = \gamma_k \quad (2-14)$$

## (2-1-6) Time Series Models [37, 59]

The time series models are classified into three main types:

### 1- Autoregressive Model AR(p)

The stationary time series can follow the autoregressive model of the order p if it can be written in terms of its past observations  $W_{t-1}, W_{t-2}, \dots, W_{t-p}$  and the random error  $u_t$  as shown below:

$$W_t = \alpha_1 W_{t-1} + \alpha_2 W_{t-2} + \dots + \alpha_p W_{t-p} + u_t \quad (2 - 15)$$

$$W_t = \sum_{j=1}^p \alpha_j W_{t-j} + u_t \quad (2 - 16)$$

where

$\alpha_j$  :  $j^{\text{th}}$  Autoregressive parameters and  $j = (1, 2, \dots, p)$

$W_t$ : Deviation of the original time series from its mean

$u_t$ : Random error,  $u_t \sim N(0, \sigma^2)$

The autoregressive model of the order p, which is denoted by AR(p), can be written in terms of the backshift operator L as shown below:

$$\alpha_p(L)W_t = u_t \quad (2 - 17)$$

where  $\alpha_p(L) = (1 - \alpha_1 L^1 - \alpha_2 L^2 - \dots - \alpha_p L^p)$  and  $L^p W_t = W_{t-p}$

The stationary condition of the AR(p) is that the roots of the equation  $\alpha_p(L) = 0$  should lie outside the unit circle.

The properties of the AR(p) are:

$$A) E(W_t) = 0 \quad (2 - 18)$$

$$B) Var(W_t) = \frac{\sigma_u^2}{1 - \alpha_1 \rho_1 - \alpha_2 \rho_2 - \dots - \alpha_p \rho_p} \quad (2 - 19)$$

$$C) Y_k = \alpha_1 Y_{k-1} + \alpha_2 Y_{k-2} + \dots + \alpha_p Y_{k-p} \quad : k \geq 1 \quad (2 - 20)$$

$$D) \rho_k = \alpha_1 \rho_{k-1} + \alpha_2 \rho_{k-2} + \dots + \alpha_p \rho_{k-p} \quad : k \geq 1 \quad (2 - 21)$$

In practice, it can be recognized if the time series follows the AR(p) model or not through the behaviour of the SACF and the SPACF where the SACF of the AR(p) model is gradually decreasing in an exponential or

sinusoidal way according to the signal  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_p)$  to approach zero, while the SPACF of the AR(p) cut off after the lag p.

## 2- Moving Average Model MA(q)

The stationary time series  $W_t$  can follow the general moving average of the order q if it can be written in the form of its past random errors as follow:

$$W_t = u_t - \beta_1 u_{t-1} - \beta_2 u_{t-2} - \dots - \beta_q u_{t-q} \quad (2 - 22)$$

$$W_t = -\sum_{\tau=1}^q \beta_\tau u_{t-\tau} + u_t \quad (2 - 23)$$

where

$\beta_\tau$  :  $\tau^{\text{th}}$  moving average parameters and  $\tau = (1, 2, \dots, q)$

the moving average model of an order q, MA(q), can be written in terms of the backshift operator as follow:

$$W_t = \beta_q(L)u_t \quad (2 - 24)$$

Where  $\beta_q(L) = (1 - \beta_1 L^1 - \beta_2 L^2 - \dots - \beta_q L^q)$

The invertible condition of the MA(q) model is that the roots of the equation  $\beta_\tau(L) = 0$  should lies outside the unit circle.

The properties of the MA(q) model are:

$$A) E(W_t) = 0 \quad (2 - 25)$$

$$B) Var(W_t) = (1 + \beta_1^2 + \beta_2^2 + \dots + \beta_q^2)\sigma_u^2 \quad (2 - 26)$$

$$C) \gamma_k = \begin{cases} (-\beta_k + \beta_1\beta_{k+1} + \beta_2\beta_{k+2} + \dots + \beta_{k-q}\beta_q)\sigma_u^2 & : k = 1, 2, \dots, q \\ 0 & : k > q \end{cases} \quad (2 - 27)$$

$$D) \rho_k = \begin{cases} \frac{(-\beta_k + \beta_1\beta_{k+1} + \beta_2\beta_{k+2} + \dots + \beta_{k-q}\beta_q)}{1 + \beta_1^2 + \beta_2^2 + \dots + \beta_q^2} & : k = 1, 2, \dots, q \\ 0 & : k > q \end{cases} \quad (2 - 28)$$

In practice, it can be recognized if the time series follows the MA(q) model or not through the behaviour of the SACF and the SPACF where the SACF of the MA(q) model is cut off after the lag q, while the SPACF of the MA(q) gradually decreasing in an exponential or sinusoidal way according to the signal  $\beta = (\beta_1, \beta_2, \dots, \beta_q)$  to approach zero.

### 3- Autoregressive Moving Average Model ARMA(p,q)

There are many time series that cannot be represented by the AR or the MA models, so that there is a need to use the mixed model that is called the Autoregressive Moving Average Model.

It can be said that the stationary time series  $W_t$  follows the Autoregressive Moving Average Model of order (p,q) if it can be written in terms of its past observations and its past random errors as shown below:

$$W_t = \alpha_1 W_{t-1} + \alpha_2 W_{t-2} + \dots + \alpha_p W_{t-p} + u_t - \beta_1 u_{t-1} - \beta_2 u_{t-2} - \dots - \beta_q u_{t-q} \quad (2 - 29)$$

$$W_t = \sum_{j=1}^p \alpha_j W_{t-j} - \sum_{\tau=1}^q \beta_\tau u_{t-\tau} + u_t \quad (2 - 30)$$

The Autoregressive Moving Average Model of order (p,q), which is denoted by ARMA(p,q) can be written in terms of the backshift operator as shown in the formula below:

$$\alpha_p(L)W_t = \beta_q(L)u_t \quad (2 - 31)$$

The stationary condition of the ARMA(p,q) model matches the stationary condition of the AR(p) model, while the invertible condition of the ARMA(p,q) model matches the invertible condition of the MA(q) model.

The properties of the ARMA(p,q) model are:

$$A) E(W_t) = 0 \quad (2 - 32)$$

$$B) \gamma_k = \alpha_1 \gamma_{k-1} + \alpha_2 \gamma_{k-2} + \dots + \alpha_p \gamma_{k-p} \quad : k \geq q + 1 \quad (2 - 33)$$

$$C) \rho_k = \alpha_1 \rho_{k-1} + \alpha_2 \rho_{k-2} + \dots + \alpha_p \rho_{k-p} \quad : k \geq q + 1 \quad (2 - 34)$$

Both the SACF and the SPACF of the ARMA(p,q) decay gradually either in an exponential or sinusoidal way.

#### (2-1-7) Autoregressive Integrated Moving Average Model ARIMA(p,d,q) [7, 10, 37, 47]

Majority of real time series  $Y_t$  are not mean stationary. In order to transform those time series to the stationary status, the difference of an order d (normally d = 1 or d = 2) is calculated for those series. For that reason, the ARIMA model of the order (p,d,q) is used to represent the difference stationary series. The ARIMA(p,d,q) can be defined by the following formula:

$$\alpha_p(L)(1 - L)^d Y_t = \beta_q(L)u_t \quad (2 - 35)$$

$$\alpha_p(L)W_t = \beta_q(L)u_t \quad (2 - 36)$$

where

$$W_t = (1 - L)^d Y_t, \quad d = \text{Degree of difference}, \quad Y_t = \text{Original time series}$$

### **(2-1-8) Box – Jenkins Methodology [7, 10, 17, 37, 59]**

Box – Jenkins methodology, which was first produced by G. E. Box and G. M. Jenkins in 1970, is considered as one of the statistical tools that can be used to build a model for the time series and predict the variations that will happen in the future based on the patterns variations that were happened in the past intervals. One of the characteristics of this methodology is that it does not require the assumption of the independence between observations, but it make use of the correlation between observations through the ARIMA models which has the ability to invert lots of the real time series.

Box – Jenkins Methodology consists of four repetitive steps:

#### **1- Step one: Model Identification (Selection of an Initial Model)**

The identification stage is the most important stage that is used to analyse the time series, in which the model that will be used to represent the stationary data will be chosen. In this step, it will be decided whether the time series is stationary or not. This can be done in many ways:

##### **A) Plot Time Series [7]**

Plotting the raw data of the time series is very important to discover the outlier points, seasonality and trends whether it is increasing or decreasing. The time series is called stationary if there are no outlier points and no fluctuations around certain level.

##### **B) Correlogram Checking [10]**

The time series can be checked whether it is mean stationary or not by checking the correlogram for the raw data and especially the SACF. If the autocorrelation coefficients are not significantly different from zero after the first and the second lags then the time series is considered mean stationary.

##### **C) Tests Application [38, 59, 60]**

The following test can be used to check if the time series is stationary or not:

- a. Portmanteau Test
- b. Augmented Dickey – Fuller Test (ADF)
- c. Phillips – Perron Test (P.P)
- d. Kwiatkowski – Phillips – Schmidt – Shin Test (KPSS)

Augmented Dickey – Fuller test will be considered in this work.

The ADF is one of the unit root tests that is used to check the non-stationary in any time series and find the degree of differences by finding the number of roots that equal to 1 which represents the differences that are required to transform the time series to the stationary status.

The statistic test ADF, represented by  $t_\zeta$ , is defined in the formula below:

$$t_\zeta = \frac{\hat{\zeta}}{SE(\hat{\zeta})} \quad (2 - 37)$$

where

$$\zeta = \alpha - 1 \quad : \quad \alpha = \sum_{j=1}^p \alpha_j \text{ and } \hat{\zeta} \text{ is least squares estimator for } \zeta$$

The statistic test  $t_\zeta$ , is used to examine the following hypothesis:

$$H_0 : \zeta = 0 \text{ (i.e. there is a unit root and the series is non stationary)}$$

against

$$H_1 : \zeta < 0 \text{ (i.e. there is no unit root and the series is stationary)}$$

for the following three models:

$$\Delta Y_t = \zeta Y_{t-1} + \sum_{I=1}^k \vartheta_I \Delta Y_t + u_t \quad (2 - 38)$$

$$\Delta Y_t = \Omega_0 + \zeta Y_{t-1} + \sum_{I=1}^k \vartheta_I \Delta Y_t + u_t \quad : \quad \Omega_0 = \text{constant} \quad (2 - 39)$$

$$\Delta Y_t = \Omega_0 + \Omega_1 t + \zeta Y_{t-1} + \sum_{I=1}^k \vartheta_I \Delta Y_t + u_t \quad : \quad \Omega_1 = \text{constant} \quad (2 - 40)$$

$$\text{where } \vartheta_I = -\sum_{j=I+1}^p \alpha_j \quad : \quad I = 1, 2, \dots, k$$



If the value of the test statistic  $t_{\zeta}$ , is greater than the critical value, it can be concluded that the raw data is stationary.

If the above-mentioned methods are applied and they are founded that the time series is stationary then the model of the time series can be identified by comparing the SACF and the SPACF of the time series with the behaviour of the ACF and the PACF as shown in the table below:

**Table (2-1)**

**Behaviour of Theoretical ACF and PACF for Stationary Process[37]**

<b>Model</b>	<b>ACF</b>	<b>PACF</b>
MA(q)	Cuts off after lag q	Exponential decay and/or damped sinusoid
AR(p)	Exponential decay and/or damped sinusoid	Cuts off after lag p
ARMA(p,q)	Exponential decay and/or damped sinusoid	Exponential decay and/or damped sinusoid

However, If the above-mentioned methods are applied and it is founded that the time series is non-stationary then the difference of the series is taken in order to make the time series mean stationary and the Box-Cox transformation will be applied as follow:

$$T(Y_t) = \frac{Y_t^\lambda - 1}{\lambda} \quad (2 - 41)$$

where  $\lambda$  is the transformation parameter to make the series variance stationary .

**2- Step two: Parameter Estimation (Estimation the Model Coefficients) [59]**

To estimate the ARIMA model, we can use one of the following methods of estimating:

- The Moments Method
- The Ordinary Least Square Method
- The Conditional Maximum Likelihood Method
- The Exact Maximum Likelihood Method

we will use that the exact maximum likelihood method as follows:

If the stationary time series  $W_t$  follow AR(1) model then the maximum likelihood function of this series is written as follow:

$$L = (2\pi\sigma_u^2)^{-\frac{N}{2}}(1 - \alpha_1^2)^{\frac{1}{2}} * \exp\left[\frac{-1}{2\sigma_u^2}\left((1 - \alpha_1^2)W_1^2 + \sum_{t=2}^N(W_t - \alpha_1 W_{t-1})^2\right)\right] \quad (2 - 42)$$

$$\frac{\partial \ln(L)}{\partial \alpha} = 0$$

$$\frac{-\hat{\alpha}_1}{1 - \hat{\alpha}_1^2} \sigma_u^2 + \sum_{t=2}^N W_t W_{t-1} - \hat{\alpha}_1 \sum_{t=3}^N W_{t-1}^2 = 0 \quad (2 - 43)$$

by solving (2-43), we have the exact maximum likelihood estimator as follows:

$$\hat{\alpha}_1 = \frac{N - 2}{N - 1} \frac{\sum_{t=2}^N W_t W_{t-1}}{\sum_{t=3}^N W_{t-1}^2} \quad (2 - 44)$$

### 3- Step three: Diagnostic Checking (Analysing the Residuals) [17]

Before using the model for prediction, it is crucial to check the reliability of the model to represent the data by testing the compatibility of the model. This can be done by applying the available data to the model and then checking the estimated residuals which represent the white noise.

The check can be done by using one of the following approaches:

#### A- Correlogram Check for Residuals

If the autocorrelation coefficients of the residuals are lied within the period  $(\frac{-1.96}{\sqrt{N}}, \frac{+1.96}{\sqrt{N}})$ , which means that the autocorrelation coefficients of the residuals are not significantly different from zero, then it can be said that the residuals represent the white noise.

#### B- Portmanteau Tests (Goodness of Fit Tests) [17, 37]

The importance of the Portmanteau Tests is that they are used to check all of the autocorrelation coefficients in a small group rather than checking ever one of them individually. Each group normally has 20 or 24 lags.

The Portmanteau tests are used to check the hypothesis:

$H_0 : \rho_J = 0 \quad \text{for all } J = 1, 2, \dots, h$

against

$H_1 : \rho_J \neq 0 \quad \text{for some } J = 1, 2, \dots, h$

where  $h$  is the number of lags

The Portmanteau tests include:

### 1- Box – Pierce Test

This test is suggested in 1970 by Box and Pierce and it is defined as:

$$Q_{BP} = N \sum_{J=1}^h r_J^2 \quad (2 - 45)$$

where  $Q_{BP} \sim \chi^2(h - g)$  and  $g$  is the number of estimated parameters in the model.

If the P-value of the test statistic  $Q_{BP}$  is greater than the significant level, this leads to the conclusion that the residuals represent the white noise.

### 2- Ljung-Box (Modified Box – Pierce) Test

This test is suggested in 1978 by Ljung and Box and it is written as follow:

$$Q_{LB} = N(N + 2) \sum_{J=1}^h \left(\frac{1}{N - h}\right) r_J^2 \quad (2 - 46)$$

where  $Q_{LB} \sim \chi^2(h - g)$

If  $Q_{LB} > \chi^2(h - g)$  then the alternative hypothesis  $H_1$  is accepted.

The tests  $Q_{BP}$  and  $Q_{LB}$  can be applied in the first step of Box – Jenkins by applying  $g = 0$ .

### 4- Step four: Forecasting

After successfully passing the three steps then the model is ready for prediction and can be written as follow:

$$\begin{aligned}
\widehat{W}_t(\ell) &= E(W_{t+h}) \\
&= \alpha_1 E(W_{t+\ell-1}) + \dots + \alpha_{p+d} E(W_{t+\ell-p-d}) - \beta_1 E(u_{t+\ell-1}) - \\
&\quad \dots - \beta_q E(u_{t+\ell-q}) + E(u_{t+\ell})
\end{aligned} \tag{2-47}$$

Where  $\ell$  is the interval that will be predicted.

### (2-1-9) Model Selection Criteria [5, 17, 37, 60]

The aim to analyse the time series by using Box – Jenkins methodology is to find the unique model that will represent the data. For that reason, certain criteria to choose the right model is suggested as follow:

#### 1- Information Criterion of Akaike

This criterion (AIC) is calculated as follow:

$$AIC = Ln \left( \frac{1}{N} \sum_{t=1}^N \hat{u}_t^2 \right) + \frac{2}{N} g \tag{2-48}$$

#### 2- Bayesian Information Criterion

This criterion (BIC) is calculated as follow:

$$BIC = Ln \left( \frac{1}{N} \sum_{t=1}^N \hat{u}_t^2 \right) + \frac{Ln(N)}{N} g \tag{2-49}$$

This criterion is denoted by SC sometimes because it was suggested by Schwarz in 1978.

#### 3- Hannan – Quinn Criterion

This criterion (H-Q), which was suggested by E. J. Hannan and B. J. Quinn in 1979, is calculated as follow:

$$H - Q = Ln \left( \frac{1}{N} \sum_{t=1}^N \hat{u}_t^2 \right) + \frac{2Ln(Ln(N))}{N} g \tag{2-50}$$

## 2.2 Spectral Analysis

### (2-2-1) Spectral Analysis

The variations in time series of specific phenomenon are usually analysed in the time domain using the autocorrelation function. However, several time series contain some variations that rely on the frequency. To analyse such series, the spectral analysis methodology is utilized, which analyses the time series in the frequency domain using the spectrum. The spectrum of time series is a distribution of the series variance as a function of frequency. The goal of spectral analysis is to study and estimate the spectrum.

### (2-2-2) Harmonic Analysis [7, 10]

Let  $W_t$ , where  $(t = 1, 2, 3, \dots)$ , be a time series that consists of a cosine wave with a specific amplitude, it is possible to write the sinusoidal model of the series in the following formula:

$$W_t = G \cos(\lambda t + \phi) + u_t \quad (2 - 51)$$

Where

$G$ : Amplitude

$\phi$ : Phase

$\lambda$ : Frequency

$u_t$ : Purely stationary process

The resulted wave from the sinusoidal model shown in equation (2-51) fluctuates around zero mean while the high points of the wave represents the peaks while the low points represent the troughs. The vertical distance between the zero and any peak or between the zero and any trough represents the amplitude. It is worth mentioning that the distance between two subsequent peaks or troughs represents the wave length, so called the period, which is usually referred to as  $P$  and can be defined in the following formula:

$$P = \frac{1}{\lambda} \quad (2 - 52)$$

This means that the wave length is the reverse of the frequency. It is possible to write the model in equation (2 - 51) in the following formula:

$$W_t = G[\cos(\lambda t) \cos(\phi) - \sin(\lambda t) \sin(\phi)] + u_t \quad (2 - 53)$$

$$W_t = A \cos(\lambda t) + B \sin(\lambda t) + u_t \quad (2 - 54)$$

Where  $A = G \cos(\phi)$  ,  $B = -G \sin(\phi)$

The part  $[A \cos(\lambda t) + B \sin(\lambda t)]$  represents the harmonic component of the series  $W_t$  while the equation (2 – 54) is called the harmonic model.

**(2-2-3) Analysis of Harmonic Model [7, 9, 13]**

If the time series  $W_t$  consists of more than one harmonic component, its harmonic model is defined in the following formula:

$$W_t = G_i \cos(\lambda_i t + \phi_i) + u_t \quad : i = 0,1,2, \dots, h \quad (2 - 55)$$

$$W_t = A_i \cos(\lambda_i t) + B_i \sin(\lambda_i t) + u_t \quad : i = 0,1,2, \dots, h \quad (2 - 56)$$

Where  $A_i = G_i \cos(\phi_i)$  ,  $B_i = -G_i \sin(\phi_i)$

By estimating each of  $A_i$  and  $B_i$  by the least squares method, we can obtain the following:

$$\widehat{W}_t = \xi_0 + \sum_{j=1}^h [\xi_j \cos(\lambda_j t) + \Omega_j \sin(\lambda_j t)] + u_t \quad (2 - 57)$$

Where  $\xi_0 = \hat{A}_0$  ,  $\xi_j = \hat{A}_j$  and  $\Omega_j = \hat{B}_j$   $j = 1,2, \dots, h$

The equation (2 – 57) is called Fourier series representation while  $\lambda_j$  is called Fourier frequencies or standard frequencies and can be defined in the following formula:

$$\lambda_j = 2\pi \frac{j}{N} \quad j = 1,2, \dots, h \quad (2 - 58)$$

From (2 – 58), we conclude that the value of frequency depends on the sample size.

If the number of observation of  $W_t$  is odd, then the harmonic model can be written as in (2 – 57) where  $h = \frac{N-1}{2}$  while  $\xi_0, \xi_j$  and  $\Omega_j$  which represents the least squares estimators of  $A_0, A_j$  and  $B_j$  can be calculated for the model shown in (2 – 57) by using the following formulas:

$$\xi_0 = \hat{A}_0 = \frac{1}{N} \sum_{t=1}^N W_t \quad (2 - 59)$$

$$\xi_j = \hat{A}_j = \frac{2}{N} \sum_{t=1}^N W_t \cos(\lambda_j t) \quad (2 - 60)$$

$$\Omega_j = \hat{B}_j = \frac{2}{N} \sum_{t=1}^N W_t \sin(\lambda_j t) \quad (2 - 61)$$

Where  $j = 1, 2, \dots, h$

If the number of observations of the series is even, the harmonic model of the series can be written in the following formula

$$W_t = \xi_0 + \sum_{j=1}^{h-1} [\xi_j \cos(\lambda_j t) + \Omega_j \sin(\lambda_j t)] + \xi_h \cos(\lambda_h t) + u_t \quad (2 - 62)$$

Where  $h = \frac{N}{2}$  and  $\xi_0, \xi_h, \xi_j, \Omega_j$  can be calculated using the following formulas :

$$\xi_0 = \hat{A}_0 = \bar{W} \quad (2 - 63)$$

$$\xi_j = \hat{A}_j = \frac{2}{N} \sum_{t=1}^N W_t \cos(\lambda_j t) \quad (2 - 64)$$

$$\xi_h = \hat{A}_h = \frac{2}{N} \sum_{t=1}^N W_t (-1)^t \quad (2 - 65)$$

$$\Omega_j = \hat{B}_j = \frac{2}{N} \sum_{t=1}^N W_t \sin(\lambda_j t) \quad (2 - 66)$$

Where  $j = 1, 2, \dots, h - 1$

By using (2 - 55) and (2 - 57), we obtain the following:

$$\xi_j \cos(\lambda_j t) + \Omega_j \sin(\lambda_j t) = G_i \cos(\lambda_i t + \phi_i) \quad (2 - 67)$$

and from that, we can obtain the estimated amplitude of the  $j^{\text{th}}$  harmonic by using the following formula:

$$\hat{G}_j = \sqrt{\xi_j^2 + \Omega_j^2} \quad (2 - 68)$$

In addition, we can obtain the estimated phase of the  $j^{\text{th}}$  harmonic by using the following formula:

$$\hat{\Phi}_j = \tan^{-1} \left( \frac{-\Omega_j}{\xi_j} \right) \quad (2 - 69)$$

where

$$j = \begin{cases} 1, 2, \dots, h & \text{if } N \text{ is odd} \\ 1, 2, \dots, h - 1 & \text{if } N \text{ is even} \end{cases}$$

#### (2-2-4) Periodogram Analysis [7, 30, 39, 40]

Periodogram is a tool that is used to analyse time series that consist of sine and cosine waves with different frequencies. The importance of the periodogram comes from its capability in searching for hidden periodicities and examining the randomness of time series. The periodogram can be defined by a plot of the intensities  $I(\lambda_j)$  against the frequencies  $\lambda_j$ . The values of intensities can be calculated using the following formulas:

1- If the number of observations of the series is odd ,then

$$I(\lambda_j) = \frac{N}{2} (\xi_j^2 + \Omega_j^2) \quad : \quad j = 1, 2, \dots, h \quad (2 - 70)$$

1- If the number of observations of the series is even ,then

$$I(\lambda_j) = \frac{N}{2} (\xi_j^2 + \Omega_j^2) + I(\lambda_h) \quad : \quad j = 1, 2, \dots, h - 1 \quad (2 - 71)$$

Where  $I(\lambda_h) = N\xi_h^2$

There are other equivalent formulas to define the periodogram but they are different as shown below:

$$I(\lambda_j) = \frac{1}{\pi N} \left| \sum_{t=1}^N W_t e^{i\lambda_j t} \right|^2 \quad (2 - 72)$$

where  $j = 1, 2, \dots, h - 1$

and



$$I(\lambda_j) = \frac{1}{\pi N} \left[ \left( \frac{1}{N} \sum_{t=1}^N W_t \cos(\lambda_j t) \right)^2 + \left( \frac{1}{N} \sum_{t=1}^N W_t \sin(\lambda_j t) \right)^2 \right] \quad (2 - 73)$$

where  $j = 1, 2, \dots, h$

The difference in defining the periodogram between researchers comes from the use of negative frequencies or the use of cyclic frequencies  $f$  instead of the angular frequencies  $\lambda = 2\pi f$ .

### (2-2-5) The Search for Hidden Periodicities [7]

It was explained in (2-2-4) that the periodogram is used to search for hidden periodicities (systematic sinusoidal components) by examining the following hypothesis:

$$H_0 : W_t = \xi_0 + u_t$$

against

$$H_1 : W_t = \xi_0 + \xi \cos(\lambda_j t) + \Omega \sin(\lambda_j t) + u_t$$

To examine the hypotheses above, one of the following tests can be utilized.

#### (2-2-5-1) Fisher's Test [7, 39, 40, 47]

This test was first derived by Fisher in 1929. The test statistic for this test is given by the following formula :

$$T = \frac{I^{(1)}(\lambda_{(1)})}{\sum_{j=1}^m I(\lambda_j)} \quad (2 - 74)$$

Where

$$I^{(1)}(\lambda_{(1)}) = \max\{I(\lambda_j)\} \quad : \quad j = 1, 2, \dots, m \quad (2 - 75)$$

and  $m=h$  if  $I(\lambda_j)$  is defined as in (2 - 70) or  $m=h-1$  if  $I(\lambda_j)$  is defined as in (2 - 71).

If the calculated value of  $T$  is less than the critical value ( $b_{alpha}$ ), the null hypothesis should be accepted meaning that the time series represent Gaussian white noise, which means that there is no hidden periodic component in the time series. It is worth mentioning that if the critical value is not available for all  $m$  values, it can be calculated using the following formula:

$$P(T > b_{alpha}) = alpha \quad (2 - 76)$$

Where  $P(T > b_{alpha}) \approx m(1 - b_{alpha})^{m-1}$  and

$alpha$  = significant level

**(2-2-5-2) Cumulative Periodogram Test [7, 13, 30]**

In 1966, Bartlett explained that the cumulative periodogram can help in uncovering the hidden periodicities (periodic randomness). This can be performed by calculating the following test statistic:

$$C(\lambda_I) = \frac{\sum_{j=1}^I I(\lambda_j)}{\sum_{j=1}^m I(\lambda_j)} \quad : I = 1, 2, \dots, m \quad (2 - 77)$$

Where the value of  $m$  and the calculation of  $I(\lambda_j)$  can be chosen in the same way of Fisher test.

After calculating  $C(\lambda_I)$ , which is usually referred to as the normalized cumulative periodogram, Kolmogorov smirnov test is applied by determining its limits as follow:

The middle line, so called theoretical line, should be plotted from (0,0) to ( $\pi$ , 1), then the upper and lower limits should be plotted as well through the following formulas

$$\text{Upper limit} = \text{theoretical line} + \frac{K_{alpha}}{\sqrt{m}} \quad (2 - 78)$$

$$\text{Lower limit} = \text{theoretical line} - \frac{K_{alpha}}{\sqrt{m}} \quad (2 - 79)$$

Where  $K_{alpha}$  can be determined by using table (2 - 2)

**Table (2-2)**  
**Coefficients for Calculating Approximate Probability Limits for Cumulative Periodogram Test[7]**

<b>alpha</b>	0.01	0.05	0.10	0.25
<b><math>K_{alpha}</math></b>	1.63	1.36	1.22	1.02

If the value of  $C(\lambda_I)$  lies outside the two limits, lower and upper limits, we conclude that the time series contain hidden periodic sinusoidal component. However, if the value of  $C(\lambda_I)$  lies between the two limits, we conclude that the data do not contain periodic randomness.

### (2-2-6) Selection of Harmonic Model

In general, the harmonic model consists of one or more harmonic components, which all should be significant. Therefore, after calculating the periodogram using the formulas in (2 – 70) or (2 – 71) according to the number of the series observations and after ensuring that hidden periodic components exist in the data using one of the two tests explained in (2-2-5), it is necessary to determine the number of significant harmonic components that consist the model. To do that, one of the following two methods can be followed:

#### 1. Method one: [13, 59]

Application of  $F_k$  statistic:

$$F_k = \frac{(N - 3) I(\lambda_k)}{2 \sum_{\substack{j=1 \\ j \neq k}}^m I(\lambda_j)} \quad : \quad k = 1, 2, \dots, m \quad (2 - 80)$$

Where the  $F_k$  statistic follows the  $F$  distribution with two degrees of freedom (N-3) and 2.

$F_k$  statistic can examine the following hypotheses:

$$H_0 : \xi_k = \Omega_k = 0$$

against

$$H_1 : \xi_k \neq 0 \quad \text{or} \quad \Omega_k \neq 0$$

By comparing the calculated  $F_k$  with the critical  $F_{alpha}(2, N - 3)$ , we can accept the alternative hypothesis if  $F_k > F_{alpha}(2, N-3)$ , which means that the harmonic component  $k$  is significant.

#### 2. Method TWO: [2, 7, 59]

The contribution percentage of the harmonic component in the variance is calculated using the following formula:

A) If the number of observations of the stationary time series is odd, the following formula should be used

$$H.C.V. = \frac{\hat{G}_j^2}{2\sigma_W^2} * 100 \quad : \quad j = 1, 2, \dots, h \quad (2 - 81)$$

Where

$$\sigma_W^2 = \frac{1}{N} \sum_{t=1}^N (W_t - \bar{W})^2$$

$$= \frac{1}{N} \sum_{j=1}^h I(\lambda_j)$$

$$= \frac{1}{2} \sum_{j=1}^h \hat{G}_j^2$$

B) If the number of observations of the stationary time series is even, the following formula should be used:

$$H.C.V. = \frac{\hat{G}_j^2}{2\sigma_W^2} * 100 \quad : j = 1, 2, \dots, h - 1 \quad (2 - 82)$$

Where

$$\sigma_W^2 = \frac{1}{2} \sum_{j=1}^{h-1} \hat{G}_j^2 + \hat{G}_h^2$$

while the contribution percentage of the harmonic component  $j=h$  can be calculated using the following formula:

$$H.C.V. = \frac{\hat{G}_h^2}{\sigma_W^2} * 100 \quad (2 - 83)$$

After calculating H.C.V values, it should be ranked descending. Then, the statistic  $F_k$ , shown in method one, should be applied on the harmonic components that have contribution percentage equal to or more than 95%.

### (2-2-7) Spectrum [10, 13, 47]

The spectrum is a tool that is used to analyse time series in the frequency domain. This tool can explain the variance distribution of time series in specific domain of frequencies. The spectrum is mathematically defined as shown below:

Kintchine and Wiener explained that for each stationary random process with autocovariance function  $\gamma_k$ , there is a monotonically increasing function  $F(\lambda)$  that satisfy the equation

$$\gamma_k = \int_0^{\pi} \cos(\lambda k) d F(\lambda) \quad (2 - 84)$$

This equation is called the spectral distribution equation.

By looking at (2 – 84), we can find that the physical interpretation of the spectral distribution function is that  $F(\lambda)$  represents the contribution to the variance of the series, which is accounted for by frequencies in the range  $(0, \lambda)$ . This function can satisfy the following conditions:

$$1- F(\lambda) = 0 \text{ for } \lambda < 0 \quad (2 - 85)$$

$$2- F(\pi) = \sigma_W^2 \quad (2 - 86)$$

Because this function is monotonically increasing, it is possible to analyse it to two functions  $F_1(\lambda)$  and  $F_2(\lambda)$  Where

$$F(\lambda) = F_1(\lambda) + F_2(\lambda) \quad (2 - 87)$$

where  $F_1(\lambda)$  is a non-decreasing continuous function and  $F_2(\lambda)$  is a non-decreasing step function.

This analysis is similar to Wold decomposition where  $F_1(\lambda)$  is related to the indeterministic component of the process and  $F_2(\lambda)$  is related to the deterministic component of the process. Consequently, for purely indeterministic processes, we can get  $F_2(\lambda) = 0$ . As a result,  $F(\lambda)$  will be continuous function at  $[0, \pi]$  for indeterministic discrete time series.

By deriving  $F(\lambda)$  with respect to  $\lambda$  at  $(0, \pi)$ , it is possible to obtain the spectrum, referred to as  $g(\lambda)$ , as shown in the following formula:

$$g(\lambda) = \frac{d}{d\lambda} F(\lambda) \quad (2 - 88)$$

By substitution of (2 – 88) in (2 – 84), we can obtain

$$\gamma_k = \int_0^{\pi} \cos(\lambda k) g(\lambda) d(\lambda) \quad (2 - 89)$$

The formula in (2–89) represents the spectral representation for autocovariance function.

By reversing the above relationship, we can get

$$g(\lambda) = \frac{1}{\pi} \sum_{k=-\infty}^{\infty} \gamma_k e^{-i\lambda k} \quad : 0 < \lambda < \pi \quad (2 - 90)$$

From (2 – 90), we find that the spectrum represents Fourier transform of the autocovariance function. In addition, because  $\gamma_k$  is an even function, it is possible to write (2 – 90) in the following formula:

$$g(\lambda) = \frac{1}{\pi} \left[ \gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos(\lambda k) \right] \quad (2 - 91)$$

There are several formulas to define the spectrum that differs among each other by the constant multiple and the range of the definition of  $g(\lambda)$ .

The most common definition of the spectrum is

$$g(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k e^{-i\lambda k} \quad : \quad -\pi \leq \lambda \leq \pi \quad (2 - 92)$$

which can be obtained by reversing the following relationship:

$$\gamma_k = \int_{-\pi}^{\pi} e^{i\lambda k} g(\lambda) d(\lambda) \quad (2 - 93)$$

The characteristics of the spectrum

1-  $g(\lambda)$  is a continuous real-valued non-negative function

i.e.  $|g(\lambda)| = g(\lambda)$

2-  $g(\lambda)$  is periodic with period  $2\pi$

i.e.  $g(\lambda) = g(\lambda + 2\pi)$

3-  $g(\lambda)$  is a symmetric even function

i.e.  $g(\lambda) = g(-\lambda)$

4-  $\int_{-\pi}^{\pi} g(\lambda) d(\lambda) = \gamma_0 \quad (2 - 94)$

This means that if the spectrum is plotted, the fourth character means that the total area under the spectrum curve represents the process variance.

### (2-2-8) Spectrum of ARMA Models [47, 56, 59]

Let  $\gamma_k$ , where  $k = 0, \pm 1, \pm 2, \dots$ , a sequence of autocovariancies. Thus, the autocovariance generating function can be defined in the following formula:

$$\gamma(L) = \sum_{k=-\infty}^{\infty} \gamma_k L^k \quad (2 - 95)$$

by comparing equation (2 - 95) with equation (2 - 92), we can obtain

$$g(\lambda) = \frac{1}{2\pi} \gamma(e^{-i\lambda}) \quad (2 - 96)$$

If  $W_t$  is indeterministic stationary time series with a zero mean, it is possible to write it using Wold representation in the following formula:

$$W_t = \psi(L)u_t \quad : \quad \psi(L) = \sum_{i=0}^{\infty} \psi_i u_{t-i} \quad , \quad \psi_0 = 1 \quad (2 - 97)$$

In addition, the autocovariance generating function for this series can be defined by the following formula

$$\gamma(L) = \sigma_u^2 \psi(L)\psi(L^{-1}) \quad (2 - 98)$$

and by assuming that the series  $W_t$  follow ARMA model, we can obtain

$$\alpha_p(L)W_t = \beta_q(L)u_t \quad (2 - 99)$$

and by comparing equation (2 - 99) with equation (2 - 97), we obtain

$$\psi(L) = \frac{\beta_q(L)}{\alpha_p(L)} \quad (2 - 100)$$

and by substituting equation (2 - 100) in equation (2 - 98), we obtain

$$\gamma(L) = \sigma_u^2 \frac{\beta_q(L)}{\alpha_p(L)} \frac{\beta_q(L^{-1})}{\alpha_p(L^{-1})} \quad (2 - 101)$$

and thus

$$\gamma(e^{-i\lambda}) = \sigma_u^2 \left| \frac{\beta_q(e^{-i\lambda})}{\alpha_p(e^{-i\lambda})} \right| \quad (2 - 102)$$

and by substituting equation (2 - 102) in equation (2 - 96), we obtain

$$g(\lambda) = \frac{1}{2\pi} \sigma_u^2 \left| \frac{\beta_q(e^{-i\lambda})}{\alpha_p(e^{-i\lambda})} \right| \quad (2 - 103)$$

The formula in (2 - 103) represent the spectrum of the model ARMA (p,q).

### **(2-2-9) The Relationship Between the Spectrum and the Autocorrelation Function [7, 10]**

Most of the time, it is useful to use the normalized form of the spectrum that can be obtained as shown below

$$A(\lambda) = \frac{g(\lambda)}{\sigma_W^2} \quad (2 - 104)$$

and by using equation (2 – 92), we obtain

$$A(\lambda) = \frac{1}{2\pi} \left[ 1 + 2 \sum_{k=1}^{\infty} \rho_k \cos(\lambda k) \right] \quad : \quad -\pi < \lambda < \pi \quad (2 - 105)$$

$A(\lambda)$  is often referred to as the normalized spectrum or the spectral density function where  $\int_0^{\pi} A(\lambda) d\lambda = 1$  which means that  $A(\lambda)$  has the same characteristics of the ordinary probability density function.

### (2-2-10) Methods of Spectrum Estimation [10]

Spectral analysis is a name that is given to the methods of spectrum estimation or the spectral density function. There are several methods including:

#### 1- The First Method [59]

By using function (2 – 92), we can obtain the spectrum estimator of the time series  $W_t$  that consists of  $N$  observations as shown in the following formulas:

$$\hat{g}(\lambda) = \frac{1}{2\pi} \sum_{k=-(N-1)}^{(N-1)} c_k e^{-i\lambda k} \quad (2 - 106)$$

$$\hat{g}(\lambda) = \frac{1}{2\pi} [c_0 + 2 \sum_{k=1}^{N-1} c_k \cos(\lambda k) ] \quad (2 - 107)$$

Where equation (2 – 107) is called the sample spectrum.

#### 2- The Second Method [7]

It is possible to rewrite equation (2 – 70) of the periodogram as follow:

$$I(\lambda_j) = \frac{N}{2} (\xi_{\lambda} - i\Omega_{\lambda})(\xi_{\lambda} + i\Omega_{\lambda}) \quad (2 - 108)$$

$$I(\lambda_j) = \frac{N}{2} d_{\lambda}^* d_{\lambda} \quad (2 - 109)$$

Where  $d_{\lambda}^*$  is the complex conjugate of  $d_{\lambda}$ . By using the values of  $\xi_j, \Omega_j$  from (2 – 60) and (2 – 61), we obtain

$$d_{\lambda} = \frac{2}{N} \sum_{t=1}^N W_t [\cos(\lambda t) - i \sin(\lambda t)]$$

$$d_{\lambda} = \frac{2}{N} \sum_{t=1}^N W_t e^{-i\lambda t}$$



$$d_\lambda = \frac{2}{N} \sum_{t=1}^N (W_t - \bar{W}) e^{-i\lambda t} \quad (2 - 110)$$

$$d_\lambda^* = \frac{2}{N} \sum_{\hat{t}=1}^N (W_{\hat{t}} - \bar{W}) e^{+i\lambda \hat{t}} \quad (2 - 111)$$

and by substituting equation (2 – 110) and (2 – 111) in equation (2 – 109), we obtain

$$I(\lambda) = \frac{2}{N} \sum_{t=1}^N \sum_{\hat{t}=1}^N (W_t - \bar{W}) (W_{\hat{t}} - \bar{W}) e^{-i\lambda(t-\hat{t})} \quad (2 - 112)$$

Since

$$c_k = \sum_{t=1}^{N-k} (W_t - \bar{W}) (W_{t+k} - \bar{W})$$

The transformation  $k = t - \hat{t}$  transforms equation (2 – 112) into

$$I(\lambda) = 2 \sum_{k=-(N-1)}^{(N-1)} c_k e^{-i\lambda k} \quad (2 - 113)$$

$$I(\lambda) = 2 \left[ c_0 + 2 \sum_{k=1}^{N-1} c_k \cos(\lambda k) \right] : \pi \leq \lambda \leq \pi \quad (2 - 114)$$

and by comparing equation (2 – 114) with equation (2 – 107), the spectrum estimator as a function of the periodogram can be given in the following formula:

$$\hat{g}(\lambda) = \frac{1}{4\pi} I(\lambda) \quad (2 - 115)$$

### 3- The Third Method: Smoothing [47, 59]

The spectrum estimator that were obtained using the first and the second method is unbiased estimator, yet not consistent. To obtain a consistent estimator of the spectrum, it is possible to smooth the periodogram or the spectrum sample in the frequency domain.

The periodogram smoothing in the frequency domain is performed by smoothing the periodogram locally in the neighboring targeted frequency  $\lambda_i$  by using the weighting function  $V_N(\lambda_i)$  as follow:

$$\hat{g}_V(\lambda_j) = \sum_{i=-m_N}^{m_N} V_N(\lambda_i) \hat{g}(\lambda_j - \lambda_k) \quad (2 - 116)$$

Where

$m_N$  represents the number of frequencies used in the smoothing and  $\lambda_j$  represents Fourier frequencies while  $V_N(\lambda_i)$  satisfies the following conditions:

- 1-  $\sum_{i=-m_N}^{m_N} V_N(\lambda_i) = 1$
- 2-  $V_N(\lambda_i) = V_N(-\lambda_i)$
- 3-  $\lim_{n \rightarrow \infty} \sum_{i=-m_N}^{m_N} V_N^2(\lambda_i) = 0$

$V_N(\lambda_i)$  is called the spectral window.

It is also possible to smooth the sample spectrum using the same method but using the following formula

$$\hat{g}_V(\lambda_j) = \int_{-\pi}^{\pi} V_N(x) \hat{g}(\lambda - x) dx \quad (2 - 117)$$

Where  $x$  represents the frequencies in  $(-\pi, \pi)$  while  $V_N(x)$  satisfies the following conditions:

- a-  $\int_{-\pi}^{\pi} V_N(x) dx = 1$
- b-  $V_N(x) = V_N(-x)$
- c-  $\lim_{n \rightarrow \infty} \frac{1}{N} \int_{-\pi}^{\pi} V_N^2(x) dx = 0$

it is worth mentioning that the spectral window is also defined as Kernel.

It is also possible to smooth the spectrum in the time domain using the lag window by applying the weighting function on the sample autocovariances as shown in the following formula:

$$\hat{g}_D(\lambda) = \frac{1}{2\pi} \sum_{k=-(N-1)}^{(N-1)} D(\vartheta) c_k e^{-i\lambda k} \quad (2 - 118)$$

The weights of the weighting function are selected in a proportional way with the lag magnitude  $k$ . The weighting function is defined by the following formula:

$$D(\vartheta) = D\left(\frac{k}{M}\right) \quad (2 - 119)$$

Where  $M$  is called the truncation point, and its value depends on the sample size  $N$ .

The weighting function is often derived from a bounded even continuous function  $D(y)$  where it satisfies the following conditions:

- a-  $|D(y)| \leq 1$
- b-  $D(0) = 1$
- c-  $D(y) = D(-y)$
- d-  $D(y) = 0$  :  $|y| < 1$

The weighting function  $D(\vartheta)$  is defined as the lag window, which is related to the spectral window  $V_N(\lambda)$  by the following formula:

$$V_N(\lambda) = \sum_{k=-M}^M D_N(\vartheta) e^{-i\lambda k} \quad (2 - 120)$$

From equation (2 – 120), we can find that the spectral window represents Fourier transformation to the lag window. Thus, it is possible to obtain the lag window by reversing Fourier of the spectral window as shown below:

$$D_N(\vartheta) = \int_{-\pi}^{\pi} V_N(\lambda) e^{i\lambda k} d\lambda \quad : k = 0, \mp 1, \mp 2, \dots, \mp M \quad (2 - 121)$$

The most frequently used windows are:

### **A- The Rectangular (Truncated) Window**

The rectangular window is defined by the following formula

$$D_N^R(\vartheta) = \begin{cases} 1 & : |\vartheta| \leq M \\ 0 & : |\vartheta| > M \end{cases} \quad (2 - 122)$$

Where  $M \leq N - 1$  and M is called the window parameter.

The rectangular spectral window can be obtained by using equation (2 – 120) as shown in the following formula

$$V_N^R(\lambda) = \frac{1}{2\pi} \frac{\sin(\lambda(M + (1/2)))}{\sin(\lambda/2)} \quad (2 - 123)$$

The rectangular spectral window is also defined as Dirichlet Kernel or Fejer Kernel.

### **B- Bartlett's (Triangular) Window**

Bartlett's lag window is defined by the following formula

$$D_N^B(\vartheta) = \begin{cases} 1 - |\vartheta|/M & : |\vartheta| \leq M \\ 0 & : |\vartheta| > M \end{cases} \quad (2 - 124)$$

While the Bartlett's spectral window is obtained by the following formula :

$$V_N^B(\lambda) = \frac{1}{2\pi M} \frac{\sin(\lambda M/2)}{\sin(\lambda/2)} \quad (2 - 125)$$

### C- The Blackman – Tukey window

The Blackman – Tukey window is defined by the following formula:

$$D_N^T(\vartheta) = \begin{cases} 1 - 2\delta + 2\delta\cos(\pi\vartheta/M) & : |\vartheta| \leq M \\ 0 & : |\vartheta| > M \end{cases} \quad (2 - 126)$$

Where  $0 \leq \delta \leq 25$

While the Blackman – Tukey spectral window can be obtained using function (2 – 120) as shown in the following formula:

$$V_N^T(\lambda) = \delta V_N^R\left(\lambda - \frac{\pi}{M}\right) + (1 - 2\delta)V_N^R(\lambda) + \delta V_N^R\left(\lambda + \frac{\pi}{M}\right) \quad (2 - 127)$$

If  $\delta = 0.23$ , the Blackman – Tukey window is called the Tukey – humming or the humming window.

### D- Parzen Window

Parzen lag window is defined by the following formula:

$$D_N^P(\vartheta) = \begin{cases} 1 - 6(\vartheta/M)^2 + 6(|\vartheta|/M)^3 & : |\vartheta| \leq (M/2) \\ 2(1 - |\vartheta|/M)^3 & : (M/2) \leq |\vartheta| \leq M \\ 0 & : |\vartheta| > M \end{cases} \quad (2 - 128)$$

While the Parzen spectral window can be obtained using equation (2 – 120) as shown in the following formula:

$$V_N^P(\lambda) = \frac{3}{8\pi M^3} \left[ \frac{\sin(\lambda M/4)}{(1/2)\sin(\lambda/2)} \right]^4 \quad (2 - 129)$$

## 4- The Fourth Method[10, 35, 59]

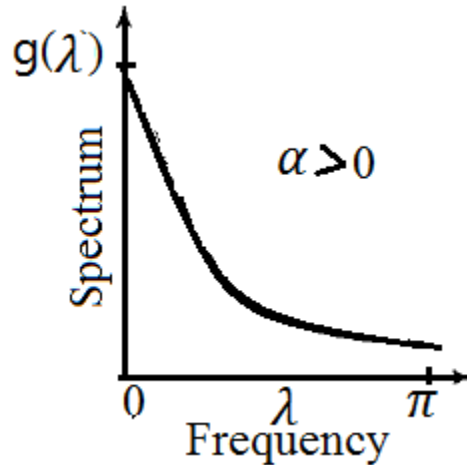
Based on equation (2 – 103), it is possible to estimate the spectrum for ARMA models as follow:

### a- The Spectrum of AR (1) Model

The spectrum estimation of AR (1) model can be obtained using the following formula:

$$\hat{g}(\lambda) = \frac{1}{2\pi} \sigma_u^2 \frac{1}{1 + \hat{\alpha}_1^2 - 2\hat{\alpha}_1 + \cos(\lambda)} \quad (2 - 130)$$

The spectrum shape of AR (1) model relies on  $\alpha_1$  signal. If  $\alpha_1 > 0$ , the spectrum is dominated by low frequency components. However, if  $\alpha_1 < 0$ , the spectrum is dominated by high frequency components. Figure (2 – 1) shows the spectrum of AR (1) model when  $\alpha_1 > 0$ .



**Figure (2-1)**

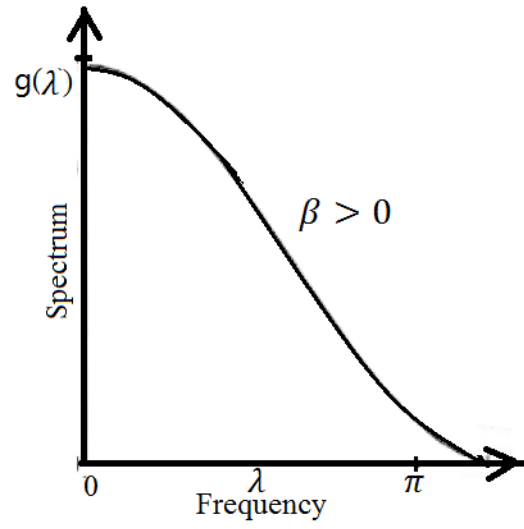
**Spectrum of AR(1) [59]**

**b- The Spectrum of MA (1) Model**

The spectrum estimation of MA (1) model can be obtained using the following formula:

$$\hat{g}(\lambda) = \frac{1}{2\pi} \sigma_u^2 (1 + \hat{\beta}_1^2 - 2\hat{\beta}_1 \cos(\lambda)) \quad (2 - 131)$$

The spectrum shape of MA (1) model also relies on  $\beta_1$  signal. If  $\beta_1 > 0$ , the spectrum is dominated by low frequency components. However, if  $\beta_1 < 0$ , the spectrum is dominated by high frequency components. Figure (2 – 2) shows the spectrum of MA (1) model when  $\beta_1 > 0$ .



**Figure (2-2)**

**Spectrum of MA(1) [59]**

## 2.3 Artificial Neural Networks

### (2-3-1) Artificial Neural Networks Definition

The Artificial Neural Network (ANN) is a computational technique designed to mimic the human brain behaviour in order to perform a particular task. ANN processes information in parallel throughout special processing units called nodes or neurons. The importance of ANN is its ability to learn and model linear and nonlinear relationships by using previous examples of both inputs and/or outputs of the relationships.

One of the main functions of the ANN is to understand the process in which the human brain processes and analyses information and obtains the data features by learning, differentiation, recognising, and makes use of all of these features in order to build a mathematical model which is capable of analysing data and then predicting outputs.

ANN is used in different fields starting from medicine where it is used for disease diagnosis and early warnings for instance.

### (2-3-2) Artificial Neural Network Structure [15, 65]

In general, the ANN consists of three levels:

#### 1- Input Level

This level consists of one layer called the input layer, which consists of a number of units called the input units.

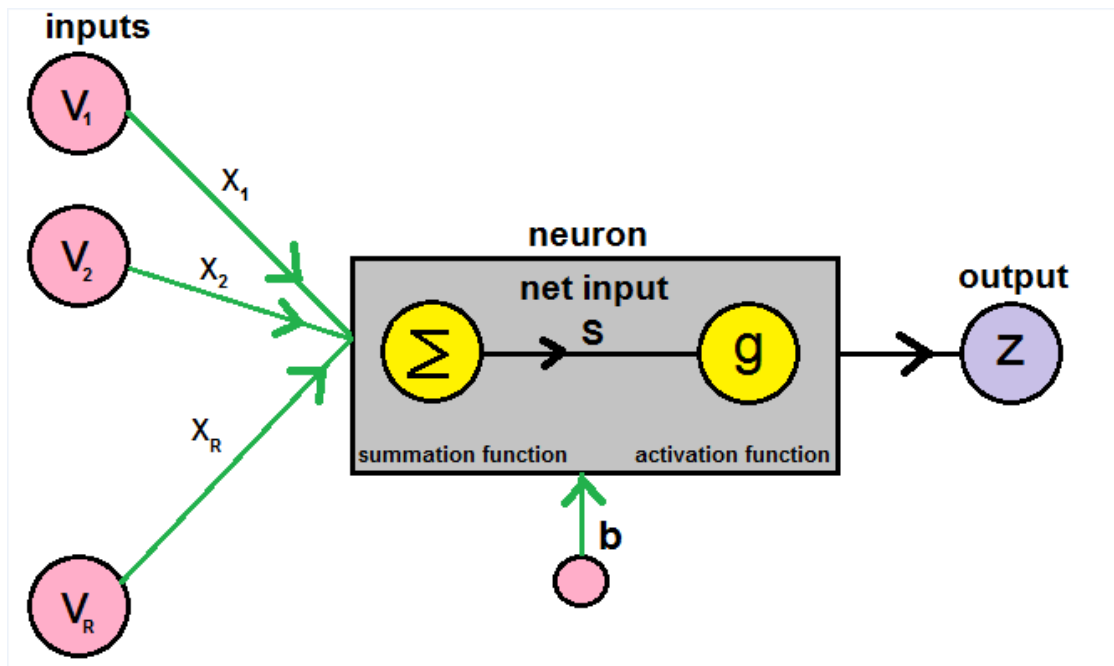
#### 2- Hidden Level

This level consists of one layer or more called the hidden layer(s). Every hidden layer has its own processing units called the hidden nodes.

#### 3- Output Level

This level consists of one layer called the output layer. This layer consists of a number of processing units called the output nodes.

These three levels are connected to each other by a connections strength called the weights that connects each level to the next or previous level. The processing units, which is called neurons or nodes, consist of two parts as shown in figure (2-3):



**\*Figure (2-3)**

### **Artificial Neuron**

#### **a: Summation Function**

This function collects the input signals to the nodes and puts it in a linear combination in order to produce one signal called the net input as shown below:

$$\begin{aligned}
 s &= x_1 v_1 + x_2 v_2 + \dots + x_R v_R \\
 &= \sum_{j=1}^R x_j v_j
 \end{aligned}
 \tag{2-132}$$

Where

$s$ : net input

$v_j$  : input variable

$x_j$  : weights

Normally, a bias node is added to the inputs nodes, and then equation (2-132) can be written as:

$$s = \sum_{j=1}^R x_j v_j + b \quad :
 \tag{2-133}$$

where  $b$  represents the bias.

#### **\*Work of the researcher**



## b: Activation Function

This function represents the second part of the processing unit, which is a mathematical formula that can be linear or nonlinear. This function processes the net input ( $s$ ) and then produces the output of the processing node in a range that fits with the limits of the activation function of that node. The importance of the activation function is to prevent the output of the processing node and then the output of the particular layer from exceeding the limits of the activation function and reaching high values that can stop the learning process or the overall net to be collapsed [21]. The activation function is used only in the hidden layers and the output layer, and it is sometimes called transformation or squashing.

In theory, every continuous function can be considered an activation function if it is differentiable but in practice every bounded, monotonically increasing and differentiable function can be used as an activation function [11].

The commonly used activations functions in the processing nodes are:

### 1- Linear Function

$$g(s)=s$$

(2-134)

$$g(s)=1$$

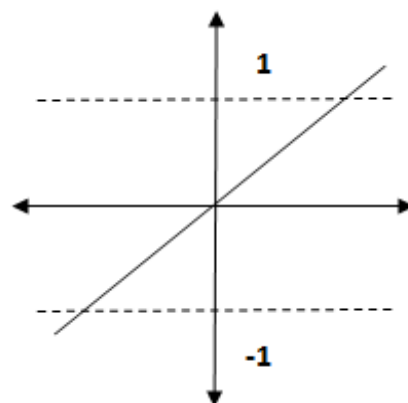


Figure (2-4)

Linear Function

## 2- Sigmoid Function

$$g(s) = \frac{1}{1 + e^{-s}} \quad : \quad 0 < g(s) < 1 \quad (2 - 135)$$

$$\dot{g}(s) = g(s)(1 - g(s))$$

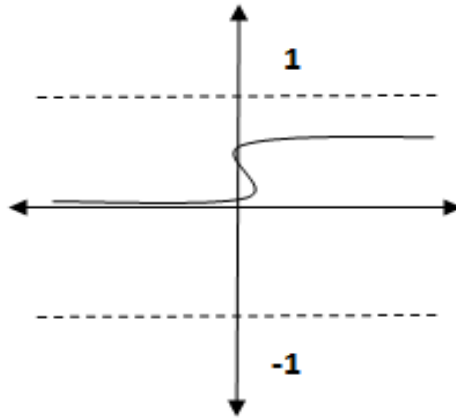


Figure (2-5)

## Sigmoid Function

## 3- Hyperbolic Function

$$g(s) = \frac{e^s - e^{-s}}{e^s + e^{-s}} \quad : \quad -1 < g(s) < 1 \quad (2 - 136)$$

$$\dot{g}(s) = 1 - g^2(s)$$

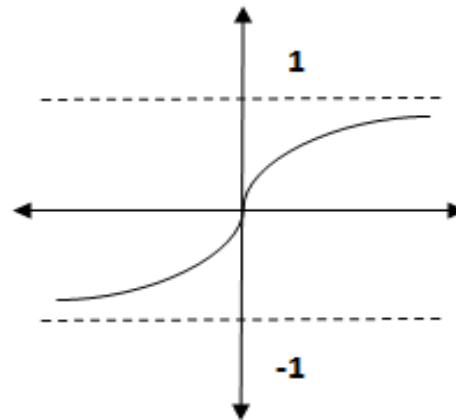


Figure (2-6)

## Hyperbolic Function

It is important to mention that the activation function in the input layer has nothing to do with the input signal and there is no processing at the input layer of the neural network. The only function of the input layer is to collect the input and pass it to the next layer (hidden layer). The number of the input units matches the number of the input variables of the neural network.

### **(2-3-3) Artificial Neural Networks Architecture [15, 53, 65]**

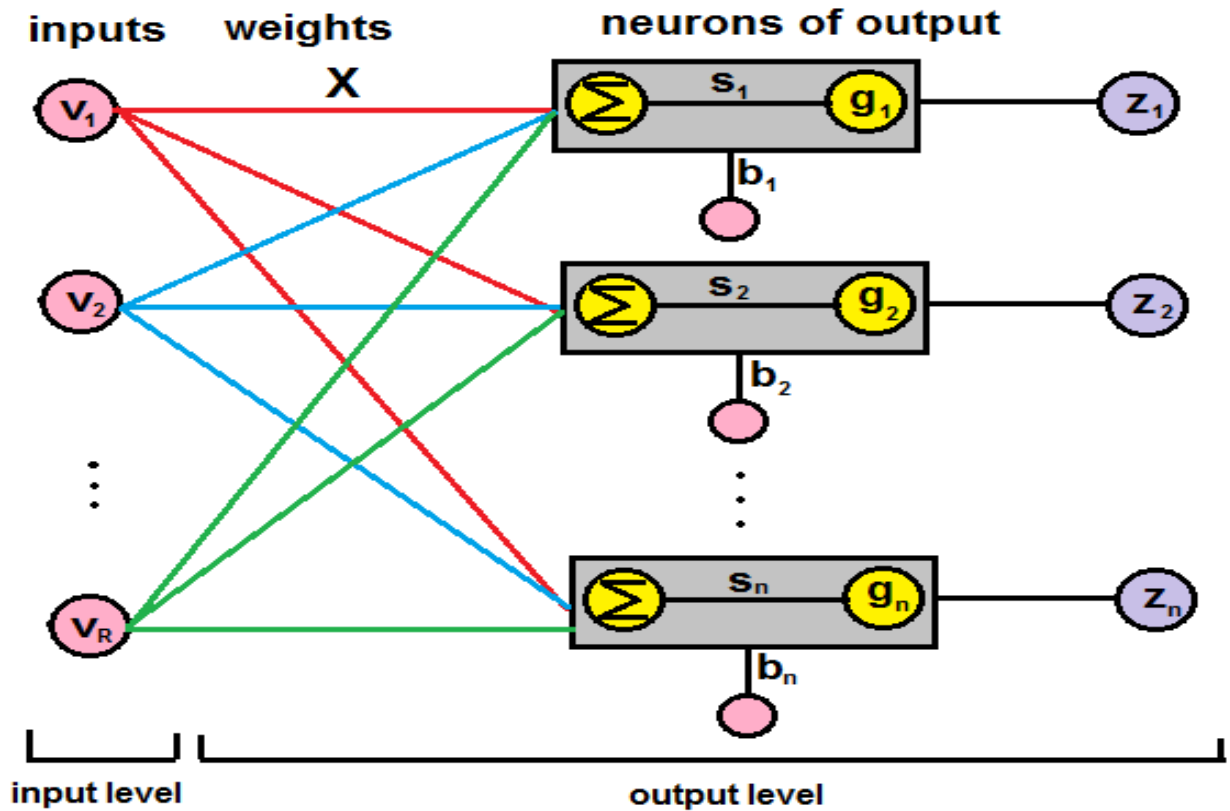
The architecture of the artificial neural networks means how the nodes form the layers of the network and the layers are connected to each other. The ANN can be classified according to the number of the layers into two main types:

#### **(2-3-3-1) Single Layer Networks [15, 53]**

This type of ANN has only input and output layers, which means that there is no hidden layer. This type of ANN has one layer of weights that connects the input and the output layers. When applying the input signal  $v_j$  to this type of ANN, the output signal will be calculated as follow:

$$\begin{aligned} z_i &= g(s_i) \\ &= g\left(\sum_{j=1}^R x_j v_j + b_i\right) \quad : \quad i = 1, 2, \dots, n \end{aligned} \quad (2 - 137)$$

This type of ANN is used to solve the problem of patterns classification. Perceptron and Adaline are types of this ANN. Figure (2-7) show the single-layer ANN.



**\*Figure (2-7)**  
**Single-Layer Artificial Neural Network**

**(2-3-3-2) Multi-Layer Networks [15, 65]**

This type of ANN consists of three levels: the input level, the hidden level and the output level. The hidden level can be one hidden layer of nodes or more. This type of ANN has the ability to solve complex problems that cannot be solved by the single layer network. The reason for that is the existence of the hidden level that contains one or more hidden layers that gives the ANN more ability and flexibility to model the relationship between the inputs and the outputs. This type of ANN takes much more time to learn than the single layer ANN, but it gives much more powerful models than the single layer ANN.

**\*Work of the researcher**

When applying input to this type of ANN, the out is calculated as follow:

$$\mathbf{Z}^{m+1} = g^{m+1}(\mathbf{X}^{m+1}\mathbf{Z}^m + \mathbf{b}^{m+1}) \quad : \quad m = 0,1,2, \dots, N \quad (2 - 138)$$

Where:

N: number of hidden layers

$\mathbf{Z}^{m+1}$ : output vector of the layer m+1

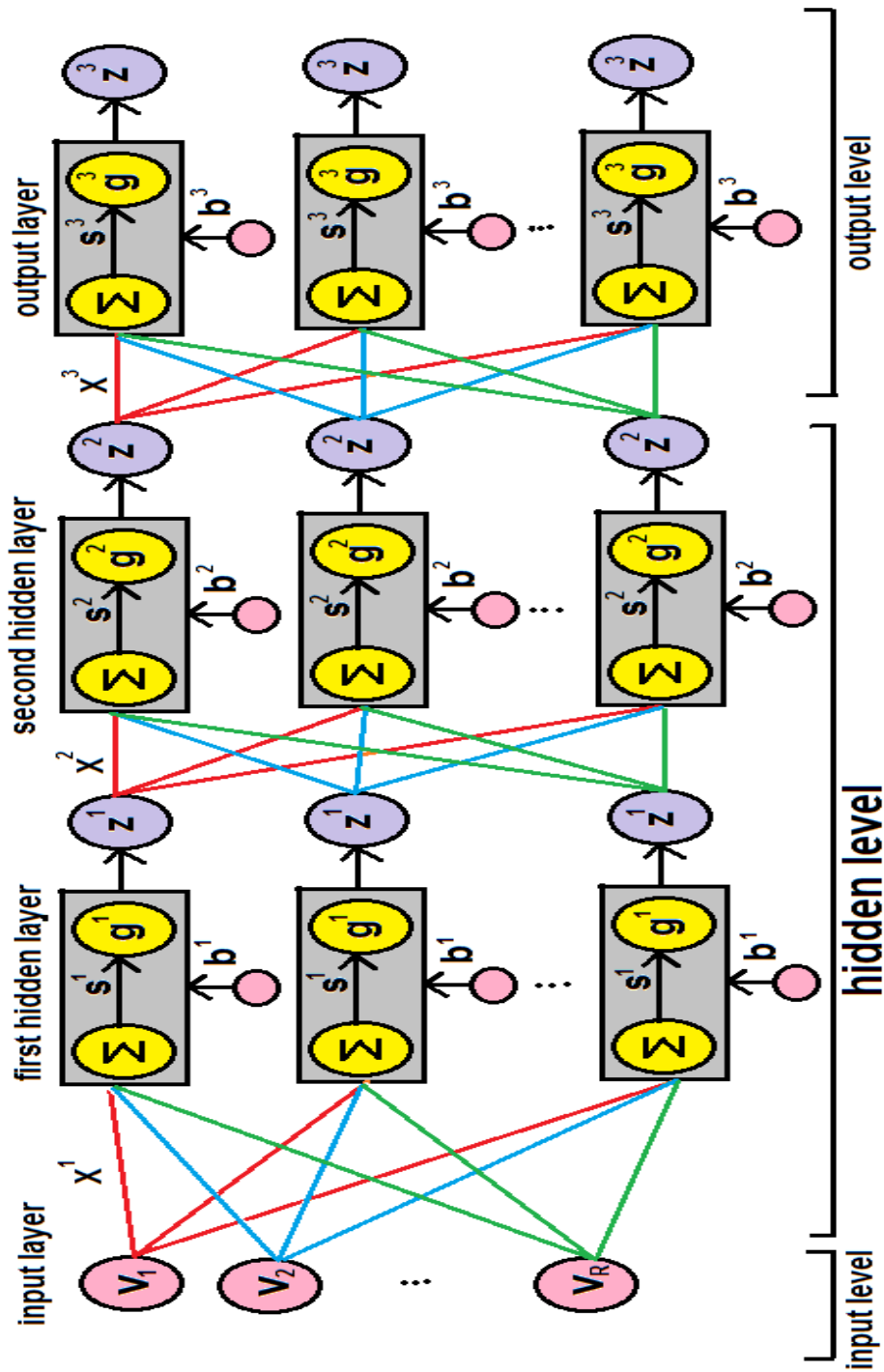
$g^{m+1}$ : activation function of the layer m+1

$\mathbf{b}^{m+1}$ : bias vector of the layer m+1

$\mathbf{X}^{m+1}$ : weights matrix of the layer m+1

$\mathbf{Z}^0 = \mathbf{V}$ , and  $\mathbf{V}$  is the input vector.

The multi-layer network is shown in figure (2-8) below:



**\*Figure (2-8)**

**Multi-layer Artificial Neural Network**

**\*Work of the researcher**

#### **(2-3-4) Training Algorithms [15, 28, 45, 65]**

The training in ANN means adjusting the values of the weights and biases in the network, and this process is sometimes called the learning process. The main aim of the training is to teach the network to do specific tasks.

The training algorithms are divided into three main categories:

##### **(2-3-4-1) Supervised Training [15, 65]**

During the supervised training, the training data is applied as two pairs of vectors: the first pair is the input vector and the second pair is the desired vector. When the input vector is applied to the network, the ANN starts comparing the output of the network with the desired data. The ANN then uses the comparison result (error) to update the weights of the network and the biases in order to reduce the error or the difference between the network output and the desired output.

The error is calculated as follow

$$\text{Error} = \text{target output} - \text{actual output} \quad (2-139)$$

Perceptron learning rule is an example of the supervised training.

##### **(2-3-4-2) Unsupervised Training [15, 45]**

The unsupervised learning is used when there is only an input vector and there is no output vector. In this case, the ANN will figure out the properties of the input values and then stimulate its nodes and weights based on the training algorithm and the input vectors.

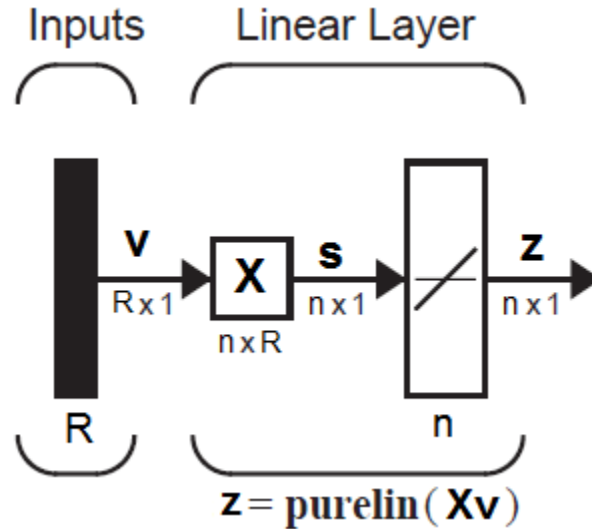
There some types of ANN that uses this type of training such as Hamming Network. The unsupervised training sometimes called the competitive training.

##### **(2-3-4-3) Reinforcement (Graded) Training [15, 28]**

This type of training is considered an intermediate training form between the supervised and unsupervised training. Data is applied to the network in this type of training as an input vector only, exactly as in the unsupervised training, and then the training algorithm is applied which is considered the performance indicator as mentioned in the supervised training. This type of training is not commonly used, and it is used in control systems applications and the self-organising neural training.

**(2-3-5) Hebb Learning Rule [15, 53, 65]**

Hebb learning rule is the first rule that had been used to train the ANN in 1949, proposed by Donald Hebb to train the network and adapt the weights of it. The linear associator is one of the networks that are trained by the Hebb Learning Rule as shown in figure below:



**Figure (2-9) [15]  
Linear Associator**

After applying the input vector  $V$  for the linear associator, then the output vector  $Z$  is calculated as follow:

$$Z = XV \quad (2-140)$$

where

$X$  : weights matrix

The output vector can also be obtained as follow:

$$z_i = \sum_{j=1}^R x_{ij} v_j \quad (2-141)$$

It can be concluded from equation (2-141) that the weights vector represents the connection between the input vector  $V$  and the output vector  $Z$ . This explains the Hebb's postulate: "if two neurons on either side of a synapse are activated simultaneously, the strength of the synapse will increase". This mean that if both  $Z_i$  and  $V_j$  are positive then  $X_{ij}$  should increase. Thus, the mathematical expression of the Hebb's postulate will be:

$$x_{ij}^{new} = x_{ij}^{old} + \eta f_i(z_{iq}) g_j(v_{jq}) \quad (2-142)$$

Where

$v_{jq}$  : represents the  $j^{\text{th}}$  element in the input vector  $V_j$

$z_{iq}$  : represents the  $i^{\text{th}}$  element in the input vector  $Z_i$

$\eta$  : positive constant called the learning rate

$f_i g_j$  : activation functions



for simplification, equation (2-142) can be written as:

$$x_{ij}^{new} = x_{ij}^{old} + \eta z_{iq} v_{jq} \quad (2-143)$$

From equation (2-143), it can be concluded that the change in the weight is not only increased when both  $Z_i$  and  $V_j$  are positive, but also increased when both of them are negative, and then decreased when they have different polarity (one is positive and the other is negative).

Hebb's Rule is classified into two types:

### (2-3-5-1) Unsupervised Hebb Learning Rule [15, 53]

Equation (2-143) is considered as a definition to the Hebb Rule for unsupervised learning because this method does not require any information about the desired output. This rule can be written in vector form as:

$$\mathbf{X}^{new} = \mathbf{X}^{old} + \mathbf{z}_q \mathbf{v}_q^T \quad (2-144)$$

Where  $\eta$  was assumed to be 1 for simplification.

### (2-3-5-2) Supervised Hebb Learning Rule [15, 65]

The supervised Hebb Learning was obtained by replacing the network output in equation (2-143) by the desired output as follow:

$$x_{ij}^{new} = x_{ij}^{old} + d_{iq} v_{jq} \quad (2-145)$$

Where  $\eta$  is assumed to be 1 for simplification.

Equation (2-145) can be re-written in a vector notation:

$$\mathbf{X}^{new} = \mathbf{X}^{old} + \mathbf{d}_q \mathbf{v}_q^T \quad (2-146)$$

If (input/output) pairs,  $\{ \mathbf{v}_1, \mathbf{d}_1 \}$ ,  $\{ \mathbf{v}_2, \mathbf{d}_2 \}$ , ...,  $\{ \mathbf{v}_Q, \mathbf{d}_Q \}$ , was applied to the neural network assuming that the initial weights equal to zero so equation (2-146) can be re-written as follow:

$$\mathbf{X} = \sum_{q=1}^Q \mathbf{d}_q \mathbf{v}_q^T \quad (2-147)$$

and equation (2-147) can be written in matrix form

$$\mathbf{X} = \mathbf{D} \mathbf{V}^T \quad (2-148)$$

Where  $\mathbf{D} = [\mathbf{d}_1 \ \mathbf{d}_2 \ \dots \ \mathbf{d}_Q]$  is the desired output, and

$\mathbf{V} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_Q]$  is the input vector.

If  $\mathbf{v}_k$  is an input to the network, then the output vector is:

$$\mathbf{z} = \mathbf{X} \mathbf{v}_k$$

$$= \left( \sum_{q=1}^Q \mathbf{d}_q \mathbf{v}_q^T \right) \mathbf{v}_k$$

$$= \sum_{q=1}^Q \mathbf{d}_q (\mathbf{v}_q^T \mathbf{v}_k) \quad (2 - 149)$$

Now there are two cases:

**Case 1:** if the vectors  $\mathbf{v}_q$  are orthonormal then:

$$\mathbf{z} = \mathbf{d}_k \quad (2 - 150)$$

This means that the output of the network matches the desired output, which means that Hebb's rule will produce the correct output for every input.

**Case 2:** if the vectors  $\mathbf{v}_q$  are not orthonormal then:

$$\mathbf{z} = \mathbf{d}_k + \sum_{q \neq k}^Q \mathbf{v}_q^T \mathbf{v}_k \quad (2 - 151)$$

Where the second term on the right hand side of equation (2 - 161) represents the error which means that Hebb's rule will not produce the right output in this case.

One of the methods that are used to reduce the error caused by the input variables in case of the non-orthogonal is the Pseudoinverse rule as follow:

$$\mathbf{A}^+ = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \quad (2 - 152)$$

Where  $\mathbf{A}^+$  represents the Pseudoinverse of the matrix  $\mathbf{A}$ . The importance of this rule is to decrease the performance index:

$$H(\mathbf{X}) = \sum_{q=1}^Q \|\mathbf{d}_q - \mathbf{X} \mathbf{v}_q\|^2 \quad (2 - 153)$$

When the input vectors  $\mathbf{v}_q$  are non-orthogonal, by solving the equation

$$\mathbf{X} \mathbf{V} = \mathbf{D} \quad (2 - 154)$$

by using the pseudoinverse rule to find the weights matrix  $\mathbf{X}$ .

One of the main problem of the Hebb's Rule is that it leads to high weights. There are some ways that can be followed to reduce the weights:

a- Make the learning rate lower than 1 as follow

$$\mathbf{X}^{new} = \mathbf{X}^{old} + \eta \mathbf{d}_q \mathbf{v}_q^T \quad : \eta < 1 \quad (2 - 155)$$

b- Add a decay term that makes the learning rate behave like a smoothing filter

$$\mathbf{X}^{new} = (1 - \gamma)\mathbf{X}^{old} + \eta \mathbf{d}_q \mathbf{v}_q^T \quad (2 - 156)$$

If  $\gamma = 0$  then Hebb's rule will be a standard rule, and if  $\gamma = 1$  then Hebb's rule will forget the old input patterns and only remember the most recent patterns, so that the  $\gamma$  is normally a positive constant with a value of less than 1 to prevent the weights from a non-bounded increment.

Equation (2 - 155) can be re-written by replacing the desired output by the difference between the actual output and the desired output as follow:

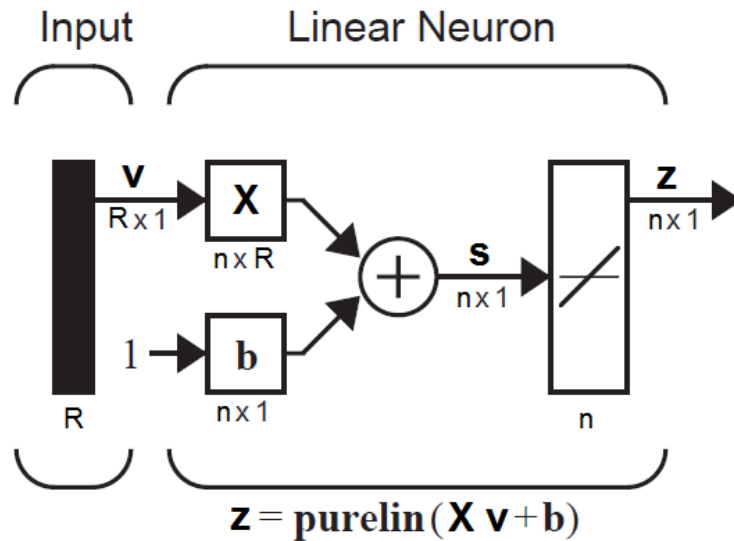
$$\mathbf{X}^{new} = \mathbf{X}^{old} + \eta(\mathbf{d}_q - \mathbf{z}_q)\mathbf{v}_q^T \quad (2 - 157)$$

Equation (2 - 157) is called the Delta Rule, which is working to reduce the mean square error and then provide results similar to the pseudoinverse results.

### **(2-3-6) Least Mean Square (LMS) Algorithm [15, 28, 65]**

This algorithm is an example of the supervised training and it is considered as an introduction to the back propagation algorithm. This algorithm was proposed in 1960 by Widrow-Hoff so that it is known also as a Widrow-Hoff algorithm. The LMS algorithm is considered as more powerful and more applicable than the perceptron algorithm.

When applying the input vectors and the targeted vectors as pairs  $\{v_1, d_1\}$ ,  $\{v_2, d_2\}$  ...  $\{v_Q, d_Q\}$  to the ANN and calculating the network output, then the LMS algorithm starts updating the weights of the network to reduce the mean square error. The LMS algorithm can be applied to many networks such as the Adaptive Linear Neural Network which is written as ADALINE and has a basic structure that matches the perceptron except that the ADLINE has a linear activation function as shown in figure below



**Figure (2-10) [15]  
Adaptive Linear Neural Network**

The output of the ADALINE can be calculated as follow:

$$\begin{aligned} \mathbf{z} &= g(\mathbf{X}\mathbf{v} + \mathbf{b}) \\ &= \mathbf{X}\mathbf{v} + \mathbf{b} \end{aligned} \quad (2 - 158)$$

The weights and biases of the ADALINE can be updated by using LMS algorithm.

By using the vectors

$$\mathbf{P} = \begin{bmatrix} \mathbf{v} \\ 1 \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} \mathbf{x} \\ b \end{bmatrix} \quad : \quad \mathbf{x} = [x_{1,1} \ x_{1,2} \ \dots \ x_{1,R}]^T$$

Where:

$\mathbf{P}$  = the input vector

$\mathbf{Y}$  = the parameters vector (weights and biases)

The performance index (mean square error) can be written as:

$$\begin{aligned} H(\mathbf{Y}) &= E[e^2] \\ &= E[(d - \mathbf{Y}^T \mathbf{P})^2] \end{aligned} \quad (2 - 159)$$

The main idea of the LMS algorithm is to estimate the mean square error  $H(\mathbf{Y})$  by:

$$\begin{aligned} \hat{H}(\mathbf{Y}) &= (d(k) - z(k))^2 \\ &= e^2(k) \end{aligned} \quad (2 - 160)$$

where the expectation of error squared in equation (2 – 159) is replaced by the squared error in each iteration  $k$  as shown in equation (2 – 160).

The gradient estimate, or what is called as the stochastic gradient, can be estimated as follow:

$$\hat{\nabla}H(\mathbf{Y}) = \nabla e^2(k) \quad (2 - 161)$$

The first element  $R$  of  $\nabla e^2(k)$  are derivatives of the weights, and  $R+1$  of  $\nabla e^2(k)$  are derivatives of the biases which means:

$$\begin{aligned} [\nabla e^2(k)]_j &= \frac{\partial}{\partial x_{1,j}} e^2(k) && : j = 1, 2, \dots, R \\ &= 2e(k) \frac{\partial}{\partial x_{1,j}} e(k) \end{aligned} \quad (2 - 162)$$

$$\begin{aligned} [\nabla e^2(k)]_{R+1} &= \frac{\partial}{\partial b} e^2(k) \\ &= 2e(k) \frac{\partial}{\partial b} e(k) \end{aligned} \quad (2 - 163)$$

where

$$\frac{\partial}{\partial x_{1,j}} e(k) = -v_j(k) \quad (2 - 164)$$

$$\frac{\partial}{\partial b} e(k) = -1 \quad (2 - 165)$$

Thus, the gradient of the mean square error can be written as:

$$\begin{aligned} \hat{\nabla}H(\mathbf{Y}) &= \nabla e^2(k) \\ &= -2e(k)\mathbf{P}(k) \end{aligned} \quad (2 - 166)$$

By substituting the value of the estimated gradient in the steepest descent algorithm which has the following formula:

$$\mathbf{Y}_{k+1} = \mathbf{Y}_k - \eta g_k \quad (2 - 167)$$

Where

$$g_k = \nabla H(\mathbf{Y})|_{\mathbf{Y} = \mathbf{Y}_k}$$

We get

$$\mathbf{Y}_{k+1} = \mathbf{Y}_k + 2\eta e(k)\mathbf{P}(k) \quad (2 - 168)$$

$${}_1\mathbf{x}(k + 1) = {}_1\mathbf{x}(k) + 2\eta e(k)\mathbf{v}(k) \quad (2 - 169)$$

$$b(k + 1) = b(k) + 2\eta e(k) \quad (2 - 170)$$

Both equations (2 – 169) and (2 – 170) represent the LMS algorithm for updating the weights and the biases for any ANN with single layer (including ADALINE). If ANN with single layer has more than one output node, then the LMS algorithm can be written as:

$${}_i\mathbf{x}(k + 1) = {}_i\mathbf{x}(k) + 2\eta e_i(k)\mathbf{v}(k) \quad (2 - 171)$$

$$b_i(k + 1) = b_i(k) + 2\eta e_i(k) \quad (2 - 172)$$

Where  $e_i(k)$  is the  $i^{\text{th}}$  element at the iteration  $k$

The LMS algorithm can be also written in a matrix form:

$$\mathbf{X}(k + 1) = \mathbf{X}(k) + 2\eta \mathbf{e}(k)\mathbf{v}_k^T \quad (2 - 173)$$

$$\mathbf{b}(k + 1) = \mathbf{b}(k) + 2\eta \mathbf{e}(k) \quad (2 - 174)$$

The LMS algorithm normally known as delta rule or an Approximate Gradient Descent Algorithm.

### (2-3-7) Backpropagation Algorithm [15, 53, 65]

The Backpropagation Algorithm is considered a generalization of the Least Mean Square algorithm, and it is used to train the multi-layer networks. This algorithm is sometimes called the steepest descent backpropagation algorithm because it is an approximation of the steepest descent algorithm. The Backpropagation Algorithm (BP or SDBP) is considered one of the most commonly used algorithms among all the supervised training algorithms in neural networks [26] [41]. The training of the neural networks by using the BP algorithm has three main stages:

#### (2-3-7-1) Forward Propagation Stage [15, 53]

In this stage, the inputs are applied to the network and the weights are randomly generated with small values in addition to identifying the learning rate within a period (0,1). The data is processed starting from the input layer then the hidden layer(s) and ending at the output layer. The inputs in the node (i) can be described as:

$$s_i^m = \sum_{j=1}^{n^{m-1}} x_{i,j}^m z_j^{m-1} + b_i^m \quad (2 - 175)$$

Where  $n^{m-1}$  represents the number of the nodes in the hidden layer (m-1).

The weights are adapted by using steepest descent algorithm to approximate the mean square error as follows:

$$x_{i,j}^m(k+1) = x_{i,j}^m(k) - \eta \frac{\partial}{\partial x_{i,j}^m} \hat{H} \quad (2-176)$$

$$b_i^m(k+1) = b_i^m(k) - \eta \frac{\partial}{\partial b_i^m} \hat{H} \quad (2-177)$$

Because the error is indirect function of the weights in the hidden layer, the chain rule is used to calculate the partial derivatives:

$$\frac{\partial}{\partial x_{i,j}^m} \hat{H} = \frac{\partial}{\partial s_i^m} \hat{H} \times \frac{\partial}{\partial x_{i,j}^m} s_i^m \quad (2-178)$$

$$\frac{\partial}{\partial b_i^m} \hat{H} = \frac{\partial}{\partial s_i^m} \hat{H} \times \frac{\partial}{\partial b_i^m} s_i^m \quad (2-179)$$

Based on the definition of the inputs in equation (1-175), we get

$$\frac{\partial}{\partial x_{i,j}^m} s_i^m = z_j^{m-1} \quad , \quad \frac{\partial}{\partial b_i^m} s_i^m = 1 \quad (2-180)$$

by set

$$\delta_i^m \equiv \frac{\partial}{\partial s_i^m} \hat{H} \quad (2-181)$$

Where  $\delta_i^m$  is the sensitivity of the  $\hat{H}$  to changes in the  $i^{\text{th}}$  element of the net input at layer  $m$ .

By using equations (2-180) and (2-181), the equations (2-178) and (2-179) can be written as follow:

$$\frac{\partial}{\partial x_{i,j}^m} \hat{H} = \delta_i^m z_j^{m-1} \quad (2-182)$$

$$\frac{\partial}{\partial b_i^m} \hat{H} = \delta_i^m \quad (2-183)$$

By substituting equation (2-182) and (2-183) in the equations (2-176) and (2-177) respectively, the final formula of to update the weights and biases can be written as follow:

$$x_{i,j}^m(k+1) = x_{i,j}^m(k) - \eta \delta_i^m z_j^{m-1} \quad (2-184)$$

$$b_i^m(k+1) = b_i^m(k) - \eta \delta_i^m \quad (2-185)$$

Equation (2 – 184) and (2 – 185) can be written in a matrix form as follows:

$$\mathbf{X}^m(k + 1) = \mathbf{X}^m(k) - \eta \boldsymbol{\delta}^m (\mathbf{z}^{m-1})^T \quad (2 - 186)$$

$$\mathbf{b}^m(k + 1) = \mathbf{b}^m(k) - \eta \boldsymbol{\delta}^m \quad (2 - 187)$$

where

$$\boldsymbol{\delta}^m \equiv \frac{\partial}{\partial \mathbf{s}^m} \hat{H} = \left[ \frac{\partial}{\partial s_1^m} \hat{H} \quad \frac{\partial}{\partial s_2^m} \hat{H} \quad \dots \quad \frac{\partial}{\partial s_{n^m}^m} \hat{H} \right]^T \quad (2 - 188)$$

### (2-3-7-2) Backward Propagation Stage [15, 65]

The sensitivity is calculated in this stage starting from the last layer and backward to the hidden layer(s) and ending at the input layer. This calculation is based on the errors of the output layer and the weights that are calculated in the forward propagation. This process is called the backpropagation since it describes the recurrent relationship where the sensitivity at the layer  $m$  is calculated based on the sensitivity at the layer  $m+1$ . In order to describe the recurrent process, the Jacobian matrix with size  $n^m \times n^{m+1}$  is defined as:

$$\frac{\partial \mathbf{s}^{m+1}}{\partial \mathbf{s}^m} = \frac{\partial s_i^{m+1}}{\partial s_j^m} \quad \text{where } i = 1, 2, \dots, n^{m+1} \quad , \quad j = 1, 2, \dots, n^m$$

$$\begin{aligned} \frac{\partial s_i^{m+1}}{\partial s_j^m} &= \frac{\partial}{\partial s_j^m} \left( \sum_{l=1}^{n^m} x_{i,l}^{m+1} z_l^m + b_i^{m+1} \right) \\ &= x_{i,j}^m \dot{g}^m(s_j^m) \end{aligned} \quad (2 - 189)$$

Where

$$\dot{g}^m(s_j^m) = \frac{\partial}{\partial s_j^m} g^m(s_j^m) \quad (2 - 190)$$

The Jacobian matrix can then be written as:

$$\frac{\partial}{\partial \mathbf{s}^m} \mathbf{s}^{m+1} = \mathbf{X}^{m+1} \dot{\mathbf{G}}^m(\mathbf{s}^m) \quad (2 - 191)$$

Where  $\dot{\mathbf{G}}^m(\mathbf{s}^m)$  is a diagonal matrix  $n^m \times n^m$  and the main diagonal elements are:

$$\dot{g}^m(s_j^m) \quad : \quad j = 1, 2, \dots, n^m$$



by using the chain rule, the recurrent relation can be written in a matrix form:

$$\boldsymbol{\delta}^m = \dot{\mathbf{G}}^m(\mathbf{s}^m)(\mathbf{X}^{m+1})^T \boldsymbol{\delta}^{m+1} \quad (2 - 192)$$

Where  $m=1,2,\dots,M-1$

It is clear that the sensitivity  $\boldsymbol{\delta}^m$  is calculated starting from the last layer and ending at the first layer:

$$\boldsymbol{\delta}^M \rightarrow \boldsymbol{\delta}^{M-1} \rightarrow \dots \rightarrow \boldsymbol{\delta}^1$$

The sensitivity at the last layer (output layer) is calculated as follows:

$$\begin{aligned} \delta_i^M &= \frac{\partial}{\partial s_i^M} \hat{H} \\ &= \frac{\partial}{\partial s_i^M} e^T e \\ &= \frac{\partial}{\partial s_i^M} \sum_{j=1}^{n^M} (d_j - z_j)^2 \\ &= -2(d_j - z_j) \dot{g}^M(s_i^M) \end{aligned} \quad (2 - 193)$$

Where

$$\dot{g}^M(s_i^M) = \frac{\partial}{\partial s_i^M} g^M(s_i^M) \quad (2 - 194)$$

Equation (2 - 193) can be written as a matrix form:

$$\boldsymbol{\delta}^M = -2\dot{\mathbf{G}}^M(\mathbf{s}^M)(\mathbf{d} - \mathbf{z}) \quad (2 - 195)$$

### (2-3-7-3) Weights Adaptation Stage [53, 65]

After the forward and backward propagation, the process of adapting weights and biases is started based on the results of the previous two stages:

$$\mathbf{X}^m(k+1) = \mathbf{X}^m(k) + \Delta\mathbf{X}^m(k) \quad (2 - 196)$$

$$\text{Where } \Delta\mathbf{X}^m(k) = -\eta \boldsymbol{\delta}^m (\mathbf{z}^{m-1})^T$$

$$\mathbf{b}^m(k+1) = \mathbf{b}^m(k) + \Delta\mathbf{b}^m(k) \quad (2 - 197)$$

$$\text{Where } \Delta\mathbf{b}^m(k) = -\eta \boldsymbol{\delta}^m$$

By using momentum filter, the weights and biases can be adapted as follows:

$$\mathbf{X}^m(k+1) = \mathbf{X}^m(k) + \gamma \Delta \mathbf{X}^m(k-1) - (1-\gamma)\eta \boldsymbol{\delta}^m(\mathbf{z}^{m-1})^T \quad (2-198)$$

$$\mathbf{b}^m(k+1) = \mathbf{b}^m(k) + \gamma \Delta \mathbf{b}^m(k-1) - (1-\gamma)\eta \boldsymbol{\delta}^m \quad (2-199)$$

Where  $0 < \gamma \leq 1$

### (2-3-8) Criteria of Model Selection [59]

The following criteria are used to compare between the models to choose the best model to represent the data.

1- Mean Square Error (MSE)

$$MSE = \frac{1}{N} \sum_{t=1}^N (W_t - \widehat{W}_t)^2 \quad (2-200)$$

Where

N: number of observations

$W_t$ : the real observations

$\widehat{W}_t$ : the estimated observations

2- Mean Absolute Error (MAE)

$$MAE = \frac{1}{N} \sum_{t=1}^N |W_t - \widehat{W}_t| \quad (2-201)$$

3- Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{1}{N} \sum_{t=1}^N \left| \frac{W_t - \widehat{W}_t}{W_t} \right| \quad (2-202)$$

# **Chapter Three**

**The Application Side:**

**- Artificial Neural**

**Networks**

**-Spectral Analysis**

### **3. The Application Side**

#### **(3-1) Data Set**

The exchange rate can be defined as a number of units of a certain currency that is exchanged with one unit of another currency. In addition, the exchange rate can also be defined as the value of two different currencies that is obtained from the relationship between them. For example, the Iraqi dinar is the national currency in Iraq, which is strongly related to the US dollar at a certain exchange rate as the federal budget of Iraq is based on US dollar. The exchange rate between the Iraqi Dinar and the US Dollar witnessed many fluctuations during the 1990s because of the war and the conflicts afterwards. The exchange rate has a high level of importance because the national economic is highly dependent on it.

The inside economic balance represents the nearly fixed prices with a slight economic growth, while the outside economic balance represents the ratio between the outcome or external payments and the general income. In other words, the economic balance represents the general commercial exchange with the outside world and the move of capital from and to Iraq.

Under the golden pricing system, the exchange rate between any two currencies is counted based on the golden value of each currency compared to the other currency. This system helps to limit the fluctuation in exchange rates to the minimum levels with very little tolerance based on the price of the gold which is semi-constant.

After the independence of the currency value from gold, the exchange rates witnessed high fluctuation rates, and they become highly dependent on the interest rate, economic growth and many other factors.

The exchange rate data of the Iraqi dinar and the US dollar has been gathered from the Central Bank of Iraq from 30/01/2004 to 30/12/2014.

During this study, we have used R and SPSS software to analyse the data.

**Table (3-1)**

**Exchange Rate of Iraqi Dinar vs US Dollar**

<b>Month</b>	<b>2004</b>	<b>2005</b>	<b>2006</b>	<b>2007</b>	<b>2008</b>	<b>2009</b>	<b>2010</b>	<b>2011</b>	<b>2012</b>	<b>2013</b>	<b>2014</b>
<b>1</b>	1467	1453	1483	1323	1224	1179	1185	1185	1206	1226	1222
<b>2</b>	1409	1459	1480	1299	1225	1178	1185	1185	1236	1231	1222
<b>3</b>	1423	1469	1480	1290	1222	1178	1185	1185	1240	1255	1222
<b>4</b>	1443	1474	1481	1284	1216	1179	1185	1187	1263	1267	1218
<b>5</b>	1462	1473	1485	1275	1212	1187	1185	1196	1250	1270	1222
<b>6</b>	1460	1468	1485	1269	1205	1180	1185	1197	1241	1237	1213
<b>7</b>	1463	1476	1486	1261	1202	1184	1185	1197	1253	1218	1214
<b>8</b>	1463	1480	1488	1253	1196	1184	1185	1199	1248	1209	1213
<b>9</b>	1463	1481	1488	1249	1188	1183	1185	1200	1228	1211	1204
<b>10</b>	1463	1475	1486	1245	1185	1183	1185	1200	1200	1220	1207
<b>11</b>	1463	1477	1467	1240	1183	1183	1188	1200	1207	1218	1200
<b>12</b>	1462	1479	1394	1216	1180	1185	1195	1218	1222	1222	1205

### **(3-2) Artificial Neural Network Methodology**

The success of designing artificial neural network depends on the understanding of the problem that needs to be solved, and identifying the input variables that are required to build the network for prediction purposes.

It is commonly known that the data collection process depends on the problem that needs to be solved, while in neural network design, the data collection depends on the training algorithm that will be used to train the neural network. If the training algorithm is one of the supervised algorithms then there is a need to have both the input and output data. However, if the training algorithm is unsupervised algorithm then no need to have the output data where the input data will be enough.

#### **(3-2-1) Identifying the Input Nodes**

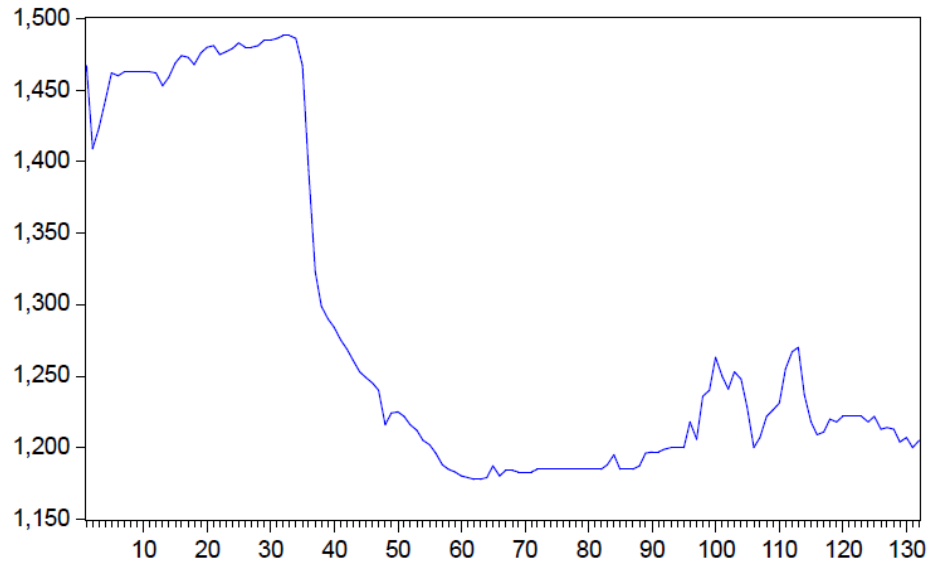
The most important aspect of designing neural networks is to identify the nodes that are required in each level. The number of input nodes is the same as the number of the input variables. In the case of causal problems data such as the regression problems, it is easy to identify the number of the input nodes where it will be the same as the number of the input variables. However, in time series data there is no rule to identify the input variables. For this reason, the Box – Jenkins methodology will be applied to choose the input variables.

- **Applying Box – Jenkins Methodology**

- 1- Model Identification**

In this stage, the time series  $Y_t$  is plotted for the data of the Iraqi Dinar exchange rate according to the US dollar as shown in figure (3-1)

Y



**Figure (3-1)**  
**Exchange Rate of Iraqi Dinar vs US Dollar**

By examining figure (3-1), it is clearly shown that the time series is not fluctuating or oscillating around a certain level, and it has a general decreasing trend that proves the time series  $Y_t$  is not mean stationary and not variance stationary.

It is also clearly shown in figure (3-2) below that the SACF is slowly exponentially decreasing toward zero and is not cut off after the second lag or the third lag which also means that the time series is not mean stationary.

Correlogram of Y

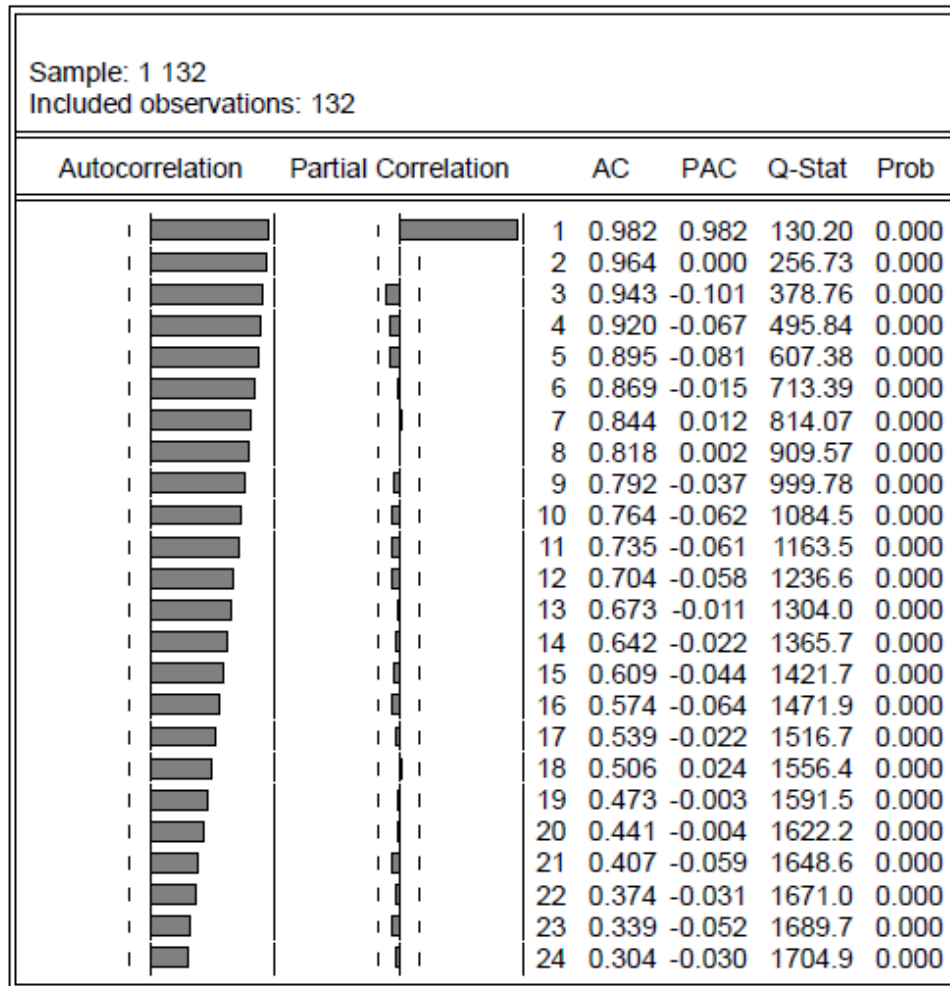


Figure (3-2)

Correlogram of the Original Time Series  $Y_t$

The non-stationarity of the time series  $Y_t$  has also been proved by applying the ADF, P.P, KPSS tests as shown in table (3-2) below



**Table (3-2)**  
**Tests Results of the Original Time Series  $Y_t$**

<b>Test</b>	<b>Model</b>	<b>Test Statistic</b>	<b>Critical Values</b>	<b>p-Value</b>
ADF	without constant	-0.837279	-1.943304	0.3514
	With constant	-0.979861	-2.883756	0.7591
	With constant and time trend	-1.307575	-3.444756	0.8817
P.P	without constant	-1.397307	-1.943304	0.1504
	With constant	-1.46627	-2.883756	0.5477
	With constant and time trend	-1.200101	-3.444756	0.9059
KPSS	With constant	0.937165	0.643000	
	With constant and time trend	0.277682	0.146000	

By checking the P-value in table (3-2) for the three models of the tests ADF and P.P, the null hypothesis is accepted which means that the time series  $Y_t$  has a unit root and this means that the time series is non-stationary.

By comparing the value of the test statistic with the critical value in table (3-2) for the two estimated models in the KPSS test, then the alternative hypothesis is accepted which means that the time series  $Y_t$  is non-stationary.

In order to achieve the stationary in variance, the Box-Cox transformation is applied as given in the equation (2-41) where  $\lambda=0$ . This means that the logarithmic transform of the time series  $Y_t$  has been calculated in order to be variance stationary and then the first difference is taken in order to be mean stationary.

To test the stationary of transformed data, we checked the correlogram of the transformed series  $W_t = (1 - L)(\text{Log}(Y_t))$  as shown in figure (3-3), where we notice that SACF is cut off after the first lag which means that the time series  $W_t$  is mean stationary.

Correlogram of W

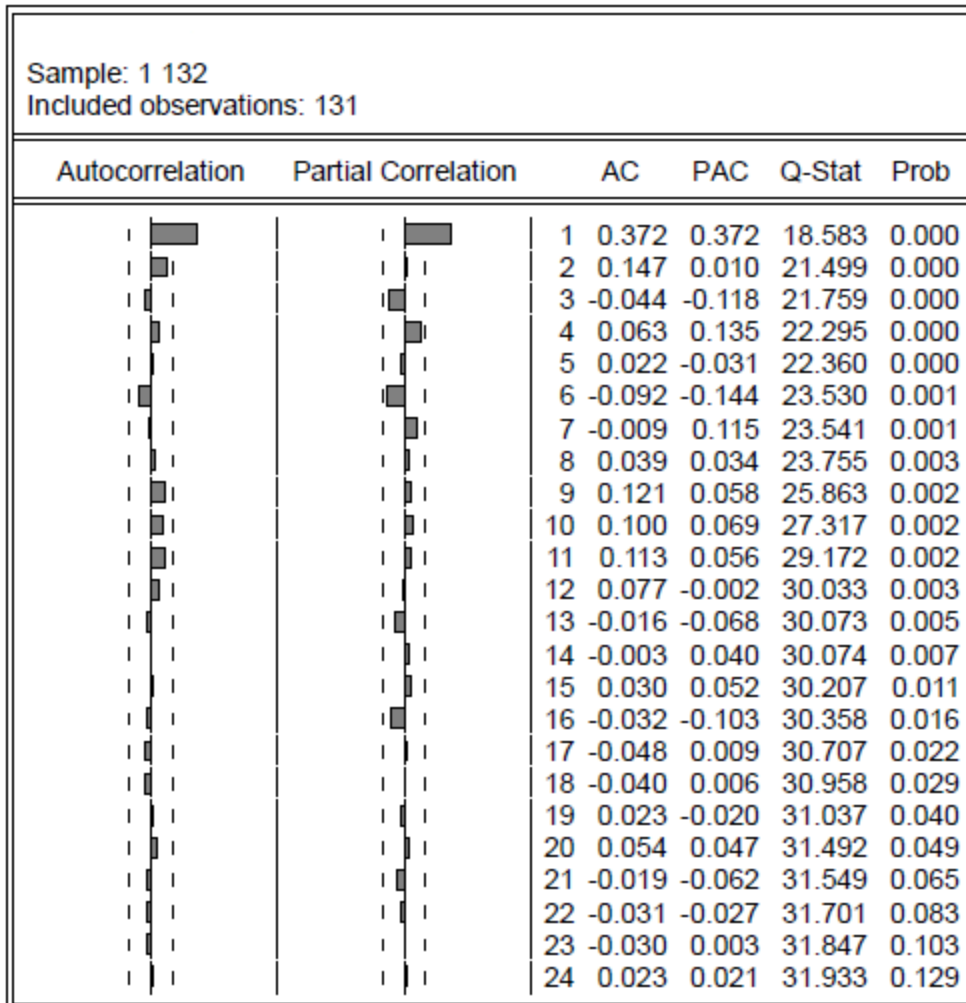


Figure (3-3)

Correlogram of the Transformed Time Series  $W_t$

The ADF, P.P, and KPSS tests have been applied to the series  $W_t$  and the results are shown in table (3-3):

**Table (3-3)**

**Tests Results of the Transformed Time Series  $W_t$**

<b>Test</b>	<b>Model</b>	<b>Test Statistic</b>	<b>Critical Values</b>	<b>p-Value</b>
ADF	without constant	-8.18208	-1.943304	0.0000
	With constant	-8.197312	-2.883756	0.0000
	With constant and time trend	-8.149873	-3.444756	0.0000
P.P	without constant	-8.344295	-1.943304	0.0000
	With constant	-8.382012	-2.883756	0.0000
	With constant and time trend	-8.343332	-3.444756	0.0000
KPSS	With constant	0.214289	0.643000	
	With constant and time trend	0.098436	0.146000	

By checking the P-value in table (3-3) for the three models of the tests ADF and P.P, the alternative hypothesis is accepted which means that the time series  $W_t$  has no unit root (there is no need to take further differences) which means that it is stationary.

By comparing the value of the test statistic with the critical value in table (3-3) for the two estimated models in the KPSS test, then the null hypothesis is accepted which means that the time series  $W_t$  is stationary.

After checking the stationarity of the time series  $W_t$ , the figure (3-3) has been checked again in order to identify the model. Based on the behavior of the SACF and the SPACF, the ARIMA (1,1,1) model has been recognized and identified as a primary model to represent the data.

## 2- Parameters Estimation

In this stage, the suggested primary model has been estimated by the Exact Maximum Likelihood Estimation method. Some models have been suggested which are very close to the primary model as: ARIMA(1,1,0), ARIMA(2,1,0), ARIMA(2,1,2), ARIMA(2,1,1), ARIMA(1,1,2) and ARIMA(0,1,1) model as shown in table (3-4) .

To select the best model to represent the data, we calculated all the criteria including AIC, H-Q, BIC, MSE, MAE and MAPE of each estimated model as shown in table (3 – 5).

**Table (3-4)**  
**Estimated Parameters of ARIMA(p,d,q) Models**

ARIMA	Parameters		P-Value
(2,1,2)	$\hat{\alpha}_1$	-0.180371	0.1145
	$\hat{\alpha}_2$	-0.501118	4.14e-06
	$\hat{\beta}_1$	0.635411	1.28e-015
	$\hat{\beta}_2$	0.834250	7.00e-034
(1,1,2)	$\hat{\alpha}_1$	-0.332391	0.0773
	$\hat{\beta}_1$	0.836455	4.24e-07
	$\hat{\beta}_2$	0.515426	4.44e-09
(1,1,0)	$\hat{\alpha}_1$	0.431265	6.57e-08
(1,1,1)	$\hat{\alpha}_1$	0.404445	0.0293
	$\hat{\beta}_1$	0.0338880	0.8672
(2,1,1)	$\hat{\alpha}_1$	0.866848	0.5128
	$\hat{\alpha}_2$	-0.227888	0.6809
	$\hat{\beta}_1$	-0.412057	0.7584
(0,1,1)	$\hat{\beta}_1$	0.0347870	3.32e-05
(2,1,0)	$\hat{\alpha}_1$	0.443303	5.82e-013
	$\hat{\alpha}_2$	-0.0253968	0.7719

**Table (3-5)**  
**Estimated Criteria of ARIMA(p,d,q) Models**

<b>ARIMA</b>	<b>AIC</b>	<b>H-Q</b>	<b>BIC</b>	<b>MSE</b>	<b>MAE</b>	<b>MAPE</b>
(2,1,2)	-856.127	-850.285	-841.751	8.2251E-05	0.005552	0.077613
(1,1,2)	-855.403	-850.733	-843.905	8.4133E-05	0.00549	0.076745
(1,1,0)	-850.17	-847.833	-844.419	8.8425E-05	0.005308	0.07414
(1,1,1)	-848.21	-844.705	-839.585	8.8409E-05	0.005307	0.074141
(2,1,1)	-846.496	-841.823	-834.996	8.8286E-05	0.005339	0.074591
(0,1,1)	-845.428	-843.092	-839.678	9.0732E-05	0.0054472	0.076101
(2,1,0)	-848.239	-844.734	-839.614	8.8399E-05	0.005309	0.074167

by checking the P-value of the estimated parameters for every model in table (3-4), the following models have not been considered:

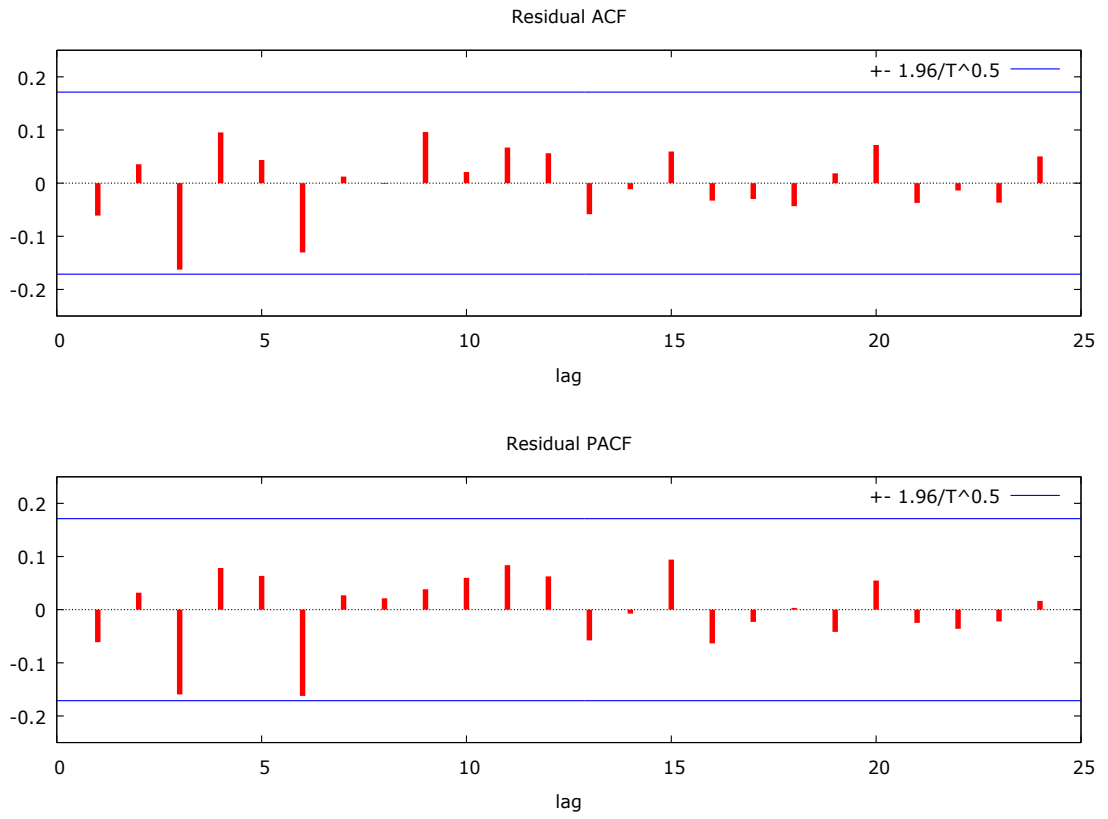
ARIMA(2,1,0), ARIMA(2,1,2), ARIMA(2,1,1), ARIMA(1,1,2), ARIMA(1,1,1).

The reason for not considering the above models is because they have some non-significant parameters.

Based on the value of the criteria in table (3-5) for ARIMA (1,1,0) and ARIMA (0,1,1) model, the ARIMA (1,1,0) model has been chosen as the best model to represent the time series. It is also noted that the estimated results in table (3-4) shows that the ARIMA (1,1,0) model satisfies the stationary condition  $|\alpha_1| = |0.431265| < 1$ .

### **3- Diagnostic Checking**

In this stage, both SACF and SPACF have been calculated and plotted for the residuals series of the ARIMA(1,1,0) model as shown in figure (3-4) and table (3-6) below:



**Figure (3-4)**  
**Correlogram of Residuals**

**Table (3 – 6)**

**Autocorrelation Coefficients of the Residuals**

Residual autocorrelation function  
\*\*\*, \*\*, \* indicate significance at the 1%, 5%, 10% levels  
using standard error  $1/T^{0.5}$

LAG	ACF	PACF	Q-stat.	[p-value]
1	-0.0611	-0.0611		
2	0.0356	0.0320	0.6711	[0.413]
3	-0.1628 *	-0.1595 *	4.2781	[0.118]
4	0.0954	0.0785	5.5278	[0.137]
5	0.0436	0.0636	5.7907	[0.215]
6	-0.1304	-0.1623 *	8.1607	[0.148]
7	0.0123	0.0269	8.1820	[0.225]
8	-0.0017	0.0213	8.1824	[0.317]
9	0.0962	0.0384	9.5039	[0.302]
10	0.0209	0.0599	9.5669	[0.387]
11	0.0669	0.0837	10.2172	[0.422]
12	0.0562	0.0626	10.6791	[0.471]
13	-0.0584	-0.0579	11.1820	[0.513]
14	-0.0114	-0.0075	11.2015	[0.594]
15	0.0595	0.0941	11.7331	[0.628]
16	-0.0329	-0.0635	11.8971	[0.687]
17	-0.0298	-0.0230	12.0331	[0.742]
18	-0.0434	0.0033	12.3232	[0.780]
19	0.0185	-0.0419	12.3764	[0.827]
20	0.0718	0.0548	13.1848	[0.829]
21	-0.0372	-0.0248	13.4037	[0.859]
22	-0.0138	-0.0360	13.4342	[0.893]
23	-0.0367	-0.0221	13.6509	[0.913]
24	0.0503	0.0165	14.0623	[0.925]

It is shown in figure (3-4) that all the autocorrelation coefficients of the residuals are not significantly different from zero and they lie within the period  $(\frac{-1.96}{\sqrt{N}}, \frac{+1.96}{\sqrt{N}})$  at significance level of  $\alpha = 0.05$ .

This means that the residuals represent white noise.

The Ljung – Box test has been applied in this stage as well and the results are shown in table (3-7) below:

**Table (3-7)**

**Results of Ljung-Box Test**

<b>Ljung-Box Statistic (<math>Q_{LB}</math>)</b>	<b><math>\chi^2(df = 23, \alpha = 0.05)</math></b>	<b>P-value</b>
14.0623	35.172	0.9251

By checking the P-value of the test statistic, the null hypothesis is accepted which means that the residuals represent the white noise.

Based on the Box – Jenkins methodology to represent the time series of the Iraqi Dinar exchange rate against the US Dollar, it can be concluded that the neural network input variable is the  $Y_{t-1}$  variable which means that the input level of the neural network has only one node.

**(3-2-2) Identifying the Output Nodes**

The output node is set to be only one because the goal is to predict one-step-a head. The backpropagation algorithm has been chosen to train the neural network so that the input variable is set to be  $Y_{t-1}$  and the output variable is set to be  $Y_t$ .

**(3-2-3) Data Processing**

It is rare to feed the neural network with both input and output data directly, and normally the data is being processed and scaled between lower bound and upper bound. [50] showed that the data processing is essential to simplify the learning of the neural networks and meet the requirements of the training algorithm. According to [63], showed that the data preprocessing helps avoid the calculations problems.

The commonly used formula to preprocess data is:

$$f(v) = (c - b) \frac{v - \min(V)}{\max(V) - \min(V)} + b \quad [63]$$

Where:

c , b : constants

V: data vector



Usually,  $b = 0$  and  $c = 1$  which results in normalized formula that scale the data to be between 0 and 1. Another adjusted normalized formula can be done by setting  $b = -1$  and  $c = 1$  which scale the data to be between -1 and +1.

There are other formulas such as the standardized formula that scale the data between -1 and +1, and the simple formula that scales the data between 0 and 1.

### **(3-2-4) Activation Functions Selection**

It is important to choose the activation functions before applying any data to the network in order to choose the right preprocessing formula for the data before applying it to the network.

There is no standard rule to choose the activation functions in the hidden layer or the output layer of the network. Both [49,62] and [61,62] chose a specific activation function for each node in the layer. However, the majority of the networks have one type of activation function for all the nodes in a certain layer. Most papers reported the use of the logistic function in the hidden layer, while in the output layer the linear function was used when the training algorithm is conducted for classification.

The non-linear activation functions are normally used when the training algorithm is used for prediction. [21] showed that the non-linear activation function is more efficient for financial data. Because there is no rule to choose the activation function, five models have been designed in this thesis:

First Model: hyperbolic activation function for the hidden layer nodes and sigmoid activation function for the output layer node.

Second Model: sigmoid activation function for the hidden layer nodes and sigmoid activation function for the output layer node.

Third Model: hyperbolic activation function for the hidden layer nodes and Identity activation function for the output layer node.

Forth Model: sigmoid activation function for the hidden layer nodes and Identity activation function for the output layer node.

Fifth Model: hyperbolic activation function for the hidden layer nodes and hyperbolic activation function for the output layer node.

In tables (3-8) to (3-12), the data processing formula is set based on the activation function in the output layer.

The models are shown in tables (3-8) to (3-12):

**Table (3-8)**  
**First Model: ANN(1)**

<b>Activation Function for Hidden Layer</b>		<b>Hyperbolic tangent</b>					
<b>Activation Function for Output Layer</b>		<b>Sigmoid</b>					
<b>Data Preprocessing Formula</b>		<b>Normalized</b>					
<b>Number of Hidden Nodes</b>	<b>MSE</b>	<b>MAE</b>	<b>MAPE</b>	<b>Number of Hidden Nodes</b>	<b>MSE</b>	<b>MAE</b>	<b>MAPE</b>
1	355.3003	13.8103	1.0669	6	244.4384	10.8591	0.8442
2	239.2133	10.7915	0.8421	7	240.6842	10.4926	0.8174
3	314.3509	12.3766	0.96	8	320.4959	12.2783	0.9625
4	363.1231	14.0481	1.0874	9	268.5468	11.936	0.9279
5	299.6979	12.6116	0.9751	10	276.1918	11.8453	0.9259

**Table (3-9)**  
**Second Model: ANN(2)**

<b>Activation Function for Hidden Layer</b>		<b>Sigmoid</b>					
<b>Activation Function for Output Layer</b>		<b>Sigmoid</b>					
<b>Data Preprocessing Formula</b>		<b>Normalized</b>					
<b>Number of Hidden Nodes</b>	<b>MSE</b>	<b>MAE</b>	<b>MAPE</b>	<b>Number of Hidden Nodes</b>	<b>MSE</b>	<b>MAE</b>	<b>MAPE</b>
1	308.7995	12.6732	0.98	6	400.6688	13.8814	1.0779
2	399.0621	14.5312	1.0978	7	366.3151	14.5746	1.1027
3	337.6484	12.7363	1.0029	8	272.8888	9.5703	0.7261
4	246.7104	10.5956	0.8088	9	297.0384	12.5679	0.9652
5	256.438	11.5061	0.8838	10	272.4871	11.658	0.9101

**Table (3-10)**  
**Third Model: ANN(3)**

<b>Activation Function for Hidden Layer</b>		<b>Hyperbolic tangent</b>					
<b>Activation Function for Output Layer</b>		<b>Identity</b>					
<b>Data Preprocessing Formula</b>		<b>Normalized</b>					
<b>Number of Hidden Nodes</b>	<b>MSE</b>	<b>MAE</b>	<b>MAPE</b>	<b>Number of Hidden Nodes</b>	<b>MSE</b>	<b>MAE</b>	<b>MAPE</b>
1	177.339	7.6599	0.5865	6	184.5194	8.0842	0.6196
2	173.3501	7.6115	0.5843	7	196.4128	8.754	0.6751
3	198.175	8.6334	0.6629	8	179.8835	7.9118	0.6056
4	183.9683	8.0917	0.6195	9	182.1894	7.9497	0.6101
5	183.2493	7.9365	0.6103	10	184.5148	8.1362	0.6227

**Table (3-11)**  
**Forth Model: ANN(4)**

<b>Activation Function for Hidden Layer</b>		<b>Sigmoid</b>					
<b>Activation Function for Output Layer</b>		<b>Identity</b>					
<b>Data Preprocessing Formula</b>		<b>Normalized</b>					
<b>Number of Hidden Nodes</b>	<b>MSE</b>	<b>MAE</b>	<b>MAPE</b>	<b>Number of Hidden Nodes</b>	<b>MSE</b>	<b>MAE</b>	<b>MAPE</b>
1	211.7407	9.3136	0.7118	6	182.8886	8.0308	0.6238
2	222.3407	9.6718	0.7421	7	198.721	8.6888	0.6715
3	185.2495	8.2581	0.6293	8	181.8466	7.9717	0.6123
4	195.651	8.7934	0.6728	9	220.4262	9.1451	0.7009
5	191.5193	8.686	0.6657	10	185.0319	8.1435	0.6259

**Table (3-12)**  
**Fifth Model: ANN(5)**

<b>Activation function for hidden layer</b>		<b>Hyperbolic tangent</b>					
<b>Activation function for output layer</b>		<b>Hyperbolic tangent</b>					
<b>Data preprocessing formula</b>		<b>Adjusted normalized</b>					
<b>Number of Hidden Nodes</b>	<b>MSE</b>	<b>MAE</b>	<b>MAPE</b>	<b>Number of Hidden Nodes</b>	<b>MSE</b>	<b>MAE</b>	<b>MAPE</b>
1	298.7304	12.7281	0.9971	6	307.0816	12.5883	0.992
2	183.2822	8.6717	0.6709	7	284.4868	11.752	0.9224
3	175.6466	8.3228	0.6452	8	267.8135	11.5292	0.899
4	274.034	11.7873	0.9162	9	281.4044	11.748	0.923
5	283.5786	11.9748	0.9335	10	304.3154	12.8228	1.0107

**Table (3-13)**

**Minimum Values of the Criteria**

<b>Model</b>	<b>MIN MSE</b>	<b>Number of Hidden Nodes</b>	<b>MIN MAE</b>	<b>Number of Hidden Nodes</b>	<b>MIN MAPE</b>	<b>Number of Hidden Nodes</b>
ANN(1)	239.2133	2	10.4926	7	0.8174	7
ANN(2)	246.7104	4	9.5703	8	0.7261	8
ANN(3)	173.3501	2	7.6115	2	0.5843	2
ANN(4)	181.8466	8	7.9717	8	0.6123	8
ANN(5)	175.6466	3	8.3228	3	0.6452	3

**(3-2-5) Identifying the Number of the Hidden Layers**

In theory, a neural network with one hidden layer and enough number of nodes in this layer, can estimate any continuous function [21]. However, in practice, the neural network is normally trained by using one hidden layer, and if the target is not achieved then it will be trained by using two hidden layers. Many studies showed that any neural network with more than four hidden layers will not get better results so that the models is table (3-8) to table (3-12) will be designed with one hidden layer.

### (3-2-6) Identifying the Number of the Hidden Nodes

There is no specific rule to choose the number of the nodes in the hidden layer. However, the common approach to choose the number of the nodes in the hidden layer is the error and trial approach where the training process starts with small number of nodes in the hidden layer and then the number is increased as the mean square error decreases. In addition, there are some formulas that were suggested by researchers to identify the number of the nodes in the hidden layer according to the number of the input nodes (R):

- 1- Number of hidden nodes = R [62]
- 2- Number of hidden nodes = 2R [62]
- 3- Number of hidden nodes = 2R +1 [65]
- 4- Number of hidden nodes = R/2 [62]
- 5- Number of hidden nodes =  $\frac{(B_t - 1) \times e}{B_i + 2}$  [27]

Where

$B_t$  : number of input nodes

$B_i$  : size of the training data

$e \leq 0.125$

- 6- Number of hidden nodes =  $\beta \sqrt{B_i B_o}$  [43]

Where

$B_o$  : the number of the output nodes

$\beta$  : multiplication factor

In this work, the formulas 1, 2 and 3 that are mentioned above, have been used in the models in tables (3-8) to table (3-12). In addition, the error and trial method has been used to choose between 4 to 10 nodes for every model.

### (3-2-7) Data Partitioning

Before applying it to the neural network, the data is divided into two groups called: the training data and the testing data. The training data is normally bigger than the testing data and it is used by the neural network to learn the patterns. The size of the testing data is normally 10% to 30% of the overall size of the data. The testing data is used to check the ability of the neural network to learn the patterns of the data [62].

The models in tables (3-8) to (3-12) has been designed with a training data set of sizes 100% of the raw data which means that the ability of the neural network will be examine based on the training data set only.

### **(3-2-8) Training**

After identifying the requirements of the neural network, the training process is then started. The main goal of the training process is to find the optimal weights that produce the lowest error which means the lowest difference between the output of the neural network and the desired output. The training process requires the following parameters to be set: the initial weights, the learning rate and the momentum. Setting the initial weights values is the first thing to do in the training process and normally they are randomly set to small values. [15] showed that the initial weights based on the type of the neural network. For multi-layer neural networks, the initial weights are chosen to be between -0.5 and +0.5 if the training data set are within the period [-1, +1] [15].

The initial weights have been randomly set for every model in the tables (3-8) to (3-12) by using SPSS software.

The learning rate  $\eta$  and the momentum  $\gamma$  are normally set in a random way where  $0 < \eta < 1$  and  $0 < \gamma \leq 1$ . Initially, the learning rate is set to 0.5 and the momentum is set to 0.9 for all the models in tables (3-8) to (3-12).

When the training process starts to find the optimal weights, there should be a termination criterion to stop the process and below are some of them:

1- Time

2- Relative Cyclic Error

The relative cyclic error is the error allowed in every iteration, and when the calculated error is equal to or less than the cyclic error then the training process will stop.

3- Number of Iterations

There should be a realistic number of iteration which is normally set to be 10000 iterations. However, some researcher reported to 3000 iterations[55] and 5000 iterations[23].

All the three above-mentioned criteria have be used to stop the training in all of the models that are mentioned in tables (3-8) to (3-12) where the time is set to 15 minutes, the cyclic error is set to 0.0001 and the maximum number

of iterations is set to 10,000 iteration. When one of the criteria conditions is met, the training process will be stopped.

### **(3-2-9) Implementation**

All the models that are mentioned in tables (3-8) to (3-12) have been implemented and then the MSE, MAE, and MAPE have been calculated for each model.

Comparing the ability of the models in tables (3-8) to (3-12) to extract the minimum value of the MSE, MAE, and MAPE to the five models in table (3-13), it can be seen that the minimum value of the MSE, MAE and MAPE is in the third model in table (3-10) regardless of the number of the hidden nodes.

Further checking the table (3-10), it can be seen that the best activation function in the hidden layer is the hyperbolic tangent activation function, and of the output layer is the linear activation function. Thus, the suitable data preprocessing approach is the normalized formula. Regarding the number of the nodes in the hidden layer, it can be seen from table (3-10) that the minimum values of MSE, MAE, and MAPE appear when the number of the nodes in the hidden layer are 2. This means that the second formula of choosing the number of the nodes in the hidden layer is successful.

Based on the third model in table (3-10) with 2 nodes in the hidden layer, the value of the momentum was changed while the learning rate and the other model requirements are fixed in order to set the right value of the momentum that is suitable for the learning rate of 0.5 as shown in table (3-14)

**Table (3-14)**

**Values of the Criteria During Momentum Change and Fixed Learning Rate**

<b>Learning Rate</b>	<b>Momentum</b>	<b>Time</b>	<b>MSE</b>
0.5	0.9	0:00:00.09	173.3501
0.5	0.8	0:00:00.08	184.3754
0.5	0.7	0:00:00.05	171.2072
0.5	0.6	0:00:00.05	170.5924
0.5	0.5	0:00:00.17	188.2508
0.5	0.4	0:00:00.37	180.5767
0.5	0.3	0:00:00.08	173.5609
0.5	0.2	0:00:00.03	171.6374
0.5	0.1	0:00:00.09	197.1893

From table (3-14), it is clearly shown that the best momentum value is 0.6 based on the MSE criterion, and the value of the momentum effects on the training time where this time reaches its maximum value when the momentum was 0.2, while in momentum 0.6 the least MSE can be achieved in a suitable time.

Based on the results that are obtained previously, the data has been divided into two sets: the training set and the testing set as mentioned in table (3-15)



**Table (3-15)**

**Values of the Criteria When Data Partitioning**

<b>Training Data</b>	<b>Testing Data</b>	<b>MSE</b>
100	0	173.3501
90	10	184.357
80	20	183.1978
70	30	180.9971
60	40	188.8886
50	50	179.1047
40	60	198.5476
30	70	189.0676
20	80	204.1265
10	90	185.3909

From table (3-15), it can be seen that the lowest error can be achieved based on the MSE when the training set is 100% of the overall data while the maximum MSE is achieved when the training set is 20% of the overall data.

Based on the results of the final model with hyperbolic tangent activation function for the hidden layer and linear activation function for the output layer in addition to a momentum of 0.6, the final model has been used in two cases: the first one where the data was not processed, and the second one when the data was processed using the normalized preprocessing formula, and the results are shown in table (3-16).

**Table (3-16)**

**Values of Criteria for Final Model with and without Processing**

<b>Final Model</b>	<b>MSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Without Processing</b>	13345.0652	99.8602	7.5583
<b>With Processing by Normalized Formula</b>	170.5924	7.5583	.5800

By checking the above table and based on the values of the MSE, MAE and MAPE, it can be found that criteria is highly increased when the training data is not processed which means that the data should be processed before being applied to the network.

Applying the last model, the optimal weights and biases are shown in table (3-17) that gives the lowest MSE = 170.5924, MAE = 7.5583 and MAPE = 0.5800.

**Table (3-17)**

**Final Values of Weights and Biases**

<b>Hidden Layer Weights and Biases</b>		<b>Output layer Weights and Biases</b>	
$x_{1,1}^1$	.254	$x_{1,1}^2$	207
$x_{1,2}^1$	.974	$x_{1,2}^2$	1.102
$b_1^1$	-.441	$b^2$	.275
$b_2^1$	-.187		

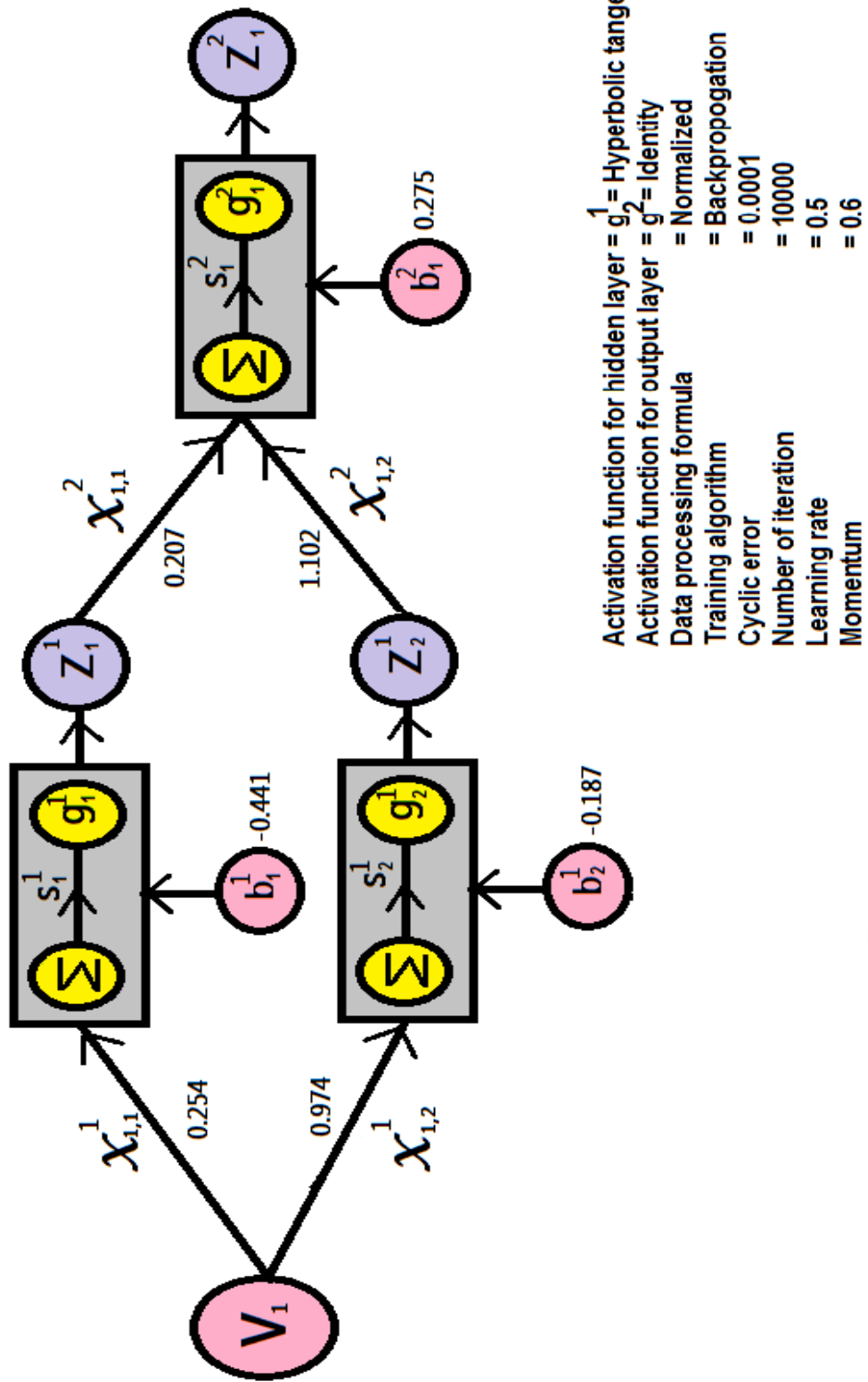


Figure (3-5)

The Final Artificial Neural Network Model

### **(3-3) Spectral Analysis**

The Spectral Analysis is approach based on analyzing the stationary data in the frequency domain, so that the most important step is to generate the frequencies of the stationary time series. Based on the stationary time series  $W_t$  that resulted from Box–Jenkins methodology (3-2-1), the Fourier frequencies have been generated by using the formula (2-58) as shown in table (3-18) where 65 standard frequencies have been generated because the number of observations of the  $W_t$  is odd.

#### **(3-3-1) Search for Hidden Periodicities**

The use of the spectral analysis to build harmonic model to represent the data, mainly depends on the existence of the periodicities (harmonic components) hidden in the stationary data. So that the detection of the hidden periodicities in the data is a crucial step that govern whether to use the spectral analysis or not.

In order to detect the hidden periodicities in the  $W_t$  time series, the periodogram has been calculated by using the formula (2-70) as shown in table (3-18) and figure (3-5) where both  $\Sigma_j$  and  $\Omega_j$  have been estimated by using formula (2-60) and (2-61) respectively.

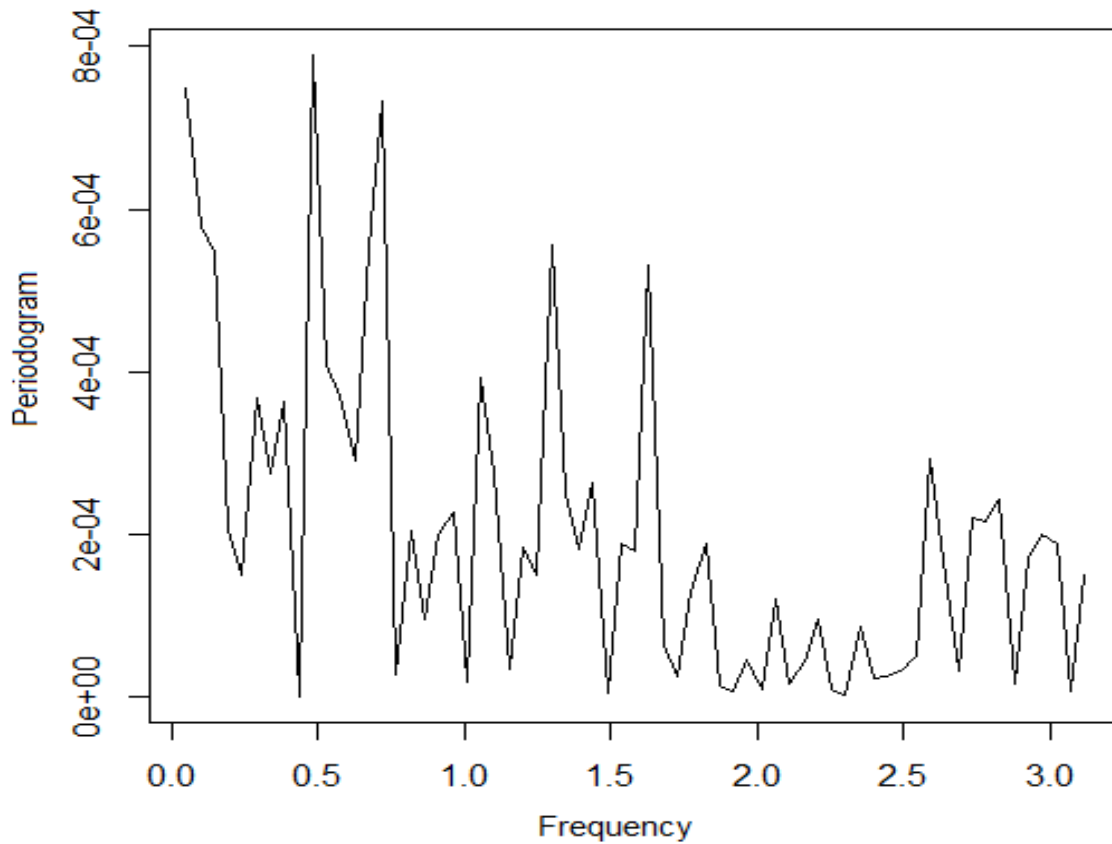
In addition to that, the wave length of the  $W_t$  time series has been calculated by using the formula (2-52) as shown in table (3-18) below:

**Table (3-18)**

**Periodogram Analysis of  $W_t$**

<b>j</b>	<b>Frequency</b>	<b>Period</b>	$\xi_j$	$\Omega_j$	<b>Amplitude</b>	<b>Phase</b>	<b>Periodogram</b>
1	0.04796325	20.8492975	7.079433e-04	-3.307630e-03	0.0033825436	1.35994424	7.494249e-04
2	0.09592649	10.4246488	1.431931e-03	2.602334e-03	0.0029702810	-1.06776213	5.778783e-04
3	0.14388974	6.9497658	-1.660042e-03	2.361398e-03	0.0028865100	0.95806557	5.457421e-04
4	0.19185299	5.2123244	-1.057219e-03	-1.432414e-03	0.0017803154	-0.93497568	2.076037e-04
5	0.23981623	4.1698595	6.907379e-04	-1.340426e-03	0.0015079326	1.09497412	1.489379e-04
6	0.28777948	3.4748829	8.760026e-04	2.201576e-03	0.0023694553	-1.19210338	3.677379e-04
7	0.33574273	2.9784711	-1.694825e-03	1.152857e-03	0.0020497585	0.59732830	2.751989e-04
8	0.38370597	2.6061622	-7.703295e-04	-2.223622e-03	0.0023532748	-1.23730538	3.627326e-04
9	0.43166922	2.3165886	1.112947e-04	6.335595e-05	0.0001280644	-0.51751207	1.074232e-06
10	0.47963247	2.0849298	-7.756589e-04	3.382998e-03	0.0034707814	1.34541046	7.890342e-04
11	0.52759571	1.8953907	-1.071563e-03	2.255256e-03	0.0024968837	1.12723366	4.083551e-04
12	0.57555896	1.7374415	1.226869e-03	-2.036024e-03	0.0023770990	1.02848120	3.701143e-04
13	0.62352221	1.6037921	1.523896e-03	-1.452157e-03	0.0021049983	0.76129763	2.902317e-04
14	0.67148545	1.4892355	-2.849056e-03	5.581426e-04	0.0029032121	0.19345444	5.520760e-04
15	0.71944870	1.3899532	-3.324620e-03	-3.815690e-04	0.0033464453	-0.11427072	7.335146e-04
16	0.76741195	1.3030811	-6.058610e-05	-6.516940e-04	0.0006545042	-1.47809570	2.805862e-05
17	0.81537519	1.2264293	3.992304e-04	1.720010e-03	0.0017657352	-1.34272555	2.042173e-04
18	0.86333844	1.1582943	-5.347976e-04	1.077578e-03	0.0012029894	1.11011643	9.479051e-05
19	0.91130169	1.0973314	-3.786098e-04	-1.707657e-03	0.0017491249	-1.35261267	2.003932e-04
20	0.95926493	1.0424649	3.652009e-04	-1.827602e-03	0.0018637327	1.37356885	2.275142e-04
21	1.00722818	0.9928237	-2.190613e-04	5.006453e-04	0.0005464737	1.15833728	1.956050e-05
22	1.05519143	0.9476953	-2.440567e-03	-1.415041e-04	0.0024446655	-0.05791519	3.914535e-04
23	1.10315467	0.9064912	-1.399170e-03	-1.496164e-03	0.0020484589	-0.81888554	2.748501e-04
24	1.15111792	0.8687207	6.120694e-04	-3.864795e-04	0.0007238752	0.56321041	3.432169e-05
25	1.19908117	0.8339719	-1.507334e-03	7.399063e-04	0.0016791421	0.45631752	1.846784e-04
26	1.24704441	0.8018961	-1.252173e-03	-8.397729e-04	0.0015076987	-0.59075700	1.488917e-04
27	1.29500766	0.7721962	2.248463e-03	-1.856158e-03	0.0029156316	0.69011101	5.568095e-04
28	1.34297091	0.7446178	1.040600e-03	-1.651796e-03	0.0019522497	1.00862326	2.496388e-04
29	1.39093415	0.7189413	-1.153633e-03	-1.196109e-03	0.0016617904	-0.80347318	1.808813e-04
30	1.43889740	0.6949766	-2.000057e-03	-1.215087e-04	0.0020037445	-0.06067807	2.629820e-04
31	1.48686065	0.6725580	-1.684346e-04	2.189918e-04	0.0002762746	0.91516007	4.999460e-06
32	1.53482389	0.6515405	1.661626e-03	3.536203e-04	0.0016988374	-0.20968754	1.890362e-04
33	1.58278714	0.6317969	1.703586e-05	-1.653322e-03	0.0016534096	1.56049267	1.790615e-04
34	1.63075039	0.6132146	-1.036664e-03	-2.648826e-03	0.0028444593	-1.19775397	5.299571e-04
35	1.67871363	0.5956942	8.291315e-04	-5.393994e-04	0.0009891465	0.57676847	6.408590e-05
36	1.72667688	0.5791472	2.320018e-04	5.888433e-04	0.0006328991	-1.19547665	2.623676e-05
37	1.77464012	0.5634945	3.076818e-04	-1.367872e-03	0.0014020490	1.34954405	1.287561e-04
38	1.82260337	0.5486657	7.966094e-04	-1.493402e-03	0.0016925823	1.08077200	1.876467e-04
39	1.87056662	0.5345974	3.951427e-04	-2.576959e-04	0.0004717467	0.57789160	1.457669e-05
40	1.91852986	0.5212324	-3.668132e-05	-3.362969e-04	0.0003382915	-1.46215162	7.495895e-06
41	1.96649311	0.5085195	2.758703e-04	-7.867621e-04	0.0008337260	1.23355143	4.552899e-05
42	2.01445636	0.4964118	-3.849164e-05	-3.839125e-04	0.0003858373	-1.47086878	9.751014e-06
43	2.06241960	0.4848674	1.228237e-03	-5.733382e-04	0.0013554644	0.43673462	1.203421e-04
44	2.11038285	0.4738477	-4.790360e-04	-4.077087e-05	0.0004807679	-0.08490562	1.513953e-05
45	2.15834610	0.4633177	-5.107497e-04	-6.366272e-04	0.0008161859	-0.89467013	4.363344e-05
46	2.20630934	0.4532456	1.210339e-03	-7.011997e-05	0.0012123689	0.05786945	9.627441e-05
47	2.25427259	0.4436021	2.256535e-04	-3.054342e-04	0.0003797493	0.93450455	9.445724e-06
48	2.30223584	0.4343604	-1.495696e-04	-1.186235e-04	0.0001908994	-0.67051872	2.386990e-06
49	2.35019908	0.4254959	1.096871e-03	-3.403344e-04	0.0011484568	0.30085888	8.639142e-05
50	2.39816233	0.4169860	5.478249e-04	-2.295263e-04	0.0005939650	0.39675844	2.310803e-05
51	2.44612558	0.4088098	3.180891e-04	-5.428895e-04	0.0006292135	1.04079514	2.593208e-05
52	2.49408882	0.4009480	-3.528393e-04	6.147261e-04	0.0007087904	1.04973040	3.290614e-05
53	2.54205207	0.3933830	3.334777e-04	8.110359e-04	0.0008769188	-1.18069359	5.036862e-05
54	2.59001532	0.3860981	2.075319e-03	3.923407e-04	0.0021120793	-0.18684569	2.921876e-04

55	2.63797856	0.3790781	1.148417e-03	-1.121643e-03	0.0016052868	0.77360450	1.687899e-04
56	2.68594181	0.3723089	-5.626839e-04	-4.030420e-04	0.0006921387	-0.62157207	3.137817e-05
57	2.73390506	0.3657771	1.820756e-03	-1.943188e-04	0.0018310956	0.10632180	2.196157e-04
58	2.78186830	0.3594706	1.727725e-03	-5.538987e-04	0.0018143426	0.31024183	2.156155e-04
59	2.82983155	0.3533779	1.556185e-04	-1.921047e-03	0.0019273403	1.48996571	2.433090e-04
60	2.87779480	0.3474883	3.654281e-04	3.333941e-04	0.0004946608	-0.73959027	1.602715e-05
61	2.92575804	0.3417918	1.150607e-03	1.125550e-03	0.0016095833	-0.77439011	1.696947e-04
62	2.97372129	0.3362790	1.726059e-03	-2.961115e-04	0.0017512740	0.16989970	2.008859e-04
63	3.02168454	0.3309412	1.081432e-03	-1.309441e-03	0.0016982730	0.88047699	1.889106e-04
64	3.06964778	0.3257703	-3.246632e-04	8.699341e-05	0.0003361161	0.26179988	7.399799e-06
65	3.11761103	0.3207584	1.507766e-03	-1.383658e-04	0.0015141012	0.09151244	1.501589e-04



**Figure (3-6)**

### **Periodogram**

It is clear from figure (3-6) that the changes in periodogram are not random, and they have a certain pattern. It is also noted from figure (3-6) and table (3-18) that there are three distinctive peaks at the frequencies:

$$\lambda_1 = 0.04796325$$

$$\lambda_{10} = 0.47963247$$

$$\lambda_{15} = 0.71944870$$

this indicates that there is a periodic in the data. In order to get more precise results, the Fisher test has been conducted, which is given in formula (2-74), and the result is:

$$T = 0.05961395$$

And because there is no critical value for Fisher test at  $h = 65$ , then the critical value has been calculated by using formula (2-76) at significant level  $\alpha = 0.05$  and the result is:

$$b_{\alpha} = 0.1059853503$$

By comparing the critical value  $b_{\alpha}$  with the Fisher statistic  $T$ , then the null hypothesis is accepted which means that the data of the  $W_t$  is random and there is no hidden periodicities.

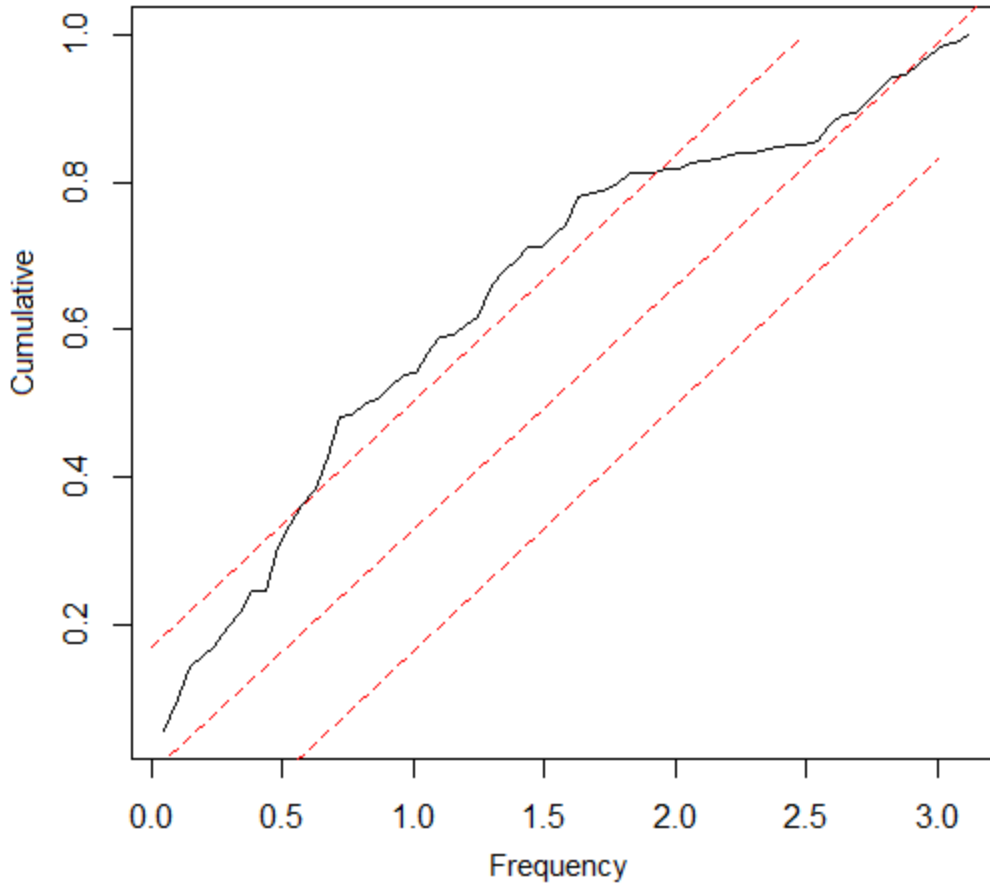
Due to the results conflicts between the periodogram and the Fisher test, the cumulative periodogram test has been applied by calculating the cumulative periodogram by using the formula (2-77) as shown in table (3-19) and then apply the Kolmogorov Smirnov test by plotting the theoretical line from the point  $(0,0)$  to the point  $(\pi, 1)$ . The upper limit and lower limit have also been plotted by using the formulas (2-78) and (2-79) respectively as shown in figure (3-6).

**Table (3-19)**

**Cumulative Periodogram of  $W_t$**

<b>I</b>	<b>Frequency</b>	<b>Cumulative Periodogram</b>	<b>I</b>	<b>Frequency</b>	<b>Cumulative Periodogram</b>
1	0.04796325	0.05662134	34	1.63075039	0.78135100
2	0.09592649	0.10028182	35	1.67871363	0.78619289
3	0.14388974	0.14151430	36	1.72667688	0.78817516
4	0.19185299	0.15719940	37	1.77464012	0.79790307
5	0.23981623	0.16845211	38	1.82260337	0.81208035
6	0.28777948	0.19623583	39	1.87056662	0.81318167
7	0.33574273	0.21702796	40	1.91852986	0.81374800
8	0.38370597	0.24443351	41	1.96649311	0.81718786
9	0.43166922	0.24451467	42	2.01445636	0.81792458
10	0.47963247	0.30412862	43	2.06241960	0.82701679
11	0.52759571	0.33498110	44	2.11038285	0.82816063
12	0.57555896	0.36294436	45	2.15834610	0.83145727
13	0.62352221	0.38487226	46	2.20630934	0.83873110
14	0.67148545	0.42658329	47	2.25427259	0.83944475
15	0.71944870	0.48200256	48	2.30223584	0.83962509
16	0.76741195	0.48412247	49	2.35019908	0.84615223
17	0.81537519	0.49955171	50	2.39816233	0.84789811
18	0.86333844	0.50671342	51	2.44612558	0.84985736
19	0.91130169	0.52185374	52	2.49408882	0.85234352
20	0.95926493	0.53904314	53	2.54205207	0.85614902
21	1.00722818	0.54052100	54	2.59001532	0.87822469
22	1.05519143	0.57009651	55	2.63797856	0.89097729
23	1.10315467	0.59086227	56	2.68594181	0.89334800
24	1.15111792	0.59345538	57	2.73390506	0.90994064
25	1.19908117	0.60740840	58	2.78186830	0.92623105
26	1.24704441	0.61865762	59	2.82983155	0.94461378
27	1.29500766	0.66072628	60	2.87779480	0.94582468
28	1.34297091	0.67958725	61	2.92575804	0.95864564
29	1.39093415	0.69325339	62	2.97372129	0.97382318
30	1.43889740	0.71312249	63	3.02168454	0.98809596
31	1.48686065	0.71350021	64	3.06964778	0.98865504
32	1.53482389	0.72778247	65	3.11761103	1.00000000
33	1.58278714	0.74131112			





**Figure (3-7)**  
**Cumulative Periodogram of  $W_t$**

From figure (3-7), it is clear that the cumulative periodogram lies outside the upper limit which means that the series  $W_t$  is not random and has hidden periodic components. Based on the cumulative periodogram results and the periodogram plot, it is concluded that the series  $W_t$  is not random and it has hidden harmonic components.

### **(3-3-2) Choosing the Harmonic Model**

After checking that there are hidden harmonic components,  $F_k$  statistic was calculated using formula (2 – 80), as shown in table (3 – 20), in order to find the significant harmonic components in the harmonic model.

**Table (3-20)** **$F_k$  Statistic Application Results**

<b>k</b>	<b>Frequency</b>	<b>F-Statistic</b>	<b>k</b>	<b>Frequency</b>	<b>F-Statistic</b>
1	0.04796325	3.841263532	34	1.63075039	2.669436583
2	0.09592649	2.921839197	35	1.67871363	0.311388428
3	0.14388974	2.752365939	36	1.72667688	0.127117126
4	0.19185299	1.019842640	37	1.77464012	0.628702526
5	0.23981623	0.728369720	38	1.82260337	0.920394664
6	0.28777948	1.828973830	39	1.87056662	0.070561791
7	0.33574273	1.358951182	40	1.91852986	0.036266160
8	0.38370597	1.803378278	41	1.96649311	0.220910593
9	0.43166922	0.005194758	42	2.01445636	0.047184775
10	0.47963247	4.057155804	43	2.06241960	0.587240968
11	0.52759571	2.037417789	44	2.11038285	0.073289431
12	0.57555896	1.841133116	45	2.15834610	0.211682804
13	0.62352221	1.434848268	46	2.20630934	0.468935871
14	0.67148545	2.785700301	47	2.25427259	0.045706437
15	0.71944870	3.754928909	48	2.30223584	0.011544125
16	0.76741195	0.135962744	49	2.35019908	0.420481252
17	0.81537519	1.002945982	50	2.39816233	0.111931911
18	0.86333844	0.461655906	51	2.44612558	0.125638041
19	0.91130169	0.983876631	52	2.49408882	0.159510775
20	0.95926493	1.119362529	53	2.54205207	0.244482597
21	1.00722818	0.094722730	54	2.59001532	1.444736099
22	1.05519143	1.950520308	55	2.63797856	0.826708920
23	1.10315467	1.357191961	56	2.68594181	0.152086419
24	1.15111792	0.166390435	57	2.73390506	1.079846238
25	1.19908117	0.905629647	58	2.78186830	1.059851496
26	1.24704441	0.728141247	59	2.82983155	1.198527339
27	1.29500766	2.810633769	60	2.87779480	0.077591585
28	1.34297091	1.230307102	61	2.92575804	0.831197742
29	1.39093415	0.886751384	62	2.97372129	0.986333087
30	1.43889740	1.297400132	63	3.02168454	0.926683895
31	1.48686065	0.024183504	64	3.06964778	0.035800975
32	1.53482389	0.927308872	65	3.11761103	0.734409593
33	1.58278714	0.877707418			

By comparing the statistic  $F_k$  for every harmonic component with the critical value of the distribution F at a degree of freedom (2,128), we obtain three significant harmonic components at the frequencies  $\lambda_1$ ,  $\lambda_{10}$  and  $\lambda_{15}$  because the calculated F at these frequencies are greater than the critical value  $F(2,128) = 3.00$  so that the alternative hypothesis is accepted at these frequencies only.

The harmonic model to represent the data of the  $W_t$  can be written as follow:

$$\widehat{W}_t = \sum_{l=1,10,15} [\xi_l \cos(\lambda_l t) + \Omega_l \sin(\lambda_l t)] + u_t \quad \text{where } t = 1, 2, \dots, N$$

### (3-3-3) Check the Suitability of Harmonic Model

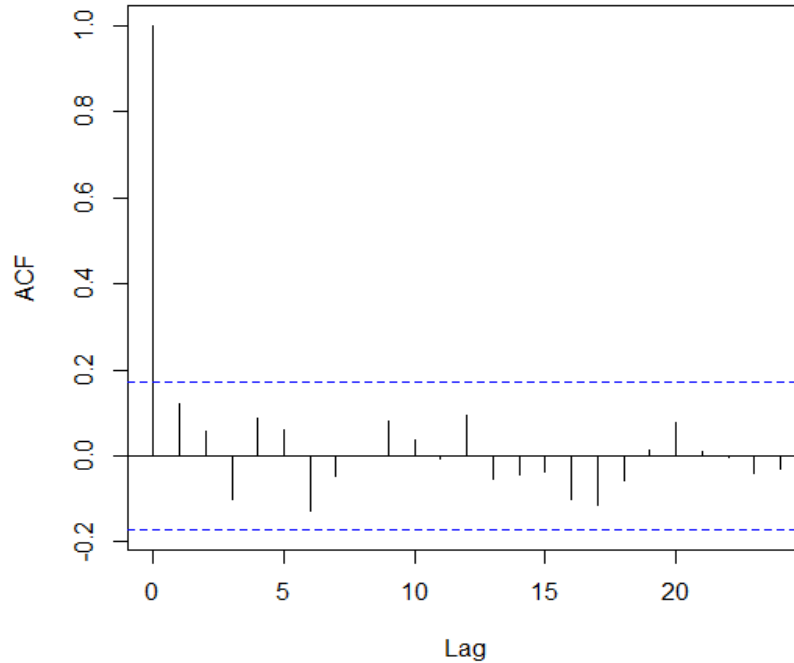
It is assumed that the residuals of the harmonic model are purely random and are not containing hidden periodic component, and to check that the following tests are applied on the residuals series:

- 1- Calculate and plot the SACF of the residuals and then apply the Ljung – Box test by using the formula (2-46) and this is shown in figure (3-7), table (3-21), and table (3-22) respectively.

**Table (3-21)**

**Autocorrelation Coefficients of Residuals**

<b>Lag=k</b>	<b>SACF= <math>r_k</math></b>	<b>Lag=k</b>	<b>SACF= <math>r_k</math></b>
1	0.120	13	-0.053
2	0.058	14	-0.043
3	-0.101	15	-0.036
4	0.089	16	-0.100
5	0.061	1	-0.114
6	-0.127	18	-0.057
7	-0.048	19	0.014
8	0.001	20	0.077
9	0.081	21	0.011
10	0.036	22	-0.003
11	-0.007	23	-0.041
12	0.093	24	-0.031



**Figure (3-8)**  
**Autocorrelation Function for Residuals**

**Table (3-22)**  
**Box-Ljung Test**

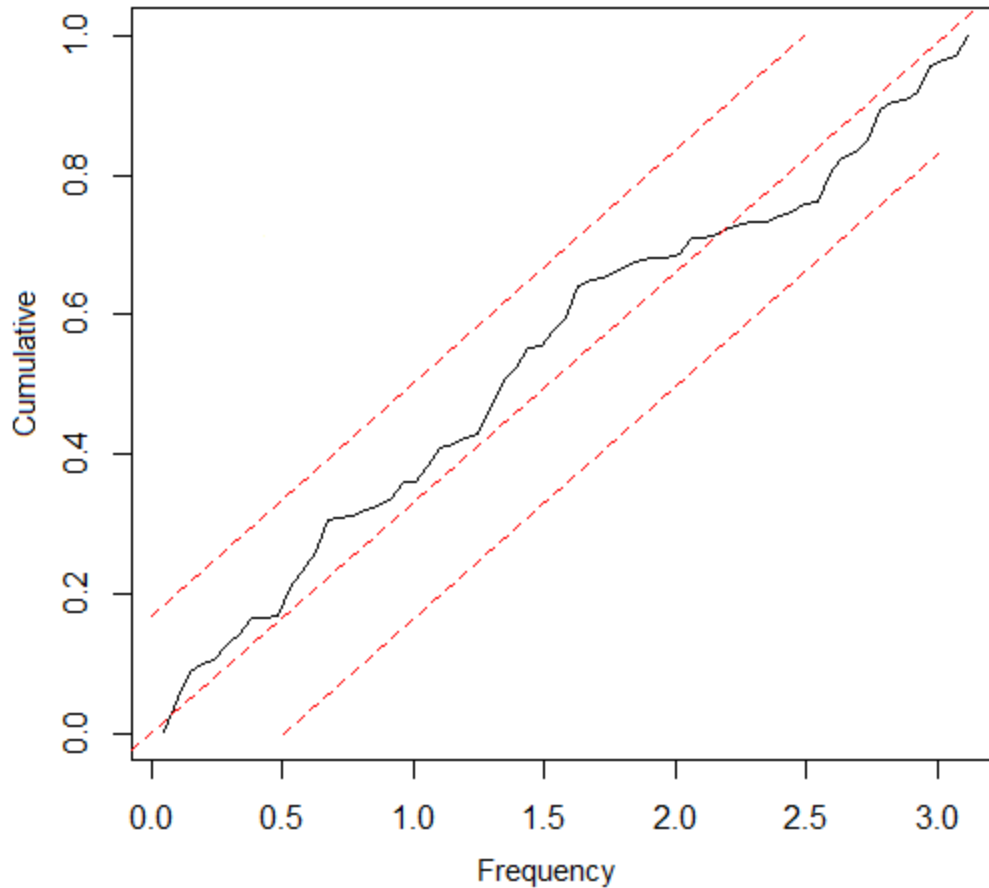
$(Q_{LB})$	$\chi^2(df = 18, \alpha = 0.05)$	p-value
25.244	28.869	0.3926

From figure (3-8) and table (3-22), it can be concluded that the residuals represent the white noise

- 2- The Fisher test has been applied to the residual series by using the formula (2-74) and the results is  $T = 0.05270805$ . by comparing the value of the statistic T with the critical value  $b_{\alpha} = 0.1059853503$ , it can be concluded that the residuals are not containing any hidden periodic component.
- 3- The cumulative periodogram test has been applied to the residuals by using the formula (2-77) as shown in table (3-23). Then the Kolmogorov Smirnov has been applied as shown in figure (3-9).

**Table (3-23)**  
**Cumulative Periodogram of Residuals**

<b>I</b>	<b>Frequency</b>	<b>Cumulative Periodogram</b>	<b>I</b>	<b>Frequency</b>	<b>Cumulative Periodogram</b>
1	0.04796325	0.002456905	34	1.63075039	0.642122953
2	0.09592649	0.049364103	35	1.67871363	0.649462408
3	0.14388974	0.086998336	36	1.72667688	0.653215022
4	0.19185299	0.097763410	37	1.77464012	0.660930585
5	0.23981623	0.104318645	38	1.82260337	0.673069633
6	0.28777948	0.127416094	39	1.87056662	0.678540542
7	0.33574273	0.142645394	40	1.91852986	0.681636745
8	0.38370597	0.164689285	41	1.96649311	0.682242938
9	0.43166922	0.166435225	42	2.01445636	0.686899243
10	0.47963247	0.168892129	43	2.06241960	0.710910509
11	0.52759571	0.209350655	44	2.11038285	0.711466656
12	0.57555896	0.232573413	45	2.15834610	0.717534644
13	0.62352221	0.260324792	46	2.20630934	0.724477876
14	0.67148545	0.305013206	47	2.25427259	0.730865416
15	0.71944870	0.307470111	48	2.30223584	0.732205159
16	0.76741195	0.310084383	49	2.35019908	0.734456228
17	0.81537519	0.320399035	50	2.39816233	0.741049085
18	0.86333844	0.325930308	51	2.44612558	0.748517399
19	0.91130169	0.335472151	52	2.49408882	0.760232087
20	0.95926493	0.359116209	53	2.54205207	0.762898112
21	1.00722818	0.359307275	54	2.59001532	0.803427931
22	1.05519143	0.384229510	55	2.63797856	0.825174296
23	1.10315467	0.410139975	56	2.68594181	0.832473328
24	1.15111792	0.413911360	57	2.73390506	0.852261452
25	1.19908117	0.423220586	58	2.78186830	0.893125736
26	1.24704441	0.429838196	59	2.82983155	0.906569946
27	1.29500766	0.468594839	60	2.87779480	0.907918926
28	1.34297091	0.508039434	61	2.92575804	0.920170941
29	1.39093415	0.522508111	62	2.97372129	0.957748156
30	1.43889740	0.551785197	63	3.02168454	0.967072495
31	1.48686065	0.555959171	64	3.06964778	0.972255970
32	1.53482389	0.579790583	65	3.11761103	1.000000000
33	1.58278714	0.595965747			



**Figure (3-9)**

**Cumulative Periodogram of Residuals**

From figure (3-9), it can be concluded that the residuals are not containing any hidden harmonic component. This proves the efficiency of the harmonic model to represent the data.

By comparing the spectral analysis methodology and the neural network methodology in prediction of the exchange of the Iraqi dinar vs the US Dollar, it can be clearly seen that the neural network methodology is better than the spectral analysis methodology according to the three criteria below:

**Table (3-24)**

**Comparison between Artificial Neural Networks and Spectral Analysis**

<b>Methodology</b>	<b>MSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Artificial Neural Networks</b>	170.5924	7.5583	.5800
<b>Spectral Analysis</b>	394.7326	11.2156	.81900

## **(3-4) Conclusions and Recommendations**

### **(3-4-1) Conclusions**

The following conclusions have been discovered:

- 1- The series of the exchange rate (raw data) is not mean stationary as stated in figures (3-1) and (3-2) respectively.
- 2- The raw data is not variance stationary as shown in figure (3-1) and table (3-2).
- 3- The criterion AIC tends to overestimate the models while the BIC criterion successfully identifies the best model to represent the data as stated in table (3-5).
- 4- The Box–Jenkins methodology showed that the best model to represent the data is the ARIMA(1,1,0) model .
- 5- Based on the MSE, MAE, and MAPE, the best activation function among all is the hyperbolic tangent activation function for the hidden layer, and the linear activation function for the output layer as stated in table (3-13) .
- 6- The best way to choose the number of the hidden nodes in the hidden layer that consists of one hidden layer is to choose the second formula (2R) as stated in (3-2-6) and in table (3-10).
- 7- In the neural networks methodology, if the sample size is small then it is not preferable to split the data into two portions (training set and testing set) as stated in table (3-15).
- 8- There is no specific rule to choose the value of the momentum and the learning rate, so that it is preferred to start with more than one value and then decide which one is the best because the value of the momentum affects the training time and the cyclic error as stated in table (3-14).
- 9- Processing the data before applying it to the network is crucial and needs to be done in advance as stated in table (3-16).

10- The best neural network model to predict the exchange rate of the Iraqi Dinar vs the US Dollar is the design with hyperbolic activation function in the hidden layer, linear activation function in the output layer, 2 nodes in the hidden layer, learning rate = 0.6, momentum = 0.5, and with weights and biases as shown in table (3-17).

11- The stationary exchange rate series contains hidden periodicities as stated in figures (3-5) and (3-6) respectively.

12- The Fisher test fails in detecting the hidden periodic component in the exchange rate series which is clearly shown in section (3-3-1) .

13- The statistical approaches aim at helping the specialist in financial and banking science to make the proper decisions that achieve the desired interest and promoting their business management.

14- The artificial neural networks methodology is better than the spectral analysis methodology in making accurate predictions on the exchange rate of the Iraqi dinar against the USdollar.



### **(3-4-2) Recommendations**

- 1- Comparing the backpropagation networks and the Jordan or Elman networks in predicting the exchange rate.
- 2- Applying the hybrid methodology to predict the exchange rate. Then comparing the hybrid model and the pure neural network model to choose the best.
- 3- Conducting a research on choosing the truncation point in the spectral windows and the lag windows as show in (2-2-10) in order to estimate a consistent spectrum.
- 4- We recommend obtaining of the advanced statistical approaches by the specialists in financial and banking science to study the behavior of the exchange rate, to build future predictions and to make proper decisions.

# References

- [1] Agami, N., Atiya, A., Saleh, M. and El-Shishiny, H., 2009. A neural Network Based Dynamic Forecasting Model for Trend Impact Analysis. Elsevier, 76(7), PP.952-962.
- [2] Al-douri, A., 1991. The Spectral Analysis and Harmonic Analysis of the Time Series with Application from the Earthquakes in Sulaimaniya Governorate. Master Thesis, University of Baghdad, Department of Statistics, Iraq.
- [3] Al\_qabbi, A., 2001. Using Spectral Analysis in Bivariate Autoregressive. P.HD Dissertation, University of Baghdad, Department of Statistics, Iraq.
- [4] Al\_zaidy, A., 2004. Spectral Analysis of Nucleic-acid Chain DNA. P.HD Dissertation, Al-Mustansiria University, Department of Statistics, Iraq.
- [5] Al\_zubay, O., 2005. Identification and Diagnostic Checking for Low Order Mixed Models. P.HD Dissertation, University of Baghdad, Department of Statistics, Iraq.
- [6] Bloomfield, P., 2000. Fourier Analysis of Time Series: An Introduction. John Wiley & sons, U.S.A.
- [7] Box, G., Jenkins, G. and Reinsel, G., 2008. Time Series Analysis: Forecasting and Control. John Wiley & Sons, New Jersey.
- [8] Broersen, P., 2002. Automatic Spectral Analysis with Time Series Models. IEEE, 51(2), PP.211-216.
- [9] Chatfield, C., 1975. The Analysis of Time Series: Theory and Practice. John Wiley & Sons, New York.
- [10] Chatfield, C., 2004. The Analysis of Time Series: An Introduction. Chapman & Hall/CRC, New York.
- [11] Chen, T. and Chen, H., 1995. Universal Approximation to Nonlinear Operators by Neural Networks with Arbitrary Activation Functions and Its Application to Dynamical Systems. IEEE, 6(4), PP. 911- 917.

- [12] Fernández, M., Gázquez, J., Rose, M. and Castellano, N., 2016. Optimization of the Periodogram Average for the Estimation of the Power Spectral Density (PSD) of Weak Signals in the ELF Band. Elsevier, 78(1), PP.207 – 218.
- [13] Fuller, W., 1996. Introduction to Statistical Time Series. John Wiley & Sons, New York.
- [14] Garrett, L. and Fogarty, K., 2012. Using Box-Jenkins Modeling Techniques to Forecast Future Disease Burden and Identify Disease Aberrations in Public Health Surveillance Report. P.HD Dissertation, Western Michigan University.
- [15] Hagan, M. and Demuth, H., 2008. Neural Network Design. Amazon, 2nd Edition.
- [16] Halliday, D., Rosenberg, J., Rigas, A. and Conway, B., 2009. A Periodogram –Based Test for Weak Stationarity and Consistency Between Sections in Time Series. Elsevier, 180(1), PP.138-146.
- [17] Hanke, J. and Wichern, D., 2009. Business Forecasting. PHI Learning, U.S.A.
- [18] Jenkins, G. and Watts, D., 1986. Spectral Analysis and Its Application. Holden-Day, U.S.A.
- [19] Jeong, H., 2012. The Spectral Analysis of Non Stationary Categorical Time Series Using Local Spectral Envelope. P.HD Dissertation, University of Pittsburgh, Department of Statistics.
- [20] Jentsch, C. and Pauly, M., 2012. A Note on Using Periodogram-Based Distances for Comparing Spectral Densities. Elsevier, 82(1), PP.158 – 164.
- [21] Kaastra, L and Boyd, M., 1996. Designing a Neural Network for Forecasting Financial and Economic Time Series. Elsevier, 10(3), PP. 215-236.

- [22] Kalogirou, S., 2001. Artificial Neural Networks in Renewable Energy Systems Applications: A Review . Elsevier, 5(4), PP.373 – 401.
- [23] Kamruzzamana, J. and Sarkerb, R., 2004. ANN-Based Forecasting of Foreign Currency Exchange Rates. Researchgate, 3(2), PP.49-58.
- [24] Kang, S., 1992. An Investigation of the Use of Feedforward Neural Networks for Forecasting. Ph.D. Dissertation, Kent State University, USA.
- [25] Khan, M., 2013. Forecasting of Gold Prices (Box Jenkins Approach). International Journal of Emerging Technology and Advanced Engineering, 3(3), PP.662-670.
- [26] Khashei, M. and Bijari,M., 2010. An Artificial Neural Network (p,d,q) Model for Time Series Forecasting, Elsevier, 37(1), PP.479 – 489.
- [27] Kihoro, J., Otieno, R. and Wafula, C., 2004. Seasonal Time Series Forecating: A Comparative Study of ARIMA and ANN Models. Unesco, 5(2), PP.41- 49.
- [28] Konar, A., 2000. Artificial Intelligence and Soft Computing : Behavioral and Cognitive Modeling of the Human Brain. CRC Press LLC, United States of America.
- [29] Kondratenko, V. and Kuperin, Y., 2003. Using Recurrent Neural Networks To Forecasting of Forex. Cornell University Library.
- [30] Koopmans, L., 1995. The Spectral Analysis of Time Series. Academic Press, United Kingdom.
- [31] Kulkarni, M., Patil, S., Rama, G., and Sen, P., 2008. Wind Speed Prediction Using Statistical Regression and Neural Network. Journal of Earth System Science, 117(4), PP.457–463.
- [32] Kumar, P., Murthy, M., Eashwar, D. and Venkatdas, M., 2002. Time Series Modeling Using Artificial Neural Networks. Journal of Theoretical and Applied Information Technology, 4(12), PP.1259-1264.

- [33] Lepage, K., 2009. Some Advances in the Multitaper Method of Spectrum Estimation. P.HD Dissertation, Department of Mathematics and Statistics, Canada.
- [34] Levich, R. & Thomas, L., 1993. The Significance of Technical Trading-Rule Profits in the Foreign Exchange Market: A Bootstrap Approach. Elsevier, 12(5), PP.451- 474.
- [35] Mahmood, S., 2017. Some Robust and Non-Parametric Estimators for Power Spectrum According to the ARMA Models with Application. P.HD Dissertation, University of Baghdad, department of Statistics, Iraq.
- [36] Meng, X., Jia, M. and Wang, T., 2014. Neural Network Prediction of Biodiesel Kinematic Viscosity at 313 K. Elsevier, 121, PP.133-140.
- [37] Montgomery, D., Jennings, C. and Kulahci, M., 2015. Introduction to Time Series Analysis and Forecasting. John Wiley & Sons, New Jersey.
- [38] Nasser, K., 2015. Traditional Methods and Alternative Methods to Test the Unit Root for the Stationary of Time Series : a comparative study. P.HD Dissertation, Al-Mustansiria University, Department of Statistics.
- [39] Novick, S., 1995. Analysis of Fisher's Test for Hidden Periodicities. Master Thesis, Leigh University, Department of Applied Mathematics.
- [40] Nowroozi, A., 1967. Table for Fisher's Test of Significance in Harmonic Analysis. Geophysical journal international, 12(5), PP.517-520.
- [41] Oancea, B. and Ciucu, S., 2014. Time Series Forecasting Using Neural Networks. Cornell University Library, PP.1403-1408.
- [42] Ojo, J., 2008. On the Performance and Estimation of Spectral and Bispectral Analysis of Time Series Data. Asian Journal of Mathematics and Statistics, 1(3), PP.150-158.
- [43] Palit, A. and Popovic, D., 2005. Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications. Springer, U.S.A.

- [44] Philip, A., Taofiki, A. and Bidemi, A., 2011. Artificial Neural Network Model for Forecasting Foreign Exchange Rate. *World of Computer Science and Information Technology Journal*, 1(3), PP.110 – 118.
- [45] Pissarenko, D., 2002. *Neural Networks For Financial Time Series Prediction: Overview Over Recent Research: A project*. Researchgate.
- [46] Preuß, P. and Hildebrandt, T., 2013. Comparing Spectral Densities of Stationary Time Series with Unequal Sample Sizes. *Elsevier*, 83(4), PP.1174 – 1183.
- [47] Priestley, M., 1989. *Spectral Analysis and Time Series*. Academic Press, Great Britain .
- [48] Reinert, G., 2010. *Time Series*. <http://www.stats.ox.ac.uk/~reinert>.
- [49] Schneburg, E., 1990. *Stock Price Prediction Using Neural Networks: A Project Report*. Elsevier, 2(1), PP.17-27.
- [50] Sharda, R. and Patil, R., 1992. Connectionist Approach to Time Series Prediction: An Empirical Test. *Springer*, 3(5), PP.317–323.
- [51] Shumway, R. and Stoffer, D., 2011. *Time Series Analysis and Its Applications With R Examples*. Springer, New York.
- [52] Spangl, B. and Dutter, R., 2007. Estimating Spectral Density Functions Robustly. *ResearchGate, Revstat – Statistical Journal*, 5(1), PP.41– 61.
- [53] Suzuki, K., 2011. *Artificial Neural Networks Methodological Advances and Biomedical Applications*. InTech, India.
- [54] Tang, Z. and Fishwick, P., 1993. Feed-forward Neural Nets as Models for Time Series Forecasting. *ORSA Journal of Computing*, PP.374-385.
- [55] Tarsauliya, A., Kant, S. & Kala,R.,(2010), "Analysis of Artificial Neural Network for Financial Time Series Forecasting" , *International Journal of Computer Applications*, Vol. 9, No.5.

- [56] Yaffee, R. and McGee, M., 1999. Introduction to Time Series Analysis and Forecasting with Applications of SAS and SPSS. Academic Press, United Kingdom.
- [57] Yan, L. and Choon, O., 2000. Neural Networks Forecasting on Electricity Consumption in Malaysia. ResearchGate, Malaysia.
- [58] Yixin, Z. and Zhang, J., 2010. Stock Data Analysis Based on BP Neural Network. IEEE.
- [59] Wei, W., 2006. Time Series Analysis :Univariate and Multivariate Methods. Pearson Addison Wesley, United States of America.
- [60] Wolters, J. and Kirchgassner, G., 2007. Introduction to Modern Time Series Analysis. Springer, New York.
- [61] Wong, F., 1991. Time Series Forecasting Using Backpropagation Neural Networks. Elsevier, 2(4), PP.147-159.
- [62] Zhang, G., Patuwo, E. and Hu, M.Y., 1998. Forecasting with Artificial Neural Networks: The State of the Art. Elsevier, 14(1), PP.35-62.
- [63] Zhang, G., 2003. Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. Elsevier, 50, PP.159 – 175.
- [64] Zhang, G. and Qi, M., 2005. Neural Network Forecasting for Seasonal and Trend Time Series. Elsevier, 160(2), PP.501–514.
- [65] Zurada, J., 1992. Introduction to Artificial Neural Systems. West Publishing Company, United States of America.



## الخلاصة

ان التقلبات التي تحدث في سعر الصرف وخصوصا الارتفاعات والانخفاضات المفاجئة غير المتوقعة لها تأثير كبير على الاقتصاد القومي لأي دولة ومنها العراق، لذا فإن التنبؤ الدقيق بسعر صرف الدينار العراقي مقابل الدولار الامريكي يعتبر من الامور المهمة والمؤثرة في عملية التخطيط والمحافظة على التوازن والاستقرار الاقتصادي للعراق.

يهدف هذا البحث الى المقارنة بين اسلوب التحليل الطيفي والشبكات العصبية الاصطناعية في التنبؤ بسعر صرف الدينار العراقي مقابل الدولار الامريكي للفترة الزمنية من 30/01/2004 الى 30/12/2014 وتم الاعتماد على متوسط مربع الخطأ MSE ومتوسط مطلق الخطأ MAE ومتوسط مطلق الخطأ النسبي MAPE كمعايير احصائية للمفاضلة بين النموذجين وقد بينت النتائج ان اسلوب الشبكات العصبية الاصطناعية افضل من اسلوب التحليل الطيفي في التنبؤ.



جمهورية العراق  
وزارة التعليم العالي والبحث العلمي  
جامعة القادسية  
كلية علوم الحاسوب والرياضيات  
قسم الرياضيات

# مقارنة بين التحليل الطيفي والشبكات العصبية الاصطناعية لتحديد أفضل نموذج للتنبؤ مع التطبيق

رسالة مقدمة إلى

مجلس كلية علوم الحاسوب والرياضيات – جامعة القادسية  
كجزء من متطلبات نيل درجة ماجستير في علوم الرياضيات

من قبل

**ايمان يوسف عبدعون الشباني**

بإشراف

**أ. د. محمد حبيب الشاروط**

2017 م

1439 هـ

رقم الايداع في دار الكتب والوثائق ببغداد ..... لسنة ٢٠١٧