Journal of AL-Qadisiyah for computer science and mathematics Vol.9 No.2 Year 2017
ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504

Prediction by using Artificial Neural Networks and Box-Jenkins methodologies: Comparison Study

Emaan Yousif Abdoon
Al -Qadisiyah University
College of Computer Science & Information Technology
Department of Mathematics
emaan.yousif1991@gmail.com

Mohammed Habib Al- Sharoot
Al –Qadisiyah University
College of Administration & Economics
Department of Statistics
m.alsharood@gmail.com

Recived : 9\7\2017 Revised : // Accepted : 21\8\2017

Abstract

The variations in exchange rate, especially the sudden unexpected increases and decreases, have significant impact on the national economy of any country. Iraq is no exception; therefore, the accurate forecasting of exchange rate of Iraqi dinar to US dollar plays an important role in the planning and decision-making processes as well as the maintenance of a stable economy in Iraq. This research aims to compare Box-Jenkins methodology to neural networks in terms of forecasting the exchange rate of Iraqi dinar to US dollar based on data provided by the Iraqi Central Bank for the period 30/01/2004 and 30/12/2014.

Based on the Mean Square Error (MSE), the Mean Absolute Error (MAE), and the Mean Absolute Percentage Error (MAPE) as criteria to compare the two methodologies, it was concluded that Box-Jenkins is better than neural network approach in forecasting.

Keywords: Time series analysis, Autoregressive Moving Average models, Artificial neural network, Backpropagation algorithm.

1- Time series [1,2,3]

Time series is a sequence of observations of a specific phenomenon throughout a previous time period. Usually, these observations are dependent and organized according to time. Time series can be classified to two types: stationary and non-stationary time series. The word stationary refers to the absence of growth in the data meaning that the data fluctuate around a constant level without any increasing or decreasing trend. Therefore, time series can be stationary if it has the two following conditions:

a. mean stationary

$$E(W_t) = \mu \quad for\ all\ t \ (1)$$

b. variance stationary

$$Var\,(W_t) = E(W_t - \mu)^2$$

$$= \gamma_0 \quad for\ all\ t \ (2)$$

In practice, most of time series, especially economic series, are non-stationary and difficult to model. Therefore, series that are not mean stationary can be transformed to stationary by taking the differences of d degree as follow:

$$(1-L)^d Y_t = Y_t - Y_{t-d} \qquad (3)$$

Where L is the backshift operator. The degree of differences usually equals to 1 or 2.

The model of time series is a function that relates the current value of time series to the past values and adds the random error. This model is divided to three types:

A. Autoregressive models [4]

These models are often referred to as AR(p) and can be written in the following formula:

$$W_t = \alpha_1 W_{t-1} + \alpha_2 W_{t-2} + \cdots + \alpha_p W_{t-p} + u_t \qquad (4)$$

α_j: Autoregressive parameters and j = (1,2, …, p)

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

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

B. Moving Average Models [5]

These models are often referred to as MA (q) and can be written in the following formula:

$$W_t = u_t - \beta_1 u_{t-1} - \beta_2 u_{t-2} - \cdots - \beta_q u_{t-q} \qquad (5)$$

β_τ : Moving average parameters and τ = (1, 2, …,q)

C. Autoregressive Moving Average Models [3,6]

These models are often referred to as ARMA (p, q) and can be written in the following formula:

$$W_t = \alpha_1 W_{t-1} + \alpha_2 W_{t-2} + \cdots + \alpha_p W_{t-p} + u_t - \beta_1 u_{t-1} - \beta_2 u_{t-2} - \cdots - \beta_q u_{t-q} \qquad (6)$$

2- Box-Jenkins methodology [7,8,4]

Introduced by Box and Jenkins in 1970, Box-Jenkins approach is one of the most distinguished statistical approaches to analyse the time series of specific phenomenon and forecasting the possible variations that may occur in the future based on previous observations of the phenomenon.

Box-Jenkins methodology mainly relies on representing stable data using autoregressive moving average models. This methodology consists of three repetitive phases including identification, parameters estimation, and diagnostic checking. The identification phase includes the selection of the proper model to represent the data and determining its rank after examining the stability of the data in the mean and variance by plotting the orginal time series, plotting the autocorrelation function of the data, and applying one of the three tests that include Augmented Dickey – Fuller (ADF), Phillips – Perron (P.P), and Kwiatkowski – Phillips – Schmidt – Shin (KPSS).

If the time series is not mean stationary, appropriate number of differences can be taken to accomplish stationary in the mean. If the time series is not variance stationary, specific transformations can be taken, such as the log or the square root, to accomplish stability. In the second phase, the model is estimated using one of the estimation methods, such as Ordinary Least Square method, moments method, the conditional maximum likelihood, and the exact maximum likelihood method.

In the third phase, the randomization of residuals of the estimated model is examined by plotting the autocorrelation function of the residuals or applying Box- Pierce test, often referred to as Q_{BP}, or Ljung -Box test, often referred to as Q_{LB}. When the model pass the third phase successfully, the model is used to forecast the future values of the studied phenomenon using the following formula:

$$\widehat{W}_t(l) = E(W_{t+l})$$

$$= \alpha_1 E(W_{t+l-1}) + \cdots + \alpha_{p+d} E\left(W_{t+l-p-d}\right)$$

$$-\beta_1 E(u_{t+l-1}) - \ldots - \beta_q E\left(u_{t+l-q}\right) + E(u_{t+l}) \qquad (7)$$

where l represents the length of the forecasted time period.

3-Artificial neural networks

[9,10,11,12,13,14,15,16]

Artificial neural network is a computational technique that simulates the way human brain uses to perform a specific task. It consists of several processing units defined as neurons or nodes that are organized in three levels (layers) including input level, hidden level that consists of one or more hidden layer, and the output level. the nodes of each level are associated with the next level through connection force called weights, which work on saving the acquired knowledge from the training of network.

The nodes at the input level are called the input nodes and so on for the other levels. In addition, there is a node called the bias b that has a positive value of one and has the same constant role in the regression model.

Each node of the hidden and output levels is provided with an activation function, linear or nonlinear, that works on processing the input signal and preventing the output of the processing node from reaching high value, which may stop the network leading to a failure of the training process.

Neural networks are divided to two types including single layer and multiple layer networks. The single layer network does not have the hidden level and contains one layer of weights that connects the input level to the output level. when applying the input signal v_j in this type of network, we can obtain the output signal z_i through the following formula:

$$z_i = g\left(\sum_{j=1}^{R} x_j\, v_j + b_i\right) \quad : \; i = 1,2,\ldots,n \qquad (8)$$

Where

x_j: weight

As for multi-layer networks, it has the ability to solve more complex problems because it contains the hidden level with one hidden layer or more. The output signal is obtained in this type of networks as follows

$$\boldsymbol{Z}^m = \boldsymbol{g}^m(\boldsymbol{X}^m \boldsymbol{Z}^{m-1} + \boldsymbol{b}^m) : \; m = 1,2,\ldots,M \; (9)$$

Where:

M: number of layers

Z^{m-1}: output vector of the layer m-1

g^m : activation function of the layer m

b^m: bias vector of the layer m

X^m: weights matrix of the layer m

Z^0 = V, and V is the input vector.

(3-1) Backpropagation algorithm [17,18]

The backpropagation algorithm is a generalization of the least mean squares algorithm where it is used to train multiple layers networks. It is often referred to as BP and is considered one of the most used algorithms among the supervised learning algorithms. It consists of three stages:

1- **Feed forward propagation stage** [10,19]
In this stage, inputs $\mathbf{Z}^0$ = $\mathbf{V}$ are applied on the network and random initial weights $\mathbf{X}^1$ are generated in small values. In addition, this stage includes determining the learning rate η and the momentum γ by small value that falls within (0, 1) and (1, 0] respectively. Data are processed starting from the input layer to the output layer throughout

the hidden layers. The outputs of the network are obtained using the following formula:

$$z_j^m = g^m\left(s_j^m\right) \quad m = 1,2,\dots,M \quad (10)$$

$$= g_j^m\left(\sum_{h=1}^{n^{m-1}} x_{j,h}^m\, z_h^{m-1} + b_j^m\right) \qquad (11)$$

Which can be written as a matrix

$$\boldsymbol{Z}^m = \boldsymbol{g}^m(\boldsymbol{s}^m) \qquad (12)$$

$$= \boldsymbol{g}^m(\boldsymbol{X}^m \boldsymbol{Z}^{m-1} + \boldsymbol{b}^m) \qquad (13)$$

2-Backward Propagation Stage [10,16]

In this stage, the sensitivities $\boldsymbol{\delta}$ is calculated from the last layer, which represents the output layer, to the first layer throughout the hidden layers.

Sensitivity $\boldsymbol{\delta}^{\mathbf{M}}$ in the last layer M can be calculated using the following formula:

$$\boldsymbol{\delta}^M = -2\dot{\boldsymbol{G}}^M(\boldsymbol{s}^M)(\boldsymbol{d} - \boldsymbol{z}) \qquad (14)$$

Where

d: desired output

z: calculated output by network

$$\dot{\boldsymbol{G}}^M(\boldsymbol{s}^M) = \begin{bmatrix} \dot{g}_1^M(s_1^M) & 0 & \cdots & 0 \\ 0 & \dot{g}_2^M(s_2^M) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \dot{g}_{n^M}^M(s_{n^M}^M) \end{bmatrix} \text{ and}$$

$$\dot{g}_j^M\left(s_j^M\right) = \frac{\partial}{\partial s_j^M}\, g_j^M\left(s_j^M\right)\; j = 1,2,\dots,n^M \quad (15)$$

While sensitivity $\boldsymbol{\delta}^m$ in the hidden layer m, where m= 1,2, …, M-1, can be calculated using the following formula:

$$\boldsymbol{\delta}^m = \dot{\boldsymbol{G}}^m(\boldsymbol{s}^m)(\boldsymbol{X}^{m+1})^T \boldsymbol{\delta}^{m+1} \qquad (16)$$

Where

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

$$\dot{\boldsymbol{G}}^m(\boldsymbol{s}^m) = \begin{bmatrix} \dot{g}_1^m(s_1^m) & 0 & \cdots & 0 \\ 0 & \dot{g}_2^m(s_2^m) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \dot{g}_{n^m}^m(s_{n^m}^m) \end{bmatrix}$$

and

$$\dot{g}_j^m\left(s_j^m\right) = \frac{\partial}{\partial s_j^m}\, g_j^m\left(s_j^m\right) \;:\; j = 1,2,\dots,n^m \qquad (17)$$

3- Updating weights stage [19,16]

After passing the feed forward propagation and the Backward Propagation stages, the stage of updating the weights and biases begins using the following formulas:

$$\boldsymbol{X}^m(k) = \boldsymbol{X}^m(k-1) + \Delta\boldsymbol{X}^m(k-1) \qquad (18)$$

$$\boldsymbol{b}^m(k) = \boldsymbol{b}^m(k-1) + \Delta\boldsymbol{b}^m(k-1) \qquad (19)$$

Where

1- In case of not using the momentum:

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

$\Delta\boldsymbol{b}^m(k-1) = -\eta\boldsymbol{\delta}^m$

2- In case of using the momentum:

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

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

4- Data

The data used in this study is a time series of exchange rate of Iraqi dinar to the US dollar and was provided by the Iraqi Central Bank. The data consist of 132 monthly observations from 30/01/2004 to 30/12/2014 as show in table (1) below.

Table 1: Exchange rate of Iraqi dinar to the US dollar

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

5- Box- Jenkins methodology application

The first stage: identification

The time series of the exchange rate data was plotted as shown in figure (1) below

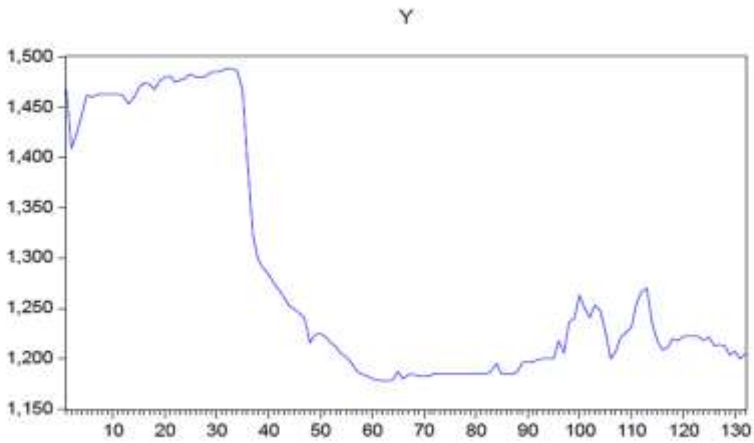
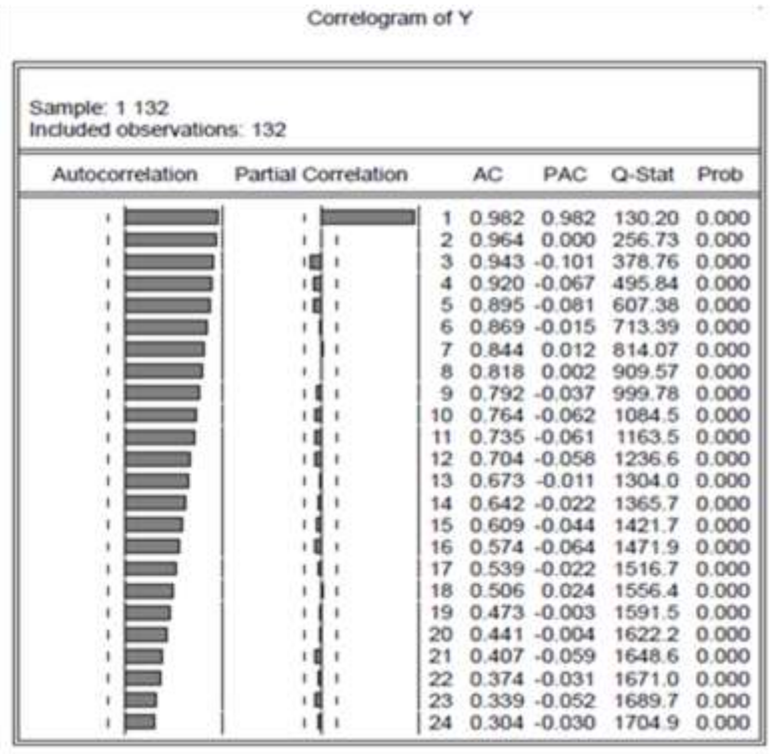


Figure 1: Exchange rate of Iraqi dinar to the US dollar

By looking at figure (1), we notice that the data Y_t does not fluctuate around constant level, and it takes a decreasing trend which indicates that the time series is not mean stationary and not variance stationary.

To check the accuracy of results about the stationary of the time series in the mean, autocorrelation function and partial autocorrelation function were plotted for the raw data as shown in figure (2) below.

Journal of AL-Qadisiyah for computer science and mathematics Vol.9 No.2 Year 2017
ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504



Correlogram of Y

Sample: 1 132
Included observations: 132

Lag	AC	PAC	Q-Stat	Prob
1	0.982	0.982	130.20	0.000
2	0.964	0.000	256.73	0.000
3	0.943	-0.101	378.76	0.000
4	0.920	-0.067	495.84	0.000
5	0.895	-0.081	607.38	0.000
6	0.869	-0.015	713.39	0.000
7	0.844	0.012	814.07	0.000
8	0.818	0.002	909.57	0.000
9	0.792	-0.037	999.78	0.000
10	0.764	-0.062	1084.5	0.000
11	0.735	-0.061	1163.5	0.000
12	0.704	-0.058	1236.6	0.000
13	0.673	-0.011	1304.0	0.000
14	0.642	-0.022	1365.7	0.000
15	0.609	-0.044	1421.7	0.000
16	0.574	-0.064	1471.9	0.000
17	0.539	-0.022	1516.7	0.000
18	0.506	0.024	1556.4	0.000
19	0.473	-0.003	1591.5	0.000
20	0.441	-0.004	1622.2	0.000
21	0.407	-0.059	1648.6	0.000
22	0.374	-0.031	1671.0	0.000
23	0.339	-0.052	1689.7	0.000
24	0.304	-0.030	1704.9	0.000

Figure 2: Correlogram of raw data Y_t

By looking at figure (2), we notice that the autocorrelation function is slowly decreasing toward the zero and does not cut after the first and the second lag, which indicates that the raw data is not mean stationary.

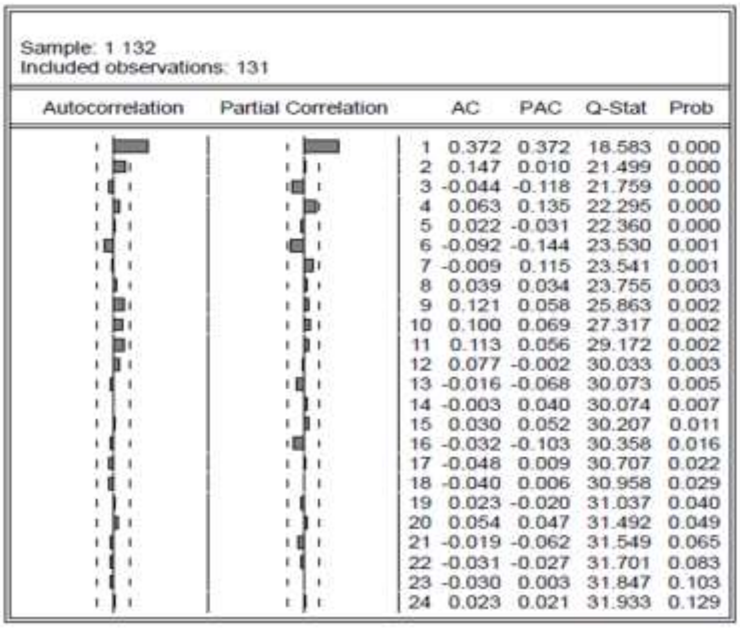
To increase accuracy in the results about the stationary of the time series in the variance, ADF, P.P, and KPSS tests were applied as shown in table (2) below.

Test	Model	Test Statistic	Critical Values	p-Value
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 of each model of the estimated models for both tests ADF and P.P at 0.05 significance level (alpha), we accept the null hypothesis and concluded that the time series has unit root meaning that it is not variance stationary. In addition, through ADF test, we concluded that the time series needs to take differences.

By comparing the calculated value of KPSS test statistic to the critical value, we accept the alternative hypothesis, which means that the time series is not variance stationary.

From the results of plots and tests, we conclude that the time series is not mean stationary and not variance stationary. Therefore, the log transformation was applied then the first difference was taken to accomplish stationary in the series in the variance and the mean respectively as shown in figure (3) which shows that the autocorrelation function of the transformed data is cut after the first lag, which indicates that the time series is mean stationary.


Correlogram of W

Sample: 1 132
Included observations: 131

	AC	PAC	Q-Stat	Prob
1	0.372	0.372	18.583	0.000
2	0.147	0.010	21.499	0.000
3	-0.044	-0.118	21.759	0.000
4	0.063	0.135	22.295	0.000
5	0.022	-0.031	22.360	0.000
6	-0.092	-0.144	23.530	0.001
7	-0.009	0.115	23.541	0.001
8	0.039	0.034	23.755	0.003
9	0.121	0.058	25.863	0.002
10	0.100	0.069	27.317	0.002
11	0.113	0.056	29.172	0.002
12	0.077	-0.002	30.033	0.003
13	-0.016	-0.068	30.073	0.005
14	-0.003	0.040	30.074	0.007
15	0.030	0.052	30.207	0.011
16	-0.032	-0.103	30.358	0.016
17	-0.048	0.009	30.707	0.022
18	-0.040	0.006	30.958	0.029
19	0.023	-0.020	31.037	0.040
20	0.054	0.047	31.492	0.049
21	-0.019	-0.062	31.549	0.065
22	-0.031	-0.027	31.701	0.083
23	-0.030	0.003	31.847	0.103
24	0.023	0.021	31.933	0.129

Figure 3: Correlogram of transformed data W_t

It becomes clear from the results of ADF, P.P, and KPSS tests of the transformed data, shown in table (3) below, that the time series is variance stationary. In addition, the results of ADF indicate that the series does not need to take anymore differences, which indicates that the series is mean stationary.

Journal of AL-Qadisiyah for computer science and mathematics Vol.9 No.2 Year 2017
ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504

Table 3: The tests results of transformed data W_t

Test	Model	Test Statistic	Critical Values	p-Value
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	

In this stage, the suggested primary model ARIMA(1,1,1) has been estimated by the Exact Maximum Likelihood method. Some models have been suggested which are very closes 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 (4) below

4: estimated parameters of ARIMA (p,d,q) models

ARIMA	parameters		P-Value	ARIMA	parameters		P-Value
(2,1,2)	$\hat{\alpha}_1$	0.180371	0.1145	(2,1,1)	$\hat{\alpha}_1$	0.866848	0.5128
	$\hat{\alpha}_2$	0.501118	06e-4.14		$\hat{\alpha}_2$	0.227888-	0.6809
	$\hat{\beta}_1$	0.635411	e-0151.28		$\hat{\beta}_1$	0.412057-	0.7584
	$\hat{\beta}_2$	0.834250	e-0347.00	(0,1,1)	$\hat{\beta}_1$	0.0347870	e-053.32
(1,1,2)	$\hat{\alpha}_1$	0.332391-	0.0773	(2,1,0)	$\hat{\alpha}_1$	0.443303	3.82e-06
	$\hat{\beta}_1$	0.836455	07e-4.24		$\hat{\alpha}_2$	-0.0253968	0.7905
	$\hat{\beta}_2$	0.515426	09e-4.44				
(1,1,0)	$\hat{\alpha}_1$	0.431265	08e-6.57				
(1,1,1)	$\hat{\alpha}_1$	0.404445	0.0293				
	$\hat{\beta}_1$	0.0338880	0.8672				

To select the best model among the estimated models to represent the data, some statistical criteria including AIC, BIC, H-Q, MSE, MAE, MPE, and MAPE were calculated as shown in table (5) below.

Table 5: Estimated criteria of ARIMA (p, d, q) models

ARIMA	Criterion		ARIMA	Criterion	
(2,1,2)	AIC	-856.1266	(2,1,1)	AIC	-846.4964
	H-Q	-850.2850		H-Q	-841.8231
	BIC	-841.7506		BIC	-834.9956
	MSE	8.2251e-005		MSE	8.8286e-005
	MAE	0.0055517		MAE	0.005339
	MPE	-0.014625		MPE	-0.012791
	MAPE	0.077613		MAPE	0.074591
(1,1,2)	AIC	-855.4026	(0,1,1)	AIC	-845.4281
	H-Q	-850.7329		H-Q	-843.0915
	BIC	-843.9054		BIC	-839.6777
	MSE	8.4133e-005		MSE	9.0732e-005
	MAE	0.005490		MAE	0.0054472
	MPE	-0.012031		MPE	-0.015472
	MAPE	0.076745		MAPE	0.076101
(1,1,0)	AIC	-850.1698	(2,1,0)	AIC	-848.2394
	H-Q	-847.8332		H-Q	-844.7344
	BIC	-844.4194		BIC	-839.6138
	MSE	8.8425e-005		MSE	8.8399e-005
	MAE	0.0053075		MAE	0.005309
	MPE	-0.011707		MPE	-0.012026
	MAPE	0.074140		MAPE	0.074167
(1,1,1)	AIC	-848.2102			
	H-Q	-844.7053			
	BIC	-839.5846			
	MSE	8.8409e-005			
	MAE	0.0053073			
	MPE	-0.011881			
	MAPE	0.074141			

Based on the significance of estimated parameters shown in table (4) and the values of criteria shown in table (5), ARIMA (1,1,0) model was selected to represent the data. We can also conclude from table (4) that ARIMA (1,1,0) model fulfills the stationary condition $|\alpha_1| = |0.431265| < 1$.

The third stage: Diagnostic Checking

To ensure the efficiency of ARIMA (1,1,0) model in representing the data, the model residuals were tested by calculating and plotting the autocorrelation function of the residuals as shown in figure (4) and table (6).

By looking at figure (4) and table (6), we notice that all autocorrelation cofficients of the model residuals fall within trust limits and do not significantly differ from zero, which indicates that the residuals represent the white noise.

Journal of AL-Qadisiyah for computer science and mathematics Vol.9 No.2 Year 2017
ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504

Table 6: Autocorrelation coefficients of residuals for ARIMA (1,1,0) model

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]

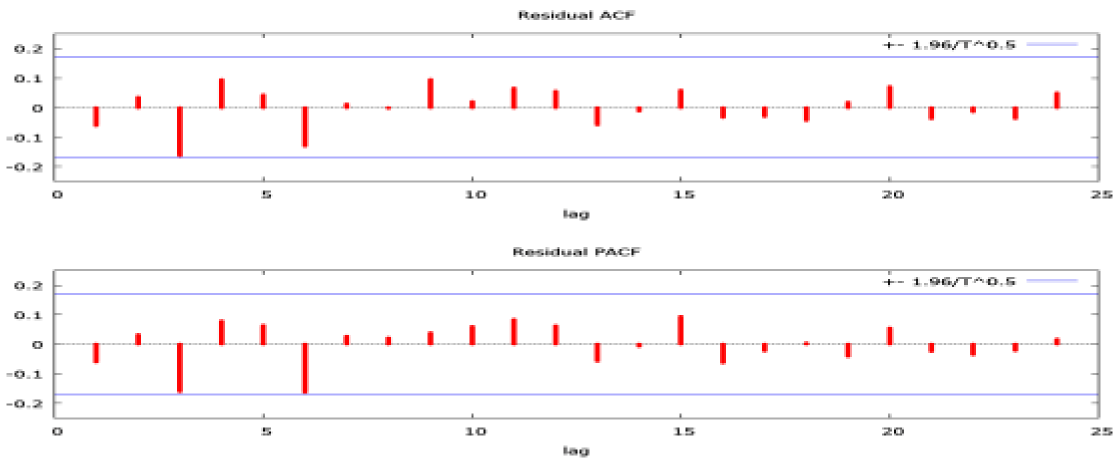


Figure 1: Correlogram of residuals for ARIMA (1,1,0) model

Journal of AL-Qadisiyah for computer science and mathematics Vol.9 No.2 Year 2017
ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504

To increase the accuracy of the results, Ljung-Box test was applied on the model residuals **as shown in table (7) below.**

Table 7: Ljung-Box test results

النموذج	df	Q_{LB}	$\chi^2(df\ ,alpha = 0.05)$	p-value
ARIMA(1,1,0)	23	15.4305	35.172	15.4305

By testing the P-value, we accept the null hypothesis and conclude that the residuals of ARIMA (1,1,0) model are completely random and represent the white noise. Therefore, ARIMA (1,1,0) model represent the best model to estimate the exchange rate

6- Artificial neural network methodology application

The main step in designing the neural network model of a specific time series is determining the number of input variables. Based on the results of Box-Jenkins methodology that showed ARIMA (1,1,0) model as the best model to represent the data, we conclude that the input variables include the Y_{t-1} only. Therefore, the number of input nodes equal to one. Because the goal here is to predict one-step-a head, one.

output node was set in the output layer, which include one variable Y_t. In addition, one hidden layer was determined for the hidden level. By choosing backpropagation algorithm to train the network, a 0.5 learning speed and 0.9 momentum were selected, and we include 100% of the data for the training due to the small sample size.

Because there is no constant rule to select the activation functions in both the hidden and output layers, 5 models including ANN(1), ANN(2), ANN(3), ANN(4), and ANN(5) were built with different activation functions as shown in table (8). Based on the activation function in the output layer of each model, the processing formula was determined. The number of hidden nodes in each models were also determined based on the try and error approach and the following formulas:

Number of hidden nodes = R

Number of hidden nodes = 2R

Number of hidden nodes = 2R +1

Where R represents the number of the input nodes.

Journal of AL-Qadisiyah for computer science and mathematics Vol.9 No.2 Year 2017
ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504

Table 8: Artificial neural networks models

First Model ANN(1)	Activation function for hidden layer				Hyperbolic tangent			
	Activation function for output layer				sigmoid			
	Data preprocessing formula				normalized			
	Number of Hidden Nodes	MSE	MAE	MAPE	Number of Hidden Nodes	MSE	MAE	MAPE
	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
Second Model ANN(2)	Activation function for hidden layer				sigmoid			
	Activation function for output layer				sigmoid			
	Data preprocessing formula				normalized			
	Number of Hidden Nodes	MSE	MAE	MAPE	Number of Hidden Nodes	MSE	MAE	MAPE
	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
Third Model ANN(3)	Activation function for hidden layer				Hyperbolic tangent			
	Activation function for output layer				Identity			
	Data preprocessing formula				normalized			
	Number of Hidden Nodes	MSE	MAE	MAPE	Number of Hidden Nodes	MSE	MAE	MAPE
	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
Forth Model ANN(4)	Activation function for hidden layer				sigmoid			
	Activation function for output layer				identity			
	Data preprocessing formula				normalized			
	Number of Hidden Nodes	MSE	MAE	MAPE	Number of Hidden Nodes	MSE	MAE	MAPE
	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
Fifth Model ANN(5)	Activation function for hidden layer				Hyperbolic tangent			
	Activation function for output layer				Hyperbolic tangent			
	Data preprocessing formula				Adjusted normalized			
	Number of Hidden Nodes	MSE	MAE	MAPE	Number of Hidden Nodes	MSE	MAE	MAPE
	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

Regardless the number of the hidden nodes, by examining table (8) and based on MSE, MAE, and MAPE criteria, the best model among the designed models is ANN (3). Therefore, we conclude that the best activation function for the hidden layer is the bipolar function and the linear function for the output layer. In addition, by testing the third model, it is apparent that the best number of nodes for the hidden layer is 2.

Therefore, we conclude that the best formula to determine the number of hidden nodes is 2R.

Based on the third model ANN (3) with 2 nodes and constant requirements of the other network except the sample size, data was divided to two sets including the training and the testing with specific portions as shown in table (9). The network was retrained again and the results are shown in table (9).

Table 9: Values of the criteria when data partitioning

Training Data	Testing Data	MSE	MAE	MAPE
100	0	173.3501	7.6115	0.5843
90	10	184.357	8.2431	0.6318
80	20	183.1978	7.2871	0.5619
70	30	180.9971	7.763	0.5966
60	40	188.8886	7.575	0.5855
50	50	179.1047	7.9445	0.6062
40	60	198.5476	9.2811	0.7166
30	70	189.0676	8.8814	0.674
20	80	204.1265	8.046	0.6266
10	90	185.3909	7.3814	0.573

By examining table (9), we conclude that the inclusion of all the data in the training leads to the lowest potential error. This is clear through the MSE criterion that reach its lowest value when 100% of the data is included in the training, which indicates a safe primary selection of the data size.

Based on the third model with 2 hidden nodes and constant network requirements except the momentum value, different models were designed with different momentum values as shown in table (10) below.

Table 10: Values of the criteria during momentum change and fixed learning rate

Learning Rate	Momentum	Time	MSE	MAE	MAPE
0.5	0.9	0:00:00.09	173.3501	7.6115	0.5843
0.5	0.8	0:00:00.08	184.3754	8.1804	.6289
0.5	0.7	0:00:00.05	171.2072	7.4626	.5748
0.5	0.6	0:00:00.05	170.5924	7.5583	.5800
0.5	0.5	0:00:00.17	188.2508	8.3374	.6406
0.5	0.4	0:00:00.37	180.5767	7.9174	.6067
0.5	0.3	0:00:00.08	173.5609	7.5861	.5816
0.5	0.2	0:00:00.03	171.6374	7.3010	.5630
0.5	0.1	0:00:00.09	197.1893	8.9972	.6925

By examining table (10), we conclude that the momentum value affects the training time and the error calculated by the network. We also conclude that the best value of momentum that gives the lowest possible error at appropriate time with 0.5 learning rate is 0.6.

Based on the last modifications of the third model of the network, a comparison was conducted between the network training without data processing and the network training with data processing by using normalized formula . The results are shown in table(11).

Table 11: Values of criteria for final model without processing

Final Model	MSE	MAE	MAPE
Without Processing	13345.0652	99.8602	7.5583
With Processing	170.5924	7.5583	.5800

By examining table (11), we conclude that data processing is a crucial step before providing the network with data. This is apparent through the values of MSE, MAE, and MAPE criteria as shown in table (11).

Therefore, the best model of the neural network that can be used in the estimation is the third model with 2 hidden nodes and 0.6 momentum at 0.5 learning rate.

By comparing the calculated values of MSE, MAE, and MAPE for both models ARIMA (1,1,0) and AAN (3) using Box-Jenkins and the neural network respectively, we conclude that the best methodology in forecasting the exchange rate is Box-Jenkin methodology as shown below.

methodology	MSE	MAE	MAPE
Box-Jenkins	156.0669	6.867803	0.005323
artificial neural networks	170.5924	7.5583	.5800

7- Conclusions:

1- The series of exchange rate of Iraqi dinar to the US dollar is non-stationary in the mean and the variance.
2- The best model in forecasting the exchange rate using Box-Jenkins methodology is ARIMA (1,1,0).
3- The best model of the artificial neural network to forecast the exchange rate using backpropagation algorithm is the network designed with one variable (Y_t), hyperbolic activation function in the hidden layer and linear activation function in the output layer, learning rate of (0.5), (0.6) momentum, and two hidden nodes in one hidden level.
4- Based on the MSE, MAE and MAPE criterion, it is apparent that Box-Jenkins methodology is better than the neural network in forecasting the exchange rate of Iraqi dinar to the US dollar.

8- Recommendations:

1- Compare the backpropagation network and the Jordan or Elman network in predicting the exchange rate.

2- Apply the hybrid methodology to predict the exchange rate. Then compare the hybrid model and the pure neural network model to choose the best.

References

1. Box,G.B.;Jenkins,G.M.&Reinsel,G.C. ;(2008); "Time Series Analysis: Forecasting and Control" John Wiely & Sons, New Jersey.
2. Hanke, J.E. & Wichern,D.W.;(2009); "Business Forecasting" PHI Learning; U.S.A.
3. Wei,W.W.;(2006);"Timè Series Analysis :Univariate and Multivariate Methods" ;United states of America; Pearson Addison Wesley.
4. Montgomery,D.C.; Jennings,C.L. & Kulahci,M.;(2015); "Introduction to time series analysis and forcasting" ;John Wiley & Sons; New Jersey.
5. Shumway, R.H. & Stoffer, D.S. ;(2011); "Time Series Analysis and Its Applications With R Examples" ;Springer; New York.
6. Wolters, J. & Kirchgassner,G.;(2007(; "Introduction to modern time series analysis" ;Springer; New York.
7. Chatfield, C.;(2004); "The analysis of time series: An introduction" ;NewYork; Chapman & Hall/CRC.
8. Garrett,L.C. & 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.
9. Agami,N.;Atiya,A.;Saleh,M.&El Shishiny, H.;(2009); Aneural network based dynamic forecasting model for Trend Impact Analysis" ;Elsevier; Vol.76; No.7; P.P: 952-962.
10. Hagan, M.T. & Demuth, H.B.;(2008); "Neural Network Design" ;Amazon; 2nd Edtion .
11. Khashei,M. & Bijari,M.;(2010); "An artificial neural network (p,d,q) model for time series forecasting" ;Elsevier; Vol. 37; No.1; P.P: 479 – 489.
12. Kulkarni, M.A.; Patil, S.; Rama, G.V. & Sen, P.N.;(2008); "Wind speed prediction using statistical regression and neural network" ;Journal of Earth System Science; Vol. 117; No.4; P.P: 457–463.
13. Meng, X.; Jia,M. & Wang,T.;(2014); "Neural network prediction of biodiesel kinematic viscosity at 313 K" ;Elsevier; Vol. 121; P.P: 133-140.
14. Oancea, B. &Ciucu, S.C.;(2014);"Time series forcasting using neural networks" ;Cornell university library; P.P:1403-1408.
15. Yixin,Z. & Zhang,J. ;(2010);"Stock Data Analysis Based on BP Neural Network" ;IEEE.
16. Zhang,G.P. & Qi,M. ;(2005); "Neural network forecasting for seasonal and trend time series" ;Elsevier; Vol. 160; No. 2; P.P: 501 – 514.
17. Zurada, J.M.;(1992); "Introduction to Artificial Neural Systems" ;west ppublishing company; United States of America.
18. 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
19. Suzuki,K. ;(2011).; "Artificial Neural Networks- Methodological Advances and Biomedical Applications" ; InTech; India.

Journal of AL-Qadisiyah for computer science and mathematics Vol.9 No.2 Year 2017
ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504

التنبؤ باستخدام اساليب الشبكات العصبية الاصطناعية وبوكس جينكنز: دراسة مقارنة

ايمان يوسف عبد عون
جامعة القادسية
كلية علوم الحاسوب وتكنولوجيا المعلومات
قسم الرياضيات
emaan.yousif1991@gmail.com

محمد حبيب الشاروط
جامعة القادسية
كلية الادارة واقتصاد
قسم الاحصاء
m.alsharood@gmail.com

المستخلص :

ان التقلبات التي تحدث في سعر الصرف ولا سيما الارتفاعات والانخفاضات المفاجئة وغير المتوقعة لها تأثير كبير على الاقتصاد القومي لأي دولة ومنها العراق. لذا فأن التنبؤ الدقيق بسعر صرف الدينار العراقي مقابل الدولار الامريكي له اثر كبير في عملية التخطيط واتخاذ القرار بالإضافة الى المحافظة على التوازن والاستقرار الاقتصادي للعراق .

يهدف هذا البحث الى المقارنة بين اسلوب بوكس جينكنز واسلوب الشبكات العصبية الاصطناعية في التنبؤ بسعر صرف الدينار العراقي مقابل الدولار الامريكي بالاعتماد على البيانات المأخوذة من البنك المركزي العراقي للفترة من 30/01/2004 الى 30/12/2014 .

وبالاعتماد على متوسط مربع الخطأ MSE ،متوسط مطلق الخطأ MAE ومتوسط مطلق الخطأ النسبي MAPE كمعايير احصائية للمفاضلة بين الاسلوبين تم التوصل الى ان اسلوب بوكس جينكنز أفضل من اسلوب الشبكة العصبية الاصطناعية في التنبؤ.