



Available online at [www.qu.edu.iq/journalcm](http://www.qu.edu.iq/journalcm)

JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS

ISSN:2521-3504(online) ISSN:2074-0204(print)



# ***Use of Exponential Holt Model and the Box-Jenkins Methodology in Predicting the Time Series of Cement Production in Sudan***

***Emad Yagoub Hamid<sup>a</sup> , Mohammed Hassan Osman Abushaba<sup>b</sup>***

<sup>a</sup> Department of Mathematics Science, University of Shagra – Eldowadmy, KSA, Email: [yagoub2020@yahoo.com](mailto:yagoub2020@yahoo.com).

<sup>b</sup> Department of Applied Statistics, University of Nile Valley, Sudan, Email: [dr.m.abushaba@hotmail.com](mailto:dr.m.abushaba@hotmail.com).

## ARTICLE INFO

### *Article history:*

Received: 03 /12/2019

Revised form: 03 /02/2020

Accepted : 04 /02/2020

Available online: 19 /02/2020

### *Keywords:*

Box-Jenkins, Holt exponential model, Time series, Industrial production.

## ABSTRACT

This study examines the prediction in time series of industrial production, showing how the accuracy of predictions obtained by the statistical models used in the study, and the ability of these models to deal with changes in time series data and the changes on the performance of these models. The study aimed mainly to find the most suitable and best statistical forecasting models to be used in the process of forecasting cement production in Sudan. Using the Box-Jenkins methodology and the exponential smoothing method (Holt model), a number of models were constructed for the study sample of the time series of cement production in Sudan. The results of the study showed that the Box-Jenkins model outperformed the Holt model, where the comparative criteria indicators indicated the preference of the ARIMA model (2,1,1). Standards. The results showed that the results of future predictions obtained using the selected model ARIMA (2,1,1) are all within the limits of confidence, which indicates the quality and accuracy of this model in the prediction. Prediction values have indicated an increase in productivity over the next 10 years. Random changes were the most important factors that directly affected the performance of both models, but they were more effective in reducing the efficiency of the Holt exponential model compared to the Box-Jenkins model.

MSC :

DOI : 10.29304/jqcm.2020.12.1.666

## 1 . Introduction

Prediction is a major goal when studying and analyzing time series data for different phenomena. There are many statistical methods and models that are used in studying time series. These models are used to analyze changes in time series data over a certain period of time, analyze the causes of these changes and find out the nature and degree of its effect and then predict the future of these phenomena. These statistical models differ in their performance and predictive ability, and depend on the quality of their performance and the accuracy of their results on a number of factors such as the quality and nature of the data under study and the changes that affect these data, and there are some other conditions and standards that must be available before using some of these models.

Corresponding author Emad Yagoub Hamid

Email addresses: [yagoub2020@yahoo.com](mailto:yagoub2020@yahoo.com)

Communicated by Qusuay Hatim Egaar

In order to obtain good results and accurate predictions when analyzing time series data, attention should be given to the model selection stage and to experimenting with a number of statistical tools and models to obtain the appropriate model along with the studied phenomenon data.

Exponential preamble and Box-Jenkins models are statistical methods that are frequently used in the construction of time-series prediction models, the two models have characteristics if some data standard is available, and the reason of using these models is for its methodology and the way of dealing with these data to choose the right model, which often leads to good results and accurate predictions.

The two methods differ in simplicity, complexity and methodology in constructing prediction models, varying in their ease of application and the degree of accuracy of their results. Each method is compatible with certain types of data and produces more accurate results than those according to the nature of these data. Box-Jenkins models follow an integrated methodology that goes through a number of stages to reach the best model. These stages help to raise the efficiency of the selected models and thus the accuracy of the predictions obtained. But the exponential preamble models are experimental models where different values of preamble parameters are experimented at the model selection stage to reach the best model, and it is used to reduce the changes in series data, it gives high efficiency with some types of data that is not suffer form changes and sharp fluctuations.

These methods are frequently used in the analysis of time series of industrial production of these advantages, because the industrial production sector data is effected with some changes and factors which need special care when analyzing and choosing the appropriate models to predict with.

The industrial production sector is considered to be one of the most important economic sectors that have an effective influence on the development of the national economy .This sector has a great deal of interest in building prediction models as these models play an important and effective role in knowing the reality and future of this sector.

The cement industry in Sudan is one of the successive industries in industrial sector and has a considerable effect in national income. The analysis of the reality of this industry and the production in future gives image about the ability of local factories to cover the high demand of cement, and ability to increase its production in future for exporting.

## **2. The importance of the study**

Because of the economic importance of the cement industry in supporting the Sudanese economy and the effective role of statistical prediction methods and models in analyzing reality and planning for the future.

## **3. The problem of the study**

The accuracy of the predictions obtained when analyzing the time series is directly affected by the nature and quality of the data and the changes that affect this data as well as the statistical model used to predict these data. Therefore, the selection of the appropriate statistical tool with the nature and quality of data and attention to the changes that affect this data is one of the most important stages of time series analysis. Through this study we want to identify the following questions:

How can exponential boot models and Box Jenkins methodology be used to predict time series? What are the steps to be followed and the conditions that must be taken to reach the appropriate results and accurate predictions when using these models? What depends on the predictive accuracy of these models? How does the nature and quality of data influence the performance of these models? How well do these models handle changes in time series data?

How well are the Box-Jenkins and exponential models in the time series analysis of industrial production appropriate? What is the most accurate and accurate model for predicting time series data for cement production in Sudan?

#### 4. Objectives of the study

The main objectives of this study are:

Highlight the role and importance of statistical methods of forecasting in future policy-making.

Recognize the predictive power and accuracy of Box-Jenkins and exponential models in predicting time series of industrial production.

Predict the quantities of cement produced in the coming period and provide a statistical reference on the future of cement production in Sudan.

Understand the advantages and disadvantages of Box-Jenkins and Asian boot models in dealing with time series of industrial production.

Comparison of Box-Jenkins models with boot models and knowledge of agreements and differences in time series prediction.

#### 5. Study hypotheses

In order to answer the study questions, the following hypotheses will be tested:

The accuracy of prediction results when analyzing time series data is influenced by the statistical model used to predict.

The accuracy of the prediction results is influenced by the nature and quality of the time series data under study.

Changes in time series data have a direct effect on the performance of Box-Jenkins and exponential models.

Box-Jenkins models are more efficient than exponential models in predicting time series data for cement production in Sudan.

#### 6. Methodology

To achieve the objectives of this study and verify its hypotheses, the analytical statistical approach will be followed. A number of prediction models are constructed using the exponential preliminary method and the Box-Jenkins methodology for time series data under study. First, a graph of the time series data is drawn to see the shape of the series and to detect the changes that affect the series and whether the series is static, has a general trend or is unstable in contrast. Using the Box-Jenkins methodology, a number of ARIMA models will be constructed, including the selection of the appropriate model with data. Using the exponential preamble method, a number of exponential preamble models will be piloted to reach the best model that can predict string data. In the last stage, the selected models are compared using the two methods to reach the most efficient model that can predict the production of cement in Sudan.

In order to analyze the studied time series data and build the required models, some ready statistical packages such as (SPSS25, Minitab18.1, Matlab.R2018b ,Eviews10).

#### 7. Previous Studies

1- Tabboush El-Elja study (2018): The effectiveness of the Box Jenkins and Holt Winter methods in predicting the sales of the National Electricity and Gas Corporation (Sonelgaz) Tissemsilt branch. This study aimed to identify the predictive power of the Box-Jenkins and Holt Winter methods in predicting the sales of the National Electricity and Gas Corporation "Sonelgaz" and then compare the two previous methods, to determine which is better and more accurate.

The study found that the Box Jenkins method is the best and closest to the predictive compromise values for sales of the National Electricity and Gas Corporation.

2-Adnan Karim and Hussein Ali Hashem (2018: The use of some statistical methods for the purposes of predicting lost electrical energy in Karbala province. This study aimed to predict the loss of electrical energy by selecting the best model for prediction through the models of Box Jenkins.

The study found that there is a clear trend in the series as well as containing the seasonal vehicle as it repeats itself every 12 months. The study concluded that the efficient and appropriate model for the representation of time series

data is the seasonal multiplier model SARIMA (0,1,1) (0,1,1) 12, depending on the differentiation criteria (AIC, BIC, H-Q).

3 - Amal Gilani study (2017) : The use of the preliminary predication to predict the production and consumption of electricity in Sudan (1989-2015 m). This study aimed to build an accurate scientific statistical model for the production and consumption of electricity in Sudan and then use it in the prediction process.

The study found that the exponential model is the best among the models used in the study to achieve moral significance. The study also showed that the growth rates were convergent, ie, what is produced is consumed.

4 - Mohammed Omar (2015) : Prediction of quantities of crude oil production in Libya using the specified models (exponential models) during the period 1972 - 2013. The aim of this study was to introduce statistical prediction models (Smoothing Models) and to clarify the steps necessary to do them. It also aimed to compare statistical models with each other, in order to choose the best to predict oil production in Libya during the period 2014-2024.

The study reached a number of results including:

The Holt model is the best model to represent the data of the oil production chain in Libya, giving the lowest value of the average error box (MSE). Based on the predictions reached by the research shows that the production of Libyan crude oil will see a decline in the future, and this decline in production due to the political situation experienced by Libya since 2011.

5- Mohammedin Abkar Zakaria study (March 2012): Estimation of the Demand Function in Sudan 2008-1990. The study aims to indicate that the level of demand encourages increased production and the establishment of new factories in the country, and that the growing urban growth in the country, led to significant changes in the price and the trend to import from abroad.

The study concluded that the explanatory power of the proposed model reached 88% according to the modified determination coefficient, which indicates the quality of harmonizing the model.

6 - Abeer Hassan Al-Jubouri (2010): This study aimed at predicting oil prices and their impact on the global financial crisis. One of the most important results of the study is that the double exponential preamble model (Holt method) achieved better predictive accuracy than the Box-Jenkins model and therefore the Holt model is better than the Box-Jenkins model of the data of that study.

## 7. Theoretical Background:

### Cement industry in Sudan

Cement industry is one of the leading industries in the map of the Sudanese industry, has been linked to the establishment of the first development project in the field of energy and irrigation in Sudan (Sennar Damp), where a cement plant was built specifically to meet the needs of building this Damp. Then the establishment of Atbara Cement Factory in 1947 with a production capacity of (400) thousand tons, then Rabak Cement Factory in 1964 with a production capacity (100) thousand tons. The productivity of these factories did not meet the needs of the country, and covered the shortage of imports, which amounted to about (85%) of the needs. (Ministry of Industry, 2009).

To replenish this shortage, a number of factories were established in cooperation with the private sector: Al Salam Cement Company October 2008, Atbara New Factory February 2010, Berber Factory in May 2010, Al Takamol Factory August 2010 and Al Shamal Cement Factory December 2010. Then Nile Cement Factory (RBC) rehabilitated. These new factories have contributed to raising the productivity of this essential commodity and now cover nearly 70% of the local needs, and there are ongoing efforts by these factories to raise production capacity and meet the high demand of this commodity. All this contributes to raising the of industrial sector to the country's GDP by 23%. Where the cement industry is one of the most important products of the industrial sector.

This clear growth and significant effect of this industry leads us to pay more attention to planning and statistical analysis tools in this sector in order to analyze reality and read the future.

## 8.Exponential Smoothing Models

Exponential smoothing is a statistical procedure that dealing with confusion or random errors in time series, and can be defined as a process of polishing or smoothing data in which there is confusion, and it is a kind of estimation process that proved successful by studying the cases that depend or change with time. Exponential smoothing methods are important methods for estimating time series (Alaa 2013).

Preface is meant to try to minimize changes in string values around the curve line representing the overall direction of the chain. The exponential preamble is a set of experimental methods in which different values of the preamble coefficients are experimented and the coefficients are chosen that keep the sum of the squared deviations of the estimated values from the real values to a minimum.

Exponential smoothing methods are divided into the following types:

Simple Exponential Smoothing

Double Exponential Smoothing

Triple-Exponential Smoothing

If the time series does not contain a general trend or seasonal variations, simple exponential smoothing models are preferred in the prediction process. If the series has a general trend, it is preferable to use double exponential boot patterns. If the series contains a general trend and seasonal variations, it is preferable to use triple exponential models.

## 9.Exponential Smoothing Models

### Firstly: Simple Exponential Smoothing

This model is used to predict when time series data do not have a significant general trend or recurring seasonal pattern. This form is written as follows:

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha) \hat{y}_t \rightarrow (1)$$

value at time  $t + 1$ .  $\hat{y}_t$  : estimated value at time  $t$ .  $y_t$  : Real time value at time  $t$ , ( $t = 0, 1, 2, \dots, n$ ).

The value of the smoothing factor in the simple exponential smoothing is determined empirically, where a value is chosen that keeps the sum of squares of error deviations to a minimum.

The preamble coefficient value plays an important role in the final prediction amount and also in reducing the error value of the model. (Montgomery & Lynwood, 1976)

In general, you choose a small value for  $\alpha$  when data is highly volatile, and a higher value is chosen when the data is almost stable or changes are not severe. Brown has suggested a special range for  $\alpha$  ranging of ( $0.1 \leq \alpha \leq 0.3$ ).

### Secondly: Double Exponential Smoothing

A double exponential smoothing method is a generalization of a simple exponential boot method if there is a general linear trend in the data series. We pave the original data according to the general trend and then the data obtained in the first stage. Its methods include:

Brown Linear Model: This model is expressed in the following relationships

$$\hat{y}_{t+1} = a_t + b_t \rightarrow (2)$$

$$a_t = 2s'_t - s''_t \rightarrow (3)$$

$$b_t = \frac{\alpha}{1 - \alpha} (s'_t - s''_t) \rightarrow (4)$$

$$s'_t = \alpha y_t + (1 - \alpha) s'_{t-1} \rightarrow (5)$$

$$s''_t = \alpha s'_t + (1 - \alpha) s''_{t-1} \rightarrow (6)$$

where at , bt: parameters of the model,  $s'_t, s''_t$  : represents the value of a single and double exponential preamble at time t.

Holt's Linear Model: This method is used if time series data have a general linear composite direction, and this method is characterized by the  $\alpha$  and  $\gamma$  coefficients (each ranging from zero to one) .The Holt model consists of an exponential smoothing component  $S_t$  and a directional component  $b_t$  .Holt's linear model is given by the following relationships:

$$s_t = \alpha y_t + (1 - \alpha)(s_{t-1} + b_{t-1}) \rightarrow (7)$$

$$b_t = \gamma(s_t - s_{t-1}) + (1 - \gamma)b_{t-1} \rightarrow (8)$$

$$\hat{y}_{t+1} = s_t + b_t \rightarrow (9)$$

**Where:**

$S_t$ : The time value of the time series,  $b_t$  :The value of linear trend for time series data.

$\alpha$ : constant part of smooth level,  $\gamma$ : constant smooth of the general trend.

To use this model, you must specify the value of ( $\alpha, \gamma, s_0, b_0$ ), they are selected and tested experimentally in order to achieve the least possible deviations.

### Thirdly: Triple-Exponential Smoothing

This method is a generalization of the simple exponential smooth method if the string contains a nonlinear general trend and seasonal variations, where the original string data is booted three times in succession: general boot, directional boot, and seasonal smooth.

**10-Brown Quadratic Model:** The Brown Quadratic Model includes a single preamble coefficient used to straighten data (Brown, 1963)

Winter's Model: This model differs from the Brown model in that it includes three types of boot parameters, a general boot parameter, a directional boot parameter and a seasonal boot factor.

### 11.Box-Jenkins Models

This method was developed by the scientist George E.P. Box and Gwilym M. Jenkins in 1930 became known by their names. This method of dealing with time series data is based on Self-Regression (AR), Moving Average (MA) and ARMA models, which combine AR and MA models.

This method of analyzing time series data to build the appropriate models of this data depends on the fragmentation of the series into several components by exposing the series data on a number of stages called refinement coefficients, namely: Sleep Filter, Self Regression Filter and Moving Averages Filter. These filters filter data from time series changes to remain purely random and unpredictable.

#### 11-1 Autoregressive Models (AR)

When the current value of a string ( $y_t$ ) is a function of its value in previous periods ( $y_{t-1}, y_{t-2}, \dots, y_{t-n}$ ) and in addition to some errors, the models formed by this process are called self-regression models (Robert, 2000).

The p-slope models are referred to as AR (p) and are expressed as follows:

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + a_t \rightarrow (10)$$

Where  $(\phi_1, \phi_2, \dots, \phi_p)$ : Parameters of the autistic regression model,  $a_t$ : a random changes,  $\delta$ : form constant.

### 11-2 Moving Average Models (MA)

If the current value ( $y_t$ ) of the time series can be expressed as a function of the current random change ( $a_t$ ) and the previous random changes ( $a_{t-1}, a_{t-2}, \dots$ ), the resulting model is called the moving averages model.

The models of moving averages  $q$  are referred to as MA ( $q$ ) and are expressed as follows:

$$y_t = \delta + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \rightarrow (11)$$

Where  $(\theta_1, \theta_2, \dots, \theta_q)$  model parameters,  $(a_{t-1}, a_{t-2}, \dots, a_{t-q})$ : Random variables,  $\delta$ : The cut part.

### 11-3 Autoregressive Moving Average Models (ARMA)

The common model is ARMA ( $p, q$ ), the self-regression model - mixed moving average of grade ( $p, q$ ), where  $p$  refers to the number of autistic regression parameters in the model, and  $q$  refers to the number of parameters of moving averages. This form is expressed as follows:

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \rightarrow (12)$$

$$\phi_p(B)y_t = \delta + \theta_q(B)a_t$$

### 11-4 Autoregressive Integrated Moving Average Models (ARIMA)

Common models ARMA ( $p, q$ ) after analgesia are referred to as integrated models and are indicated by ARIMA ( $p, d, q$ ), where  $d$  indicates the number of differences needed to achieve sleep. The ARIMA formula ( $p, d, q$ ) is written as follows: (George E.E.P. Box, Gwilym M. Jenkins, 1994).

$$\phi_p(B)w_t = \delta + \theta_q(B)a_t \rightarrow (13)$$

Where:

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

$$W_t = \nabla^d y_t = (1 - B)^d y_t$$

## 12. Building Steps of Box - Jenkins Models

The Box-Jenkins methodology for building time series models consists of the following phases:

Firstly: Identification stage: This stage consists of two main steps are:

Model Recognition: The stability of the chain is checked and the necessary transformations are applied to make it stable if not. Then the model type (AR, MA, ARMA) is identified). PACF). (Box & Jenkins, 1994).

Model rank: ( $p, q$ ) in non-seasonal models and ( $P, Q$ ) in seasonal models, using the theoretical properties of ACF and PACF, along with some ranking measures, such as the self-notification standard (AIC) and the Schwartz standard Albizi (SIB). (Where ACF and PACF have approximate graphs as shown in the table below:

**Table (1): Theoretical properties of ACF and PACF for some ARIMA models**

PACF	ACF	Sample
$\phi_{kk} = 0$ , $k > p$ Cut after p of displacements	Decreasing towards zero in exponential form And / or wobbling	AR(p)
Decreasing towards zero in exponential and / or sinusoidal form	$\rho_k = 0$ , $k > q$ Cut after q of displacements	MA(q)
Decreasing towards zero exponentially And / or fluctuation after the period p-q	Decreases to zero exponentially and / or fluctuates After the period q-p	ARIMA(p,0,q)

Source: Robert, Opcit, p148

Second: Estimation Stage: After identifying a specific model or a set of models that may fit the data of the series under study, we find estimated values for the parameters of these models.

There are a range of estimation methods that are used to find parameter values for ARIMA models, such as the least-squares method, the greatest ability method, and other estimation methods.

Third: Diagnostic Checking Stage: In this stage, the validity of the selected models is checked and their suitability for the series data is subjected to a number of statistical tests. If the model passes these tests, it is usable. There are a number of tests used at this stage, the most important of them are:

1. Residuals analysis: Residual analysis of the selected model is analyzed by some tests to verify that it meets the conditions imposed on random changes in ARIMA models. We perform the following tests on the remainder:  
- Make sure there is no self-correlation, done in two ways: using the Ljung-Box test. Or depending on the confidence limits of the self-correlation and partial self-correlation parameters of the residues, if all the estimated values of the self-correlation and partial self-correlation parameters of the residues fall within a 95% confidence interval then the model is representative of the time series data.

- Make sure that the residuals are normal distributed, this is done by using many methods: Goodness of Fit Test and Kolmogorov- Smirnov Test.

- Deleting of parameters: to find out the unnecessary ones in the model, to study the standard error for parameters and to study the sample correlation of these parameters.

- Adding of parameters: to know if the model containing suitable number of parameters and to study the improvement of this addition.

Comparison Criterias: there are some standards is used to distinguish between the models, after passing all previous tests , to choose the best models some criterias are calculated, the model that have a minimum value is to be chosen.

The common Criterias are:

Automatic Information Criteria (AIC) is sometimes called the Akaike's Information Criteria and is calculated from the following formula:

$$AIC(k) = nLn\sigma_a^2 + 2K \rightarrow (14)$$

where: K: the number of estimated parameters in the form,  $\sigma_a^2$ : the variance of the error, n: the number of views.

- Schwartz Bayesian Criteria (SBC): Calculated from the following formula:

$$SBC = nL_n\delta_a^2 + KLn(n) \rightarrow (15)$$



**Fourth: Forecasting Stage:** After identifying the appropriate models and estimating their parameters and checking their validity, the final model chosen is used to generate future predictions.

- Prediction error: Prediction error is defined as the difference between the predicted value and the value seen. The prediction error is used to calculate a number of criteria to measure the quality of the prediction.

Prediction accuracy criteria: These criteria are used to assess the predictive power of the estimated model. The smaller the result of these criteria, the more predictable the model will be. These include: Mean Square Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE).

### 13. Applied study

This section deals with all the practical analysis of the time series data for cement production in Sudan. The most suitable for the application.

### 14. Description of the study data:

The data of this study are annual data for cement production in Sudan during the period from 1992 to 2017. These data were obtained from the reports of the concerned factories in addition to the annual reports of the Bank of Sudan.

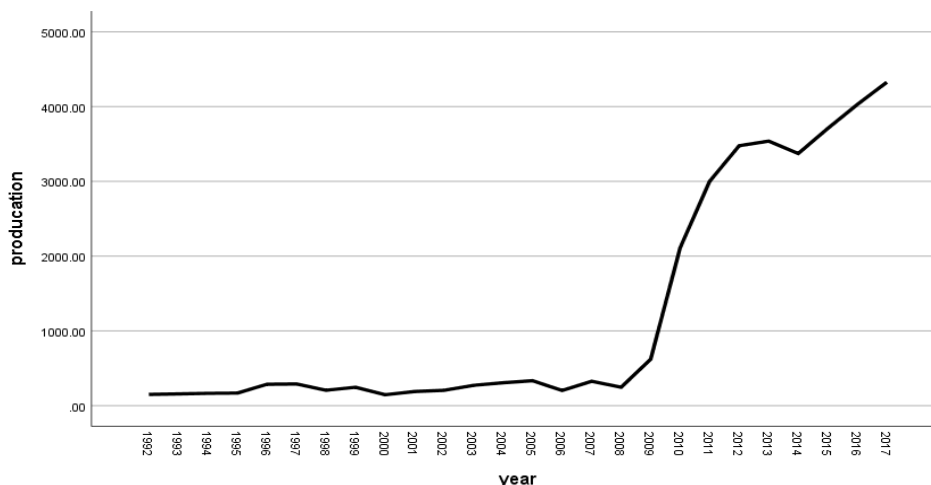
**Table (2): Descriptive statistics of the study data**

N	Minimum	Maximum	Mean	Std. Deviation
Production thousand tons / year 26	145.80	4326.10	1233.8538	1547.99378

The source is prepared by the researcher using the SPSS program.

Table (2) shows the descriptive statistics of these data, where the average production of cement during the said period (1233.8538) thousand tons with a standard deviation (1547.99378). The lowest productivity (145.80) thousand tons / year and the highest productivity (4326.10) thousand tons / year.

Figure (1): Annual cement production in Sudan during the period (1992 - 2017).



The source is prepared by the researcher using the SPSS program.

Figure (1) shows the time curve of the data series during the study years. That indicates that the series is unstable without general trend, with some random frequencies, and without any cyclic steady changes.

**Build and analyze Box-Jenkins models:**

- Time series stability check:

In order to ensure the stability of the chain we perform a test (Dickey -fuller) or what is known as testing the root of the unit on the time series data to produce cement.

Table (3): Dickey -fuller Test Results

Test	Model	Test Statistic	5% Critical Value	p-value
D.F	Cement production	0.223528	2.991878	0.9227

Source prepared by the researcher using the program Eviews Null Hypothesis: sugar has unit root.

From Table (3) we note that the value of p-value = 0.9227, which is greater than the statistical significance level 0.05, this means acceptance of the null hypothesis which provides for the existence of the root of the unit. We conclude that the time series is unstable.

**- Model recognition stage:**

At this stage, we study the ACF and PACF functions to determine the chain dormancy and to identify the appropriate models for the study data. Box-Ljung Q is used to check the chain dormancy.

Table (4): Self-correlation and partial self-correlation functions of cement production.

Autocorrelation .6	Partial .5 Correlation	ACF .4	PACF .3	Q-Stat .2	Prob .1
.  *****  .12	.  *****  .11	0.879 .10	0.879 .9	22.484 .8	0.000 .7
.  ****  .18	. *  .  .17	0.733 .16	-0.173 .15	38.771 .14	0.000 .13
.  ****  .24	. *  .  .23	0.586 .22	-0.077 .21	49.651 .20	0.000 .19
.  ***  .30	.   .  .29	0.452 .28	-0.038 .27	56.414 .26	0.000 .25
.  **  .36	. *  .  .35	0.309 .34	-0.141 .33	59.733 .32	0.000 .31
.  *  .42	. *  .  .41	0.156 .40	-0.152 .39	60.618 .38	0.000 .37
.   .  .48	. *  .  .47	0.010 .46	-0.088 .45	60.621 .44	0.000 .43
. *  .  .54	.   .  .53	-0.101 .52	0.016 .51	61.032 .50	0.000 .49
. *  .  .60	.  *  .  .59	-0.141 .58	0.179 .57	61.883 .56	0.000 .55
. *  .  .66	.   .  .65	-0.161 .64	-0.026 .63	63.069 .62	0.000 .61
. *  .  .72	. *  .  .71	-0.184 .70	-0.081 .69	64.717 .68	0.000 .67
. *  .  .78	.   .  .77	-0.199 .76	-0.021 .75	66.768 .74	0.000 .73

Significant if |Correlation|> 0.285714

The source is prepared by the researcher using the SPSS program.

In Table (4) of the correlation coefficients for ACF, PACF we find that: The values of the ACF coefficients are mostly significant, where they placed outside the confidence limits, and the first coefficient of the PACF placed outside the confidence limits. The value of the Q test and the associated significant value indicate the significance of the ACF and PACF coefficients, and we conclude that the series is unstable. In order to convert it to a stable chain, the first difference was taken and the ACF, PACF and Q test were recalculated as shown below:

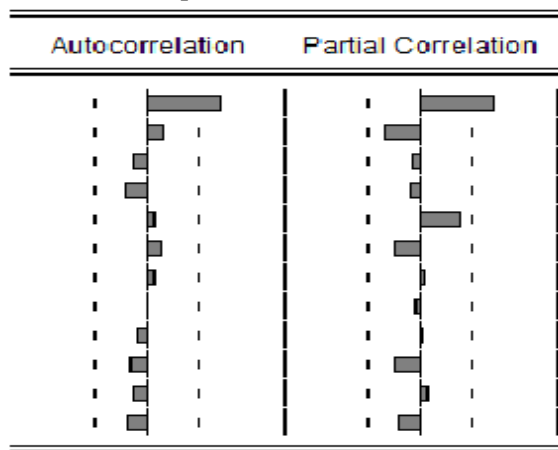
Table (5): Self-correlation and partial self-correlation functions of the first series of cement production

Autocorrelation .84	Partial .83 Correlation	ACF .82	PACF .81	Q-Stat .80	Prob .79
.  ****  .90	.  ****  .89	0.560 .88	0.560 .87	8.8068 .86	0.003 .85
.  *  .96	. **  .  .95	0.123 .94	-0.277 .93	9.2474 .92	0.010 .91
. *  .  .102	.   .  .101	-0.101 .100	-0.052 .99	9.5590 .98	0.023 .97
. *  .  .108	. *  .  .107	-0.172 .106	-0.067 .105	10.508 .104	0.033 .103
.   .  .114	.  **  .  .113	0.055 .112	0.298 .111	10.612 .110	0.060 .109
.  *  .  .120	. *  .  .119	0.100 .118	-0.191 .117	10.967 .116	0.089 .115

.   .   .126	.   .   .125	0.044 .124	0.034 .123	11.039 .122	0.137 .121
.   .   .132	.   .   .131	-0.003 .130	-0.025 .129	11.039 .128	0.199 .127
. *  .   .138	.   .   .137	-0.074 .136	0.022 .135	11.268 .134	0.258 .133
. *  .   .144	. *  .   .143	-0.127 .142	-0.192 .141	11.990 .140	0.286 .139
. *  .   .150	.   .   .149	-0.113 .148	0.058 .147	12.606 .146	0.320 .145
. *  .   .156	. *  .   .155	-0.143 .154	-0.169 .153	13.670 .152	0.322 .151

Significant if |Correlation|> 0.285714 Source: Output of the program E-views.

Figure (2): the functions of self-correlation and partial self-correlation of the first series of cement production.



Source: Output of the program E-views.

Table (6): Dickey and fuller test results after the first difference

Test	Model	Test Statistic	5% Critical Value	p-value
D.F	Cement production	-2.165655	-1.955681	0.0318

Significant if |Correlation|> 0.285714 Source: Output of the program E-views

From table (6) we find the value of p-value = 0.0318, which is smaller than the level of statistical significance 0.05, so we reject the null hypothesis which states that the root of the unit, which means that the time series became static at this stage.

As can be seen from table (5) of the functions (ACF) and (PACF), which were calculated after taking the first difference, that most of the coefficients are not significant and fall within the limits of confidence, and the value of the test Q and the associated moral value indicates the non-significant meaning Is equal to zero, and we deduce from this the stability of the chain after taking the first difference.

In Table 5 and Figure 2 we can filter the ARIMA model (2,1,1) to represent the time series of cement production. Also, some other models of statistical significance from ARIMA models will be nominated in preparation for examination and comparison between them and the proposed model.

**- Model parameters estimation stage:**

At this stage we will estimate the parameters of the proposed models and calculate some statistical criteria to differentiate between them, and it is the chosen model that gives the best match of these criteria.

Table (7): Results of the differentiation criteria for the possible models

NO .163	ARIMA .162	MSE .161	RMSE .160	MAPE .159	BIC .158	AIC .157
1 .170	(1.1.1) .169	168. 87937.04	282.829 .167	42.678 .166	11.805 .165	164. 347.9804
2 .177	(1.1.2) .176	175. 81357.58	272.055 .174	42.430 .173	11.856 .172	171. 349.2266
3 .184	(2.1.1) .183	182. 74670.83	269.593 .181	38.687 .180	11.838 .179	178. 347.9555

4 .191	(2.1.2) .190	189. 74747.48	276.550 .188	39.756 .187	12.015 .186	185. 349.9369
--------	--------------	---------------	--------------	-------------	-------------	---------------

The source is prepared by the researcher software using SPSS, MATLAB, e-views

Note that from Table7, the optimal model that expresses the quantities produced by cement is ARIMA (2,1,1). This model obtained the lowest values of the standards (RMSE, AIC, MAPE, MSE). These metrics are used to calculate the magnitude of prediction errors and the best model is the one with the lowest magnitude of prediction errors and the lowest values of information criteria.

The table below gives the capabilities of the selected model ARIMA (2,1,1):

Table (8): ARIMA (2,1,1) estimation results for the cement production chain

ARIMA Model Parameters						
		Estimate	SE	t	Sig.	
production- Model_1	Constant	-40572.719	13908.205	-2.917	0.009	
	AR	Lag 1	1.278	0.208	6.135	0.000
		Lag 2	-0.583	0.178	-3.275	0.004
	Difference	1				
	MA	Lag 1	0.994	4.091	0.243	0.810
	Numerator	Lag 0	20.319	6.936	2.929	0.008

Source: Prepared by the researcher based on the results of the program spss

Depending on the results of Table (8), the estimated model can be written as follows:

$$y_t = -40572.719 + 1.278y_{t-1} - 0.583y_{t-2} + a_t - 0.994a_{t-1} \rightarrow (16)$$

**- Model diagnosis and testing stage:**

At this stage, we are testing the accuracy of the nominated model to determine its suitability to represent time series data and its usability to predict. This is done by studying the random errors (residues) of the model and making sure that they are random and there is no self-correlation between them and they are distributed normal distribution.

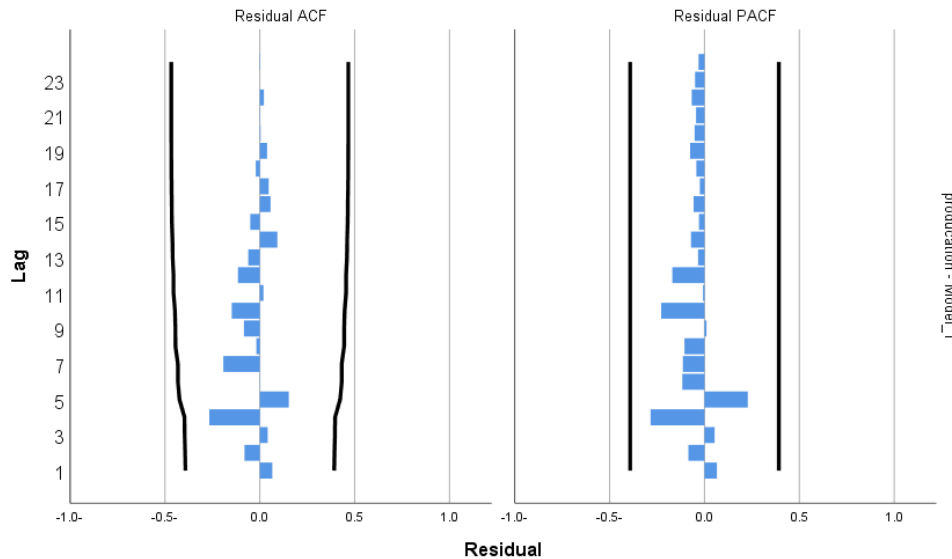
Table 9: Ljung-Box Q test results for ARIMA (2,1,1) series of cement productivity series.

192. Test Ljung-Box (Q)						
199. NULL Rejected	198. P-value	197. Test statistics	196. Critical Value	195. Lag	194. DF	193. Significance level
206. false	205. 0.9858	204. 8.7281	203. 31.4104	202. 20	201. 20	200. 0.05

Source: Prepared by the researcher based on the results of the program matlab.

From Table (9) we find that the probability value associated with the Ljung-Box Q test is 0.9858 and this value is greater than the statistical significance level 0.05, which leads us to accept the zero hypothesis which provides for the independence of random variables, i.e., there is no subjective correlation between errors and it is random. This is confirmed by the ACF and PACF functions of the residues where all of their coefficients fall within the confines of confidence, as shown in Figure 3.

Figure (3): The functions of self-correlation and partial self-correlation of residues



Source: Prepared by the researcher based on the results of the program SPSS

To ascertain the nature of the probability distribution of the remains, we use the non-parameter test Kulmogorov-Smirnov, where we test the null hypothesis that the data follow the normal distribution.

Table 10: Kolmogorov-Smirnov Test Results

	Tests of Normality					
	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
residual from production -Model_1	.173	25	0.052	.903	25	.021

Source: Prepared by the researcher based on the results of the program SPSS

Table (10) shows that the residues are subject to normal distribution, that the probability value is 0.052 which is greater than significant  $\alpha = 0.05$  so we accept the zero hypnosis .

Table (11): Random Test Results

Noise residual from production-Model_1	
Test Value a	12.72
Z	0.008
Asymp. Sig. (2-tailed)	0.993

The source is prepared by the researcher using the SPSS program.

The Run test is used to check whether the residues are random .From Table (11) we find that the test is not significant since the value, which is greater than this leads us to accept the hypothesis of non-randomness residuals. From the above results, we conclude that the proposed ARIMA model (2,1,1) is the most suitable for the cement production chain data in Sudan and can be used in forecasting.

**- Forecasting phase:**

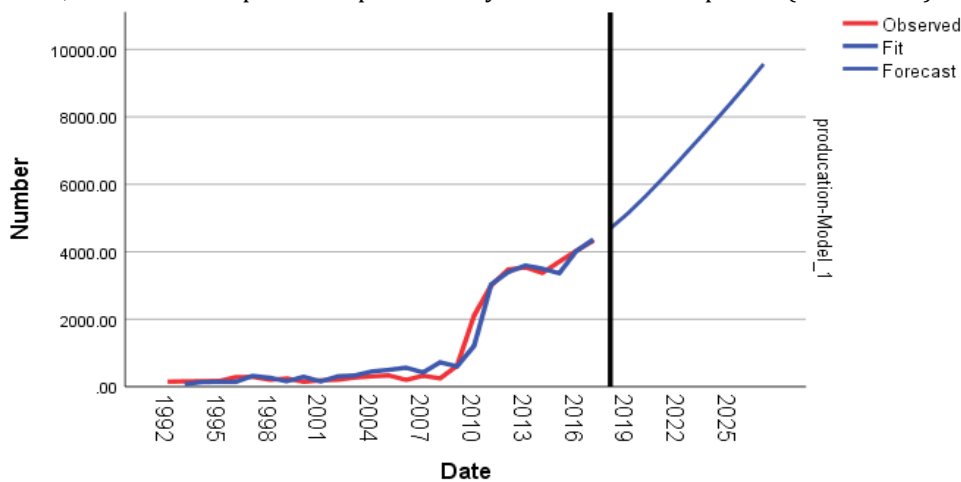
After passing the ARIMA (2,1,1) diagnostic tests, it can be used to predict the future values of the cement production chain .Table 10 represents the predictive values for 2018-2027.

Table (12): Predicted cement production for the period (2018-2027) using ARIMA model (2,1,1)

Model	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027
Forecast	4692.54	5127.16	5615.29	6138.22	6680.65	7233.89	7795.80	8368.68	8956.71	9563.93
UCL	5248.55	6045.03	6726.47	6726.47	7865.63	8416.40	8979.52	9553.95	10141.17	10746.80
LCL	4136.52	4209.30	4504.11	4960.98	5495.67	6051.39	6612.09	7183.41	7772.25	8381.06

The source is prepared by the researcher using the SPSS program.

Figure (4): The actual, estimated and predicted productivity of cement for the period (2018-2027) using the model



Source: Prepared by the researcher by the program spss

Table (12) gives the predicted values for the next ten years, and Figure (4) represents the actual, estimated and predicted values using ARIMA model (2,1,1) for the cement production chain. Increasing trend in production in the coming years. As shown in the figure that there is a convergence between the estimated series and the actual series, and we conclude from the quality of this model in its representation of the study data.

**Building and analyzing of the exponential Holt model:**

Depending on the nature of the time series data in question, the appropriate exponential model is the Holt model, because there is a general trend in the data and the data are annual and there are no seasonal changes. The results and analysis of this model are as follows:

-Model parameters:

Table (13): Results of Estimating Holt Exponential Model  
Exponential Smoothing Model Parameters

Model		Estimate	SE	t	Sig.	
production-Model_1	No Transformation	Alpha (Level)	1.000	.204	4.897	.000
		Gamma (Trend)	.999	.457	2.186	.039

Source: From the outputs of the e-views program

Table (13) above gives the estimated values of the model parameters, and the t-statistic for measuring the significance of these parameters .The probability values (sig) associated with the test statistic are all below the

statistical significance level (0.05), which means the significance of these parameters .From Table 12, the estimated model can be written as follows:

$$\hat{y}_{t+1} = y_t + 0.99(l_t - l_{t-1})y_{t-2} + 0.1b_{t-1} \rightarrow (17)$$

**-Morality of the model:**

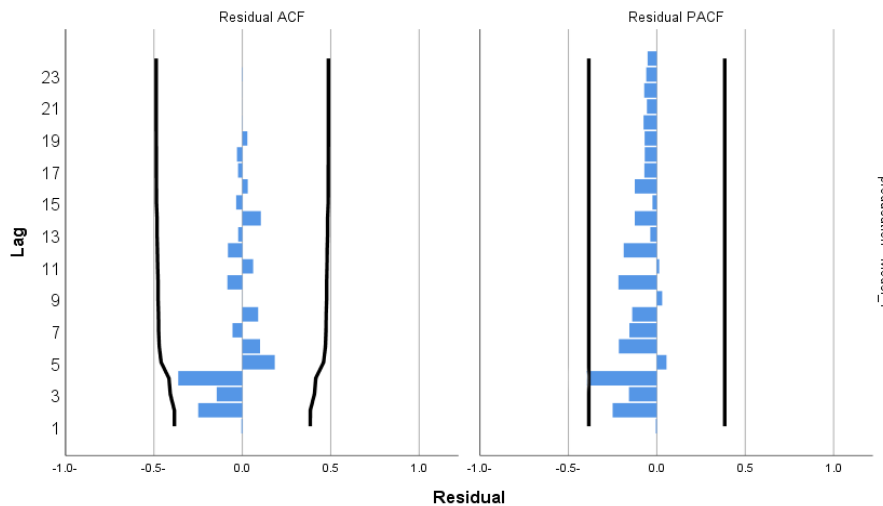
Table (14): Measurements related to predictive accuracy tests

Model	Ljung-Box Q(18)		
	Statistics	DF	Sig.
production-Model_1	10.702	16	.827

Source: From the outputs of the e-views program

Table (14) gives the statistical Ljung-Box Q, and the probability value associated with the test 0.827, which is greater than the level of statistical significance 0.05, we conclude the independence of random variables, that is, the lack of self-correlation between errors.

**Figure (5): The functions of self-correlation and partial self-correlation of residues**



Source: spss.

We can see from Figure (5) that the coefficients of the functions of self-correlation and partial self-correlation of the remainder are all within the confines of confidence, as they follow the pattern of the White Noise chain (i.e., it is independent and distributed naturally).

As a result of the statistical Ljung-Box Q and ACF and PACF coefficients for the remainder, we infer the significance of the Holt model built.

**-Model accuracy measures:**

Table (15): Measurements of predictive accuracy tests for Holt model.

208.	207. Model Fit statistics				
209. Model Type	214. MSE	213. RMSE	212. MAPE	211. MAE	210. BIC
220. Holt	219. 103462	218. 334.537	217. 31.668	216. 201.191	215. 11.876

Source: From the outputs of the e-views program

Table (15) gives measurements related to predictive accuracy tests for Holt model .These measures will be used when comparing the Holt model with the ARIMA model (2,1,1).

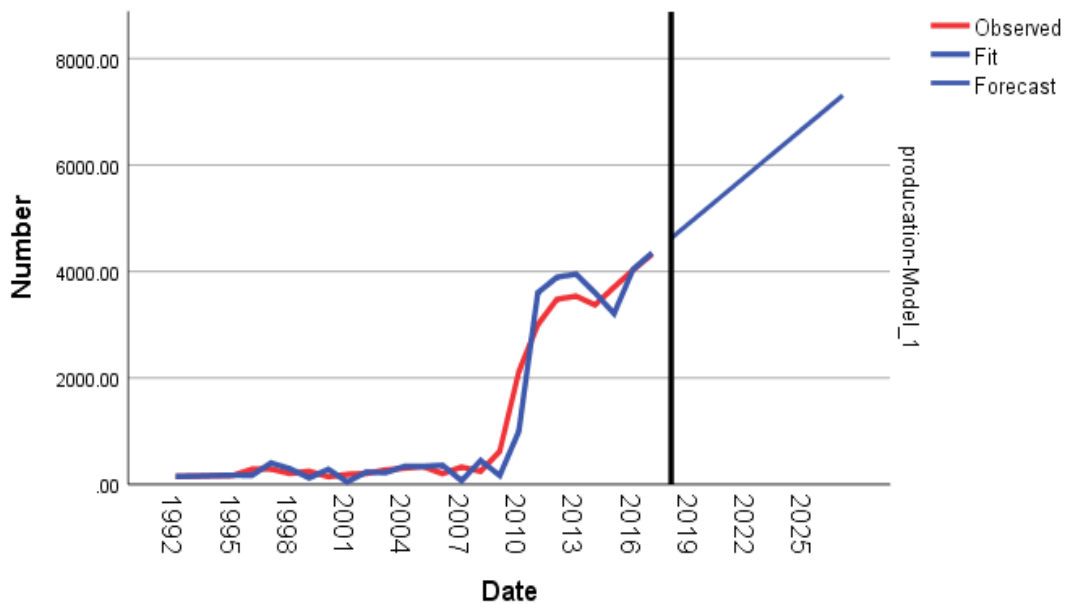
**-Prediction using the Holt model:**

Table (16): Predicted Cement Production for the Period (2018-2027) Using Exponential Holt Model

Model		2018	2019	2020	2021	2022	2023	2024	2025	2026	2027
Production-Model_1	Forecast	4624.23	4922.35	5220.48	5518.61	5816.73	6114.86	6412.99	6711.12	7009.24	7307.37
	UCL	5314.68	6465.49	7802.11	9297.29	10932.73	12695.19	14574.57	16562.86	18653.47	20840.89
	LCL	3933.78	3379.21	2638.85	1739.92	700.74	-465.46	-1748.60	-3140.63	-4634.99	-6226.15

Source: Prepared by the researcher by the program SPSS.

Figure (6): The actual, estimated and predicted production of cement for the period (2018-2027) using Holt Model



Source: Prepared by the researcher by the program SPSS.

Table 16 gives the predicted values of cement production for the next ten years using the exponential Holt model Figure 6 represents the actual, estimated and predicted values using the same model.

The predicted values show an increasing trend in production in the coming years .The figure also shows the approximation between the estimated series and the actual series, which is an indication of the validity of this model in its representation of the study data.

The appearance of negative values in the lower confidence threshold for forecasting in the last five years of the series reflects the weakness of exponential preliminary models in long-period predictions.

**Comparison of Box-Jenkins and Holt models:**

We will compare the ARIMA (2,1,1) model based on the Box-Jenkins methodology and the Holt model based on the exponential smoothing method, to determine the accuracy of the two methods and the trade-off between them to choose the best one to use in predicting cement production in Sudan. This will be done by using some statistical criteria in Table (17) below.



Figure (7): Actual, Estimated and Predicted Cement Production for the Period (2018-2027) Using Holt and Box Jenkins Model

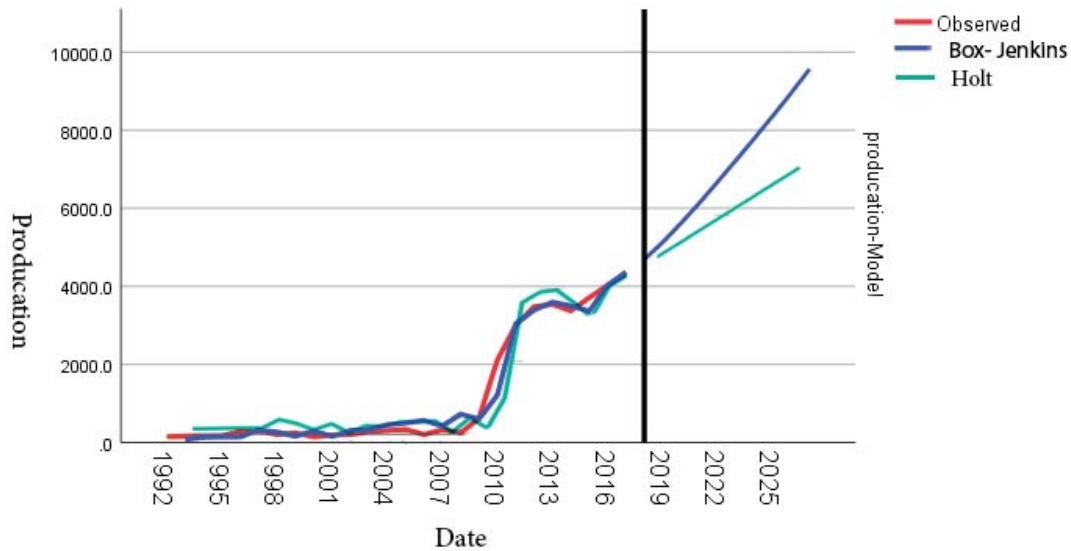


Table (17): Statistical Comparative Criteria between ARIMA (2,1,1) and Holt Exponential

The scale	Model ARIMA(2,1,1)	Model Holt
MSE	74670.83	103462
RMSE	269.593	335.537
MAPE	38.687	31.668
BIC	11.838	11.876

Source: Prepared by the researcher program e-views.

Table (17) shows some comparison criteria between the two models, and we note from these values that the model ARIMA (2,1,1) has exceeded the Holt exponential model in three of these criteria, which means that it is the most efficient and the most appropriate model for predicting cement production in Sudan. From Figure 7, the two lines estimated by the two models are close to the actual data line, which confirms the validity of the two models, with some superiority to the line representing the Box-Jenkins model compared to the Holt model line.

Random variations are one of the most significant factors that have directly reduced the efficiency of the Holt exponential model compared to the Box-Jenkins model.

### 14. Conclusion

In order to obtain an efficient statistical forecasting model that gives accurate forecasts of cement production in Sudan, the Box-Jenkins methodology and the exponential smoothing method (Holt model) were applied to the study data and constructed two prediction models. According to the results of the differentiation criteria, the ARIMA model (2,1,1) It is the best model for predicting cement production. This model outperformed the exponential Holt model in comparison criteria (MSE, RMSE) and the performance of the two models was convergent in the information standard (BIC) with a slight preference for the Box-Jenkins model.

Future forecast values obtained using the selected ARIMA model (2,1,1) indicated that there will be an increase in productivity over the coming years. All prediction values were within confidence limits, confirming the quality and validity of this prediction model.

Although the Holt exponential model was significant and passed all validity tests, when used in the calculation of prediction values for the next ten years, it showed some negative values within the minimum confidence threshold of the prediction for the last five years, which indicates the poor performance of these models in the long-term prediction.

Random variations were the most important factors that directly affected the performance of the two models, but they were more effective in reducing the efficiency of the Holt exponential model compared to the Box-Jenkins model.

## 15. References:

- 1-Akaike, H. (1973), "Information theory and extension of the maximum likelihood principle", Budapest.
- 2-Berry, Adnan Majed. (2002). Statistical prediction methods, Part 1. King Saud University, KSA. P. 16.15.
- 3-Brown, (1963). "Smoothing forecasting and prediction of discrete time series. Englewood Cliffs", N.J. Prentice Hall.
- 4-George .E.P. Box, Gwilym M. Jenkins&, Goegory C. Reninsed, (1994) Time Series Analysis Forecasting and Control, Prentice – Hall International, USA.
- 5-Gilani, Amal. (2017) Using the exponential predicate to predict the production and consumption of electricity in Sudan (1989-2015), Master Thesis, University of Gezira. p38.
- 6-Ibrahim, Al-Shaimaa. (2015). Box and Jenkins models applicable to spss. Faculty of Commerce, Damietta University, Egypt, p3.
- 7-Karim, Adnan and Hashem Hussein Ali (2018). The use of some statistical methods for the prediction of lost electric power in Karbala Governorate. University of Karbala Scientific Journal, 16 (2): p101-115.
- 8-Montgomery, D.C & Lynwood, A. J (1976), Forecasting and time Series Analysis, Mc Graw- Hill book company, New York.
- 9-Robert A.Yaffe and Monnie McGee, (2000), Introduction to Time Series Analysis & Forecasting, Academic Press, New York.
- 10-The Ministry of Industry. (2009). A report on the cement industry in Sudan. Khartoum. P. 23.