Academic Journal of International University of Erbil

Journal Homepage: https://ue.edu.krd/ojs/index.php/public

PRINT-ISSN: 2519-6928

Time Series Models to Forecast Barley Crop Production in The Kurdistan Region of Iraq

Lanja Assi Hayder[1], Hezha Mohammed Noor Mohammed[2], Jawher Jumaa Ali[3]*

[1]Accounting Departement, College of Administration and Economics, Charmo University, Sulaimani, Kurdistan Region, IRAQ

[2,3]Business Administration Department, college of Administration and Economics, Charmo University, Sulaimani, Kurdistan Region, IRAQ

*Corresponding Author: Lanja.Assi@chu.edu.iq

Received 02 Feb 2025; Accepted 04 Mar 2025; Available online Apr 2025

ABSTRACT: Time series are important methods in analyzing phenomena and events during specific time periods and predicting future values that contribute to giving an estimated picture of the phenomenon. Therefore, the study aimed to use the Box Jenkins methodology to forecasting barley production in the Kurdistan Region of Iraq for the period (2024-2030) based on the time series of barley production in Kurdistan for the period (1981-2023). It was found through studying the time series of barley production quantities for the period (1981-2023) that the time series is unstable in the mean and variance; after taking the second difference of the data to convert it to a stable series in the mean and then taking the natural logarithm transformation to convert it to a stable series around the variance, it was found that the appropriate model for predicting barley production in the Kurdistan Region is ARIMA (1,2,1) among the group of proposed models for having the lowest values for the criteria (BIC, SBC, H_Q). The study recommends using the models that have been developed to predict barley production in the Kurdistan Region and adopting the predictions resulting from these models in developing future plans. The study also recommends applying the Box-Jenkins methodology to deduce and develop models to predict other crops, based on the actual analysis of the time series as explained in the study.

Keywords: ARIMA Model, Barley Production, Box-Jenkins, Time Series forecasting.

1 INTRODUCTION

Forecasting is a fundamental objective in statistical, economic, and agricultural research, as it predicts future variable values to support planning and policy formulation, including decisions related to national productivity and imports. Forecasting is a statistical method that can help decision-makers make their future decisions with great accuracy.

Barley is one of the main crops in the Kurdistan Region and comes in second place in importance after wheat. Barley cultivation in Kurdistan depends mainly on the climatic conditions and geographical nature of the region, which is characterized by a semi-arid climate and fertile soil in many of its regions. It is one of the main crops grown in the region and is used as a source of food, whether for humans or as animal feed, especially in rural areas that depend on livestock. This crop is grown in the provinces of the Kurdistan Region because it can withstand cold and frost, as well as relatively high salinity, and therefore it is grown in all provinces of the region.

At the global level, this crop ranks fourth after wheat, rice, and corn in terms of importance and production in the world[1]. Therefore, it is considered one of the basic and main commodities that is used as fodder or food worldwide, as it is considered one of the main food sources for citizens in addition to its use as one of the sources of animal feed due to its high protein content [2].

Barley is grown in different areas, where the average cultivated area during the period (1981-2023) was (736,012) thousand dunums, and the average production during this period was (171,710) thousand tons [3].

This researcher seeks through this research to predict the future of barley production in the Kurdistan region for the next seven years in light of the data and figures obtained from official sources using the moving average autoregressive model.

The overall structure of this paper is divided into nine sections. The first section is a brief introduction and presents the research problems and objectives; the second section presents a literature review of some similar studies. The third section presents a review of the time series models, and then the fourth section presents a time series stationary test; the fifth section presents a Box-Jenkins model, and the sixth section presents a model selection criterion, and the seventh section presents forecasting. In the pre-final section, the description and analysis of the barley production time series from (1980-1981) to (2022-2023) are included. In the last section, which is the ninth section, the conclusion of the study is discussed.

1.1 IMPORTANCE OF RESEARCH

The importance of this research comes from employing forecasting methods to analyze the time series of barley production and directing the results of the analysis towards supporting decisions and policies related to the future of barley production in the Kurdistan Region. The most prominent forecasting methods used in this context are autoregressive models and integrated moving averages, which are known for their high accuracy in forecasting.

1.2 RESEARCH PROBLEM

Barley production is one of the main factors that contribute to supporting the economy. Therefore, it is necessary to conduct scientific studies aimed at discussing barley production and proposing appropriate solutions to overcome the problems and challenges that hinder the growth of its production and improve its quality in a way that enhances its contribution to the economy.

1.3 RESEARCH OBJECTIVE

This study aims to analyze the time series of barley production in the Kurdistan Region, focusing on developing the most suitable ARIMA model to obtain a stable time series and then using it to forecast barley production for the period 2024–2030, based on official production data from 1981 to 2023.

2 LITERATURE REVIEW

Statistical tools can address a wide range of issues and assist decision-makers in reaching the best possible outcomes through data analysis of economic or environmental events. Future projections using time series are becoming more significant since they help achieve necessary goals. Consequently, one of the better models in this area is that developed by Box and Jenkins. Much research has been done using ARIMA and Box-Jenkins's methodology; however, not much has been done on barley production, especially in Kurdistan.

For example, [4] has used ARIMA for wheat production in Iraq by using a time series based on the Box-Jenkins model., intending to make a forecast of wheat production for the next five years from 2020 to 2024 by relying on time series data for the study period 1980 - 2018. The research results showed that the best model for forecasting wheat production in Iraq is ARIMA (1,0,2). The results also showed that wheat production in Iraq will increase in the following years.

In another study, [5] has used the Box-Jenkins models to foresee the overall amount of municipal waste for the years 2019–2024. By analyzing the time series for the years 2008–2018, the ARIMA model (1,0,0) was found to be the ideal model for this data. This model was then used to estimate the amount of municipal waste for the years 2019–2024.

Similarly, [6] has used ARIMA to forecast tobacco in Zimbabwe. The researchers focused on the time series analysis of tobacco yield (1980-2018) using (ARIMA) models to forecast 2019-2023 yield, and ARIMA (1, 1, 0) was identified as the best model.

[7] aimed to forecast the food gap and production of wheat in Iraq by using Box-Jenkins. They found that the best method for forecasting wheat production for the period 2016-2025 is ARIMA (4,1,3), and for the food gap, the ARIMA model (1.0.1) was the best model for the same period. They stated that there is an increase in the production of wheat in Iraq during the coming years (2016-2025), while the food gap for wheat is continuous.

Meanwhile, [8]worked on forecasting the production of Iraq's maize crop using ARIMA models. Using the semi-annual data for the period 1980-2021, the production of the maize crop was forecasted for 5 years, starting from 2022 to 2026. They found that the ARIMA (4, 1, 12) model can be used for accurate and high predictive forecasting of maize production in Iraq.

3 METHODOLOGY

3.1 TIME SERIES

Time series are defined as a set of values arranged according to time and represent the phenomenon over successive and equal time periods, such as annual, monthly, or daily periods. Time series analysis aims to understand the nature of the changes in the values of the phenomenon over time, which facilitates making future estimates and predictions. This analysis includes identifying the basic components, such as the general trend that reflects long-term changes in the value of the phenomenon and the seasonal trend that highlights changes in values over short periods (less than a year), such as seasonal, monthly, and daily changes. In addition, there are periodic changes that occur regularly over periods longer than a year, which are related to economic and political developments[9].

Time series were used in data analysis through the Box-Jenkins model, which requires the availability of a long time series. To achieve this, a time series for barley production in the Kurdistan Region was relied upon, with 43 observations for the period between 1981 and 2023. The data were analyzed using statistical methods, with an attempt to stabilize them through modification and deletion, whether from the effect of variance or the general trend [4].

3.1.1 AUTOCORRELATION FUNCTION

It is a measure used to determine the degree of relationship between the values of a time series with themselves over different time periods with different shifts. This function is also used to study the behavior and stability of time series through successive time correlation (k). The correlation function also helps in understanding the basic pattern of the series and its components and is symbolized by the form ρ_k, The mathematical formula for the correlation function is as follows [10]:

$$\rho_k = Correlation\ (Y_t, Y_{t+k}) = Correlation(Y_t, Y_{t-k}) \tag{4}$$

$$\rho_k = \frac{Cov(Y_t, Y_{t+k})}{\sqrt{Var(Y_t)Var(Y_{t+k})}} = \frac{E[(Y_t-\mu)(Y_{t+k}-\mu)]}{\sqrt{E(Y_t-\mu)^2 E(Y_{t+k}-\mu)^2}} \tag{5}$$

$$\rho_k = \frac{\gamma_k}{\gamma_0} \qquad k = 0, \mp 1, \mp 2, \mp 3, \ldots$$

3.1.2 PARTIAL AUTOCORRELATION FUNCTION

The partial autocorrelation function $\emptyset_{kk}$ is defined as the function that explains the relationship between the two variables $(Y_t Y_{t+k})$ after excluding the effect of the variables that lie between them. This means that the partial autocorrelation coefficient can be used to measure the degree of the relationship between any two variables after removing the effect of the intermediate variables between them [11].

$$Corr\ ((y_t, y_{t+k} | y_{t+1}, y_{t+2}, \ldots\ldots, y_{t+k-1}) \tag{6}$$

To estimate the partial autocorrelation coefficients of the sample, we follow the following mathematical formula [12].

$$\emptyset_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \emptyset_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \emptyset_{k-1,j}\rho_j}, \qquad k = 2, 3, \ldots\ldots \tag{7}$$

$$\emptyset_{kj} = \emptyset_{k-1,j} - \emptyset_{kk}\emptyset_{k-1,k-1}, \qquad j = 1, 2, \ldots\ldots, k-1 \tag{8}$$

$$\emptyset_{kk} = \frac{Cov[(y_t - \hat{y}_t), (y_{t+k} - \hat{y}_{t+k})]}{\sqrt{Var(y_t - \hat{y}_t)}\sqrt{Var(y_{t+k} - \hat{y}_{t+k})}} \tag{9}$$

ρ_j is an estimate of the autocorrelation coefficients, and the partial autocorrelation function is an important additional tool in addition to the autocorrelation function. It helps us to determine the stability of the time series and to choose the degree of the appropriate AR, MA, or ARIMA model that represents the time series data better.

3.2 TIME SERIES STATIONARITY TESTS

A necessary condition for using the Box-Jenkins approach is that the time series be stationary. Stationarity is an essential property for analyzing time series and selecting the most appropriate mathematical model for them. One of the characteristics that distinguishes a stationary time series is time homogeneity, which is the condition where the shape of the series in one time period occasionally equals that of the series in another time period. A series may be considered stationary if any of the following apply [5]:

1. $E(Y_t) = E(Y_{t+k}) = \mu$ (1)
2. $Var(Y_t) = E[(Y_t - E(Y_t)]^2 = Var(Y_{t+k}) = E[(Y_{t+k} - E(Y_{t+k})]^2 = \gamma(0) = \delta^2$ (2)
3. $Cov(Y_t, Y_{t+k}) = E[(Y_t - \mu)(Y_{t+k} - \mu)] = Cov(Y_{t+k}, Y_{t+k+s}) = \gamma(k)$ (3)

There are several statistical methods that can be used to infer the stability of a time series, the most important of which are:

3.2.1 GRAPH

A graph is a reliable means of determining the nature of a time series, whether it is stationary or not, by observing the fluctuations of the series over time and determining whether there is a general trend or not. However, this method may not be conclusive at times, so tests based on statistical measures are resorted to obtain more accurate results [13].

3.2.2 UNIT ROOT TEST

It is one of the modern methods used to test the stationarity of time series, and this is done by fulfilling the stationarity condition when the unit root of the series lies within the unit circle. One of the most prominent methods used to test the stationarity of time series is the Dickey-Fuller test, which is based on estimating the following model [5].

$$\nabla Y_t = \gamma Y_{t-1} + \sum_{j=1}^{k} \emptyset_j \nabla Y_{t-j} + e_t \qquad (10)$$

The above equation is known as the Dickey-Fuller test. Dickey and Fuller developed their test in 1981, and it is known as the Augmented Dickey-Fuller test. It is more efficient than the simple Dickey-Fuller test[14], the Augmented Dickey-Fuller test is based on the following two hypotheses:

$$H_0: \gamma = 0 \ \ The\ time\ series\ is\ non-stationary.$$
$$H_1: \gamma \neq 0 \ \ The\ time\ series\ is\ stationary.$$

The test statistic is applied to the estimated value of γ and the value of t is calculated as follows:

$$T_{Cal} = \frac{\hat{\gamma}}{\delta\hat{\gamma}} \qquad (11)$$

Where:
$\hat{\gamma}$: is an estimate of parameter γ.
$\delta\hat{\gamma}$: represents the standard deviation.

By comparing the calculated value of the t-statistic with the table value of the Augmented Dickey-Fuller test, if the absolute value of the calculated t-statistic is greater than the table value, we reject the null hypothesis, and the series is stationary. However, if the calculated value of the t-statistic is less than the table value, we do not reject the null hypothesis, in which case the series is not stationary [15].

To determine the degree of stability that the series reaches, we repeat the test after taking the first difference. If the series becomes stationary after the first difference, the data are integrated to the first degree. Thus, to (d) of tests. If the series is found to be non-stationary according to the aforementioned tests, the data must be processed using the Box-Cox methodology, based on the following formula [16] :

$$y(\lambda) = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda} & if\ \lambda \neq 0 \\ Log\ y & if\ \lambda = 0 \end{cases} \qquad (12)$$

3.3 BOX-JENKINS TIME SERIES MODELS

Box-Jenkins models are important statistical methods in time series analysis. These models are used to represent time series of certain phenomena and to predict the future values of these phenomena. Among these models are [17]:

3.3.1 THE AUTOREGRESSIVE MODEL (ARP)

In this model, the current variable Yt depends on the values of the same variable in previous periods Y_{t-1}, Y_{t-2}….., Y_{t-p}. In other words, in an autoregressive model of order P, the current values Yt are derived from the weighted average of the previous values with an interval of P. For this reason, this model is known as an autoregressive of order AR(p). This model can be expressed using the following equation [18]:

$$Y_t = \emptyset_0 + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots \ldots + \emptyset_p Y_{t-p} + e_t \qquad (13)$$

Here:

$\emptyset_0$: Is the constant term.

$\emptyset_1 \ldots \ldots \emptyset_p$: Is the Parameters of the autoregressive models.

e_t: Is the random error.

We can write the model by using Bach Shift Operator

$$\boldsymbol{\emptyset_p(\beta) Y_t = e_t} \qquad (14)$$

Where:
$\emptyset_{\boldsymbol{p}}(\boldsymbol{\beta}) = (1 - \emptyset_1\beta - \emptyset_2\beta^2 - \cdots - \emptyset_p\beta^p)$

3.3.2 THE MOVING AVERAGES (MAQ)

This model is based on the current values of the time series based on the random errors that occurred in previous time periods. In other words, the model forms a relationship between the current value and the random errors of the past. The general formula for the model is [6]:

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots \ldots . \theta_q e_{t-q} \qquad (15)$$

The model can be written using the back shift operator as follows:

$$Y_t = \theta_q(\beta)e_t \qquad (16)$$

Where:
$\theta_q(\beta) = (1 - \theta_1\beta - \theta_2\beta^2 - \cdots - \theta_q\beta^q)$

3.3.3 THE AUTOREGRESSIVE MOVING AVERAGES MODE (ARMA PQ)

The processes of autoregressive and moving average models can be combined to create a merged autoregressive moving average process in the following way:

$$Y_t = \emptyset_0 + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots + \emptyset_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \qquad (17)$$

which is created by processes called $MA(q)$ and $AR(P)$. Such processes are called ARMA processes of order (p, q). One benefit of modeling time series data with a $ARMA$ process over only utilizing a MA or AR process is that a $ARMA$ process may be able to effectively represent a time series with fewer parameters. Using the simplest model that nevertheless describes the data is one of the main objectives of statistical modeling; this is known as the principle of parsimony [19]. The model can be written using the back shift operator as follows:

$$\emptyset(\beta)Y_t = \theta(\beta)e_t \qquad (18)$$

3.3.4 AUTOREGRESSIVE INTEGRATED MOVING AVERAGE MODEL ARIMA

When a time series is non-stationary, it must be transformed into a stationary time series before the mathematical model can be built. This is done by calculating the differences or transformations required to make the series stationary; the number of differences required is known as the degree of integration. The autoregressive moving average model is then transformed into an autoregressive moving average integrated model [20]. To transform a non-stationary time series into a stationary series, the backward difference operator is used, which is defined as follows [10]:

$$\nabla Y_t = (1 - \beta)Y_t = Y_t - Y_{t-1} \qquad (19)$$
$$\nabla \mathrm{d} Y_t = (1 - \beta)^d Y_t \qquad (20)$$

If we replace Yt with $\nabla^d Y_t$ in the previous equation, we get a new model that is able to handle a certain type of non-stationary time series, known as homogeneous time series. This model is called the "non-stationary mixed model", and is denoted by ARIMA (p, d, q). Where p denotes the autoregressive order, d represents the order of the differences, and q denotes the order of the moving averages. The mathematical formula for the ARIMA model is written as follows [7]:

$$\emptyset(\beta)\nabla^d Y_t = \emptyset(\beta)(1 - \beta)^2 Y_t = \theta(\beta)e_t \qquad (21)$$

Since
$\emptyset(\beta) = (1 - \emptyset_1\beta - \emptyset_2\beta^2 - \cdots - \emptyset_p\beta^p)$
$\theta(\beta) = (1 - \theta_1\beta - \theta_2\beta^2 - \cdots - \theta_q\beta^q)$

3.3.5 THE BOX JENKINS METHODOLOGY

Autoregressive integrated moving average (ARIMA) models can be diagnosed, prepared, tested, and used systematically, provided at least 30 data are available. Four iterative processes are followed: model identification, model testing, model diagnostic checking, and forecasting, which evaluates the model's suitability by examining residuals [21].

3.3.3.1MODEL IDENTIFICATION

The ARIMA model's rank and the stability of the original time series are established in this step, which is crucial in creating time series models. The partial autocorrelation function (PACF) and autocorrelation function (ACF) are useful tools for determining the best model for a series. Identifying issues in the series' initial observations is essential before studying it. Examining autocorrelation and partial autocorrelation functions alone cannot uncover every issue and the reasons behind its instability. While the autocorrelation function can identify broad trends, it may not identify other causes like variance instability [22]

3.3.3.2 MODEL ESTIMATION

During the estimation step, methods like ordinary least squares (OLS) and maximum likelihood estimation (MLE) are used to estimate the model parameters. Following the identification of the ARIMA model's orders, historical data is used to estimate the model's parameters. The maximum likelihood of the observed data given the model is found by maximizing the values of the model parameters', and this is the most widely used approach to parameter estimation [8].

3.3.3.3 MODEL DIAGNOSTIC CHECKING

After the models are estimated, they are diagnosed by estimating the residuals of the estimated models in order to choose the model that best represents the data. This is done by calculating the autocorrelation coefficient (ACF) and the partial autocorrelation coefficient (PACF). If the plot of the correlation coefficients all falls within the 95% confidence interval, this is evidence that the model is suitable for prediction purposes [13].

3.4 CHOOSING THE BEST CRITERIA

There are also other criteria that can be used to examine the model, the most important of which are [9] [21] :.

3.4.1 AKAIKE'S INFORMATION CRITERION (AIC)

This test is used as a tool to diagnose and select the optimal model and can be calculated according to the following formula:

$$AIC = nln(\sigma_E^2) + 2k \qquad (22)$$

Where:
K: Number of model parameters.
σ^2: represents the mean squared error.
N : Number of Observations.

3.4.2 SCHWARTZ BAYESIAN CRITERIA (SBC)

The criterion (SBC) is known mathematically by the following equation:

$$SBC = nln(\sigma_?^2) + k\ln(n) \qquad (23)$$

3.4.3 HANNAN – QUINN (H-Q)

$$H - Q = ln(\sigma_E^2) + 2kc\ln(\ln n)/n \qquad (24)$$

After conducting experiments on a number of models, the best model is the one with the lowest value that can be used.

3.5 FORECASTING

Forecasting is the final stage in time series analysis and can only be reached after completing all the necessary statistical tests to diagnose the chosen model. In this stage, the proposed model is practically applied to obtain the expected predictive values for the time series. After determining the model values (p, d, q) and estimating the model, it is used for forecasting by substituting the current and past values of the dependent variable Y_t and the residuals as an estimated value for the error term, with the aim of obtaining the future value$Y_{t+1}, Y_{t+2}, \dots$, and this is known as forecasting for future periods[6] .

4 RESULT AND DISCUSSION

In this section the research results with explanation provided

4.1 DATA DESCRIPTION & ANALYSIS

Table 1 shows barley production in the Kurdistan Region during the planting seasons of 1980-1981 to 2022-2023 [3]. We note clear fluctuations in the abundance and scarcity of production quantities for the study period, and there are significant fluctuations in barley production over the years. For example, there is a sharp decline in some years, such as 1983–1984 and 2007–2008, followed by a sharp increase in later years, such as 1984–1985 and 2008–2009. This suggests that productivity may be affected by seasonal, economic, or environmental factors. And from 2000 to 2007, production rose sharply from about 87,039 tons in 2000 to 610,118 tons in 2007. This may indicate improved farming practices, increased cultivated areas, or improved weather conditions. After 2007, production experienced some decline, dropping to 69,733 metric tons in 2008 and then experiencing some relative stability in subsequent years. Then, from 2010 to 2013, production was relatively low, with figures ranging from 213,955 to 41,740 tons. This reflects a period of instability in production, which may warrant further analysis on the reasons behind this decline. In 2014, production started to increase again and reached 204,758 tons in 2015. In subsequent years, there was a trend towards stability and growth of barley production. Production in 2022 was about 175,142 tons, and the amount produced between 2017 and 2023 has been in an unstable situation with figures ranging from 90,941 to 281,323 tons. This highlights the need for more effective strategies to improve and stabilize productivity.

Table 1. The quantity of barley production in the Kurdistan region of Iraq for the period (1981-2023)

NO	Year	Barley production	N0	Year	Barley production	N0	Year	Barley production
1	1980-1981	146,323	16	1995-1996	103,218	31	2010-2011	56,549
2	1981-1982	98,406	17	1996-1997	71,944	32	2011-2012	41,740
3	1982-1983	113,878	18	1997-1998	104,108	33	2012-2013	140,020
4	1983-1984	38,237	19	1998-1999	47,050	34	2013-2014	194,637
5	1984-1985	257,220	20	1999-2000	87,039	35	2014-2015	204,758
6	1985-1986	164,276	21	2000-2001	226,546	36	2015-2016	183,655
7	1986-1987	101,722	22	2001-2002	340,892	37	2016-2017	90,941
8	1987-1988	51,106	23	2002-2003	191,979	38	2017-2018	128,676
9	1988-1989	37,933	24	2003-2004	278,569	39	2018-2019	281,323
10	1989-1990	235,999	25	2004-2005	437,249	40	2019-2020	271,153
11	1990-1991	79,172	26	2005-2006	325.754	41	2020-2021	158,875
12	1991-1992	124,819	27	2006-2007	610,118	42	2021-2022	175,142
13	1992-1993	107,430	28	2007-2008	69,733	43	2022-2023	185,698
14	1993-1994	116,624	29	2008-2009	370,952			
15	1994-1995	118,108	30	2009-2010	213,955			

Source: Kurdistan Region Statistics Office

In general, Table 1 shows the inflation trend of barley production in the Kurdistan Region with the stages of growth and decline. It is important to assess the factors affecting this crop, including agricultural policies, climate conditions, and agricultural practices, in order to establish effective strategies to improve the sustainability of production and increase it in the future.

For analyzing the serious EViews 9 has been use through the research. First of all, before analyzing the collected data and figuring out a proper model of the production of barley in the Kurdistan region of Iraq, the data would be illustrated in the following chart:

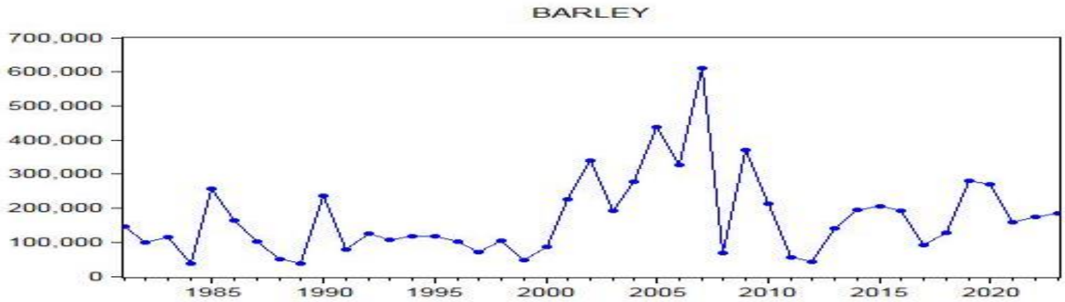
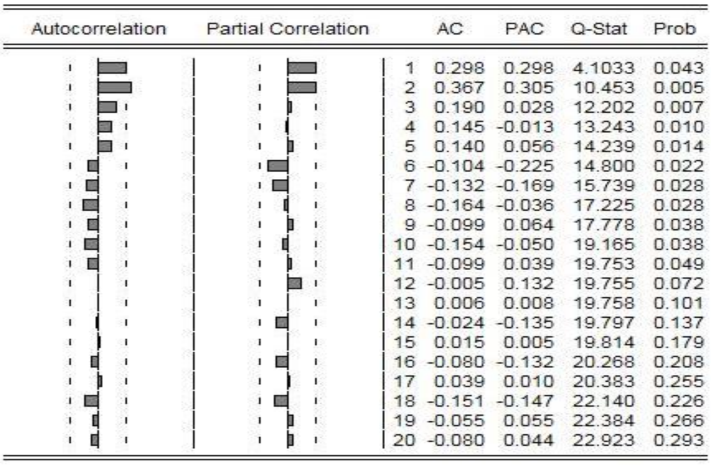


FIGURE 1. Time Series plot of total Barley product (1980 - 1981 up to 2022- 2023)

The above figure represents the time series plot of the total barley production in Kurdistan for the period of 1980-1981 up to 2022-2023, which was done to envision the data before considering performing any statistical test. The plot shows that there was a significant fluctuation in the barley production. We can see that from 1980 to 2000, the barley production was to some extent stable with slight fluctuations, and then it rose up sharply until 2007 but later decreased. Also, from 2010 and after, it continued to fluctuate slightly. In addition, between 2002 and 2011, barley production was high compared to the other years in the period. The autocorrelation and partial auto correlation function of the data shown below.



Lag	AC	PAC	Q-Stat	Prob
1	0.298	0.298	4.1033	0.043
2	0.367	0.305	10.453	0.005
3	0.190	0.028	12.202	0.007
4	0.145	-0.013	13.243	0.010
5	0.140	0.056	14.239	0.014
6	-0.104	-0.225	14.800	0.022
7	-0.132	-0.169	15.739	0.028
8	-0.164	-0.036	17.225	0.028
9	-0.099	0.064	17.778	0.038
10	-0.154	-0.050	19.165	0.038
11	-0.099	0.039	19.753	0.049
12	-0.005	0.132	19.755	0.072
13	0.006	0.008	19.758	0.101
14	-0.024	-0.135	19.797	0.137
15	0.015	0.005	19.814	0.179
16	-0.080	-0.132	20.268	0.208
17	0.039	0.010	20.383	0.255
18	-0.151	-0.147	22.140	0.226
19	-0.055	0.055	22.384	0.266
20	-0.080	0.044	22.923	0.293

FIGURE 2. The autocorrelation function and partial autocorrelation of the time series of barley production in Kurdistan

4.2 STATIONARY

From figure 1 we can see that the data is nonstationary, and also almost all the autocorrelations up to lag 20 that are roughly different from zero in figure 2 confirm nonstationary.

The unit root test represented by the Augmented Dickey-Fuller test was used:

H_0: Series is non-stationary (contains a unit root)

H_1: Series is stationary

The Augmented Dicky Fuller applied to test the stationarity of the series considering both intercept and Trend & Intercept, The ADF test result is shown below:

Table 2. ADF test for the original data

ADF test		Intercept	Trend & Intercept
ADF-test statistics		-2.690992	-2.808675
P-value		0.0842	0.2025
Test critical Values	0.01 level	-3.600987	-4.198503
	0.05 level	-2.935001	-3.523623
	0.1 level	-2.605836	-3.192902

The p-values obtained are greater than 0.05, we fail to reject the null hypothesis and conclude that the data is non-stationary. Therefore, the problem of instability must be addressed by:

<u>**First**</u>**:** We work to solve the instability in the variance by finding the transformation fitting to series data using the Box-Cox method, for the barley production series data lambda is close to zero as it is shown below:

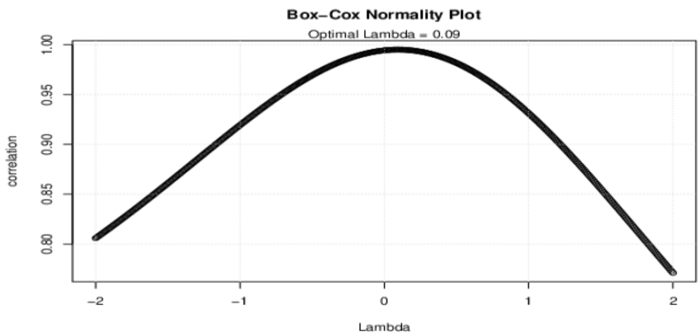


FIGURE 3. The Box-Cox Normality Plot

It is noted from the above figure (figure 3) that lambda is close to zero and equal to 0.09. Therefore, we use the natural log transformation. The graph of the barley production series after taking the natural log becomes:

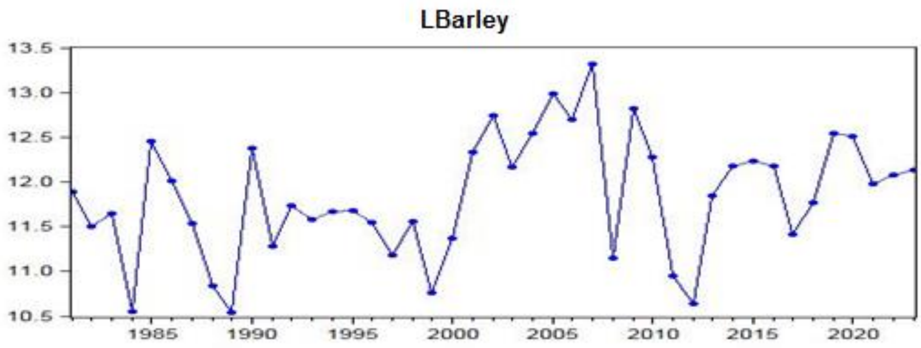


FIGURE 4. Time Series plot of the transformed data (LBarley) of the total barley product

From the above figure we can see that the data remains nonstationary.

Second: After treating the non-stationarity in the variance to a good extend we notice from Figure 4 that the time series still suffers from non-stationarity in the mean, so we work on taking the first difference of the transformed time series to make the time series stable in the mean.

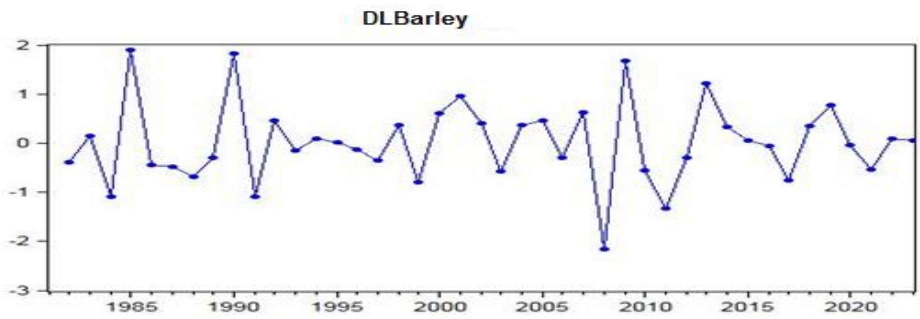


FIGURE 5. Time Series plot of the transformed data of the total Barley product after the first difference

We notice from Figure 5 that the new series became more fluctuated around the mean but does not become stationary after taking the first difference, so we go to the second difference and we plot the series again:

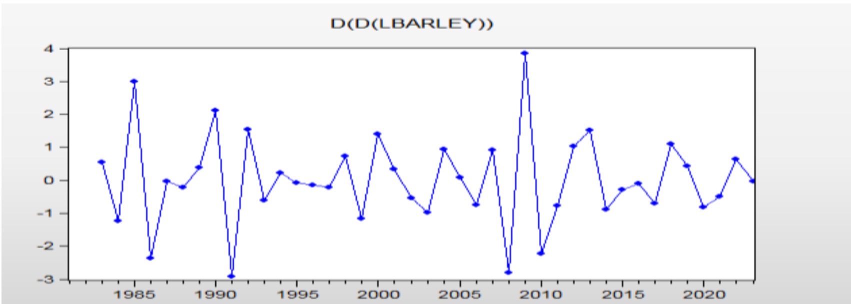


FIGURE 6. Time Series plot of the transformed data of the total Barley product after the second difference

From figure 6 we can see that the series after taking the natural log and two differences becomes stationary. For more certainty, the ADF test was applied to the new series as presented in the following table:

Table 3. ADF test for the series after log transformation and second difference

ADF test		Intercept	Trend & Intercept
ADF-test statistics		-7.484250	-7.366881
P-value		0.0000	0.0000
Test critical Values	0.01 level	-3.621023	-4.226815
	0.05 level	-2.943427	-3.536601
	0.1 level	-2.610263	-3.200320

Table 3 also confirms the stability of the time series as the test statistic reached a significant level (p-value = 0.0000), which is less than 0.05, and this indicates the rejection of the null hypothesis, i.e., the series is stationary.

Also, before starting to study the periodic behavior of the stable series, it is necessary to study the probability distribution of the time series, therefore, the Jarque-bera test was used and the results of the test shown in Figure 7 and Table 4.

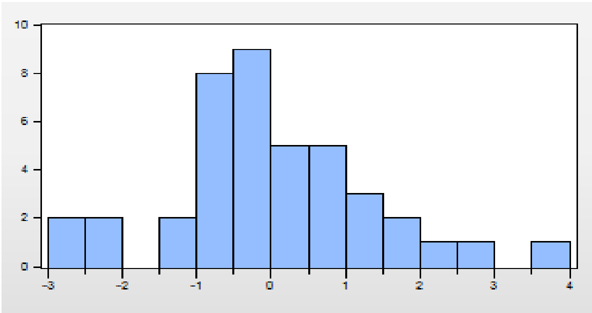


FIGURE 7. Time Series plot of the transformed data of the total Barley product after the second difference

From the normality histogram it can be seen that the transformed data has a normal distribution shape, also, the Jarque-Bera normality test from the EViews program gives results shown in the following table:

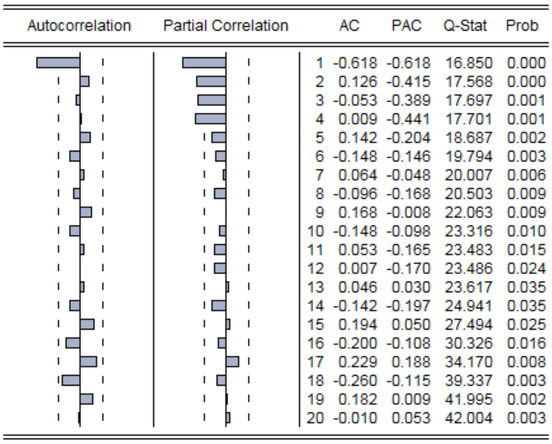
Table 4. Jarque-Bera test results

Series: D(D(LBARLEY)) Sample: 1981 – 2023 Observations: 41			
Mean	0.011103	Skewness	0.267522
Median	-0.069471	Kurtosis	3.827347
Maximum	3.840378	Jarque-Bera	1.658408
Minimum	-2.920211	Probability	0.436396
Std.Dev.	1.374698		

From Table 4, it can be seen that the value of the Jarque-Bera is 1.1658, and the p-value is 0.433, which is greater than alpha 0.05, it indicates that the transformed series is approximately normally distributed.

4.3 Model Identification

In this step, the autocorrelation and partial autocorrelation function graphs of the new series (after taking the log and two differences) are plotted



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.618	-0.618	16.850	0.000
		2	0.126	-0.415	17.568	0.000
		3	-0.053	-0.389	17.697	0.001
		4	0.009	-0.441	17.701	0.001
		5	0.142	-0.204	18.687	0.002
		6	-0.148	-0.146	19.794	0.003
		7	0.064	-0.048	20.007	0.006
		8	-0.096	-0.168	20.503	0.009
		9	0.168	-0.008	22.063	0.009
		10	-0.148	-0.098	23.316	0.010
		11	0.053	-0.165	23.483	0.015
		12	0.007	-0.170	23.486	0.024
		13	0.046	0.030	23.617	0.035
		14	-0.142	-0.197	24.941	0.035
		15	0.194	0.050	27.494	0.025
		16	-0.200	-0.108	30.326	0.016
		17	0.229	0.188	34.170	0.008
		18	-0.260	-0.115	39.337	0.003
		19	0.182	0.009	41.995	0.002
		20	-0.010	0.053	42.004	0.003

FIGURE 8. Results showing ACF and the PAC for the new series (DDL Barley)

It can be seen from Figure 8 that the autocorrelation coefficient of the new sequence is considerably non-zero when the lag order is 1, and it is out of the confidence band. And it is basically in the confidence interval when the lag order is greater than 1, so q can be taken 1. The partial autocorrelation coefficient is significantly nonzero when the lag order is equal to 1, and it is also different from 0 when the lag order is 2, 3 and 4, so p = 1, p = 2, p =3 or p = 4 can be considered. Considering that the judgment is very subjective, to establish a more accurate model, the range of values of p and q are chosen, and multiple ARIMA (p, q) models are established.

4.4 SELECTING THE CANDIDATE MODELS

Table 4 lists the candidates of ARIMA (p, d, q) models for different parameters, the model parameter values with their probabilities. In addition to the model R-squared, adjusted R-squared, Akaike information criterion, Schwarz criterion, and the Hannan-Quinn criteria.

It should be noted that while the Akaike info criterion (AIC) value, the Schwarz criterion (SC) value, and the Hannan-Quinn Criteria (HQ) value are typically used to choose the proper ARIMA model. The model with the minimum (AIC, SC, and HQ) measure is the best model. However, there are more model selection criteria for the ideal ARIMA model besides the minimal AIC, SC, and HQ values, which should be considered as significance of the coefficients and adjusted R squared. Considering all these points, the preferred model is the ARIMA (1,2,1) model. The model has the lowest (AIC, SC, and HQ) values, the highest adjusted R squared and the AR and MA coefficients are significant. Accordingly, the best model has AR value of order 1 (p = 1) and MA of order 1 (q = 1) after taking the log and two differences (d = 2) to the original series.

It can be seen from the t statistics of the model coefficients and their P values that the parameter estimates of the explanatory variables of the model are significant at the significance level of 0.01. Only the constant term is not significant, which means that the constant term is not important in the model. Also, the Durbin-Watson statistics of the model is 2.13 which means the model residuals does not have autocorrelation problem.

The model is used to fit values the real and fitted of the new series data, as well as the residual of the series, and the result is shown in Figure 9.

Table 5. Summary of the Candidate ARIMA models

Candidate ARIMA (p,d,q) Models	Consonant	AR	MA	R-squared	Adjusted R-squared	Criteria AIC	SBC	H-Q
ARIMA(0,2,1)	0.001834 **(0.8676)**		-0.999795 **(0.0000)**	0.642748	0.633587	2.517901	2.601490	2.548339
ARIMA (1,2,0)	0.003369 **(0.9753)**	-0.618313 **(0.0000)**		0.383768	0.367552	3.086373	3.170817	3.116906
ARIMA (2,2,0)	0.034280 **(0.8946)**	0.126847 **(0.4371)**		0.016408	-0.010175	3.560845	3.646156	3.591453
ARIMA(3,2,0)	-0.045679 **(0.8260)**	-0.053344 **(0.7321)**		0.003296	-0.024390	3.471963	3.558151	3.502628
ARIMA (4,2,0)	0.013538 **(0.9503)**	0.008878 **(0.9534)**		0.000099	-0.028470	3.595906	3.721289	3.641563
ARIMA (1,2,1)	-0.007050 **(0.7036)**	-0.464347 **(0.0019)**	-1.468476 **(0.0000)**	0.814947	0.804944	1.933393	2.060059	1.979191
ARIMA (2,2,1)	-0.001797 **(0.8753)**	-0.051930 **(0.7506)**	-0.999978 **(0.0000)**	0.653813	0.634580	2.567895	2.695861	2.613808
ARIMA(3,2,1)	-0.001101 **(0.9210)**	-0.069045 **(0.6582)**	-0.999842 **(0.0000)**	0.654825	0.635101	2.464191	2.593474	2.510189
ARIMA(4,2,1)	0.004508 **(0.7266)**	0.034241 **(0.8273)**	-0.999962 **(0.0000)**	0.629074	0.607255	2.478896	2.609511	2.524944

4.5 MODEL ESTABLISHMENT AND INSPECTION

The estimated results with the ARIMA model are as follows:

Table 6. Estimation results of the ARIMA model

Variable	Coefficient	Std. Error	t-Statistics	Prob.
Consonant	-0.007050	0.018385	-0.383450	0.7036
AR(1)	-0.464347	0.139185	-3.336187	0.0019
MA(1)	-1.468476	0.247740	-5.927497	0.0000
R-squared	0.814947		Mean dependent var.	-0.002188
Adjusted R-squared	0.804944		S.D. dependent var.	1.389541
S.E. of Regression	0.613692		Akaike info criterion	1.933393
Sum Squared Resid.	13.93488		Schwarz criterion	2.060059
Log Likelihood	-35.66785		Hanan-Quinn criterion	1.979191
F-Statistics	81.47135		Durbin Watson stat	2.138063
Prob(F-statistics)	0.000000			
Inverted AR Roots	-0.46		Estimated MA process is noninverted	
Inverted MA Roots	1.47			

The final model of the new series (DDL Barley) is ARIMA (1, 2, 1), and the following equation displays the specific form of the model. The estimated value of the variance of the corresponding error term is 0.613692.

$$\boldsymbol{ln\hat{y}_t} = -0.0070 - 0.4643\ \boldsymbol{lny_{t-1}} - 1.468\ \boldsymbol{e_{t-1}}$$

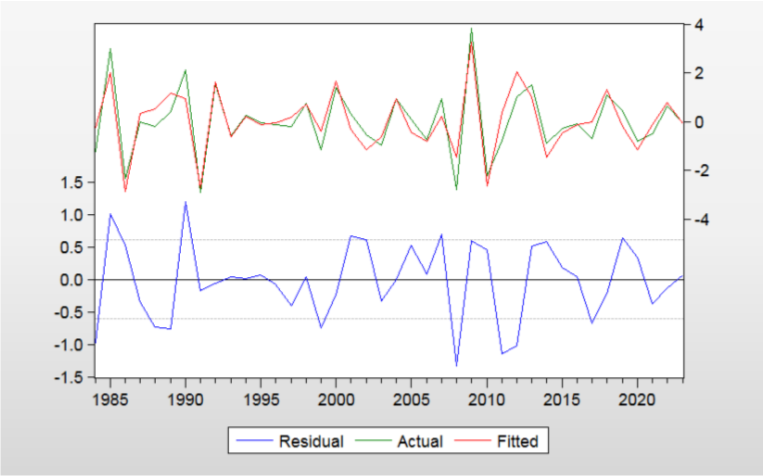


FIGURE 9. Actual series, fitted series and residual series of the DDL Barley sequence.

The Jarque-Bera test was conducted to determine whether the residuals are normally distributed or not, as shown in Figure 10 and Table7. The P-value was found to be greater than 0.05, indicating the acceptance of the null hypothesis that the residuals are normally distributed.

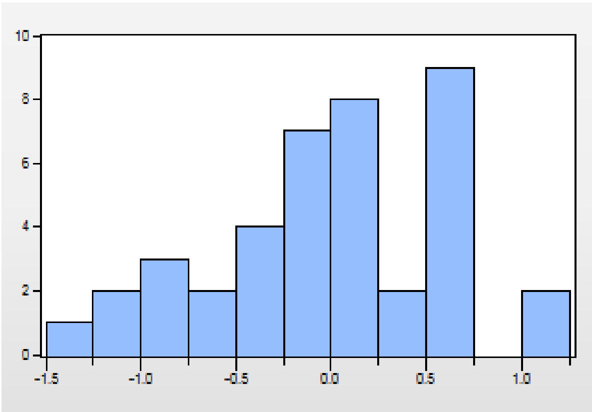


FIGURE 10. Jarque-Bera test of the residuals

From the normality histogram it can be seen that the transformed data has a normal distribution shape, also, the Jarque-Bera normality test from the EViews program gives results shown in the following table:

Table 7. Jarque-Bera test results of the residuals

Series: Residuals Sample: 1984 – 2023 Observations: 40			
Mean	-0.024981	Skewness	-0.231671
Median	0.019110	Kurtosis	2.509191
Maximum	1.186501	Jarque-Bera	0.759297
Minimum	-1.333785	Probability	0.684102
Std.Dev.	0.509191		

From the above table, it can be seen that the value of the Jarque-Bera is 0.759297, and the p-value is 0.684102, which is greater than alpha 0.05, it indicates that the model residuals are approximately normally distributed.

4.6 FORECASTING

Dynamic forecast mode is applied to forecast barley production in Kurdistan for the seven years from 2023-2024 to 2029-2030 period, the results are listed in Table 8.

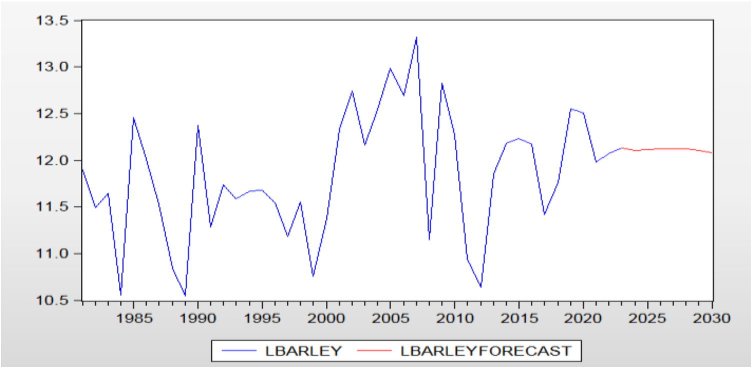


FIGURE 11. Plot of 2023 – 2024 to 2029-2030 barely production forecasts

Figure 11 shows the data after log transformation with the forecasted values of barley production until 2030. The predicted values for annual forecasts for the next 7 years are summarized in Table 8.

Table 8. Predicted values for annual forecasts for 7 years

	Year	Forecast (logBarley)	Forecast Barley (Ton)
1	2023-2024	12.11139	181 932
2	2024-2025	12.11727	183 005
3	2025-2026	12.10058	179 976
4	2026-2027	12.08405	177 026
5	2027-2028	12.05713	172 324
6	2028-2029	12.02470	166 825
7	2029-2030	11.98451	160 253

It is apparent from Table 7 and Figure 11 that the forecasted production of barley in Kurdistan fluctuates less between highs and lows, fewer than the production of this crop during the study period, which extended from 1980-1981 to 2022-2023. The expected barley production is likely to decrease from 181932 tons in 2023-2024 and 183005 tons in 2024-2025 to 160253 tons in 2029-2030.

The forecasting for the period 2024-2030 is very crucial for the government. One side from the Economic Diversification of the Kurdistan Regional Government (KRG) has been focusing on enhancing agricultural output to reduce reliance on oil revenues. The forecasting result in this study showed that the barley production will decrease in the future years; this result does not align with government diversification efforts, contributing to economic stability and growth. Also, barley is a crucial component of livestock feed. Enhanced production can support the livestock sector, ensuring a stable supply of feed, which in turn bolsters food security and supports rural livelihoods. Realizing the crop production benefits will require coordinated efforts between the government and agricultural producers to implement supportive policies, ensure price stability, and promote sustainable farming practices.

CONCLUSION

1. In the present study, the non-stationary time series of barley production in Kurdistan for the 1980-1981 to 2022–2023 period were converted into a stationary time series after transforming data by natural log and then taking the second difference of the transformed data, depending on the Augmented Dickey Fuller test for testing the stationarity of the data.

2. After transforming the sequence to a stationary sequence, a normality test is applied to ensure that the distribution of the data is normal.

3. Several ARIMA models were applied to find the best model. Depending on the Akaike info criterion (AIC) value, the Schwarz criterion (SC) value, and the Hannan-Quinn Criteria (HQ). The model with the minimum (AIC, SC, and HQ) measure is the best model. In addition, other important objects such as R-squared and adjusted R-squared of the models considered in indicating the proper ARIMA model.

4. The study showed that the data provide an ARIMA (1,2,1) model to forecast barley production in Kurdistan.

5. For checking the chosen ARIMA model, the graph of both the residuals and the actual series data vs the fitted series founded by the ARIMA model drawn. In addition, the normality test applied on the residuals.

6. The chosen ARIMA model has been used to forecast the barley production for the next seven years, from 2023-2024 to 2029-2030.

7. Suggestions for further study can be testing the ARIMA model with other essential crop productions in Kurdistan, including rice, wheat, or maize, for next years, to see if the outcomes align with government crop production plans. In addition, applying different time series forecasting models, such ARIMAX can be useful to examine their efficacy for agricultural forecasts can better than ARIMA.

REFERENCES

[1] N. A. Abdul Karem,"Estimating Function of Supply for Barley Crop in Iraq for the Period (1990- 2014)," Euphrates J. Agric. Sci., vol. 9, no. 2, pp. 236–243, 2017.

[2] Z. H. Mahmood, "economical analysis of barley supply response in iraq, irrigated, and rained area during 1990-2007," Anbar J. Agric. Sci., vol. 8, no. 4, pp. 244–254, 2010.

[3] Kurdistan Regional Statistical Office, ‘Summary of crop data 1969-2023’, Kurdistan Region Statistics Office. [Online]. Available: https://krso.gov.krd/content/upload/1/root/پوختەی-داتای-بەروبوومەکانی-هاوینە-و-زستانە1969-2023.pdf

[4] A. R. H. Al-Moola, "Forecasting of Wheat Production in Iraq Using Box- Jenkins Model for Period 2020-2024," Anbar J. Agric. Sci., vol. 18, no. 1, pp. 109–118, 2020, doi: 10.32649/ajas.2020.170516.

[5] L. K. Hamza and I. M. Alwan, "Using Time Series Analysis to Predicting of the Amount of Solid Waste for Baghdad Governorate From 2008 - 2018," in Al-Rafidain University College For Sciences, 2020, pp. 270–284.

[6] T. W. Mapuwei, J. Ndava, M. Kachaka, and B. Kusotera, "An Application of Time Series ARIMA Forecasting Model for Predicting Tobacco Production in Zimbabwe’" Am. J. Model. Optim., vol. 9, no. 1, pp. 15–22, 2022, doi: 10.12691/ajmo-9-1-3.

[7] I. K. Mustafa and O. K. Jbara, "Forecasting the food gap and production of wheat crop in Iraq for the period (2016-2025)," Iraqi J. Agric. Sci., vol. 49, no. 4, pp. 560–568, 2018, doi: 10.36103/ijas.v49i4.63.

[8] H. A. Drebee, N. A. A. Razak, and A. A. Mohsen, "Maize Production Forecasting in Iraq : A Box- Jenkins Approach for the Period of 2022-2026," in Earth and Environmental Science, 2023. doi: 10.1088/1755-1315/1259/1/012128.

[9] S. N. Yaequb, A. A. Aziz, H. M. Saeed, and H. O. Faqe, "Forecasting For Silver Closing Price and Modifying Predictions by Using Wavelet Transformation," Polytech. J. Humanit. Soc. Sci., vol. 3, no. 1, pp. 129–142, 2022. doi: 10.25156/ptjhss.v3n1y2022.pp129-142.

[10] H. A. H. Al Sarhan, "Using some statistical methods to predict the lost electrical energy (applied study)," University of Karbala, 2018.

[11] M. M. Faqe and A. S. Hamad, "Use the time-series models to predict the numbers of children with Respiratory Infections Streams(RIS) in the governorate of Sulaymaniyah," Univ. Kirkuk J. Adm. Econ. Sci., vol. 5, no. 1, pp. 201–219, 2015.

[12] B. S. Abdulhaq, "A statistical Model for Forecasting the Needs of Electrical Power in Erbil- Iraq," Mansoura University, 2014.

[13] A. H. M. Albasri, "Using ARIMA models to forecast the volume of cargo handled in Iraqi ports An applied study in the general company of Iraqi ports," J. Econ. Adm. Sci., vol. 26, no. 120, pp. 452–474, 2020.

[14] A. H. Farhan, "Using dickey _ fuller expanded test for testing variables of investment function in Iraq," J. Econ. Adm. Sci., vol. 25, no. 114, pp. 1–19, 2019. doi: 10.33095/jeas.v25i114.1752.

[15] N. J. Mohammed and A. T. Jabbar, "Forecasting Air Pollution Using Time Series," J. Madenat Al-Elem, vol. 13, no. 1, pp. 303–318, 2021.

[16] W. A. Hadba, "comparison method process capability indices for normal and non-normal data by use simulating," Kirkuk Univ. J. /Scientific Stud., vol. 11, no. 4, pp. 112–140, 2016.

[17] B. K. Ahmed, S. A. Rahim, B. B. Maaroof, and H. A. Taher, "Comparison Between ARIMA And Fourier ARIMA Model To Forecast The Demand Of Electricity In Sulaimani Governorate," Qalaai Zanist Sci. J., vol.

5, no. 3, pp. 908–940, 2020. doi: 10.25212/lfu.qzj.5.3.36.

[18] M. M. ELSayed, H. M. Mostafa, and Sanaa H.M. Sadek, "An Analytical Study of the Current and Future Status of Agricultural Land in Egypt," Alexandria Sci. Exch. J., vol. 44, no. 2, pp. 227–243, 2023, doi: 10.21608/asejaiqjsae.2023.296518.

[19] D. H. Kadir, "Time Series Modeling to Forecast on Consuming Electricity : A case study Analysis of electrical consumption in Erbil City from 2014 to 2018," Journal of Al-Rafidain University College For Sciences, 2020, pp. 473–485.

[20] A. A. Hamad, "The forecast quantities of electrical energy consumed per month in Anbar province , for the period 2013- 2014," AL-Anbar Univ. J. Econ. Adm. Sci., vol. 7, no. 14, pp. 343–356, 2015.

[21] E. Y. Hamid and M. H. Osman Abushaba, "Use of Exponential Holt Model and the Box-Jenkins Methodology in Predicting the Time Series of Cement Production in Sudan," J. Al-Qadisiyah Comput. Sci. Math., vol. 12, no. 1, pp. 42–59, 2020, doi: 10.29304/jqcm.2020.12.1.666.

[22] S. Hussein, "Using the ARIMA method in forecasting money supply in the Iraqi economy," Gulf Econ., vol. 34, no. 35, pp. 113–136, 2018.