Journal of Economics and Administrative Sciences
2024; 30(144), pp. 517-536
P-ISSN 2518-5764
E-ISSN 2227-703X

Journal of Economics and Administrative Sciences (JEAS)

Available online at http://jeasiq.uobaghdad.edu.iq
DOI: https://doi.org/10.33095/6afw4112

Using a Hybrid Model (EVDHM-ARIMA) to Forecast the Average Wheat Yield in Iraq

Hayder Khalid Rashid Al-Sammarraie *
Department of Statistics
College of Administration and Economics,
University of Baghdad, Iraq.
E-mail: haidar.khaled2201m@coadec.uobaghdad.edu.iq
Orcid: https://orcid.org/0009-0005-4410-3129
***Corresponding author**

Lamyaa Mohammed Ali Hameed
Department of Statistics
College of Administration and Economics,
University of Baghdad, Iraq.
E-mail: lamiaa.mohammed@coadec.uobaghdad.edu.iq
Orcid: https://orcid.org/0000-0001-7213-2115

Received: 30/7/2024 Accepted: 29/8/2024 Published Online First: 1 /12/ 2024

Abstract:
Purpose: Using an appropriate analytical method to deal with time series containing linear and nonlinear compounds and minimizing nonstationary in order to obtain good modeling and more reliable forecasts.
Theoretical framework: Many methodologies have been developed in the past to perform time series forecasting, including those presented by the two scientists (Box and Jenkins) and known as (ARIMA) models (Box et al., 2015), It gives more reliable forecasts when analyzing time series for linear compounds, but they are less suitable when dealing with nonlinear compounds that characterize real-world problems. This causes an increase in forecasting error, so it has recently been demonstrated use of hybrid models that compounds linear and nonlinear the best forecast is obtained.
Design/methodology/approach: Using the hybrid model (EVDHM-ARIMA) and comparing with single model (ARIMA) and comparing the two models using the (RMSE) criterion to forecast the average yield per dunum of the harvested area for the wheat crop in Iraq for the period (2024-2033), given the strategic importance of this crop in providing food security for the country.
Findings: The showed, value of (RMSE) for the estimates obtained from the hybrid model (EVDHM-ARIMA), is the best for forecasting of the research data.
Research, Practical & Social implications: We suggest generalizing the research idea for forecasting in different fields and comparing it with other forecasting methods.
Originality/value: Adopting the forecasting values obtained from the proposed method in the annual agricultural plans and developing proactive solutions to meet the citizen's need for his daily sustenance.
Keywords: Time Series, (ARIMA) Model, Eigenvalue Decomposition (EVD), Hankel Matrix (HM), Phillips-Perron Test (PPT).
JEL Classification: C2, C220, C380, C65, C320.
Acknowledgements

P-ISSN 2518-5764
E-ISSN 2227-703X

We extend our sincere thanks to the staff of the Authority Statistics and Geographic Information Systems for their assistance in collecting data.

1 Introduction:

Increasing wheat production for any country is considered a goal for the growth of its economy due to the strategic importance this crop represents in providing food security for the country. That is any study related to the wheat crop, which suffers from constant fluctuation in the quantity of its production, and diagnosing the causes of this fluctuation in order to give the best solutions to increase and improve its production is considered among the important topics that deserve study. Given the climate change the world is witnessing and its impact on the global stock of this crop, which has turned into a consumer commodity the demand increases with the increase in population. As Iraq is witnessing a deficit in producing this crop and supplying it with imported crops for several reasons, the most important of which is the lack of Rain, water allocation for agriculture, fires for crops and their destruction, and some farmers switched to other work, including grazing in the fields, which led to a decrease in the average yield of the country in recent years. Therefore, research is required to find a number of possible solutions in order to address the problems faced increasing the production of this crop and improving the reality of Its production in quantity and quality by providing future data and giving a quick picture of what the future And what producing of this crop may witness in the country for the coming years (Abdullah & Khalil, 2021).

So you ask much of the applied research and studies in various fields requires the use of statistical methods and quantitative measurement methods to determine the characteristics of the studied phenomenon in an unbiased scientific manner, and one of the most important of these methods is time series forecasting, where forecasting is considered a scientific method that aims to reduce the element of guesswork and intuition in the decision-making process by knowing the changes that occur over time and identifying and interpreting the results and what will happen in the future to the time series. Therefore, it is considered an important tool for planning as it represents the stage that precedes the planning of work.

Many forecasting methods have been used in the past, including autoregressive and integrated moving average (ARIMA) models, which have proven their efficiency in forecasting many phenomena and in various fields, such as forecasting the number of viral hepatitis infections (Kazom & Mohammed, 2016), and forecasting the value of the gross domestic product. For the sectors (public and private) in Iraq (Rashid & Abdulrahman, 2018), forecasting the volume of goods handled in Iraqi ports (Albasri, 2020), and forecasting monthly traffic accidents (Habeeb, 2020), they are suitable when analyzing time series with linear components and are less appropriate when dealing with nonlinear components. So it was proposed to use hybrid models based on (ARIMA) models to improve the accuracy of forecasting.

A modern method has been applied to treat non-stationary time series based on the eigenvalue decomposition of Hankel matrix (EVDHM), which has proven effective in decomposing nonlinear compounds. When combined with ARIMA models, the model becomes a hybrid and is called (EVDHM-ARIMA). It is possible to improve forecasting better than using individual ARIMA models alone in the forecasting process.

The rest of this article is organized in the following way. Section II reviews some previous studies on the use of hybrid models, while Section III reviews the methods used for forecasting, which are the single model (ARIMA) and the hybrid model (EVDHM-ARIMA). As for Section IV, reviews the results of the applied aspect using a program prepared in the language (MATLAB\R2022b) of forecasting agricultural time series for the average yield per dunum for the harvested area of the wheat crop in Iraq. The time series was used from the year (1943-2023) with (81) observations measured (kg/dunum), which were collected from the Ministry of Planning/Authority Statistics and Geographic Information Systems /Directorate of Agricultural Statistics, for various years (Statistics, 2023), and this period was used to find the best model to forecast it for the next ten years up to (2033). While Section V reviews the most important findings of the article. In conclusion, we mention some brief remarks.

2 Literature review and Hypothesis Development:

(Zhang et al., 2017) studied a forecasting method using ensemble empirical mode decomposition (EEMD) in combination with the ARIMA method. To improve the accuracy of daily forecasting of hotel occupancy, as the results show that this new method, EEMD-ARIMA, can improve forecast accuracy better than the common ARIMA method, because it has the lowest values when compared to the three statistical standards (RMSE, MAPE, MAE). (Aradhye et al., 2019) studied The hybrid model (ARIMA-SVM) to deal with time series data for both linear and non-linear parts together and compared it with the single models (ARIMA, SVM, ANN) for data on the sales price inflation index and the economic index on crude oil production data. In India, where the results showed for both types of The Data shows that the hybrid model (ARIMA-SVM) gave the best results through the use of comparison standards (MSE, MAE). (Mousa & Mohammed, 2020) studied the hybrid models (ARMAX-GARCH), (ARMAX-GARCHX) and the single model (ARMAX) to forecast the unemployment rate in the United States of America, using monthly data with (216) views and comparing the models used for forecasting. According to the comparison metrics (MAPE, MAE, Q-LIKE), it was found that the hybrid model (ARMAX-GARCHX) is the best for Forecasting the data because it has the lowest value for the metrics comparison. (Sharma et al., 2021) studied the developed eigenvalue decomposition approach for Hankel matrix (EVDHM) to decompose it into subcomponents, then they used ARIMA models to obtain a hybrid model (EVDHM-ARIMA) for forecasting. The hybrid model has proven highly efficient in forecasting the rest of the methods used in which the analysis is not repeated because it produces the same components such as single-spectrum analysis (SSA) as well as experimental mode decomposition (EMD) by having the lowest value of the (RMSE) criterion in estimating future values. For daily new cases of the pandemic disease COVID-19 for India, USA and Brazil. (Wazeer & Hameed, 2022) studied a method that combines the classical method (ARIMA) and the wavelet transform technique (Wavelet), which is called the hybrid model (Wavelet-ARIMA), compared it to the single model (ARIMA), and using functions (Haar, Db4, Db6) for a period of 64 weeks, where it was found that the model (Wavelet-ARIMA) hybrid Db6 function achieves the best results for Forecasting the exchange rate of the euro against the Iraqi dinar using comparison standards (AIC, BIC, MSE, MAPE, and RMSE). (Hamel & Abdulwahhab, 2022) studied and analyzed the single models SARIMA and NARNN and the hybrid models represented by the (SARIMA-NARNN) model to Forecast the numbers of people infected with the Covid-19 virus in Iraq and compared them using the lowest value of the comparison criteria (RMSE, MAE ,MAPE), where the superiority of the hybrid model over other models was proven. (Ghosh et al., 2023) studied The Concentration of fine particles in the air with dynamic diameters, including PM2.5 and PM10. Three cities in India (Kolkata, Siliguri, and Haldia) were selected. The daily concentrations of PM2.5 and PM10 were recorded. After training and testing the model based on EVDHM-ARIMA, it was compared with some well-performing models such as (Series Long and Short-Term Memory LSTM), (Convolutional Neural Network with Hybrid Long Short-Term Memory (CNN-LSTM), and (Seasonal Integrated Autoregressive and Moving Averages Model SARIMA). It has been shown that the EVDHM-ARIMA model is characterized by high efficiency and robustness for practical applications according to the lowest value of the two criteria (MAE, RMSE).

3 Research Methodology:

3.1 ARIMA(p,d,q) Model:

The model was proposed by Box & et al. in 1970 and its formulation is based on the autoregressive model AR(p), the moving average model MA(q), and the mixed model between them ARMA(p,q). It is considered an analytical method with one variable, as this method is represented by: Its four stages (stationary of data - defining the model and estimating its parameters - diagnostic examination of the model - forecasting). Most time series that arise from practical applications are characterized by the characteristic of non-stationary in (arithmetic mean and variance) in most of their forms. If they are non-stationary in variance, transformations are taken for them. Necessities, such as taking (the square root, the logarithm, the reciprocal of a number, or the reciprocal of the square root), and in the case of non-stationary around the mean, taking the differences of degree (d), which represents the rank of the difference factor, so that the model is formulated in the form (p, d, q) (ARIMA). Autoregressive Integrated Moving Average Model, and (p, d, q) refer to the ranks of the model and are non-negative integers. It is considered the most widely used model in time series because all models can be derived from it, and here its formula is using the Back Shift Operator (B). As follows (Box et al., 2015):

$$\left(1-\varphi_1 B-\cdots-\varphi_p B^p\right)(1-B)^d Z_t=s+\left(1-\theta_1 B-\cdots-\theta_q B^q\right)\varepsilon_t \cdots (1)$$

Z_t: represents the time series.

s: constant of the model.

$\varphi_1, \varphi_2 \ldots \varphi_p$: Parameters of the AR model.

$\theta_1, \theta_2 \ldots \theta_q$: Parameters of the MA model.

$(1-B)^d$: Difference factor.

$$(B)^j Z_t = Z_{t-j} \quad , j = 1,2, \ldots \qquad (2)$$

After obtaining the best stationary for the time series and confirming it by conducting the Augmented Dickey-Fuller Test (ADFT) (Harris, 1992), the stage of forming the appropriate model begins by determining the rank of the temporary model by observing the patterns of the autocorrelation functions (ACF) and partial autocorrelation functions (PACF) (Bisgaard & Kulahci, 2011). In order to ensure that the correct order of the model is chosen, we use the best value (AIC) Akaike's Information Criterion ,which is the lowest value for this criterion (Akaike, 1974).

$$AIC = n \ln \hat{\sigma}_\varepsilon^2 + 2V \qquad (3)$$

$\hat{\sigma}_\varepsilon^2$: The amount of variance of the model's residuals.

V: Model rank.

n: Sample size.

The next stage is to estimate the model parameters through several methods for efficient estimation. Here we will use the maximum likelihood method (Wei & W.S, 2006), then we test the suitability of the model by calculating the autocorrelation coefficients and partial autocorrelation of the residuals for a set of displacements that are within confidence limits or not, and to find out whether the correlations are randomly distributed or not. We resort to the Ljung-Box test according to the following test hypothesis:

$H_0: r_1(\varepsilon) = r_2(\varepsilon) \ldots = r_m(\varepsilon) = 0$

$H_1: r_1(\varepsilon) \neq r_2(\varepsilon) \ldots \neq r_m(\varepsilon) = 0$

$$Q = n(n+2)\sum_{k=1}^{m}\frac{r_k^2(\hat{\varepsilon})}{(n-k)} \qquad (4)$$

m: Lac number under test.

$r_k(\hat{\varepsilon})$: Autocorrelation estimated from lac k.

With a Chi-square distribution (x^2 Distribution) and with a degree of freedom (m-p-q) in nonseasonal time series and with a certain level of significance (α). If the value of [Q < x^2(m-p-q ,α)] or through the value of [p-value ≥ 0.05], then we accept the null hypothesis, meaning that the model is adequate to represent the data and the errors are independent (Ljung & Box, 1978). If it is the opposite, a new model must be diagnosed, then its parameters should be estimated, and then its suitability for representing time series data can be determined. Then comes the forecasting stage after the diagnostic model passes the test of random distribution of errors.

3.2 (EVDHM-ARIMA) Model:

The hybrid model-based procedure (EVDHM-ARIMA) is applied to derive reliable forecasting based on the previous data series. After including the time series data in forming the structure of the matrix and extracting the decomposed components (CP) that are closest to stationary, the Phillips-Perron test (PPT) is applied to verify the stationary of each component obtained. If the component does not meet the stationary property, the matrix can be restructured from this component and The EVDHM method will be applied which will decompose them into other stationary components, and so on until the best stationary of the extracted components is obtained, or up to a maximum of three iterations. After that, all the chains of the extracted components are modeled using the ARIMA model for each component, and then their future forecasting is made. Finally, Combine all of these forecasting to produce a composite forecast result (Sharma et al., 2021).

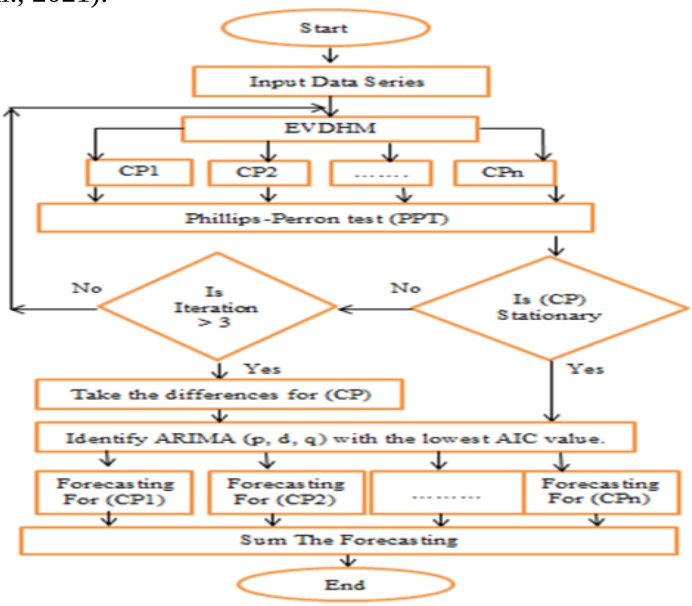

Figure (1): How to build a model (EVDHM-ARIMA) for time series forecasting.

3.2.1 Eigenvalue Decomposition of Hankel Matrix (EVDHM):

In complex, non-linear and non-stationary time series, the estimation of ARIMA models is not satisfactory enough. Also, single models are not sufficient to estimate all types of data. Therefore, a new method has been developed recently to analyze non-linear time series data with Hankel matrix eigenvalue decomposition (EVDHM). This method transforms the data series into a Hankel matrix (H), then decomposes this matrix into its eigenvalues and eigenvectors, and then by decomposing a matrix of eigenvalues into matrices such that each matrix contains a pair of distinct eigenvalues in order to effectively analyzes a data series to obtain single-component subsets. These subsets are closer to stationarity (non-stationary lower compared to the main data series) and are also well qualified to avoid the value of the integration parameter (d) which is high, given that the value of (d) for the main data series is lower if its value is taken on the extracted sub-components, and this helps in reducing forecasting errors (Sharma et al., 2021).

3.2.2 Hankel Matrix (HM):

The Hankel matrix used in the time series analysis process is a research topic for applications in various fields, including:

1. Time series forecasting: Research has revealed that using the Hankel matrix enhances the accuracy of forecasting, after converting the time series data to the Hankel matrix, analyzing it, and applying different forecasting methods to it (Ghosh et al., 2023).
2. Dimensionality reduction: Some studies use time series data by converting them to Hankel matrices and performing dimensionality reduction applications such as principal compounds analysis (PCA) or singular value decomposition (SVD) in order to simplify the analysis while preserving basic information (Yin et al., 2019).
3. Signal processing: The Hankel matrix is used in signal processing to filter and analyze time series data. It also works to improve the quality of signals when applying techniques based on the Hankel matrix to reduce noise, such as ECG signals (Sharma & Pachori, 2018a), and estimating the instantaneous frequency of the audio speech signal (Jain & Pachori, 2014).

It is a square matrix with fixed diagonals and was named after the German mathematician Hermann Hankel (Nagvanshi et al., 2023).

It is possible to use the square Hankel matrix in the process of analyzing interrupted time series by forming a square matrix of size (T×T) for the time series $(Z_{(t)}, t = 1,2, \dots\dots, a)$ and it is identical as in the formula The following (Sharma et al., 2021):

$$H=\begin{bmatrix} Z_{(1)} & Z_{(2)} & \cdot & \cdot & Z_{(T)} \\ Z_{(2)} & Z_{(3)} & \cdot & \cdot & Z_{(T+1)} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ Z_{(T)} & Z_{(T+1)} & \cdot & \cdot & Z_{(2T-1)} \end{bmatrix} \qquad (5)$$

H: Hankel matrix (square and symmetric).

a=2T-1: represents the length of the time series and must be an odd number (Sharma & Pachori, 2018b).

T=(a+1)/2: The embedding dimension, equal to the number of rows and equal to the number of columns (Singh & Pachori, 2022).

3.2.3 Eigenvalue Decomposition (EVD):

If H is a square matrix of order T×T (information matrix), then the matrix (V) is called the eigenvector matrix, and the λ matrix is the eigenvalue matrix that fulfills the following formula (Strang, 2009):

$$H_{(T*T)}\, V_{(T*T)} = V_{(T*T)}\lambda_{(T*T)} \qquad (6)$$

Thus, we get new vectors with the same direction, and their length is equal to one because) $V'_i * V_i = 1$) is orthonormal. If the matrix T has different eigenvalues, it will have T eigenvectors. The corresponding different units are orthogonal, meaning $(V'_i * V_j = 0\,, i \neq j)$. The sum of the eigenvalues of the matrix H is equal to the sum of the elements of its main diagonal, called the trace of the matrix, and the product of the multiplication of the eigenvalues of the matrix equals the value of its determinant. Since the vectors are orthogonal, then: $(VV' = I$ ، $V' = V^{-1})$ and thus the formula becomes as follows (N. H. (Ed) Timm, 2002):

$$H = V\lambda V' = V\lambda V^{-1} \qquad (7)$$

Formula No. (7) is referred to as (eigenvalue decomposition or Eigen decomposition). After obtaining these matrices, the time series analysis relies heavily on this eigenvalue matrix. The distinct pairs of eigenvalues play an important role in the decomposition process by choosing the number of decomposed components for the matrix. The sum of the size of all distinct eigenvalue pairs must constitute at least % 90 of the sum of the total eigenvalues $(\sum_1^T \lambda_i)$ as the threshold criterion. As shown in the following table (1) (Pachori, 2023):

Table (1): Method of choosing the number of decomposed components for the matrix

λ_i	Absolute Values (λ_i)	Relative Importance (%)	Eigenvalue Pair = (i , T-i+1)	Number-CP= Max (T+1/2)
$-\lambda_1$	$+\lambda_1$	(1)= $\frac{+\lambda_1}{\sum_1^T \lambda_i} * 100$	=(1)+(T)	1
$-\lambda_2$	$+\lambda_2$	(2)= $\frac{+\lambda_2}{\sum_1^T \lambda_i} * 100$	=(2)+(T-1)	2
$-\lambda_3$	$+\lambda_3$	(3)= $\frac{+\lambda_3}{\sum_1^T \lambda_i} * 100$	=(3)+(T-2)	3
. .	. .	. .	. .	. .
$+\lambda_{T-2}$	$+\lambda_{T-2}$	(T-2)= $\frac{+\lambda_{T-2}}{\sum_1^T \lambda_i} * 100$		
$+\lambda_{T-1}$	$+\lambda_{T-1}$	(T-1)= $\frac{+\lambda_{T-1}}{\sum_1^T \lambda_i} * 100$		
$+\lambda_T$	$+\lambda_T$	(T)= $\frac{+\lambda_T}{\sum_1^T \lambda_i} * 100$		
	$\sum_1^T \lambda_i$	100	Sum ≥ 90% STOP (threshold criteria)	

New eigenvalue matrix can be formed, and we will denote them with the symbol $\tilde{\lambda}_i$, using the distinct eigenvalue pair, which holds only the values of the distinct pair (i, T-i+1) and the rest of the matrix values are zeros. For example, the eigenvalue matrix of the first distinct pair can be represented as follows:

$$\tilde{\lambda}_1 = \begin{bmatrix} -\lambda_1 & 0 & \cdot & 0 & 0 \\ 0 & 0 & \cdot & . & 0 \\ . & . & \cdot & . & . \\ . & . & \cdot & . & . \\ . & . & \cdot & 0 & 0 \\ 0 & 0 & \cdot & 0 & +\lambda_T \end{bmatrix} \quad (8)$$

The eigenvalue matrix of the second discriminant pair can be represented as follows:

$$\tilde{\lambda}_2 = \begin{bmatrix} 0 & 0 & \cdot & 0 & 0 \\ 0 & -\lambda_2 & \cdot & . & 0 \\ . & . & \cdot & . & . \\ . & . & \cdot & . & . \\ . & . & \cdot & +\lambda_{T-1} & 0 \\ 0 & 0 & \cdot & 0 & 0 \end{bmatrix} \quad (9)$$

Thus, the matrices of distinct eigenvalue pairs can be represented by the number of identified principal components. In order to calculate the values of the component elements, for example, the first decomposed component is calculated by taking the average of the skewed diagonal elements of the matrix formed according to the following formula (Sharma & Pachori, 2017):

$$\tilde{H}_1 = V\tilde{\lambda}_1 V^{-1} \quad (10)$$

Where a square matrix with degree (T×T) appears to us, as shown in formula (11) and with the default symbols below, where the symbols (a, y, x, c, b, d) represent values for the elements of the first component matrix $\tilde{H}_1$, each according to its location. So that d represents the first value in the component and a represents the last value in it, The Table (2) shows the method of calculating the values as follows:

P-ISSN 2518-5764
E-ISSN 2227-703X

$$\tilde{H}_1 = \begin{bmatrix} d & b_1 & c_1 & \cdot & \cdot & \cdot \\ b_2 & c_2 & \cdot & \cdot & \cdot & \cdot \\ c_3 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & x_1 \\ \cdot & \cdot & \cdot & \cdot & x_2 & y_1 \\ \cdot & \cdot & \cdot & x_3 & y_2 & a \end{bmatrix} \quad (11)$$

Table (2): Method of calculating the values of the elements of the first decomposed component.

first component matrix	sequence of the element in the component	value of the element in the component
$\tilde{H}_1 = \begin{bmatrix} d & b_1 & c_1 & \cdot & \cdot & \cdot \\ b_2 & c_2 & \cdot & \cdot & \cdot & \cdot \\ c_3 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & x_1 \\ \cdot & \cdot & \cdot & \cdot & x_2 & y_1 \\ \cdot & \cdot & \cdot & x_3 & y_2 & a \end{bmatrix}$	1	$= \frac{d}{1}$
	2	$= \frac{b_1 + b_2}{2}$
	3	$= \frac{c_1 + c_2 + c_3}{3}$
	⋮	⋮
	2T-3	$= \frac{x_1 + x_2 + x_3}{3}$
	2T-2	$= \frac{y_1 + y_2}{2}$
	2T-1	$= \frac{a}{1}$

Thus, the values of the second component of the matrix are calculated $(\tilde{H}_2 = V\tilde{\lambda}_2V^{-1})$ and the third component of the matrix $(\tilde{H}_3 = V\tilde{\lambda}_3V^{-1})$ and so we continue for the rest of the other components with the difference In the value of symbols (a,y,x,c,b,d) from one matrix to another.

3.2.4 Philips and Perron Test (PPT):

The extended Dickey-Fuller test assumes that the error term is statistically independent and includes a constant variance, while Philips and Perron (1988) developed a non-parametric test, which takes into account the conditional variance of errors. It allows eliminating biases of features related to random fluctuations, as they adopted the same distributions. Ltd. To test the ADF (Phillips & Perron, 1988), the PPT test will be used to determine the stationary of the time series for all the decomposed components analyzed by decomposing the eigenvalue of the Hankel matrix for univariate data. The null hypothesis of the unit root is tested due to the nonparametric property by observing the value of the logical decision vector (h) and the value of The (p-value) and the methodology of this test is that when the value of (h=0) the time series is non-stationary , but if the value of (h=1) then it is stationary , and if the value of (p-value) is less than the level of significance the statistic is (0.05), then the null hypothesis is rejected and the alternative hypothesis is accepted. Here, the time series is considered stationary. However, if it is greater than (0.05), the null hypothesis is accepted, and here the time series is considered non-stationary (Sharma et al., 2021).

3.2.5 Sum Squares of Values (SS Value):

The sum of the squares of the values in the process of decomposing the eigenvalue of the Hankel matrix is an indicator to evaluate the strength of the components decomposed from the time series. When the value of this indicator for the decomposed components that are characterized by non-stationary at the last iteration of the EVDHM process appears to be less than (2.0%) of its value in the first decomposed component, then these components play an insignificant role in forming the model and forecasting it. Likewise, when a component appears and the value of this indicator is close to or equal to its value in the decomposed component, then this component is generally considered a non-stationary trend component, and further analysis of these components is unnecessary, so we must terminate it. Repetition and the tendency towards taking the difference to the value of the parameter (d) when using ARIMA models for these cases (Sharma et al., 2021). Its mathematical formula is given as follows:

$$\text{SS Value} = \sum_{t=1}^{a} z_t^{\,2} \qquad (12)$$

a: the upper limit of the time series.

z_t: Time series data.

3.2.6 Root Mean Squared Error (RMSE):

It is also known as the standard deviation of error, and it measures the forecast performance of the proposed models, and its formula is given as follows (Chai & Draxler, 2014):

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n}(Z_t - \hat{Z}_t)^2}{n}} \qquad (13)$$

$\hat{z}_t$: Estimated values.

Based on this criterion, a comparison will be made between the Box & Jenkins model represented by the single model (ARIMA), and the proposed hybrid model (EVDHM-ARIMA) to choose which one is better in giving the lowest value for the criterion (RMSE).

4 Results:

4.1 (ARIMA) Model Analysis:

First, plot the real data for the time series as in Figure (2) where it is clear that the time series is nonstationary in variance and mean, so a logarithmic transformation was taken and then the first difference was taken for it. In addition, plot the time series data after taking the natural logarithm and its first difference to achieve the condition of stationary in Figure (3) which shows the stationary of the time series. Then, the expanded Dickey-Fuller test was conducted to confirm after taking the natural logarithm and the first difference, where the value of the test statistic was shown to be equal to (-13.3510) and its probability value (P-Value) was equal to (0.001).

This is less than the value of (0.05), that is, accepting the alternative hypothesis and that the series timeframe has become stationary ,and as shown in Figure (4) for plotting the functions (ACF&PACF), which shows that there are (2) coefficients outside the zero limits of the function (ACF) and (5) limits of the function (PACF) that are used to determine the rank of the model, so several candidate models were tested up to rank (5), and then the Choose the best model that has the lowest value of the AIC criterion, and is equal to (4.674) as shown in table (3), which is the ARIMA (2,1,3) model. To ensure the stationary of the model estimated through Figure (5), we notice an almost constant fluctuation in the values of the residuals around zero, that is, the condition of independence is met, and this is what we observe in Figure (6), where the correlations appear within confidence limits. To confirm that they are not related to each other, the Ljung-Box test statistics showed a value equal to (15.6436), which is less than the value of (x^2 (20-2-3, 0.05) =24.996), as well as the value of (p-value= 0.7385>0.05). Therefore, we accept the null hypothesis that the residuals are independent.

Forecasting is the final stage of building an ARIMA model. At this stage, and through the use of the estimated model, we worked to calculate (81) estimated values for the time series within the range of series limits, as shown in Figure (7) which illustrates the behavior of the actual values with the behavior of the estimated values using the single model ARIMA (2,1,3). The values of the Errors for the estimates obtained from the model in addition to calculating the (RMSE) index are displaced as it is shown in Table (4) after taking the inverse of the natural logarithm to obtain the true values of these estimates.

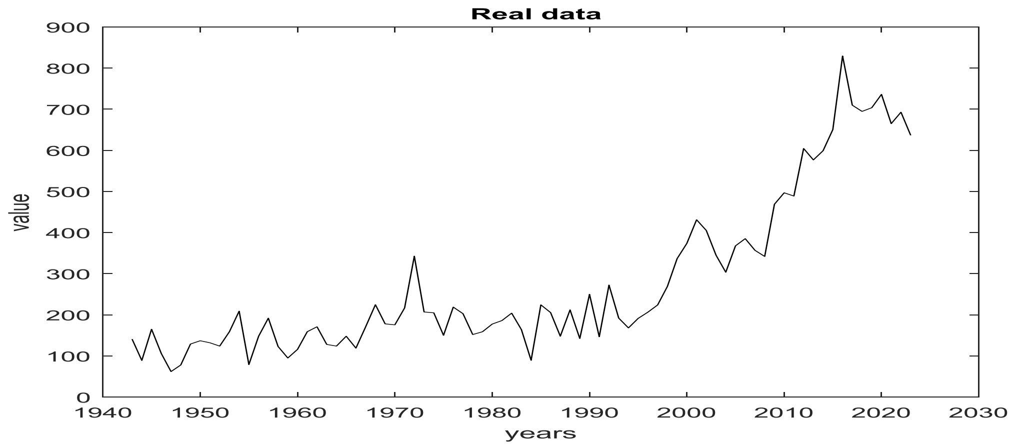

Figure (2): Plot of Real Data

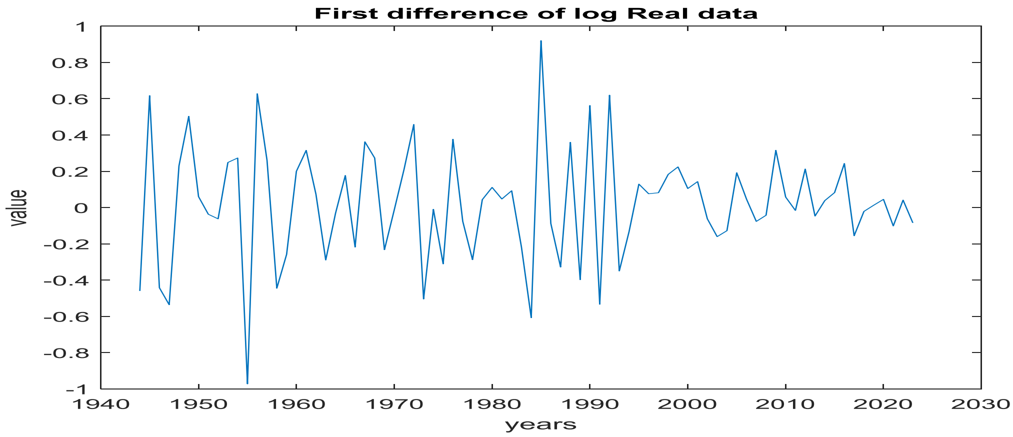

Figure (3): plotting the time series after taking the logarithm and the first difference

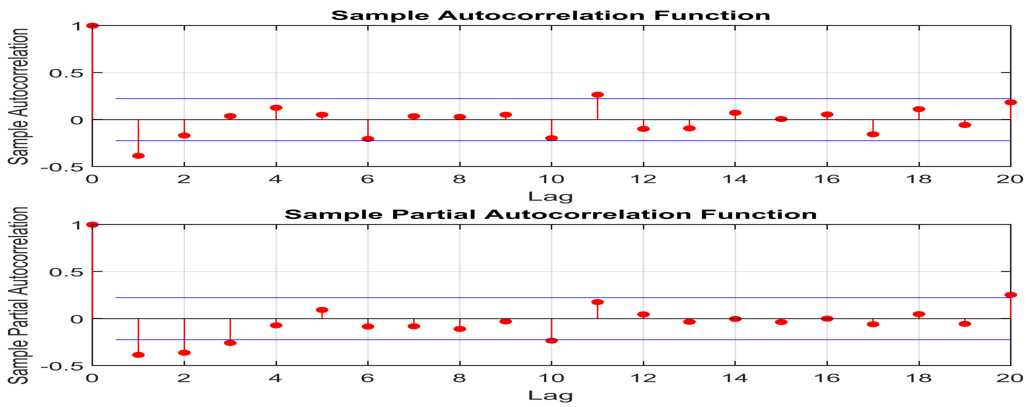

Figure (4): Plot of functions (ACF, PACF) after taking the logarithm and the first difference

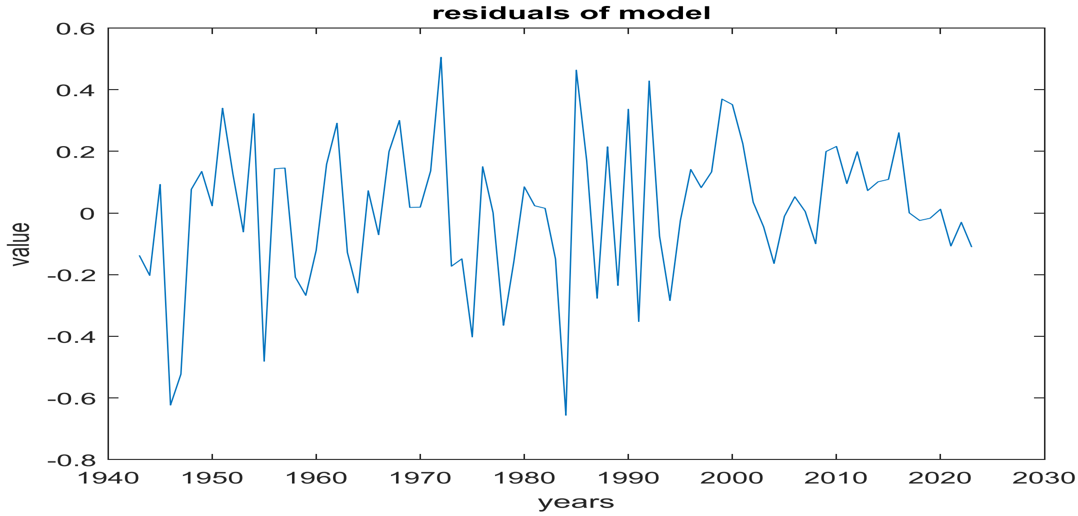

Figure (5): Residuals of Model

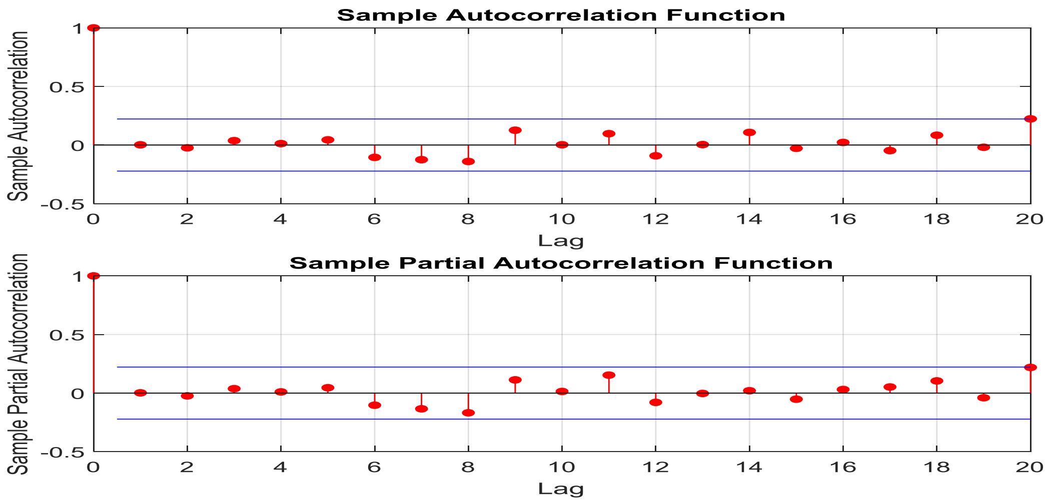

Figure (6): (ACF & PACF) of the Residuals

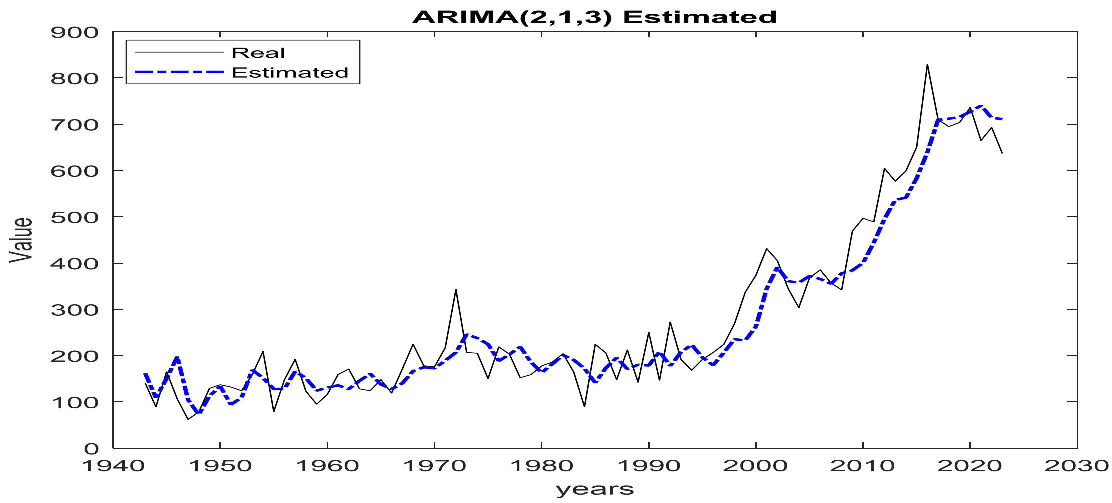

Figure (7): plotting the estimated and real values of the ARIMA model

P-ISSN 2518-5764
E-ISSN 2227-703X

Table (3): (AIC) Values for the Proposed Models

Model	AIC	Model	AIC	Model	AIC	Model	AIC
ARIMA(2,1,3)	4.674	ARIMA(5,1,3)	14.776	ARIMA(5,1,2)	17.295	ARIMA(0,1,5)	18.969
ARIMA(4,1,5)	9.077	ARIMA(4,1,4)	15.301	ARIMA(0,1,2)	17.709	ARIMA(3,1,2)	19.690
ARIMA(5,1,4)	9.801	ARIMA(3,1,5)	15.308	ARIMA(4,1,0)	17.895	ARIMA(2,1,5)	19.769
ARIMA(5,1,5)	10.825	ARIMA(4,1,2)	15.318	ARIMA(1,1,1)	18.078	ARIMA(2,1,0)	19.832
ARIMA(3,1,4)	13.332	ARIMA(5,1,1)	15.411	ARIMA(3,1,1)	18.191	ARIMA(2,1,4)	20.160
ARIMA(4,1,1)	13.493	ARIMA(0,1,1)	16.078	ARIMA(1,1,4)	18.373	ARIMA(0,1,4)	20.414
ARIMA(1,1,5)	14.662	ARIMA(3,1,0)	16.216	ARIMA(2,1,1)	18.396	ARIMA(3,1,3)	20.466
ARIMA(4,1,3)	14.697	ARIMA(2,1,2)	16.734	ARIMA(0,1,3)	18.573	ARIMA(1,1,0)	29.885
ARIMA(1,1,2)	14.738	ARIMA(1,1,3)	16.736	ARIMA(5,1,0)	18.580		

Table (4): Error values and RMSE value for the (ARIMA) model

years	Error	years	Error	years	Error	years	Error
1943	-20.8	1964	-36.7	1985	83.2	2006	19.8
1944	-20.0	1965	10.3	1986	31.9	2007	1.7
1945	14.6	1966	-8.7	1987	-47.2	2008	-35.9
1946	-91.6	1967	30.9	1988	41.0	2009	84.8
1947	-42.6	1968	58.3	1989	-37.8	2010	96.5
1948	5.8	1969	3.1	1990	71.4	2011	44.5
1949	16.3	1970	3.3	1991	-61.9	2012	109.0
1950	3.1	1971	27.7	1992	94.9	2013	40.4
1951	38.0	1972	136.0	1993	-15.1	2014	57.6
1952	14.6	1973	-39.1	1994	-55.2	2015	67.2
1953	-10.1	1974	-32.9	1995	-4.7	2016	190.1
1954	57.6	1975	-74.3	1996	27.2	2017	0.2
1955	-48.8	1976	30.5	1997	17.6	2018	-17.1
1956	19.8	1977	0.1	1998	33.6	2019	-12.1
1957	26.1	1978	-66.8	1999	104.0	2020	9.0
1958	-28.5	1979	-27.0	2000	110.7	2021	-75.2
1959	-29.1	1980	14.4	2001	86.8	2022	-21.1
1960	-15.0	1981	4.3	2002	13.9	2023	-74.4
1961	23.3	1982	3.0	2003	-15.8	**RMSE=53.56**	
1962	43.2	1983	-26.4	2004	-54.0		
1963	-17.5	1984	-82.9	2005	-4.0		

4.2 Hybrid (EVDHM-ARIMA) Model Analysis:

After conducting the process of detecting the non-stationary of the time series of the research data as a first stage, it is included in the form of a matrix and then this matrix is analyzed to obtain components that are stationary or close to stationary by using the (EVDHM) method, then the values are analyzed (SS Value) and (P-Value) for the Phelps-Perron test (PPT) to determine the selected components. Then, the process of determining the optimal values for the parameters of the ARIMA models is carried out for each identified component in order for these models to be forecasted. As a final stage, these forecasting are combined to obtain a composite forecasting result for the time series.

4.2.1 Eigenvalue Decomposition of Hankel Matrix (EVDHM):

The time series data was included in the form of a Hankel matrix according to the formula (5) and then it was analyzed into two matrices, the first for eigenvalues and the second for eigenvectors, where the number of rows reached (41) rows. These are equal to the number of columns and equal to the number of eigenvalues. After that, the number of rows was chosen. Pairs of distinct eigenvalues to show us the number of decomposed components, amounting to (12) components, which are referred to as (CP1, CP2, CP3,..., CP12) as shown in Table (5), which shows that out of (12) components, (10) components appear to us. It is characterized by the characteristic of stationary, which is (CP3, CP4,..., CP12). As for the first and second components, they are characterized by non-stationary, and this is confirmed by the results of the (PPT) test for each component obtained after completing the process of applying (EVDHM) and as in the figure (8), so we apply the decomposition. The second of these two components (CP1, CP2), which results in (CP1-1) for the decomposition of the first component (CP1) and three components upon the decomposition of the second component (CP2), which are (CP2-1, CP2-2, CP2-3), and when applying the (PPT test) once again on these components, the results show that only the component (CP2-3) is characterized by stationary. The other three components (CP1-1, CP2-1, CP2-2) failed the test, that is, they are characterized by non-stationary, as shown in Figure (9) for the second iteration, so we apply the third decomposition of these components, which results in (CP1-11) regarding the decomposition of the component (CP1-1), and one component, which is (CP2-11), regarding the decomposition of the component (CP2-1), and three Components (CP2-21, CP2-22, CP2-23) regarding the decomposition of the component (CP2-2). When the PPT test is applied again to these components that were extracted, the results show that the two components (CP2-22, CP2-23) passed the test, meaning it is stationary. However, the components (CP1-11, CP2-11, and CP2-21) did not pass the test, meaning it is non-stationary, as shown in Figure (10).

4.2.2 Analysis of (p-Value, SS Value) for non-stationary components:

We relied on studying the values of (SS Value) to evaluate the strength of the decomposed components as shown in Table (5) using (EVDHM), where we note that the value of (SS Value) for the first component (CP1) and its decomposition to (CP1-1) for the second iteration and to (CP1-11). The third iteration is approximately (8×10^6) and its p-value is (0.99). In general, it is nonstationary directional component, and this is shown in the three graphs (8), (9) and (10), respectively, so we can choose (CP1) or (CP1-1), as they are almost the same component. As for the second component (CP2), which showed us two non-stationary components, namely (CP2-1) and (CP2-2), with regard to the decomposition of the component (CP2-1) into the component (CP2-11), it is also considered one component, and this is explained through Figures (9) and (10) which show that their SS Value is close to (8×10^4), and it is also a non-stationary component based on its p-Value, so it is possible to choose (CP2-1) or (CP2-11). As for the component (CP2-2), which resulted in the non-stationary component (CP2-21), its SS value reached approximately (5×10^3). That is, a percentage of (0.07%) of the value of the first decomposed component, whose value was approximately (8×10^6), that is, less than (2.0%), so further decomposition is not necessary for these components.

Table (5): (h-Value, p-Value, SS Value) for the components extracted using (EVDHM)

<table>
<tr><th>Iteration</th><th colspan="4">Iter1</th><th colspan="4">Iter2</th><th colspan="4">Iter3</th></tr>
<tr><th>Method</th><th>CP</th><th>h-Value</th><th>P-Value</th><th>SS Value</th><th>CP</th><th>h-Value</th><th>P-Value</th><th>SS Value</th><th>CP</th><th>h-Value</th><th>P-Value</th><th>SS Value</th></tr>
<tr><td rowspan="16">EVDHM</td><td>CP1</td><td>0</td><td>0.999</td><td>8.1034×10^{6}</td><td>CP1-1</td><td>0</td><td>0.999</td><td>7.8501×10^{6}</td><td>CP1-11</td><td>0</td><td>0.999</td><td>7.7621×10^{6}</td></tr>
<tr><td rowspan="5">CP2</td><td rowspan="5">0</td><td rowspan="5">0.732</td><td rowspan="5">0.1134×10^{6}</td><td>CP2-1</td><td>0</td><td>0.771</td><td>8.2091×10^{4}</td><td>CP2-11</td><td>0</td><td>0.618</td><td>7.8883×10^{4}</td></tr>
<tr><td rowspan="3">CP2-2</td><td rowspan="3">0</td><td rowspan="3">0.072</td><td rowspan="3">0.8587×10^{4}</td><td>CP2-21</td><td>0</td><td>0.189</td><td>5.4619×10^{3}</td></tr>
<tr><td>CP2-22</td><td>1</td><td>0.001</td><td>0.3439×10^{3}</td></tr>
<tr><td>CP2-23</td><td>1</td><td>0.001</td><td>0.3207×10^{3}</td></tr>
<tr><td>CP2-3</td><td>1</td><td>0.001</td><td>0.1271×10^{4}</td><td></td><td></td><td></td><td></td></tr>
<tr><td>CP3</td><td>1</td><td>0.001</td><td>0.0128×10^{6}</td><td colspan="8"></td></tr>
<tr><td>CP4</td><td>1</td><td>0.001</td><td>0.0089×10^{6}</td><td colspan="8"></td></tr>
<tr><td>CP5</td><td>1</td><td>0.001</td><td>0.0070×10^{6}</td><td colspan="8"></td></tr>
<tr><td>CP6</td><td>1</td><td>0.001</td><td>0.0073×10^{6}</td><td colspan="8"></td></tr>
<tr><td>CP7</td><td>1</td><td>0.001</td><td>0.0080×10^{6}</td><td colspan="8"></td></tr>
<tr><td>CP8</td><td>1</td><td>0.001</td><td>0.0032×10^{6}</td><td colspan="8"></td></tr>
<tr><td>CP9</td><td>1</td><td>0.001</td><td>0.0038×10^{6}</td><td colspan="8"></td></tr>
<tr><td>CP10</td><td>1</td><td>0.001</td><td>0.0016×10^{6}</td><td colspan="8"></td></tr>
<tr><td>CP11</td><td>1</td><td>0.001</td><td>0.0032×10^{6}</td><td colspan="8"></td></tr>
<tr><td>CP12</td><td>1</td><td>0.001</td><td>0.0030×10^{6}</td><td colspan="8"></td></tr>
</table>

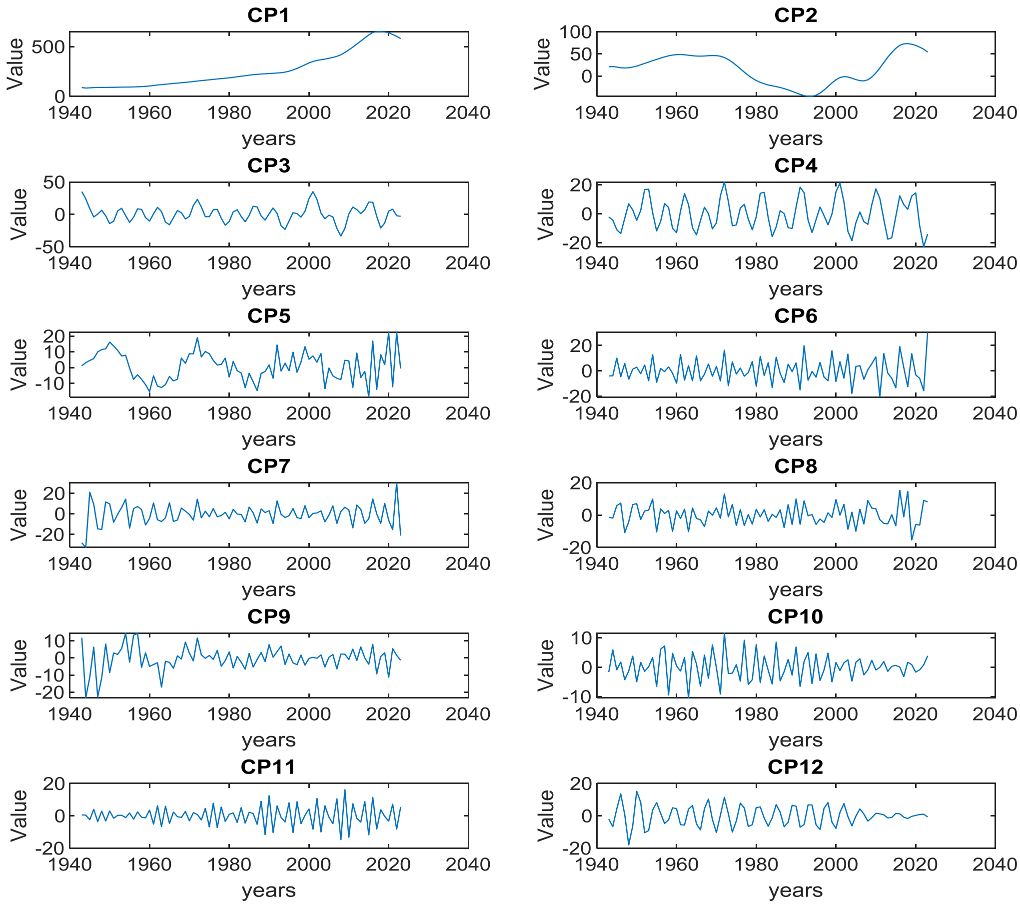

Figure (8): Drawing of the decomposed components (CP) for the first iteration

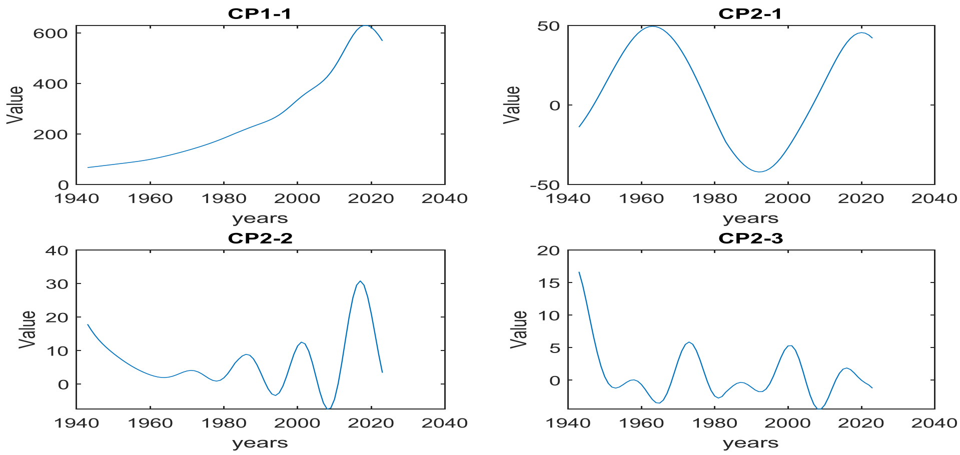

Figure (9): Drawing of the decomposed components (CP) for the second iteration

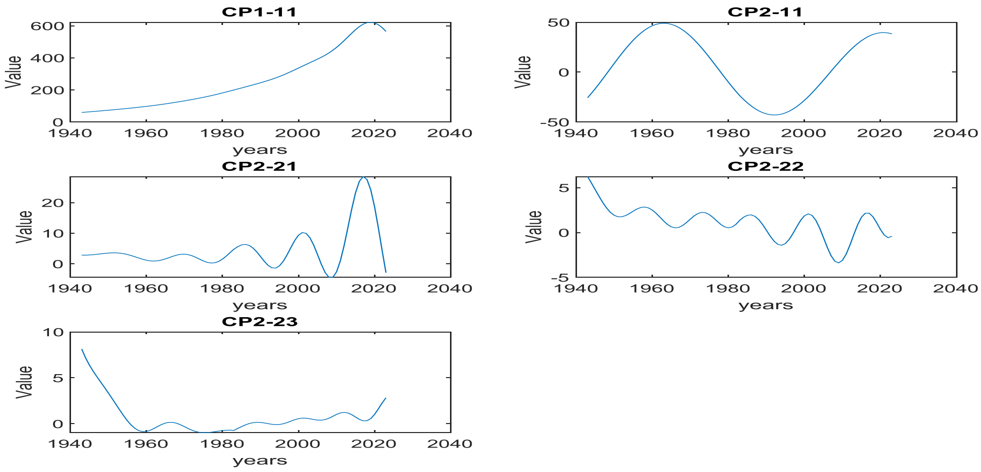

Figure (10): Drawing of the decomposed components (CP) for the third iteration

4.2.3 Diagnose the appropriate (ARIMA) models for the selected components using (EVDHM)

After identifying the decomposed components in building the model (EVDHM-ARIMA), the process of diagnosing the model for each component is carried out, where the optimal values for the ranks of the parameters (p,q) shown in Table (6) are summarized through the combination of a number of models starting from (0) to rank (5), taking the first difference (d=1) for components characterized by non-stationary , the best models were selected according to the lowest value of the rank determination criterion (AIC) as follows:

Table (6): Minimum value of the AIC standard for the ARIMA models (p, d, q) for components selected using (EVDHM)

Component	p	d	q	AIC	Component	p	d	q	AIC
CP3	4	0	5	366.23	CP11	5	0	5	134.01
CP4	4	0	5	367.90	CP12	5	0	5	237.72
CP5	5	0	4	417.35	CP1-1	3	1	4	49.20
CP6	5	0	5	434.62	CP2-1	2	1	1	-119.17
CP7	4	0	4	499.86	CP2-3	5	0	0	-182.11
CP8	5	0	4	455.11	CP2-21	4	1	5	-316.75
CP9	5	0	5	423.97	CP2-22	4	0	5	-334.39
CP10	5	0	5	318.25	CP2-23	5	0	0	-265.58

4.2.4 Forecasting using the model (EVDHM-ARIMA)

At this stage, and using personalized models, we calculated (81) estimated values within the range of series limits for each component. Then we collected these estimates to obtain a single time series of estimated values, while calculating the error values for the hybrid model, and calculating the RMSE value for these estimates, as shown in the table (7) The following:

Table (7): Error values and RMSE value for the (EVDHM-ARIMA) model

years	Error	years	Error	years	Error	years	Error
1943	51.5	1964	-23.9	1985	4.5	2006	-14.8
1944	25.8	1965	2.2	1986	11.5	2007	-13.7
1945	45.7	1966	3.4	1987	-24.1	2008	-15.3
1946	-7.4	1967	-9.1	1988	8.4	2009	12.3
1947	28.8	1968	22.1	1989	-37.4	2010	3.7
1948	13.3	1969	-17.4	1990	-2.7	2011	-15.0
1949	-22.2	1970	7.9	1991	-37.2	2012	36.9
1950	12.2	1971	13.4	1992	-1.6	2013	12.2
1951	23.7	1972	30.5	1993	-3.3	2014	-8.4
1952	-3.6	1973	12.8	1994	-14.4	2015	27.8
1953	-1.7	1974	27.4	1995	-21.0	2016	48.1
1954	34.6	1975	-6.7	1996	-50.8	2017	33.4
1955	8.2	1976	1.2	1997	-33.8	2018	18.6
1956	12.7	1977	26.7	1998	-10.2	2019	12.7
1957	35.6	1978	2.2	1999	16.7	2020	12.7
1958	-12.4	1979	18.5	2000	1.4	2021	10.2
1959	0.9	1980	2.5	2001	32.8	2022	72.6
1960	7.8	1981	-20.0	2002	-0.2	2023	19.9
1961	3.4	1982	12.9	2003	-24.6	**RMSE=22.41**	
1962	9.0	1983	-7.8	2004	-7.1		
1963	5.1	1984	-36.0	2005	-7.2		

Through comparison, it was found that using the EVDHM-ARIMA model is better in the forecasting process for data series than the single ARIMA model because it has the lowest value of the RMSE index. This model was used for Forecasting (10) observations, that is, for the period (2024-2033). It is shown in Table (8) and the real, estimated and forecasting values are presented in Figure (11) as follows:

Table (8): Forecasting values using the model (EVDHM-ARIMA)

years	Forecast	years	Forecast
2024	586.9	2029	415.8
2025	581.7	2030	423.3
2026	575.5	2031	323.6
2027	448.1	2032	309.9
2028	419.7	2033	317.9

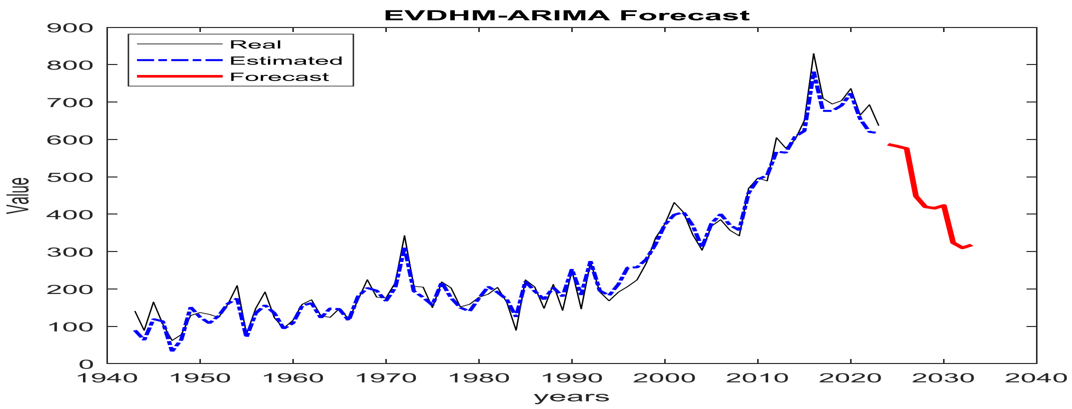

Figure (11): plotting the estimated, real, and Forecasted values of the EVDHM-ARIMA model

5. Discussion of Results:

The results indicate that the hybrid model outperforms the individual model with a high percentage in the forecasting process for research data because it has the lowest value of the RMSE index. The value of (RMSE) decreased by (31.15), after its value for the estimates was obtained from the single ARIMA (2, 1, 3) model which was equal to (53.56). The forecasting results of the hybrid model (EVDHM-ARIMA) also show that the average dunum yield for the area harvested for the wheat crop will witness a noticeable decline in the next few years, as the average dunum yield decreases from (636.6) kg/dunum in 2023 to (317.9) kg/dunum in year 2033, this is due to many reasons, the most important of which are the date method of planting, the variety grown, fertilization, irrigation, control of cultivated land, agricultural experience, harvesting methods, and others.

6. Conclusions:

We conclude that the time series of the average yield of the harvested area of wheat in Iraq takes a non-stationary course in the variance and the mean, i.e. it has a general trend, and after taking the natural logarithm and the first difference filter, the time series becomes stationary. The single model ARIMA (2, 1,3) was diagnosed after matching a set of models to the time series observations based on the lowest value of the evaluation criterion (AIC) achieved by this model. The decomposed results showed found that there are (16) decomposed components subject to the forecasting process, of which (13) are stationary components that do not take the difference filter, i.e. (d=0) and (3) components for which the first difference filter (d=1) is taken when applying the (EVDHM-ARIMA) model to the time series. After obtaining the forecasting for these components and collecting them together, it is clear that using the (EVDHM-ARIMA) model gives more reliable estimates than what we obtained when using the single model (ARIMA) for the research data, as it gave the lowest value (RMSE).

Authors' individual contribution: Conceptualization — H.H.F.; Methodology — H.H.F.; Formal Analysis — H.H.F.; Investigation — H.H.F. and W.K.; Data Curation — H.H.F.; Writing —Original Draft — H.H.F.; Writing — Review & Editing — W.K., and H.H.F.; Visualization — H.H.F.; H.A.H. & G.A.K.A.; Supervision —W.K., and H.H.F.; Project Administration —H.H.F.

Declaration of conflicting interests: The Authors declare that there is no conflict of interest.

Data Availability Statement

Data for the article can be found at:
https://docs.google.com/document/d/1H0_i9GfsSlOC6_mlHJ-9qJgu-vAQkQCK/edit?usp=sharing&ouid=110333825232813259985&rtpof=true&sd=true.

Authors Declaration:

Conflicts of Interest: None

-We Hereby Confirm That All The Figures and Tables In The Manuscript Are Mine and Ours. Besides, The Figures and Images, which are Not Mine, Have Been Permitted Republication and Attached to The Manuscript.

- Ethical Clearance: The Research Was Approved by The Local Ethical Committee in The University.

Refrences

Abdullah, H. N., & Khalil, M. H. (2021). The Impact of Some Economic Variables on the Production of Wheat Crop in Iraq for the Period(2004-2019). *Tikrit Journal of Administrative and Economic Sciences*, *17*(55 part 3), 397–413. https://www.iasj.net/iasj/article/218544

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Albasri, A. H. M. (2020). Using ARIMA models to forecast the volume of cargo handled in Iraqi ports An applied study in the general company of Iraqi ports. *Journal of Economics And Administrative Sciences*, *26*(120), 452–474. https://doi.org/10.33095/jeas.v26i120.1927

Aradhye, G., Rao, A. C. S., & Mastan Mohammed, M. D. (2019). A novel hybrid approach for time series data forecasting using moving average filter and ARIMA-SVM. *Advances in Intelligent Systems and Computing*, *813*, 369–381. https://doi.org/10.1007/978-981-13-1498-8_33

Bisgaard, S., & Kulahci, M. (2011). Time Series Analysis and Forecasting by Example. In *Choice Reviews Online* (Vol. 49, Issue 06). Wiley. https://doi.org/10.1002/9781118056943

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control (5th ed)*. John Wiley & Sons. http://library.lol/main/BB617EAC4CB4EC545575A49DBD7825DD

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7*(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

Ghosh, A. K., Das, S., Dutta, S., & Mukherjee, A. (2023). Sensing of Particulate Matter (PM 2.5 and PM 10) in the Air of Tier 1, Tier 2, and Tier 3 Cities in India Using EVDHM-ARIMA Hybrid Model. *IEEE Sensors Letters*, *7*(5), 1–4. https://doi.org/10.1109/LSENS.2023.3270905

Habeeb, A. S. (2020). Compare prediction by autoregressive integrated moving average model from first order with exponential weighted moving average. *Journal of Economics and Administrative Sciences*, *26*(120), 426–439. https://doi.org/10.33095/jeas.v26i120.1925

Hamel, A. A., & Abdulwahhab, B. I. (2022). Using a hybrid SARIMA-NARNN Model to Forecast the Numbers of Infected with (COVID-19) in Iraq. *Journal of Economics and Administrative Sciences*, *28*(132), 118–133. https://doi.org/10.33095/jeas.v28i132.2276

Harris, R. I. D. (1992). Testing for unit roots using the augmented Dickey-Fuller test. Some issues relating to the size, power and the lag structure of the test. *Economics Letters*, *38*(4), 381–386. https://doi.org/10.1016/0165-1765(92)90022-Q

Jain, P., & Pachori, R. B. (2014). Event-based method for instantaneous fundamental frequency estimation from voiced speech based on eigenvalue decomposition of the Hankel matrix. *IEEE Transactions on Audio, Speech and Language Processing*, *22*(10), 1467–1482. https://doi.org/10.1109/TASLP.2014.2335056

Kazom, J. M., & Mohammed, I. (2016). Box and Jenkins use models to predict the numbers of patients with hepatitis Alvairose in Iraq. *Journal of Economics And Administrative Sciences*, *22*(89), 407–424. https://doi.org/10.33095/jeas.v22i89.628

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297–303. https://doi.org/10.1093/biomet/65.2.297

Mousa, M. A., & Mohammed, F. A. (2020). Applying some hybrid models for modeling bivariate time series assu ming different distributions for random error with a practical application. *Journal of Economics and Administrative Sciences*, *26*(117), 442–479. https://doi.org/10.33095/jeas.v26i117.1824

Nagvanshi, S. S., Kaur, I., Agarwal, C., & Sharma, A. (2023). Nonstationary time series forecasting using optimized-EVDHM-ARIMA for COVID-19. *Frontiers in Big Data*, *6*. https://doi.org/10.3389/fdata.2023.1081639

Pachori, R. B. (2023). Time-Frequency Analysis Techniques and their Applications. In *Time-Frequency Analysis Techniques and their Applications*. CRC Press. https://doi.org/10.1201/9781003367987

Phillips, P. C. B., & Perron, P. (1988). Testing for a Unit Root in Time Series Regression. *Biometrika*, *75*(2), 335–346. https://doi.org/10.2307/2336182

Rashid, H. K., & Abdulrahman, S. A. (2018). Comparison of the statistical methods used to Forecast the size of the Iraqi GDP for the two sectors (public and private) for the period (2025-2016). *Journal of Economics And Administrative Sciences*, *24*(107), 590–613. https://doi.org/10.33095/jeas.v24i107.1314

Sharma, R. R., Kumar, M., Maheshwari, S., & Ray, K. P. (2021). EVDHM-ARIMA-Based Time Series Forecasting Model and Its Application for COVID-19 Cases. *IEEE Transactions on Instrumentation and Measurement*, *70*(December). https://doi.org/10.1109/TIM.2020.3041833

Sharma, R. R., & Pachori, R. B. (2017). A new method for non-stationary signal analysis using eigenvalue decomposition of the Hankel matrix and Hilbert transform. *International Conference on Signal Processing and Integrated Networks*, *November 2018*, 484–488. https://doi.org/10.1109/SPIN.2017.8049998

Sharma, R. R., & Pachori, R. B. (2018a). Baseline wander and power line interference removal from ECG signals using eigenvalue decomposition. *Biomedical Signal Processing and Control*, *45*(May), 33–49. https://doi.org/10.1016/j.bspc.2018.05.002

Sharma, R. R., & Pachori, R. B. (2018b). Time-frequency representation using IEVDHM-HT with application to classification of epileptic EEG signals. *IET Science*, *Measurement and Technology*, *12*(1), 72–82. https://doi.org/10.1049/iet-smt.2017.0058

Singh, V. K., & Pachori, R. B. (2022). Iterative Eigenvalue Decomposition of Hankel. *TechRxiv*, 1–9. https://doi.org/10.36227/techrxiv.21730487.v1

Strang, G. (2009). *Introduction to Linear Algebra, Fourth Edition [4th ed.]*. Wellesley Cambridge Press. http://library.lol/main/26CE0F0057C1FFF647491C51560DC5F2

Timm, N. H. (Ed). (2002). Applied Multivariate Analysis. In N. H. Timm (Ed.), *Springer*. New York, NY. https://doi.org/10.1007/b98963

Wazeer, E. A. A., & Hameed, L. M. A. (2022). Euro Dinar Trading Analysis Using WARIMA Hybrid Model. *Journal of Economics and Administrative* Sciences, *28*(131), 193–204. https://doi.org/10.33095/jeas.v28i131.2245

Wei, W., & W.S. (2006). *Time Series Analysis (2th ed). Univariate and Multivariate Methods*. Adison Westley.

Yin, X., Xu, Y., Sheng, X., & Shen, Y. (2019). Signal Denoising Method Using AIC–SVD and Its Application to Micro-Vibration in Reaction Wheels. *Sensors*, *19*(22), 5032. https://doi.org/10.3390/s19225032

Zhang, G., Wu, J., Pan, B., Li, J., Ma, M., Zhang, M., & Wang, J. (2017). Improving daily occupancy forecasting accuracy for hotels based on EEMD-ARIMA model. *Tourism Economics*, *23*(7), 1496–1514. https://doi.org/10.1177/1354816617706852