Employing the principal components in time series models and selecting the best models with Application

ID No. 552

(PP 434 - 448)

https://doi.org/10.21271/zjhs.27.2.26

Rizgar Maghded Ahmed
Department of Statistics, College of Administration and Economics, Salahaddin University-Erbil
rizgar.ahmed@su.edu.krd

Nida Salim Mala Younis
Department of Statistics, College of Administration and Economics, Salahaddin University-Erbil
nida.malayounis@su.edu.krd

Received: 01/09/2022
Accepted: 10/10/2022
Published: 17/04/2023

Abstract

In this study, principal components analysis, which is one of the methods of multivariate analysis for prediction of time series models (Box-Jenkins Model) was used by applying to electric power data (Erbil Gas Power Plant) (EGPS) which contains multivariate data (5 stations) and the data was monthly for the period from (1/1/2017) to (14/9/2021).

The idea of the research was based on applying principal component analysis to multiple time series data, obtaining the components extracted from them, and then estimating the Box-Jenkins Models.

The main conclusion is that principal component analysis is effective in reducing multiple time series data and obtaining the best models based on statistical criteria.

And finally, the best proposed model for predicting electrical energy production data in the City of Erbil is (ARIMA(2,2,2)x(2,2,0)12).

Keywords: Multivariate Analysis; Principal Component Analysis; Box-Jenkins model; prediction.

1. Introduction

Data forecasting is one of the important topics in scientific research and in many disciplines, as it is considered the cornerstone in determining and planning future policies, in this research, we discussed the issue of electrical energy, this research aims to predict multiple time series using Box-Jenkins methodology for multiple time series data after converting them into components based on the (PCA) with covariance matrix, where each component was predicted separately and then these models were developed In a multiple model that is processed to predict future data and compare it with the original data, and to achieve the objectives of the study, the descriptive analytical approach was used by describing the study variables and analyzing the results of the applied side, which was based on the statistical program (Statgraphics-19). In this research, the researchers tried diagnose the best way to predict the productivity of electrical energy in Erbil Governorate (EGPS) power station for the period from (2017 - 2021) using the method of principal components of the time series data.

2. The Study Area

The EGPS power station is located in (perdawd) village, Erbil Governorate, KRI, it contains (10) stations, (8) of which depend on gas and (2) of them depend on steam, and the city of Erbil depends on this station.

3. Methodology

In this section, the theoretical aspect of the research is presented, as well as how to link the analysis of the principal components with time series models

3.1 Principal Components Analysis:

The Principal Components Analysis is one of the branches of multivariate analysis and one of the important methods in studying a large number of variables, that is, those that pertain to a group of phenomena that are observed around a number of variables linked to each other by interrelationships, which is called multi-collinearity.

3.2 Principal Component Model

The Principal components model is so that the Eigen vectors are placed as factors in a linear combination of the studied random variables X_j, (j=1,2, 3,…,p), and it can be expressed as follows:

$$PC_j = a_{1j}X_1 + a_{2j}X_2 + \ldots + a_{pj}X_p \qquad \ldots(3.1)$$

$$PC_j = \sum_{k=1}^{P} a_{kj}X_k \qquad j,k = 1,2,\ldots,p \qquad \ldots(3.2)$$

Whereas:

PC_j: Principal Component j.

a_{kj} : the (k) parameter in the (j) component, these coefficients (a_{kj}) represent the values of the Eigen vectors (a_j) accompanying the Eigen roots (λ_i).(Dunteman, G.H., 1989.)

3.3 Time series

A time series is a set of observations of a particular phenomenon generated during time. These observations are characterized by being arranged according to their occurrence in time, and successive observations are usually not independent, that is, they depend on each other, and the lack of independence will be exploited in arriving at reliable predictions.

The time series is defined mathematically as a sequence of random variables defined within the multivariate probability space and its index is the index (t), which returns to an index set (T) and symbolizes the time series usually{Y(t); t∈T} or simply {Y(t)}. The aim of the analysis of the time series model is to understand its basic properties ((Trend (T), Cyclical (C), Seasonal (S) Irregular (I)), as well as using it to estimate and then predict the behavior of the time series in the future (Abdul, 2004), Observational time series (x_1, x_2,…, x_{n-1} , x_n) is stationary if the following conditions: (Bari, 2002)

1. Mean:

$$E(Y_t) = \mu \qquad \ldots.(3.3)$$

2. Variance:

$$var(Y_t) = E(Yt - \mu)^2 = \sigma_Y^2 \qquad \ldots(3.4)$$

3. The independence of the autocorrelation coefficients between (Y_s) , (Y_t)

$$E\left[(Y_t - \mu)(Y_s - \mu)/\sigma_Y^{2}\right] = \rho_{t-s} \qquad \ldots(3.5)$$

3.4 Autocorrelation Function(ACF)

The main statistical tool in the analysis of time series is the autocorrelation coefficient, also called series Correlation, and the term autocorrelation can be clarified on the basis that it represents the correlation between the sequential observations of the same variable during a period of time, and the content of autocorrelation is the fact that the random variable that occurs during A certain time period is related to the random variable that precedes it or follows it, i.e. the series correlation with itself or its creep by (1,2,3,…) period, and the general formula for calculating the autocorrelation of a phased series is:

$$\rho_k = \frac{E(Y_t - \mu_Y)(Y_{t+k} - \mu_Y)}{E(Y_t - \mu_Y)^2} \qquad \ldots(3.6)$$

Where (p_k) represents the autocorrelation with a Lag of k .

3.4.1 Partial autocorrelation Function (PACF)

It is used to measure the degree of correlation between (Y_{t-k} , Y_t)when the effect of delay time is (Time lags) (1,2,…,k-1) has been removed, PACF is defined as:

$$\phi_{11} = r_1 \qquad \ldots(3.7)$$

$$\phi_{kk} = \frac{r_k - \sum_{j=1}^{k-1} \phi_{k-1}\, r_{k-1,j}\, r_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j}\, r_j} \quad \ldots (3.8)$$

Where :

$$\phi_{kj} = \phi_{k-1,j} - \phi_{kk}\, \phi_{k-1,k-j} \quad \ldots (3.9)$$

3.5 Box-Jenkins Models in Time Series

3.5.1 Autoregressive Model (AR(p))

In this model the value of the variable in the current period (Y_t) depends on its value in the previous periods ($Y_{(t-1)}$)The autoregressive model is a linear regression of the time series values (dependent variable) with one or more of the previous values of the time series as variables Not supported (independent variables) and is denoted by the symbol AR (p). The model can be written in the following form: (Karakaş, 2019)

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + a_t \quad \ldots (3.10)$$

whereas:

(μ) : Constant term ($-\infty<\mu<\infty$) .

(p) : rank of the model.

(Y_t) time series observations at time t.

($\emptyset$): Model parameters, ($-1 < \emptyset < 1$).

(a_t): The random error (white noise) is distributed normally $a_t \sim WN(0,\sigma^2)$

3.5.2 Moving Average Model (MA(q))

The time series can be represented by the moving averages model, and it expresses the current value of the series (Y_t) in terms of the weighted sum of the previous values of the errors, meaning that it depends on the previous errors to represent the time series and is denoted by the symbol MA (q) and written in the following formula: (Karakaş, 2019)

$$Y_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \quad \ldots (3.11)$$

3.5.3 Integrated Mixed Model (ARIMA)

ARIMA models for an unstable linear time series, in the event that the time series is unstable on average, and so we make transformations by taking the differences and converting them to a stable series, and the differences are taken with positive integers (d), and ARIMA models (p, d, q)) are considered The most used time series models in the process of forecasting future values, in the Arima model, the future value of the variable is a linear combination of past values and past errors, and its formula is as follows: (Chakravarti et al., 1973) (Zhang, 2003) (Mumbare et al., 2014) (Wang, Y) (Tian, S 2018)

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + a_t - \theta_1 Y_{t-1} - \theta_2 Y_{t-2} - \cdots - \theta_q Y_{t-q} \quad \ldots (3.12)$$

Y_t: the Actual values.

a_t : Random errors.

$\emptyset, \theta$: Represent the parameters of the model. p, q: represents the model's rank.

3.5.4 Multiplicative Seasonal ARIMA Models

In some cases, all previous seasonal and non-seasonal models are combined to form a model that may be the best in data analysis. The seasonal autoregressive (SAR) models and the Seasonal moving averages (SMA) model are considered the multiplication model. As a product of the components of the four time series (general trend, seasonal changes, cyclical changes, episodic changes), this model assumes that the four factors interact with each other and do not move independently ARMA(p,q)x(P,Q) and its formula is as follows: (Guo, J. 2009) (Shu, 2005) (Tran 2015) (Miao 2014)

$$\emptyset(B)\Phi(B^S)Y_t = \theta(B)\Theta(B^s)a_t \qquad \ldots(3.13)$$

The general formula of the Seasonal Autoregressive Integrated Moving Average SARIMA(p,d,q)x(P,D,Q) is as follows:

$$\phi_p(B)\Phi_p(B^s)\nabla^d\nabla_s^D Y_t = \theta_q(B)\Theta_Q(B^S)a_t \qquad \ldots(3.14)$$

3.5.5 Modeling Procedure:

Box–Jenkins, 1976, proposed a method for analyzing time series data consists of four steps:

i) Model identification.

ii) Estimation of model parameters.

iii) Diagnostic checking for the identified model.

v) application of the model in forecasting purposes.[3]

3.6 Employ Principal Components in Box-Jenkins Model

The following method has been employed to find the final model adopted in prognosis using the principal components (Ladalla,2000):

$$X_t P = W_t \qquad \ldots(3.15)$$

Where:

W_t : Principal component matrix k*k, k=1,2,3…

X_t : Matrix of random variables k*k, k=1,2,3…

P : Matrix of eigen Vectors k*k, k=1,2,3…

From equation (3.15) the original data can be obtained X_t:

$$X_t = W_t P^T \qquad \ldots(3.16)$$

P^T : Transpose of the matrix P.

To simplify, suppose we have the following model:

$$W_{j,t} = \varphi_{jt} W_{j,t-1} + a_{j,t} \qquad \ldots(3.17)$$

$W_{j,t}$: Tracing Dynamic model

φ_{jt} : Parameter of Dynamic model.

$a_{j,t}$: : Random errors distributed $N(0, \sigma_j^2)$

From the equations (3.15) (3.17), we get:

$$X_t = \varphi_1 X_{t-1} + a_t \qquad \ldots(3.18)$$

φ_1: Parameter of the model is represented in the form of a matrix represented by the following equation:

$$\varphi_1 = p\lambda_i p^T \qquad \ldots(3.19)$$

and that:

$$\lambda_i = diag(\varphi_{1,1}, \varphi_{2,1}, \ldots.. \varphi_{k,1}) \qquad \ldots(3.20)$$

and that:

$$a_t \sim N(0, \Sigma)$$

$$\Sigma = pDp^T \qquad \ldots(3.21)$$

$$D = diag(\sigma_1^2, \sigma_1^2, \ldots\ldots \sigma_k^2)$$

By knowing the model that the series follows, it is possible to predict future periods and know the efficiency of the models using statistical criteria that the researcher can determine.

4-Applications

4-1. Introduction:

Implementation of the practical applied aspect requires obtaining data that is a basic pillar in reaching reliable results, and this is done through the quality of data that achieves the applied aspect.

The application methodology in this research was represented in studying the analysis of the principal components of time series and applying the Box-Jenkins models (ARIMA) where data were obtained on the productivity of electrical energy. With monthly productivity for a year, (5) units of production, with (57) views for each series, and the results were extracted using (STATGRAPHICS 19).

4-2. Data Analysis: Data analysis includes a number of steps:

4-2-1. Identification

The identification process is carried out first by test the time series, is it stationary with the mean and variance, and through the diagram we can see the nature of the fluctuation in it, where the data of the monthly produced quantity of electricity as shown in Figure (1), To test the mean and variance of the time series (5) were plotted with time as follows:

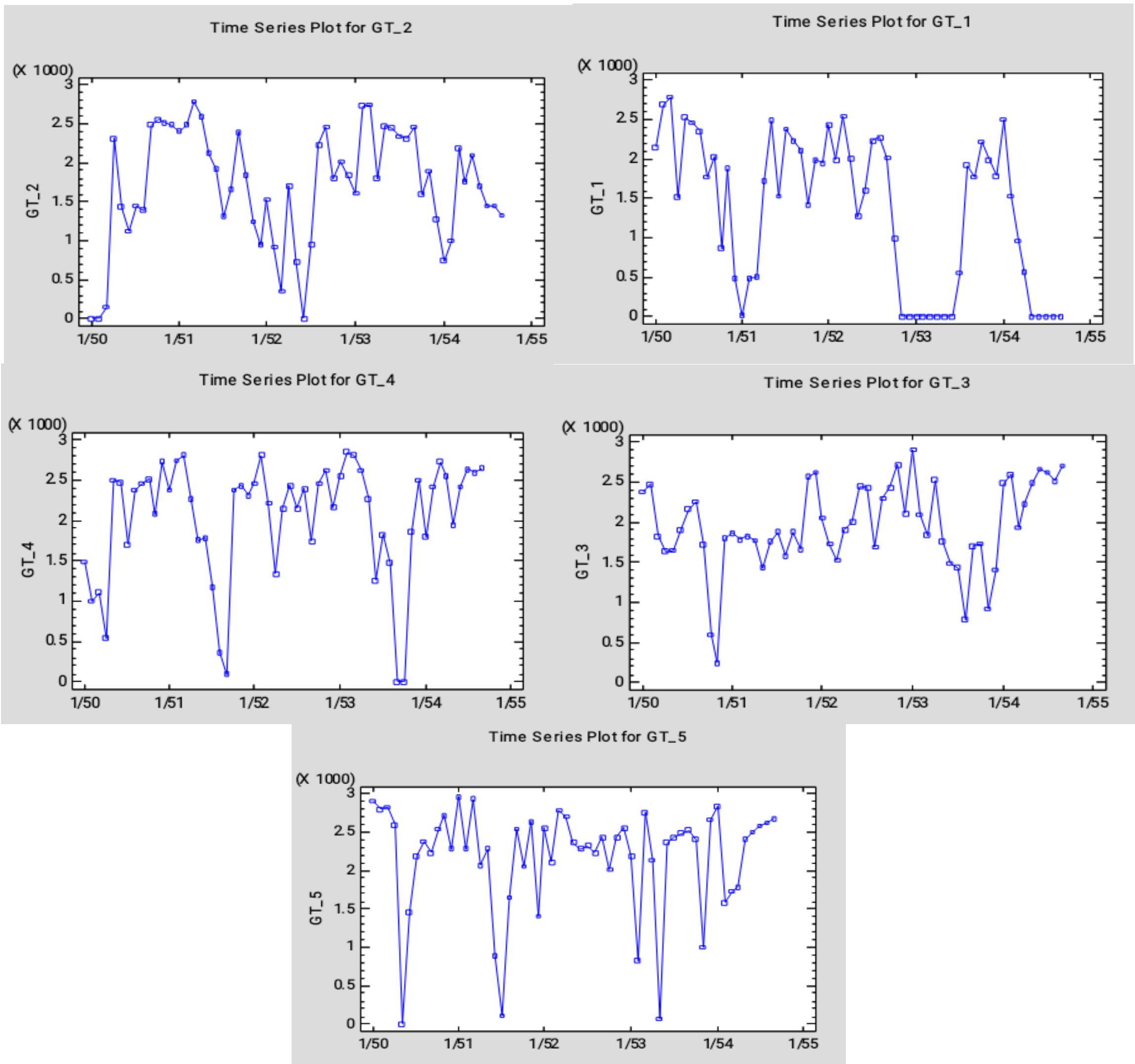


Figure (1): Time Series Plots

It is noticed from the previous figures that the time series are non-stationary in the mean and variance, and it can be ascertained about the progress of the series by observing the autocorrelation (2) and partial autocorrelation (3) functions of the series (5) and as follows:

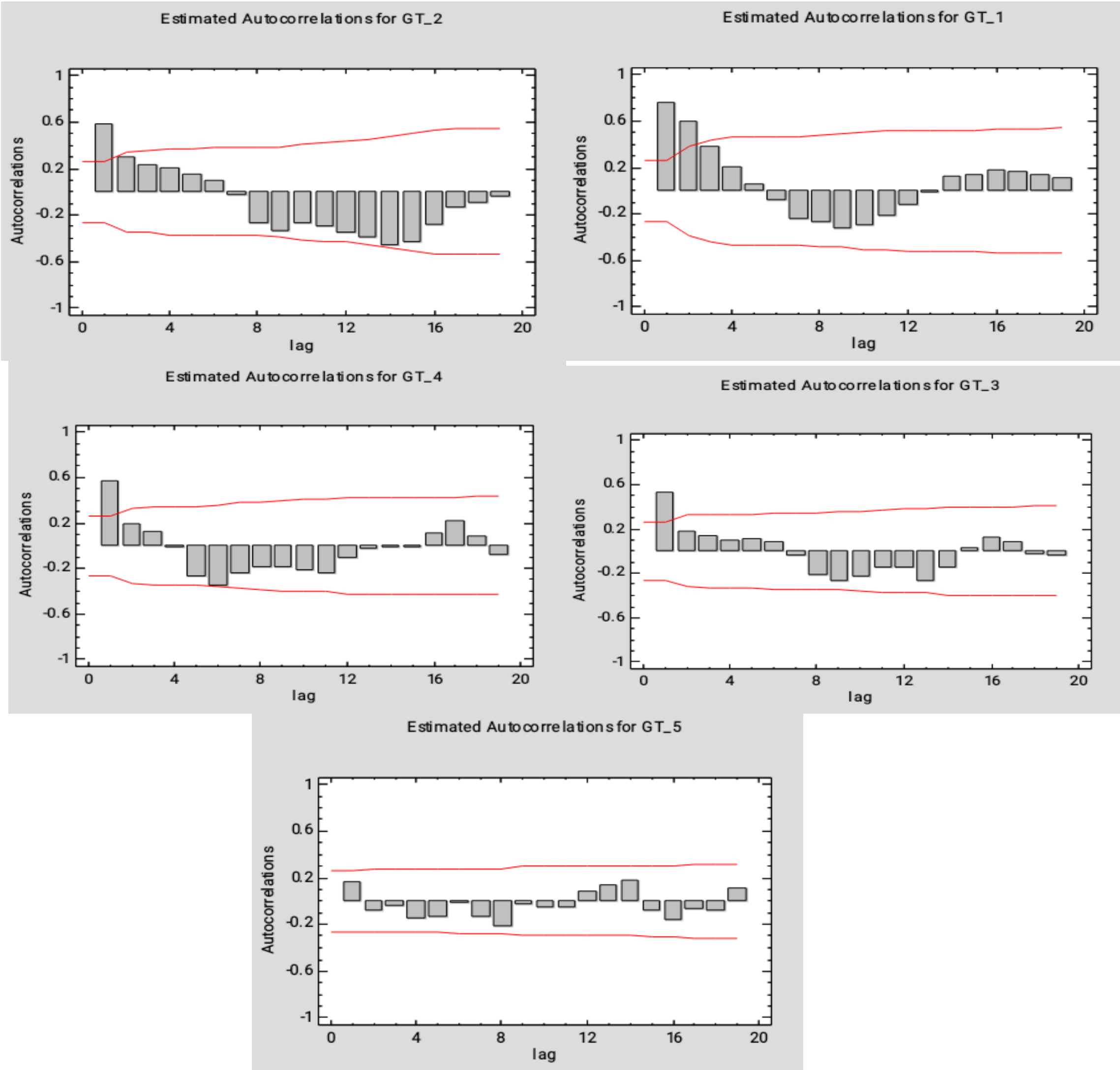
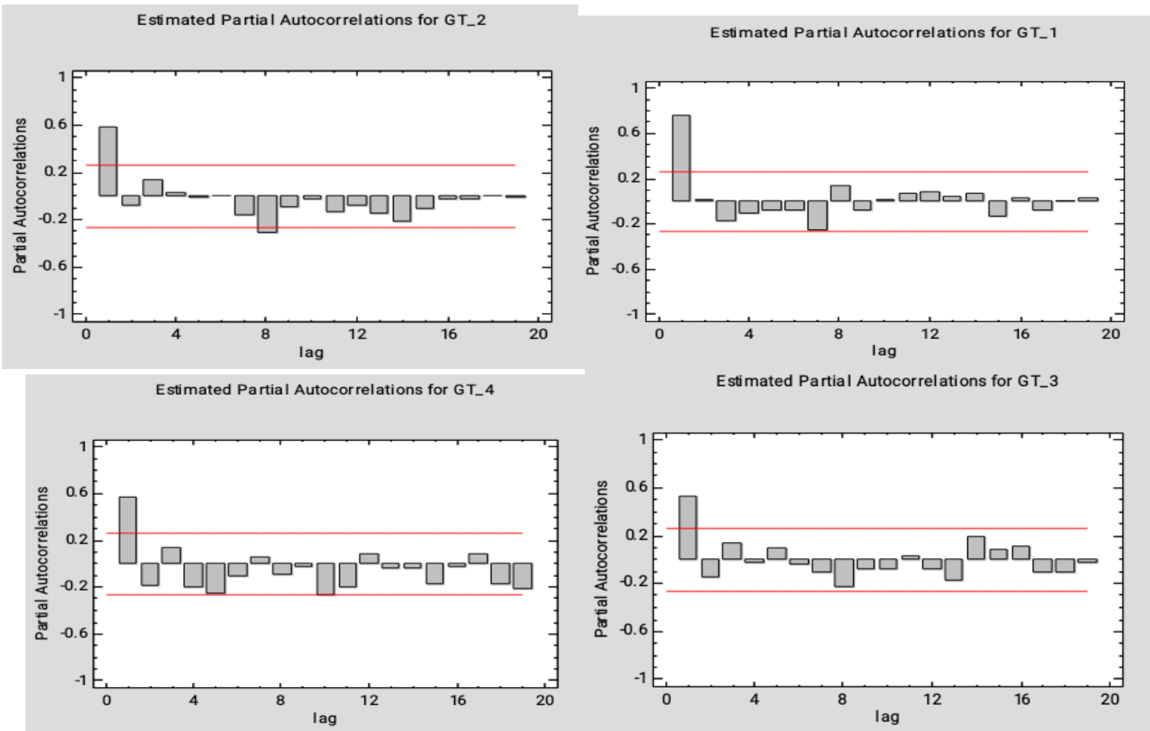


Figure (2): Autocorrelations



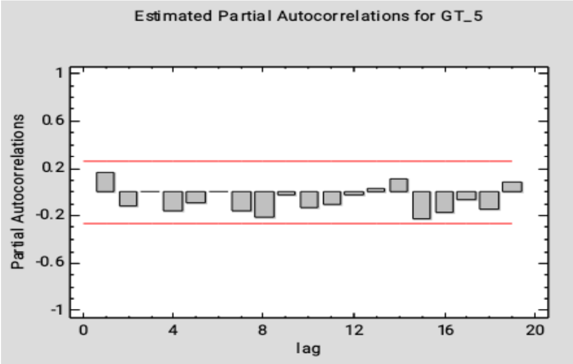


Figure (3): Partial Autocorrelations

4-2-2. Principal Component Analysis

To reduce the number of studied variables and without losing a large amount of information, principal component analysis was relied on to analyze a certain number of variables, we note in table (1) and Figure (4) that (3) principal components were extracted because (the eigenvalue has greater than the correct one) with an interpretation rate of 82.976%, which is a high percentage.

Table (1): Explain Total Variance

Component Number	Eigenvalue	Percent of Variance	Cumulative Percentage
1	1.68314	33.663	33.663
2	1.45174	29.035	62.698
3	1.01395	20.279	82.976%
4	0.622194	12.444	95.420
5	0.228982	4.580	100.000

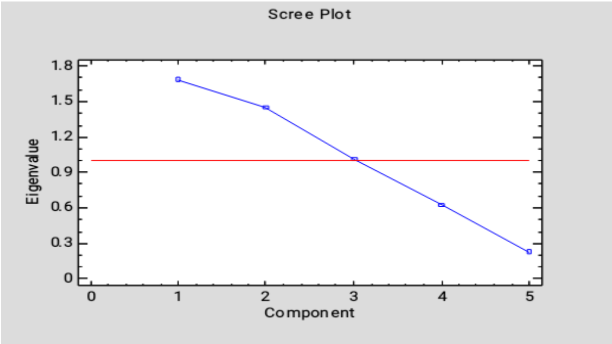


Figure (4): Scree Plot

4-2-3. Application using ARIMA models on the extracted components:

In this section, ARIMA models were used on the extracted data of the three principal components:

4-2-3-1 Application using ARIMA models on the first component:

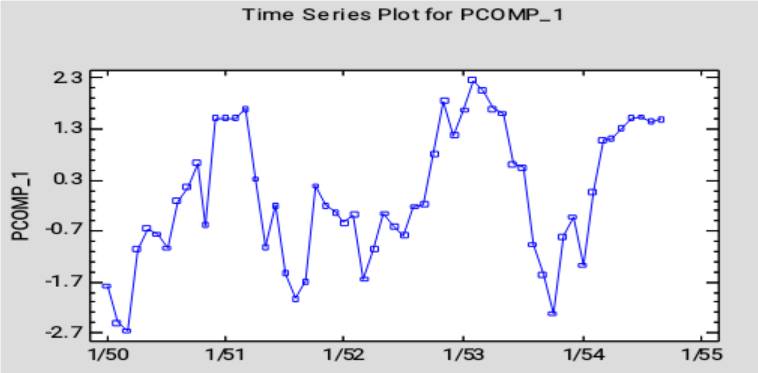


Figure (5): time series plot for first component

We notice in the figure (5) that the series is non-stationary, which indicates that the mean is non-stationary over the time, to know and make sure that the series is stationary or not? the autocorrelation function and the partial autocorrelation function of the chain were examined as in the two figures (6) & (7):

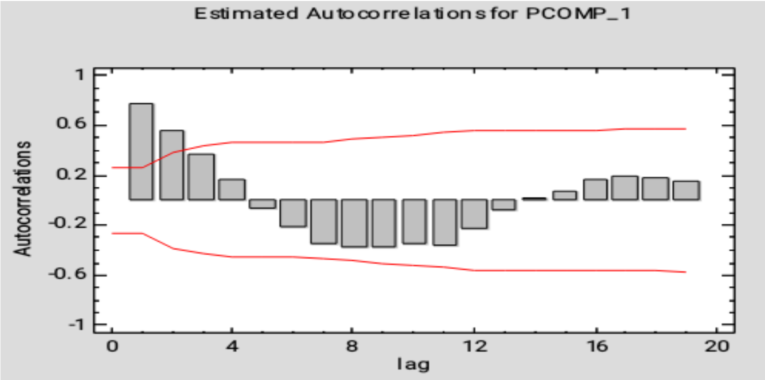


Figure (6): ACF for first component

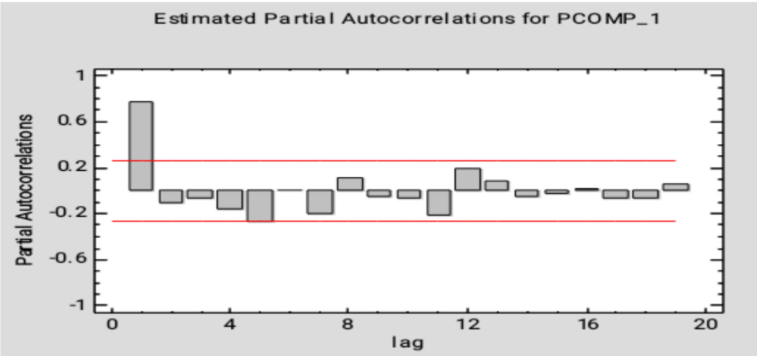


Figure (7): PACF for first component

We note that the ACF autocorrelation function is that the first shifts in the values of the autocorrelation coefficients are significant outside the confidence limits (∓0.259), and only the partial autocorrelation function PACF is outside the confidence limits.

4-2-3-2 Randomization test:

The ljung-box test was used to test the hypothesis shown in Table (6). Since the $p < 0.05$, we reject the null hypothesis and this indicates that the time series of PC1 is not random.

Table (6): The ljung box test to test the randomness of the time series of PC1

Hypothesis testing	test statistic	P-value
H_0: time series is random H_1: time series is not random	129.72	0.000

4-2-3-3 Achieving Stationary for the time series (PC1):

It was found through the drawing of the original time series for the PC1 that it is non-stationary, and to make the series stationary around the variance and about the mean, it also contains seasonal effects, as it was found that the values in the periods (24,12) are repeated, which indicates that the time series is seasonal and that it repeats itself every (12) months. After several attempts to make the time series stable, we first take the first seasonal difference and the model is doubly seasonal as in Figure (8), the autocorrelation coefficients (ACF) and partial autocorrelation (PACF) were drawn, as most The values are within confidence limits as shown:

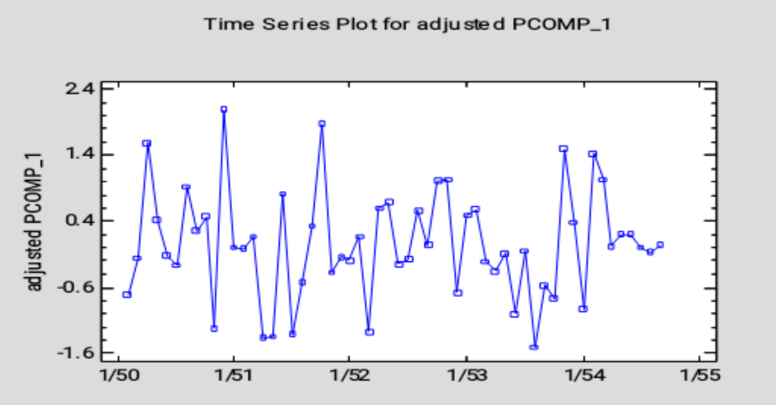


Figure (8): plot the first component time series transformed (first seasonal difference)

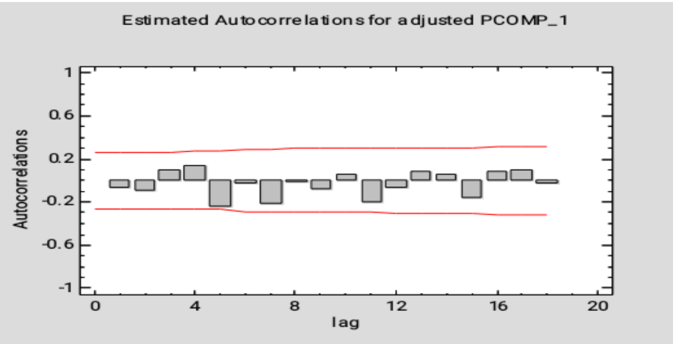


Figure (9): plot of the ACF of the transformed series (first seasonal difference)

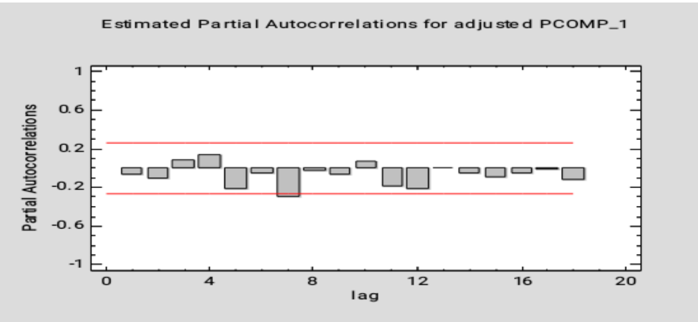


Figure (10): plot of the PACF of the transformed series (first seasonal difference)

After drawing the autocorrelation function (ACF) and the partial autocorrelation function (PACF), we notice that each of the two curves, the values of their coefficients, decrease gradually with the increase of the displacement periods and fall within the confidence range so that the time series is considered stable.

4-2-3-4 Randomization test after taking the first difference:

We note in the table (7) that after taking the necessary transformations (first seasonal difference) the data of series PC1 is random.

Table (7): ljung & box test to test the randomness of the modified time series after taking the first seasonal difference.

Hypothesis testing	test statistic	p-value
H_0: time series is random H_1: time series is not random	16.8533	0.533209

4-2-3-5 Choosing the appropriate model for the time series:

In this step, the appropriate model is determined after we have stationary in the time series of the first component and determined its degree by studying the behavior of the ACF and the PACF, (5) models were taken, in order to choose the appropriate model for the data that has the lowest value for many (criteria) that were applied to them (RMSE), Akaike Information Criterion (AIC), Baysian Information Standard (BIC) and Hanan-Queen Standard (HQC) as in the table(8):

Table (8): Suggested models for the time series for the first component

Model	RMSE	AIC	HQC	SBIC
ARIMA(2,2,2)x(2,2,0)^{12}	0.180955	-3.20849	-3.12491	-2.99343
ARIMA(2,2,0)x(2,2,0)^{12}	0.259764	-2.55561	-2.49989	-2.41224
ARIMA(2,2,1)x(2,2,0)^{12}	0.257681	-2.53663	-2.46698	-2.35741
ARIMA(2,2,0)x(2,2,1)^{12}	0.268249	-2.45624	-2.38659	-2.27702
ARIMA(2,2,1)x(2,2,1)^{12}	0.266173	-2.43669	-2.35311	-2.22164

The appropriate model is the first model, ARIMA (2,2,2)x(2,2,0)^{12}, which was chosen based on the Akaike Information Criterion (AIC), Hanan Quinn (HQC) and Schwartz Bees Standard (SBIC).

4-2-3-5 Estimation:

After determining the appropriate model, the parameters of the ARIMA model (2,2,2)x(2,2,0)^{12} were estimated, in order to know the significance of the estimated parameters as in Table (9), We note that all the estimated parameters of the appropriate model are significant ($p<0.001$):

Table (9): Estimation of Parameters Values for ARIMA Model (2,2,2)x(2,2,0)^{12}

Parameter	Estimate	Stnd. Error	t	P-value
AR(1)	-0.00164876	0.0360998	-0.0456724	0.963934
AR(2)	-1.00086	0.00372375	-268.777	0.000000
MA(1)	0.18658	0.0568016	3.28477	0.003017
MA(2)	-0.808975	0.0584686	-13.8361	0.000000
SAR(1)	-1.25691	0.0485947	-25.8652	0.000000
SAR(2)	-1.00187	0.00661437	-151.469	0.000000

3-2-3-6 Forecasting

After determining the appropriate model for the time series and estimating its parameters, we now reach the last stage, which is forecasting the production quantities for the first component, which includes production units in the EGPS station. The forecast is on a monthly basis for a Forecast of one year.

Table (10): Forecast values for the first component.

Period	Forecast	Lower 95% Limit	Upper 95% Limit
10/54	1.7557	1.37763	2.13378
11/54	4.92998	4.14758	5.71237
12/54	4.31813	3.11083	5.52542
1/55	4.1786	2.45845	5.89874
2/55	5.05642	2.70481	7.40803

3/55	5.18121	2.15156	8.21087
4/55	8.46858	4.75674	12.1804
5/55	10.8841	6.43624	15.3319
6/55	8.92454	3.65331	14.1958
7/55	10.6842	4.55098	16.8174
8/55	12.4425	5.4446	19.4403
9/55	12.3301	4.42555	20.2347

The table shows the future values that all the predicted values are between the upper and lower bounds, with a confidence of 95%.

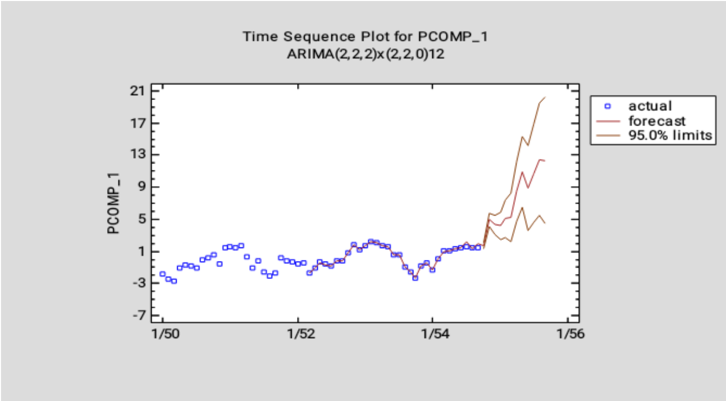


Figure (11): forecast values for the first component and confidence limits.

The same steps above were applied to the second and third components, and the results were as follows:

4-2-4 ARIMA models on the Second Component:

Table (11): Suggested models for the time series for the second component

Model	RMSE	AIC	HQC	SBIC
ARIMA(2,2,2)x(2,2,0)12	0.514273	-1.11947	-1.0359	-0.904417
ARIMA(2,2,2)x(2,2,1)12	0.518498	-1.06802	-0.970516	-0.817124
ARIMA(2,1,0)x(2,2,2)12	0.527868	-1.06729	-0.983713	-0.852234
ARIMA(2,2,2)x(2,2,2)12	0.528351	-0.995287	-0.883849	-0.708543
ARIMA(1,0,0)x(2,2,2)12	0.567394	-0.957964	-0.888314	-0.778749

After determining the appropriate model, the parameters of the ARIMA model $(2,2,2)x(2,2,0)^{12}$ were estimated, in order to know the significance of the estimated parameters as in Table (12):

Table (12): Estimation of Parameters Values for ARIMA Model (2,2,2)x(2,2,0)12

Parameter	Estimate	Stnd. Error	t	P-value
AR(1)	-0.578814	0.0740673	-7.8147	0.000000
AR(2)	-0.973181	0.0424652	-22.9172	0.000000
MA(1)	1.47241	0.0837192	17.5875	0.000000
MA(2)	-0.580915	0.0693088	-8.38154	0.000000
SAR(1)	-0.00814576	0.0108006	-0.754196	0.457777
SAR(2)	0.997621	0.0105285	94.7545	0.000000

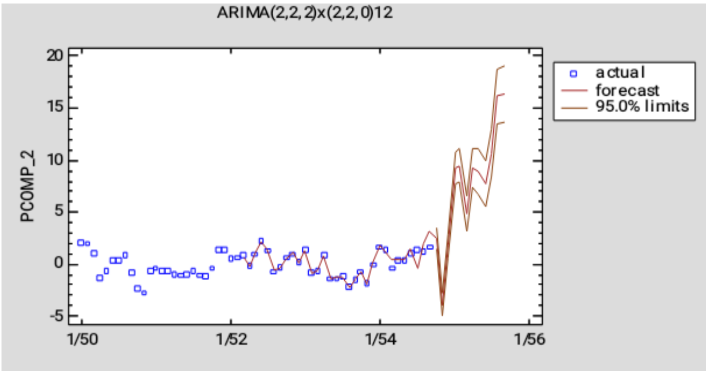


Figure (12): The original time series and the predicted values in the ARIMA(2,2,2)x(2,2,0)12 model.

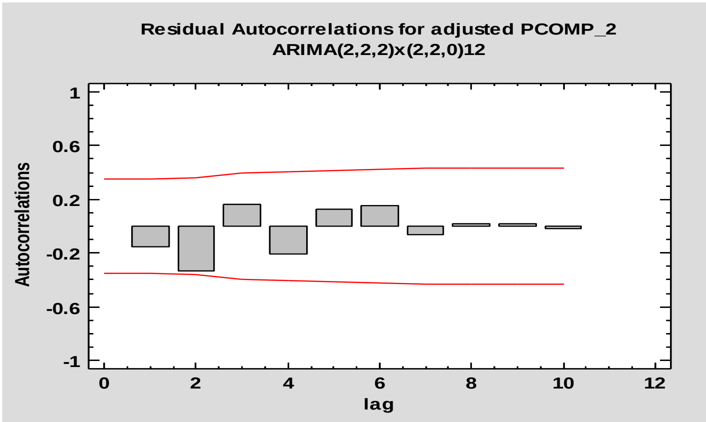


Figure (13): Autocorrelation of the residuals of the ARIMA (2,2,2)x(2,2,0)12 model.

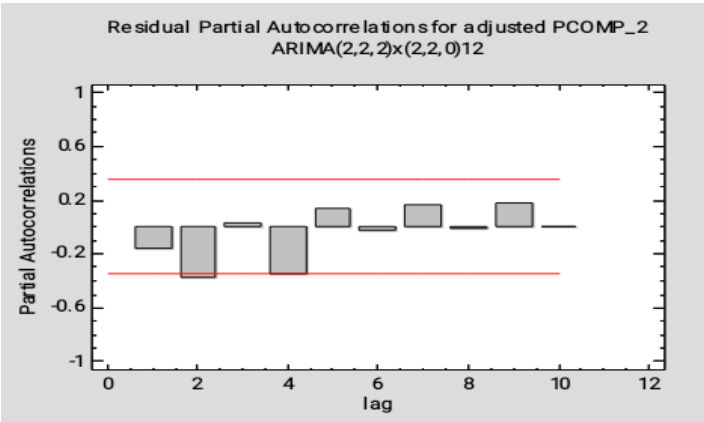


Figure (13): The PACF of the residuals of the ARIMA $(2,2,2)x(2,2,0)^{12}$ model.

Table (13): Forecast values for the second component.

Period	Forecast	Lower 95% Limit	Upper 95% Limit
10/54	2.54565	1.47466	3.61664
11/54	-3.80888	-4.88127	-2.73649
12/54	1.72324	0.601445	2.84504
1/55	9.22223	7.69154	10.7529
2/55	9.44044	7.78468	11.0962
3/55	4.86993	3.17778	6.56208
4/55	9.21663	7.37225	11.061

5/55	8.86193	6.67385	11.05
6/55	7.68736	5.49922	9.8755
7/55	10.6652	8.44787	12.8826
8/55	16.1437	13.5107	18.7766
9/55	16.3141	13.623	19.0051

4-2-5 ARIMA models on the Third Component:

Table (14): Suggested models for the time series for the third component

Model	RMSE	AIC	HQC	SBIC
ARIMA(2,1,2)x(2,2,0)12	0.399718	-1.62347	-1.53989	-1.40841
ARIMA(2,1,2)x(2,2,1)12	0.399472	-1.58961	-1.4921	-1.33871
ARIMA(2,1,2)x(2,2,2)12	0.401984	-1.54198	-1.43055	-1.25524
ARIMA(2,0,1)x(2,2,0)12	0.471664	-1.32754	-1.25789	-1.14832
ARIMA(0,0,2)x(2,2,0)12	0.489263	-1.28936	-1.23364	-1.14599

Table (15): Estimation of Parameters Values for ARIMA Model (2,1,2)x(2,2,0)[12]

Parameter	Estimate	Stnd. Error	t	P-value
AR(1)	0.870261	0.0779859	11.1592	0.000000
AR(2)	-0.971534	0.0530477	-18.3143	0.000000
MA(1)	1.56539	0.0686402	22.8057	0.000000
MA(2)	-0.630029	0.0566347	-11.1244	0.000000
SAR(1)	-1.83538	0.0763487	-24.0394	0.000000
SAR(2)	-0.983574	0.013598	-72.3324	0.000000

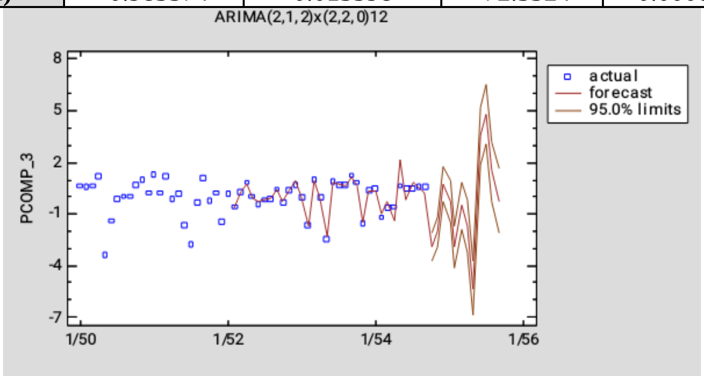


Figure (15): forecast values for the third component and confidence limits.

Table (16): Forecast values for the third component.

Period	Forecast	Lower 95% Limit	Upper 95% Limit
10/54	-2.91658	-3.74641	-2.08675
11/54	-2.02813	-2.89566	-1.16059
12/54	0.756745	-0.261128	1.77462
1/55	-0.338919	-1.54959	0.871748
2/55	-2.90106	-4.11173	-1.69039
3/55	-0.533547	-1.92745	0.860362
4/55	-1.70038	-3.24026	-0.160496
5/55	-5.29098	-6.83161	-3.75035
6/55	3.53251	1.87033	5.1947
7/55	4.80126	3.08114	6.52139
8/55	1.48514	-0.256761	3.22704
9/55	-0.226752	-2.11262	1.65911

4-2-6. Comparison between the models in components:

Based on (RMSE) to comparison between the selected models in each component's component, as shown in the table (17):

Table (17): Comparison of the extracted models

NO.	Components	Models	RMSE
1	First component	ARIMA(2,2,2)x(2,2,0)12	0.180955
2	Second component	ARIMA(2,2,2)x(2,2,0)12	0.514273
3	Third component	ARIMA(2,1,2)x(2,2,0)12	0.399718

We notice in the table (17) that the first component series model ARIMA(2,2,2)x(2,2,0)12 ranks first, the third component series model ARIMA(2,1,2)x(2,2,0)12 ranks second, and the second component series model ARIMA(2,2,2)x(2,2,0)12 ranks third.

4-2-7. Select the best model and forcasting data of EGPS :

In this section, the best model is selected, which is (ARIMA(2,2,2)x(2,2,0)12) based on the statistical criteria in table (17), and the data on electric power production for the city of Erbil were forcasting for a period of (12) months for 2022, the table (18) and graph (16) show that.

Table (18) Forecast Table for EGPS
Model: ARIMA(2,2,2)x(2,2,0)12 with constant

Period	Forecast	Lower 95% Limit	Upper 95% Limit
Period	*Forecast*	*Limit*	*Limit*
58.0	2423.14	2345.45	2500.83
59.0	2845.58	2724.93	2966.23
60.0	2579.56	2454.11	2705.0
61.0	2794.95	2669.5	2920.39
62.0	2370.52	2235.31	2505.72
63.0	2820.57	2637.44	3003.69
64.0	2946.14	2724.83	3167.46
65.0	2364.79	2134.57	2595.01
66.0	2783.37	2549.49	3017.24
67.0	3415.87	3161.04	3670.7
68.0	3793.02	3488.19	4097.86
69.0	3627.24	3285.36	3969.11

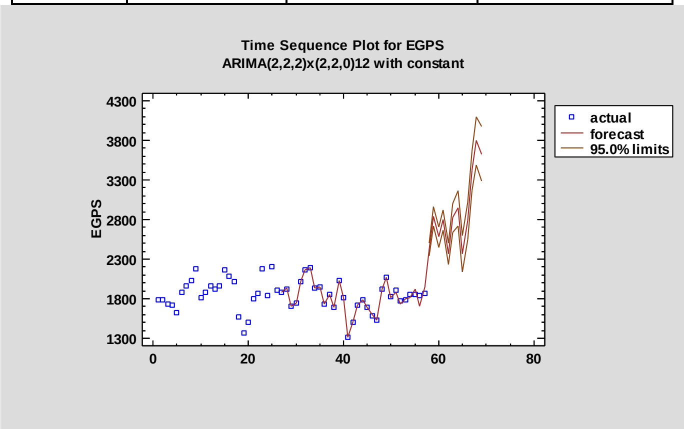


Figure (16): forecast values for the best model ARIMA(2,2,2)x(2,2,0)12

4-2-8. Conclusions:

1- In terms of time series models, we note that the estimated models are close to each other, which means that it can be relied upon in forecasting electrical energy data.

2- There are seasonal effects on the data of 12.

3- It can be concluded that the application of Box Jenks models on the extracted components data gave good results in terms of predicting data.

4-The main components extracted have importance in terms of order (the first component, then the second component ... etc.) and here we note that the estimated model for the first component has less RMSE.

5-The best proposed model for forecasting electrical energy production data in the city of Erbil is: ARIMA(2,2,2)x(2,2,0)12.

5. References

- Abdul-Ahad, Manahil Daniel, (2004): "The immune estimation in the first-order autoregressive model", Master's thesis, College of Computer Science and Mathematics, University of Mosul.
- Bisgaard, S. and Kulahci, M., 2011. Time series analysis and forecasting by example. John Wiley & Sons.
- Dunteman, G.H., 1989. Principal components analysis (No. 69). Sage.
- Guo, J., Peng, Y., Peng, X., Chen, Q., Yu, J. and Dai, Y., 2009, August. Traffic forecasting for mobile networks with multiplicative seasonal ARIMA models. In 2009 9th International Conference on Electronic Measurement & Instruments (pp. 3-377). IEEE.
- KARAKAŞ, G., 2021. Forecasting of Natural Honey Yield in Turkey through ARIMA Model. Gaziosmanpaşa Üniversitesi Ziraat Fakültesi Dergisi, 38(3), pp.166-172.
- Ladalla, N., Josef (2000). "Multivariate Time Series Analaysis in Principal Components Space", The 32nd Symposium on interface: computing science and statistic, New Orleans, Louisiana.
- Miao, D., Qin, X. and Wang, W., 2014, October. The periodic data traffic modeling based on multiplicative seasonal ARIMA model. In 2014 Sixth International Conference on Wireless Communications and Signal Processing (WCSP) (pp. 1-5). IEEE.
- Smith, L.I., 2002. A tutorial on principal component.
- Tian, S., Fu, Y., Ling, P., Wei, S., Liu, S. and Li, K., 2018, November. Wind power forecasting based on arima-lgarch model. In 2018 International Conference on Power System Technology (POWERCON) (pp. 1285-1289). IEEE.

بەکارهێنانی پێکهاتە سەرەکییەکان لە مۆدێلی زنجیرە کاتییەکان و هەڵبژاردنی باشترین مۆدێلەکان لەگەڵ ئەپلیکەیشن

ڕزگار مەغدید ئەحمەد
هەولێر-بەشی ئامار و زانیاریەکان، کۆلێژی کارگێری و ئابووری، زانکۆی سەلاحەددین
rizgar.ahmed@su.edu.krd

نیداء سلیم مەلە یونس
هەولێر-بەشی ئامار و زانیاریەکان، کۆلێژی کارگێری و ئابووری، زانکۆی سەلاحەددین
rizgar.ahmed@su.edu.krd

پوختە

لەم توێژینەوەیەدا، شیکاری پێکهاتە سەرەکییەکان کە یەکێکە لە شێوازەکانی شیکاری فرەگۆڕاو بەکارهێنرا بۆ پێشبینیکردنی مۆدێلی زنجیرە کاتییەکان (مۆدێلی بۆکس-جێنکنز) بە بەکارهێنانی لەسەر داتاکانی کارەبا (وێستگەی کارەبای غازی هەولێر) (EGPS) کە داتای فرەگۆڕاو لەخۆدەگرێت (5 وێستگە) و داتاکان مانگانە بوون بۆ ماوەی (2017/1/1) تا (2021/9/14)، بیرۆکەی توێژینەوەکە لەسەر بنەمای جێبەجێکردنی شیکاری پێکهاتە سەرەکیەکان لەسەر داتای فرەگۆڕاوی زنجیرە کاتییەکان و بەدەستهێنانی پێکهاتە دەرهێنراوەکان پاشان خەمڵاندنی مۆدیلی (بۆکس -جێنکز) لەسەر پێکهاتەکان, دەرەنجامی سەرەکی ئەوەیە کە شیکاری پێکهاتەی سەرەکی کاریگەرە لە کەمکردنەوەی چەندین داتای زنجیرە کاتیی و خەمڵاندنی مۆدێلی کاریگەر پشت بەست بە پێوەری ئاماری.

لە کۆتاییدا، باشترین مۆدێل هەلبذیردرا بۆ پێشبینی داتای بەرهەمر هێنانی کارەبای شاری هەولێر کة ئةویش:ARIMA(2,2,2)x(2,2,0)12

ووشە کلیلەکان: شیکاری فرەگۆڕاو؛ شیکاری پێکهاتەی سەرەکی؛ مۆدێلی بۆکس-جێنکنز؛ پێشبینیکردن.

توظيف المكونات الرئيسية في نماذج السلاسل الزمنية واختيار أفضل النماذج مع التطبيق

رزكار مغديد أحمد
قسم الأحصاء والمعلوماتية، كليه الأدارة و الأقتصاد، جامعة صلاح الدين - أربيل
rizgar.ahmed@su.edu.krd

نيداء سليم ملا يونس
قسم الأحصاء والمعلوماتية، كليه الأدارة و الأقتصاد، جامعة صلاح الدين - أربيل
nida.malayounis@su.edu.krd

ملخص

في هذه الدراسة، تم استخدام تحليل المكونات الرئيسية ، وهو أحد طرق التحليل متعدد المتغيرات للتنبؤ بنماذج السلاسل الزمنية (Box-Jenkins Model) من خلال التطبيق على بيانات الطاقة الكهربائية (محطة توليد الطاقة الغازية في أربيل) (EGPS) والتي تحتوي على بيانات متعددة المتغيرات (5 محطات) وكانت البيانات شهرية للفترة من (2017/1/1) الى (2021/9/14)، استندت فكرة البحث إلى تطبيق تحليل المكونات الرئيسية على بيانات السلاسل الزمنية المتعددة ، والحصول على المكونات المستخرجة منها ، ثم تقدير نماذج Box-Jenkins.

و اخيرا تم اختيار افضل النماذج المقترحة في التنبؤ لبيانات انتاج الطاقة الكهربائية في مدينة اربيل هي:ARIMA(2,2,2)x(2,2,0)12

الاستنتاج الرئيسي هو أن تحليل المكون الرئيسية فعال في تقليل بيانات السلاسل الزمنية المتعددة والحصول على أفضل النماذج بالأستناد على معايير إحصائية.

الكلمات الدالة: تحليل متعدد المتغيرات. تحليل المكونات الرئيسية؛ نموذج Box-Jenkins ؛ التنبؤ.