Journal of Nature, Life and Applied Sciences
Volume (3), Issue (4): 30 Dec 2019
P: 90 - 102

AJSRP
ISSN: 2522-3356

مجلة العلوم الطبيعية والحياتية والتطبيقية
المجلد (3)، العدد (4): 30 ديسمبر 2019 م
ص: 90 - 102

Application of SARIMA Model for Forecasting Consumption of Electricity in Gezira State, Sudan (2006-2018)

Nada Mohammed Ahmed Alamin

Department of Mathematics || Faculty of Science || University of Hafr Albatin || KSA

Abstract: The purpose of the research is to reach the forecast of monthly electricity consumption in Gezira state, Sudan for the period (Jun 2018 - Dec 2020) through the application to the historical data of electric power consumption (Jan 2006-May 2018) obtained from the National Control Center, which has been applied in the research methodology of seasonal Autoregressive Integrated Moving Average due to seasonal behavior in the data, good forecast has been given by SARIMA (2, 1, 7) (0, 1, 1), which has been examined its quality using the Thiel coefficient. The study recommended the use of the model of seasonal Autoregressive Integrated Moving Average in data with Seasonal behavior due to its simple application and accuracy of the results reached.

Keywords: seasonal Autoregressive Integrated Moving Average(SARIMA), Thiel coefficient, Electricity consumption.

تطبيقات نماذج ساريما للتنبؤ باستهلاك الكهرباء في ولاية الجزيرة، السودان (2018 -2006)

ندى محمد أحمد الأمين

كلية العلوم || جامعة حفر الباطن || المملكة العربية السعودية

الملخص: الهدف من البحث الوصول إلى التنبؤ بالاستهلاك الشهري للطاقة الكهربائية في ولاية الجزيرة للفترة من يونيو 2018-ديسمبر 2020 من خلال التطبيق علي بيانات استهلاك الطاقة الكهربائية (يناير 2006- مايو 2018) المتحصل علها من مركز التحكم القومي وقد طبقت في البحث نموذج الانحدار الذاتي التكاملي المتوسط المتحرك الموسمية نظراً للسلوك الموسمي في البيانات وقد تم التوصل إلى تنبؤات جيدة بواسطة النموذج (0، 1، 1)(2، 1، 7) SARIMA تم فحص جودته باستخدام معامل ثايل وقد اوصت الدراسة باستخدام اسلوب الانحدار الذاتي التكاملي المتوسط المتحرك الموسمي للبيانات ذات السلوك الموسي نظراً لسهوله تطبيقه ودقة النتائج المتوصل الها.

الكلمات المفتاحية: نموذج الانحدار الذاتي التكاملي المتوسط المتحرك الموسمي، معامل ثايل، استهلاك الكهرباء.

Introduction:

The shocking fact about the electricity supply in Sudan is that the country today has a large areas do not enjoy electricity services and live in total darkness in the west, central and eastern Sudan, and there are some areas has been exit of productive sectors such as industry, agriculture and a large part of the services even in The capital, so It was necessary to be precise in determining the volume of electricity consumption in Sudan in general and in Gazira state in particular, given the importance of the mandate in the Sudanese economy.

Time series modeling and forecasting has fundamental importance to various practical domains. Thus a lot of active research works is going on in this subject during several years.

The Research problem:

The consumption of electric energy is subject to several factors that make each year different for the following year and each month different from the others. The most immediate problem is lack of studies in delineating and forecast. The study fetches the most appropriate model to predict the production of electricity in Gezira state, where it gives more accurate results.

Research questions:

1. Can we suggest the most suitable statistical time series model from which we can predict the amount of consumption of electric energy in Gezira state?
2. Is it possible to predict the behavior (stationary- seasonal) of electric power consumption?

Research hypothesis:

Predictions by seasonal models (SARIMA) of such data are more accurate than those derived from normal ARIMA models.

Research objectives:

The main objective of this study is to reach an appropriate model to forecast the consumption of electricity using SARIMA model the more effective and simple model for forecasting and the Sub objectives is:

1. To suggest the most suitable statistical time series model from which we can predict the amount of consumption of electric energy in Gezira state.
2. To identify the order of SARIMA model that fits the precipitation time series data.
3. To use the SARIMA model to forecast precipitation.
4. To predict the behavior of electric power production.

research tools and methods:

The research based its methodology on the theoretical side that dealt with the SARIMA model in the analysis of time series and support the theoretical side of the applied side, which relied on actual data on electricity consumption in Gezira state, Sudan to reach the best results for forecasting. The last part of the research included the conclusions and recommendations, program is used to analyze the data SPSS release 18.

Theoretical framework:

The Time Series allows you to build custom nonseasonal or seasonal ARIMA (Autoregressive Integrated Moving Average) models also known as Box-Jenkins (Box, Jenkins, and Reinsel, 1994) models with or without a fixed set of predictor variables.

Non-Seasonal ARIMA model:

This method has three variables to account for

P = Periods to lag for, P helps adjust the line that is being fitted to forecast the series

D = In an ARIMA model we transform a time series into stationary one(series without trend or seasonality) using differencing. D refers to the number of differencing transformations required by the time series to get stationary. Stationary time series is when the mean and variance are constant over time. It is easier to predict when the series is stationary(1).

Differencing is a method of transforming a non-stationary time series into a stationary one. This is an important step in preparing data to be used in an ARIMA model.

The first differencing value is the difference between the current time period and the previous time period. If these values fail to revolve around a constant mean and variance then we find the second differencing using the values of the first differencing. We repeat this until we get a stationary series

The best way to determine whether or not the series is sufficiently differenced is to plot the differenced series and check to see if there is a constant mean and variance.

Q = This variable denotes the lag of the error component, where error component is a part of the time series not explained by trend or seasonality.

Trend, seasonality, cycles and residuals:

One simple method of describing a series is that of classical decomposition. The notion is that the series can be decomposed into four elements:

1. Trend (Tt): long term movements in the mean.
2. Seasonal effects (It): cyclical fluctuations related to the calendar.
3. Cycles (Ct): other cyclical fluctuations (such as a business cycles).
4. Residuals (Et): other random or systematic fluctuations(2).

The idea is to create separate models for these four elements and then combine them, either additively

(1) Sangarshanan ," Time series Forecasting — ARIMA model"
https://towardsdatascience.com/@sangarshananveera?source=post_page

(2) Samira Muhammad Salh, Salahaddin A.Ahmed, " Box –Jenkins Models For Forecasting The Daily Degrees Of Temperature In Sulaimani City", Journal of Engineering Research and Applications, feb 2014

Xt = Tt + It + Ct + Et

or multiplicatively

Xt = Tt ·It ·Ct ·Et.

Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

ACF is the proportion of the auto covariance of Y_t and Y_{t-k} to the variance of a dependent variable Y_t.

Partial autocorrelation function (PACF) is the simple correlation between and minus the part explained by the intervening lags

After plotting the ACF plot we move to Partial Autocorrelation Function plots (PACF). A partial autocorrelation is a summary of the relationship between an observation in a time series with observations at prior time steps with the relationships of intervening observations removed.

The partial autocorrelation at lag k is the correlation that results after removing the effect of any correlations due to the terms at shorter lags.

Seasonal ARIMA model:

Seasonality is a particular type of autocorrelation pattern where patterns occur every season like monthly, yearly, etc. Seasonality must also be corrected before a time series model can be fitted and we can see the way that we can correct the seasonality problem.

Seasonal Index: Seasonal index represents the extent of seasonal influence for a particular segment of the year. The calculation involves a comparison of the expected values of that period to the grand mean.

A seasonal index is how much the average for that particular period tends to be above (or below) the grand average. Therefore, to get an accurate estimate for the seasonal index, we compute the average of the first period of the cycle, and the second period, etc, and divide each by the overall average. The formula for computing seasonal factors is:

$S_i = D_i/D$,

where:

Si = the seasonal index for i^{th} period,

Di = the average values of i^{th} period,

D = grand average,

i = the i^{th} seasonal period of the cycle.

A seasonal index of 1.00 for a particular month indicates that the expected value of that month is 1/12 of the overall average. A seasonal index of 1.25 indicates that the expected value for that month is

25% greater than 1/12 of the overall average. A seasonal index of 80 indicates that the expected value for that month is 20% less than 1/12 of the overall average[3].

The problem with ARIMA?

Autoregressive Integrated Moving Average, or ARIMA, is a forecasting method for univariate time series data.

As its name, it join both an autoregressive and moving average elements. The integrated element refers to differencing allowing the method to support time series data with a trend.

A problem with ARIMA is that it does not support and cannot deal with seasonal data. That is a time series with a repeating cycle every season.

ARIMA can deal with data that is either not seasonal or has the seasonal component removed, by adjusting using several methods such as seasonal differencing.

Definition of SARIMA (Seasonal Autoregressive Integrated Moving Average):

SARIMA or Seasonal ARIMA, is an extension of ARIMA that explicitly supports univariate time series data with a seasonal component.

As the name of the model, this model is used when the time series exhibits seasonality. This model is similar to ARIMA models or we can say the SARIMA model is the type of ARIMA, we just have to add in a few parameters to account for the seasons.

It adds three new parameters to specify the autoregression (P), differencing (D) and moving average (Q) for the seasonal component of the series, as well as an additional parameter for the period of the seasonality.

We can write SARIMA as

ARIMA(p, d, q)(P, D, Q)m, where

p : the number of autoregressive

d : degree of differencing

q : the number of moving average terms

m : refers to the number of periods in each season here m is optional.

(P, D, Q): represents the (p, d, q) for the seasonal part of the time series.

(3) Sitio Espejo para Amrica Latina , Sitio en los E.E.U.U. Handbook of Time-Critical Decision Making for Business Administration, 1994-2015

Goodness of fit

Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are two measures goodness of fit. They measure the trade-off between model fit and complexity of the model.

AIC = 2k + n [Ln(2(pi) RSS/n) + 1],

AIC = 2k + n Log(RSS/n),

AIC = k + n [Ln(2(pi) RSS/(n-k)) +1],

RSS is the Residual Sum of Squares and K is the number of model parameters

BIC = k log(n)- 2log(L($\hat{\theta}$)).

Here n is the sample size; the number of observations or number of data points you are working with.

k is the number of parameters which your model estimates, and θ is the set of all parameters.

L($\hat{\theta}$) represents the likelihood of the model tested, given your data, when evaluated at maximum likelihood values of θ. You could call this the likelihood of the model given everything aligned to their most favorable.

A lower AIC or BIC value indicates a better fit for the model.

Thiel's inequality coefficient:

Thiel's inequality coefficient, also known as Thiel's U, provides a measure of how well a forecasting of estimated values compares to observed values, Thiel's U is calculated as:

$$U = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}x^2} + \sqrt{\frac{1}{n}\sum_{i=1}^{n}y^2}}$$

Thiel's inequality coefficient is useful for comparing different forecast methods, The closer the value of U is to zero, The better the forecast method.

Results and Discussion:

Analyzing and forecasting the data must have many stages can be summarized in the following steps:

1. Data stationary:

a time series data for Consumption of Electricity in Gezira State, Sudan(Jan 2006- April 2018) should be stationary, so the first step in model identification is to make sure that the series is stationary, and this is done by looking at the time series plot of the data, extracting Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF), as well as drawing the confidence limits for the autocorrelation and partial autocorrelation function to determine the behavior of the data.

The figure (1) below show a sequence chart of the consumption of electricity in Gezira state from January 2006 until April 2018, as we see the data have a trend.

Differencing is necessary when trends are present, series with trends are typically nonstationary and ARIMA modeling assumes stationarity.

Consumption of electricity in Gezira state from January 2006 until April 2018 by watts:

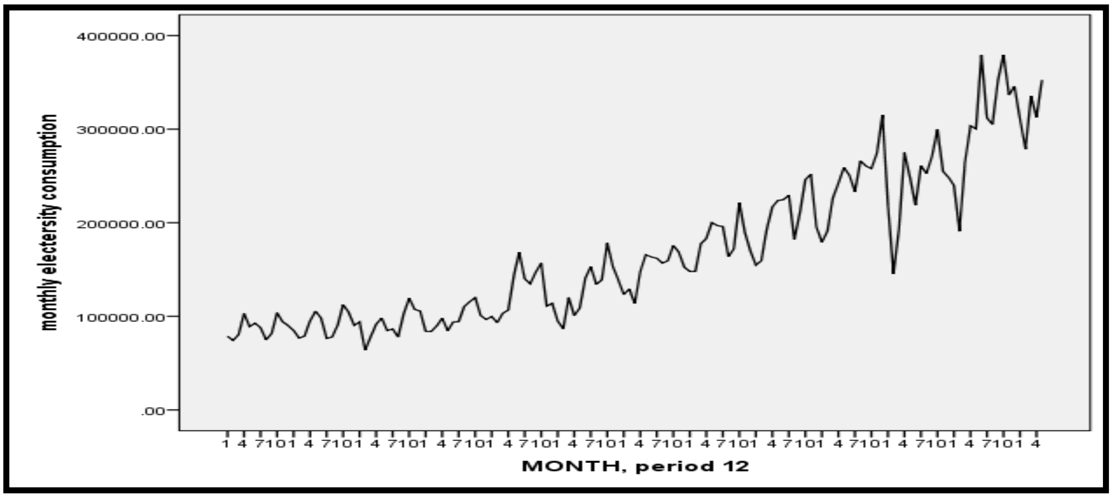

Fig (1) Consumption of electricity in Gezira state from January 2006 until April 2018 by watts

Source: SPSS program

2. Data Transformations:

In time series graph the data shown seasonally behavior, then strong need to take the seasonally difference from the data and another diagram will appear after we generate date variables to establish periodicity and to distinguish between historical, validation, and forecasting periods.

Consumption of electricity in Gezira state from January 2006 until April 2018 by watts after take the seasonally difference:

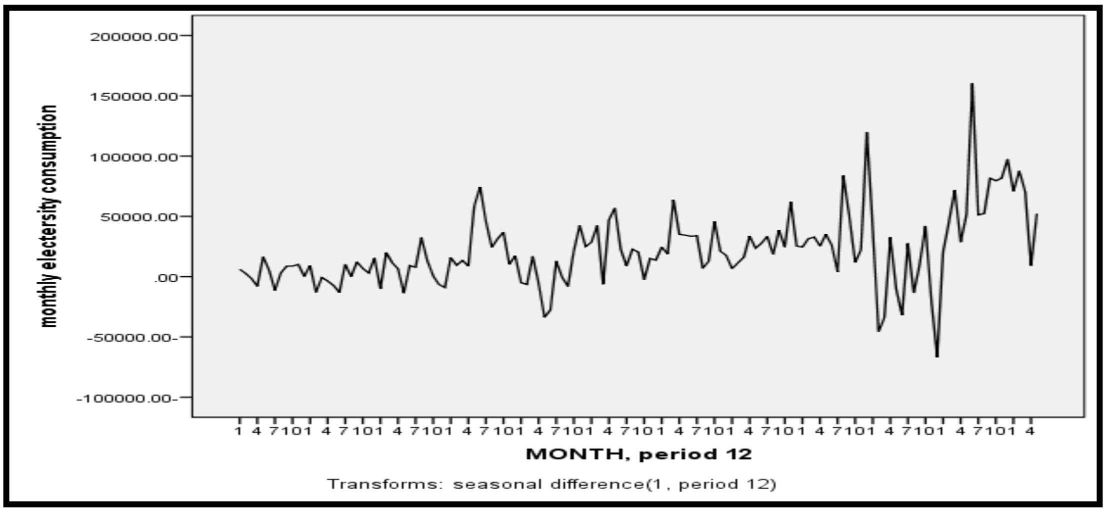

Fig (2) Consumption of electricity in Gezira state from January 2006 until April 2018 by watts after take the seasonally difference. Source: SPSS result program

3. Create Time Series

The Create Time Series allows to create new variables based on existing numeric time series variables. These transformed values are useful in many time series analysis procedures.

Table (1) Monthly electricity consumption summary

Case Processing Summary		
Item		**monthly electricity consumption**
Series or Sequence Length		149
Number of Missing Values in the Plot	User-Missing	0
	System-Missing	0

Source: SPSS program result

4.Building Models and Producing Forecasts

By using SPSS program two procedures for accomplishing the tasks of creating models and producing forecasts is:

- The time series modeler
- The apply time series models.

Estimated residual ACF and PACF(autocorrelation function and Partial Autocorrelation):

working with ARIMA process of order (p, d, q) ans SARIMA (p, d, q) (P, D, Q), the problem Which faced us is to choose the most appropriate values for p, d, q, P, D and Q to specify the ARIMA or SARIMA model. This problem is partly resolved by examining both the autocorrelation function, and the partial autocorrelation function for the time series.in figure (3) the ACF and PACF within the confidence interval for SARIMA (2, 1, 7) (0, 1, 1) model.

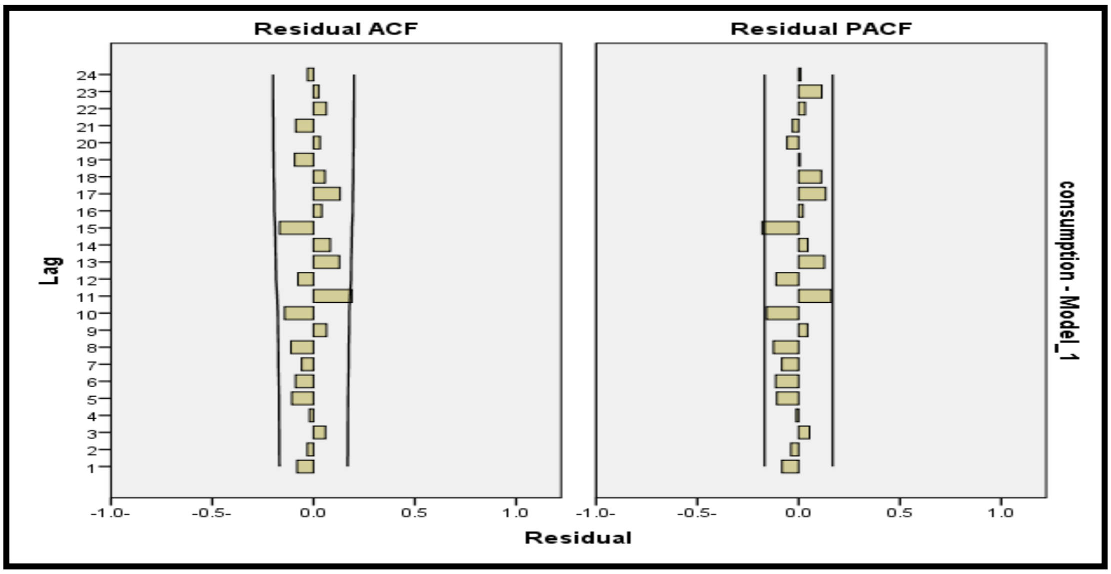

Fig (3) ACF and PACF ARIMA model (2,1,7) (0,1,1)

The best model was selected by comparing the BIC value. Model (2, 1, 7) (0, 1, 1) which was selected as the most appropriate model since it has the least values of the information criterions (BIC).

Table (2) ARIMA (2,1,7) (0,1,1) model description

Model Description			
			Model Type
Model I	monthly consumption	Model_1	ARIMA(2, 1, 7)(0, 1, 1)

Source: SPSS program result

Table (3) ARIMA (2,1,7) (0,1,1) model statistics

Fit Statistic	Mean
Stationary R-squared	.721
R-squared	.957
RMSE	16954.999
MAPE	7.816
MaxAPE	39.171
MAE	12105.860
MaxAE	55609.216
Normalized BIC	19.766

Source: SPSS program result

Table (4) ARIMA (2,1,7) (0,1,1) model fit statistic

Model Statistics							
Model	Model Fit statistics			Ljung-Box Q(18)			Number of Outliers
	Stationary R-squared	R-squared	Normalized BIC	Statistics	DF	Sig.	
consumption	.721	.957	19.766	27.956	14	.014	4

Source: SPSS program result

Data forecasting:

Forecasts for model often start after the last point in the range of the requested estimation period, and end at the last period for which non-missing values of all the predictors are available or at the end date of the requested forecast period.

The main aim of this study was to formulate a SARIMA model that fits our precipitation data and use the model to forecast. Having identified SARIMA (2, 7, 1) (0, 1, 1) as the most appropriate SARIMA model, we used the model to forecast the next 31 monthly precipitation values.

The results of forecasted consumption presented in table below. The behavior of electricity consumption has shown in Figure (6). In this figure, the time is measured along the horizontal axis and the vertical axis measures the level of monthly average.

Table (5) consumption of electricity in Gezira state forecasting using SARIMA model (2,7,1) (0,1,1)

Forecast			
	Model		
	monthly electricity consumption-Model_1		
	Forecast	UCL	LCL
Jun 2018	352836.24	386179.52	319492.95
Jul 2018	364221.96	401892.54	326551.38
Aug 2018	357243.33	395471.84	319014.81
Sep 2018	385371.24	425360.35	345382.14
Oct 2018	410570.28	453414.57	367725.99
Nov 2018	381506.06	426472.42	336539.71
Dec 2018	377373.97	424030.99	330716.94
Jan 2019	352069.79	399242.47	304897.11
Feb 2019	322185.65	369928.12	274443.19
Mar 2019	377386.25	426001.19	328771.31
Apr 2019	416221.70	465669.50	366773.89
May 2019	406726.39	456888.87	356563.91
Jun 2019	402622.71	457990.33	347255.10
Jul 2019	414822.32	472608.32	357036.33

Forecast			
Aug 2019	408433.96	467369.56	349498.36
Sep 2019	436401.27	496945.37	375857.17
Oct 2019	461280.69	523777.93	398783.44
Nov 2019	431754.73	495968.67	367540.79
Dec 2019	426754.06	492520.80	360987.31
Jan 2020	401602.52	468408.37	334796.67
Feb 2020	372005.52	439865.78	304145.25
Mar 2020	427155.65	496200.09	358111.21
Apr 2020	465896.18	536094.11	395698.25
May 2020	456417.56	527702.86	385132.25
Jun 2020	452345.26	528707.69	375982.82
Jul 2020	464539.35	543573.38	385505.32
Aug 2020	458140.62	538665.78	377615.45
Sep 2020	486109.75	568565.86	403653.64
Oct 2020	510992.60	595694.80	426290.39
Nov 2020	481466.04	568188.82	394743.26
Dec 2020	476464.23	565049.63	387878.83

Source: SPSS program result

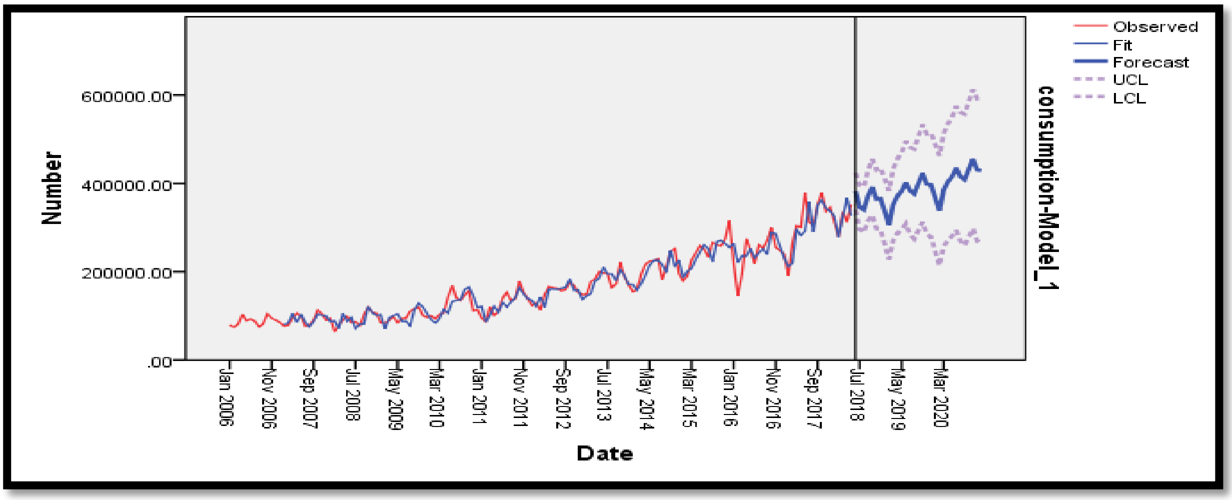

Fig (4) Plot for consumption of electricity in Gezira state after apply time series model

Source: SPSS program result

prediction of consumption of electricity in Gezira state using ARIMA model (2,1,7) (0,1,1):

Table (6) consumption of electricity in Gezira state real data and forecasting using ARIMA model (2,1,7) (0,1,1):

month	Real value	forecasting
Jun 2018	350382	352836.2

month	Real value	forecasting
Jul 2018	315642	364222
Aug 2018	377997	357243.3

We use Thiel's inequality coefficient to evaluate prediction of consumption of electricity in Gezira state.

Thiel's inequality coefficient define as:

Table (7)Thiel's inequality calculation using ARIMA model (2,1,7) (0,1,1)

Thiel inequality coefficient using ARIMA models(2,1,7) (0,1,1):						
	x	y	x^2	y^2	x-y	$(x-y)^2$
jun	350382	352836.2	1.22767E+11	1.24493E+11	-2454.33	6023735.749
jul	315642	364222	99629853857	1.32658E+11	-48579.99	2360015331
aug	377997	357243.3	1.42882E+11	1.27623E+11	20753.61	430712245
sum			3.65279E+11	3.84774E+11		2796751312
sum/3			1.2176E+11	1.28258E+11		932250437.3
sqr(sum/3)			348940.7886	358131.1888		30532.77644
U	0.043181992					

Source: EXCEL program result

As we see before when we use SARIMA model (2, 1, 7) (0, 1, 1) U = 0.043 This value is close to zero, which indicates the accuracy of the forecasts obtained.

Conclusion:

With the help of the SARIMA model, the future consumption data will be available. The forecast consumption show an increasing trend, with due consideration to seasonality. For this government may be advised to plan the production process, so that the electricity get a stable and continues. The consumption during September, October, November and December months has been observed to be high and government can plan to increase production during these months.

Certainly the existence of specialized models in seasonal data gives better models and predictions than normal models, as these models are developed to solve the problem of seasonality in the data.

In this paper, Using the time series data of monthly average consumption of electricity in Gezira state, the study build a Seasonal ARIMA (2, 1, 7) (0, 1, 1) model. It could be successfully used for modeling as well as forecasting of average monthly consumption of electricity in Gezira state.

Recommendation:

1. Attention must be paid to the fact that the electricity generation, supply and distribution strategy in general is a process characterized by long-term strategic planning and requires an enormous amount

of funding and a high degree of qualification in management, training and experience, Therefore, it was necessary to build strong, accurate and specialized data quality models such as SARIMA models.

2. Sudan in general and Gezira state in particular needs to build a large, basic and cheap production generating force that meets demand at the base level of supply.
3. create more generators from the various energy sources available in the country.

References:

1- Sangarshanan, " Time series Forecasting — ARIMA model" https://towardsdatascience.com/@sangarshananveera?source=post_page

2- Samira Muhammad Salh, Salahaddin A.Ahmed, " Box –Jenkins Models For Forecasting The Daily Degrees Of Temperature In Sulaimani City", Journal of Engineering Research and Applications, feb 2014

3- Sitio Espejo para Amrica Latina, Sitio en los E.E.U.U. Handbook of Time-Critical Decision Making for Business Administration, 1994-2015.

4- Hemant Sharma and S. S. "Burark, Bajra Price Forecasting in Chomu Market of Jaipur District: An Application of SARIMA Model", Agricultural Situation in India, 61 (11), 2015

5- S. C. Hillmer and G. C. Tiao, "An ARIMA model based Approach to Seasonal Adjustment", Journal of the American Statistical Association, 72 (377), 1982

6- Haque, Mainul, Harnhirun, Somachai, and Shapiro, Daniel (1990). A time series analysis of causality between aggregate merger and stock prices: the case of Canada. Routledge, Taylor and Francis Group. Issue 7, Vol. 27

7- Kallberg, Jarl, and Pasquariello, Paolo (2008). Time Series and cross – sectional excess comovement in stock indexes. Journal of Empirical Finance, Issue 15.