استخدام نموذج ARIMA للتنبؤ بانتاج الطاقة الكهربائية في استراليا
Using ARIMA model to forecasting with production of electrics in Australia

مدرس مساعد . مشتاق طالب حسين
جامعه الانبار / كلية الادارة والاقتصاد
Asst.Lecture . Mushtaq .T.H
Al-Anbar University/College of Administration and Economics

الملخص

هذا البحث يتعامل مع نماذج ARIMA المستخدمة في تحليل السلاسل الزمنية ، حيث ان هذه النماذج تمتاز بمرونة ودقه عاليتين في تحليل السلاسل الزمنية .

الجانب النظري يركز على المفاهيم العامة لنماذج ARIMA والادوات الاحصائية المستخدمة في تحليل السلسله الزمنية .

الجانب العملي يوضح الجزء التطبيقي لبحثنا باستخدام بيانات ماخوذة من الموقع الالكتروني (www.maths .monash.edu.au/Hyndman/forecasting/) ، هذه البيانات تمثل الطاقة الكهربائية المنتجة(بالاشهر) في استراليا من(كالنون الثاني 1956 الى اب 1995) ، بعد ذلك تم تشخيص النموذج المناسب للبيانات وبالتالي بناء النموذج والتنبؤ بانتاج الطاقة الكهربائية لسنتين قادمتين (ايلول 1995 الى اب 1997).

Abstract

This research to deal with ARIMA models it used in time series analysis. Whereas this models are distinct with high flexible and accuracy in time series analysis.

Theoretical part is concentrate on general concepts for ARIMA models, and statistical tools that have proved useful in analyses time series.

Practice part is clear the practice part to our research, by using data taken from web page (www.maths.monash.edu.au/hyndman/forecasting/) .

This data is representing the production of electrics capacity in australia in (months) from (jan 1956 to aug 1995) after that, to be accomplishment checking the adequate model to this data and so building the model and forecasting with production of electrics for two year ahead (sep 1995 to aug 1997).

1- Introduction

1.1 Time series analysis

A time series is a collection of observations made sequentially in time. The statistical methodology dealing with the analysis of such a sequence of data is called “time series analysis” . in many other areas of statistical analysis the observations are assumed statistically independent, time series methods are used to analyze dependent observation. It is this dependence which characterizes the dynamic or the “ memory” of the underlying system. This dependence /dynamic/memory property of the system enables the prediction of future values of the system from past values once we quantify the dependence. the nature of the dependence or dynamics distinguishes one time series from another .[Alankang 1980 p.2].

1.2 The objective of time series analysis

The initial objective of time series analysis is to make inference about the properties or basic features of the stochastic process from the information contained in the observed series. The first step in the analysis is usually to form certain summary statistics, but the eventual aim is to construct a model from the data.

Once a model has been obtained, it can be used to generate synthetic data for future study, forecast values of the series, and evaluate control related systems. [Alankang 1980 p.2]

2- The aim of study

The aim of this study is computing forecasting values for production of electrics in Australian from (sep 1995 to agu 1997), by using Box-Jenkins methodology .

1.3 The Box-Jenkins approach

The Box-Jenkins approach lets the data speak for itself. It provides an objective and systematic approach to modeling and forecasting discrete time series. The Box-Jenkins approach strives for the “best model” for forecasting or control purposes based upon time series data. Their concept of “best model” is not the usual one found by minimization of the fitting errors, a model is adequate if all parameters are statistically significant and the errors from the model are independently distributed.

The Box-Jenkins methodology is a power full approach to the solution of many time series analysis problem. [Ajoy 2005 P.3]

3.2 Examining correlation in the time series data

In this section we consider analysis that can be applied to our time series to determine its statistical properties. [Spyros 1998 P.410]

3.2.1 The autocorrelation function (ACF)

The key statistic in time series analysis is the (correlation of time series with itself, lagged 1,2, or more periods). The formula of autocorrelation function as follows: [Spyros 1998 P.411]

$$r_k = \frac{\sum_{t=k+1}^{n}(y_t - \overline{y})(y_{t-k} - \overline{y})}{\sum_{t=1}^{n}(y_t - \overline{y})^2} \qquad \text{..............(3-1)}$$

3.2.2 White noise model

Any forecasting model should have forecast errors containing on white noise, equation (3-2) is a simple random model . [Box 1976 P.220]

$$y_t = c + e_t \qquad \text{……………..(3-2)}$$

Where :
C is a constant term
et is error component

3.2.3 The partial autocorrelation function (PACF)

Partial autocorrelations are used to measure the degree of association between yt and yt-k , when the effect of other time lags 1,2,…,k-1 .

The value of this can be seen in the following simple example. Suppose there was a significant autocorrelation between yt and yt-1 . then there will also be a significant correlation between yt-1 and yt-2 , since they are also one time unit a part, consequently, there will be a correlation between yt and yt-2 because both are related to yt-1 . so to measure the real correlation between yt and yt-2 this is what partial autocorrelation . [Spyros 1998 P.413]

3.2.4 Recognizing seasonality in time series

Seasonality is defined as a pattern that repeat itself over fixed interval of time. In general, seasonality can be found by identifying a large autocorrelation coefficient or a large partial autocorrelation coefficient at the seasonal lag. So for monthly data, large autocorrelations might also been at lag 12,24 , and even lag 36 . [Spyros 1998 P.415-416]

3.3 Examining stationary of time series data

Stationary is meaning on growth or decline in the data. The data must be roughly horizontal a long the time axis. In other words the mean and the variance for the data are constant.

The autocorrelation function also displays a typical pattern for a non – stationary series, with a slow decrease in the size of the autocorrelations. The autocorrelation for one time lag r1 is very large and positive. The autocorrelation for two time lags is also large and positive, but not as large as r1 . because the random error components will begin to dominate the autocorrelations . [Joseph 2006 P.9]

3.3.1 Removing non-stationary in time series

Trends, or other non-stationary pattern in the level of a series result in positive autocorrelations that dominate the autocorrelation. One way of removing non-stationary is through the model of differencing. We define the differenced series as the change between each observation in the original series. [Alankang 1980 p.11]

$$\bar{y}_t = y_t - y_{t-1} \qquad \text{……………(3-3)}$$

In case the series contains on seasonality patterns, for this case equation (3-4) is adequate .

$$\bar{y}_t = y_t - y_{t-s} \qquad \text{……………..(3-4)}$$

Where : s is the number seasons, for monthly data (s=12), for quarterly data (s=4). If the series non-stationary in the variance then we will take transformation* to the series.

3.3.2 Back shift operator

The back shift operator is convenient for describing the process of differencing. first difference can be written as: $\bar{y}_t = y_t - y_{t-1} = y_t - By_t = (1-B)y_t$, the second difference is a first difference of the first differences would be denoted 1-B2 . for monthly data , B12 is used and the notation is B12yt =yt-12 . a seasonal difference follow by difference can be written as : (1-B)(1-Bs)yt . [Spyros 1998 P.420]

3.4ARIMA models for time series

We discussion the following equation

$$y_t = b_0 + b_1 y_{t-1} + b_2 y_{t-2} + ... + b_p y_{t-p} + e_t \quad ....(3\text{-}5)$$

In this equation we denote that b0,b1,…..,bp are represented the parameters for this equation while is the forecast variable which is dependent on previous values yt-1 ,yt-2,….,yt-p , and therefore the name auto regression (AR) . the following equation while

$$y_t = b_0 + b_1 e_{t-1} + b_2 e_{t-2} + ... + b_q e_{t-q} + e_t \quad .... (3\text{-}6)$$

For this equation (3-6) , we denote that the forecast variable yt is dependent on previous values for the error term et . because it is defined

Transformation* *:* the main approach for achieving stationary in variance is through a logarithm or power transformation of the data.

as a moving average of the error series e_t (MA) for this season is called (ARIMA) , autoregressive integrated moving average .

The general non-seasonal model is known ARIMA(p,d,q):

AR: p=order of the autoregressive part .

I: d= degree of first difference involved.

MA:q= order of the moving average part .

And the general seasonal model is known as ARIMA(p,d,q) $(P,D,Q)_s$

(p,d,q) is represent non-seasonal part of the model

$(P,D,Q)_s$ is represent seasonal part of the model, when s is number of periods per season. [Martins 2007P. 19-20]

3.4.1 High order moving average models

The general MA model of order q can be written as following:

$$y_t = c + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - ..... - \theta_q e_{t-q} = c + e_t - \sum_{j=1}^{q} \theta_j e_{t-j} \quad ....(3\text{-}7)$$

Where:

c = constant term

$\theta_j =$ jth moving average parameter

et-q=the error term at time t-q

Using back shift operator equation (3-7) can be written as:

$$y_t = (1 - \theta_1 B - \theta_2 B^2 - ..... - \theta_q B_q) e_t + c$$

Where is θ_j restricted between -1 and 1

3.5 How to recognize pattern in ACF and PACF

If we want recognize suitable patterns for time series data, that is does through ACF and PACF which is compute of data. This process is clear by the following table. [Milan 2007P. 265]

process	ACF	PACF
AR(1)	Exponential decay: on positive	Spike at lag 1, then cuts of to zero

	side if ϕ_1>0 and alternating in sign starting on negative side if ϕ_1<0.	spike positive if ϕ_1>0, negative if ϕ_1<0.
AR(p)	Exponential decay or damped sine-wave the exact pattern depends on the sign and sizes of ϕ_1,….., ϕ_p .	Spikes al lag 1 to p then cuts off to zero.
MA(1)	Spike at lag 1 then cuts off to zero: spike positive if $\theta_1 < 0$, negative if $\theta_1 > 0$	Exponential decay:on negative side if $\theta_1 > 0$ and alternating in sign starting on positive side if $\theta_1 < 0$.
MA(q)	Spikes at lags 1 to q then cuts off to zero.	Exponential decay or damped sine-wave the exact pattern depends on the sign and sizes of $\theta_1,\ldots\ldots,\theta_q$.

Table (3-1) How to recognize pattern in ACF and PACF

4-Practice part

4.1 Plot the data

Plot the data is clear that the data is non-stationary in the mean and in the variance because the variation in the magnitude of the fluctuations with time is referd to non-stationary in the variance as show in figure (4-1).

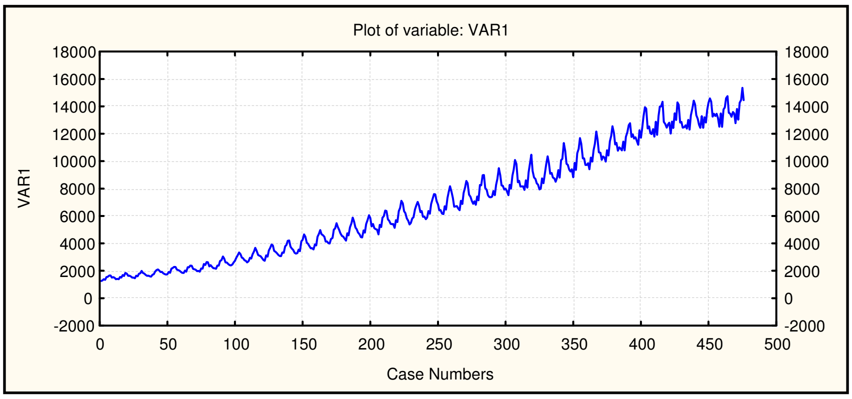

Figure (4-1) non-stationary in the variance

First step is stabilize the data in the variance by using transformation function Wt=log(Yt) ,show figure (4-2) stationary in the variance .

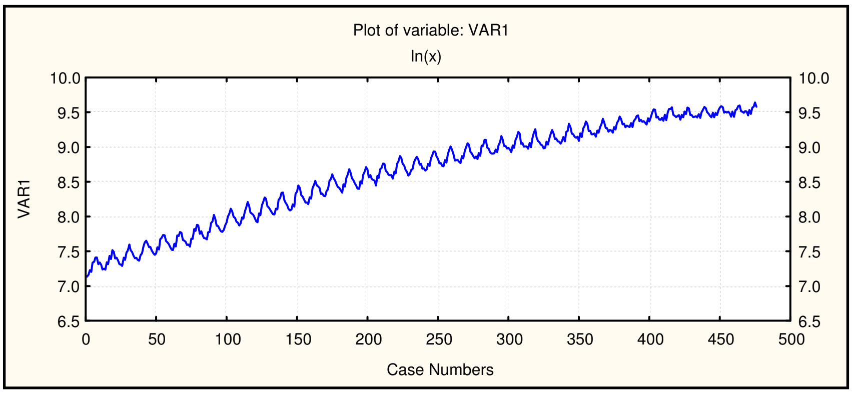

Figure (4-2) stationary in the variance

After that, the data is become stabilize in the variance but it is remain non-stationary in the mean, because there are several a rounds and declines in the data, to make stationary, we take first difference, because the data in figure (4-1) show the seasonality in the series and this evidence on the seasonality data. Figure (4-3) show the stationary in the mean after take seasonality difference.

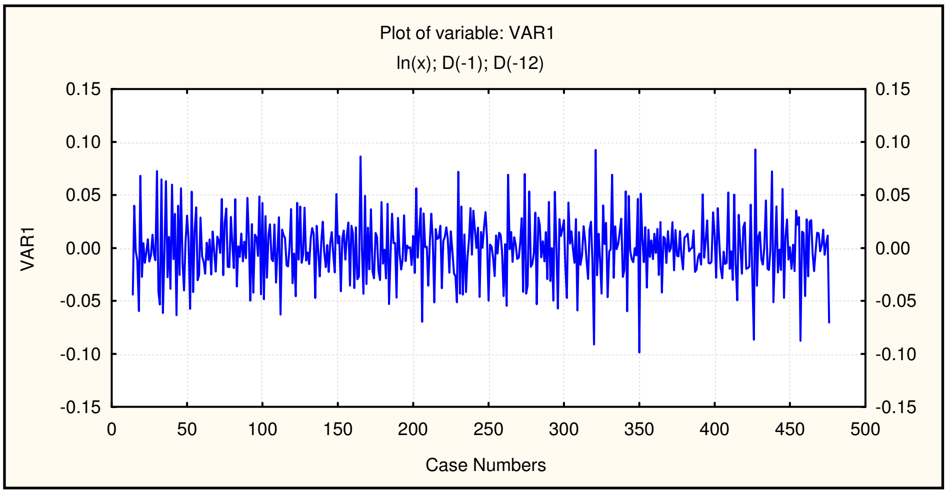

Figure (4-3) stationary in the mean

4.2 Model selection

In this point, we will examine the data ACF & PACF, after get on the stationary for time series examine ACF &PACF figure (4-4 a)&(4-4 b) are shows the ACF &PACF respectively for stationary data .note that PACF is exponentially decay for first few lags, in the ACF, the r_1 is significant reinforcing the non- seasonal MA(1) model and r_{12} is significant a seasonal MA(1) model. And we end up with the tentative identification: [Box 1976 P.289]

ARIMA $(0,1,1)(0,1,1)_{12}$

Or

$$(1-B)(1-B^{12})y_t = (1-\theta_1 B)(1-\Theta_1 B^{12})e_t \quad \text{..........(4-1)}$$

Non-seasonal Difference — Seasonal difference — Non-seasonal MA(1) — Seasonal MA(1)

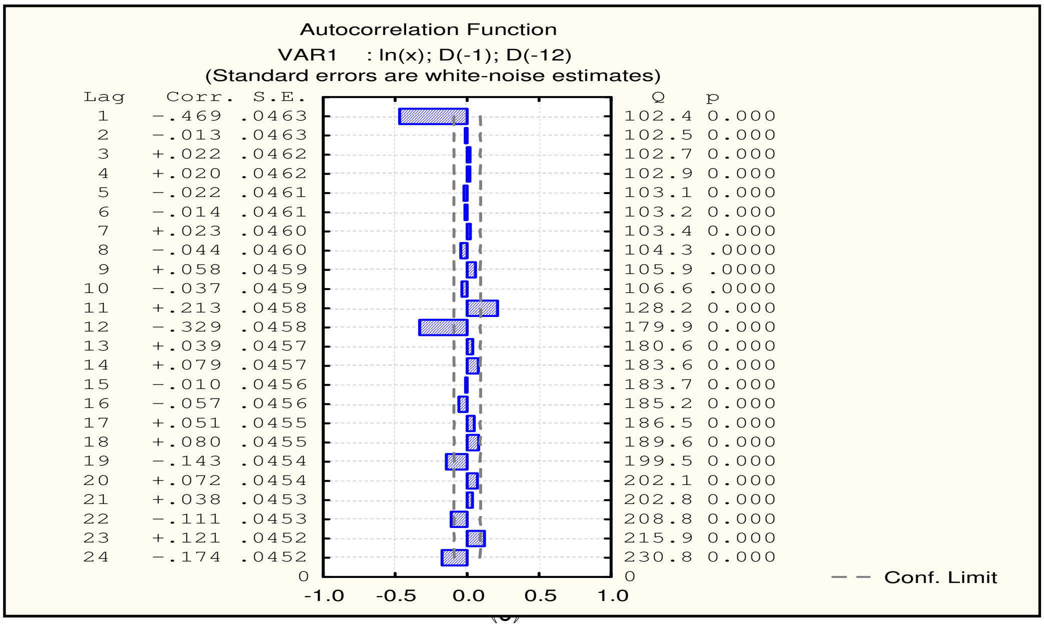

Figure (4-4 a) ACF

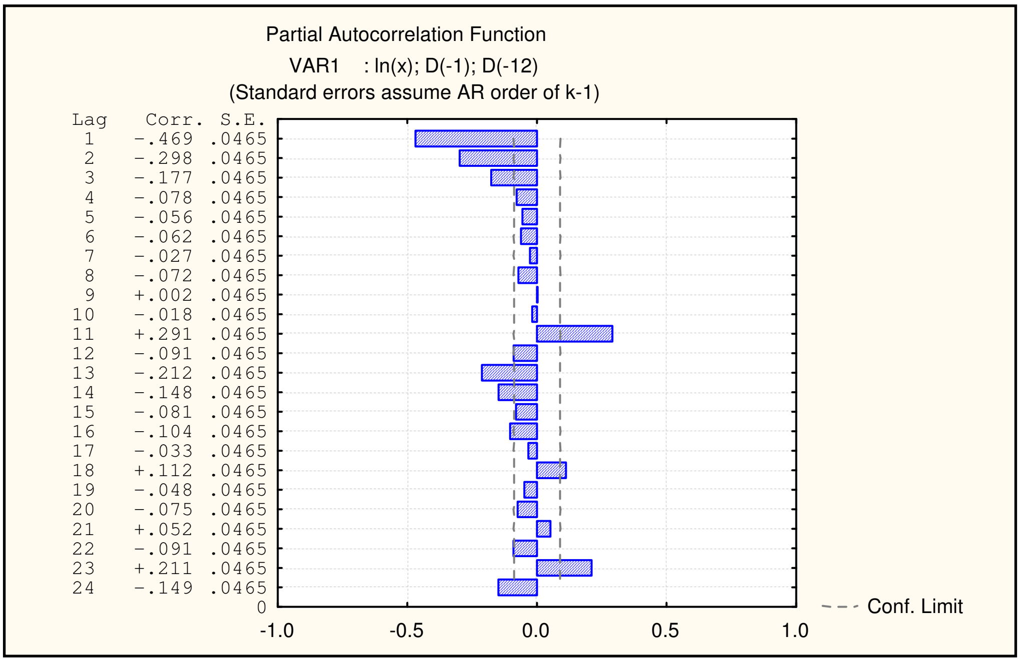

Figure (4-4 b) PACF

Note that this model is some time called the (air line model) because it was applied to international air line data by Box& Jenkins (1970). It is one of most commonly used seasonal ARIMA models .

4.3 Estimating the parameters

After make a tentative model identification for our problem the MA parameters non-seasonal, seasonal Should be determined.

Computer program for fitting ARIMA models will automatically find appropriate initial estimate of the parameters and then successively refine them until the optimum values of the parameters are found using either the least square or maximum likelihood criteria. The parameters for our model (0,1,1)(0,1,1)12 . Are clear from the following table.

Parameters	Estimate	Std.error	Z	P-value
θ_1	0.66418	0.03607	18.413	0.000
Θ_1	0.66264	0.03494	18.965	0.000

Table (4-1) parameters for ARIMA model $(0,1,1)(0,1,1)_{12}$

4.4 identification revisted

After estimated an ARIMA model, it is necessary to revisit to see if the selected model can be improved, there are many criteria to this purpose ,some of these criteria are mean square error (MSE),mean absolute percentage error, and **Akiake's information criteria (AIC) *.**

To see that the following table clear some ARIMA models for the (production of electrics in Australia), by taking minimum (MSE,MAPE and AIC) of ARIMA model, for these ARIMA models, the model (0,1,1)(0,1,1)12 , is the best ARIMA model for our time series .

ARIMA model	**MSE**	**MAPE**	**AIC**
$(0,1,1)(0,1,2)_{12}$	6631299	16.085	-850.04617

$(0,1,1)(0,1,1)_{12}$	**3770881**	**12.043**	**-853.48265**
$(0,1,2)(0,1,2)_{12}$	5483764	14.791	-849.02081
$(0,1,2)(0,1,1)_{12}$	3879010	12.364	-851.47951

Table (4-2) some ARIMA models for production of electrics

4.5 Diagnosis checking

For a good forecasting model, the residuals after fitting the model should be simply white noise, there force, if the ACF & PACF of the residuals are obtained, we would hope to find no significant autocorrelation and no significant partial autocorrelation. The residuals from the fitted our ARIMA model $(0,1,1)(0,1,1)_{12}$ are analyzed in figure (4-5)(4-6) respectively . as show following :

Akiake's information criteria (AIC)* , the form to AIC is: AIC=(n(1+log(2Π))+(n log σ^2 +2m)) .where m=p+q+P+Q be the number of terms estimated in the model .then we can choose the values of p,d,P and Q by minimizing (AIC). σ^2 is the variance of the residuals and n is the number of observations in the series .

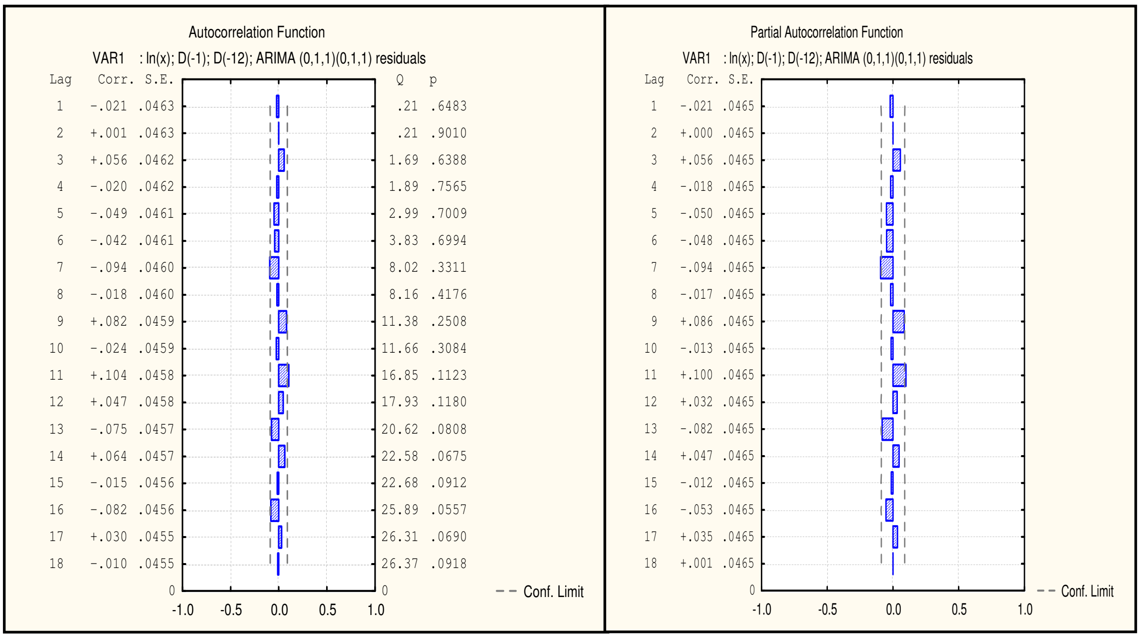

Figure (4-5) residuals for ACF
Figure (4-6) residuals for PACF

We show that the spikes of ACF and PACF are within the limit of ACF &PACF

4.6 Forecasting with ARIMA model

An ARIMA (0,1,1)(0,1,1)12 model is described as

$$(1-B)(1-B^{12})y_t = (1-\theta_1 B)(1-\Theta_1 B^{12})e_t \qquad \text{..........(4-2)}$$

Forecasting equation for above model we can obtain by simple algebra operation, for the model above , forecasting equation is:

$$y_t = y_{t-1} + y_{t-12} - y_{t-13} - \theta_1 e_{t-1} - \Theta_1 e_{t-12} + \theta_1 \Theta_1 e_{t-13} + e_t \qquad \text{...(4-3)}$$

In this research, consider the ARIMA (0,1,1)(0,1,1)12 fitted to the (production of electrics in Australian) . This model will be used to forecast for two years ahead.

Now in order to forecast for two years ahead (24 months) from (Sep 1995 to Aug 1997) , forecast values are computing by statistica program v. 6 as show in table (4-3).

month	year	Forecasting values
Sep	1995	14117
Oct	1995	13806
Nov	1995	13722
Dec	1995	13879
Jan	1996	14223
Feb	1996	13671
Mar	1996	14196
Apr	1996	14260
May	1996	14618
Jun	1996	14660
Jul	1996	15207
Aug	1996	14409
Sep	1996	14453
Oct	1996	14496
Nov	1996	14541
Dec	1996	14585
Jan	1997	14629
Feb	1997	14674
Mar	1997	14718
Apr	1997	14763
May	1997	14808
Jun	1997	14853
Jul	1997	14898
Aug	1997	14944

Table (4-3) forecasts for the production of electrics

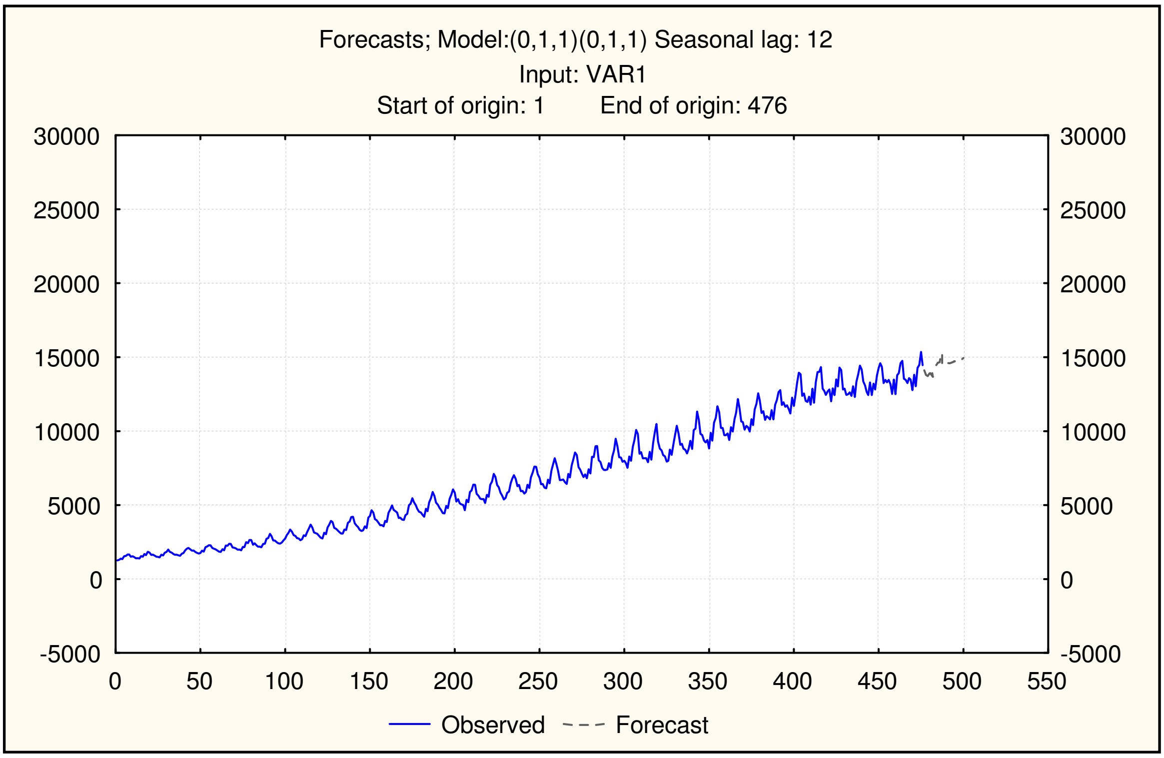

Figure (4-7) forecast for the production of electrics

5. Recommendation and Conclusion

1-The model $(0,1,1)(0,1,1)_{12}$ it was adequate ARIMA model, because it is minimum mean square error, mean absolute percentage error and minimum Akiake's information criteria .

2- The work is continuous to practice the ARIMA models on data taken from our country.

References

1- Ajoy .K. palit & Dobrivoje popovice " **Computational intelligence in time series forecasting**" 2005

2- Alankang.chin-sheng "**identification of autoregrresive integrated moving average time series**" Arizona state university a dissertation presentedin partial ful fillment of the reqirements for the degree doctor of philosophy .may 1980

3-Box G.E.P and Jenkins G.M " **time series analysis forecasting and control** "(1976).

4- Joseph George (2006)" **The Box-Jenkins forecasting technique**".

5- Martins.Levy (2007**)" forecasting and time series methods**".

6- Milan Marcek (2007)" **Economic time series forecasting** :Box-Jenkins methodology and signal processing approach"

7- Paulo Cortez & Miguel Rocha (2004)" **Evolving time series forecasting ARIMA models**".

8-Spyros Makridakis and Steven C. Wheelwright "**forecasting methods and applications**" 3rd edition (1998).