Production of Sugarcane Forecasting using ARIMAX Model

M Gopinath, M Kavithamani

Abstract— The main objective of this study is to analyze the production of sugarcane in India and to forecast the sugarcane production using Auto Regressive Integrated Moving Average with Exogenous variable or Inputs model (ARIMAX). This model mainly focused in the area of forecasting and it gives the accurate prediction. Data has been collected from Sugarcane Breeding Institute, Coimbatore. The ARIMAX model was introduced by Box and Tiao in their study. At present this model was used minimum amount of people in their studies. When we compare with ARIMA model, the ARIMAX model gives the greater accuracy.

Index Terms—Sugarcane, Prediction, ARIMA, ARIMAX, India.

I. INTRODUCTION

Sugarcane also called as 'Saccharum officinarum' which is grown all over in India. It also gives more employment opportunity to the millions of people who all are related to this field. Sugarcane was developed in 'New Guinea' 1000 years before forming of sugarcane have been started in vedic period. In earliest day the cultivation of sugarcane is found in 1400-1000 B.C in Indian writings. Now a day's world accepted that India is the home for saccharum species.

There are different climatic place for sugarcane cultivation in all over India. i.e tropical region and subtropical region. There are five climatic area is identified for the purpose of development of sugarcane. (i). Northwest area (ii). North central area (iii). Northeast area (iv). Penisular area (v). Coastal area. The type of a zone called tropical given 45% and 55% of total amount of production in India. In subtropical zone the production of sugarcane is vise-versa (55% and 45%).

The growth of sugarcane is high when the climatic condition should be in between latitude 36.70 northern area and 31.00 for southern areas of the equator enlarging tropical to subtropical. The weather condition alone decides the productivity of sugarcane and juice conditions. In sugarcane production Maharashtra plays vital role in India, it covers the area of 9.4 lakh, with produces 61.32 million ton and Tamilnadu is also produce sugarcane high. The lower production of sugarcane is only in the state Bihar. One of the largest agro-based production industries is sugarcane.

Gur and kandsari in sugarcane:

Gur is the legitimate commodity of sugarcane and it is also called as unrefined sugar form. Gur is like brown coloured around full of sucrose, so that the colour of gur being brown. Making of gur is much easier than making of sugar, because no need of frequent refinement process to make gur. The gur is mainly used to make sweet items and it is usually the consuming commodity. In India the production of gur is abot 60% , and the country Brazil is a top most exporter of jaggery.

Kandsari is a classical organic unprocessed sugar that consist fragment minerals and fibre. Kandsari is made with a help of cold raw juice of sugarcane. Without adding any chemicals the organic form of biggest sugar called kandsary and it is a kind of unrefined sugar in a large size.

Objective:

The main objective of our study is to establish a relevant ARIMAX model to be measure the production of sugarcane forecasting using ARIMAX model. The major objective of this study is based on Autoregressive Integrated Moving Average with exogenous variable (ARIMAX) to several kind of sugarcane production such as seed and feed(%) and Gur & khansari as exogenous variable

II. REVIEW OF LITERATURE

There are several author have been worked under this ARIMAX model but in the area of agriculture there are very few of them have been worked under the ARIMAX model. Mohammed Amir Hamjah (2014) have applied this model with a title " Climatic Effects on major pulse in Bangladesh an application of Box-jenkins ARIMAX model "In this study they conclude that instead of using other models such as ARIMA,SARIMA etc.. suggested that ARIMAX for best prediction. Uyothu Amekauma Victor-Edema, ProofIsaac Didi Essi (2016) in the title" The ARIMAX for Nigerian non-oil Export" in this study the author have been explained about ARIMAX model fitting, reason for choosing X as a exogenous variable and also given a figure to show the roots of characteristics polynomial. Mofeng YANG, Jiaohong XIE, Peipei MAO, Chao WANG, Zhirui YE (2018) on the topic "Application of ARIMAX on Forecasting Freeway Traffic Flow" have been discussed about ARIMAX model implementation, Stationary test, Series differencing, ARIMA parameter estimation, in that they discovered both ARIMA and ARIMAX model which gives better performance in urban areas compare to sub-urban areas. Durka peter and Pastorekova Silvia worked under the topic" ARIMA vs ARIMAX – based on these two analysis which approach is enhanced to analyse and forecast the macroeconomic time series models?" have been

Revised Manuscript Received on September 14, 2019.

M Gopinath, Department of Mathematics and Statistics, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu, India.(E-mail: gopimalaisamy@gmail.com)

M Kavithamani, Department of Mathematics and Statistics, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu, India.(E-mail: statisticskavi@yahoo.co.in)

distinguished between ARIMA and ARIMAX and finally they proved that ARIMA model is nearly seems to be an scientific model rather than ARIMAX model.

III. METHODOLOGY

A time series means that a number that represents the measures of certain activity more equally gaps time interval. It is on among the method which is mainly focused historical record in an equal interval space with consistency in measurement.

Method of Moving Average:

Method of Moving average models was proposed by Slutsky (1927) and Wold (1938). The series of moving average can be written as

$$Y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \theta_3\varepsilon_{t-3} - \ldots - \theta_q\varepsilon_{t-q} \quad (1)$$

This is called as method of moving average in the order q (MA q) here Y_t is the series which contains original data and ε_t is the error term of the series.

Method of Auto Regressive Process:

The Auto Regressive process was first proposed by Yule (1926) this regressive process is specially satisfies the following equation

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \ldots + \phi_p Y_{t-p} \quad (2)$$

The present value Y_t is called as linear combination for p (AR P) also ε_t is the assumption of independent of $Y_{t-1}, Y_{t-2}, Y_{t-3} \ldots Y_{t-q}$.

The Auto Regressive Integrated Moving Average (ARIMA):

The ARIMA methodology was proposed by the author Box and Jenkins in the year 1976. This ARIMA model is also called as Box- Jenkins model. This model is based on the error term of the time series. To get conclude time series the data should be stationary which means, the mean, variance and covariance are constants over the time period. For this we can write the ARIMA model equation as follows.

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \ldots \phi_p Y_{t-p} + \phi_1 Y_{t-s} + \phi_2 Y_{t-2s} + \ldots + \phi_p Y_{t-ps} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q} - \Theta_1 a_{t-s} - \Theta_2 a_{t-2s} - \ldots \Theta_q a_{t-qs} \quad (3)$$

With a help of back shift operator (lag) we write the above equation as follows:

$$\phi_p(B)\phi_p(B^s) z_t = \theta_q(B)\Theta_q(B^s) a_t \quad (4)$$

Here:

$$z_t = (1-B)^d (1-B^s)^D \ln(Y_t)$$

$\phi_p(B)$ – *Non seasonal operator of Autoregressive process AR(p)*

The Auto Regressive Integrated Moving Average with Exogenous variable (ARIMAX):

This model is a development of Autoregressive Integrated Moving Average (ARIMA). This ARIMA consist of three parameters specifically p,d and q, p represent term of auto regressive, q represents the term of moving average and d is used to make the term stationary. The ARIMA (p,q) model is expressed as follows:

$$\phi(L).\Delta^d Y_t = \theta(L)\varepsilon_t \quad (5)$$

Where

$\phi(L)$ *is the autoregressive polynomial and it is* exp *ressed as*

$$\phi(L) = (1 - \phi_1 L^1 - \phi_2 L^2 - \ldots - \phi_p L^p \quad (6)$$

And

$\theta(L)$ *is called the moving average polinomial and it is* exp *ressed as*

$$\theta(L) = (1 - \theta_1 L^1 - \theta_2 L^2 - \ldots \theta_q L^q \quad (7)$$

Y_i is the dependent variable in time t, Δ^d is represent as degree of difference and ε_t is the process of white noise. By adding equation (6) we get

$$\phi(L).\Delta^d Y_t = \varphi(L) X_t + \theta(L)\varepsilon_t$$

Where (8)

$$\varphi(L) = (1 - \varphi_1 L^1 - \varphi_2 L^2 - \ldots - \varphi_r L^r)$$

The equation (8) represents the usual form of ARIMAX (p,d,q) model.

Model Fitting:

In ARIMAX model also called Box-Jenkins with exogenous variable model adopts certain approach like identification of the problem, forecasting and diagnostic checking also checked the given information, missing value of the data are stationary. In this work I applied Augmented Dickey-Fuller (ADF) test to obtain the conformation of the stationary series. Here applied Autocorrelation function (ACF) and partial autocorrelation function (PACF) have combined together to get the order of AR and MA of this model. With a help of Akaike information criterion (AIC) the suitable model has been selected for this study.

IV. RESULTS AND DISCUSSION

The following figures are explains about the production of sugarcane for million tons, their seed and feed in percentage and gur & khandsari in percentage. In the figure -1, shows that there is a gradual increase in the pattern of production of sugarcane over a period of time. The figure-3 shows that there is gradual increase in seed and feed for a while and suddenly it comes very low for a particular year and started increasing trend. The figure-5 shows that gradual decrease from the starting year itself. The first difference of the variable production of sugarcane, seed and feed and gur & khandasari with (p-value = 0.0000).

The ARIMAX model gives the effects of production on seed and feed and gur and khandsari. The autocorrelation function (ACF) explains that autoregressive model capable of order 1. i.e. AR(1) when the partial autocorrelation function (PACF) Explains that a moving average capable order 2. i.e. MA(2). Comparatively the models make use of Akaike information criteria (AIC), the appropriate model has been developed as ARIMAX (1, 1, 2).

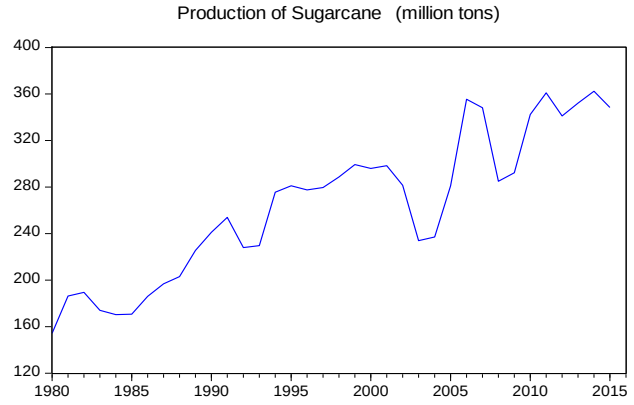


Figure – 1

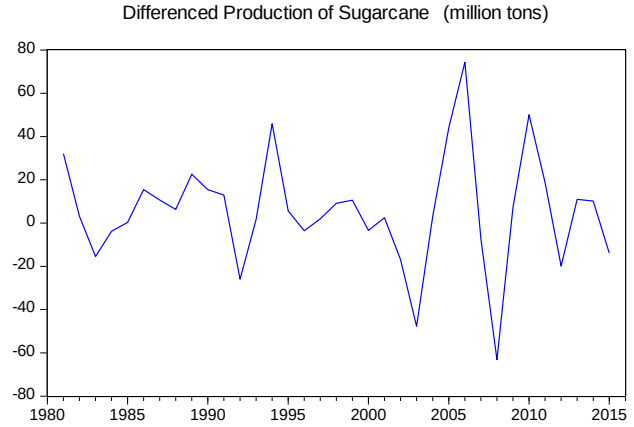


Figure – 2

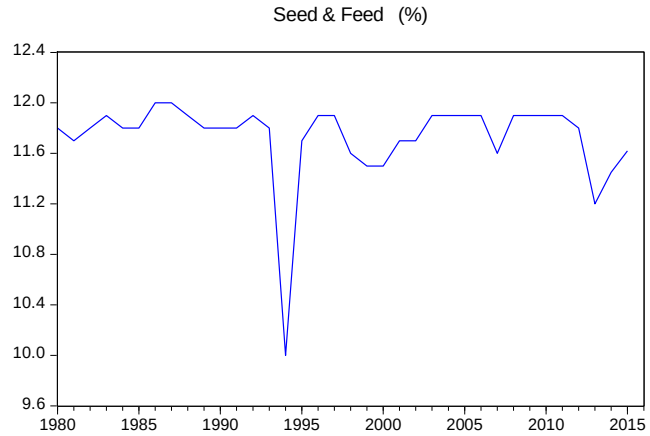


Figure – 3

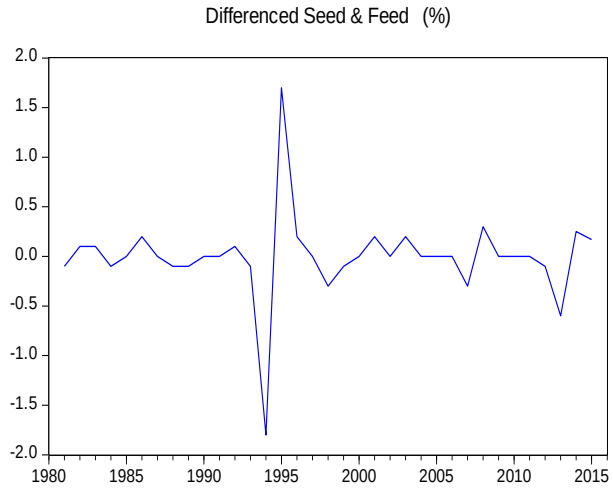


Figure – 4

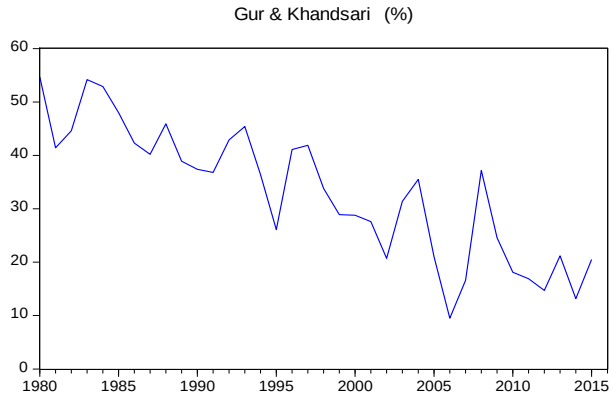


Figure – 5

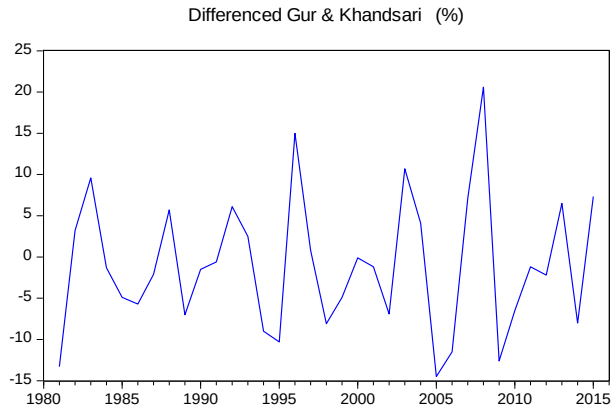


Figure – 6

The ADF test for the variable production of sugarcane (p-value = 0.7691), seed and feed (p-value = 0.0002) and gur & khansari (p-value = 0.8096) these values have obtained at 0.05 level of significance, as non-stationary.

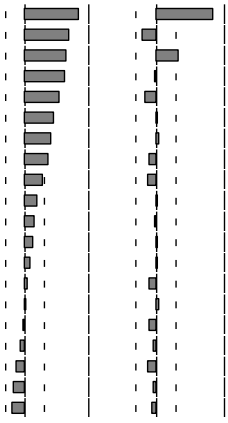
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.879	0.879	34.048	0.000
		2	0.729	-0.192	58.064	0.000
		3	0.673	0.357	79.096	0.000
		4	0.650	-0.017	99.201	0.000
		5	0.570	-0.161	115.10	0.000
		6	0.473	0.007	126.38	0.000
		7	0.423	0.043	135.64	0.000
		8	0.380	-0.101	143.37	0.000
		9	0.294	-0.122	148.13	0.000
		10	0.207	0.011	150.57	0.000
		11	0.159	-0.022	152.05	0.000
		12	0.134	0.031	153.14	0.000
		13	0.098	0.018	153.75	0.000
		14	0.042	-0.085	153.86	0.000
		15	0.006	0.045	153.87	0.000
		16	-0.022	-0.090	153.90	0.000
		17	-0.063	-0.039	154.19	0.000
		18	-0.125	-0.120	155.38	0.000
		19	-0.176	-0.040	157.87	0.000
		20	-0.208	-0.049	161.49	0.000

Table - 1

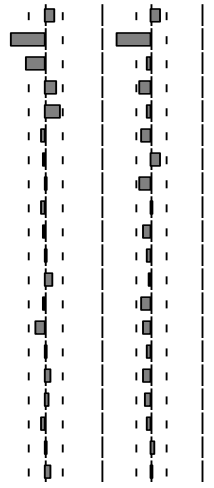
Autocorr...	Partial Correl...		AC	PAC	Q-Stat	Prob
		1	0.184	0.184	1.4641	0.226
		2	-0.639	-0.697	19.520	0.000
		3	-0.364	-0.071	25.527	0.000
		4	0.212	-0.223	27.618	0.000
		5	0.274	-0.097	31.213	0.000
		6	-0.055	-0.182	31.362	0.000
		7	-0.048	0.207	31.479	0.000
		8	-0.015	-0.252	31.492	0.000
		9	-0.087	0.051	31.900	0.000
		10	-0.049	-0.167	32.033	0.000
		11	0.034	-0.083	32.102	0.001
		12	0.126	-0.031	33.049	0.001
		13	-0.026	-0.176	33.091	0.002
		14	-0.162	-0.162	34.777	0.002
		15	-0.011	-0.059	34.785	0.003
		16	0.121	-0.144	35.805	0.003
		17	0.061	-0.128	36.074	0.004
		18	-0.067	-0.058	36.418	0.006
		19	0.027	0.064	36.475	0.009
		20	0.098	-0.014	37.284	0.011

Table – 2

Table – 1 represent the auto correlation function for the data. It shows that our data is not stationary. Then move on to the next step called first differenced auto correlation. Table – 2 shows that the data is stationary.

ADF Test

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.452008	0.8896
Test critical values:	1% level	-3.615588	
	5% level	-2.941145	
	10% level	-2.609066	

*MacKinnon (1996) one-sided p-values.

Table – 3

ADF Test for difference

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-9.164991	0.0000
Test critical values:	1% level	-3.615588	
	5% level	-2.941145	
	10% level	-2.609066	

*MacKinnon (1996) one-sided p-values.

Table – 4

To check further our data is stationary, the Phillips-Perron Unit root test is also conducted and the result has been given in the table – 3. It shows the data is non-stationary therefore we conduct the first differenced ADF test, hence it is found that the data became stationary and it has been shown in table – 4.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(PRODUCTION OF SUGARCANE)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SUGARCANE_(-1)	-0.055341	0.060142	-0.920166	0.3651
D(SUGARCANE_(-1))	0.364963	0.135243	2.698562	0.0115
D(SUGARCANE_(-2))	-0.661789	0.134349	-4.925895	0.0000
C	21.66137	16.12981	1.342941	0.1897

R-squared	0.520284	Mean dependent v	4.816455
Adjusted R-squared	0.470658	S.D. dependent var	26.00468
S.E. of regression	18.91993	ADF	8.831522
Sum squared resid	10380.95	SBC	9.012916
Log likelihood	-141.7201	Hannan-Quinn	8.892555
F-statistic	10.48415	Durbin-Watson stat	2.035684
Prob (F-statistic)	0.000077		

Table - 5

Based on the diagnostic test we have selected the ARIMAX model as bellow Table – 6. It shows that the ARIMAX (1, 2, 2) and the SMA (18) also found. With a help of diagnostic testing we have chosen thesis p and q. The quality of the ARIMAX model have chosen with a help of the program and it is listed in the below table.

Dependent Variable: D(PRODUCTION_OF_SUGARCANE_)
Method: ARMA Maximum Likelihood (OPG-BHHH)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	67.53746	92.31947	0.731562	0.4705
SEED FEED	-5.232068	7.755292	-0.674645	0.5054
GUR KHANDSARI	-0.013975	0.285143	-0.049011	0.9613
AR(1)	0.322231	0.180058	1.789594	0.0843
MA(2)	-1.000000	3033.849	-0.000330	0.9997
SMA(18)	-1.000000	21638.91	-4.62E-05	1.0000
SIGMASQ	156.3222	1612560.	9.69E-05	0.9999

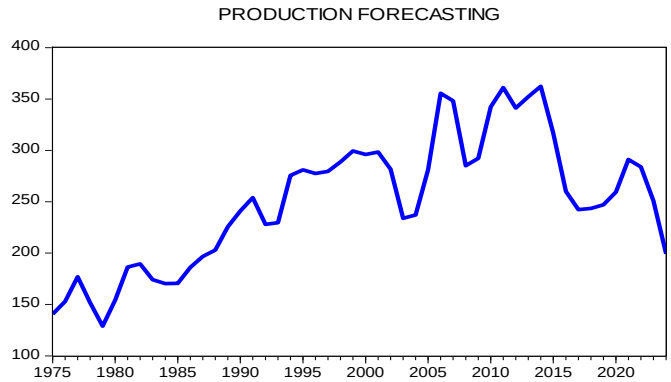
R-squared	0.755405	Mean dependent	5.54857
Adjusted R-squared	0.702992	S.D. dependent var	25.6496
S.E. of regression	13.97865	Akaike info criterion	9.03038
Sum squared resid	5471.276	Schwarz criterion	9.34145
Log likelihood	-151.0317	Hannan-Quinn criter.	9.13776
F-statistic	14.41253	Durbin-Watson stat	1.85114
Prob (F-statistic)	0.000000		

Table - 6

Forecasting:

This forecasting is based on the past time series data (1975 to 2016) the forecasted values are shown in the table – 7.The forecasted diagram - 7 is also given.

Forecasting of Sugarcane from 2017-2025



Year	Production of Sugarcane (million tons)	Year	Production of Sugarcane (million tons)	Year	Production of Sugarcane (million tons)
1975	140.60	1992	228.03	2009	292.30
1976	153.00	1993	229.66	2010	342.38
1977.	176.96	1994	275.54	2011	361.03
1978	151.65	1995	281.1	2012	341.19
1979	128.83	1996	277.56	2013	352.14
1980	154.24	1997	279.54	2014	362.33
1981	186.35	1998	288.72	2015	348.44
1982	189.50	1999	299.32	**2016**	**260.05**
1983	174.07	2000	295.95	**2017**	**242.27**
1984	170.31	2001	298.42	**2018**	**243.47**
1985	170.64	2002	281.57	**2019**	**246.97**
1986	186.09	2003	233.86	**2020**	**259.45**
1987	196.73	2004	237.08	**2021**	**291.15**
1988	203.03	2005	281.17	**2022**	**283.84**
1989	225.56	2006	355.52	**2023**	**250.92**
1990	241.04	2007	348.18	**2024**	**199.25**
1991	253.99	2008	285.02	**2025**	**196.63**

V. CONCLUSION

In this article, we have constructed a special model called ARIMAX. Based on this model we came to know the future values which means that the forecasting value of sugarcane production 2015 to 2026. The figure – 7 is clearly explains the future production of sugarcane per million also Table – 7 explains that the amount of change in future production. This indicates that there will be a gradual decrease production of sugarcane; we should take the necessary steps in order to improve the production.

REFERENCES

1 Box, G. E. P. – Jenkins, G. M. – Reinsel, G. C.: Time Series Analysis: Forecasting and Control. John Wiley & Sons Inc., New York, 2008.
2 Box, G. E. P. – Tiao, G. C.: Intervention analysis with applications to economic and environmental problems. Journal of the American Statistical Association. 1975, vol. 70, no. 349, p. 70-79.
3 Marek, L.: Transfer function models. Acta Oeconomica Pragnesia. 2000, vol. 8, no. 3, p. 83-94
4 Pankratz, A.: Forecasting with Dynamic Regression Models. Wiley-Interscience, 1991.
5 Rublíková, E. – Marek, L.: Linear transfer function model for outflow rates. Ekonomické rozhľady. 2001,vol,30,no. 4, p.457 – 466.
6 De Felice, M., Alessandri A., Ruti P. M.,(2013). Electricity demand forecasting over Italy: potential benefits using numerical weather prediction models. Electric Power Systems Research. Retrieved on 16/10/2015 www.researchgate.net/publication/263927372
7 Wangi K., Singhasivanon, P., Silawan, T., Lawpoolsri, S., White, N. J.,& Kaewkungwal, J. (2010). Development of temporal modeling for forecasting and prediction of malaria infections using time series and ARIMAX analysis: a case study in endemic districts of Bhutan. Malaria Journal. 9(251).
8 8.L.R.Dhumne,"Electroniccommerce:acurrenttrend,"InternationalJournalonInformationTechnologyManagement,2012.
9 X. Q. Zhang and S. W. Chen, "Forecast of Chinas Forestation Area Based on ARIMA Model," Chinese Forestry Science & Technology, vol.5, no.2,pp. 50–55,2010.
10 K. W. Wang, C. Deng, and J. P. Li, "Hybrid methodology for tuberculosis incidence time-series forecasting based on ARIMA and a NAR neural network," Epidemiology & Infection, vol. 1, 2017.
11 S.Wei, D.Zuo, and J.Song, "Improving prediction accuracy of river discharge time series using a Wavelet-NAR artificial neural network," Journal of Hydro informatics, vol. 14, no. 4, pp. 974– 991,2012.
12 P. Geng, L. Na, and L. Ben-Fu, "Research on the Prediction of E-commerce Transaction Volume-Based on the Prediction Model of Search Data and Commodity Classification in Station,"Management Modernization, vol.2,pp.30–32,2014