e-ISJN: A4372-3114
p-ISJN: A4372-3115

ISSN: 2321-7782 (Online)
ISSN: 2347-1778 (Print)
Impact Factor: 7.327

Volume 6, Issue 2, February 2018

International Journal of Advance Research in Computer Science and Management Studies

Research Article / Survey Paper / Case Study

Available online at: www.ijarcsms.com

Special Issue: National Conference on Management, Economics & Social Science (NCMESS 2018)

Organized by: Department of Business Administration, ST. JOSEPH'S COLLEGE (AUTONOMOUS), Tiruchirappalli - 620002, India

Forecasting for the Potato production in India by using ARIMA Models

P. Rajalakshmi[1]
Research Scholar (M. Sc; M. Phil)
Department of Statistics
St. Joseph's college (Autonomous)
Trichy, India

Dr. Lilly George[2]
Asst. Professor (M. Sc; M. Phil, Ph .D)
Department of Statistics
St. Joseph's college (Autonomous)
Trichy, India

Abstract: ***Forecasting, being useful in risk management, requires a suitable model, chosen from several available techniques. The study is hence directed at finding a suitable forecasting model. A suitable model is one which is applicable to the product and data available. Selection of a suitable model requires determining efficiency of different models in predicting future outcomes and selecting the model which best suits the job of prediction.***

The objectives of the study are as follows.

- ***To develop a suitable forecasting ARIMA models for Potato production in India***
- ***To study the forecasting ability of univariate ARIMA models***
- ***To suggest an optimal model, Best forecast models selected.***

I. INTRODUCTION

- Origin of Potato :

South America is known to be native of Potato. In 1537, the Spaniards first came into contact with Potato in one of the villages of Andes. In Europe, Potato was introduced between 1580 A.D. to 1585 A.D. in Spain, Portugal, Italy, France, Belgium and Germany. In India it was introduced by the Portuguese sailors during early 17th century and its cultivation was spread to North India during the British period.

- Potato Production in the World :

Potato is grown in more than 100 countries in the world. China ranks first, followed by Russia and India. China, India, USA, Ukraine, Germany and Poland put together constitute more than 62 per cent of total global production.

- Potato Production in India In India:

Potato is cultivated in almost all states under diverse agro-climate conditions. About 85 per cent of Potatoes are cultivated in Indo-gangetic plains of North India. The states of Uttar Pradesh, West Bengal, Punjab, Bihar and Gujarat accounted for more than 80 per cent share in total production.

India is the 3rd largest potato producer in the world, after China at 1 and Russia at 2 and nited States at4.India always ranks high in the production of Potato. This vegetable is mainly used for making curry across India. Fried potato chips are also popular especially when multinational company entered Indian market. Basically potato carries enormous amount of Starch ,

some protein , little fat and rich in minerals, vitamins. This product with potato attractive products like fried potato chips, dehydrated potato chips and coloured potato flour can be manufactured.

Potato production has increased more than 85% since 1960 due to both increased production area and yield. The per capita potato consumption in India has risen from 12 kg/capita/year in the early nineties to over 16 kg/capita currently, with a slight decline in recent years (Source: FAOSTAT) However, the potato processing industry is expanding fast. The sector developing most rapidly is the snack foods sector, including potato chips. Market leader is Frito-lay with a 45 % market share. Haldiram's has a 27% market share. recently ITC(Indian Tobacco company) , has made huge inroad in the CPG market and has managed to get a market share of 11% with its potato chip "Bingo" in just 6 month. Also a dairy manufacturer (Amul) just announced to move into the snack market.

Intriguing aspect of the potato supply chain in India is the strong vertical integration: ITC bought earlier this year the Australian company Technico, that developed technology for rapid multiplication and variety improvement. Also the company Merino Industries (dehydrated potato products among many other products) has its own tissue culture laboratories for multiplication and potato variety development.

Although Central Potato Research Institute (CPRI) certainly has done a good job in developing suitable varieties for processing for the Indian cultivation conditions, the degree of involvement of processing companies in the multiplication and further development offers a lot of promise for the future potato processing potential in India.

Potato farmers in the Indian state Punjab are headed for another season of dwindling profits because the cold storage facilities are still flooded with old crop - and several other states are in a similar situation. Potato farmers in South-west Uttar Pradesh (Aligarh, Hathras, Mathura, Agra, Firozabad, Etawah, Mainpuri, Kannauj and Farrukhabad) account for roughly a fifth of India's total potato production.

II. PREDICTOR OF POTATO PRODUCTION IN INDIA USING ARIMA MODELS

ARIMA modelling:

In general, an ARIMA model is characterized by the notation ARIMA (p,d,q) where, p, d and q denote orders of auto-regression integration (differencing) and moving average respectively. Time series is a linear function of past actual values and random shocks. For instance, given a time series process $\{Y_i\}$, a first order auto-regressive process is denoted by ARIMA (1,0,0) or simply AR(1) and is given by

$$Y_i = \mu + \varphi_1 Y_{i-1} + \varepsilon_t$$

and a first order moving average process is denoted by ARIMA (0,0,1) or simply MA(1) and is given by

$$Y_i = \mu - \theta_1 \varepsilon_{i-1} + \varepsilon_t$$

Alternatively, the model ultimately derived, may be a mixture of these processes and of higher orders as well. Thus a stationary ARMA (p, q) process is defined by the equation

$$Y_i = \varphi_1 Y_{i-1} + \varphi 2Y_{i-2} + \ldots + \varphi_p Y_{i-p} - \theta_1 \varepsilon_{i-1} - \theta_2 \varepsilon_{i-2} + \ldots - \theta_q \varepsilon_{i-q} + \varepsilon_q$$

where ε_i' s are independently and normally distributed with zero mean and variance σ^2

for t = 1,2,...n. Note here that the values of p and q, in practice lie between 0 and 3. The degree of differencing of main variable Y_i .

Box-Jenkins(BJ)Methodology:

(i) Identification

The foremost step in the process of modelling is to check for the stationary of the series, as the estimation procedures are available only for stationary series. There are two kinds of stationary, viz., stationary in 'mean' and stationary in 'variance'. A cursory look at the graph of the data and structure of autocorrelation and partial correlation coefficients may provide clues for the presence of stationary. Another way of checking for stationary is to fit a first order autoregressive model for the raw data and test whether the coefficient 'φ_1' is less than one. If the model is found to be non-stationary, stationary could be achieved mostly by differencing the series. Or go for a Dickey Fuller test. Stationary in variance could be achieved by some modes of transformation, say, log transformation. This is applicable for both seasonal and non-seasonal stationary.

Thus, if 'X_i' denotes the original series, the non-seasonal difference of first order is $Y_i = X_i - X_{i-1}$

followed by the seasonal differencing (if needed)

$Z_i = Y_t - Y_{i-s} = (X_i - X_{i-1}) - (X_{i-s} - X_{i-s-1})$

The next step in the identification process is to find the initial values for the orders of seasonal and non-seasonal parameters, p, q, and P, Q. They could be obtained by looking for significant autocorrelation and partial autocorrelation coefficients (see section 5 (iii)). Say, if second order auto correlation coefficient is significant, then an AR (2), or MA (2) or ARMA

(2) model could be tried to start with. This is not a hard and fast rule, as sample autocorrelation coefficients are poor estimates of population autocorrelation coefficients. Still they can be used as initial values while the final models are achieved after going through the stages repeatedly.

(ii) Estimation

At the identification stage one or more models are tentatively chosen that seem to provide statistically adequate representations of the available data. Then we attempt to obtained precise estimates of parameters of the model by least squares as advocated by Box and Jenkins. Standard computer packages like SAS, SPSS, GRELl etc. are available for finding the estimates of relevant parameters using iterative procedures.

iii) Diagnostics

Different models can be obtained for various combinations of AR and MA individually

iv) Plot of residual ACF

Once the appropriate ARIMA model has been fitted, one can examine the goodness of $\sqrt{\text{fit}}$ by means of plotting the ACF of residuals of the fitted model. If most of the sample autocorrelation coefficients of the residuals are within the limits $\pm 1.96 / \sqrt{N}$ where N is the V number of observations upon which the model is based then the residuals are white noise indicating that the model is a good fit.

III. Review of Literature

Analysis of the area, production and productivity of Potato in the state:

Singh (1993) productivity of Potato crop under riverbed cultivation is about 330 quintal per hectare which is about 50 per cent higher than under field situations. Cultivation of Potato both under riverbed and fields is a profitable proposition but it

requires heavy investment too. Farmers face many constraints in the availability of inputs. Area has potential to produce even high yields of Potato which may be achieved by relaxing the constraints in farm supplies.

Singh and Mathur (1994) in their study "Growth and instability in production and prices of Potato in India, Agricultural Situation in India" assessed instability in Potato production in India by using the co-efficient of variation. It was found that the area and production were unstable because of the response of Potato production to prices of competing crops and the adoption of modern technology, respectively.

(http://ageconsearch.umn.edu/bitstream/113945/2/my%20thesis.pdf)

Analysis :

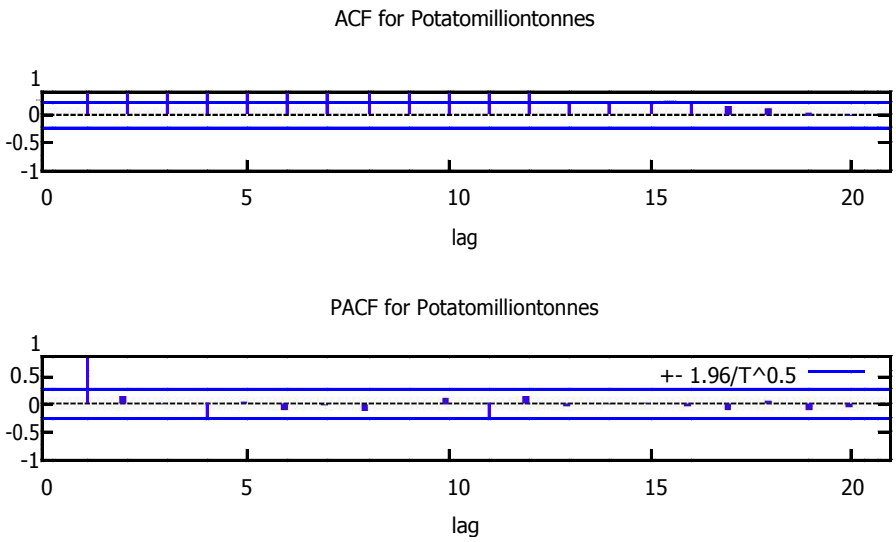
Autocorrelation function for Potatomilliontonnes

***, **, * indicate significance at the 1%, 5%, 10% levels

using standard error 1/T^0.5

LAG	ACF	PACF	Q-stat. [p-value]
1	0.9451 ***	0.9451 ***	53.6450 [0.000]
2	0.9063 ***	0.1216	103.8638 [0.000]
3	0.8675 ***	0.0016	150.7291 [0.000]
4	0.8114 ***	-0.1815	192.5087 [0.000]
5	0.7660 ***	0.0250	230..4540 [0.000]
6	0.7078 ***	-0.1296	263.4930 [0.000]
7	0.6517 ***	-0.0195	292.0622 [0.000]
8	0.5894 ***	-0.1166	315.9078 [0.000]
9	0.5292 ***	0.0008	335.5301 [0.000]
10	0.4851 ***	0.1082	352.3669 [0.000]
11	0.4157 ***	-0.2106	365.0034 [0.000]
12	0.3695 ***	0.1226	375.2048 [0.000]
13	0.3209 **	-0.0493	383.0778 [0.000]
14	0.2668 **	-0.0139	388.6451 [0.000]
15	0.2288 *	0.0138	392.8369 [0.000]
16	0.1816	-0.0477	395.5412 [0.000]
17	0.1327	-0.1064	397.0215 [0.000]
18	0.0933	0.0464	397.7719 [0.000]
19	0.0435	-0.1229	397.9397 [0.000]
20	-0.0015	-0.0630	397.9399 [0.000]

Plot:



The ACF graph for potato product dies out slowly (exponentially delaying), with one spike PACF that cuts off after lag1.data is non stationary . therefore, the first model for potato product is initially identified as ARIMA (1,1,0),ARIMA(0,1,1) .

Model 1: ARIMA, using observations 1951-2006 (T = 56)

Dependent variable: (1-L) Potatomilliontonnes

Standard errors based on Hessian

	Coefficient	*Std. Error*	*z*	*p-value*	
Const	*0.380624*	*0.124093*	*3.067*	*0.0022*	***
phi_1	*−0.537553*	*0.112548*	*−4.776*	*<0.0001*	***

Mean dependent var	0.364821	S.D. dependent var	1.699092
Mean of innovations	−0.003174	S.D. of innovations	1.418649
Log-likelihood	−99.21457	Akaike criterion	204.4291
Schwarz criterion	210.5052	Hannan-Quinn	206.7848

	Real	*Imaginary*	*Modulus*	*Frequency*
AR				
Root 1	-1.8603	0.0000	1.8603	0.5000

Test for ARCH of order 4 -

Null hypothesis: no ARCH effect is present

Test statistic: LM = 12.6544

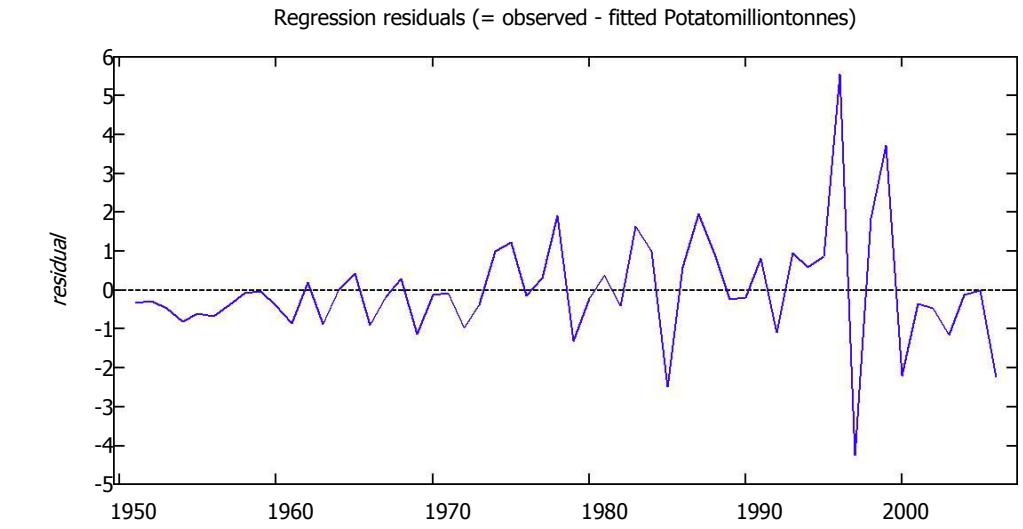
with p-value = P(Chi-square(4) > 12.6544) = 0.013094

Test for autocorrelation up to order 4

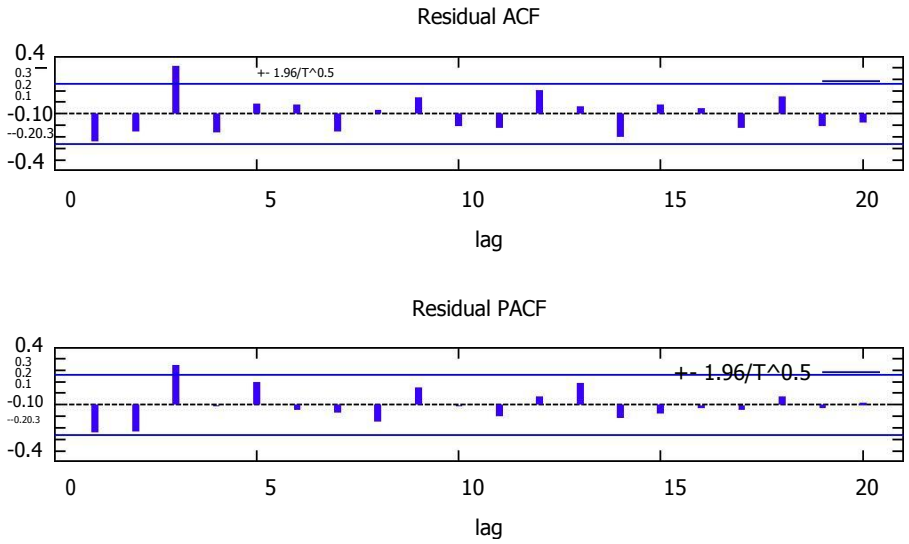
Ljung-Box Q' = 17.0657,

with p-value = P(Chi-square(3) > 17.0657) = 0.0006851

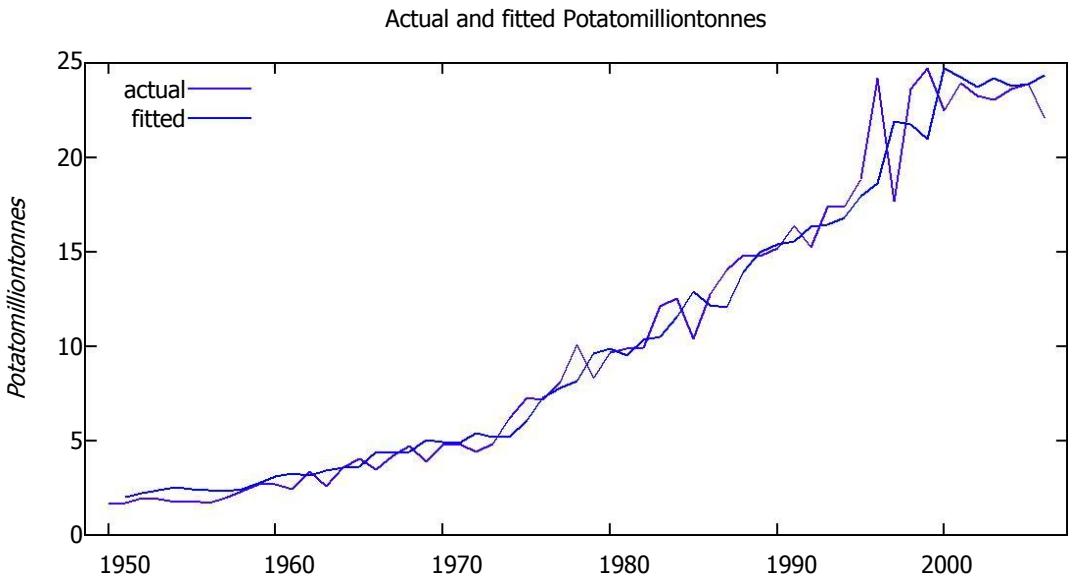
REGRESSION RESIDUALS PLOT:



RESIDUAL PLOT:



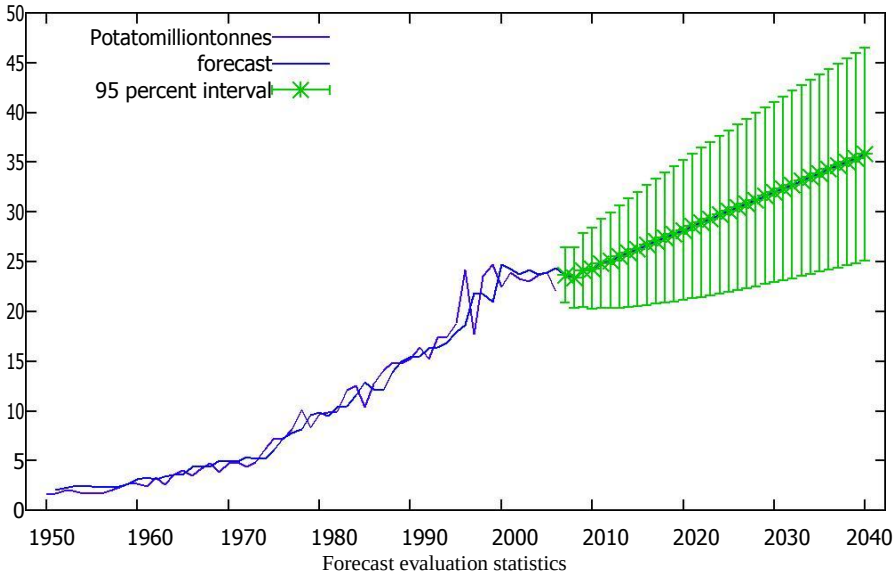
ACTUAL AND FITTED PLOT:



For 95% confidence intervals, z(0.025) = 1.96

Obs	prediction	std. error	95% interval
2007	23.6536	1.41865	(20.8731, 26.4341)
2008	23.3983	1.56300	(20.3349, 26.4617)
2009	24.1208	1.89190	(20.4127, 27.8288)
2010	24.3176	2.07229	(20.2560, 28.3792)
2011	24.7970	2.28557	(20.3174, 29.2767)
2012	25.1246	2.45653	(20.3098, 29.9393)
2013	25.5337	2.62833	(20.3823, 30.6851)
2014	25.8990	2.78345	(20.4435, 31.3545)
2015	26.2879	2.93348	(20.5384, 32.0374)
2016	26.6641	3.07460	(20.6380, 32.6902)
2017	27.0471	3.21035	(20.7549, 33.3392)
2018	27.4264	3.34016	(20.8798, 33.9730)
2019	27.8077	3.46533	(21.0158, 34.5997)
2020	28.1880	3.58602	(21.1595, 35.2165)
2021	28.5688	3.70284	(21.3114, 35.8262)
2022	28.9493	3.81605	(21.4700, 36.4286)
2023	29.3300	3.92602	(21.6352, 37.0249)
2024	29.7106	4.03298	(21.8061, 37.6151)
2025	30.0912	4.13717	(21.9825, 38.2000)
2026	30.4719	4.23881	(22.1639, 38.7798)
2027	30.8525	4.33807	(22.3500, 39.3549)
2028	31.2331	4.43510	(22.5405, 39.9258)
2029	31.6137	4.53006	(22.7350, 40.4925)
2030	31.9944	4.62307	(22.9333, 41.0554)

Forecast plot:



Forecast evaluation statistics

Mean Error	-0.0031737	Root Mean Squared Error	1.4188
Mean Absolute Error	0.94123	Mean Percentage Error	-5.1853
Mean Absolute Percentage Error	12.084	Theil's U	0.98962
Bias proportion, UM	5.0034e-006	Regression proportion, UR	0.0015442
Disturbance proportion, UD	0.99845		

ARIMA(0,1,1) (model II)

Model 2: ARIMA, using observations 1951-2006 (T = 56)

Dependent variable: (1-L) Potato million tonnes

Standard errors based on Hessian

	Coefficient	*Std. Error*	*Z*	*p-value*	
Const	0.395463	0.0853605	4.633	<0.0001	***
theta_1	−0.552195	0.0885772	−6.234	<0.0001	***

Mean dependent var	0.364821	S.D. dependent var	1.699092
Mean of innovations	−0.012886	S.D. of innovations	1.389249
Log-likelihood	−98.05317	Akaike criterion	202.1063
Schwarz criterion	208.1824	Hannan-Quinn	204.4620

	Real	*Imaginary*	*Modulus*	*Frequency*
MA				
Root 1	1.8110	0.0000	1.8110	0.0000

Test for ARCH of order 4 -

Null hypothesis: no ARCH effect is present

Test statistic: LM = 11.8868

with p-value = P(Chi-square(4) > 11.8868) = 0.0182132

Test for autocorrelation up to order 4

Ljung-Box Q' = 12.2537,

with p-value = P(Chi-square(3) > 12.2537) = 0.006563

Test for normality of residual -

Null hypothesis: error is normally distributed

Test statistic: Chi-square(2) = 17.8743

with p-value = 0.000131414

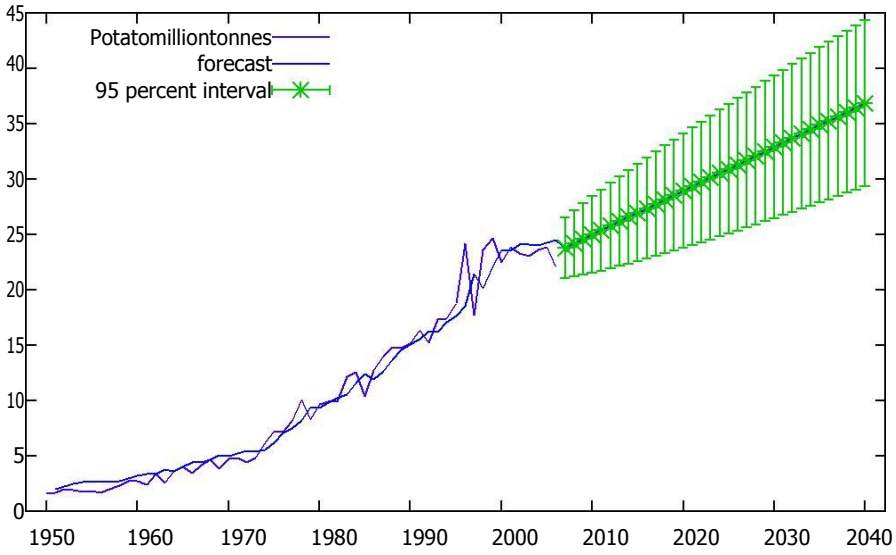
FORECAST:

For 95% confidence intervals, z(0.025) = 1.96

Obs	prediction	std. error	95% interval
2007	23.8137	1.38925	(21.0909, 26.5366)
2008	24.2092	1.52218	(21.2258, 27.1926)
2009	24.6047	1.64440	(21.3817, 27.8276)
2010	25.0001	1.75815	(21.5542, 28.4460)
2011	25.3956	1.86497	(21.7403, 29.0509)
2012	25.7910	1.96599	(21.9378, 29.6443)
2013	26.1865	2.06208	(22.1449, 30.2281)
2014	26.5820	2.15388	(22.3604, 30.8035)
2015	26.9774	2.24192	(22.5833, 31.3715)
2016	27.3729	2.32664	(22.8128, 31.9330)
2017	27.7684	2.40837	(23.0480, 32.4887)
2018	28.1638	2.48742	(23.2886, 33.0391)
2019	28.5593	2.56404	(23.5339, 33.5847)
2020	28.9547	2.63843	(23.7835, 34.1260)
2021	29.3502	2.71078	(24.0372, 34.6632)
2022	29.7457	2.78125	(24.2945, 35.1968)

2023	30.1411	2.84998	(24.5553, 35.7270)
2024	30.5366	2.91709	(24.8192, 36.2540)
2025	30.9321	2.98269	(25.0861, 36.7780)
2026	31.3275	3.04688	(25.3557, 37.2993)
2027	31.7230	3.10974	(25.6280, 37.8180)
2028	32.1184	3.17136	(25.9027, 38.3342)
2029	32.5139	3.23180	(26.1797, 38.8481)
2030	32.9094	3.29113	(26.4589, 39.3599)
2031	33.3048	3.34942	(26.7401, 39.8696)
2032	33.7003	3.40670	(27.0233, 40.3773)
2033	34.0958	3.46304	(27.3083, 40.8832)
2034	34.4912	3.51847	(27.5951, 41.3873)
2035	34.8867	3.57305	(27.8836, 41.8897)
2036	35.2821	3.62680	(28.1737, 42.3906)
2037	35.6776	3.67977	(28.4654, 42.8898)
2038	36.0731	3.73199	(28.7585, 43.3876)
2039	36.4685	3.78349	(29.0530, 43.8840)
2040	36.8640	3.83429	(29.3489, 44.3791)

Forecast plot:



Forecast evaluation statistics

Mean Error	-0.012886	Root Mean Squared Error	1.3895
Mean Absolute Error	0.95412	Mean Percentage Error	-7.2441
Mean Absolute Percentage Error	13.423	Theil's U	1.157
Bias proportion, UM	8.5996e-005	Regression proportion, UR	0.018016
Disturbance proportion, UD	0.9819		

Overall, we can see that ARIMA(1,1,0) provide a good fit for potato production in India. Its gives a fairly accuracy forecasting. However, although forecast from 2017-2040 are within the 95% interval, the graph shows that the green line of actual data has gradually moving out of the confidence interval.

IV. Conclusion

- In this research study, researcher analyzed and obtained the forecast of potato production in India using ARIMA models. The result of the study conclude that ARIMA(1,1,0)model is the most appropriate model for forecasting Potato production in India.
- The forecast model and the forecast graph that the potato production is rapidly increasing with the passage of year.

- Overall, we can see that ARIMA(1,1,0) provide a good fit for potato production in India. Its gives a fairly accuracy forecasting. However, although forecast from 2017-2040 are within the 95% interval, the graph shows that the green line of actual data has gradually moving out of the confidence interval.

References

1. Damodar N Gujarati, Dawn C Porter, Sangeetha Gunasekar "Basic Econometric" fifth edition 2012.
2. David M. Levine , Timothy C. Krehbiel, Mark L. Berenson "Business Statistics A First Course " Third Edition 2007. P. No: 567-612.
3. Naval Bajpai, " Business Statistics " 2010. P. No: 571-615
4. Box, G.E.P. and Jenkins G.M. (1976), "Time Series Analysis,forecasting and Control, Holden-Day, San Francisco.Brent, M., and Mehmet P. (2010), "Simple ways to forecast inflation: What works best?", Trade Publication,17, 1-9.
5. Fannoh, R., Orwa G., and Mung'atu J. K.. (2014), "Modeling the Inflation Rates in Liberia SARIMA Approach", International Journal of Science and Research, 3, 1360-1367.
6. Kibunja H., Kihoro J. and Orwa G.(2014), "Forecasting Precipitation Using SARIMA Model: A Case Study of Mt.
7. Kenya Region", International Institute for Science, Technology and Education, 4(11), 50-58
8. Martinez E. Z., and Soares E.A. (2011), "Predictingthe number of cases of dengue infection in Ribeirão Preto, São Paulo State, Brazil, using a SARIMA model", Revista da Sociedade Brasileira de Medicina Tropical, 44(4), 436-440.
9. OtuA. O., Osuji G. A., Opara J., Ifeyinwa M. H., and Iheagwara A.I. (2014), "Application of SARIMA models in modelling and forecasting Nigeria's inflation rates", American Journal of Applied Mathematics and Statistics, 2, 16-28.
10. Saz G. (2011), "The efficacy of SARIMA models in forecasting inflation rates in developing countries: The case for Turkey", International Research Journal of Finance and Economics, 62, 111-142.
11. Webster, D. (2000), "New Universal Unabridged Dictionary", Barnes and Noble Books.
12. Makridakis, Wheelwright and Hyndman 1998, Forecasting Methods and Applications (3rd Edition), John Wiley & Sons, Inc.
13. Selvanathan, E.A., June 1991, "A Note on the Accuracy of Business Economists'.

