IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 12, Issue 3 Ver. VII (May. - Jun. 2016), PP 01-07
www.iosrjournals.org

Time Series Analysis of Index of Industrial Production of India

Salam Shantikumar Singh[1], T. Loidang Devi[2], Tanusree Deb Roy[3]
[1] *Department of Statistics, Manipur University, India*
[2] *Department of Statistics, Manipur University, India*
[3] *Department of Statistics, Assam University, Silchar, Assam, India*

Abstract:*Industry sector plays a very important role in the economic growth of the India. Composition & production of goods from Indian industrial depends on many factors and due to that factors, there may be short-term (seasonal) and long-term (trend) variations in index if industrial production (IIP) in India. Therefore, the study on the effects of seasonal and trend on IIP in India is the prime objective of the present study. The data set of IIP is from National Data Sharing & Accessibility Policy (NDSAP), Government of India, it comprises of general IIP and 26 sub-sectors of industrial productions. A well-known time series model, ARIMA (p, d, q) model is used to analyze the time series data. The finding suggests that both seasonal and trend effects are present in IIP.A forecast of the future IIP of India is made by this model after adjusting the effects of short-term and long-term variations.*
Keywords:*IIP (Index of Industrial Production), Seasonal, Trend, ARIMA, Forecast*

I. Introduction

1.1 General Introduction

The index of industrial production (IIP) is an index for measuring the growth of various sectors of industry in an economy of a country such as mining, electricity, manufacturing etc[1]. The all India IIP is a composite indicator that measures the short term changes in months in the volume of a basket of industrial products during a given period with respect to that in a chosen base year[2]. The industrial productions are often influenced by seasonal fluctuations, trend fluctuations as well as calendar & trading day effects, which cover relevant short and long term movements of the time series.

The climatic and natural factors may also influence on consumption and production behavior of industries. The time series of industrial products may also be affected by moving holidays and festivals such as Holi, Durga Puja, Diwali, Ramjan, Easter, Christmas etc. and variations in fiscal years of the country which may bring seasonality in the time series.

Industry accounts for 26 percent of GDP and employs 22 percent of the total workforce in India[3]. According to World Bank[4], India's industrial manufacturing GDP output in 2015 was 6th largest in the world on current US Dollar basis ($559 billion) and 9th largest on inflation adjusted constant 2005 US Dollar basis ($197.1 billion). The Indian industrial sector underwent significant changes as a result of the economic liberalization in India Economic Reforms of 1991, which removed import restrictions, brought in foreign competition, lead to the privatization of certain government own public sector industries, liberalized the FDI regime, improved infrastructure and led to an expansion in the production of fast-moving consumer goods[5].

1.2 Review of Literature

The relationship between resource abundance and economic growth is considerable attention from a wide range of researchers over last couple of years. It is confirmed by Sahoo et.al. (2014)[6] that mineral export, industrial production and economic growth are co-integrated, indicating an existence of long run equilibrium relationships among variables. They also suggest that there is a long run Granger causality relationship running from economic growth and industrial production to the mineral export of India.

Mondalet. al., (2014)[7]have conducted a study on the effectiveness of autoregressive integrated moving average (ARIMA) model, on fifty six Indian stocks from different sectors. They have accurately modeled for

---

[1]*Wikipedia*, Index of Industrial Production. retrieved from https://:en.wikipedia.org/wiki/index_of_industrial_production dated 30/04/2016.
[2]CSO,*Study on internal consistency of all India index of industrial production (IIP) data*. Ministry of Statistics & Programme Implementation, GoI.
[3] India-All Economic Indicators, available *in*" https://www.quandl.com/collections/india-economy-data" retrieved on 26th April, 2016.
[4] World Bank (2015),*GDP and its breakdown at current price in US Dollar*,United Nations Statistics Division.
[5]*Economic Reforms in India: Task Force Report*. University of Chicago, 2013, p 9-12.
[6]A. K.Sahoo, Dukhabandhu&N. C. Sahu, Mining export, industrial production and economic growth: a cointegration and casuality analysis of India. *Resources Policy*, 42, 2014, 27-34
[7]P.Mondal, L. Shit, and S. Goswami, Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices*International Journal of Computer Science, Engineering & Applications(IJCSEA)*, 4(2), 2014, 13-29.

forecasting of stock prices in National Stock Exchange of India by using (ARIMA) model. Accuracy of prediction of stock prices for next few months is above 85 percent indicating that ARIMA model gives good accuracy of prediction.

Daniel et.al. (2013)[8], focuses on the regression analysis and time series analysis of the petroleum product sales in masters energy oil and gas. From the regression analysis, it is observed that the petroleum product sales are affected only by environmental factors. However, the effect of environmental factors is not significant after removing the seasonal and trend variations in the data by using time series analysis. It shows that time series analysis explain the influence of seasonal and trend on the petroleum product sales.

In the study of Jha K., et.al. (2013)[9], autoregressive integrated moving average (ARIMA) model has been used to model the growth pattern traffic intensity and forecasting because it can minimize the error resulting from the estimation with varying parameters. From their finding it is suggested that the traffic volume forecasting using ARIMA model is better accuracy than any other traditional approaches such as regression modeling. The efficiency of the time series method improves considerably as the traffic data in analysis is moved closer to the forecasting window i.e. efficacy of this method improves with short term forecasting.

Wojewwodzki M (2010)[10] examines the short and long run causality between the growth rate of household's savings in the consolidated banking system in Poland and Polish economic growth rate using industrial production index as proxy variable. The different methods of economic time series analysis such as cointegration, ganger causality were used and followed by VAR, ARMA and GARCH econometric procedures in order to find out the causalities and the most accurate forecast for saving growth for the next 5 quarters. They suggested that there is one way ganger causality from the output growth to saving growth. The changes in the savings are influenced far more by the changes in the output than vice versa. Therefore the policymakers should be focused on boosting the economic growth than on savings by promoting the factors of production.

1.3Rationale of the study

Industry sector plays a very important role in economic growth of the India. Consumption and production of goods from Indian industrial dependence on many factors and due to that factors there is seasonal (short term) and trend (long term) variations on industrial products. Thus, it is high time to determine the effects of seasonal and trend on index of industrial production in India. Furthermore, forecasting of the growth of industrial production is equally important and this will be a key tool for economic policymaking, business cycle of the country.

1.4 Objectives of the study

The main objectives of the study are to

- Compare the current monthly/Annually Index of Industrial Production (IIP) with the previous month/year.
- To identify the real movements and turning points in industrial productions
- To provide more reliable short-term and long term forecasts of industrial productions.

II. Methodology

1.2 Data

Data is from National data Sharing and Accessibility Policy (NDSAP), Government of India[11]. This data of IIP is growth rates of different components of industrial productions with the base year of 2004-05. The data set comprises of 26 sub-sectors of industrial productions. In the present study, only 'general' (all subsectors of industry)since April, 2005 up to February, 2016, has been used for analyzing and forecasting the IIP of India.

2.2 Time series analysis:

A time series is a set of statistics, usually collected at regular intervals of time. Time series data occur naturally in many application areas such as monthly data for index of industrial productions, stock prices, unemployment, rainfall and weather condition etc. suppose

$$Y_t = IIP at time t$$

The time series data should be studied differently due to the dependency and correlation between the data. It is known that the majority of statistical methods are based on the assumption of independence, which does not hold for most of the time series data.

[8] E.C Daniel, et.al.,Regression and time series analysis of petroleum product sales in master energy oil and gas,*International Journal of Sciences, Engineering and Technology Research (IJSETR)*, 2(6),(2013), 1285-1303.

[9]K.Jha, et.al, Time series analysis of traffic data.*Modern Traffic and Transportation Engineering Research (MTTER*), 2(1), 2013, 60-67.

[10]M.Wojewodzki, The household's saving and the industrial production: a time series analysis for Poland. *Lingnan Journal of Banking, Finance & Economics*, 2, 2010, 88-108.

[11]National data Sharing and Accessibility Policy (NDSAP), Government of India, retrieved from https://data.gov.in/catalog/time_series_indices_industrial_production.

2.2.1 Components of time series
A time series data is comprised of the following components in general

- Trend (T_t)- long term movements at the level of the series
- Seasonal effects (S_t) $-$ cyclical fluctuations related to the calendar (including moving holidays and working day effects)
- Cycles (C_t) $-$cyclical fluctuations longer than a year (e.g. business cycles)
- Irregular (I_t) $-$other random or short term unpredictable fluctuations.

Therefore the general model of time series can be expressed as

$$X_t = S_t + T_t + C_t + I_t$$

The idea of seasonal adjustment it to create separate models for those components and then combine them additively.
Thus the seasonal adjusted series is derived as

$SA_t = X_t - S_t or SA_t = T_t + I_t$
Or multiplicatively $X_t = S_t - T_t - I_t$
where$SA_t = \frac{X_t}{S_t} = T_t \times I_t$

2.3 Model selection
The classical regression is often insufficient for explaining all of the interesting dynamics of a time series. The introduction of correlation as a phenomenon that may be generated through lagged linear relations leads to proposing the autoregressive (AR) and autoregressive moving average (ARMA) models. Adding the non-stationary models to the mix leads to the autoregressive integrated moving average (ARIMA) models popularized in the landmark work by Box and Jenkins (1976)[12].
Autoregressive models are based on the idea that the current value of the time series x_t can be explained as a function of p past values$x_{t-1}, x_{t-2}, \dots \dots \dots, x_{t-p}$, where p determines the number of steps into the past needed to forecast the current value.
Definition1: An autoregressive model of order p, abbreviated AR(p) is of the form

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots \cdots + \varphi_p x_{t-p} + w_t \quad (1)$$

wherex_t is stationary and $\varphi_1, \varphi_2, \cdots \cdots, \varphi_p$ are non-zero constants; w_t is Gaussian white noise series with mean 0 and variance σ_w^2.
If mean function of x_t of μ, replace x_t by $x_t - \mu$ in (1)

$$x_t - \mu = \varphi_1(x_{t-1} - \mu) + \varphi_2(x_{t-2} - \mu) + \cdots \cdots + \varphi_p(x_{t-p} - \mu) + w_t \quad (2)$$

$$or x_t = \alpha + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots \cdots + \varphi_p x_{t-p} + w_t$$

where$\alpha = (1 - \varphi_1 - \varphi_1 - \cdots \cdots - \varphi_p)$
By using backshift operator, the AR(p) model in (1) can be expressed as

$$\left(1 - \varphi_1 B - \varphi_1 B^2 - \cdots \cdots - \varphi_p B^p\right) = w_t \quad (3)$$

$$or \varphi(B) x_t = w_t \quad (4)$$

$where \varphi(B) = \left(1 - \varphi_1 B - \varphi_1 B^2 - \cdots \cdots - \varphi_p B^p\right)$is called autoregressive operator.

An alternative to the autoregressive representation in which the x_t on the left hand side of the equation are assumed to be combined linearly, the moving average model of order q, abbreviated as MA(q), assumes the white noise w_t on the right hand side of the defining equation are combined linearly to form the observed data.
Definition2: The moving average model of order q, denoted by MA(q) model is defined to be

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots \cdots + \theta_p w_{t-q} \quad (5)$$

Where there are q lags in the moving average and $\theta_1, \theta_2, \cdots, \theta_q$ ($\theta_q \neq 0$) are parameters.
We may also write (5) as

$$x_t = \theta(B) w_t \quad (6)$$

Where $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots \cdots + \theta_p B^q$ (7)
is moving average operator.
Definition 3: A time series $\{x_t, t = \pm 1, \pm 2, \cdots \cdots\}$ is ARMA (p, q) model if it is stationary and

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots \cdots + \varphi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots \cdots + \theta_p w_{t-q} \quad (8)$$

Or $x_t = \alpha + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots \cdots + \varphi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots \cdots + \theta_p w_{t-q}$ (9)
In many situations, time series can be thought of as being composed of two components, a non-stationary trend component and a zero-mean stationary component. For example, model of the form

$$x_t = \mu_t + y_t$$

[12] G.E.P Box, and G.M. Jenkins, *Time Series Analysis, Forecasting and Control* (Holden-Day, San Francisco, USA1976).

Where $\mu_t = \beta_0 + \beta_1 t$ and y_t is stationary. Differencing such a process will lead to a stationary process

$$\nabla x_t = x_t - x_{t-1} = \beta_1 + y_t - y_{t-1} = \beta_1 + \nabla y_t \tag{10}$$

Definition 4: A process x_t is said to be ARIMA (p,d,q) if

$$\nabla^d x_t = (1-B)^d x_t \tag{11}$$

Is ARMA(p, q). In general, we will write the model as

$$\varphi(B)(1-B)^d x_t = \theta(B) w_t \tag{12}$$

III. Results

The index of industrial production in India has been categorized into 26 sub-groups. But general industrial production (all industrial subsectors) is used for analysis of time series from April, 2005 up to Feb, 2016. Due to the presence of seasonal, trend and irregular variations in the time series data of index of industrial production as well as non-stationary ARIMA (p,d,q) model has been applied for estimating the parameters and forecasting. After identifying the suitable orders of autoregressive, differencing (transforming non-stationary into stationary) and moving average, the best ARIMA model is ARIMA (1, 1, 4). The following table 1 and table 2 show the goodness of fit ARIMA model.

Table 1: Model Fit

Fit Statistic	Mean	Minimum	Maximum
Stationary R-squared	0.439	0.439	0.439
R-squared	0.911	0.911	0.911
RMSE	7.187	7.187	7.187
MAPE	3.636	3.636	3.636
MaxAPE	9.496	9.496	9.496
MAE	5.760	5.760	5.760
MaxAE	16.040	16.040	16.040
Normalized BIC	4.169	4.169	4.169

Table 2: Model Statistics

Model	Number of Predictors	Model Fit statistics							Ljung-Box Q(18)		
		Stationary R^2	R^2	RMSE	MAE	MaxAPE	MaxAE	Normalized BIC	Statistics	DF	Sig.
General-Model	0	0.439	0.911	7.187	5.760	9.496	16.040	4.169	61.942	13	<0.001

Stationary R-squared compares the stationary part of the model to a simple mean model. The stationary R-squared is found to be 0.439 (positive) which indicates that the model under consideration is better than the baseline model (simple mean model). The R^2 is an estimate of the proportion of the total variation in the series explained by the model. Therefore the applied ARIMA (1,1,4) can explain 91.1 of the data by the model. RMSE (root mean square error) measure of how much a dependent series varies from its model-predicted level, expressed in the same units as the dependent series. Thus the variation of the original IPP time series data from model-predicted level is 7.187 and it is considerably good. On the other hand, MAPE (Mean absolute percentage error) measure how much a dependent series varies from its predicted- model level expressed in the independent units therefore be used to compare series with different units. In the present study, the time series data is of the same units of measurement and hence therefore the estimated MAPE given in the table 1 has no any role for testing goodness of fit model. The estimated MAE of the model is 5.760 indicating that the series varies from the model-predicted level reported in the original series units. The largest forecasted error is 9.496 percentage which is measured by MaxAPE (maximum absolute percentage error). Therefore the worst case scenario of forecasting the IIP by using ARIMA model is less than 10 percent. Similarly, MAxAE measures the worst case scenario of forecasting the time series.

The normalized BIC (Bayesian Information Criterion) measures the overall fit of a model that attempts to account for model complexity. It includes a penalty for the number of parameters in the model and the length of the series. Finally the Ljung-Box test which is found to be highly significant statistically. Hence the fitted model is suitable to the time series data.

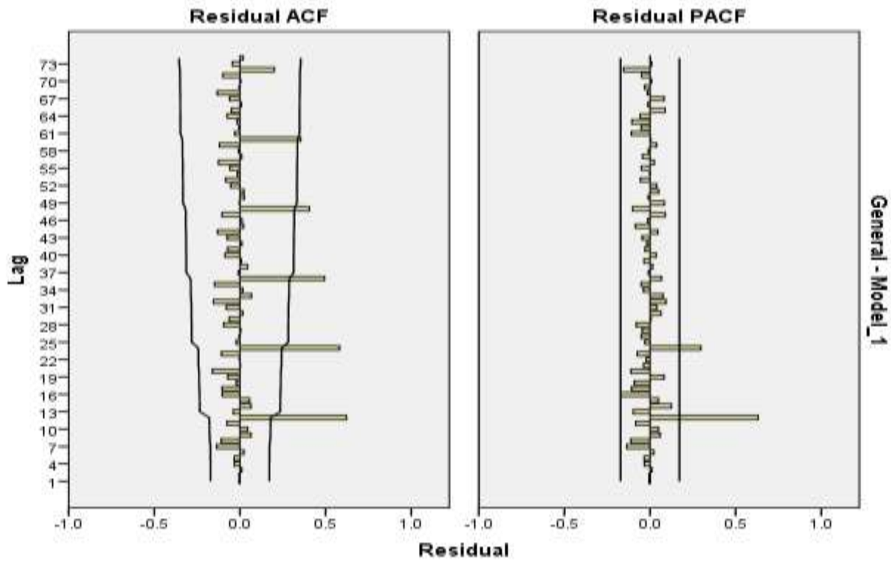

Fig.1: Residual ACF and PACF

Autocorrelation means the correlation between time series and the same time series lag measured by autocorrelation function (ACF). Partial autocorrelations are also correlation coefficient between the basic time series and the time series lag after eliminating the influence of the members between and it is measured by partial autocorrelation function (PACF). From both ACF & PACF plots in figure 1, it is revealed that ACF is exponentially declining and spikes in the first lag or more lags of the PACF, therefore it presents the autoregressive process. Further, ACF spikes at first few lags and PACF is exponentially declining which indicates the presence of moving average processes. Further, it is observed that the ACF is cutting at lag 1and PACF is tailing off. This suggests that IIP time series data follows AR (1). Again PACF is cutting off at lag 4 and AFC is tailing off. Therefore, the data follows MA(4) model.

Table 3: ARIMA Model Parameters

					Estimate	SE	t	Sig.
General-Model	General	No Transformation	Constant		0.628	0.166	3.776	0.000
			AR	Lag 1	-0.136	0.254	-0.534	0.594
			Difference		1			
			MA	Lag 1	0.586	0.236	2.486	0.014
				Lag 2	-0.253	0.203	-1.248	0.214
				Lag 3	-0.010	0.147	-0.071	0.944
				Lag 4	0.391	0.106	3.682	0.000

The estimated coefficient of AR and MA in the ARIMA (1, 1, 4) model are shown in the table 3.
Using maximum likelihood estimation method, the model for IIP x_t , the estimated model is

$$x_t = 0.628_{0.001} - 0.185_{0.277} x_{t-1} + 0.586_{0.236} \widehat{w}_{t-1} - 0.253_{0.203} \widehat{w}_{t-2} - 0.010_{0.147} \widehat{w}_{t-3} + 0.391_{0.106} \widehat{w}_{t-4} + \widehat{w}_t$$

It has been revealed from the above model that the negative correlation between the current month IIP and the previous month indicating that if there was less production of industrial goods in the previous month then the production will be increased in the consecutive and it will be negatively impactupto the previous three months but positively impact with the previous 4th months.

Table 4: Estimated Distribution Parameters

		Noise residual from General-Model
Normal Distribution	Location	0.01
	Scale	0.296

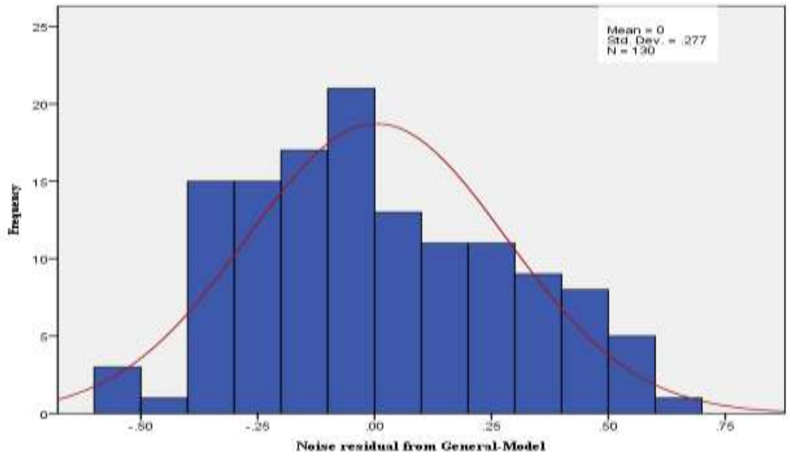

Fig. 2: Histogram of Noise residual

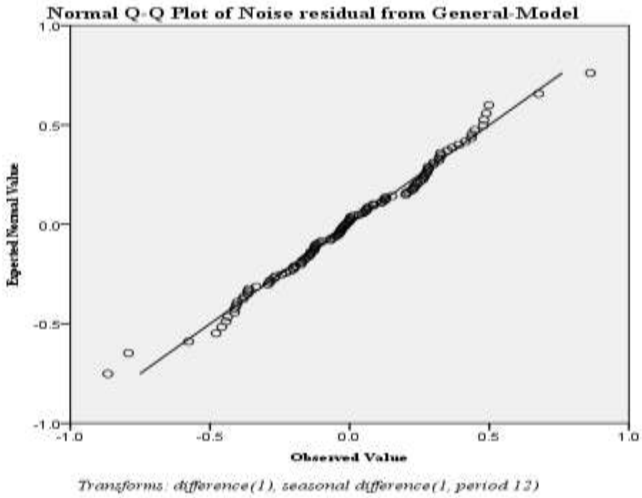

Fig. 3: Q-Q of Noise residual

From the above table 4, it is known that the location parameter (mean) of the residual (error component) is approximately zero with constant variance (scale parameter). And the normality test of the error (residual) conducted by drawing histogram of the residuals and the Q-Q plot. Normal curve shown in figure 2 suggests that the error distribution is approximately normal. Moreover, from the Q-Q plot in figure 3, it is confirmed that the error distribution in the IIP time series follows normal since almost all points in the plot is closure to the straight line except a few points. Therefore the assumption of the error component in the ARIMA (1, 1, 4) is fulfilled and thus this model can forecast the future trend of the IIP.

Table 5: Forecasting

Model		Mar 2016	Apr 2016	May 2016	Jun 2016	Jul 2016	Aug 2016	Sep 2016	Oct 2016	Nov 2016	Dec 2016	Jan 2017
General-Model	Forecast	192	191	188	188	189	190	190	191	191	192	193
	UCL	206	205	205	207	208	209	210	211	212	213	214
	LCL	178	176	171	169	170	170	170	170	171	171	171

For each model, forecasts start after the last non-missing 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, whichever is earlier.

The table 5 highlights the prediction of the IIP for future months of the times series data starting from March, 2017 up to January, 2017 by ARIMA (1,1,4) model. It is observed that the industrial production in India (General sub-sector) would likely to be declined the rainy season (i.e. from May to July) but it will increase considerably from August and onward. It may be due to the lack of raw materials and inconvenience to explore the natural resources during the rainy season which are very essential for increase in productivity of an industry.The following figure 2 depicts the observed time series trend and fitted time series model. And the forecasting trend of IIP of India is also shown by portioning a straight vertical line.

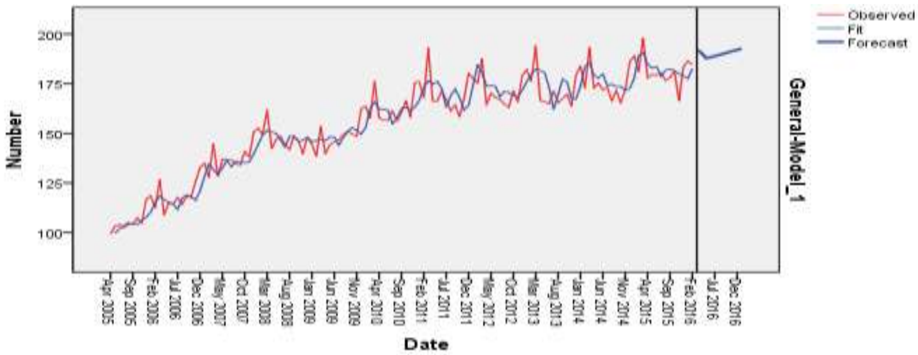

Fig. 4: Fitted Time Series Model

IV. Scope Of The Future Study

In the present study, only one sub-sector out of 26 sub-sectors of the IIP is used to analyze for determining the effects of short term, long term and random components by applying univariate ARIMA model and forecast has been made only for general category of industrial production. Thus, to see the whole nature of industrial production in India, all sub-sectors of the industrial production can be studied by using a suitable multivariate time series model and which will show the real picture of growth of India's industrial production. Moreover, productivity of different sub-sectors (26) may vary from time to time, season to season and years to years. Therefore, characteristics of time series for different sub-sectors can also be studied separately.

V. Recommendation

Industrial production of a country plays an important role in growth and development of a country in the competitive world of today. India is a developing country and its economy is totally depending on agriculture but agriculture production is declining day by day due increasing in population which indirectly effect in reducing the cultivable land. Moreover, a country can be prosperous without industrialization because it will fill the gap of agriculture in the nation. Therefore a system and considerable growth of industrial production is very much needed in the country. To maintain a sustainable growth of industrial production in India, economist, government agency and policymakers should be familiar with the trend and variation of index of industrial production. Thus, a time series analysis of IIP would like to guide for all stakeholders for future growth industry in India.

Reference

[1]. *Wikipedia*, Index of Industrial Production,Retrieved from https://:en.wikipedia.org/wiki/index_of_industrial_production dated 30/04/2016.
[2]. CSO, *Study on internal consistency of all India index of industrial production (IIP) data*. Ministry of Statistics & Programme Implementation, GoI.
[3]. India-All Economic Indicators, available *in*" https://www.quandl.com/collections/india-economy-data" retrieved on 26th April, 2016.
[4]. World Bank (2015), *GDP and its breakdown at current price in US Dollar*, United Nations Statistics Division.
[5]. *Economic Reforms in India: Task Force Report*. University of Chicago, 2013, p 9-12.
[6]. K. Sahoo, Dukhabandhu& N. C. Sahu, Mining export, industrial production and economic growth: a cointegration and casuality analysis of India. *Resources Policy*, 42, 2014, 27-34
[7]. P. Mondal, L. Shit, and S. Goswami, Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices *International Journal of Computer Science, Engineering & Applications(IJCSEA)*, 4(2), 2014, 13-29.
[8]. E.C Daniel, et.al., Regression and time series analysis of petroleum product sales in master energy oil and gas, *International Journal of Sciences, Engineering and Technology Research (IJSETR)*, 2(6), (2013), 1285-1303.
[9]. K. Jha, et.al, Time series analysis of traffic data. *Modern Traffic and Transportation Engineering Research (MTTER*), 2(1), 2013, 60-67.
[10]. M. Wojewodzki, The household's saving and the industrial production: a time series analysis for Poland. *Lingnan Journal of Banking, Finance & Economics*, 2, 2010, 88-108.
[11]. National data Sharing and Accessibility Policy (NDSAP), Government of India, retrieved from https://data.gov.in/catalog/time_series_indices_industrial_production.
[12]. G.E.P Box, and G.M. Jenkins, *Time Series Analysis, Forecasting and Control* (Holden-Day, San Francisco, USA1976).