A STUDY OF FORECAST OF RECOVERED CASES OF COVID-19 VIRUS IN INDIA BY ARIMA METHOD

U.V. Kale[1], M.O. Wankhade[2] and H.S. Lunge[3]
[1,3]Dept. of Statistics, Shri Shivaji Science College Amravati, MS, India
[2]Dept. of Statistics, G. S. Science, Arts & Commerce College Khamgaon, MS, India
[1]uvkale1964@gmail.com, mowankhade289@gmail.com

ABSTRACT

The aim of this paper is to study and analyze recovered cases of Covid19 in India. The data related to recovered cases of covid19 disease is obtained from government website https://www.covid19india.org from 29th March to 12th August 2020. The data is analyzed in Minitab19 statistical software using ARIMA (Autoregressive integrated Moving Average) method. The forecasts of recovered case along with 95% confidence limits for next three months i.e. from 12th August 2020 are generated under ARIMA model. The trend of forecasts so obtained is further analyzed. Also Randomness of error terms in the model is tested by Lunge Box chi-square statistic. The significance of Autocorrelation and Partial Autocorrelation is tested with the help of Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). We observed that the trend of forecasts of recovered cases of covid19 in India is linear in nature. The recovery rate per 100 patients will range from 83.51to 85.20 during 12th August to 10 November 2020.

Keywords: *ARIMA, Covid-19, Recovered Case, ACF, PACF, Modified Box-Pierce, India.*

1. Introduction

The World Health Organization (WHO) proclaimed corona virus pandemic in the month of March 2020. Government of India declared nationwide lockdown from 25th March 2020 to prevent the spread of the corona virus infection in the country. The number positive case of covid19 patients were very few till second week of April 2020. But later on the number of patients covid-19 increased at a significant rate. Many researchers of the world are working on prediction of future cases of patients, recovery rate, death rate etc. due to covid19 pandemic disease. M.H.D.M. Ribeiro et. al. (2020) developed the various model viz. ARIMA, CUBIST, RF, RIDGE, SVR, GP and stacking-resembling learning approaches, forecasts of covid-19 cases for a short period in 10 Brazilian states. A concept of Gaussian process was put forward by L. Li et.al. (2020) to analyze the transmission and simulation of covid-19 cases Hubei Province and F. J. D. Perez et.al. (2020) assumed Gompertz Growth curve model for prediction covid-19 patients in South Korea and Hubei-China, Italy and Spain. SIR (Susceptible Infected and Recovered) concept for prediction of covid-19 cases in open source codes from Milan Batista was developed by J. Luo (2020). He adopted a technique of Data Driven Prediction model using SIR. Lead-lag Relation concept of two time series was suggested by J. Stubinger et.al. (2020) for forecasting of covid19 infected cases in 10 most prominent countries. A rigorous study of various models for transmission of corona virus disease was made by S. Khajanchi et.al. (2020) using cumulative number of covid-19 confirmed cases in India. They developed a compartmental mathematical model for dynamics of transmission of the disease by applying six compartmental concepts viz. Uninfected, asymptomatic, symptomatic infected, unreported infected, quarantined and recovered (R) patients termed as SAIUQR, model in Indian continental and advanced first order nonlinear differential equation for forecasting of covid-19 cases in the different states of India perticularly Jharkhand, Gujarat, Andhra Pradesh, and Chandigarh. A Genetic Algorithm based model for generating future forecasts of confirmed cases and deceased cases of covid-19 the states of Maharashtra, Gujarat and Delhi and India as a whole was formulated and adopted by R. Salgotra et.al. (2020). S. He, et.al. (2020) composed a SEIR stochastic model for forecasting of covid-19 cases by grouping the population into sub-populations as susceptible (S), Exposed (E), Infected (I) and Recovered (R). A. Bhattacharjee et. al. (2020) propounded the method of prediction of covid-19 cases by combination of recovery and case load rate on analyzing cumulative confirmed cases, recovered cases and number of deaths. A Logistic Model for prediction of confirmed case of covid19 in India was suggested by A. Das (2020) by applying the concept of minimization of root mean square error and also smoothing the data by moving averages. A light was thrown on the epidemiological pattern with respect to prevalence of incidence of covid-19 disease by S. Roy et.al. (2020). They presented ARIMA model and analyzed spatial distribution of covid-19 cases in India on GIS environment platform.

2. Methodology

The data relating to daily recovered cases of Covid-19 is taken from government of India website covid19.org from 24th March to 12th August 2020 (Table-1). The missing values are estimated by taking an arithmetic mean of preceding and succeeding observations. The time series data is analyzed in MINITAB19 (2019) statistical Package by applying an Autoregressive Integrated Moving Average (ARIMA) Method of time series. Also the trend and forecasts related to recovered cases of covid19 for next 90 days i.e. from 13th August 2020 to 10th November 2020 is analyzed. An appropriate trend model is fitted for the forecasts of recovered cases of covid19 disease. Random nature of error terms in the are tested with Modified Box-Pierce (Lunge Box Chi-Square) statistic and significance of autocorrelation and partial autocorrelation is explained by Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) respectively.

The mathematical form of ARIMA (p, d, q) model is as (1);

$$Z_t = \emptyset_0 + \epsilon_t + \sum_{i=1}^{p} \varphi_i Z_{t-i} + \sum_{j=1}^{q} \emptyset_j \epsilon_{t-j} \quad ....(1)$$

Where Z_t and are the observed value of response variable.

ϵ_t denotes error term at time period t.

$$\epsilon_t \sim \text{i.i.d.}\, N(0, \sigma^2) \quad ....(2)$$

In ARIMA (p, d, q) model the constants p, d, and q are always integers greater than or equal to zero. A constant p is the order of autoregressive, d the order of integrated differencing and q the order of the moving average respectively.

φ_i (i=1, 2, , , p) and $\emptyset_j$ (j=0,1, 2, , , q) are the parameters of ARIMA model to be estimated.

The ARIMA(p, d, q) model with lags is shown in (3);

$$\varphi(L)(1-L)^d Z_t = \emptyset(L)\, \epsilon_t \quad ....(3)$$

ACF and PACF are defined as (5) and (6).

Coefficient of Autocorrelation at the k^{th} lag is;

$$\rho_k = \frac{r_k}{r_0} \quad ....(4)$$

Where $r_k = \text{Covariance}(Z_t, Z_{t+k})$ ....(5)

Covariance term at the lag zero is given as (6);

$$r_0 = \text{Variance}(Z_t)$$

2.1 Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

Ljung-Box Chi-Square Statistic Q is to test randomness of error terms is;

$$Q = n(n+2)\sum_{k=1}^{m} \frac{\hat{r}^2{}_k}{n-k} \quad ....(7)$$

Where $\hat{r}_k$ is the sample autocorrelation at the k^{th} lag, and m is the number of lags to be tested for independence of residual terms in the model.

Table-1: Daily Recovered Cases of Covd-19 in India from 24th march to 12th August 2020

SN	Date	Rcvd	SN	Date	Rcvd	SN	Date	Rcvd	SN	Date	Rcvd
1.	24-Mar	2	39.	01-May	962	77.	08-Jun	5171	115.	16-Jul	22867
2.	25-Mar	0	40.	02-May	831	78.	09-Jun	5723	116.	17-Jul	17486
3.	26-Mar	0	41.	03-May	911	79.	10-Jun	6275	117.	18-Jul	23532
4.	27-Mar	0	42.	04-May	1082	80.	11-Jun	6044	118.	19-Jul	23918
5.	28-Mar	0	43.	05-May	1295	81.	12-Jun	7263	119.	20-Jul	24303
6.	29-Mar	91	44.	06-May	1385	82.	13-Jun	8091	120.	21-Jul	27589
7.	30-Mar	42	45.	07-May	1475	83.	14-Jun	7358	121.	22-Jul	32875
8.	31-Mar	19	46.	08-May	1111	84.	15-Jun	7292	122.	23-Jul	33326
9.	01-Apr	21	47.	09-May	1414	85.	16-Jun	7226	123.	24-Jul	32514
10.	02-Apr	22	48.	10-May	1669	86.	17-Jun	6890	124.	25-Jul	32013

11.	03-Apr	39	49.	11-May	1787	87.	18-Jun	10741	125.	26-Jul	31512
12.	04-Apr	56	50.	12-May	1905	88.	19-Jun	9029	126.	27-Jul	34354
13.	05-Apr	43	51.	13-May	1963	89.	20-Jun	9052	127.	28-Jul	35683
14.	06-Apr	65	52.	14-May	1594	90.	21-Jun	9075	128.	29-Jul	32886
15.	07-Apr	81	53.	15-May	2234	91.	22-Jun	10879	129.	30-Jul	34720
16.	08-Apr	96	54.	16-May	4012	92.	23-Jun	10462	130.	31-Jul	36554
17.	09-Apr	70	55.	17-May	3247	93.	24-Jun	13089	131.	01-Aug	51368
18.	10-Apr	151	56.	18-May	2482	94.	25-Jun	13983	132.	02-Aug	40355
19.	11-Apr	186	57.	19-May	3032	95.	26-Jun	14106	133.	03-Aug	43070
20.	12-Apr	114	58.	20-May	3113	96.	27-Jun	14229	134.	04-Aug	51220
21.	13-Apr	141	59.	21-May	3131	97.	28-Jun	11631	135.	05-Aug	50681
22.	14-Apr	167	60.	22-May	3280	98.	29-Jun	13497	136.	06-Aug	50141
23.	15-Apr	144	61.	23-May	3283	99.	30-Jun	12565	137.	07-Aug	50387
24.	16-Apr	258	62.	24-May	3285	100.	01-Jul	12064	138.	08-Aug	52135
25.	17-Apr	273	63.	25-May	3012	101.	02-Jul	13241	139.	09-Aug	54474
26.	18-Apr	426	64.	26-May	3585	102.	03-Jul	14417	140.	10-Aug	47362
27.	19-Apr	422	65.	27-May	3434	103.	04-Jul	14746	141.	11-Aug	52561
28.	20-Apr	419	66.	28-May	3171	104.	05-Jul	15829	142.	12-Aug	57759
29.	21-Apr	703	67.	29-May	3737	105.	06-Jul	2250			
30.	22-Apr	394	68.	30-May	4303	106.	07-Jul	16836			
31.	23-Apr	642	69.	31-May	4928	107.	08-Jul	18122			
32.	24-Apr	542	70.	01-Jun	3882	108.	09-Jul	19408			
33.	25-Apr	442	71.	02-Jun	4531	109.	10-Jul	20289			
34.	26-Apr	585	72.	03-Jun	4461	110.	11-Jul	19981			
35.	27-Apr	580	73.	04-Jun	4390	111.	12-Jul	18198			
36.	28-Apr	636	74.	05-Jun	4770	112.	13-Jul	19587			
37.	29-Apr	690	75.	06-Jun	5433	113.	14-Jul	20976			
38.	30-Apr	826	76.	07-Jun	5191	114.	15-Jul	20646			

Source: https://www.covid19india.org

Rcvd: Recovered cases

2.2 ARIMA(1, 1, 1) Model: For Recovered cases of covid-19

Assuming the parameters of the model with p =1, d=1 and q=1 an ARIMA model for analysis of recovered cases is represented as ARIMA(1, 1, 1). The results are presented as Table-2 below;

Table-2: Estimates parameters at Each Iteration

Iteration	SSE	Parameters				
0	1174719593	0.100	0.100	0.100	0.100	408.979
1	965667991	-0.050	0.017	0.250	0.183	493.008
2	939709299	0.020	-0.108	0.349	0.033	522.031
3	914549181	0.091	-0.223	0.456	-0.117	544.810
4	884695208	0.120	-0.330	0.549	-0.267	584.084
5	861442504	0.088	-0.370	0.662	-0.390	619.740
6	860046680	0.151	-0.301	0.740	-0.388	555.073
7	859806313	0.201	-0.301	0.789	-0.421	530.736
8	859776918	0.221	-0.299	0.809	-0.432	520.256
9	859774603	0.225	-0.300	0.812	-0.435	518.908
10	859774428	0.226	-0.300	0.813	-0.436	518.410
11	859774415	0.226	-0.300	0.814	-0.436	518.300
12	859774414	0.226	-0.300	0.814	-0.437	518.264

Table-3: Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.226	0.403	0.56	0.576
AR 2	-0.300	0.181	-1.66	0.100
MA 1	0.814	0.401	2.03	0.045
MA 2	-0.437	0.219	-1.99	0.049
Constant	518	165	3.13	0.002

3. Results and Discussion

As the p-values of AR1 and AR2 parameters of the model (Table-3) are greater than 0.05 except for MA1, MA2 and constant term thus it indicates that the parameters AR1and AR2 are insignificant while MA1, MA2 and constant term are significant at 5% level of significance in the model.
ARIMA model for Recovered cases of covid-19 is obtained as under

$Z_t(\text{recovered}) = 518 + 0.226 * X_{t-1} - 0.3 * X_{t-2} + 0.814 *\epsilon_{t-1} - 0.437 *\epsilon_{t-2}$ ....(9)

Table-4:Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

Lag	12	24	36	48
Chi-Square	24.27	44.01	52.88	54.62
DF	7	19	31	43
P-Value	0.001	0.001	0.008	0.110

As the p-values of Ljung-Box Chi-Square Statistic (Table-4) for the 12th , 24th and 36th lags are less than 0.05 indicate that an error terms are not random in the model at these lags at 5% level. It means that error terms are having certain pattern for the recovered case of covid-19 in India at 12th, 24th and 36th lags. As the p-value of Ljung-Box Chi-Square Statistic at 48th lag is greater than 0.05 indicates that the error terms are random at 48th lag at 5% level of significance.

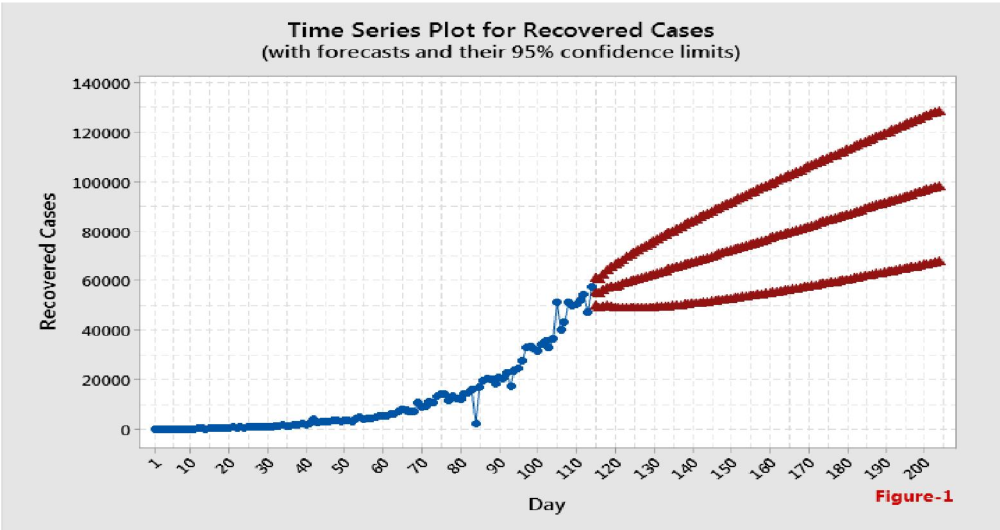
Forecasts of Recovered Cases Under ARIMA(1, 1, 1) Model:

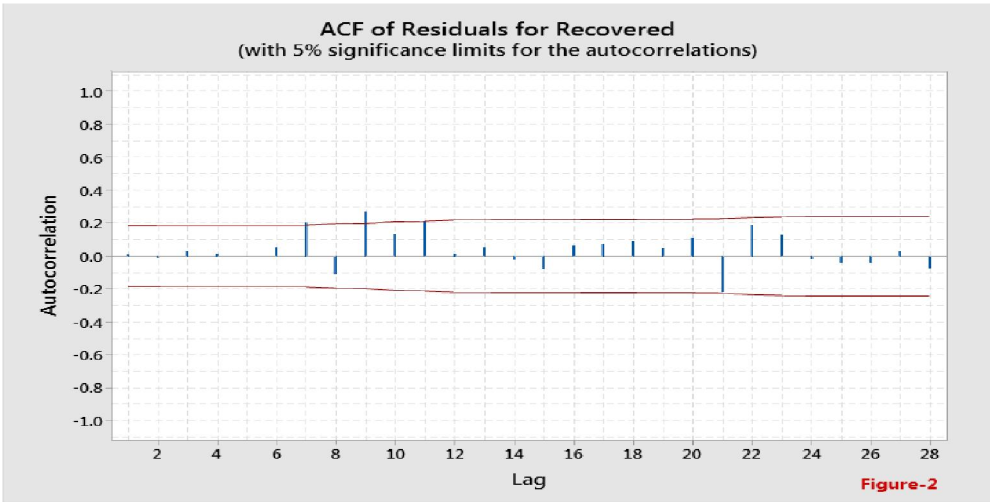
The prediction of daily recovered cases for next 90 days from 13th August to 10th November 2020 along with 95% confidence limits are presented in table-5 below. the plot of time series data along with forecasts and confidence limits is shown in figure-1. From figure-1 it is observed that predicted recovered cases of covid-19 has an increasing tendency.

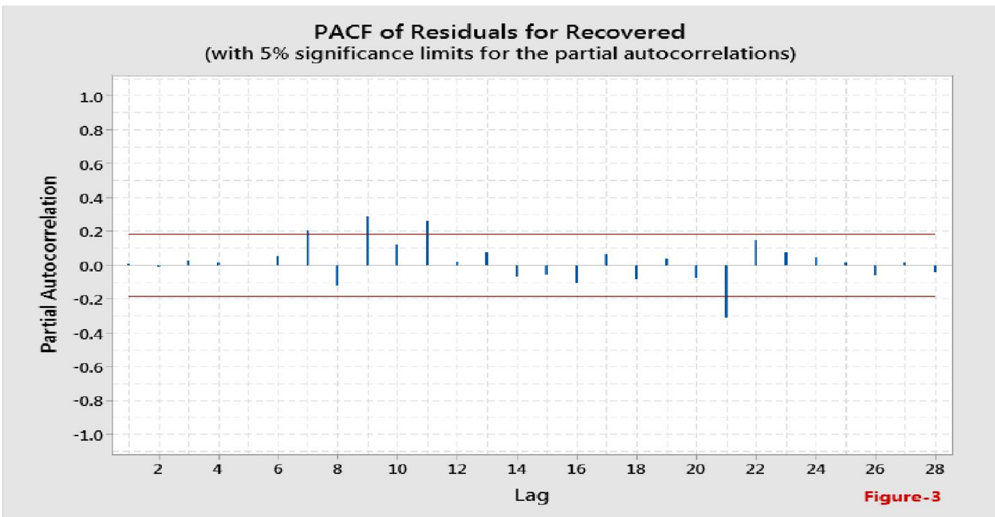
Table-5: Forecasts from period 12th August to 10th November 2020

Daily Recovered		95% Limits		Daily Recovered		95% Limits		Daily Recovered		95% Limits	
Day	Forecast	Lower	Upper	Day	Forecast	Lower	Upper	Day	Forecast	Lower	Upper
13-Aug	55034	49503.7	60566	12-Sep	69385	51187.0	87582	12-Oct	83858	58566.6	109149
14-Aug	54421	48439.2	60405	13-Sep	69867	51389.1	88345	13-Oct	84340	58846.6	109834
15-Aug	55619	49210.6	62029	14-Sep	70350	51595.3	89104	14-Oct	84823	59128.1	110517
16-Aug	56593	49393.3	63793	15-Sep	70832	51805.6	89858	15-Oct	85305	59411.2	111199
17-Aug	56972	48968.0	64975	16-Sep	71315	52019.7	90609	16-Oct	85788	59695.9	111879
18-Aug	57283	48644.8	65921	17-Sep	71797	52237.5	91356	17-Oct	86270	59982.0	112558
19-Aug	57758	48570.4	66946	18-Sep	72279	52458.8	92100	18-Oct	86752	60269.5	113235
20-Aug	58290	48568.9	68011	19-Sep	72762	52683.5	92840	19-Oct	87235	60558.5	113911
21-Aug	58786	48544.5	69028	20-Sep	73244	52911.5	93577	20-Oct	87717	60848.9	114585
22-Aug	59257	48520.9	69992	21-Sep	73727	53142.6	94311	21-Oct	88200	61140.7	115259
23-Aug	59732	48528.5	70936	22-Sep	74209	53376.7	95041	22-Oct	88682	61433.8	115930
24-Aug	60217	48564.3	71869	23-Sep	74692	53613.8	95769	23-Oct	89165	61728.2	116601
25-Aug	60702	48616.1	72787	24-Sep	75174	53853.7	96494	24-Oct	89647	62023.9	117270
26-Aug	61184	48680.0	73688	25-Sep	75656	54096.4	97216	25-Oct	90129	62320.8	117938
27-Aug	61666	48757.0	74574	26-Sep	76139	54341.6	97936	26-Oct	90612	62618.9	118605
28-Aug	62148	48847.0	75449	27-Sep	76621	54589.5	98653	27-Oct	91094	62918.3	119270
29-Aug	62631	48948.6	76313	28-Sep	77104	54839.7	99368	28-Oct	91577	63218.8	119934
30-Aug	63113	49060.4	77166	29-Sep	77586	55092.4	100080	29-Oct	92059	63520.6	120598
31-Aug	63596	49181.5	78010	30-Sep	78069	55347.4	100790	30-Oct	92542	63823.4	121260
01-Sep	64078	49311.5	78844	01-Oct	78551	55604.7	101497	31-Oct	93024	64127.4	121921
02-Sep	64560	49449.8	79671	02-Oct	79033	55864.2	102203	01-Nov	93506	64432.4	122580
03-Sep	65043	49595.7	80490	03-Oct	79516	56125.7	102906	02-Nov	93989	64738.5	123239

04-Sep	65525	49748.8	81302	04-Oct	79998	56389.4	103607	03-Nov	94471	65045.7	123897
05-Sep	66008	49908.7	82107	05-Oct	80481	56655.0	104306	04-Nov	94954	65353.9	124553
06-Sep	66490	50074.9	82905	06-Oct	80963	56922.7	105004	05-Nov	95436	65663.2	125209
07-Sep	66973	50247.0	83698	07-Oct	81446	57192.2	105699	06-Nov	95919	65973.4	125864
08-Sep	67455	50424.8	84485	08-Oct	81928	57463.5	106392	07-Nov	96401	66284.7	126517
09-Sep	67937	50608.0	85267	09-Oct	82410	57736.7	107084	08-Nov	96883	66596.9	127170
10-Sep	68420	50796.2	86043	10-Oct	82893	58011.6	107774	09-Nov	97365	66910.0	127822
11-Sep	68902	50989.3	86815	11-Oct	83375	58288.3	108462	10-Nov	97848	67224.1	128472







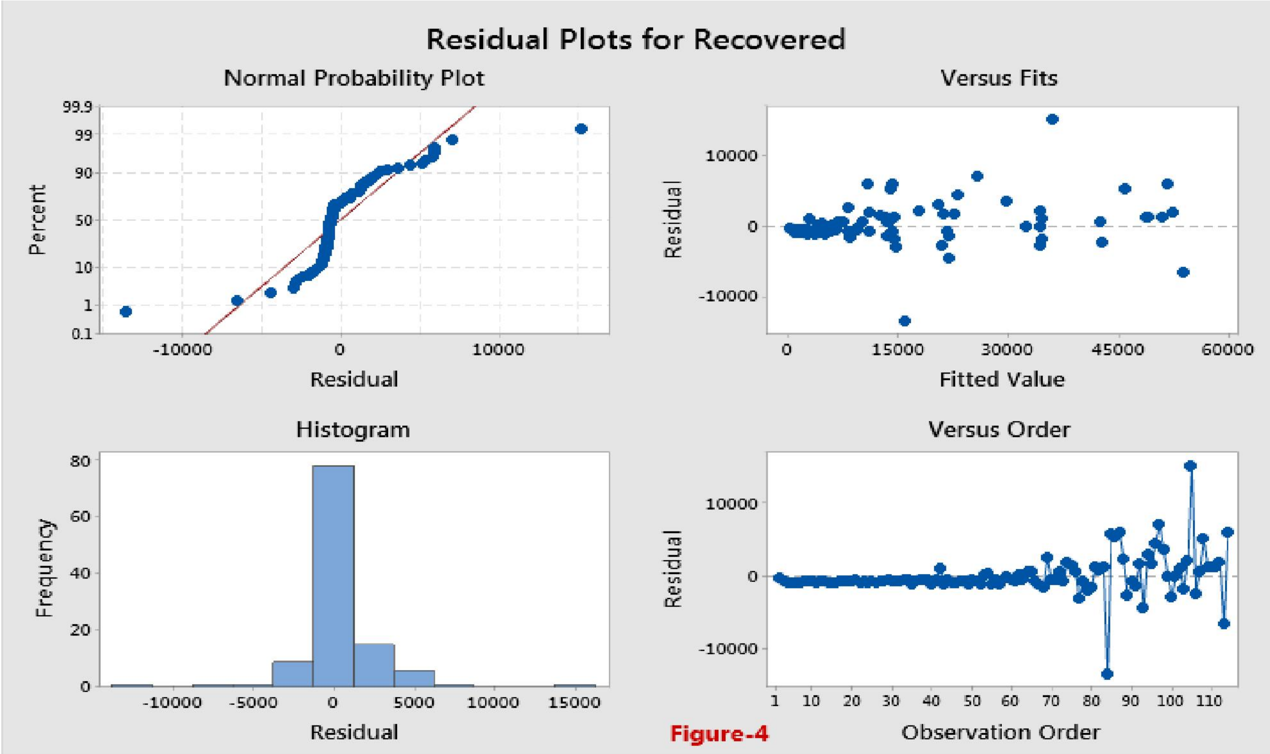


Figure-4

There is significant autocorrelation at lag numbers 5, 9, 11 and 19 at 5% level of significance. (Figure-2). Partial autocorrelation is significant at lag numbers 5, 9 11 and 21 at 5% level of significance. (Figure-3).

Trend Analysis for Forecast of Recovered Cases:

The trend model which explains the data best is the linear trend model (Figure-5). Thus the linear trend model is fitted for the predicted confirmed cases of covid-19 in India.

Fitted Trend Equation: Linear Model

$Z_t = 54393.5 + 483.020 * t$ ....(10)

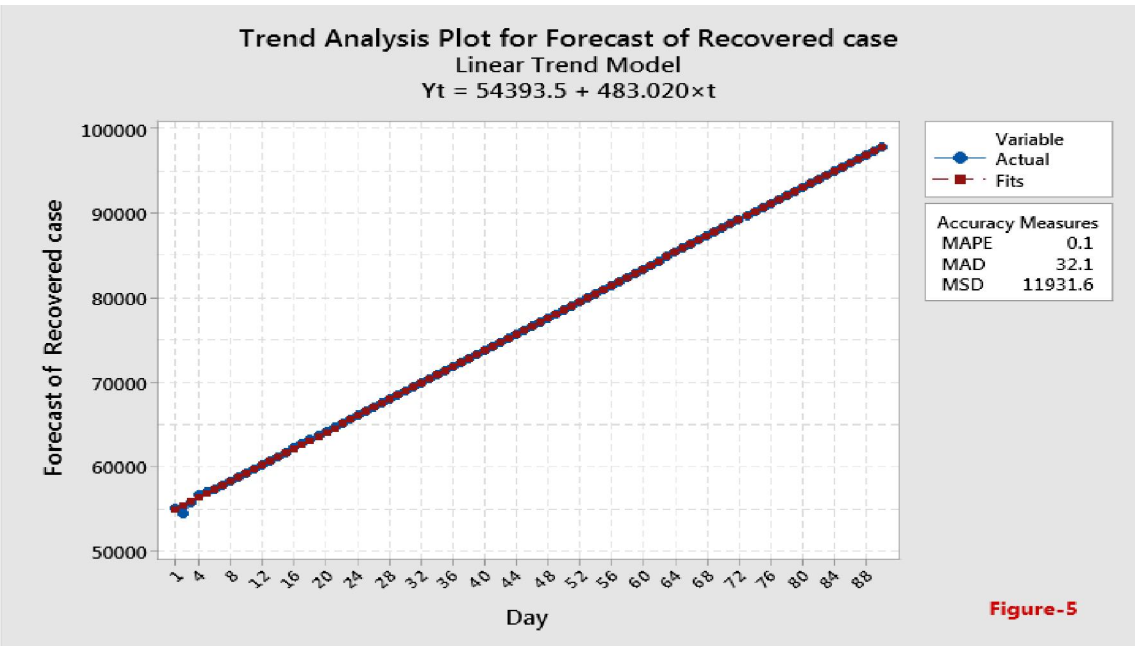


Figure-5

Table-6: Rate of Recovery

Rate of recovery per 100 confirmed cases									
13-Aug	83.51	31-Aug	85.27	18-Sep	85.26	06-Oct	85.23	24-Oct	85.22
14-Aug	84.84	01-Sep	85.30	19-Sep	85.26	07-Oct	85.23	25-Oct	85.21
15-Aug	84.07	02-Sep	85.27	20-Sep	85.25	08-Oct	85.23	26-Oct	85.21
16-Aug	86.36	03-Sep	85.29	21-Sep	85.25	09-Oct	85.23	27-Oct	85.21
17-Aug	84.98	04-Sep	85.27	22-Sep	85.25	10-Oct	85.23	28-Oct	85.21
18-Aug	85.68	05-Sep	85.28	23-Sep	85.25	11-Oct	85.23	29-Oct	85.21
19-Aug	84.91	06-Sep	85.27	24-Sep	85.25	12-Oct	85.23	30-Oct	85.21
20-Aug	85.57	07-Sep	85.28	25-Sep	85.25	13-Oct	85.23	31-Oct	85.21
21-Aug	85.12	08-Sep	85.27	26-Sep	85.25	14-Oct	85.23	01-Nov	85.21
22-Aug	85.47	09-Sep	85.27	27-Sep	85.24	15-Oct	85.22	02-Nov	85.21
23-Aug	85.17	10-Sep	85.27	28-Sep	85.24	16-Oct	85.22	03-Nov	85.21
24-Aug	85.40	11-Sep	85.27	29-Sep	85.24	17-Oct	85.22	04-Nov	85.21
25-Aug	85.22	12-Sep	85.26	30-Sep	85.24	18-Oct	85.22	05-Nov	85.21
26-Aug	85.36	13-Sep	85.27	01-Oct	85.24	19-Oct	85.22	06-Nov	85.21
27-Aug	85.24	14-Sep	85.26	02-Oct	85.24	20-Oct	85.22	07-Nov	85.20
28-Aug	85.33	15-Sep	85.26	03-Oct	85.24	21-Oct	85.22	08-Nov	85.20
29-Aug	85.26	16-Sep	85.26	04-Oct	85.24	22-Oct	85.22	09-Nov	85.20
30-Aug	85.31	17-Sep	85.26	05-Oct	85.24	23-Oct	85.22	10-Nov	85.20

4. Conclusions

An ARIMA method of time series for predicting the forecasts of covid19 confirmed cases has an increasing trend. For the period i.e. from 13th August to 10th November 2020 the daily number of recovered cases of covid-19 in India will be having an increasing tendency. It is predicted that the recovered cases on 10th of November will 97848 with confidence coefficient (67224 to 128472). Prediction of recovered cases with 95% confidence limits are shown in Table-5. It is observed that the rate of recovery of covid-19 patients in India will be around 85% to 86% of confirmed cases for the period of 90 days from 13th August to 10th November 2020 in India. Prediction of recovered case and recovery rate of covid19 cases will certainly helps the government to form the policies for the nation.

References

1. M.H.D.M. Ribeiro, R.G. da Silva and V.C. Mariani, L. D. S. Coelho (2020), Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil, Chaos, Solitons and Fractals, 135, 109853.
2. L. Li, Z. Yang, Z. Dang, C. Meng, J. Huang, H. Meng, D. Wang, G. Chen, J. Zhang, H. Peng, Y. Shao (2020), Propagation analysis and prediction of the

COVID-19, Infectious Disease Modelling, 5, 282-292.

3. F. J. D. Perez, D. Chinarro, R. P. Otin, Adib (2020), Guardiola Mouhaffel, Growth forecast of the covid-19 with the Gompertz function, Case study: Italy, Spain, Hubei (China) and South Korea, International Journal of Advanced Engineering Research and Science (IJAERS) Vol-7, Issue-7, July 2020, 67-77.
4. J. Luo (2020), When Will COVID-19 End? Data-Driven Prediction, Data-Driven Innovation Lab (http://ddi.sutd.edu.sg) Singapore University of Technology and Design (http://www.sutd.edu.sg)
5. J. Stubinger and L. Schneider (2020), Epidemiology of Coronavirus COVID-19: Forecasting the Future Incidence in Different Countries, Healthcare 2020, 8, 99, 2-15
6. S. Khajanchi and K. Sarkar (2020), Forecasting the daily and cumulative number of cases for the COVID-19 pandemic in India, Chaos 30, 071101 (2020); doi: 10.1063/5.0016240, 1-16
7. R. Salgotra, M. Gandomi, A. H. Gandomic, (2020), Time Series Analysis and Forecast of the COVID-19 Pandemic in India using Genetic Programming, Chaos, Solitons and Fractals 138 , 109945, 2-15
8. S. He, S. Tang, L. Rong (2020), A discrete stochastic model of the COVID-19 outbreak: Forecast and control, Mathematical Biosciences and Engineering Volume 17, Issue 4, 2792–2804.
9. A. Bhattacharjee, M. Kumar, K. K. Patel (2020), When COVID-19 will decline in India? Prediction by combination of recovery and case load rate, Article in press, Clinical Epidemiology and Global Health xxx (xxxx) xxx–xxx https://doi.org/10.1016/j.cegh.2020.06.004, Volume 9, January–March 2021, Pages 17-20
10. A. Das (2020), COVID-19 Predictions for India An Initial Attempt, (Technical Report) Professor, Department of Computer Science and Engineering Indian Institute of Technology Kharagpur India URL: http://cse.iitkgp.ac.in/_abhij/
11. S. Roy, G. S. Bhunia, P. K. Shit (2020), Spatial prediction of COVID‑19 epidemic using ARIMA techniques in India, Modeling Earth Systems and Environment https://doi.org/10.1007/s40808-020-00890-y