ISSN: 0974-5823 Vol. 6 (Special Issue, Nov.-Dec. 2021)

International Journal of Mechanical Engineering

Suicide in India: A Time Series Analysis

Pinky Roy[1], Bidyut Bikash Boruah[2], Kabir Barhoi[3] and Kuki Kalpita Mahanta[4]

[1]Research Scholar, Department of Statistics, Dibrugarh University-786004, India

[2]Department of Statistics, Dibrugarh University-786004, India

[3]Department of Mathematics, Dibrugarh University-786004, India

[4]Asst. Professor, Department of Statistics, Dibrugarh University-786004, India

Abstract: Suicide is a very complex and multifactorial act of intentionally causing harm to own. As compared to younger people suicide occurs more often in older, also in late childhood and adolescence. Suicide results in direct loss of lives, along with bad psychological and socio – economic effects. The factors which are responsible for suicide are – mental disorders, specific personality characteristics, stress, depression, genetic loading, exposure to inspiring models and availability of means of committing suicide. In the study, we examine the trend and structural breakpoints of suicide in India for 53 years from 1967 to 2019 and forecast it for next 15 years using ARIMA (Autoregressive integrated Moving Average) model. Forecasting is a valuable one to take necessary actions to prevent and make necessary actions for this type of activities.

Key words**:** Suicide, Trend, Breakpoints, ARIMA, Forecast.

I. Introduction

The word suicide comes from the Latin word Sucidium, means "to kill oneself". Literally, suicide is the act of intentionally causing one's own death (1). In the world, more than 800000 people die due to suicide each year. In the year 2015 mortality rate due to suicide was 10.7 per 100000. 1.4% of total deaths caused due to suicide globally (2,3). As compared to woman, many more men are died due to suicide and are higher in richer countries (4). Attempt of suicide are much more frequent than actual suicide. It is to be estimated that 2.5% of the population make at least one suicide attempt during lifetime (6,7).

Suicide rates vary from region to region. Most of the suicide occurs in low to middle income groups (8). Worldwide, rates of suicide are higher among older than the younger generation (9). Suicidal behavior among personalities in India becomes one of the most important issues because its incidence manifests a rising trend day by day (10). The rates of suicide increase day by day, so it's a matter of concern among the scientist and for policy makers. Everyone agrees that various factors contribute to suicide; ultimately each suicide is caused by highly unique, dynamic and complex interplay of biological, genetic, psychological and social factors (11). The risk factors which are associated with suicide are like stress, depression, failure in examination, alcohol use disorder, financial problems, relational break up etc. (12-15). A groundbreaking treatise of the phenomenon of suicide from a socio – cultural perspective was authored by Emile Dukheim in 1897 (16).

A sequence of numerical data, that occurs at uniform intervals over period of times are considered as time series data. the time series analysis is mainly used for seeing the past trend and for future forecasting (Monfared et al.2013(17)). ARIMA (Autoregressive Integrated Moving Average) is the most effective measure for analyzing time series data. The forecasting of suicide rates or numbers of deaths due to suicide may be very important to taking different precautional measures. It will help scientists and policy makers to take necessary actions to decrease the rate of that.

In this paper we examine the trend of suicide rates in India for 53 years from 1967-2019. Appropriate models are also being used to forecast the suicide rates for next 15 years. These analyses are done for data dataset of India as a whole and also gender wise.

II. Data source and methods

Data used in this paper are secondary in nature and collected from NCRB (National Crime Record Bureau), India. NCRB is a nodal agency that functions under the Ministry of Home Affairs, Government of India and publishes data on suicide and accidental deaths yearly basis. We collected data on suicide from 1967 to 2019 of 53 years, which are freely available.

The ARIMA (AutoregressiveIntegrated Moving Average) has been used here for forecasting. This is an advanced model, that requires long term data to predict the future occurrences. This model decomposes the past data into an Autoregressive (AR) process, which taking care of the past values, an Integrated (I) process, which is used for making the data stationary and a Moving average (MA) process, which accounts for previous error terms making it easier to forecast.

III. Data Analysis:

In this study we consider the numbers of death by committing suicide in India for 53 years from 1967 to 2019. These 53 years data are collected from the report published by National Crime Records Bureau, Ministry of Home Affairs, Government of India in the year 2020. In table -1 we show the basic descriptive statistics obtained from the data.

Table 1: Basic Descriptive Statistics of the Suicide data (year 1967-2019)

Measures	Observations		
	Male	**Feale**	**Total**
Minimum	22637 (year =1967)	15237	38217
Maximum	97613 (year=2019)	47746	139123
1st Quartile	26923 (year=1980)	18212	44732
3rd Quartile	75702(year=2006)	42192	118112
Mean	52787.53	31443	84232
Median	49851	34393	84244
SE Mean	3522.23	16122	5072.995
LCL Mean	3522.252	282077.1	74052.35
UCL Mean	59855.44	34678.29	94411.77
Standard Deviation	25642.39	11737.61	36931.96
Skewness	0.269744	-0.159530	93236.00
Kurtosis	-1.520194	-1.71385	-1.65193

Test for Normality Check of the Suicide Data:

The Jarque- Bera Test is performed to test the normality of the data with the help of the R software and the results are shown in Table – 2.

Table 2: Jarque Bera test to check normality

Hypothesis	**Category**	**Jarque Bera Test Statistic**	**Degrees of freedom**	**p- value**
H_0: the dataset is normally distributed H_0: the dataset is not normally distributed	For male	5.4054	2	0.16702
	For female	6.352	2	0.14175
	For total	5.7317	2	0.15693

Since p – values are greater than 0.05, so we have to accept the hull hypothesis that the data are normally distributed.

The time series plot of the death by Suicide in Fig – 1, shows an upward stochastic trend.

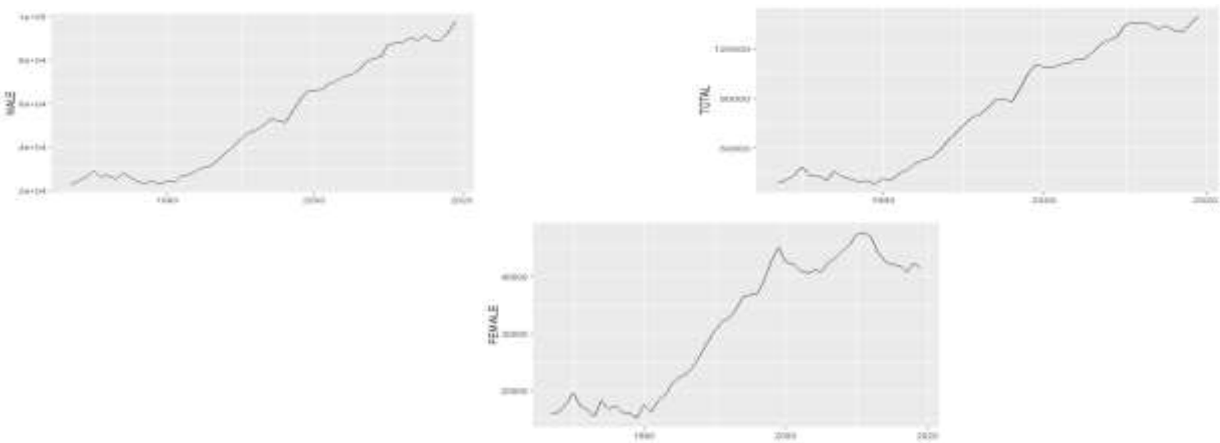


Fig1: Time series plot of death by suicide of male, female and total from the year 1967-2019

Test for detection of structural break points:

To detect the structural break points in the datasets, here we perform the F- test with the help of R software and the results are shown in Table 3,

Table 3 : F-test for detection of structural breakpoints for male

Hypothesis	Test Statistic	p- value
H_0 : Absence of structural break points H_1: Presence of structural break points	F = 241.63	p-value < 2.2e-16
	F = 374.26	p-value < 2.2e-16
	F = 262.34	p-value < 2.2e-16

From table 3, the p- values are less than 0.05, therefore, we may reject the null hypothesis indicating some structural break points in the dataset. According to R software, the structural break points for male, female and total suicide cases are shown in table 4, 5 and 6 respectively.

Table 4 : Structural break points (for male suicide)

Observation number	Year	Male suicide
17	1983	85577
24	1990	117932
31	1997	164876
39	2005	224806
46	2012	306061

From the table 4, there are five structural breakpoints for male.

Table3.5: Structural break points (for female suicide).

Observation number	Year	Female suicide
16	1982	43016
23	1989	54113
30	1996	63988

Table 5 indicates that, there are only three structural breakpoints in case of female suicide in 53 years.

Table3.6: Structural break points (for total suicide).

Observation number	Year	Suicide in India
17	1983	128576
24	1990	174401
31	1997	233903
40	2006	314704

Table 6 shows that, there are four structural breakpoints for suicide in India from 1953 to 2019.

Testing of stationary:

We have to plot the Auto Correlation Function (ACF) and Partial Auto Correlation Function(PACF) to test the stationary of the collected datasets with R-software and are shown in Fig-2 and Fig -3 and Fig-4

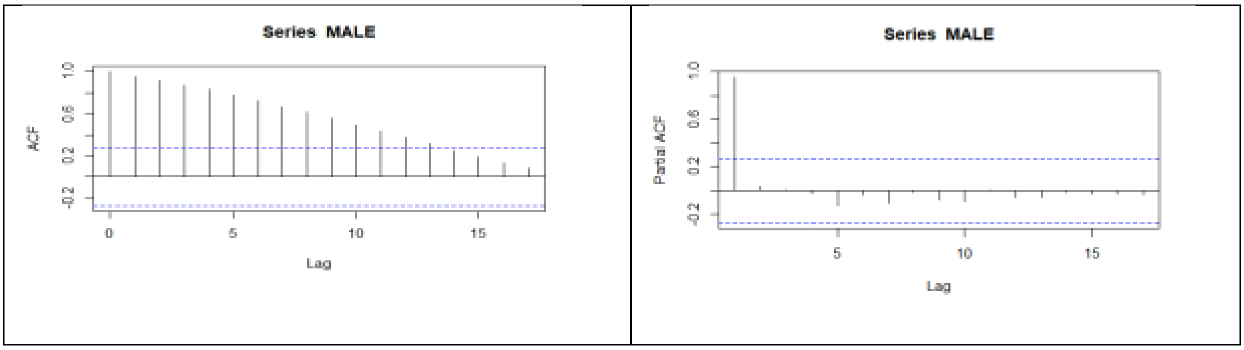


Fig 2: ACF and PACF for male suicide

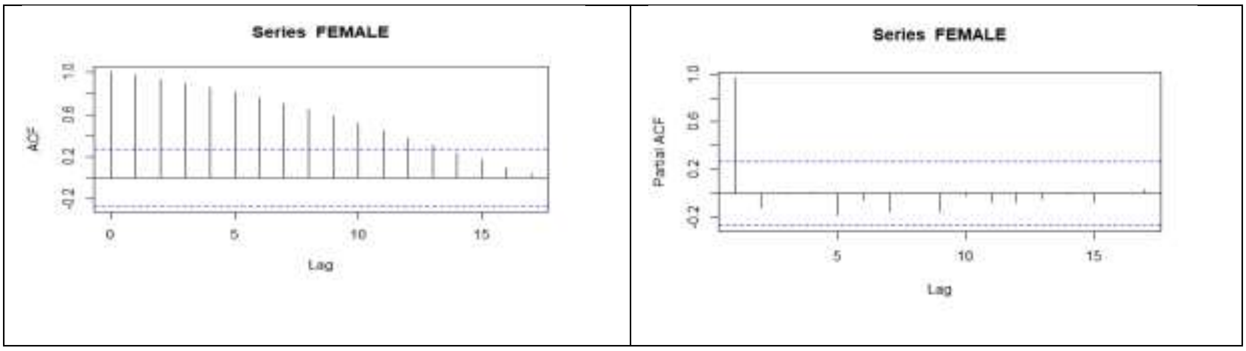


Fig 3: ACF and PACF for female suicide

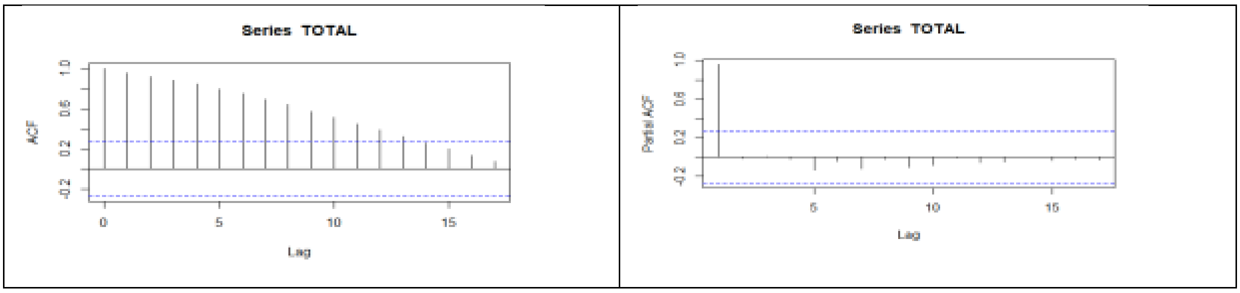


Fig 4: ACF and PACF for suicide in India

From Fig -2, Fig- 3 and Fig -4 it is seen that, for ACF most of the bars crosses the upper limits and also for PACF one bar cross the upper limit, so according to the criteria of ACF and PACF our dataset of nos. of deaths by suicide is not stationary.

Now, with the help of R software the Augmented Dickey – Fuller test can be used to test it statistically and the results are shown in table 9.

Table 7: Augmented Dickey – Fuller Test

Hypothesis		Test Statistic	Lag order	p- value
H_0 : Data is not stationary H_1 : Data is stationary	Male Death	-3.2414	3	0.0902
	Female Death	-1.3226	3	0.8471
	Total Death	-2.9684	3	0.1847

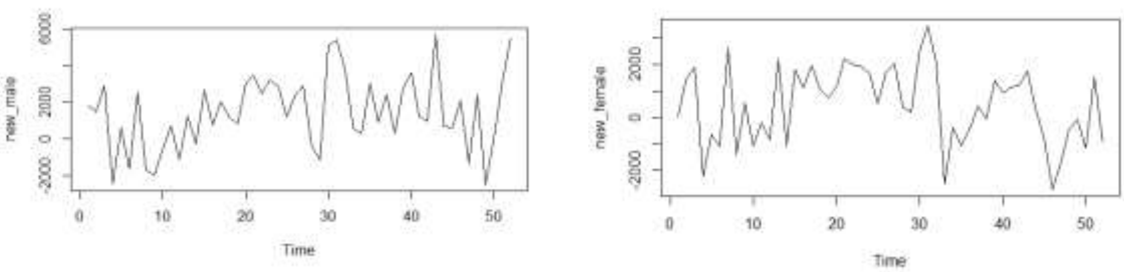
From table 7, it is seen that p – values are greater than 0.05, so null hypothesis should be accepted (data sets are not stationary)

Process of making data stationary:

The stochastic trend from the dataset should be removed to make the dataset stationary. To do that, we take the double difference i.e.

New_male = $\Delta Y_t - \Delta Y_{t-1}$

After taking the double differences the data follows the pattern shown in Fig 5



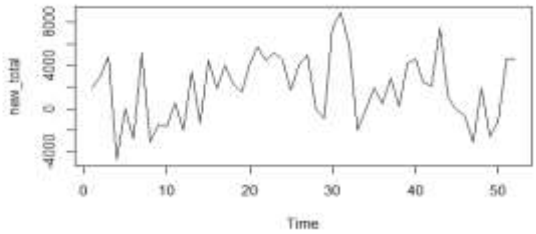


Fig 5: Time series plot of new _ male, new _ female and new _ death data after taking the log differences

For checking the stationary of the datasets after taking the double differences, we graphically plot the ACF and PACF with the help of R software and are shown them in following figures Fig 6

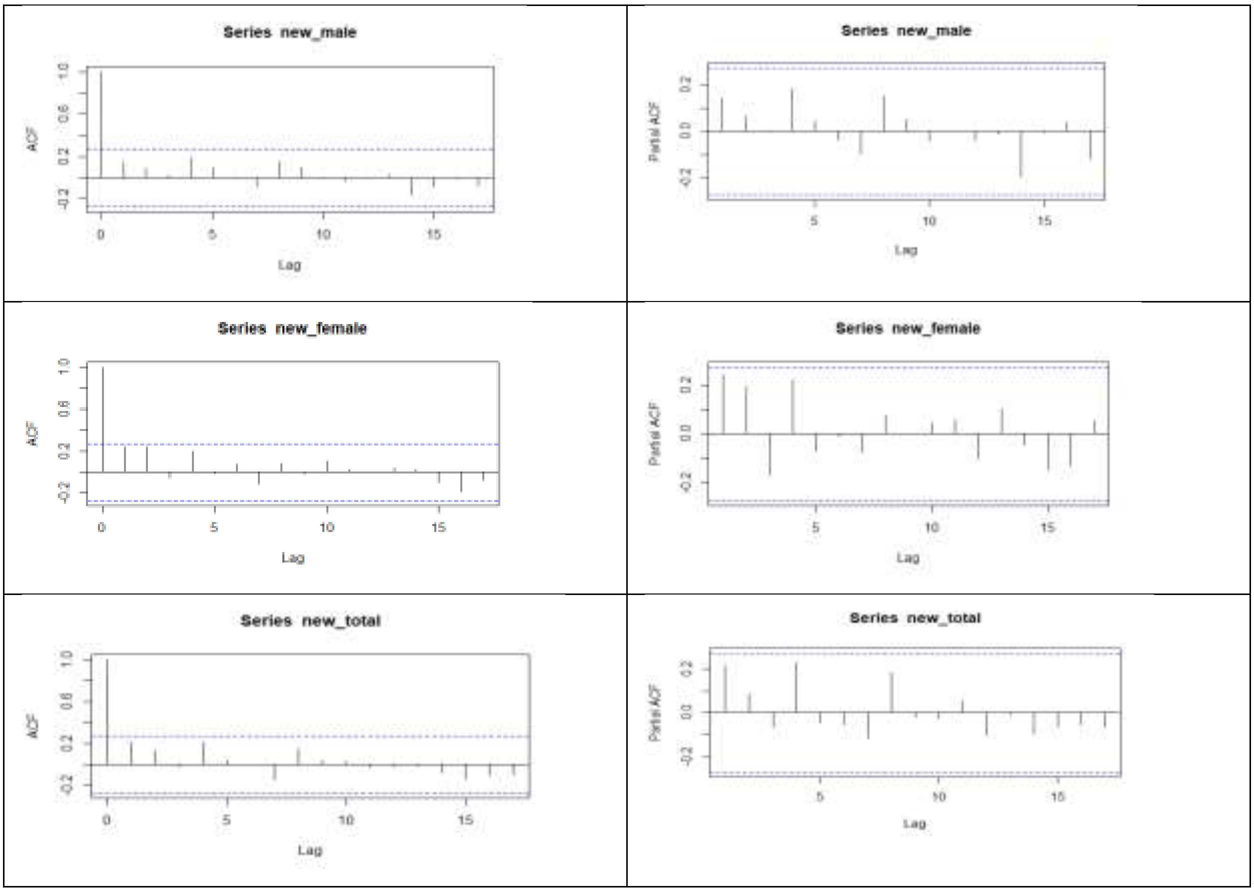


Fig 6: ACF and PACF of suicide data after taking the double differences

From Fig 6, it is seen that for all three datasets most of the bars are within the limits , so now the data becomes stationary on the basis of the criteria of ACF and PACF. Here we use Augmented Dickey Fuller test to test the stationarity of the dataset statistically. The results are shown in the Table 8.

Table 8: Augmented Dickey Fuller test

Hypothesis		**Test Statistic**	**Lag order**	**P - value**
H_0 : Data is not stationary	Male Death	-5.4756	3	0.01834
H_1 : Data is stationary	Female Death	-6.0269	3	0.01293
	Total Death	-5.8442	3	0.01511

From the table 8, it is seen that p – values for all are less than 0.05, so we reject our null hypothesis that data is not stationary.

ARIMA Model Fitting:

The datasets are become stationary after taking the log differences. Therefore, in case of fitting the ARIMA(p, d, q) model, the order of d is identified as 2. For identification of best fitted models for the datasets we propose five different ARIMA(p, d, q) models keeping d = 2 as constant. Here the best model to be chosen on the basis of AIC and Log – likelihood criteria.

The estimated models for considering the dataset of male suicide over the years

Model 1 = ARIMA(1,2,0) = $-0.4584y_{t-1} + e_t$

Model 2 = ARIMA(0,2,1) =$-0.8715e_{t-1} + e_t$

Model 3 = ARIMA(1,2,1) =$0.0752y_{t-1} - 0.8943e_{t-1} + e_t$

Model 4 = ARIMA(1,2,2) = $0.9003y_{t-1} - 1.7701e_{t-1} + 0.7706e_{t-2} + e_t$

Model 5 = ARIMA(2,2,1) =$0.0811y_{t-1} + 0.0156y_{t-2} - 0.900e_{t-1} + e_t$

The estimated models for considering the dataset of female suicide over the years

Model 1 = ARIMA(1,2,0) = $-0.4961y_{t-1} + e_t$

Model 2 = ARIMA(0,2,1) = $-0.8024e_{t-1} + e_t$

Model 3 = ARIMA(1,2,1) =$0.1488y_{t-1} - 0.8660e_{t-1} + e_t$

Model 4 = ARIMA(1,2,2) = $0.7276y_{t-1} - 1.4913e_{t-1} + 0.4312e_{t-2} + e_t$

Model 5 = ARIMA(2,2,1) =$0.2132y_{t-1} + 0.2128y_{t-2} - 1.000e_{t-1} + e_t$

The estimated models for considering the dataset of suicide rate in India over the years

Model 1 = ARIMA(1,2,0) = $-0.4418y_{t-1} + e_t$

Model 2 = ARIMA(0,2,1) = $-0.8321e_{t-1} + e_t$

Model 3 = ARIMA(1,2,1) =$0.1771y_{t-1} - 0.9134e_{t-1} + e_t$

Model 4 = ARIMA(1,2,2) = $0.7194y_{t-1} - 1.5205e_{t-1} + 0.5128e_{t-2} + e_t$

Model 5 = ARIMA(2,2,1) =$0.2172y_{t-1} + 0.1102y_{t-2} - .870e_{t-1} + e_t$

Table 9: AIC and Log – Likelihood of the fitted ARIMA model(Male)

Model	**ARIMA order**	**AIC**	**Log Likelihood**
Model 1	ARIMA(1,2,0)	983.32	-489.66
Model 2	ARIMA(0,2,1)	971.51	-483.75
Model 3	ARIMA(1,2,1)	945.11	-469.55
Model 4	ARIMA(1,2,2)	921.13	-456.56
Model 5	ARIMA(2,2,1)	937.73	-464.86

From the table 9, it is seen that AIC value is minimum and Log – likelihood is maximum for ARIMA(1,2,2) model. So it considered as the best model and is used for forecasting future deaths.

Table 10: AIC and Log – Likelihood of the fitted ARIMA model(Female)

Model	**ARIMA order**	**AIC**	**Log Likelihood**
Model 1	ARIMA(1,2,0)	931.68	-463.84
Model 2	ARIMA(0,2,1)	934.67	-465.33
Model 3	ARIMA(1,2,1)	901.81	-447.9
Model 4	ARIMA(1,2,2)	882.73	-437.36
Model 5	ARIMA(2,2,1)	901.84	-446.92

Here also for the dataset of female deaths AIC value is minimum for ARIMA (1,2,2) model, so we use this model for forecasting.

Table 11: AIC and Log – Likelihood of the fitted ARIMA model(Total)

Model	**ARIMA order**	**AIC**	**Log Likelihood**
Model 1	ARIMA(1,2,0)	931.68	-463.84
Model 2	ARIMA(0,2,1)	934.67	-465.33
Model 3	ARIMA(1,2,1)	901.81	-447.9
Model 4	ARIMA(1,2,2)	882.73	-437.36
Model 5	ARIMA(2,2,1)	901.84	-446.92

From the table 13 it is seen that for the data of total deaths due to suicide in India, ARIMA(1,2,2) model shows best by comparing the values of AIC and log likelihood.

Diagnostic checking of ARIMA(1,2,2) models:

To get an idea about the significance of the AR and MA coefficients, we perform the Z-test with the help of R software and the results are shown in table

Table 12: Z – test of the AR and MA coefficients(male)_

Coefficients	**Estimate**	**Std. Error**	**Z value**	**Pr(>|z|)**
AR1	-0.423946	0.127796	-3.3174	0.0009087 ***
MA1	-1.966820	0.087544	-22.4666	< 2.2e-16 ***
MA2	0.999987	0.087841	11.3841	< 2.2e-16 ***

Table 13: Z – test of the AR and MA coefficients(female)

Coefficients	**Estimate**	**Std. Error**	**Z value**	**Pr(>|z|)**
AR1	-0.463973	0.127522	-3.6384	0.0002744 ***
MA1	-1.991897	0.083484	-23.8596	< 2.2e-16 ***
MA2	0.999963	0.083299	12.0045	< 2.2e-16 ***

Table 14: Z – test of the AR and MA coefficients(total)

Coefficients	**Estimate**	**Std. Error**	**Z value**	**Pr(>|z|)**
AR1	-0.463973	0.127522	-3.6384	0.0002744 ***
MA1	-1.991897	0.083484	-23.8596	< 2.2e-16 ***
MA2	0.999963	0.083299	12.0045	< 2.2e-16 ***

In table 12, 13 and 14 all the p- values are less than 0.01, so it is said that AR and MA coefficients are significant for ARIMA(1,2,2).

Measure of accuracy:

Table 15: Error Measures for male ARIMA(1,2,2)

Error measures	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
	-0.0021	0.04637728	0.03587609	174.2142	270.299	0.4005469	0.05937517

Table 16: Error Measures for female ARIMA(1,2,2)

Error measures	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
	-0.0007	-0.00071	0.04106328	94.26261	156.3559	0.4061925	0.09318845

Table 17: Error Measures for total ARIMA(1,2,2)

Error measures	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
	-0.0016	0.04603487	0.03492619	63.38916	196.9121	0.3980652	0.0802338

Forecasting with the help of ARIMA(1,2,2) Model for deaths by committing suicide:

The forecasted deaths(point forecast) of males, females and also for the total population along with 80% and 95% upper and lower confidence intervals for the next 15 years based on ARIMA(1,2,2) model ase constructed using R software and are shown in table

Table 18: forecasted values and confidence intervals for males

Year	Point Forecast	80% CI(Lower Limit)	80% CI(Upper Limit)	95% CI(Lower Limit)	95% CI(Upper Limit)
2020	99392.4	9 6811	101973	95445	103339
2021	101137	97217	105058	95141	107134
2022	102852	97732	107973	95021	110683

2023	104540	98275	110805	94958	114121
2024	106202	98821	113584	94914	117491
2025	107843	99362	116323	94873	120813
2026	109463	99896	119031	94831	124095
2027	111066	100421	121710	94786	127345
2028	112652	100939	124364	94739	130564
2029	114223	101451	126995	94690	133756
2030	115781	101958	129604	94641	136921
2031	117328	102462	132193	94593	140062
2032	118863	102963	134763	94546	143180
2033	120389	103463	137315	94503	146276
2034	121907	103963	139851	94464	149350

From table 18 and fig 7 it is observed that rate of suicide for males gradually increase in future.

Table 19: Forecasted values and confidence intervals for females

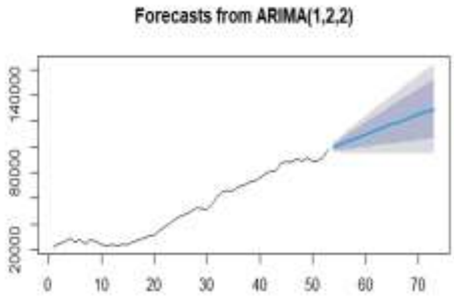
Year	Point Forecast	80% CI(Lower Limit)	80% CI(Upper Limit)	95% CI(Lower Limit)	95% CI(Upper Limit)
2020	41613	39807	43420	38851	44376
2021	41828	38932	44723	37399	46256
2022	42110	38211	46009	36147	48073
2023	42441	37594	47288	35028	49854.
2024	42808	37060	48556	34018	51598
2025	43201	36595	49808	33097	53305
2026	43613	36186	51040	32254	54972
2027	44039	35825	52253	31477	56601
2028	44475	35505	53445	30757	58193
2029	44918	35219	54617	30085	59751
2030	45366	34962	55770	29455	61278
2031	45819	34731	56906	28862	62775
2032	46274	34521	58026	28300	64247
2033	46731	34330	59131	27766	65695
2034	47189	34156	60222	27256	67122

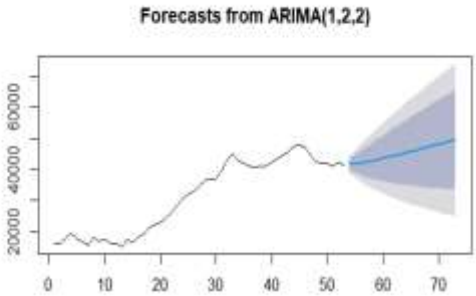
Table 20: Forecasted values and confidence intervals for total nos. of suicide in India

Year	Point forecast	80% CI(lower limit)	80% CI (upper limit)	95% CI (lower limit)	95% CI(upper limit)
2020	141539.5	137603	145475	135519	147559
2021	143827.5	137628	150026	134347	153308
2022	146023	137779	154267	133415	158631
2023	148152	137996	158308	132620	163684
2024	150233	138271	162195	131939	168527
2025	152280	138603	165956	131363	173196
2026	154302	138990	169613	130885	177719
2027	156306	139430	173182	130496	182116

2028	158297	139918	176677	130189	186406
2029	160280	140452	180108	129955	190604
2030	162255	141026	183484	129788	194722
2031	164226	141638	186814	129681	198771
2032	166193	142284	190102	129627	202759
2033	168158	142960	193356	129621	206695
2034	170121	143664	196579	129658	210584

The following fig. shows the forecast of female suicide





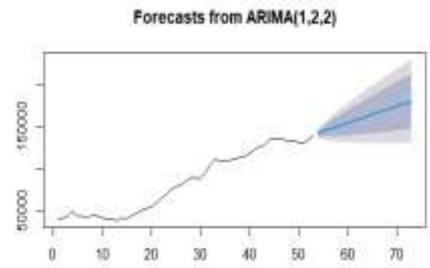


Fig 7 : Forecasted plot of male death by committing suicide Fig 8: forecasted plot of female suicide Fig 9: Forecasted plot of rate of suicide in India

From table 19 and fig 8, we observed that rate of suicide for females increase in future but its lower than the male increase rate. From table 20 and fig 9 observed that rate of suicide is gradually increase in India as a whole.

Conclusion:

In this paper, the yearly deaths due to suicide in India have been studied using time series models. On the basis of the Jarque- Bera test statistic, it is observed that data are normally distributed. Here the numbers of suicide show the upward stochastic trend. The datasets show structural breakpoints. Double differentiation has to be considered to make the data stationary for fitting of ARIMA models. The results of the estimation and diagnostics analysis revealed that used models are adequately fitted to the historical datasets, Augmented Dickey Fuller test is used here to check the adequacy. So we may concluded that, the model ARIMA (1,2,2) to be best fitted model for all the three datasets of numbers of suicide in India from 1967-2019. The forecasted numbers of deaths by suicide for next 15 years also show upward trend. Therefore, the forecasting will help the government and policy makers to take necessary actions to prevent these kind of activities, because unnaturaland unpredictable deaths create great loss to the family and for the community as whole.

REFERENCES:

1. G. Turecki, D. Brent, Suicide and suicidal behavior, Lancet, Vol. 387, pp. 1227 – 1239, 2016.
2. WHO preventing suicide: A GLOBAL IMPERATIVE, Geneva: World Health Organization, 2014. http://www.who.int/mental_health/suicide-preventation/world_report_2014/en/
3. WHO global health observatory. Geneva: World Health Organisation,2017, http:www.who.int/gho.
4. K. Hawton, K. V. Heeringen, Suicide, Lancet, Vol. 373, pp.1372-1381, 2009.
5. D. De. Leo, Can we rely on suicide mortality data?, Crisis, Vol.36, pp-1-3, 2015.

6. G. Borges, M. K. Nock et al., Twelve month prevalence of and risk factors for suicide attempts in the World Health Organisation World Mental Health Surveys, J. Clin Psychiatry, Vol. 71, pp. 1617-1628, 2010.

7. M. K. Nock, G. Borges et al, Cross- national prevalence and risk factors for suicidal identification, plans and attempts, Br J Psychiatry, Vol. 192, pp. 98-105, 2008.

8. WHO Mental Health, World Health Organisation: Geneva, 2018.

9. P. Varnik, Suicide in the world, Int J Environ Res Public Health, Vol. 9, pp. 760-771, 2012.

10. M. K. Panday, C. Kaur, Investigating suicidal trend and its economic determinants: Evidance from India, Asarc wp. 2009.

11. K. H. Van, The suicidal process and related concepts, Understanding Suicidal Behaviour, John Wiley & Sons Ltd., pp.136-59, 2001.

12. A. V. Rao, Attempted suicide, Indian J Psychiatry, Vol. 7, pp.253-267, 1965.

13. A. V. Rao, Marriage, parenthood, sex and suicidal behavior, Indian J. Psychiatry, Vol. 16,pp. 92-104, 1974.

14. D. N. Nandi, S. P. Mukherjee et al, Is suicide preventable by restricting the availability of lethal agents, A rural survey of West Bengal? Indian J Psychitry, vol. 21, pp. 251-256, 1979.

15. A. F. Khan, B. Anand, G. M. Devi, Psychological autopsy of suicide : A cross sectional study, Indian J Psychiatry, Vol 47, pp. 73-81, 2005.

16.E. Durkheim, Suicide: A study in sociology, New York : Free Google Press, 1897.

17. A. B. Monfared et al, Prediction of fatal road traffic crashes in Iran using Box – Jenkins time series model, J Asian Sci Res, vol. 3(4), pp.425-430, 2013.

18. National Crime Record Bureau, Ministry of Home Affairs annual report on accidental deaths and sucides in India from 1967 to 2019.