A REPORT
ON

PREDICTING COVID-19 DEATHS FOR INDIA USING TIME-SERIES ANALYSIS

By
Group 7

Under Supervision of
Dr. Rishi Kumar

Course Code: ECON F342
Course Title: APPLIED ECONOMETRICS

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI HYDERABAD CAMPUS

(May 2023)

GROUP DETAILS

Group Number: 7

S.No.	GROUP MEMBERS	ID NO.
1.	ALAMURI SIRI CHANDANA	2020B3A70854H
2.	ARUNDHATI BAJAJ	2020B3A71933H
3.	P SAI SHRUTHI	2020B3A70904H
4.	SAIRA DAS	2020B3A31498H
5.	SWETHA KRISHNA SRIRAM	2020B3A71252H
6.	PODDAR MEHUL ASHISH	2020B3A72158H
7.	HRITIK CHHABRA	2020B3A71929H
8.	PRATYUSH TRIPATHI	2020B3AA1838H

Acknowledgement

We would like to express our sincere gratitude to Dr. Rishi Kumar for giving us the opportunity to work on this assignment and expand our knowledge beyond the classroom. His constant guidance and support was invaluable for the completion of this project. His advice was extremely crucial at every stage in this project. We would like to express our gratitude to him once again for providing us with this magnificent opportunity to gain hands-on experience and put what we had learned in the classroom to use. We certainly owe him a lot for his time and patience, and his continuous direction, both in this project, and in the course as a whole.

1 Introduction

The COVID-19 pandemic has had a significant impact on every arena, throwing the world into unexpected chaos. The virus that originated from Wuhan, China, spread around the globe at the speed of light, causing large-scale economic disruption and loss of life. A study by the World Bank estimated that the pandemic caused a global economic contraction of 5.2% in 2020, with the number of people living in abject poverty rising by over 100M. Another study by Hogan et al. (2020), published in The Lancet stated that the pandemic could cause an additional 1.15 million deaths due to tuberculosis and 45,000 deaths from HIV over the next five years owing to the disruptions caused in the healthcare supply chain due to the pandemic. With 600M confirmed cases and 6M deaths, COVID-19 is still considered an ongoing pandemic today. While scientists and economists scramble to find methods to mitigate this unprecedented damage to global health and the global economy, accurate data-driven models have become the need of the hour to aid in decision-making and resource allocation by governments. Several studies have employed the use of time series models and analysis techniques to forecast COVID-19 outcomes for the same.

Time series techniques, like the usage of ARIMA (Autoregressive Integrated Moving Average) models, are used to make predictions for the future based on past observations. They do so by identifying patterns and trends in existing data. ARIMA models are known to be robust and flexible and, when correctly chosen, provide a good fit for data with high variations. ARIMA models capture both long-term and short-term trends in data, and are hence well suited for forecasting COVID-19 outcomes which are affected by immediate triggers and show long term patterns. The COVID-19 pandemic has evolved rapidly over time and has shown seasonal differences as well. The ARIMA model can thus be used to capture the dynamic nature of the pandemic, which remains invisible to the naked eye. Several studies have previously used ARIMA models to make predictions about COVID-19. A study by Zhang et al. (2020) used ARIMA models to predict trends in COVID-19 deaths in China, the epicenter of the pandemic and found satisfactory results. Another study by Kavak et al. (2020) used ARIMA models to forecast COVID-19 deaths in Turkey and reported accurate predictions as well. Our report aims to build on their research and use ARIMA models to predict COVID-19 deaths in **India** based on data from December 31, 2019, to August 1, 2020.

We believe that our report's findings can provide invaluable insights for policymakers and healthcare professionals to prepare for surges in demand for healthcare resources and medicines and to work on mitigation strategies to reduce the adverse economic effects caused by the spread of the virus. Forecasts of COVID-19 deaths can also help government officials make decisions on measures like social distancing, lockdowns, vaccination campaigns, etc., to be implemented. Overall, our report has significant importance in managing the COVID-19 pandemic in India; its results can help officials make informed decisions, thus improving public health outcomes, and it can also help the general public stay alert and aware of future possibilities.

2 Literature Review

<u>Paper 1: Predicting number of Covid19 deaths using Time Series Analysis (ARIMA MODEL)</u>

<u>Author: Navid Mashinchi</u>

About the paper:

Using time series analysis and ARIMA modelling, the article published in Towards Data Science aims to construct a predictive model for the number of COVID-19 deaths.

Methodology:

From January to August 2020, the study examined daily data on COVID-19 deaths in the United States. Based on past observations in the time series data, the ARIMA model was used to forecast the number of COVID-19 deaths. Metrics such as MAE and RMSE were used to assess the model's performance.

Key Findings:

The study discovered that the ARIMA model accurately predicted the number of COVID-19 deaths in the United States, with an MAE of 246 and an RMSE of 324. According to the analysis, COVID-19 deaths climbed quickly in March and April 2020, then declined in subsequent months.

Conclusion:

The study's outcomes imply that time series analysis and ARIMA modeling can be helpful to techniques for estimating the frequency of COVID-19 deaths. These findings can help public health officials establish measures to restrict the spread of COVID-19 and target resources to areas with higher anticipated fatality rates. However, it is essential to note that factors such as testing rate changes and control measure implementation variations may influence the model's accuracy. More research is needed to refine and improve the model's accuracy.

Paper 2: Time series analysis and forecasting of coronavirus disease in Indonesia using the ARIMA model and PROPHET

Authors: Christophorus Beneditto Aditya Satrio , William Darmawan , Bellatasya Unrica Nadia , Novita Hanafiah

About the paper:

We employed time series analysis in this study to forecast COVID-19 cases in Indonesia using two alternative models, ARIMA and Prophet. The study's goal was to give accurate estimates of COVID-19 cases and deaths in Indonesia and assess each model's effectiveness in forecasting COVID-19 cases.

Methodology:

The study analyzed daily data from March 2020 to May 2021 in Indonesia, including the number of confirmed cases, deaths, and recoveries. To forecast the daily COVID-19 cases in Indonesia, we used the ARIMA model and the Prophet model. We used statistical measures such as MAPE, root mean square error RMSE, and mean absolute error MAE to compare the performance of both models.

Key Findings:

The ARIMA and Prophet models performed well in projecting COVID-19 instances in Indonesia, according to the findings of our study. However, the Prophet model outperformed the ARIMA model regarding accuracy and prediction error. The Prophet model had a MAPE of 5.61%, whereas the ARIMA model had a MAPE of 7.92%.

Furthermore, the Prophet model provided a more detailed analysis of the trend and seasonality of COVID-19 cases in Indonesia, indicating a drop in patients following the peak in January 2021. The model also revealed that the daily number of incidents climbed considerably in May 2021.

Conclusion:

The study found that the Prophet model was more accurate than the ARIMA model in projecting COVID-19 instances in Indonesia. The PROPHET model also provides a more extensive examination of Indonesia's COVID-19 trend and seasonality.

Paper 3: Time series analysis and prediction COVID-19 affected patients by ARIMA model using machine learning

Authors: Fuad Ahmed Chyon,* Md. Nazmul Hasan Suman, Md. Rafiul Islam Fahim, and Md. Sazol Ahmmed

About the Paper:

The article aimed to create a predictive model for the number of COVID-19 cases using time series analysis and ARIMA modeling.

Methodology:

From March to September 2020, the study used daily data on the number of COVID-19 cases in Bangladesh. Based on past observations in the time series data, the ARIMA model was used to forecast the number of COVID-19 cases. Machine learning approaches such as Decision Trees, Random Forests, and Artificial Neural Networks were also used to evaluate their predictive accuracy with the ARIMA model.

Key findings:

The ARIMA model could adequately predict the number of COVID-19 cases in Bangladesh, with an MAE of 1461 and an RMSE of 1908. According to the study, COVID-19 patients grew dramatically in June and July 2020, then declined in subsequent months. Furthermore, the ARIMA model outperformed the Decision Tree, Random Forest, and Artificial Neural Network models in terms of predictive performance, according to the study.

Conclusion

The study's outcomes imply that time series analysis and ARIMA modeling can be helpful to techniques for estimating the frequency of COVID-19 cases. The comparison with machine learning models emphasizes the significance of adopting proper modeling techniques that are appropriate for the nature of the data.

Paper 4: COVID-19 Time Series Forecasting of Daily Cases, Deaths Caused and Recovered Cases using Long Short Term Memory Networks

Authors: Suraj Bodapati; Harika Bandarupally; M Trupthi

About the Paper:

The paper published in the IEEE Access journal presents a study on time series forecasting of COVID-19 daily cases, deaths caused, and recovered patients using Long Short Term Memory (LSTM) networks.

Methodology:

To train and assess their proposed methodology, the authors used publicly available datasets of daily COVID-19 cases, fatalities, and recovered cases from several nations. After preprocessing the dataset, the features were retrieved using the sliding window technique. The extracted features were then used to train LSTM networks, and the models' performance was assessed using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).

Key Findings:

The study discovered that LSTM networks could accurately estimate the daily cases, fatalities, and recovered cases of COVID-19. When the proposed methodology's performance was compared to existing state-of-the-art methods, it was discovered that LSTM networks beat them in terms of accuracy and efficiency. The study also demonstrated that the predicted findings might be utilized to help decision-makers control and manage the epidemic.

Conclusion:

The article concludes that LSTM networks effectively anticipate daily COVID-19 cases, fatalities, and recovered cases. The proposed methodology can be utilized to generate reliable and accurate projections to help decision-makers control and manage the epidemic. The study also emphasizes the significance of applying modern machine-learning techniques to analyze and predict the development of infectious illnesses.

Paper 5: Time series forecasting of new cases and new deaths rate for COVID-19 using deep learning methods

Authors: Nooshin Ayoobi, Danial Sharifrazi, Roohallah Alizadehsani, Afshin Shoeibi, Juan M. Gorriz, Hossein Moosaei, Abbas Khosravi, Saeid Nahavandi, Abdoulmohammad Gholamzadeh Chofreh, Feybi Ariani Goni, Jiří Jaromír Klemeš, Amir Mosavi

About the Paper:

The study "Time series forecasting of new cases and new deaths rate for COVID-19 using deep learning methods" was published in the journal PLOS ONE.

Methodology:

To train and assess their proposed methodology, the authors used publicly accessible datasets of new COVID-19 cases and new death rates from several nations. The MACD and RSI indicators were used to extract features from the preprocessed dataset. Deep learning approaches such as LSTM, GRU, and CNN were then used to forecast the number of new cases and deaths.

Key Findings:

The study discovered that deep learning approaches such as LSTM, GRU, and CNN were highly accurate in estimating new cases and fatality rates for COVID-19. When the proposed methodology's performance was compared to that of other cutting-edge methods, it was discovered that deep learning methods beat them in terms of accuracy and efficiency. The study also demonstrated that the predicted findings might be utilized to help decision-makers control and manage the epidemic.

Conclusion:

The paper concludes that deep learning methods such as LSTM, GRU, and CNN effectively forecast new COVID-19 cases and deaths. The proposed methodology can be utilized to generate reliable and accurate projections to help decision-makers control and manage the epidemic.

<u>Paper 6: COVID-19 Pandemic Prediction using Time Series Forecasting Models</u>

<u>Authors: Naresh Kumar; Seba Susan</u>

About the Paper:

The purpose of this study was to forecast the spread of the COVID-19 pandemic in different regions of India using time series forecasting algorithms. ARIMA, SARIMA, and LSTM machine learning algorithms were used to forecast the number of cases for September 2020. According to the study, the LSTM model outperformed the ARIMA and SARIMA models in predicting COVID-19 instances. The study suggested using LSTM models to forecast the spread of the pandemic.

Methodology:

From January to August 2020, the study used daily COVID-19 case data from several areas in India. ARIMA, SARIMA, and LSTM time series forecasting models were used to preprocess and analyze the data. The models were trained on data from January to August 2020 and used to forecast the number of COVID-19 cases for September 2020. The performance of the models was assessed using the MAPE and RMSE.

Key Findings:

According to the study, the LSTM model outperformed the ARIMA and SARIMA models in predicting COVID-19 instances. The MAPE and RMSE of the LSTM model were 7.67% and 2821.23, respectively, lower than those of the ARIMA and SARIMA models. The study also discovered that COVID-19 instances in India strongly depended on cases reported the day before. The LSTM model projected a significant increase in COVID-19 cases in various Indian locations, later validated by data.

Conclusion:

The study suggests that time series forecasting models can be used to anticipate the spread of the COVID-19 pandemic and that LSTM models outperform standard forecasting models. The paper recommended that policymakers and health professionals utilize LSTM models to predict the pandemic's future spread and to plan and allocate resources accordingly.

Paper 7: Prediction of confirmed cases of and deaths caused by COVID-19 in Chile through time series techniques: A comparative study

Authors: Claudia Barría-Sandoval, Guillermo Ferreira, Katherine Benz-Parra, Pablo López-Flores

About the Paper:

The study examines different time series forecasting techniques in estimating the number of confirmed cases and deaths caused by COVID-19 in Chile. The study was carried out during the COVID-19 pandemic, which afflicted millions worldwide, to provide precise predictions to assist the Chilean government in making educated decisions.

Methodology:

From March 2020 to December 2020, the researchers examined daily data on verified COVID-19 infections and deaths in Chile. The information was gathered from the official website of Chile's Ministry of Health. The performance of five-time series forecasting algorithms was examined in the study: ARIMA, SARIMA, ETS, TBATS, and Facebook's Prophet. Each technique's performance was assessed using three metrics: MAE, MSE, and RMSE.

Key Findings:

The study discovered that Facebook's Prophet and TBATS beat the other strategies in estimating the number of confirmed COVID-19 infections and deaths in Chile. Prophet had the lowest MAE and MSE values for both confirmed cases and deaths, whereas TBATS had the lowest RMSE values. The study also discovered that as the prediction horizon rose, the accuracy of the predictions reduced, showing that shorter-term predictions were more accurate.

Conclusion:

The study revealed that time series forecasting techniques can reliably predict the number of confirmed COVID-19 cases and deaths in Chile. Prophet and TBATS were discovered to be the most accurate techniques, with prediction accuracy decreasing as the forecast horizon grew. The study's findings can help the Chilean government make informed judgments and conduct suitable COVID-19 control measures.

3 Data and Methodology

Our data is collected from the Our World in Data website. We created a datasheet that consists of 3 variables, dates from 31st December 2019 to 1st August 2020, total deaths, and new deaths in India.

- Date: Displays the dates that will be used in our analysis.
- total_deaths: Indicates the overall number of Covid-19 deaths in India.
- new_deaths: Indicates how many Covid-19 deaths occur in India each day.

Our aim was to predict the total number of deaths due to Covid-19 in India from August 1st to August 21st, 2020, and from August 1st to November 1st, 2020. We are going to use time series analysis to predict the deaths. For this prediction, we are going to use the ARIMA model.

ARIMA model:

The Autoregressive Integrated Moving Average (ARIMA) model is used for the analysis of time series data and statistical forecasting. ARIMA analysis tells us the strength of the dependent variable when compared to other variables. In this model, the forecasting of future time series movement is determined by its own past values, past values of residuals, and the difference between consecutive observations and actual values.

The following is what the ARIMA model's parameters signify:
"p": Abbreviation for auto-regressive. When past data is utilized to forecast future data, this process is known as autoregression.
"d": Integrates, as indicated. Integration accounts for how much differencing will be applied to the time series.
"q": The abbreviation for moving average. When you take the various intervals, you average them to create the moving average.

Tests used to check the model's validity:

- **<u>Augmented Dickey-Fuller Test :</u>**

  The Augmented Dickey-Fuller (ADF) test is a statistical hypothesis test that is typically utilized to test for the presence of a unit root in time series data. A unit root indicates that the stochastic trend of a time series does not converge to a stable long-run mean. In other terms, a unit root indicates a non-stationary time series.

  The Dickey-Fuller (DF) test regresses the first difference of a time series on its lagged values to find a unit root. The ADF test builds on this. The ADF test extends the DF test by adding more lagged differences to the regression equation, allowing for more complex time series data.

  The ADF test generates a test statistic and a p-value, which are used to test the null hypothesis that a time series has a unit root. The null hypothesis is rejected, and the time series is considered stationary if the p-value is below the significance level (typically 0.05).

- **<u>Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) Plots:</u>**

  These plots help in determining the ARIMA model's order. The ACF plot displays a time series correlation with its lagged values. It displays the series' correlation coefficients up to a particular number of lags. The ACF plot helps determine the ARIMA model's Moving Average (MA) order. The model needs an MA(k) term if a time series has significant autocorrelation at lag k. After removing intermediate lags, the PACF plot shows the correlation between the time series and its lagged values. The PACF plot helps determine the ARIMA model's Autoregressive (AR) component order. An AR(k) term may be needed in the model if a time series has considerable partial autocorrelation at lag

k. The number of times the time series must be differenced to become stationary determines the order of the integrated component (I).

- **Ljung-Box test:**

  The Ljung-Box test is a statistical procedure is used to check for the presence of serial correlation, also referred to as autocorrelation, in a given dataset. The autocorrelation of a time series is a measure of how closely related successive observations are. Autocorrelation in a time series indicates a correlation between the present value and one or multiple past values.The Box-Ljung test tests the null hypothesis that data has no autocorrelation against the alternative hypothesis that it does. It compares the total of squared autocorrelations at various lags to a theoretical distribution that assumes no autocorrelation. If the sum of squared autocorrelations differs sufficiently from the null hypothesis, the null hypothesis is rejected in favour of the alternative hypothesis that there is some autocorrelation.

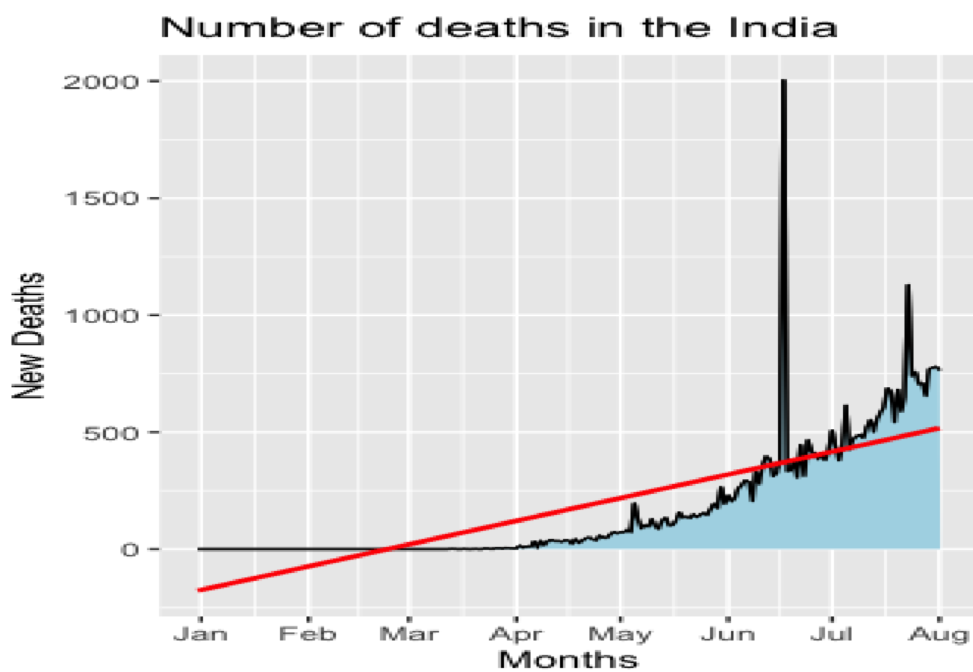
4 Results and Discussion

```
> summary(dat_demo)
     Index               dat_demo     
 Min.   :2019-12-31   Min.   :   0.0  
 1st Qu.:2020-02-22   1st Qu.:   0.0  
 Median :2020-04-16   Median :  36.0  
 Mean   :2020-04-16   Mean   : 169.8  
 3rd Qu.:2020-06-08   3rd Qu.: 290.5  
 Max.   :2020-08-01   Max.   :2003.0  
```

Our time series index begins on December 31, 2019, and it ends on August 1, 2020, according to the summary of our data. Additionally, we can see that India experiences 169.8 deaths on average each day. The highest death toll was recorded in 2003.

Time Series Analysis:

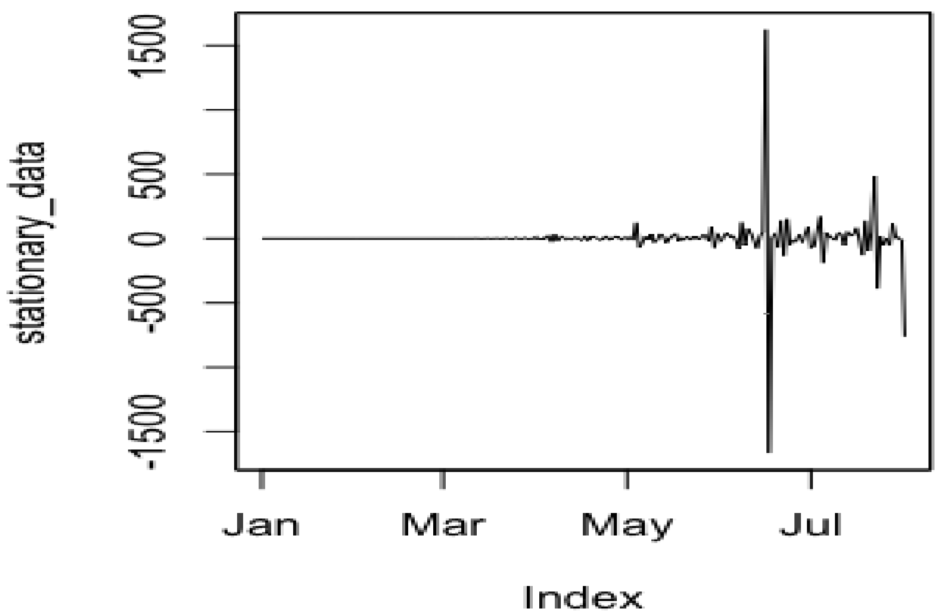
We used the zoo library, a well-known R package for this particular statistical technique, to create the date sequence before performing the time series analysis on our data. Then we determine if any time series analysis assumptions are violated by our data. The data must be stationary, so we must ensure this. In other words, the mean, variance, and covariance must be the same for all time intervals.



The mean line is rising, as shown by the red line. The fact that the distance between the data points and the mean line varies along the x axis allows us to see that variance is not constant as well. To make our data stationary, we must therefore transform it. There are various methods to make data stationary depending on the data.

For time series analysis, variance can be stabilised using data transformations like logarithms. The mean can be stabilised, however, with the aid of differencing. Every dataset has specific requirements, and for our data, we only used differencing.

Differencing is the process of calculating the variation between eight consecutive observations. Our data's graph changed once we applied differencing, as shown below:



Here, we can see that the mean has become constant as a result of the data transformation. We also carried out the Augmented Dickey-Fuller Test (ADF test), which determines whether or not our data is stationary, to see if it has become stationary.

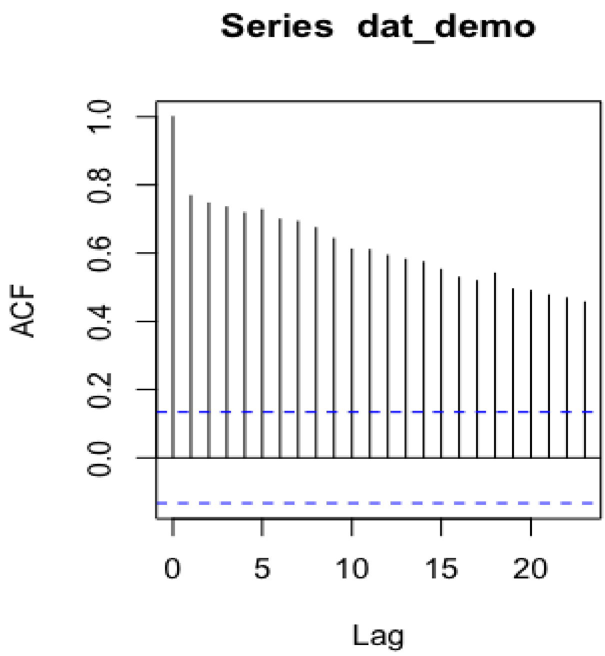
H_0: The absence of a unit root is the null hypothesis.
H_A: The time series might be stationary as an alternative theory.

```
> adf.test(as.matrix(stationary_data))

        Augmented Dickey-Fuller Test

data:  as.matrix(stationary_data)
Dickey-Fuller = -9.6619, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

On a significance level of 0.05, it is clear that the p-value for our test is 0.01 (or less). As a result, we discounted the null hypothesis. To put it another way, we can now continue the analysis because our data is stationary.

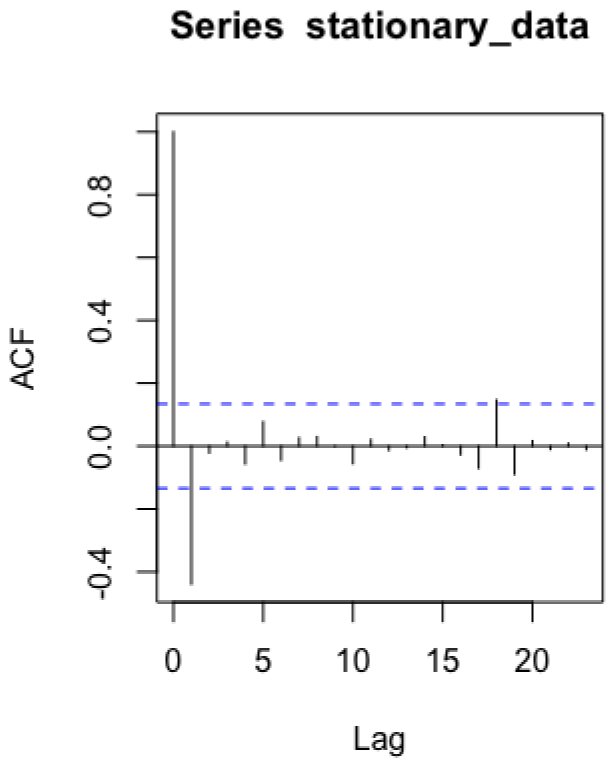
The ACF and PACF functions, which stand for autocorrelation function and partial autocorrelation function, respectively, determine how we choose our p, d, and q values.

The ACF graph for non-stationary data is displayed as follows.

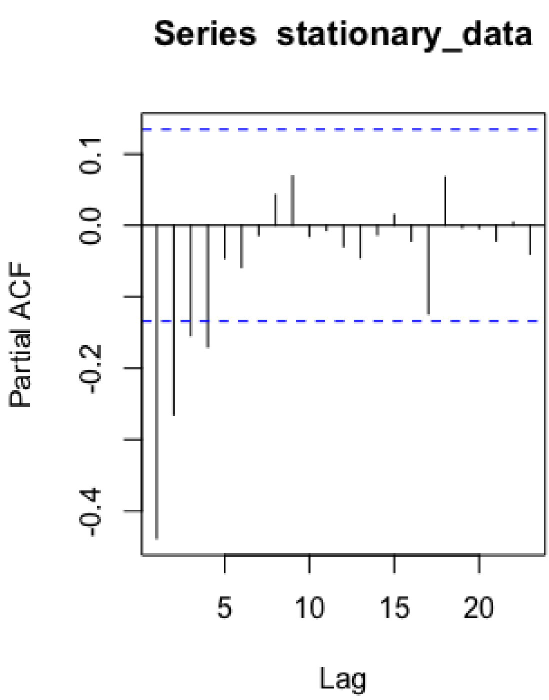


Here, we can see that all of our values are above the blue line. The values must be inverted and below the blue line to achieve the desired result.

Now, we can see what our converted stationary data's ACF and PACF are by looking at them.



Series stationary_data
ACF
0.8
0.4
0.0
-0.4
0
5
10
15
20
Lag



Series stationary_data
Partial ACF
0.1
0.0
-0.2
-0.4
5
10
15
20
Lag

Now that we have ACF and PCF graphs, we can see that the majority of the lines have inverted lines as well as not exceeding the blue line.
Using the acf graph, the q value can be chosen, and the pacf graph can be used to determine the p value. In both scenarios, we select the digit that comes before the first inverted line. In accordance with the data, you can apply that rule and assess how well the model fits, or you can apply the auto.arima function, that returns the top model based on AIC or BIC.

Our group made the decision to use the auto.arima function, which gave us the model shown below.

```
Series: dat_demo
ARIMA(0,1,1) with drift

Coefficients:
          ma1   drift
      -0.8610  3.0921
s.e.   0.0344  1.3516

sigma^2 = 19133:  log likelihood = -1358.26
AIC=2722.51   AICc=2722.62   BIC=2732.61
```

The auto.arima function selected p=0, d=1 and q=1.

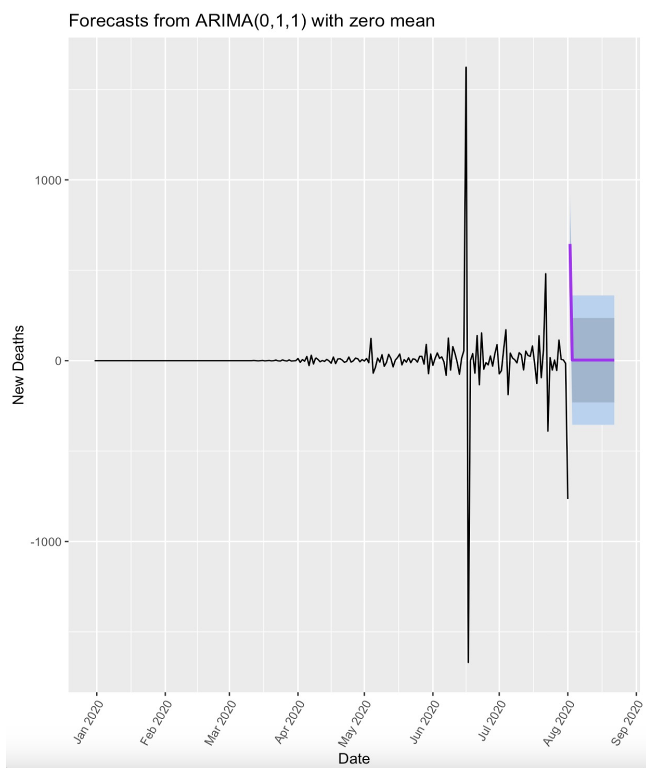
We apply this model to predict the number of deaths for the next 21 and 90 days according to previous time series data and it has been found to be very close to the actual death count data released by WHO.

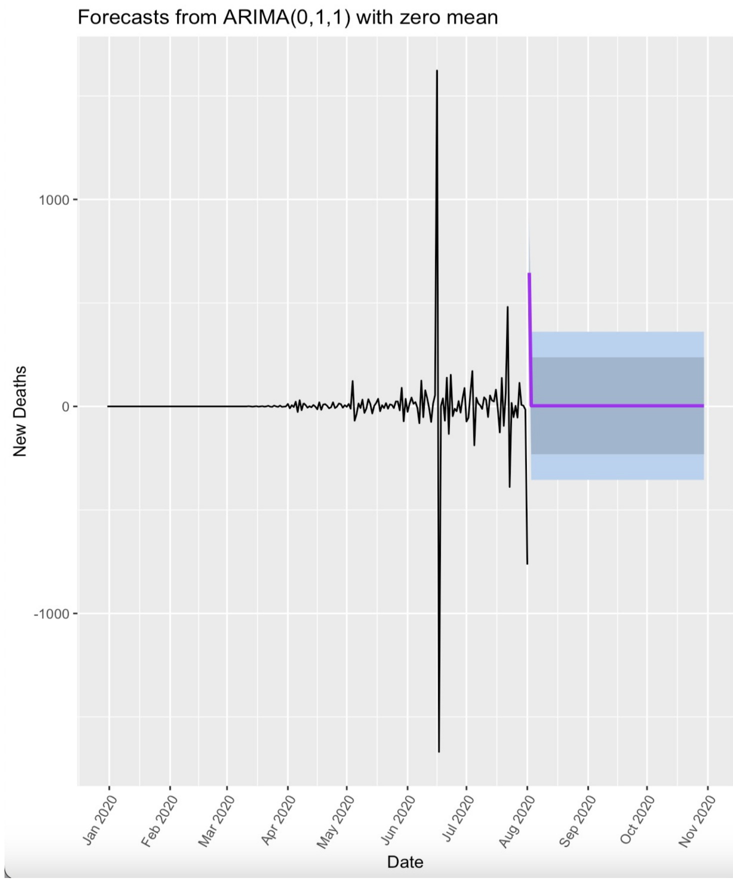
From R we have found out that the number of deaths predicted by our model for the next 21 days and 90 days are 20419((Total cumulative Deaths 56930) and 103347 (Total cumulative Deaths 139958) deaths respectively.

On the other hand, according to WHO actual death count for the next 21 days and 90 days is 19000 (total deaths 55794) and 87000(total deaths 123000).

Date:	Our projection	Actual data(WHO)
2020/08/01 – 2020/08/21	New Deaths: 20,419	New Deaths: 19,000
2020/08/01 – 2020/10/31	Total Deaths: 139958	Total Deaths: 123000

Below we can see the graphs for the forecast:





The purpose of the project was to compare the performance of our model to the actual death statistics provided by the WHO.

In order to test for the existence of autocorrelation in financial data, the Box-Ljung test is frequently used in time series analysis, especially in econometrics and finance; as a result, we carry out the aforementioned analysis on our model and verify the results.

A common technique for evaluating the ARIMA model's suitability is the Ljung-Box test, which looks at the residuals' autocorrelation. The residuals should be in line with being white noise, which is indicated by a high p-value.

```
> LBtest

	Box-Ljung test

data:  res
X-squared = 9.378, df = 20, p-value = 0.9781
```

The aforementioned test results indicate that the fitted ARIMA model has a substantial p-value of 0.9781, indicating that there is no significant autocorrelation among the residuals, indicating that the model is a good fit for the data.

To make sure the model is capturing all the information present in the data, it is crucial to make sure the residuals of the model are white noise when fitting an ARIMA model (i.e., evenly and independently distributed with a mean of zero and constant variance).

Note that the Ljung-Box test results alone should not be used to determine whether an ARIMA model is adequate. To confirm that the ARIMA model is an appropriate fit for the data, additional diagnostic checks should be made, such as examining the residuals' ACF and PACF, analysing their distribution, and looking for structural breaks or outliers.

5 Conclusion

This study tries to forecast the number of deaths due to covid-19 in India using time series data from December 31, 2019, to August 1, 2020. The dataset lists 214 observations and 3 independent variables, the date, total deaths up to the date, and new deaths on that date.

We employed the first differencing method to make our non-stationary dataset stationary in order to apply the ideas of the ARIMA model to our analysis. We discovered that India is bound to see 103,347 new deaths in the next 90 days duration, which is around 74% of the total number of fatalities up to that point. This prediction was made using the cleaned and updated dataset and the auto.arima function. Despite being slightly different, our results were close enough to the WHO's prediction.

The findings from this paper, as discussed in previous sections, can have significant policy implications and stir the minds of public health officials to act upon the impending danger that covid-19 poses to the country. Resource allocation towards catering to hospital needs, organizing sanitation awareness programs, and setting vaccination drives should be the priority of the concerned authorities.

The future scope of this research can be including socioeconomic factors such as age distribution, comorbidities, etc., to the model so that more targeted interventions can be made to reduce the number of deaths due to covid-19.

6 References

1. Mashinchi, N. (2020, December 29). Predicting number of covid19 deaths using time series analysis (Arima model). Medium. Retrieved May 4, 2023, from https://towardsdatascience.com/predicting-number-of-covid19-deaths-using-time-series-analysis-arima-model-4ad92c48b3ae
2. Christophorus Beneditto Aditya Satrio a et al. (2021) Time series analysis and forecasting of coronavirus disease in Indonesia using Arima model and prophet, Procedia Computer Science. Elsevier. Available at: https://www.sciencedirect.com/science/article/pii/S1877050921000417 (Accessed: May 4, 2023).
3. Chyon, F. A., Suman, M. N. H., Fahim, M. R. I., & Ahmmed, M. S. (2022). Time series analysis and predicting COVID-19 affected patients by ARIMA model using machine learning. Journal of virological methods, 301, 114433. https://doi.org/10.1016/j.jviromet.2021.114433
4. S. Bodapati, H. Bandarupally and M. Trupthi, "COVID-19 Time Series Forecasting of Daily Cases, Deaths Caused and Recovered Cases using Long Short Term Memory Networks," 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2020, pp. 525-530, doi: 10.1109/ICCCA49541.2020.9250863.
5. Ayoobi, N., Sharifrazi, D., Alizadehsani, R., Shoeibi, A., Gorriz, J. M., Moosaei, H., Khosravi, A., Nahavandi, S., Gholamzadeh Chofreh, A., Goni, F. A., Klemeš, J. J., & Mosavi, A. (2021). Time series forecasting of new cases and new deaths rate for COVID-19 using deep learning methods. Results in Physics, 27, 104495. https://doi.org/10.1016/j.rinp.2021.104495
6. N. Kumar and S. Susan, "COVID-19 Pandemic Prediction using Time Series Forecasting Models," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-7, doi: 10.1109/ICCCNT49239.2020.9225319.
7. Barría-Sandoval, C., Ferreira, G., Benz-Parra, K., & López-Flores, P. (2021). Prediction of confirmed cases of and deaths caused by COVID-19 in Chile through time series

techniques: A comparative study. PLOS ONE, 16(4), e0245414. https://doi.org/10.1371/journal.pone.0245414
8. https://ourworldindata.org/covid-deaths