International Journal of Early Childhood Special Education (INT-JECSE)
DOI:10.9756/INTJECSE/V14I6.316 ISSN: 1308-5581 Vol 14, Issue 06 2022

Prediction of COVID-19 Very critically ill (Equivalent to Deaths of the patients) in the United States of America using ARIMA Model in R programming

1st Sahoo Kalyan Kumar, 2nd R.Venkat Munni Reddy& 3rd Vaibhava Patil

Professor, School of Business, The Management University of Africa,Nairobi, Kenya
Professor,Manipal Academy of Higher Education, Karnatak, India.
Asst. Professor,International Institute of Management Sciences,Pune, India
* *Corresponding author:*R.Venkat Munni Reddy

Abstract: The worldwide epidemic is COVID-19. This illness originally spread in Wuhan, Hubei's capital. First instance traced to 55-year-old on November 17, 2019. The sickness spread globally. North Seattle had the first COVID case in January 2020. In April, the number of patients peaked, causing about 4000 deaths. This study attempts to anticipate the number of very seriously sick in the country by October and compare it to the country's predictions. Cleaning raw data into rows and columns helps in analysis. This project uses time-series analysis with the statistical and economic ARIMA model. ARIMA may be used to interpret and forecast time-series data. This model works well with non-stationary data. The model will take at least 21 days to estimate new extremely severely unwell. ARIMA is accomplished using the statistical programming language R. The interactive R shiny application visualizes the ARIMA model. This approach may be applied in different countries to estimate illness mortality. This model helps project the desired number of days by inputting the value in R shiny using data science.

Keywords: Covid19, ARIMA Model, R Programming, Data Science

1. Introduction

Corona Virus (2019) COVID-19 is one of the contagious diseases which occurs with respiratory and vascular problems and is also known as SARS-CoV-2. This was first spotted in Wuhan, China. This coronavirus is now widespread across the countries. Common symptoms for this virus include cough, fever, breathing problems, with lack of taste and smell. COVID-19 spreads from one person to others through the respiratory tract when an infected person sneezes, sings, or coughs. A new infection occurs when particles having virus are exhaled by an affected person through droplets from respiration. When these droplets prolong in the air for long-duration, this becomes an airborne transmission. The estimation of people getting infected from an affected person was four till September. The people getting infected can be through having symptoms or without symptoms. The WHO has published a standard testing procedure called the SWAB or nasopharyngeal swab test as a standard testing procedure for confirming the disease. There are various preventive measures suggested by WHO which includes social distancing, wearing a mask at public, rubbing hands with sanitizer washing hands with soap for 20 seconds and ventilating the indoors. Currently, the world has 1,225,913 Very critically ill (Equivalent to Deaths of the patients)s with more and more active cases daily.

In the United States of America currently, there are 9,802,274 Coronavirus cases are registered with 239,842 Very critically ill (Equivalent to Deaths of the patients)s in total with New York facing the highest Very critically ill (Equivalent to Deaths of the patients) among them. COVID-19 has dramatically reduced the economy of the country by falling to -32.9%. The fact suggests that COVID-19 will affect 70 million people in the USA.

The COVID-19 outspread has made a significant downfall to the US economy since the Great depression. With almost 30.2 million Americans are unemployed after the outbreak of the virus. The USA is one among the countries in the world which are severely affected by the Coronavirus following India in second and Brazil third.

Objectives

This project attempts to predict the Very critically ill (Equivalent to Deaths of the patients)s in the USA in the upcoming days. The data is taken from "Our World in Data" (https://ourworldindata.org/covid-deaths). From the data set, the variables relevant to the project is collected in the spreadsheet. The variable which will be predicting the Very critically ill (Equivalent to Deaths of the patients)s is "Date", "Location", "Total Very critically ills (Equivalent to Deaths of the patients)", " Very critically ill (Equivalent to Deaths of the patients)s". Time series analysis is performed using the variables from

the spreadsheet. The model chosen for time series analysis is ARIMA (Auto-Regressive Integrated Moving Average model). ARIMA model involves three parts, namely, Auto-Regressive part (AR), the order of integration (I), Moving Average (MA).

The R programming software is the tool through which the time series analysis is done to predict and forecast the number of Very critically ill (Equivalent to Deaths of the patients)s in America.

R shiny a visualization tool available in R software which is used for visualizing the data with dynamic interactions to enter the number of days for which the model will predict the number of Very critically ill (Equivalent to Deaths of the patients).

Scope of research

The scope of the project is to predict the Very critically ill (Equivalent to Deaths of the patients) due to COVID-19 in the USA. This project can be extended for an overall view for other countries affected by this disease and dynamic visualisation of all the countries total Very critically ill (Equivalent to Deaths of the patients) and expected new cases could be interactively designed also with various other techniques. This project also lends an upper hand in performing the time series using different other techniques which prominently allows to forecast and predict the values for the future.

Assumptions

This project is carried out by assuming only three variables from the entire data set of the "Our World in Data" and the data is scraped to the need. The data index starts from January 2020, and the information which is loaded in the dataset is assumed to be reliable for the analysis.

Limitations

There are few limitations where the project cannot be extended to as the frequency of the data is limited, which restricts the data to give decomposition results. The data is presumed to have the correct information from the source. The prediction results are based on the inferred data, which is taken as reliable.

"Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using prediction model under current public health interventions" Issued in the year 2020. In this paper, Saleh l. Alzahrani, Ibrahim A. Aljamaan and Ebrahim A. Al-Fakih state the world is seeing an alarming coronavirus increase and how bad the entire pandemic is affected the world nations. The ARIMA model has been implemented to fore view the daily cases of Coronavirus in Saudi Arabia for four weeks. Results suggest that cases will continue to grow and 7668 cases per day are expected, and in a total of 127,129 cases for four weeks. This paper also predicts that across 2 million people around the globe will miss the holy pilgrimage of Mecca and Medina.

"ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India" Issued In the year 2020. Farhan Mohammad Khan and Rajiv Gupta COVID-19 cases are rising in number with more and more fatalities increasing. The test data is taken from January, and it is trained with the data available from April. A NAR network is built to compare the precision of the model, which is predicted. The source is taken from Health, and Family ministry welfare has been studied for 50 days, and it projects that there will be approximately 1500 cases per day will be registered as of April 04, 2020. The model with ARIMA (1,1,0) is selected with the help of BIC values.

"Short term forecasting COVID-19 cumulative confirmed cases: Perspectives of Brazil" Issued in June 2020. Matheus Henrique Dal Molin Riberio, Ramon Gomes De Silva, Viviana Cocco Mariani and Leandro dos Santos Coelho state that, developing an ephemeral forecasting models will allow forecasting the number cases in the future. Also, to create a strategic plan to avoid very critically ill (Equivalent to Deaths of the patients). The prediction with various models like ARIMA, random forest and support vector regression and their error percentage is known, and they are ranked in the order of SVR, ARIMA and random forest. With this ranking order, the future prediction for COVID cases can be monitored.

"Prediction of the COVID-19 Pandemic for the top 15 affected countries: Advanced Autoregressive Integrated Moving Average (ARIMA) model". Issued in April 2020. Ram Kumar Singh, Meenu Rani, Akshaya Srikanth Bhagabathula, Ranjit Sah et.al., state that coronavirus pandemic has affected more than 200 countries and more than 2,800,000 people were affected as of April 24, 2020. This study aims the top 15 countries of the confirmed cases through spatial mapping. A comparative model has been done for 15 countries with the parameters like confirmed infections, fatalities, and recoveries and an ARIMA model. This study predicts the trajectories for two months of the spread of Coronavirus.

"Estimation of COVID-19 prevalence in Italy, Spain and France" Issued in August 2020. Zeynep Ceylan states that there is a need to observe and predict the Coronavirus to stop the spread. ARIMA models are formulated, and the one with the lowest MAPE is selected. This study suggests that ARIMA is the best suitable model for predicting COVID-19. This analysis tells the regions in Italy, Spain, and France to be more cautious.

"Application of the ARIMA model on the COVID-19 epidemic dataset" Issued in April 2020. Domenico Benvenuto, Marta Giovanetti, et. al., states that with various models are available in the probable evolution of this epidemic. The ARIMA model is performed considering multiple factors and analyses and the time series analysis is performed in real-time.

The review clearly says that the COVID-19 pandemic is the worst outburst the globe has seen. The fatalities and the cases are increasing day by day in huge number. The vaccine for the virus is still in process, and it should be rolled out to stop the spread of the virus. The reviews also suggest that the ARIMA model is one of the significant models that can be used to predict the uni-variate time series using any mode of implementation. The less rate and accurate prediction make the ARIMA model as the best among the different time series models available.

2. Methodology

Data Collection

The data has been collected from Our World in Data website, which in real-time updates, the fatalities, and the new cases of COVID-19 for all the countries in the world. The COIVD-19 data point is secondary data from which the data of the USA has been scrapped for the project. The data contains 34,033 rows and 34 columns, from this only four variables, are considered for the analysis that is "Date", "Location", "Total_Very critically ill (Equivalent to Deaths of the patients)s", "New_Very critically ill (Equivalent to Deaths of the patients)s". The data is limited till August 01, 2020. This is taken as the test data, and from here the information has been predicted for 21 days using the ARIMA model in R programming.

Tools and techniques used

- R Programming Software
- Microsoft Excel
- Time Series modelling (ARIMA model)
- R shiny application

The secondary data is collected from Our world in data, and the information is cleansed to four variables for the analysis and put in an excel file. The data is taken from January 01, 2020, to August 01, 2020, for the study. The missing values in the data were cleaned and filled. This data is then imported to R programming software where the time series modelling using ARIMA is performed. The data has been converted to "date" format to convert the data to time series.

Data Definitions:

Date- It represents the date which is used in the analysis. The class of the data initially was a factor which is converted to date class for performing time series analysis.

Location- It represents the name of the country. This variable is a factor.

Total Very critically ill (Equivalent to Deaths of the patients)- It represents the total number of Very critically ill (Equivalent to Deaths of the patients)s in the USA due to COVID-19. The class of this variable is numeric.

New Very critically ill (Equivalent to Deaths of the patients)- It represents the number of COVID-19 Very critically ill (Equivalent to Deaths of the patients)s in the USA per day. The class of this variable is numeric.

These variables are imported to R software, and the basic descriptive statistics are observed followed by plotting the trend of the data over the period to see the mean and variance. The data is then converted to time series, and the ARIMA modelling is done to forecast the Very critically ill (Equivalent to Deaths of the patients) in the USA.

There are various libraries used in R programming like,

Ggplot () – It is a data visualization package for the statistical language R. ggplot is an implementation from the Grammar of Graphics by Leland Wilkinson's which says graphics are the primary form of data visualization which breaks up graphs into schematic components.

Tseries ()- It is abbreviated as time series analysis and computational finance. It is used to compute the time series data and perform the various test as needed for the computation.

Forecast () – It provides method and tools for displaying and analyzing univariate time series forecast for various models, including the ARIMA model.

Shiny ()- It is an interactive web application with R. It automatically binds the input and the output. It has a wide range of prebuilt widgets to make beautiful, reactive, responsive, and robust applications.

The converted time series data is decomposed to find the seasonality and the trend of the data. Still, considering the frequency parameters, the data has very few periods to complete the frequency cycle, so the Holts Winter smoothing is used to see the plot of actual versus the fitted plot.

The data is a random walk model, and the unit root is performed to observe the stationarity of the dataset. The unit root test is a measure of stochasticity. The stationarity can be dealt with various methods like Dicky Fuller test, Augmented Dicky Fuller test and KPSS statistics. The Augmented

Dicky fuller test tests the null hypothesis and suggest that there is a unit root in the data. The alternative theory suggests the data be stationary. The corresponding p-value results in 0.6555, which is more than 0.05 indicate that the information is not stationary.

Generally, the time series with the seasonality or a trend will affect the values of the time-series data. There are two ways to make time-series data stationary. One is by differencing and another by transforming. For the univariate time series analysis, the data is made stationary by differencing to first or to the second differencing. The information is differenced once, and the augmented dicky fuller test is observed. The p-value after first differencing is 0.01, which is very much less than 0.05, which makes the data to be stationary by accepting the alternative hypothesis.

The Auto-correlation and Partial auto-correlation plots are observed. The ACF plots tell how well the present values are related to the past values, and the PACF plots are used to see the correlation with the residuals.

The time-series data is then modeled using ARIMA with best (p,d,q) value which is determined by the lowest Akaike Information Criteria (AIC). The best fitted ARIMA is obtained, and then data is predicted to the number of days as required to report the new Very critically ill (Equivalent to Deaths of the patients) in the USA. The modelled ARIMA model is dynamically visualized through R shiny application which is bind with input and output parameters, and the model is viewed through the R server.

Time series analysis

Time series can be defined as a sequence of values of which the variable is spaced equally in the time intervals. Also, time-series can be said as a set of data points with timely ordered observations of the quantitative nature of a collective phenomenon for the given period. The usage of these models is to understand the and structure that produces the observed data and to fit a model with a notion of forecasting, monitoring or even for feed forward or feedback controls.

Time-series analysis is used in many fields of application such as.

- Prediction of the economic data
- Sales forecasting
- Stock market analytic
- Analysis on the budget
- Yield Projection
- Census analysis

There are numerous techniques to model and forecast the time series data, and fitting of these data points is done with the help of various time series models. The method which is used to predict the Very critically ill (Equivalent to Deaths of the patients) in the USA due to COVID-19 is done with the help of Box-Jenkins ARIMA model.

Stationarity

A time-series is a sequence of finite values taken at successively spaced points in time. One of the common approaches in the analysis of time-series data is to take the observed data points as part of the realization of a random process. There are two definitions before defining random processes,

- Probability of space
- Random Variable

With defining stochastic processes, stationarity can be said as the statistical process which does not change over time. Stationarity of any kind is property of a random process, and it can be a finite or infinite realization of the values.

The primary classification of stationarity can be divided into weak stationarity and strong stationarity,

Strong stationarity: It requires the regular shift of the finite-values distribution involved in a random process. The distribution of a finite sequence of random variables will remain the same as we shift it along with the time index.

Weak stationarity: It requires the regular shift in time of the first significance and the auto covariance. In this process, data has a constant mean, and the covariance between any two points is t and t-k.

Unit Root Test

A unit root is also known as a difference stationary or a unit root process. If a random trend is found in a time-series, then it is called a random walk. The time-series data with a trend depicts unit root is present, then data becomes unpredictable, showing no systematic observations.

The unit root is a test to check the stationarity in time-series data. Time-series stationarity states that if there is a shift in time distribution will not change, and unit-roots are the reason for non-stationarity.

These tests have low statistical power. Few of the standard test which involves in determining the stationarity is Dicky Fuller Test, Augmented Dicky Fuller test, KPSS statistics.

Augmented Dicky Fuller Test (ADF):

The augmented dicky fuller test is a test for stationarity and predominant in the unit root. The ADF is computed using serial correlation. The ADF test is more potent than Dicky Fuller test as it can perform more complex models.

Hypotheses:

The hypotheses which are taken for the ADF test is

H0: The null hypothesis test tells the data has a unit root.

H1: Alternative hypothesis tells the data is stationary.

In general, if the p-value is less than 5%, the null hypothesis can be rejected.

3. ARIMA Model

ARIMA modelling is an approach to model the ARIMA process. This model uses the late time-series observations and an error term to forecast the upcoming values. This model involves the auto-regressive model AR(p) and moving average model MA(q). The AR(q) takes the preceding values of the regress and to make predictions. MA(q) takes the mean and errors of the past to make the predictions. This method was suggested by Box and Jenkins in 1970, and the prediction procedures were given by (Hyndman et al., 2008).

The necessary steps involved in ARIMA modelling are,

- The data is transformed to stabilize the variance and differencing to remove the remaining seasonality.
- The process to fit the data is selected.
- The model coefficients provide the best fit for the data. Coefficients are known with the help of Akaike Information Criteria (AIC).
- The model is computed to get the forecasts for the number of days as required.

The key components involved in the ARIMA model are,

AR- Auto regression- This model uses the relationship between the observed values and the lagged terms.

I-Integrated- It is used for differentiating the variables to make the time series data stationary.

MA- Moving Average uses the relationship between practical terms and error terms moving average, which is implemented on the data which is lagged.

The components of the ARIMA model is (p,d,q),

P- The number of lagged values that are present in the model, and they are known as the order of lag

Q- It is the moving average sequence, and the order is specified for the moving average

D- The number of times the time series data has been done differenced, and it is given by the degree of difference.

The best coefficients are selected based on the Akaike Information criteria (AIC). The AIC will take each model and takes the lowest value. The best value is that which overfits or under fits the data.

Akaike Information Criteria relies on the formula,

$$AIC = -2(log\ likelihood) + 2\ K$$

where,

- Number of parameters in the model is K
- Log-likelihood is the measure of a good fit. Higher the value better the fit.

For small sample sizes,

$$AIC_c = -2(log\ likelihood) + 2\ \ K\ + (2\ K(K+1)/(n-K-1))$$

Where,

- n=size of the sample
- K= no. of parameters for the model
- Log-likelihood is a measure of fit.

4. Results and discussion

The time-series data from our world dataset is projected for checking the new Very critically ill (Equivalent to Deaths of the patients) in the USA due to the coronavirus for a minimum period of 29 days and extended as required. The cleaned dataset with supporting variables is used to do the ARIMA model. This projected model is also visualised in R shiny application for dynamic interaction.

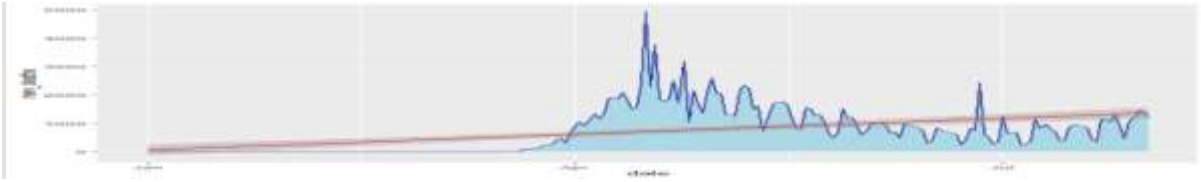
Interpretation

The data is imported to the R programming software, and the information is transformed into the required class, and then the summary of the data is observed.

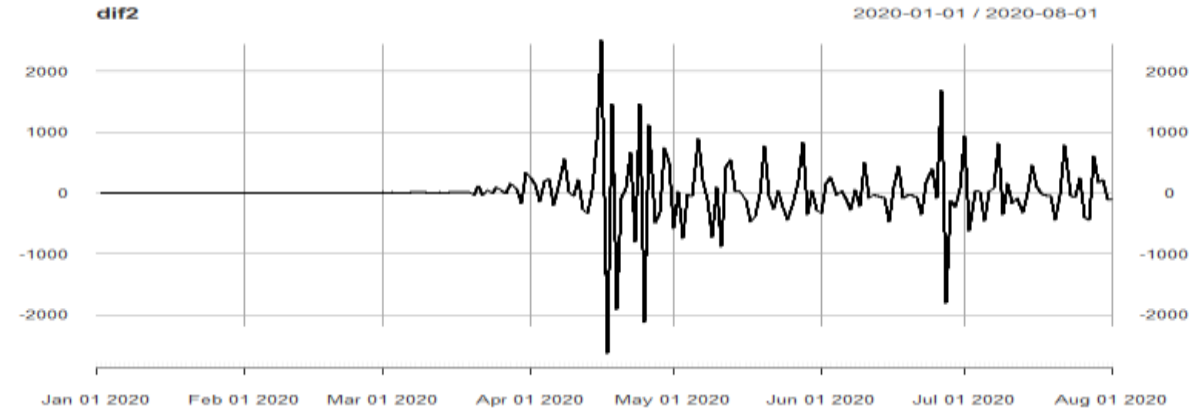
Summary:

```
> summary(covid)
     date              location          total_deaths       new_deaths
 Min.   :2020-01-01   Length:214         Min.   :     0   Min.   :   0.0
 1st Qu.:2020-02-23   Class :character   1st Qu.:     0   1st Qu.:   0.0
 Median :2020-04-16   Mode  :character   Median : 32135   Median : 551.0
 Mean   :2020-04-16                      Mean   : 53669   Mean   : 716.4
 3rd Qu.:2020-06-08                      3rd Qu.:110884   3rd Qu.:1170.5
 Max.   :2020-08-01                      Max.   :153314   Max.   :4928.0
```

The summary of the data is observed from the above table. For the given interval of time-series data, the mean is 716.4, and the maximum is found to be 4928, which indicates the highest number of cases.
The graph has been plotted concerning the number of very critically ill (Equivalent to Deaths of the patients) in the USA with respect to the months of COVID-19 to check the mean and variance of the data.



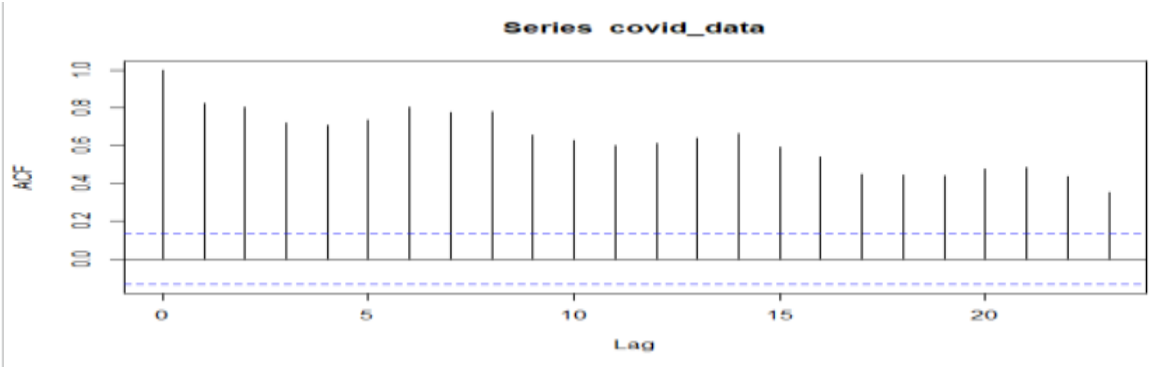
This graph shows that the data is not stationary, and the data has some trend, so it follows a random walk model. The model must be made stationary to proceed with the time series forecasting. The red line tells the mean is increasing. The variance of the data is also not constant as we can see, the distance between the data points and the mean vary with a considerable difference. This is followed by transforming the data to make it stationary.
With the transformed time series data is checked fort stationary, there are various algorithms available to convert the data to make it stationary. For univariate time-series data, the data is stabilized by taking the difference of the data—the differencing means computing the difference between consecutive observations.



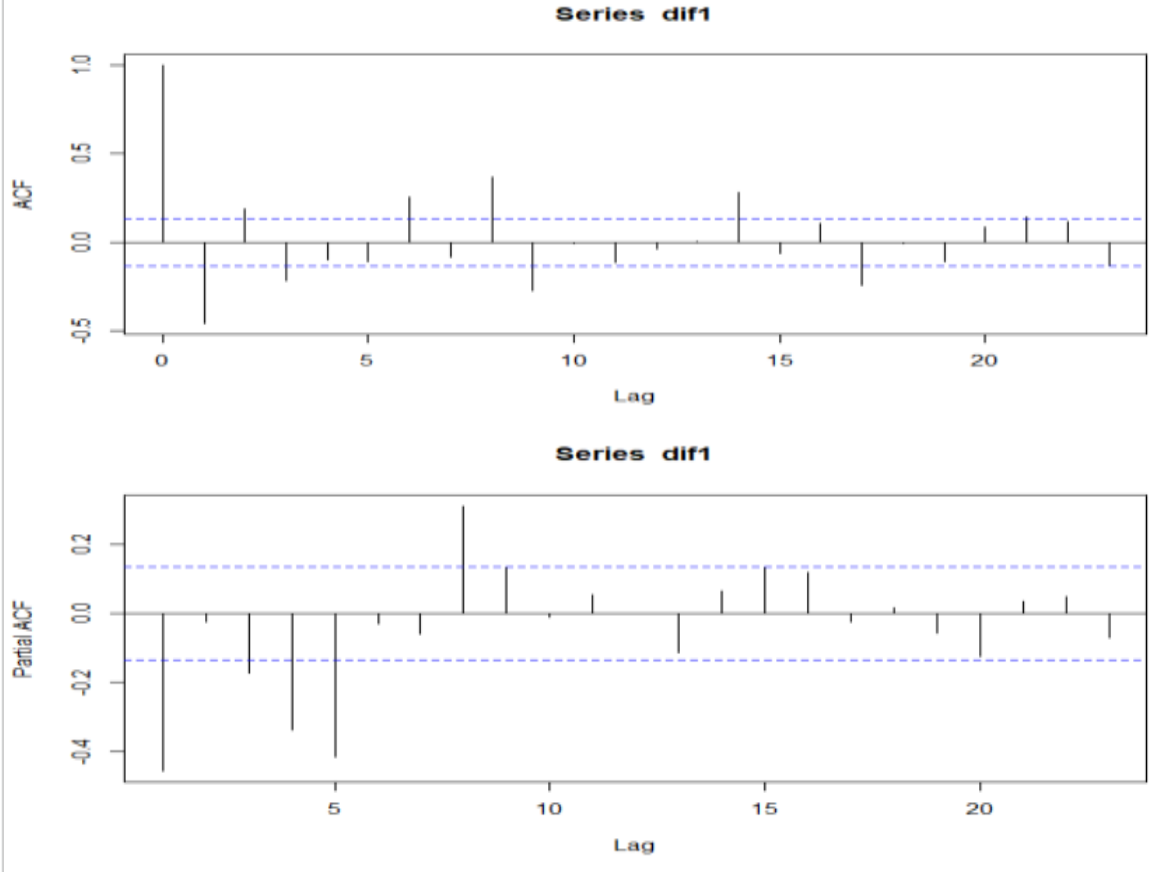
The difference data shows that the mean is constant now. Now, the data is further proceeded to check the stationarity. Typically, stationarity of a random walk model is seen through the unit root test. The most used unit root test is the ADF test or the Augmented Dicky Fuller test. This tells the time series data is stationary or not.
The hypothesis to follow the ADF test is,
H0: The null hypothesis means there is a unit root
H1: Alternative hypothesis tells the time series is stationary.
The ADF test is carried out for the differenced time series data, and p-values are determined from the results.

```
        Augmented Dickey-Fuller Test

data:  dif1
Dickey-Fuller = -10.452, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

The result of the p-value is 0.01, and is less than 0.05, thereby rejecting the H0 we can say that the data is stationary. The model further proceeds for the time series analysis.
The time-series data is modelled with the help of ARIMA. The ARIMA model forecasts based on its previous values, and it has three parameters p,d,q. These values are selected based on the ACF and PACF functions.
The ACF plot before the data is stationary has so many points exceeding the blue line.



The ACF and PACF plots after differencing and by making the data stationary we can see that most of the values are inside the blue line, which is extended to find the p, d, q values.



The Q values are taken from the ACF graph, and the P values are taken from the PACF graph. In this project, the costs are determined with the help of auto ARIMA function by selecting the best values for p, d, q.
The p,d,q values chosen by auto ARIMA with the help of AIC the best values are selected as p=3, d=1, q=2
Best model: ARIMA (3,0,2) with zero mean,

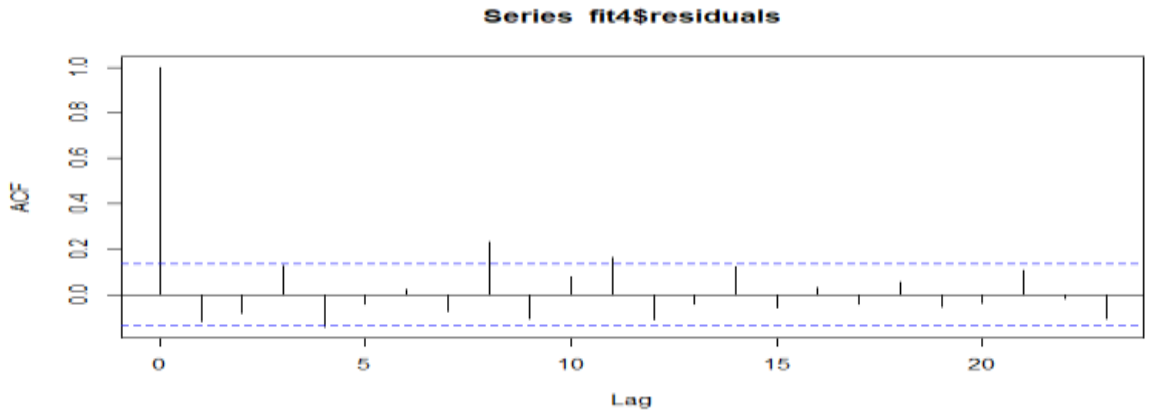
```
Series: dif1
ARIMA(3,0,2) with zero mean

Coefficients:
          ar1      ar2      ar3      ma1     ma2
      0.6082  -0.1778  -0.6058  -1.2614  0.8281
s.e.  0.0601   0.0688   0.0565   0.0462  0.0477

sigma^2 estimated as 126664:  log likelihood=-1552.18
AIC=3116.36   AICc=3116.76   BIC=3136.52
```

The result of the ARIMA is observed from the above table. With the lowest AIC values, the coefficients of the AR and MA are noted from the table, which forms the equation of the time series analysis. Few other information like Log-Likelihood and the error terms are also known.

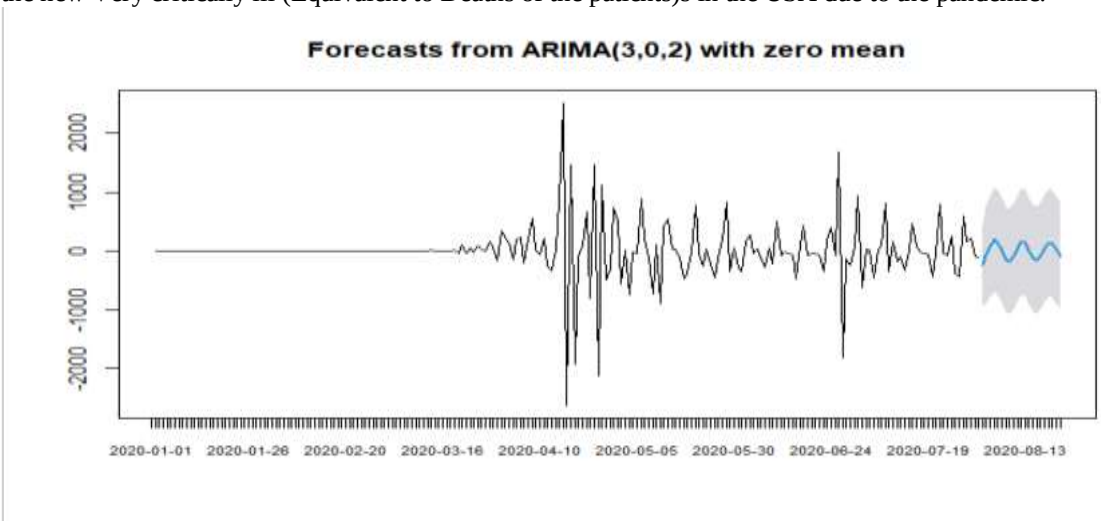
To check the lack of fit of the model the Q statistics with Box test and Ljung-Box test is carried out,

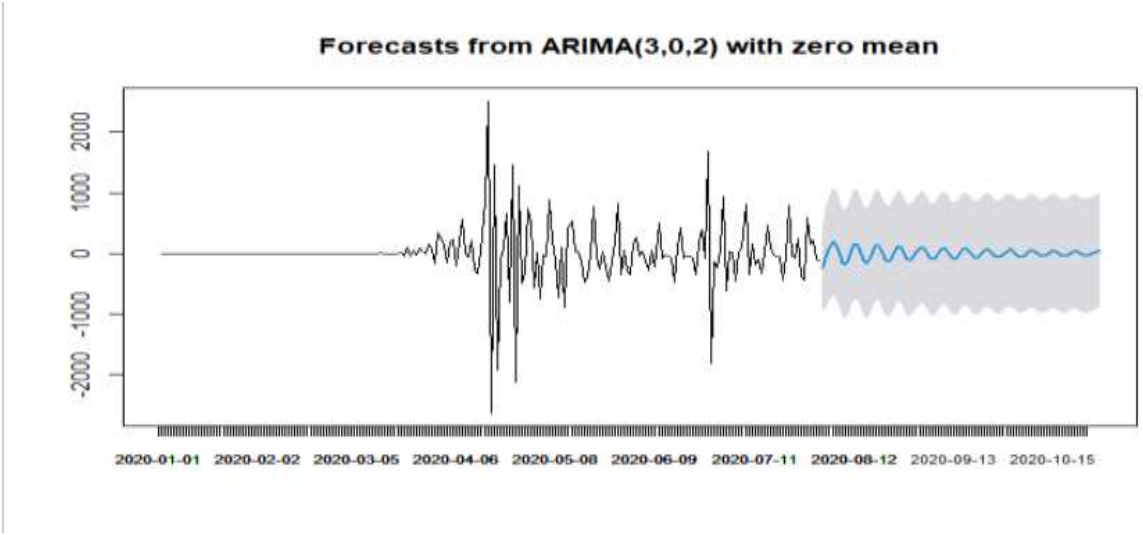
```
        Box-Ljung test

data:  dif1
X-squared = 45.327, df = 1, p-value = 1.667e-11
```

The p-value observed here is more than 0.05, which says that the model is perfect; the data points with the data points fitting under the line. The ACF residuals are observed for the excellent fit.



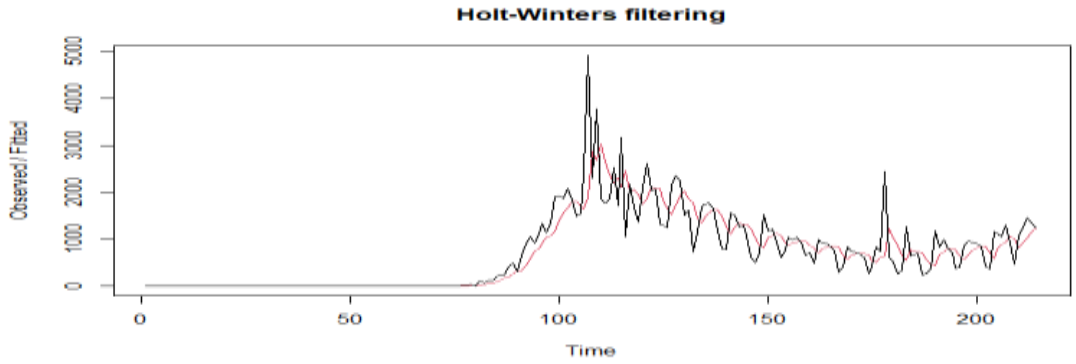
The ARIMA model is then fitted to see the forecast plot for 21 days and 90 days to see the increase in the new Very critically ill (Equivalent to Deaths of the patients)s in the USA due to the pandemic.



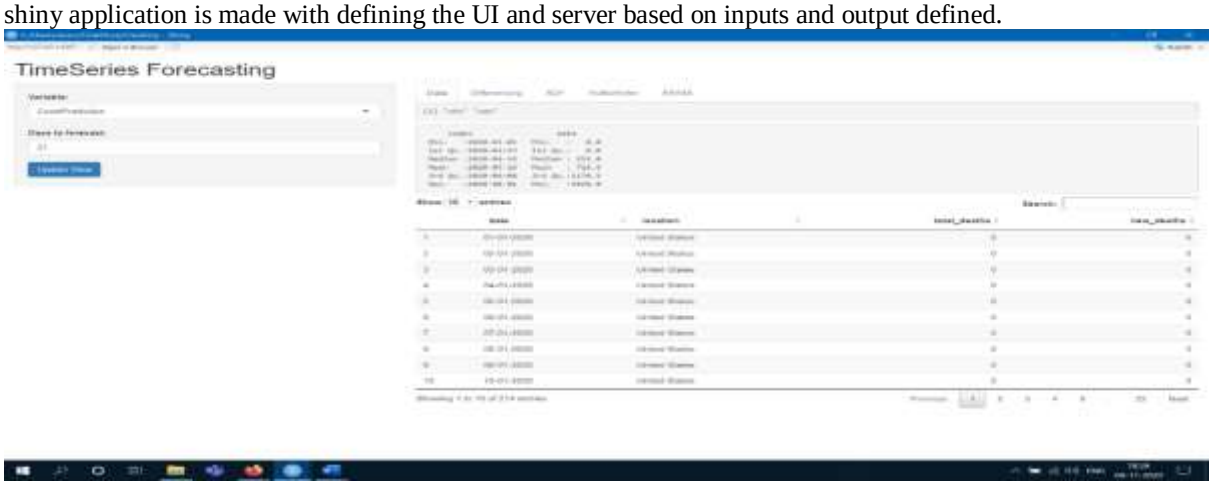


From the ARMIA forecasting, it is seen that the forecasted value of Very critically ill (Equivalent to Deaths of the patients)s for 21 days is 18633.15 which is approximated to 18633, and the number Very critically ill (Equivalent to Deaths of the patients)s when predicted for 90 days it is 82848.3 which is compared to 82848. From this, it can be highlighted that the effect of COVID-19 in the USA is spiking in number.

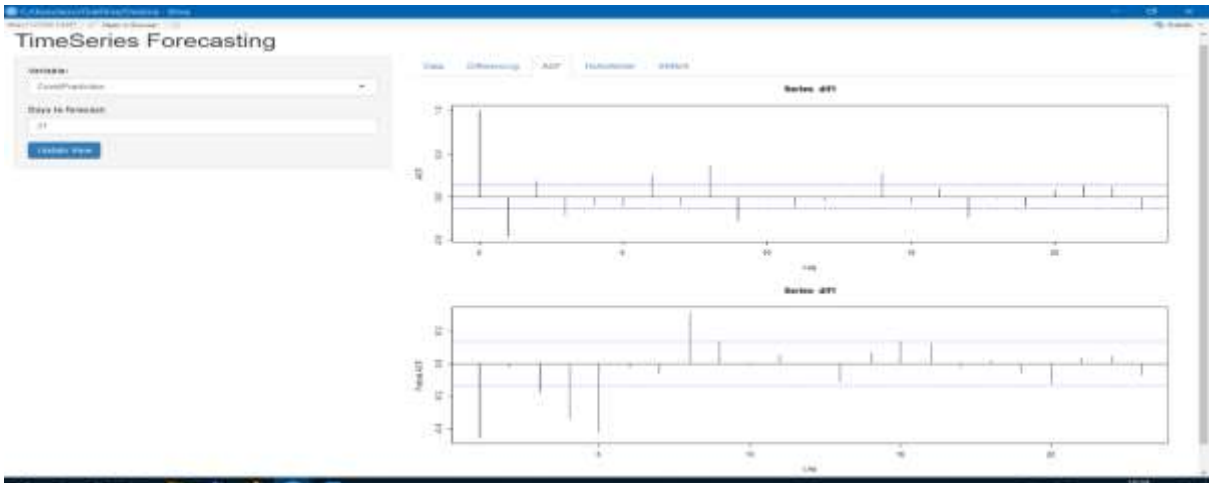
The Holts winter seasonal method is checked to determine the seasonality. The Holts-Winter forecasting equation comprises of three smoothing equations- one for the trend, one for the seasonal and one for the level. These parameters define the smoothing curve for the model. In this model, the data points lack the seasonality and the trend, so the level of the decomposition is observed between the actual and fitted data points.



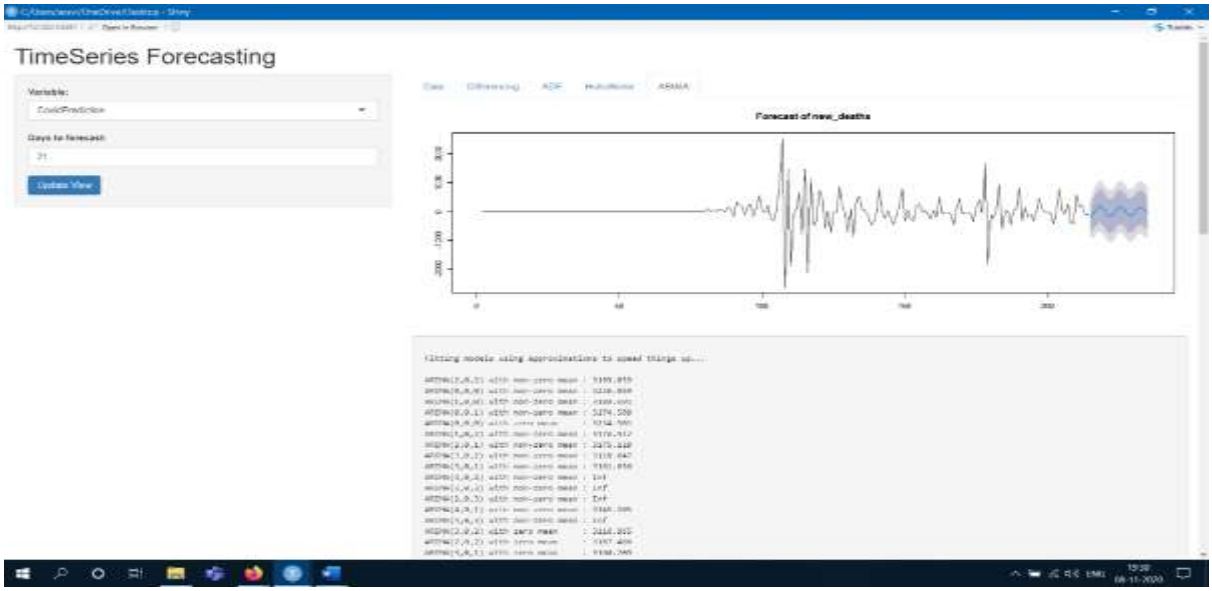
The following ARIMA model is dynamically visualised with the help of R shiny application. The R shiny dashboard with interactive buttons is built with the area given to enter the days to forecast. R shiny application is made with defining the UI and server based on inputs and output defined.

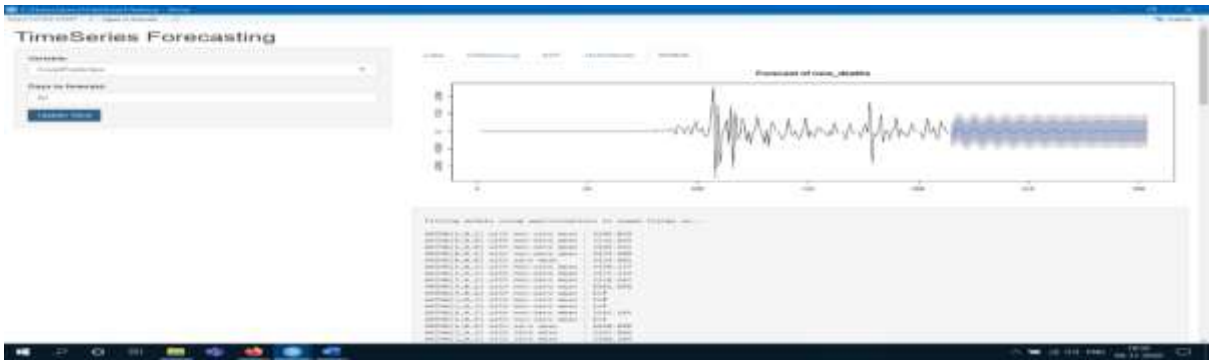



TimeSeries Forecasting


TimeSeries Forecasting


TimeSeries Forecasting
Holt-Winters filtering





5. **Findings**

The model suggests that the number of Very critically ill (Equivalent to Deaths of the patients) in the USA is increasing day by day. The parameters are said to be a good fit for future predictions. The ARIMA model is found to be the best model for prediction among the time series models. The ARIMA model in this project is projected with the best estimated provided by the AIC criterion. There is a massive increase in the cases of COVID-19 in the USA urges the need for vaccine immediately. It is found that projected value has a small deviation from the predictions posted by other sources.

6.Suggestions

This model can be performed with different other time series models like ETS etc. The visualization of the model can be improved further with more dynamic interactions. This model can be implemented for various other countries where the pandemic has worst affected. The prediction can be compared with various other sources for reliability.

6. Conclusion

The time-series analysis using the ARIMA model has shown significant results for predicting the number of very critically ill (Equivalent to Deaths of the patients) in the USA due to the pandemic. The observed effects for 21 days and 90 days come to be 18633 and 82848, which is an alarming increase in the cases. This model can be further projected to the number of days required to get an estimate of the number of Very critically ill (Equivalent to Deaths of the patients) in the USA. This model is compared with media projection which predicted the number of cases would increase to 19000 by August 21, 2020. This model predicts 18633 with a minimal deviation of 367 from the compared data. The difference is due to the skewness and the adequacy of the data from which it is built.

References:

- Li H., Liu M., Yua H, Tang S., Tang C (2020). Coronavirus disease 2019 (COVID19): Current status and future perspectives. International Journal of Antimicrobial Agents. Vol.55, Issue-5.

- Chaurasia V., Pal S., (2020). Application of machine learning time series analysis for prediction Covid-19 pandemic. Research on Biomedical Engineering, 2020.

- Devaraj J., Elevarsam R.M., Pugazhendhi R., Shafiullah G.M. (2021). Forecasting of Covid19 cases using deep learning models: Is it reliable and practically significant? Result in Physics.Vol.21. Feb.2021.

- Masfar G., Matrodji M. (2020). Interdependence of Loan and Deposit Volumes at Government-Owned, Private, and Joint Venture Banks in Indonesia during 2003-2017. Proceeding on International Conference of Science Management Art Research Technology, 2020.

- Lijing Wang, Xue Ben, Aniruddha Adiga, Adam Sadilek (2020). "Using Mobility Data to Understand and Forecast COVID19 Dynamics", Cold Spring Harbor Laboratory.

- Fernando Rojas, Olga Valenzuela, Ignacio Rojas. "Estimation of COVID-19 dynamics in the different states of the United States using Time Series clustering", Cold Spring Harbor Laboratory.

- Messis A., Adjebli Ahmed, Ayeche Riad, Ghidouche Abderrezak, Ait-Ali Djida (2020). Forecasting daily confirmed COVID-19 cases in Algeria using ARIMA model, Cold Spring Harbor Laboratory.

- Sahoo Kalyan K., Muduli K.K., Luhach A.K, Poonia R.C. Pandemic COVID-19: An empirical analysis of impact on Indian higher education system. Journal of Statistics and Management Systems, 2021.

- Sahoo Kalyan K., Mishra P.C., M.Kumar (2020), PhD Journey Made Easy. 1st Edition. Orange Publication, Kolkatta.