

India Covid 19 Predictions Using Box- Jenkins Model and Correlation Analysis between Confirmed, Death, Tested and Vaccinated Cases Using Panel Regression & Pearson’s Coefficient

Suraj Jayasimha

Abstract— Covid 19 Pandemic has affected millions worldwide over the last two years and we are still grappling with the disease. A research was conducted on the India state wise COVID19 related data (Confirmed, Death, Tested and Vaccination Cases) available from source https://api.covid19india.org/.(1) A time series model (in this case Box Jenkins Model or ARIMA) was used to predict the confirmed cases, Death cases and Tested cases for a period of 8 weeks from 8th July 2021. Also, a correlation analysis using panel regression and Persons Correlation Coefficient was performed to establish any correlation between Confirmed cases, Death Cases, Tested cases and Vaccination both independently and with Joint effect. . The actual data was compared until 12th August 2021 with the predicted values from the model. The variances between predicted values and actual data were quite minimal. While the time series is usually best in predicting the numbers closer to the period, the predictions can be useful to understand the number of cases to make an action plan to combat the future cases. Panel Regression analysis and Person’s correlation coefficient was used to establish correlation if any between Confirmed cases, Death cases, Tested Cases and Vaccination. While it was possible to establish positive correlation between Confirmed cases and Tested cases, Death Cases and Tested Cases using both the Panel Regression analysis and Person’s correlation coefficient, there was no correlation established between Confirmed cases and vaccinated cases with a 3-month lag and Death Cases and Vaccinated cases with a 3-month lag using both the methods. This may be due to the fact the vaccination started in full force only in the middle of the 2021, the effects of vaccination will be seen post June 2021. A scatter plot was also drawn to validate the results.

Index Terms— Box-Jenkins Model, Panel Regression, Pearson’s Correlation Coefficient, P Value, Scatter Plot.

I. INTRODUCTION

The Covid 19 is an ongoing global pandemic caused by severe acute respiratory syndrome coronavirus 2. This pandemic has resulted in severe global, social and economic disruption, including global recession . It has led to widespread shortages, disruption, panic buying and deaths in large numbers due to one disease.

It has also caused job losses, increased poverty, and general decrease in quality of mental and physical health among majority of the population. Government and Private offices , Educational institutions , Malls , Cinemas , Eating places and

other public areas have been partially or fully closed. Many events have been cancelled or postponed. Social Media has been used to spread incorrect information on Covid building fear in people. The Public health system of many countries have collapsed unable to handle the numerous covid patients requiring immediate hospital admission including intensive care. Lack of knowledge of the treatment in the initial days of the pandemic saw many deaths and panic among the population. Slowly the situation improved, however in few countries like India, United States, Brazil etc, the cases and deaths only saw an increase This research report aims to predict the incidence of covid cases in India for a period of 8 weeks from the 08/07/2021 using Box Jenkins model. The research also aims to establish any correlation between covid cases, Tested and vaccination using regression analysis.

II. DATA & METHODS

A Comma Separated Value (CSV) dataset which contained cumulative datewise, state wise covid data in the following format was used:

- The dates starting from 30th Jan 2020 to 8th July 2021.
• 30,752,108 confirmed cases, 405,967 death cases and 458,914,056 Tested cases over a 17.5-month period.
• Vaccinated data from 16th Jan 2021 to 8th July 2021, a total of 353,617,202 cases

Figure 1

Sample csv data set

Table with 6 columns: Date, State, Confirmed, Recovered, Deceased, Other. It contains 14 rows of sample data for various dates and states like Kerala, Delhi, Telangana, Rajasthan.

A. Data Cleaning, Transformation and Analysis

Data Cleaning the Data available from the public source was cumulative, whereas the requirement was for daily non-cumulative data. Data had null or zero values in between. The cumulative data in few cases was not in increasing values. The order of the data was date wise, whereas state wise data

Suraj Jayasimha, Dhirubhai Ambani International School

India Covid 19 Predictions Using Box- Jenkins Model and Correlation Analysis between Confirmed, Death, Tested and Vaccinated Cases Using Panel Regression & Pearson's Coefficient

was needed. The vaccination data was in landscape format with dates as columns and states in rows unlike the covid daily data and Tested data, thereby requiring the vaccination data to be transformed.

Data Transformation The three data sets were first converted to daily non-cumulative data using formulas and checks. Null data was set to the previous value or zero depending on the formula. The complete data was first filtered state wise, copied and then transformed to usable data for running the R model. The vaccination data which was in landscape format was transposed to portrait format to be consistent with other data. The three data sets were also aligned and made consistent such that it was sorted date and state wise. The daily data was then converted to weekly data to enable the future predictions easier. Also, when comparing the vaccination data with confirmed cases and death cases – a 3-month lag was used

Data Summary and Analysis

Figure 2 Covid Cumulative Confirmed Cases Period 30th Jan 2020 to 8th July 2021

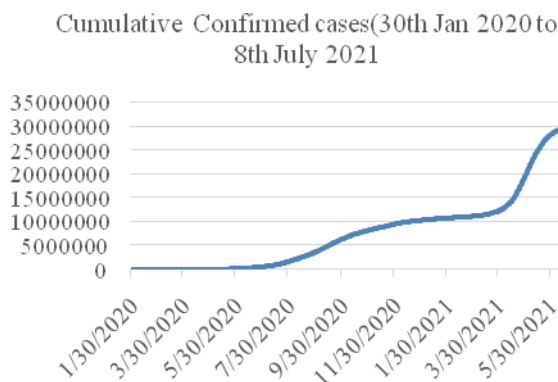


Figure 2– The sharp increase in Covid cases in India started post June 2020 reaching the peak during Oct -Nov 2020, Post Nov 2020 to March 2021, it appeared that the curve had flattened. However, Post March 2021, the start of the “second wave” saw another set of steep increase in covid cases until May 2021 declining post May 2021.

Figure 3 Covid Cumulative Death Cases Period 30th Jan 2020 to 8th July 2021

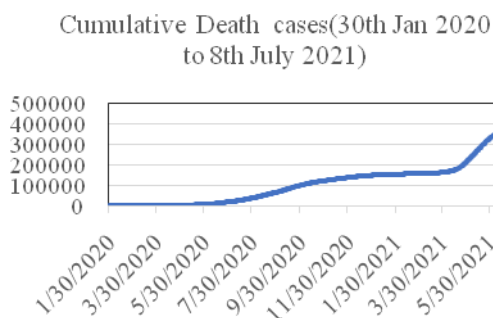


Figure 3 – The trend of total death cases followed the total confirmed cases in wave 1. with respect to covid wave 2, the total death cases appear to have a sharper increase and has a later flattening when compared to the confirmed cases. It implies that while the confirmed cases have fallen, the death continued to rise for some more time. The total death cases for the period 30th Jan 2020 to 8th July 2021 across the country is 405,967

Figure 4 Covid Cumulative Tested Cases Period 30th Jan 2020 to 8th July 2021

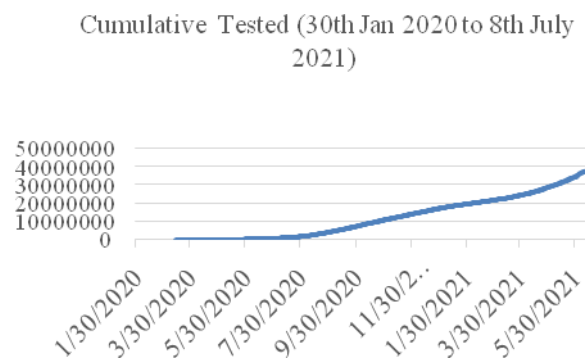


Figure 4 – The Tested cases until June 2020 was not substantial which may be due to the lack of availability of Test kits. The Tested cases picked up post June 2020 and there has been no slack in the Tested cases since then. A total of 427,016,605 tests has been conducted from 30th Jan 2020 to 8th July 2021 across the country

B. Implementation Objectives and Methods

Objective 1 - Predict the incidence of covid confirmed cases for an 8-week period post 8th July 2021 using box-Jenkins model and compare the predicted data to actual incidences of confirmed cases

Objective 2 Predict the incidence of covid death cases for an 8-week period and compare the predicted data to actual incidences of death cases

Objective 3 Predict the incidence of covid Tested cases for an 8-week period and compare the predicted data to actual incidences of tested cases

R and RStudio were used to carry out statistical analysis of the data. R is the most popular programming language in the field of data science for Statistical analysis

Figure 5 Sample R Code using Box Jenkins model for arriving at predictions

```
> library(readxl)
> library(forecast)
> library(dynlm)
> options(scipen=7)
> COVID_19_data <- read_excel("D:\Covid-19\COVID-19 data_07_07_2021 - Weekly.xlsx",
+ col_types = c("numeric", "date", "text",
+ "numeric", "numeric", "numeric"))
> Cases<-as.data.frame(COVID_19_data[,3:4])
> Deaths<-as.data.frame(COVID_19_data[,c(3,5)])
> Tests<-as.data.frame(COVID_19_data[,c(3,6)])
> Unique_States<-unique(COVID_19_data$Region)
> Forecastcases<-matrix(0,nrow = length(Unique_States),ncol=8)
> forecastdeaths<-matrix(0,nrow = length(Unique_States),ncol=8)
> forecaststests<-matrix(0,nrow = length(Unique_States),ncol=8)
>
> #ARIMA with Box-Cox Transformed Cases
for (i in 1:length(Unique_States))
+ {
+ data<-Cases[Cases$Region==Unique_States[i],]
+ ARIMA_Model<-auto.arima(data$Daily_Confirmed_cases,stepwise = FALSE,max.order = 10, lambda = "auto", ic
+ ="aic",biasadj = TRUE)
+ Eight_Week_predict<-as.vector(forecast(ARIMA_Model,8)$mean)
+ forecastcases[i,]<-Eight_Week_predict
+ }
> CasesEstimate = as.matrix(cbind(Unique_States,round(forecastcases,0)))
> colnames(CasesEstimate)<- c("States","Week 1","Week 2","Week 3","Week 4","Week 5","Week 6","Week 7","Week 8")
> CasesEstimate States Week 1 Week 2 Week 3 Week 4 Week 5 Week 6 Week 7 Week 8
```

Objective 4 -Establish correlation using Panel regression and Pearson’s correlation coefficient between Covid confirmed cases and Tested cases

Objective 5 -Establish correlation using Panel regression and Pearson’s correlation coefficient between Covid Death cases and Tested cases

Objective 6-Establish correlation using Panel regression and Pearson’s correlation coefficient between Covid Confirmed cases and Vaccination with a 3-month lag

Objective 7-Establish correlation using Panel regression and Pearson’s correlation coefficient between Covid Death cases and Vaccination with a 3-month lag

Objective 8-Establish the joint effect (if any) of Deaths and Covid Confirmed Cases on Vaccinations at a lag of 3 months using panel regression .

Figure 5 Sample R Code used for Panel Regression analysis and Person’s Correlation Coefficient

```
>#Panel regression with Tests
>library(plm)
>COVID_19_data1<-COVID_19_data
>names(COVID_19_data1)<-c("SNo","Date","Region","DCC","DDC","DT")
>library(lmtest)
>panell1 <- plm(DCC ~ DT,
+
+ data = COVID_19_data1,
+ index = c("Region", "Date"),
+ model = "within",
+ effect = "twoways")
>coefest(panell1, vcov. = vcovHC, type = "HC1")

t test of coefficients:

Estimate Std. Error t value Pr(>|t|)
DT 0.081726 0.025974 3.1465 0.001673 **
...
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Pearson's Correlation coefficient:

>cor.test(COVID_19_data$Daily Confirmed cases ,COVID_19_data$Daily Tested ,method = "pearson")

Pearson's product-moment correlation

data: COVID_19_data$Daily Confirmed cases and COVID_19_data$Daily Tested
t = 39.679, df = 2420, p-value < 2.2e-16
Alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval: 0.6090676 0.6513584
Estimates: cor 0.6278167
```

III. RESULTS

Objective 1- Comparing the predicted confirmed cases from Box Jenkins model to the actual confirmed published cases

Figure 6 Comparing the actual confirmed cases with the projected confirmed cases arrived from the model

	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
	08-07-2021	15-07-2021	22-07-2021	29-07-2021	05-08-2021	12-08-2021	19-08-2021	26-08-2021	02-09-2021
Confirmed Cases Projected (Box Jenkins Model)	30,752,108	31,041,733	31,349,299	31,712,893	32,121,777	32,580,156	33,134,435	33,750,871	34,424,745
Actual Published Data	30,752,108	31,025,878	31,291,706	31,571,299	31,855,787	32,117,067			
% variance	0.00%	-0.05%	-0.18%	-0.45%	-0.83%	-1.42%			

Figure 6 analysis -The Week 1 confirmed cases from the model and the actual numbers are very close with a variance of only -0.05%. The week two is also seeing a variance of -0.18 % and week 3 ,4 & 5 at - 0.45% , -0.83% & -1.42 % respectively . The time series is usually best in predicting the numbers closer to the period. As the prediction period increases, so does the variance.

Objective 2- Comparing the predicted death cases from Box Jenkins model to the actual death cases published data

Figure 7 Comparing the actual death cases with the projected death cases arrived from the model

	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
	08-07-2021	15-07-2021	22-07-2021	29-07-2021	05-08-2021	12-08-2021	19-08-2021	26-08-2021	02-09-2021
Death Cases Projected (Box Jenkins Model)	405,967	411,599	418,012	424,901	432,472	441,589	451,308	461,975	474,507
Actual Published Data	405,967	412,563	419,502	423,842	426,785	430,285			
% variance	0.00%	0.23%	0.36%	-0.25%	-1.31%	-2.56%			

Figure 7 analysis -The Week 1 death cases from the model and the actual numbers are close with a variance of only 0.23%. The week two is also seeing a variance of 0.36 % and week 3 , 4 & 5 at -0.25% -1.31% & -2.56% respectively. The variance is not material, and it is a good model to predict the future cases for government to take precautions

Objective 3- Comparing the predicted death cases from Box Jenkins model to the actual death cases published data

Figure 8 Comparing the actual Tested cases with the projected Tested cases arrived from the model

	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
	08-07-2021	15-07-2021	22-07-2021	29-07-2021	05-08-2021	12-08-2021	19-08-2021	26-08-2021	02-09-2021
Tested Cases -Projected (using the Box Jenkins Model)	427,016,605	439,812,405	453,256,506	467,362,417	482,258,985	498,043,300	514,877,816	532,831,520	552,113,765
Actual Published Data	427,016,605	440,023,239	452,939,545	464,650,723	476,533,650	489470779			
% variance	0.00%	0.05%	-0.07%	-0.58%	-1.19%	-1.72%			

Figure 8 analysis - The Week 1 tested cases from the model and the actual numbers are close with a variance of only 0.05%. The week two is also seeing a variance of -0.07 % and week 3 , 4 & 5 at -0.58% -1.19% & -1.72% respectively. The variance is minimal, and the model is a good fit to predict the future tested cases

Objective 4- Correlation analysis between Covid confirmed cases and Tested cases using Panel regression analysis and Pearson’s correlation coefficient

India Covid 19 Predictions Using Box- Jenkins Model and Correlation Analysis between Confirmed, Death, Tested and Vaccinated Cases Using Panel Regression & Pearson's Coefficient

Hypothesis -

H_0 : There is no correlation between the Covid Confirmed Cases and the Tested cases.

H_1 : There exists a correlation between the Covid Confirmed Cases and the Tested cases.

t test of coefficients: Estimate Std. Error t value Pr(>|t|)

DT 0.081726 0.025974 3.1465 0.001673 **

At 5% level of significance, ($\alpha = 5\%$):

Basis the Routput, we see that on Tested for the correlation, $\text{Pr}(>|t|) < 0.05$. Hence, we reject H_0 and this suggests that there exists a correlation between Covid confirmed cases and Tested.

Pearson's product-moment correlation

data: COVID_19_data\$`Daily Confirmed cases` and COVID_19_data\$`Daily Tested` t = 39.679, df = 2420, p-value < 2.2e-16

Alternative hypothesis: true correlation is not equal to 0 95 percent confidence interval: 0.6030676 0.6513584

Estimates: cor 0.6278167

The Pearson's coefficient comes to be **0.62** which states that there exists a strong positive correlation between Covid Confirmed Cases and Tested

Figure 9 Scatter Plot between Confirmed cases and Tested cases

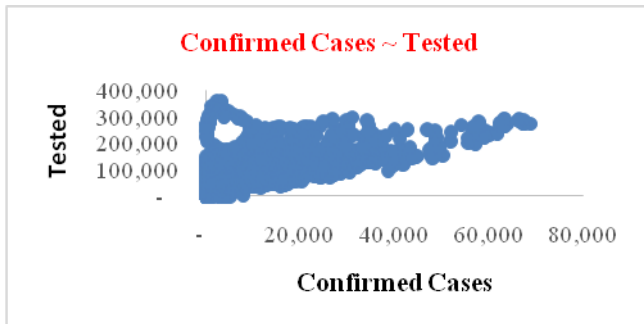


Figure 9 analysis- There is a positive correlation between Confirmed cases and Tested cases also confirmed by the scatter plot

Objective 5- Correlation analysis between Covid Death cases and Tested cases using Panel regression analysis and Pearson's correlation coefficient

H_0 : There is no correlation between the Covid Death Cases and the Tested cases.

H_1 : There exists a correlation between the Covid Death Cases and the Tested cases.

t test of coefficients:

Estimate Std. Error t value Pr(>|t|)

DT 0.0011557 0.0004808 2.4038 0.01631 *

At 5% level of significance, ($\alpha = 5\%$):

Basis the R output, we see that on Tested for the correlation, $\text{Pr}(>|t|) < 0.05$. Hence, we reject H_0 and this suggests that there exists a correlation between Covid

confirmed cases and Tested.

Pearson's product-moment correlation

data: COVID_19_data\$`Daily Deaths` and COVID_19_data\$`Daily Tested`

t = 33.982, df = 2420, p-value < 2.2e-16

Alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval: 0.5407740 0.5947275

Estimates: cor 0.5683614

The Pearson's coefficient comes to be **0.56** which states that there exists a strong positive correlation between Covid Confirmed Cases and Tested

Figure 10 Scatter Plot between Death cases and Tested cases

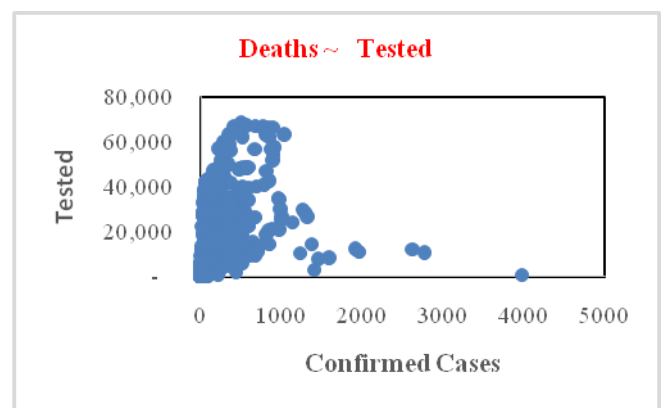


Figure 10 analysis- There is a positive correlation between Death cases and Tested cases also confirmed by the scatter plot

Objective 6- Correlation analysis between Covid Confirmed cases and Vaccinated cases with a 3-month lag using Panel regression analysis and Pearson's correlation coefficient

H_0 : There is no correlation between the Covid Confirmed Cases and the Vaccinated cases.

H_1 : There exists a correlation between the Covid Confirmed Cases and Vaccinated cases.

t test of coefficients:

Estimate Std. Error t value Pr(>|t|)

DV -0.043401 0.011791 -3.6809 0.000237 ***

At 5% level of significance, $\text{Pr}(>|t|) < 0.05$. which suggests that there exists a correlation between Covid Confirmed Cases and Vaccinations.

Pearson's product-moment correlation

data: COVID_19_vaccinated_data_lagged\$DCC and COVID_19_vaccinated_data_lagged\$DV

t = -0.63926, df = 2833, p-value = 0.5227

Alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval: -0.04880126 0.02481480

Estimates: cor -0.0120095

The Pearson's coefficient comes to be -0.012 which implies

that there exists no correlation

Figure 11 Scatter Plot between Confirmed cases and Vaccinated cases

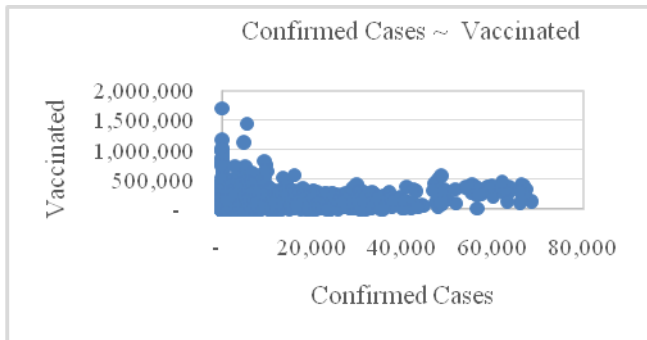


Figure 10 analysis-The points are distributed randomly across the graph. While the panel regression analysis does show a correlation, the persons coefficient does not show any, the data for vaccination is from Mid Jan 2021 and we can conclude that sufficient data is not available to establish correlation between covid confirmed cases and vaccinations

Objective 7- Correlation analysis between Covid Death cases and Vaccinated cases with a 3-month lag using Panel regression analysis and Pearson’s correlation coefficient

H₀: There is no correlation between the Covid Death Cases and the Vaccinated cases.

H₁: There exists a correlation between the Covid Death Cases and Vaccinated cases.

t test of coefficients:

Estimate Std. Error t value Pr(>|t|)
DV -0.000252399 0.000071748 -3.5179 0.0004422 ***

Pr(>|t|) in this case as well is lesser than 0.05. This suggests that there is correlation between Deaths and Vaccination with a lag of 3 months.

Pearson's product-moment correlation

data: COVID_19_vaccinated_data_lagged\$DD and COVID_19_vaccinated_data_lagged\$DV

t = 6.7556, df = 2833, p-value = 0.0000000000172

Alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval: 0.08951471 0.16197565

Estimates: cor 0.1259131

Also, the Pearson’s correlation coefficient of 0.12 implies that there is a very weak correlation or no correlation between Deaths and Vaccinations.

Figure 11 Scatter Plot between Death cases and Vaccinated cases

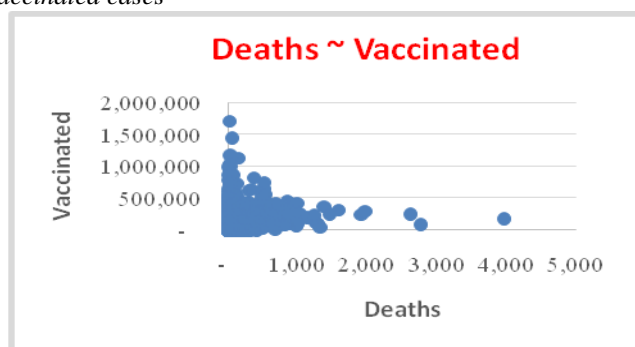


Figure 11 analysis-The points are distributed randomly across the graph. While the panel regression analysis does show a correlation, the persons coefficient does not show any, the data for vaccination is from Mid Jan 2021 and we can conclude that sufficient data is not available to establish correlation between covid death cases and vaccinations

Objective 8- Correlation analysis due to Joint effect of Covid Confirmed Cases, Death cases and Vaccinated cases with a 3-month lag using Panel regression analysis and Pearson’s correlation coefficient

H₀: There is no correlation between the Covid Confirmed, Covid Death Cases and the Vaccinated cases.

H₁: There exists a correlation between the Covid Confirmed m Covid Death Cases and Vaccinated cases

t test of coefficients:

Estimate Std. Error t value Pr(>|t|)
DCC -3.08043 0.37118 -8.2990 <2e-16 ***
DD -15.91417 15.42432 -1.0318 0.3023

The “p” values are significantly different from when run independently, hence I am unable to make any conclusions. However, from the graph and the independent relation carried out, I could not visualize any correlation between the same. This may be since the vaccination was administered only from Mid Jan -2021 and still in progress. The effect of vaccination is yet to impact the covid incidences.

IV. LIMITATIONS AND SUGGESTIONS

- The time series is usually best in predicting the numbers closer to the period. As the prediction period increases, so does the variance. It is always best to update the data with the latest numbers for closer predictive numbers
- The model has used the data as reported and has not validated the data from other sources
- The model does consider any second order impacts to the data reported
- While the data period for confirmed cases, death cases and tested cases is from 30th Jan 2020, the vaccination data is only from 15th Jan 2021.
- The Vaccination data was not sufficient to establish a valid conclusion. It is suggested that a larger data set for vaccination can be used for establishing a correlation between parameters.
- The research was conducted for the country, it is suggested that a more detailed analysis can be carried out state wise, district wise to analyse trends
- Second order impacts like availability of covid care, vaccinations, oxygen, impact of preexisting conditions etc can be analyzed as well.

V. SUMMARY

Objective 1 - The Week 1 covid confirmed cases from the model and the actual numbers are very close with a variance of only -0.05%. The week two is also seeing a variance of -0.18 % and week 3 & 4 at - 0.45% and -0.83% respectively. the variance is quite minimal, and the Government can use the Box Jenkins model to predict the future cases to manage the situation better.

Objective 2 :The Week 1 death cases from the model and the actual numbers are close with a variance of only 0.23%. The week two is also seeing a variance of 0.36 % and week 3 & 4 at -0.25% and -1.31% respectively. The Variance is quite immaterial.

Objective 3:- The Week 1 Tested cases from the model and the actual numbers are close with a variance of only 0.05%. The week two is also seeing a variance of -0.07 % and week 3 & 4 at -0.58% and -1.19% respectively. Again, it is conclusive that the prediction from the model is a good representation of the actual cases.

Objective 4 : - At 5% level of significance, ($\alpha = 5\%$), Looking at the R output, we can see that, the $Pr(>|t|)$ in case of effect"; is lesser than 0.05 for confirmed cases and Tested cases. This suggests that there is a correlation between Covid Confirmed Cases and Tested cases. Also, the Pearson's coefficient between confirmed cases and tested cases comes to be **0.62** which states that there exists a strong positive correlation between Covid Confirmed Cases and Testedcases

Objective 5 : - At 5% level of significance, ($\alpha = 5\%$), Looking at the R output, we can see that, the $Pr(>|t|)$ in case of effect"; is lesser than 0.05 for death cases and tested cases This suggests that there is a correlation between Covid Death Cases and Tested cases. Also, the Pearson's coefficient between death cases and tested cases comes to be **0.57** which states that there exists a strong positive correlation between Covid deaths and TestedCases.

Objective 6 : - At 5% level of significance, ($\alpha = 5\%$), Looking at the R output, we can see that, the $Pr(>|t|)$ in case of effect"; is lesser than 0.05 for confirmed cases and vaccination cases with a lag of 3 months. While the panel regression analysis does show a correlation, the persons coefficient does not show any. As the data for vaccination is from Mid Jan 2021 and we can conclude that sufficient data is not available to establish correlation between confirmed cases and vaccinations

Objective 7 : - At 5% level of significance, ($\alpha = 5\%$), Looking at the R output, we can see that, the $Pr(>|t|)$ in case of effect"; is lesser than 0.05 for death cases and vaccination cases with a lag of 3 months .While the panel regression analysis does show a correlation , the persons coefficient does not show any .As the data for vaccination is from Mid Jan 2021 and we can conclude that sufficient data is not available to establish correlation between Death cases and vaccinations

Objective 8 : - The "p" values of Joint effect of Covid Confirmed Cases, Death cases and Vaccinated cases with a 3-month lag of is significantly different from when run independently, hence it is difficult to make conclusions.

However, from the graph and the independent correlationanalysis carried out, it was difficult to visualize any correlation between the parameters. This may be since the vaccination was administered only from Mid Jan -2021 and still in progress. The effect of vaccination is yet to impact the covid incidences

VI. APPENDIX

For a Few Major States like Maharashtra, Karnataka and Delhi, the actual published data was compared with the Predicted Confirmed and Death cases from the model to understand the trends vis a vis the country as a whole

MAHARASTRA

Comparing the actual data with the confirmed cases for Maharashtra arrived at from the model

Figure 12 Maharashtra Covid confirmed cases confirmed cases vs the projected confirmed cases from the model

	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
	08-07-2021	15-07-2021	22-07-2021	29-07-2021	05-08-2021	12-08-2021	19-08-2021	26-08-2021	02-09-2021
Confirmed Cases - Projected (using the Box Jenkins Model)									
	6,131,976	6,193,680	6,256,182	6,322,454	6,395,393	6,479,072	6,578,012	6,697,505	6,843,301
Actual Published Data	6,131,976	6,189,257	6,245,057	6,290,156	6,336,220				
% variance	0.00%	-0.07%	-0.18%	-0.51%	-0.93%				

Figure 12 analysis: - The Week 1 confirmed cases from the model and the actual numbers are very close with a variance of only -0.07%. The week two is also seeing a variance of -0.18 % and week 3 & 4 at - 0.51% and -0.93% respectively . The trend of variance is same as that for the country.

Figure 13 Maharashtra Covid Death cases confirmed cases vs the projected Death cases from the model

	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
	08-07-2021	15-07-2021	22-07-2021	29-07-2021	05-08-2021	12-08-2021	19-08-2021	26-08-2021	02-09-2021
Death Cases - Projected (using the Box Jenkins Model)									
	124,296	126,456	128,925	131,762	135,021	138,752	142,999	147,801	153,192
Actual Published Data	124,296	126,500	131,038	132,335	133,530				
% variance	0.00%	0.08%	1.64%	0.43%	-1.10%				

Figure 13 analysis - The Week 1 death cases from the model and the actual numbers are close with a variance of only 0.08%. The week two is also seeing a variance of 1.36 % and week 3 & 4 at 0.43% and -1.10 % respectively. The trend of variance is almost the same as that for the country.

DELHI

Comparing the actual data with the confirmed cases for Delhi arrived at from the model

Figure 14 Delhi Covid confirmed cases confirmed cases vs the projected confirmed cases from the model

	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
	08-07-2021	15-07-2021	22-07-2021	29-07-2021	05-08-2021	12-08-2021	19-08-2021	26-08-2021	02-09-2021
Confirmed Cases - Projected (using the Box Jenkins Model)	1,434,873	1,435,328	1,435,744	1,436,192	1,436,708	1,437,300	1,437,853	1,438,463	1,439,240
Actual Published Data	1,434,873	1,435,553	1,435,720	1,436,144	1,436,579				
% variance	0.00%	0.00%	0.00%	0.00%	-0.01%				

Figure 14 analysis: - The variances of actual data and that from the model is 0% until week 3 and a slight variance of -0.01% for week4. The model can predict the cases with negligible variance

Figure 15 Delhi Covid Death cases confirmed cases vs the projected Death cases from the model

	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
	08-07-2021	15-07-2021	22-07-2021	29-07-2021	05-08-2021	12-08-2021	19-08-2021	26-08-2021	02-09-2021
Death Cases-Projected (using the Box Jenkins Model)	25,008	25,040	25,087	25,155	25,244	25,354	25,494	25,655	25,807
Actual Published Data	25,008	25,022	25,040	25,049	25,060				
% variance	0.00%	-0.07%	-0.19%	-0.42%	-0.73%				

Figure 15 analysis-The Week 1 death cases from the model and the actual numbers are close with a variance of only - 0.07%. The week two is also seeing a variance of -0.19% and week 3 & 4 at - 0.42% and -0.73 % respectively. . The trend of variance is almost the same as that for the country.

KARNATAKA

Comparing the actual data with the confirmed cases for Karnataka arrived at from the model

Figure 16 Karnataka Covid confirmed cases confirmed cases vs the projected confirmed cases from the model

	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
	08-07-2021	15-07-2021	22-07-2021	29-07-2021	05-08-2021	12-08-2021	19-08-2021	26-08-2021	02-09-2021
Confirmed Cases - Projected (using the Box Jenkins Model)	2,864,868	2,870,040	2,887,920	2,896,116	2,903,041	2,908,880	2,914,154	2,919,652	2,923,333
Actual Published Data	2,864,868	2,878,564	2,889,994	2,901,247	2,913,512				
% variance	0.00%	0.02%	0.07%	0.18%	-0.36%				

Figure 16 analysis- The Week 1 confirmed cases from the model and the actual numbers are very close with a variance of only -0.02%. The week two is also seeing a variance of 0.07% and week 3 & 4 at 0.18% and -0.36% respectively. while the variances are quite immaterial, the trend is in the opposite direction to the country as a whole

Figure 17 Karnataka Covid Death cases confirmed cases vs the projected Death cases from the model

	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
	08-07-2021	15-07-2021	22-07-2021	29-07-2021	05-08-2021	12-08-2021	19-08-2021	26-08-2021	02-09-2021
Death Cases-Projected (using the Box Jenkins Model)	35,663	36,018	36,461	36,906	37,407	38,180	38,986	39,925	40,937
Actual Published Data	35,663	36,037	36,293	36,491	36,705				
% variance	0.00%	0.05%	-0.46%	-1.12%	-2.09%				

Figure 17 analysis- The Week 1 death cases from the model and the actual numbers are close with a variance of only 0.05%. The week two is also seeing a variance of -0.46% and week 3 & 4 at -1.12% and -2.09 % respectively. . The trend of variance is almost the same as that for the country.

REFERENCES

- [1] Covid data source - <https://api.covid19india.org/>
- [2] <https://en.wikipedia.org/>
- [3] <https://machinelearningmastery.com/>
- [4] <https://www.tutorialspoint.com/r/>
- [5] <https://www.ncbi.nlm.nih.gov/>
- [6] <https://towardsdatascience.com/>
- [7] <https://stattrek.com/regression/>
- [8] <https://statisticsbyjim.com/glossary/>
- [9] <https://www.investopedia.com/>



Suraj Jayasimha is currently pursuing his IB Diploma at Dhirubhai Ambani International School with high level Mathematics analysis and approaches , Physics and Chemistry . He is an avid computer programmer with skills in multiple languages like Python , R programming , Java , C++ etc . He is an app developer as well having developed an app "SAMAAJ" which supports the unorganised sector to help them get jobs especially during this covid times. The app "SAMAAJ" is downloadable in both IOS and Android .