

Indo-Pak Comparison Of COVID-19 By ARIMA Model: A Secondary Analysis

Osama Ajaz

FAST National University

Muhammad Irfan (✉ irfanzafar892@gmail.com)

Liaquat National Hospital and Medical College

Muhammad Amjad

FAST National University

Ayesha Siddiqa

Research Mentor

Research Article

Keywords: ARIMA modeling, COVID-19, Forecasting, Subcontinent

DOI: <https://doi.org/10.21203/rs.3.rs-394938/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

INDO-PAK COMPARISON OF COVID-19 BY ARIMA MODEL: A SECONDARY ANALYSIS

Osama Ajaz¹, Muhammad Irfan², Muhammad Amjad³, Ayesha Siddiq⁴

1. Osama Ajaz
Lecturer, FAST-National University, Karachi, Pakistan
PhD-Statistics, (In progress), MS-Statistics & Scientific Computing
osama.ajaz@nu.edu.pk
2. Muhammad Irfan
Bio-Statistician-Liaquat National Hospital and Medical College, Karachi, Pakistan
MBA (Health Management), B.S (Statistics)
Irfanzafar892@gmail.com
3. Muhammad Amjad
Assistant Professor, FAST National University Karachi
Ph.D.-Mathematics (in Progress), M.Sc. (Industrial & Business Mathematics)
muhammad.amjad@nu.edu.pk
4. Ayesha Siddiq
Data Analyst- Research Mentor
B.S (Statistics)
siddiq516@gmail.com

Correspondence

Muhammad Irfan.
Bio-Statistician, Department of Statistics
Liaquat National Hospital and Medical College
National Stadium Road, Karachi
Email: irfanzafar892@gmail.com
Cell: 0346-3073572

Abstract

Background: Historically, the world has faced and recovered from many pandemics. The most recent global pandemic facing the entire world is Covid-19. India and Pakistan can be considered to be in the same phases of development and health spending relative to their GDP and also have similar climatic conditions. The main aim of the study is to forecast cumulative cases and deaths in Pakistan and India, which will be helpful for policy makers to plan accordingly.

Methods: The data set was obtained from the World Health Organization (WHO) website (<https://covid19.who.int>). The time period we have considered since the first corona related case and death were observed in both countries. The dataset for Pakistan covered the time period from 28th February 2020 to 28th February 2021 and for India 30th January to 28th February 2021. The Auto-Regressive integrated moving average (ARIMA) model was applied for forecasting using R-package.

Results: Our results forecasted that cumulative COVID-19 cases at the end of June 2021, at the end of September 2021 and at the end of December 2021 will be 13065792, 14704450, and 16481122 respectively in India while for Pakistan, we forecasted that at the end of June 2021, at the end of September 2021 and at the end of December 2021 will be 746963.5 873557.3 and 999766.5 respectively.

Cumulative deaths were also forecasted for Pakistan and India. We predicted cumulative deaths as at the end of June 2021 at the end of June 2021, at the end of September 2021 and at the end of December 2021 will be 170586.5, 181153.4 and 192017.5 respectively in India while for Pakistan, we forecasted that cumulative deaths at the end of June 2021, at the end of September 2021 and at the end of December 2021 will be 17890.98, 21825.26 and 25849.4 respectively.

Conclusion: Corona related cumulative cases and deaths are on the rise in both countries. The pandemic situation in India is worse than in Pakistan nevertheless both countries are at high risk. There is a sudden increasing pattern in the number of corona related cases in both countries.

Key Words: ARIMA modeling, COVID-19, Forecasting, Subcontinent

Background

The world has historically faced and recovered from many pandemics. The most recent global pandemic that the whole world is facing is Novel Coronavirus – Covid-19. The Virus has spread almost in every country at a rapid pace after initiating from China¹. Although the virus has a low mortality rate, it is highly contagious due to which the response of the country dealing with the Covid-19 pandemic can be termed as a significant determinant of the consequences of this disease². There are many articles for different countries around the world in regard to their responses to Covid-19. The current article has been aimed at sketching a comparison between the prevalent situation of Covid-19 in India and Pakistan which has resulted from their response to the pandemic.

Both the countries, India and Pakistan, can be regarded as being in the same phases of development and expenditure on healthcare with respect to their GDPs and they also have similar climatic condition³. India and Pakistan had their first cases within a month of each other as Pakistan had its first case on 26th February 2020 in Karachi whereas India got its first case on 30th January 2020 in Kerala³. Comparing the response of the two countries, it is known that India was quick to ban travels starting from early February and the country also issued the orders of a rapid lockdown in March for the purpose of containing the spread of Coronavirus⁴. In contrast, Pakistan responded slowly to the threat of Covid-19 as it's Prime Minister Imran Khan did not see the virus as a potential life-endangering threat and addressed the country downplaying the danger of Covid-19 stating that only old age people are affected by it⁵. He also encouraged the people to return to their work and mentioned that most cases of the virus can be resolved through self-quarantine. Nevertheless, realizing the danger of the virus, lockdown in Pakistan was imposed in late March.

Moving on, it can be observed that the fast response of India to the Covid-19 pandemic was significantly effective in containing the danger and destruction of the virus as in May, the reported cases of India were one-third and mortality was almost half than that of Pakistan⁶. If a comparison is drawn between the stimulus package of India and Pakistan, it can be noted that the stimulus of Pakistan was much larger than that of India in terms of GDP percentage as Pakistan offered a stimulus of 2.5% of GDP and India offered a stimulus of 0.8% of GDP⁷. However, the economic situation of Pakistan has been estimated to be more affected than that of India as according to IMF⁸, the economy of Pakistan is estimated to contract by 1.5% whereas India's economy is estimated to grow by 1.9% in the year 2020. Hence, it can be stated by comparing both the countries that India dealt with the Covid-19 pandemic in a quick and effective manner while Pakistan was slow in taking actions due to which India's situation is comparatively that of Pakistan.

Methods

Data Collection Methodology

The data set for this research is obtained from the World Health Organization (WHO) online repository (<https://covid19.who.int/>). The time period we have considered since the first corona related case and death were observed in both countries. The dataset for Pakistan covered the time period from 28th February to 28th February 2021 and for India 30th January to 28th February 2021. The following table 1 provides the summary of corona infected cases and deaths in both countries up to 28th February 2021.

Table 1: Summary of corona infected cases and deaths in both countries (Jan. 2020 -- Feb. 2021)

Country	Max Cases on a single day	Max Deaths on a single day	Cumulative cases	Cumulative deaths
Pakistan	13709	313	578797	12837
India	194121	2410	11096731	157051

The highest number of individual cases (13709) was recorded on Sunday, 14th June in Pakistan. The individual cases keep on declining after this date. In India, the highest number of individual cases (194121) was recorded on Saturday, 12 September 2020 with an overall increasing trend. The highest number of individual deaths (313) per day was observed on Monday 21th November 2020 in Pakistan while 2410 deaths were observed on Saturday 12 September 2020 in India. The above table 1 clearly reveals the fact that the number of cases and deaths in India keeps on high relative to Pakistan.

Data Analysis

The R-package “Forecast” is used to forecast the number of corona infected cases and deaths in Pakistan and India. R is an open-source programming platform that is used for analysis and model development ^{9, 10}. The Auto-Regressive integrated moving average (ARIMA) model, available in the package, is run on our dataset. The analysis and results are given in the next sections. The ARIMA model is the mixture of autoregressive and moving average processes and is specified by three parameters (p, d, q); where ‘p’ represents the order for an autoregressive part, ‘d’ represents the degree of integrated order difference I(d), and ‘q’ represents the order for Moving average part.

The ARIMA model is represented by the following linear equation:

$$Y_t = \delta + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \beta_1 e_{t-1} + \beta_2 e_{t-2} + \dots + \beta_q e_{t-q} + e_t \quad (1)$$

The above equation is assumed to be a non-seasonal time series model. See ^{11, 12} for seasonal components and other details about ARIMA modelling.

Model Estimation and Analysis

In this study, four separate ARIMA models are run for countries Pakistan and India. Two is for the number of corona infected cumulative cases and two for the number of coroneae related cumulative deaths. Fig. 1 below is the graph of confirmed cumulative corona cases in Pakistan and India for the period January 2020 to February 2021.

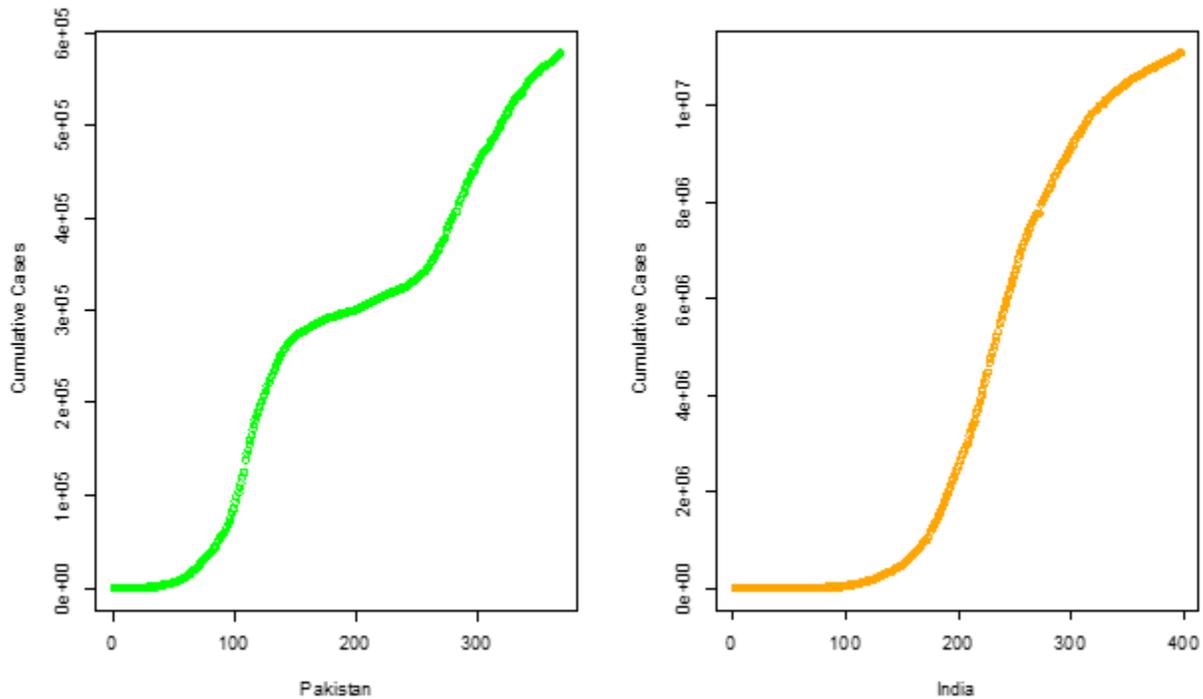


Figure-1: Number of corona infected cumulative cases in Pakistan and India

The above Fig. 1 represents the graph of cumulative corona infected cases in both countries. The corona infected cases are growing faster in India than Pakistan that is clearly seen in the above figure. The same pattern is observed in the variable of death in both countries.

Results

The Forecast package in R is used to estimate the four ARIMA models for the number of corona related cumulative cases and deaths. Table 2 represents the best-fitted models based on minimum AIC and BIC criterion. Table 3 represents the estimates of accuracy parameters for the best-fitted models.

Table 2: Best ARIMA Model Estimates based on minimum AIC and BIC

Variable	Model	AR(1)	MA(1)	MA(2)	MA(3)	MA(4)	AIC	BIC
Cases (Pak.)	ARIMA (1,2,3)	0.7770 (0.0975)	-1.9828 (0.1140)	1.3074 (0.1730)	-0.2527 (0.0837)		6292.39	6311.92
Deaths (Pak.)	ARIMA (1,2,2)	0.4703 (0.2279)	-1.5075 (0.1980)	0.6161 (0.1616)			2549.48	2564.85
Cases (India)	ARIMA (0,,2,2)		-1.3179 (0.0481)	0.4456 (0.0463)			760.31	772.24
Deaths (India)	ARIMA (0,2,4)		-1.1528 (0.0538)	0.2291 (0.0786)	-0.0544 (0.0766)	0.1148 (0.0515)	1457.53	1476.83

Table 3: Model Accuracy Parameters

Variable	Model	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Cases (Pak.)	ARIMA (1,2,3)	13.66375	971.6241	562.3845	0.8094183	1.811358	0.3575686	-0.009780305
Deaths (Pak.)	ARIMA (1,2,2)	0.1651041	27.49087	15.38458	0.4986578	1.265683	0.4147305	-0.01855311
Cases (India)	ARIMA (0,2,2)	-2317.954	18688.18	9063.402	0.3903549	1.940577	0.3226216	0.00709891
Deaths (India)	ARIMA (0,2,4)	-21.95967	270.8032	137.7186	0.3364716	1.762918	0.3086722	-0.1282017

The 306 days forecast graph (1st March 2021 to 31st December 2021) of cumulative corona cases and death are given in Figure 2. The grey shaded area represents 80% confidence interval and white shaded area represents a 90% confidence interval in the figure. The forecast graph reveals that the number of cumulative corona cases could reach 999767 in Pakistan and 16481122 in India up to 31st December 2021. The number of cumulative deaths could reach to 25849 in Pakistan and 192017in India. The forecasted values are given in the table 4 below.

Table 4: Cumulative Forecasted values from March 2021 to December 2021

	India		Pakistan	
Date	Cumulative Cases	Cumulative Deaths	Cumulative Cases	Cumulative Deaths
31-Mar-21	11574571	160427.7	621300.1	14102.22
30-Apr-21	12052311	163744.4	662781.6	15338.88
31-May-21	12560234	167205.1	705588.9	16629.8
30-Jun-21	13065792	170586.5	746963.5	17890.98
31-Jul-21	13602921	174113.8	789666.8	19205.82
31-Aug-21	14155247	177675	832321.8	20531.86
30-Sep-21	14704450	181153.4	873557.3	21825.26
31-Oct-21	15287380	184781	916124.4	23171.79
30-Nov-21	15866653	188323.6	957279	24484.13
31-Dec-21	16481122	192017.5	999766.5	25849.4

The standardized Q-Q plot and histograms for all models reflect normal curve distribution. The Ljung-Box test for all models indicates independence of residuals at 1% level of significance.

The accuracy of the forecast is assessed by the following formula:

$$\text{Accuracy} = (100 - \text{MAPE})\%$$

The best fitted models ARIMA(1,2,3), ARIMA(1,2,2), ARIMA(0,2,2) and ARIMA(0,2,4) have accuracy of 98.188 %, 98.734%, 98.059%, and 98.237% respectively.

Forecast Comparison Graph

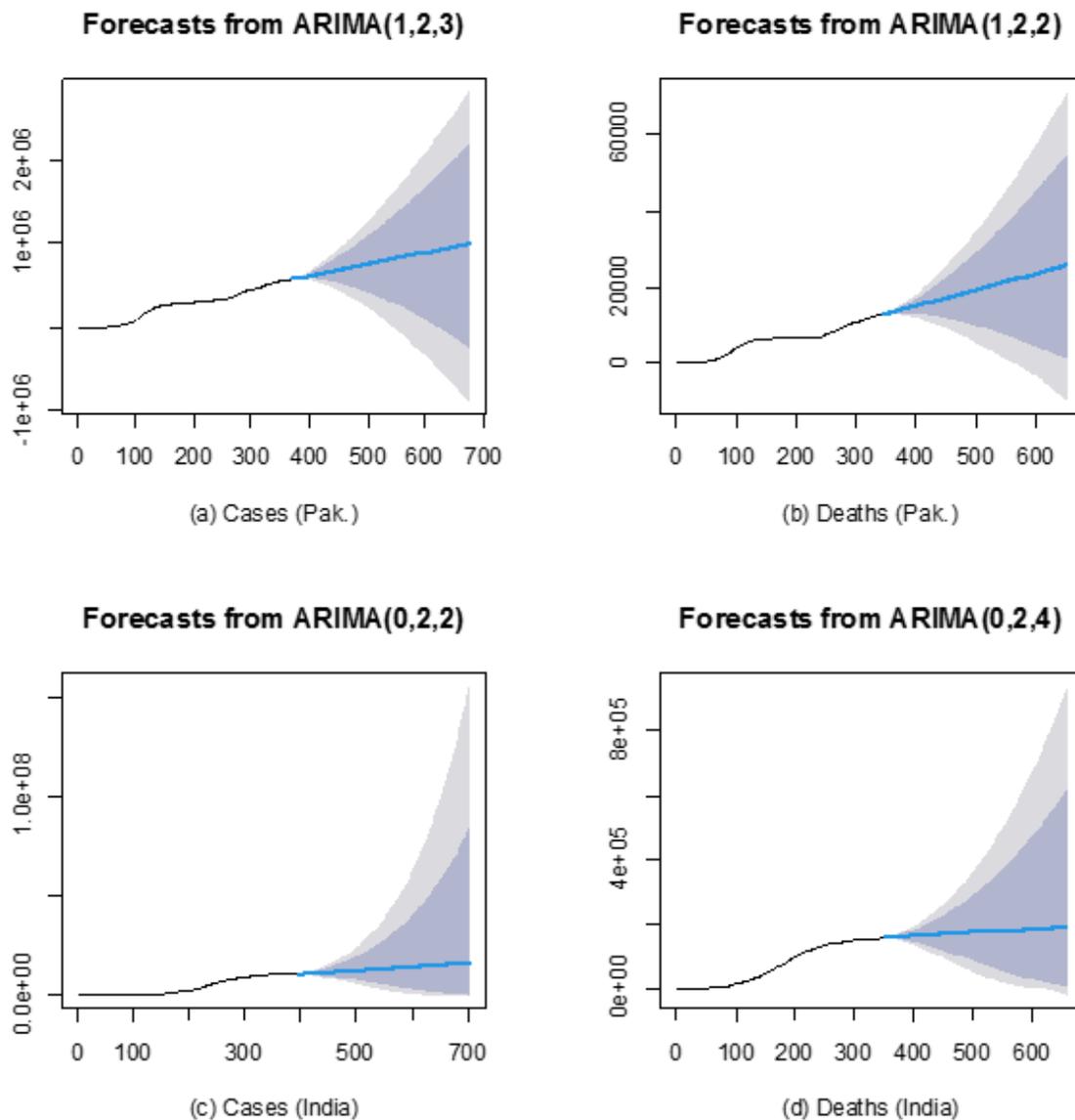


Figure 2: Forecast graphs for the number of cumulative corona cases and deaths for Pakistan and India.

Discussion

We used the WHO COVID-19 dataset for Pakistan and India. The purpose is to forecast the number of cumulative corona related cases and deaths in both countries. The AUTOARIMA technique is used, that is available in the R package “forecast”. The forecast model ARIMA is simple to use and more predictive than other regression models. It is due to the fact that the ARIMA model only considers the past lag values of dependent **variables**¹² such that cumulative cases and deaths in this research.

After the emergence of a number of cases and deaths due to coronavirus, both governments imposed lockdown to make people at home. All markets, marriage halls, restaurants, schools and universities were closed to avoid overcrowding and make lockdown effective. Schools and universities are encouraged to adopt e-learning methods. Several quarantine centres were established to treat corona infected patients. Besides the government's effort, several non-government organizations (NGOs) provided food and some amount door-to-door to needy families.

Currently, the citizens in both countries are being encouraged to take more protective measures and practice social distancing to control the COVID-19 **pandemic** ¹³⁻⁻¹⁵. Social media campaigns are being run to educate the public about this pandemic. Consequently, the spread of coronavirus is being expected to be decreased soon in the **future**.¹⁶

After this analysis, we are of the view that COVID-19 pandemic situation in Pakistan is much better than India, nevertheless, both governments must take serious measures to cope with COVID-19 pandemic. Effective lockdown, self-isolation and public awareness are necessary to further control this pandemic.

Conclusion

This research found that corona related cumulative cases and deaths are on the rise in both countries. The pandemic situation in India is worse than in Pakistan nevertheless both countries are at high risk. There is a sudden increasing pattern in the number of corona related cases in both countries. Both governments must impose effective policies to control this pandemic.

Declarations

Ethics approval and consent to participate

There is no need of ethical approval and consent as we use secondary data.

Consent for publication

Not Available

Availability of data and materials

Data is available at <https://covid19.who.int/>.

Competing interests

None

Funding

None

Authors' contributions

Osama Ajaz and Muhammad Irfan did statistical analysis.

Osama Ajaz and Muhammad Ahmed did manuscript writing.

Osama Ajaz, Muhammad Irfan and Ayesha Siddiqa did review and final approval of manuscript.

Acknowledgement

This study made possible because of data from world health organization (WHO).

References

1. Deshwal VK. COVID 19: A comparative study of Asian, European, American continent. Int J Sci Res Enginee Dev. 2020;3(2):436-40.
2. Das I, Bhattacharjee A. Trend analysis of Covid-19 cases in Indian subcontinent: Prediction comparison and psycho-social interventions. Indian Journal of Health & Wellbeing. 2020 Jan 1;11.
3. Health spending less than one percent of GDP for decades [Internet]. Pakistantoday.com.pk. 2020 [cited 8 September 2020]. Available from: <https://www.pakistantoday.com.pk/2019/07/01/health-spending-less-than-one-percent-of-gdp-for-decades/>
4. Raza S, Rasheed MA, Rashid MK. Transmission Potential and Severity of COVID-19 in Pakistan.
5. Shah A, Choksi N. Did India handle Covid crisis better or Pakistan? The answer lies in these 4 tests [Internet]. ThePrint. 2020 [cited 8 September 2020]. Available from: <https://theprint.in/opinion/did-india-handle-covid-crisis-better-or-pakistan/417098/>
6. Pasha A. Should Pakistan instate a travel ban against coronavirus emergent regions? | The Express Tribune [Internet]. The Express Tribune. 2020 [cited 8 September 2020]. Available from: <https://tribune.com.pk/article/94617/should-pakistan-instate-a-travel-ban-against-coronavirus-emergent-regions>
7. NEWS, B. and News, I., 2020. IMF India GDP Forecast: IMF Projects India's Growth Rate At 1.9% In 2020, Forecasts Global Recession Due To Covid-19 | India Business News - Times Of India. [online] The Times of India. Available at: <<https://timesofindia.indiatimes.com/business/india-business/imf-projects-indias-growth-rate-at-1-9-in-2020-forecasts-global-recession-due-to-covid-19/articleshow/75142899.cms>> [Accessed 8 September 2020].
8. Pakistan records highest single day spike in COVID-19 infections [Internet]. The Economic Times. 2020 [cited 8 September 2020]. Available from: <https://economictimes.indiatimes.com/news/international/world-news/pakistan-records-highest-single-day-spike-in-covid-19-infections/articleshow/75507645.cms>
9. Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, Petropoulos F, Razbash S, Wang E, Yasmeeen F (2020). *forecast: Forecasting functions for time series and linear models*. R package version 8.13, <https://pkg.robjhyndman.com/forecast/>.
10. HYNDMAN, Rob J.; KHANDAKAR, Yeasmin. Automatic Time Series Forecasting: The forecast Package for R. Journal of Statistical Software, [S.l.], v. 27, Issue 3, p. 1 - 22, july 2008. ISSN 1548-7660. Available at: <<https://www.jstatsoft.org/v027/i03>>. Date accessed: 20 oct. 2020. doi:<http://dx.doi.org/10.18637/jss.v027.i03>.
11. 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.

12. Box G. (2013) Box and Jenkins: Time Series Analysis, Forecasting and Control. In: A Very British Affair. Palgrave Advanced Texts in Econometrics. Palgrave Macmillan, London.
https://doi.org/10.1057/9781137291264_6
13. Ali MY, Bhatti R. COVID-19 (Coronavirus) Pandemic: Information Sources Channels for Public Health Awareness. *Asia Pacific Journal of Public Health*. 2020;32(4):168-169.
doi:10.1177/1010539520927261
14. Patrikar SR, Kotwal A, Bhatti VK, et al. Incubation Period and Reproduction Number for Novel Coronavirus 2019 (COVID-19) Infections in India. *Asia Pacific Journal of Public Health*. August 2020. doi:10.1177/1010539520956427
15. Abid K, Bari YA, Younas M, Tahir Javaid S, Imran A. Progress of COVID-19 Epidemic in Pakistan. *Asia Pacific Journal of Public Health*. 2020;32(4):154-156.
doi:10.1177/1010539520927259

Figures

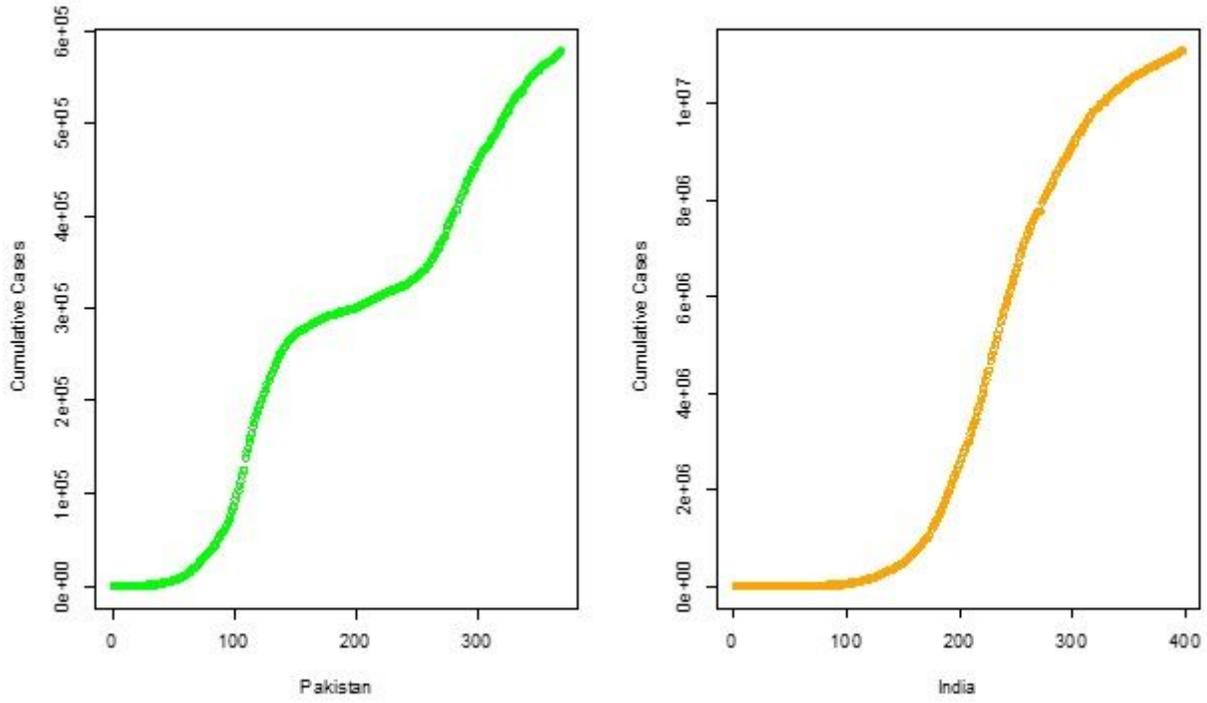
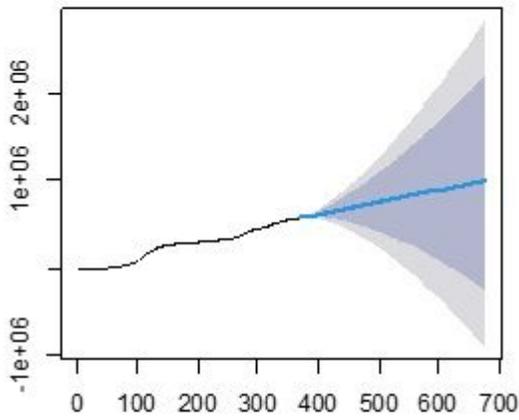


Figure 1

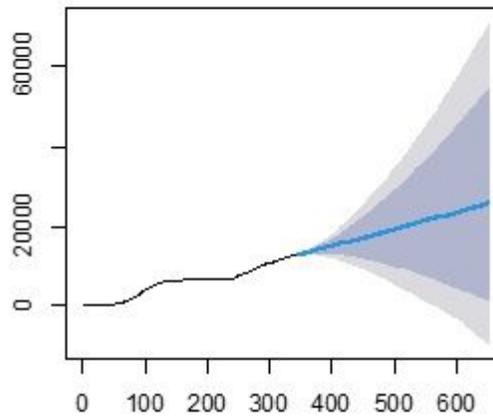
Number of corona infected cumulative cases in Pakistan and India

Forecasts from ARIMA(1,2,3)



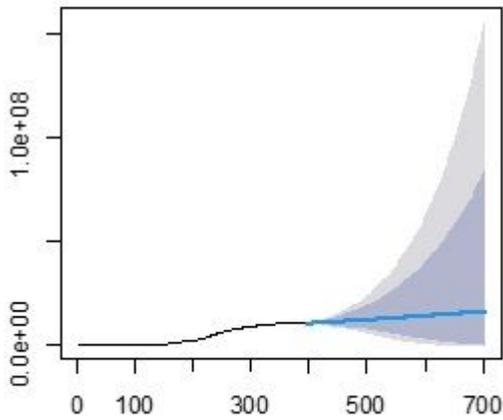
(a) Cases (Pak.)

Forecasts from ARIMA(1,2,2)



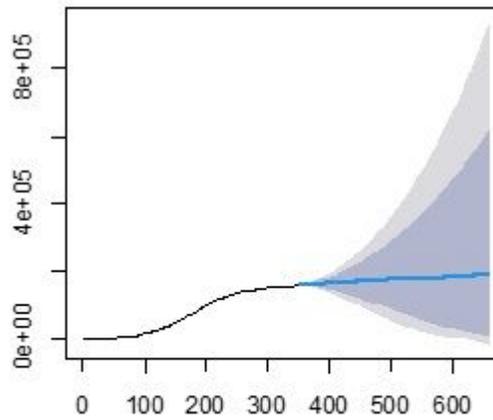
(b) Deaths (Pak.)

Forecasts from ARIMA(0,2,2)



(c) Cases (India)

Forecasts from ARIMA(0,2,4)



(d) Deaths (India)

Figure 2

Forecast graphs for the number of cumulative corona cases and deaths for Pakistan and India.