



Investigation and Forecast of COVID-19 Wave-II: A Case Study of Pakistan

Sher Akbar^{1*}, Shahid Akbar², Ali Raza³, and Muhammad Rovidad⁴

¹ Department of Statistics, Quaid-i-Azam University Islamabad, PAKISTAN.

² International Institute of Islamic Economics, International Islamic University Islamabad, PAKISTAN.

³ Department of Management Sciences, University of Haripur, Khyber Pakhtunkhwa, PAKISTAN.

⁴ Department of Economics, Bacha Khan University, Charsadda, Khyber Pakhtunkhwa, PAKISTAN.

*Corresponding Author (Tel: +92-3-349457086, Email: sakbar@stat.qau.edu.pk)

Paper ID: 12A5T

Volume 12 Issue 5

Received 23 December 2020

Received in revised form 05
March 2021

Accepted 15 March 2021

Available online 22 March
2021

Keywords:

COVID-19 second wave;
Confirmed cases; ARIMA;
Covid forecasting; Covid
model; Covid-19 infection
prediction; Covid-19
forecast; Epidemic trend;
Coronavirus projection;
Pandemic forecasting.

Abstract

This study has investigated and forecasted the number of confirmed, deaths, active, and recovered cases of COVID-19 for Wave-II using the ARIMA model in Pakistan. An exponential growth forecast for all the series has been observed under the fitted model for 1 October 2020 to 19 December 2020. Based on our empirical results of forecasted models, the numeral COVID-19 confirmed cases will increase by 1.26 times with a 95% confidence interval at the end of January 2021. Similarly, the number of deaths due to COVID-19 will increase by 1.34 times with a 95% confidence interval until January 2021. Moreover, the number of active cases of COVID-19 will increase by 0.11 times with a 95% confidence interval till the end of January 2021. Besides these, Pakistan has an 89.2% rate of recoveries which will increase by 1.38 times until January 2021. Finally, results show an alarming situation for January 2021 in Pakistan.

Disciplinary: Biostatistics, Statistics Modeling & Forecasting, Public Health.

©2021 INT TRANS J ENG MANAG SCI TECH.

Cite This Article:

Akbar, S., Akbar, S., Raza, A., and Rovidad, M. (2021). Investigating and Forecasting COVID-19 Wave-II: A Case Study of Pakistan. *International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies*, 12(5), 12A5T, 1-9. <http://TUENGR.COM/V12/12A5T.pdf> DOI: 10.14456/ITJEMAST.2021.104

1 Introduction

In the last month of 2019, a virus called COVID-19 (Coronavirus) was identified in Wuhan city believing that it is originated in the human body from a wild animal traded at a rushed marketplace (Cohen, 2020). As a result of this virus, the number of infected people arose very sharply in Wuhan city and a few days spreading into other cities of China and the rest of the world,

thus on 11 March 2020, the World Health Organization (WHO) declared COVID-19 as pandemic causing a huge number of deaths, (Organization, 2020). The most notable symptoms caused by this virus include; fever, flu, cough, diarrhea, and myalgia. It is observed to be highly communicable through respiratory droplets (a key source of spread) from the infected patients and it can also spread through human contact. The virus can sustain its existence up to 48 hours on a contacted surface. Due to the human-to-human transition, this virus is observed to grow exponentially at its very early stage and causes a much fatality rate among young children and elderly age group above 60. Various precautionary measurements like closing borders, cancellation of domestic and foreign flights, suspension of educational institutes and community services, etc. are taken worldwide in command to regulate the spread of the disease. Even a few of the countries imposed curfew costing huge economical loss to control or minimize the transmission of disease.

Till 30th June 2020 (Wave-I) worldwide, a total number of infected patients are observed as 9,843,073 with 495,760 deaths since December 2019 (WHO, 2020). In Pakistan, two infected persons due to pandemic COVID-19 were recorded on 26th February 2020, and later on, within two days the number of other cases was reported in different areas around the country. As a result, the number of confirmed cases increased exponentially and recorded as 12,775, 1,065, 1,470, 26,115, 81,985, 75,501, and 10,426 in Islamabad (ISL), Azad Jammu and Kashmir (AJK), Gilgit Baltistan (GB), Khyber Pakhtunkhwa (KP), Sindh (SIN), Punjab (PUN), and Baluchistan (BAL) till 30 June 2020, respectively (NIH, 2020). Overall, among these cases, 209,337 and 4,304 were recorded as a number of confirmed cases and deaths respectively in Pakistan. To make prevention of this pandemic disease initially government took a lot of measures like the closure of borders with its neighboring countries, closure of educational institutes, closure of domestic and international flights, closure of markets and community services, etc. However, due to Pakistan's struggling/suffering economy to stay afloat, the outbreak of the pandemic rigorously squeezed the country's economy and strapped it at the edge of bankruptcy. As a result of the pandemic outbreak, the poverty rates will around 23.4% to 59% bringing 125 million people below the poverty line (Iqbal, 2020). In view of these situations, after approximately 60 days of lockdown across the country government declared to lift the lockdown, though the number of cases was increasing on daily basis, to keep track of the economy to its earlier stage.

In October 2020 the number of infected cases due to COVID-19 started increasing again, thus identified as Wave-II in this paper, and the number death toll also started climbing. Till date 19th December 2020, the number of confirmed cases are noted as 36,117 in ISL, 7,961 in AJK, 4,822 in GB, 54,948 in KP, 205,103 in SIN, 131,428 in PUN, and 17,909 in BAL.

In order to forecast the COVID-19 dataset, several studies suggesting different models and techniques to model and get an optimal forecast of the nature of the virus (COVID-19) are conducted with uncertainties. Among these studies, Generalized Logistic Model (GLM) determines the exponential behavior of pandemic in China (Roosa et al., 2020). Similarly, Fractional Time Delay Dynamic (FTDD) system was identified better with better predictive performance (Chen et

al., 2020). Moreover, (Koczkodaj et al., 2020) projected the forecast of Covid-19 utilizing a simple exponential curve. In Reference (Maleki et al., 2020), AR models constructed based on a scale combination of normal distributions, named TP–SMN–AR models, are used to forecast the COVID-19 across the world. Further, a very devastating situation is predicted for the USA and entire Europe in April 2020 based on findings of (Kumar et al., 2020). So an optimal prediction is the most enlightened stage to make stronger strategies to stop the pandemic from getting worse. The present study uses the ARIMA model on the presented data to forecast the four series; a number of confirmed, deaths, recoveries, and active cases in Pakistan for January 2021. This will help government officials and policymakers to use the better forecast performing model in the future and modify and strengthen the planned strategies to maximize the prevention measurements.

2 Techniques for Statistical Analysis and Data Sources

In view of the study of Chen et al. (2008), Autoregressive Integrated Moving Average (ARIMA) model performs better as compared to exponential models with respect to fitting and forecast accuracy. Also, this model is used to observe and forecast the growth stage of a disease in different studies in past. Liu et al. (2011) predicted the prevalence of hemorrhagic fever through renal disease in China with the ARIMA model. Zhang et al. (2014) forecasted the data of epidemiological surveillance in China by using the ARIMA along with a support vector machine (SVM). Molina et al. (2018) predicted the bovine genital campylobacteriosis (BGC) and bovine trichomoniasis (BT) data using ARIMAX and ARIMA in Argentina, (Anokye et al., 2018) utilized ARIMA to forecast the growth nature of malaria in Kumasi. Benvenuto et al. (2020) used ARIMA to inspect the forecast projection of Covid-19. Hence, we use the ARIMA model for the purpose to forecast the numeral deaths, recoveries, active, and conformed cases.

Moreover, our data of COVID-19 (deaths, recoveries, confirmed cases, active cases) is in cumulative nature showing exponential growth, therefore, ARIMA is the most appropriate model for fitting and short-term forecasting for the given period (i.e. till the end of July). Partial Autocorrelation Function (PACF) and Autocorrelation Function (ACF) are applied to detect the lags order 'q' and 'p' for Moving Average (MA) and Autoregressive (AR) terms respectively and also to determine the differencing order 'd' to assess the model fit. Furthermore, Akaike Information Criterion is obtained to assess the model fitting by selecting the model having the minimum value of AIC. "t-series, zoo, and forecast" R-libraries developed by Quick-(2020) were used to make the statistical analysis.

The daily data is obtained from the government-owned official dashboard for a number of COVID-19 patients, the number of deaths, recovered, confirmed and active cases range from 1 October 2020 to 19 December 2020 (NIH, 2020). The data from 20 December 2020 to 31 January 2021 will be forecasted based on available data. The NIH carries health and biomedical research to manufacture vaccines and in this regard, it collects and regularly reports COVID-19 data on daily basis. The reported data includes the number of active cases, deaths, and recoveries from all four provinces (KP, Sindh, Baluchistan, and Punjab) including AJK and GB in the country.

3 Result and Discussion

Nowadays, the whole world is facing the second wave of COVID-19 pandemic situation somewhere, worst, and somewhere with good conditions. However, there is an exponential growth in the confirmed cases in Pakistan day-by-day in Wave-II.

In this regard, Table 1 demonstrates the Wave-I cumulative numeral confirmed cases in Pakistan on 8 June 2020 were 103,671, mostly reported in Punjab (PUN), followed by Sindh (SIN) 38,108 cases. Other regions also revealed an exponential increase in the number of deaths, confirmed, active, and recovered cases. There were 2,067 deaths due to COVID-19 with 34,355 got recovered from the pandemic disease. For Wave-II, Table 2 gives similar information for all regions.

Table 1. Region-wise overview of COVID-19 Cases from 26 February 2020 to 30 June 2020 (Wave-I)

Date 2020	PUN	SIN	KP	BAL	GB	AJK	ISL	Total Cases	Total Deaths	Total Recoveries	Total Active Cases
26 Feb	-	2	-	-	-	-	-	2	-	-	2
10 Mar	-	-	-	1	2	-	2	5	-	1	4
18 Mar	33	208	19	23	13	1	5	302	2	5	297
26 Mar	408	417	123	-	91	2	-	1,041	9	23	1,018
20 Apr	3,721	-	1,235	432	263	49	181	5,881	176	1,970	3,911
10 May	11,093	11,480	4,669	1,935	430	86	641	30,334	659	8,063	22,271
31 May	25,056	27,360	9,540	4,193	678	251	2,418	69,496	1,483	25,271	44,225
8 Jun	38,903	38,108	13,487	6,516	932	396	5,329	103,671	2,067	34,355	69,316
26 Jun	74,202	78,267	25,380	10,261	1,423	1,027	12,395	198,883	4,035	86,906	111,977

Table 2: Region wise overview of COVID-19 Cases from 15 July 2020 to 19 December 2020 (Wave-II).

Date 2020	PUN	SIN	KP	BAL	GB	AJK	ISL	Total Cases	Total Deaths	Total Recoveries	Total Active Cases
15 Jul	88,539	108,913	31,217	11,322	1,750	1,771	14,454	257,914	5,426	178,737	79,177
15 Aug	95,391	125,904	35,153	12,234	2,486	2,181	15,378	288,717	6,168	266,301	22,416
15 Sep	97,946	132,591	37,140	13,690	3,297	2,441	15,984	303,089	6,393	290,760	12,329
15 Oct	101,367	141,249	38,521	15,621	4,016	3,322	17,781	321,811	6,621	305,835	15,976
15 Nov	110,450	155,680	42,370	16,407	4,452	5,455	24,218	359,032	7,160	323,824	35,208
19 Dec	131,428	204,103	54,948	17,909	4,822	7,961	36,117	457,176	9,330	407,405	49,771

For estimation, we have used 2 months and 19 days of data from 1 October 2020 up to 19 December 2020 having 80 data points. Autoregressive Integrated Moving Average (ARIMA) is used to forecast the daily confirmed cases, deaths, recoveries, and active cases.

From Figure 1 it is clear that series does not possess any stationarity assumptions of having constant mean and variance. Augmented Dicky-Fuller (ADF) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests were implemented on the data to check the stationarity of the series. Due to exponential growth in the data, we took the difference of two lags ($d=2$) for each series to meet the stationarity assumption. ARIMA (0,2,1) is fitted to confirmed cases series, ARIMA (0,2,1) is fitted to the number of deaths series, ARIMA (0,2,1) is fitted to the series of recovered patients from COVID-19 while, ARIMA (0,2,2) is fitted on the active cases series respectively.

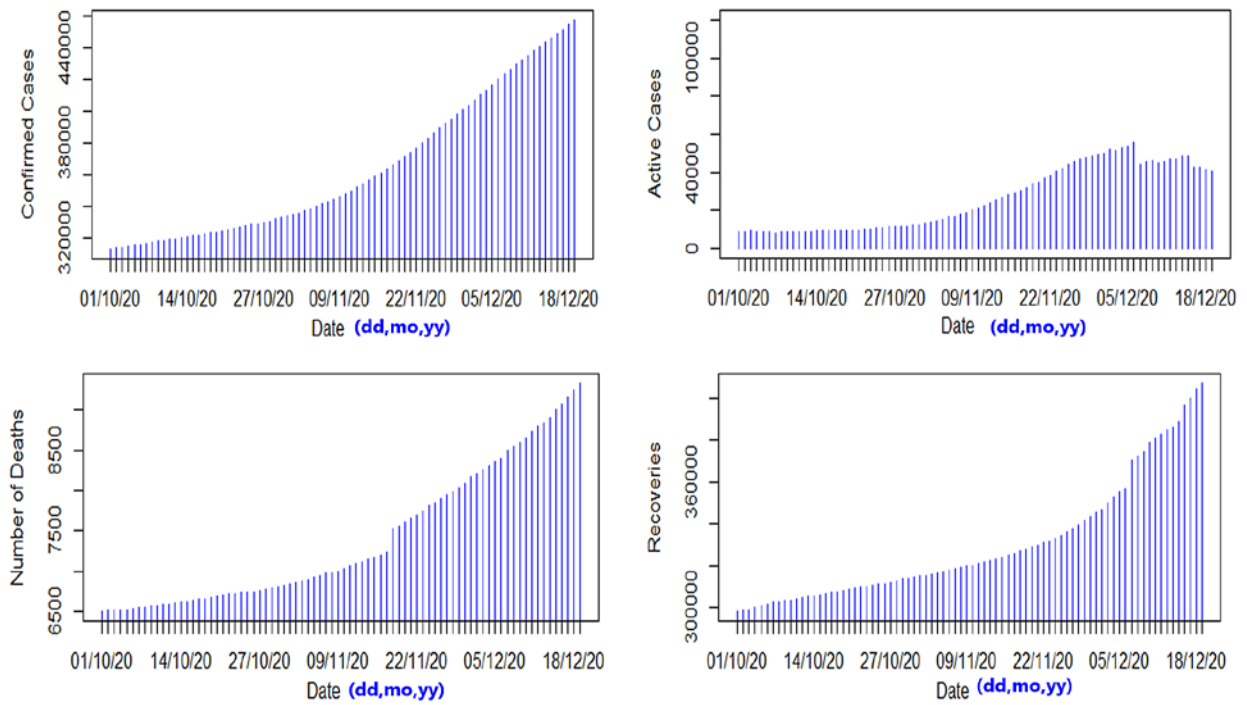


Figure 1: COVID-19 (Wave-II) Number of Active, Confirmed, Deaths, and Recovery cases in Pakistan.

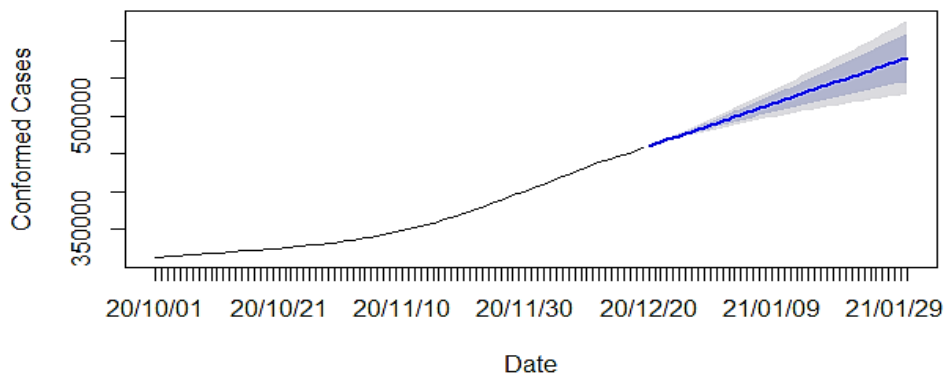


Figure 2: Forecast for the confirmed cases of COVID-19 for January in Pakistan (yy-mo-dd).

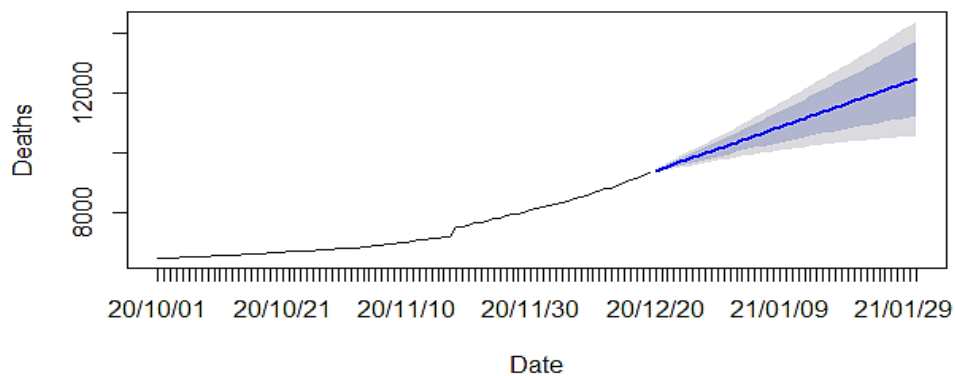


Figure 3: Forecast for the death cases of COVID-19 for January in Pakistan (yy-mo-dd).

Finally, results from the empirical analysis revealed that confirmed cases show a rapid increase with 1.26 times as compared to current cases till the end of January 2021, Figure 2. Similarly, Figure 3 shows the forecasted number of deaths may increase 1.34 times exponentially, by the end of January 2021. The forecasted model for the number of recoveries from COVID-19 shows an increase of 1.38 times more until January 2021 Figure 4. However, active cases show

exponential growth in December 2020, with an increase of 0.11 times more in January 2021 Figure 5. Note that all these forecasted values were obtained with a 95% confidence Interval (C.I).

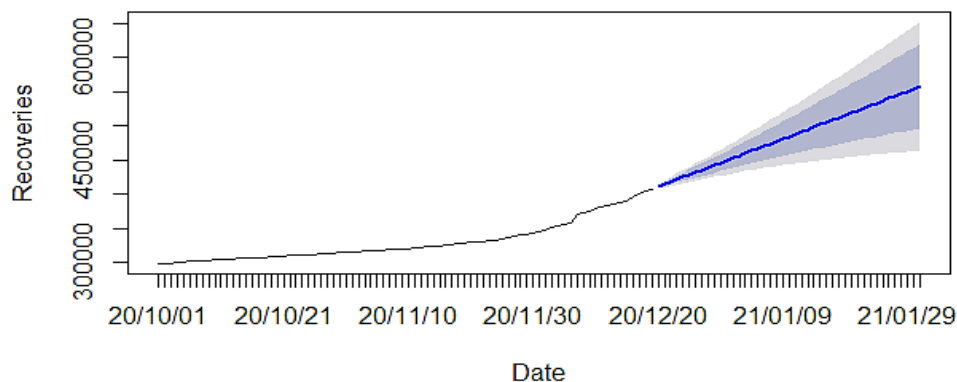


Figure 4: Forecast for the recoveries of COVID-19 for January in Pakistan (yy-mo-dd)

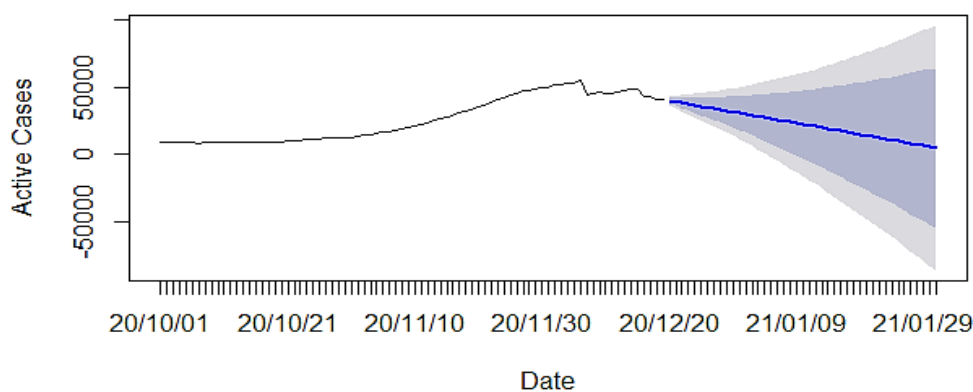


Figure 5: Forecast for the active cases of COVID-19 for January in Pakistan (yy-mo-dd)

Based on the assumption of our data accuracy, we can say that our data is more accurate as recorded and published on daily basis by (NIH, 2020) and the government officially. Thus we used the classical time series forecasting approach ARIMA, which is based on Box-Jenkins methodology (Box et al., 2015). Assuming the accuracy of data and the same trend followed by the data, we used a conventional model selection criteria that are Akaike Information Criteria (AIC). However, our study is founded on some conventions and if these conventions are not fulfilled in the future, it might lead us to great accuracy loss in the forecast. Moreover, in such epidemiological studies, uncertainty is always associated with forecasting and prediction problems thus, our study may not continue to be true in the future.

4 Conclusion

This is a formal study for forecasting the numeral COVID-19 confirmed cases, deaths, recoveries, and active cases in Pakistan. Our empirical findings suggested that there might be an exponential progression in the number of confirmed cases, if and only if the same trend continues in the country. The fatality of COVID-19 is observed from the number of deaths as it is currently increasing day by day. The number of deaths may increase up to 12,460 with a 95% confidence interval by the end of January 2021. Similarly, the rate of recovered patients from COVID-19 in Pakistan is almost 89.2% which is now increasing exponentially at 1.38 times and could be

increasing by 2.5 times more, at the end of January 2021. However, due to the uplifting of lockdown in Pakistan, the numeral confirmed cases are growing with a higher rate as linked with recovery rate with 95% confidence interval (528,483 to 625,666) shows COVID-19 rapid spread into different regions till the end of January 2021.

Pakistan is a developing country, with a lack of medical facilities resulting from the more fatal situation until and unless an antidote or medication is industrialized to treat the patients of COVID-19. Thus, based on our empirical findings, it is highly recommended for the government and health officials, to take possible precautionary measures at an emergency level to stop the speedy growth of the COVID-19 pandemic. In addition to this, the public must also maintain social distance and usage of other such precautionary measures to ensure their healthy life.

5 Availability of Data And Material

Data can be made available by contacting the corresponding authors.

6 References

- Anokye, R., Acheampong, E., Owusu, I., & Isaac Obeng, E. (2018). Time series analysis of malaria in Kumasi: Using ARIMA models to forecast future incidence. *Cogent Social Sciences*, 4(1). <https://doi.org/10.1080/23311886.2018.1461544>
- Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. (2020). Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data in Brief*, 29, 105340. <https://doi.org/10.1016/j.dib.2020.105340>
- Box GE, Jenkins GM, Reinsel GC, L. G. (2015). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.
- Chen, P., Yuan, H., Shu, X. (2008). *Forecasting crime using the ARIMA model*. In: *Proceedings - 5th international conference on fuzzy systems and knowledge discovery*, 5. FSKD., 627-630. <https://doi.org/10.1109/FSKD.2008.222>.
- Chen, Y., Cheng, J., Jiang, X., & Xu, X. (2020). *The Reconstruction and Prediction Algorithm of the Fractional TDD for the Local Outbreak of COVID-19*. 1-8. <http://arxiv.org/abs/2002.10302>
- Cohen, J. (2020). Wuhan Seafood Market may not be Source of Novel Virus Spreading Globally. *Science*, 10. <https://doi.org/10.1126/science.abb0611> ISSN 0036-8075
- NIH (2020). COVID-19 Health Advisory Platform by Ministry of National Health Services Regulations and Coordination. <https://covid.gov.pk/>.
- Iqbal, N. (2020). COVID-19 in Pakistan: Caring for the Poor and Vulnerable. *PIDE COVID-19 Bulletin*, 1(1).
- Koczkodaj, W. W., Mansournia, M. A., Pedrycz, W., Wolny-Dominiak, A., Zabrodskaa, P. F., Strzałka, D., Armstrong, T., Zolfaghari, A. H., Dębski, M., & Mazurek, J. (2020). 1,000,000 cases of COVID-19 outside of China: The date predicted by a simple heuristic. *Global Epidemiology*, 2, 100023. <https://doi.org/10.1016/j.gloepi.2020.100023>
- Kumar, P., Kalita, H., Patairiya, S., Sharma, Y. D., Nanda, C., Rani, M., Rahmai, J., & Bhagavathula, A. S. (2020). Forecasting the dynamics of COVID-19 Pandemic in Top 15 countries in April 2020 through ARIMA Model with Machine Learning Approach. *MedRxiv*, April, 2020.03.30.20046227. <https://doi.org/10.1101/2020.03.30.20046227>
- Liu, Q., Liu, X., Jiang, B., & Yang, W. (2011). Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model. *BMC Infectious Diseases*, 11. <https://doi.org/10.1186/1471-2334-11-218>.

- Maleki, M., Mahmoudi, M. R., Wraith, D., & Pho, K. H. (2020). Time series modelling to forecast the confirmed and recovered cases of COVID-19. *Travel Medicine and Infectious Disease*, March, 101742. <https://doi.org/10.1016/j.tmaid.2020.101742>
- Molina LL, Angón E, García A, Moralejo R.H., Caballero-Villalobos J, P. J. (2018). Time series analysis of bovine venereal diseases in La Pampa, Argentina. *PLoS One* 13. <https://doi.org/doi.org/10.1371/journal.pone.0201739>
- NIH (2020). National Institute of Health of Pakistan. <https://www.nih.org.pk/novel-coronavirus-2019-ncov/>
- WHO| World Health Organization (2020). Novel coronavirus (2019-nCoV). <https://www.who.int/>
- Quick-R: (2020). *Time Series*. <https://www.statmethods.net/advstats/timeseries.html>
- Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J. M., Yan, P., & Chowell, G. (2020). Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infectious Disease Modelling*, 5, 256-263. <https://doi.org/10.1016/j.idm.2020.02.002>
- Zhang, X., Zhang, T., Young, A. A., & Li, X. (2014). Applications and comparisons of four-time series models in epidemiological surveillance data. *PLoS ONE*, 9(2), 1-16. <https://doi.org/10.1371/journal.pone.0088075>.
-



Sher Akbar got an M.Phil. degree in Statistics from Department of Statistics, Qaid-i-Azam University Islamabad, Pakistan. His areas of research are Time Series Regression Analysis, Functional Data Analysis, Data Mining.



Dr. Shahid Akbar is an Econometrician at UNDP and also visiting Assistant Professor at NUST University Islamabad. He got a PhD in Econometrics from the International Institute of Islamic Economics, International Islamic University Islamabad. His research areas are Applied and Theoretical Econometrics, Panel and Times Series Econometrics, Financial Time Series Econometrics, Applied Economics, Econometric Modeling, Bayesian Econometrics, Outlier Detection Techniques for Financial and Conventional Time Series Data, Monte Carlo Simulation Methods, Asymptotic Distributions.



Ali Raza is a Lecturer at the Department of Management Sciences, The University of Haripur, Khyber Pakhtunkhwa, Pakistan. His research interests include Econometrics, Development Economics, and International Economics.



Dr. Muhammad Rovidad is Head of the Department of Economics at Bacha Khan University Charsadda. His research interests include Economic Development, Poverty Reduction, and Microfinance.