ADVANCE SOCIAL SCIENCE ARCHIVE JOURNAL
Available Online: https://assajournal.com
Vol. 04 No. 01. July-September 2025.Page#.3332-3344
Print ISSN: 3006-2497 Online ISSN: 3006-2500
Platform & Workflow by: Open Journal Systems
https://doi.org/10.5281/zenodo.17024984

A data driven approach for forecasting the COVID-19 cases in Pakistan

Aasia Rajput
Department of Statistics, Sindh Agriculture University, Tandojam, Pakistan
Naeem Ahmed Qureshi*
Department of Statistics, University of Sindh, Jamshoro, Pakistan
Corresponding Author Email: naeem.qureshi@usindh.edu.pk
Riaz Ali Buriro
Department of Statistics, Sindh Agriculture University, Tandojam, Pakistan
Zeeshan Shabbir Soomro
Department of Statistics, Sindh Agriculture University, Tandojam, Pakistan
Kainat Rajput
Institute of Food Science and Technology, Sindh Agriculture University Tandojam, Pakistan

Abstract
Aimed at forecasting the COVID-19 cases in Pakistan using probability distribution modelling approach, the present study used daily data from 1st July to 30th September 2021 were used for estimation purpose whereas the data from 1st October to 31st December 2021 were used for forecasting. Based on the two modes and positively skewed behaviour of data, Bi-model lognormal distribution was fitted. The results showed that maximum number of daily new cases and recovered cases were recorded during the first wave while the maximum of daily deaths was observed during the third wave. A retrospective analysis for evaluation of forecasts showed that the fitted model performed well leaving out-of-sample forecast error within the limit of 5% except for total recovered cases. Forecasted values for total cases and total deaths showed the decreasing pattern while increasing trend was observed for the total recovered cases.

Keywords: *COVID-19; SARS-CoV-2; coronavirus; fifth wave; phenomenological epidemiologic model; bimodal distribution; lognormal distribution*

1. Introduction

COVID-19 is now a contagious ailment that continues to spread in densely populated regions. The World Health Organization (WHO) has called COVID-19 a global pandemic, classifying as a most severe disease impacting at least 99 percent of the major nations. It was first detected in Wuhan, Hubei Province, China (Read et al., 2020). Following December 31, 2019, when such COVID-19 epidemic hit nearly 10,710,005 people all over the world and led in 517,877 casualties, the humanitarian effects of the pandemic have been mounting. Throughout Pakistan, the first two reported cases of COVID-19 were observed on February 26th, 2020 in Islamabad as well as Karachi (Sabir et al., 2023; Haq et al., 2021; Ali et al., 2020; Iftikhar & Rind, 2020; Malik, 2020). Several people have been impacted by the country's inadequate health system, and a reckless

prevailing attitudes and massive purchasing have made the upcoming days worst, though the Pakistani government executed absolute lock-down throughout the country.

COVID-19 has affected Pakistan in four different waves so far. The first wave of the pandemic occurred at the end May 2020, surged in mid-June after those daily new recorded cases and daily new deceased cases exceeded on peaks, and eventually ceased in mid-July (Ghanchi et al., 2021). There was a decrease in cases, in early September, followed by a second wave of pandemic cases from October 2020 to January 2021.The Nation's third wave occurred in mid-March 2021. Pandemic instances began to surge again in June, and Pakistan is now well into the fourth wave, as of July 2021 (Shakeel et al., 2021). The fourth wave is coming down as far positive cases and the percentage of positivity in concerned as well as the fill-up of the health system, however, said Pakistan could experience a fifth wave of infections if the pace of vaccination was not increased (Mohakud et al., 2020). In comparison to many other countries, COVID-19-related morbidity has been rather low (Chaudhry et al., 2020).

The use of statistical models to analyse and assess the course of epidemics has been around for about than just century. Epidemiologic models of increasing complexity have been established as a result of the initial pioneering studies. Pandemic models work into two different ways to describe COVID-19; a). A transmission dynamics models and b). phenomenological models affected through the SARS-CoV-2 coronavirus (Silverman, 2023; Arora et al., 2020). The basic susceptible-infected-recovered (SIR) approach, along with versions on it, has just been usually applied among some of the earlier. Approaches rely either on the logistic or Symmetric basis function have been offered for the latter. To anticipate the ultimate extent of the COVID-19 pandemic in China, Batista even used SIR and logistic approaches (Batista, 2020). To analyze the COVID-19 outbreak in India, (Mackolil, 2020) applied the SIR and logistic approaches (Mackolil, 2020). Different studies have already been conducted in Pakistan for modelling forecasting COVID-19 cases using the ARIMA model (Ali et al., 2020; Id et al., 2021; Malik, 2020).

As mentioned above, the covid-19 modelling has been done in different ways using different statistical tools and techniques such as using probability distribution approach (Ahsan-ul-Haq et al., 2022; Dar et al., 2022) but unfortunately a very little amount of literature is available regarding the modelling of corona covid-19 using distributional approach. Being a statistician, finding the best suited probability distribution of the random variable is the challenging task. Once you are sure about the probability distribution then it will become easy to forecast the events occurring at the tails of the distribution using extreme value theory (EVT). In the present study we are going to model the covid-19 pandemic in different aspects such as number of new cases each death, number of deaths, number of recovered cases by through the probability distribution modelling approach and forecast the same. The present study is designed to review the past and current status of COVID-19 in Pakistan and to forecast the COVID-19 cases based or best- fitted distribution.

2. Methodology:

The secondary data regarding the study for Pakistan were collected from the official website of the Worldometer[1], where the COVID-19 dataset are updated on the daily bases and can be easily downloaded for different countries, continents and the overall world as well. The different daily time series regarding the COVID-19 for Pakistan were downloaded starting from 1 April 2020 to

31 December 2021. These series yielded 640 observations. Statistical modelling regarding the estimation, prediction, and forecasting of new cases and total deaths have already been done using the second wave was started in Pakistan (Shakeel et al., 2021). Pakistan has already faced the four waves of COVID-19 and fifth has already been started since December, 2021. Mostly analysis of the COVID-19 has been done during different wave period hence these analyses were reported as wave-based. Keeping in view this scenario, we have also shrunk the data set limiting up to the fourth wave which was started in July, 2021 and ended in December, 2021. The total number of observations for this period was 184. It is of worth important to note that the extrapolation was done for the next three months i.e., for 1st January, 2022 to 31st March, 2022 to forecast the COVID-19 scenario for this period. After adding this extrapolation period, the total number of observations was stretched to 274. The following variables along with their notation were used in the present study: 1). $d_1(t)$ which shows daily new cases; 2). $D_1(t)$ shows total new cases, 3).$d_2(t)$ is the daily death cases, 4). $D_2(t)$ represents total death cases, 5). $d_3(t)$ shows daily recovered cases, and 6). $D_3(t)$ showing total recovered cases. Simply put, the lower case d is reserved for the new cases of each series whereas the capital D stands for the total cases of each series. The downloaded data were analyzed using MATLAB (MATRIX LABORATORY) version 2021b.

2.1 Mathematical Model

Since every random variable follows a specific probability distribution among the set of available distributions. Each distribution is characterized by its parameter(s) which controls the shape of the distribution. Different probability distributions have already been applied in the literature to estimate and forecast the different time series related to COVID-19 (Taimoor et al., 2022). Among these the most important and valuable are Lognormal distribution, Logistic and Gaussian distribution. Since each wave of COVID-19 has two modes: the first one accounts for the maximum and minimum number of cases which occur during a particular wave. The most prominent models among the available ones for COVID-19 death prediction were proposed by (Batista, 2020; Valvo, 2020). The former has used the logistic growth model and the latter used the Bimodal Lognormal distribution. The forecasted number of total deaths in Pakistan was approximately equal to those occurred in Pakistan at that time which was a clear indication that Volvo model has captured all the data dynamics.

Keeping in view the above mentioned scenario, a bimodal lognormal distribution is exploited in the present study to model the variables used. A bimodal lognormal distribution, defined as a mixture of two lognormal distributions, is expected to distribute the daily number of cases in a country over time. A mathematical expression for this distribution is as under:

$$f_i(t) = \frac{A_i}{t}\left\{\frac{\alpha_i}{\sigma_{1i}}\exp\left[-\frac{\left(t-\mu_{1i}^2\right)}{2\sigma_{1i}^2}\right] + \frac{1-\sigma_i}{\sigma_{2i}}\exp\left[-\frac{\left(t-\mu_{2i}^2\right)}{2\sigma_{2i}^2}\right]\right\} \qquad (1)$$

where *f(t)* denotes the probability density function at time *t*, *A* denotes amplitude (the maximum of the wave), $\alpha \in [0, 1]$ denotes the mixing parameter, μ_1 and σ_1 denote the mean and standard deviation of the natural logarithm of *t* for the first lognormal distribution, and μ_2 and σ_2 denote the same variables for the second lognormal distribution. The integral in time *f(t)* is the total number of daily cases:

$$F_i(t) = \frac{A_i}{2}\left\{\alpha\left[1 - erf\,\frac{(\log t - \mu_{1i})}{\sqrt{2}\sigma_{1i}}\right] + (1-\alpha_i)_1\left[1 + erf\left(\frac{\log t - \mu_{2i}}{\sqrt{2}\sigma_{2i}}\right)\right]\right\} \qquad (2)$$

The error function is denoted by *erf*(). The function *F(t)* tends asymptotically to the amplitude, *A*, which can be understood as the total number of fatalities (deaths) expected over an infinite time as *t +* increases.

2.2 Numerical Analysis

A nonlinear least-squares procedure is used to fit the available data and determine the model parameters. In all the time series used, the time variables, t, are stated in days in this example. It starts at 1 (the first day of the dataset, January 1, 2021) and ends at t_{max} = 274 (last day of simulations: 31 March 2022). The days corresponding to the end of the fitting range, t_{fit}, and the end of the extrapolation period, $t_{forecast}$, can be chosen by the user. The process calculates the days t_1 and t_2 on which d(t) reaches its first and second maximum for each location. Then, in the range between t = 1 and t = t_{fit}, the function representing the theoretical total number of deaths, F(t), is fitted against the reported total number of deaths, D(t). The process uses the MATLAB lsqcurvefit function, which requires each parameter to have a minimum, a guess, and a maximum value. These parametric values are specified in the following table.

Table 1. lsqcurvefit range of fitting parameters

Parameter	Minimum	Guess	Maximum
A (No.)	*1.10 D(t_{fit})*	*max {6 D(t_1), 2 D(t_{fit})}*	*max {12 D(t_1), 4 D(t_{fit})}*
M_1 (days)	*1*	*t_1*	*t_{max}*
σ_{t1} (days)	*0*	*1*	*t_{max}*
M_2 (days)	*1 α*	*t_2*	*t_{max}*
σ_{t2} (days)	*0*	*1*	*t_{max}*
(–)	*0*	*1 1+d(t_2)/d(t_1)*	*1*

As a result, the following six parameters are determined for each country:

A, $M_1 = \exp\left(\mu_1 - \sigma_1^2\right)$ $\sigma_{t1} = \sqrt{(\exp\sigma_1^2 - 1)\exp(2\mu_1 + \sigma_1^2)}$, $M_2 = \exp\left(\mu_2 - \sigma_2^2\right)$, $\sigma_{t2} = \sqrt{(\exp\sigma_2^2 - 1)\exp(2\mu_2 + \sigma_2^2),}$ and α .Here, M_1 and σ_{t1} respectively denote the mode and standard deviation of the first lognormal distribution. Likewise, M_2 and σ_{t2} are the same quantities for the second distribution. The obtained parameters are substituted into Equations (1) and (2). Thus, the theoretical estimates of the daily and total number of daily cases, deaths and recovered are computed between t = 1 and t = t_{max}. The procedure is repeated for each of the Pakistan. The MATLAB script saves the results obtained for each location in terms of fitted parameters, residual norm and exit flag returned by the lsqcurvefit function, actual number of deaths (at t = t_{fit}), and forecast number of deaths (at t = $t_{forecast}$). Besides, for each (selected) series, six plots are created and saved in PNG format: (i) the daily new cases and the total new cases, (ii) the daily deaths cases and total deaths cases, (iii) daily recovered cases, total recovered cases, as functions of time.

2.3 Diagnostic Checking

A first diagnostic investigation was conducted utilizing data starting from July 1, 2021 to September 30, 2021 (in-sample fitting data), to analyse the connection between model prediction and actual data. Extrapolations (out-of-sample forecasts) have been made from October 1, 2021 to December 31, 2021 every fortnightly.

2.4 Retrospective Predictions

To use a larger dataset to assess the accuracy of the model predictions, a second diagnostic checking study has been carried out. In this case, the model has been used to predict retrospectively the (known) number of total deaths, total new cases, and total new recovered as of December 2021. Different dates were chosen for the end of the fitting range: from 15 October to 31 December 2021. Table 2 shows the actual and forecast number of total deaths, total new cases, and total new recovered cases in the Country. Percentages in parentheses represent the errors of the theoretical predictions with respect to the actual data, calculated as $[D(t_{forecast}) - F(t_{forecast})] / F(t_{forecast})$.

3. Results

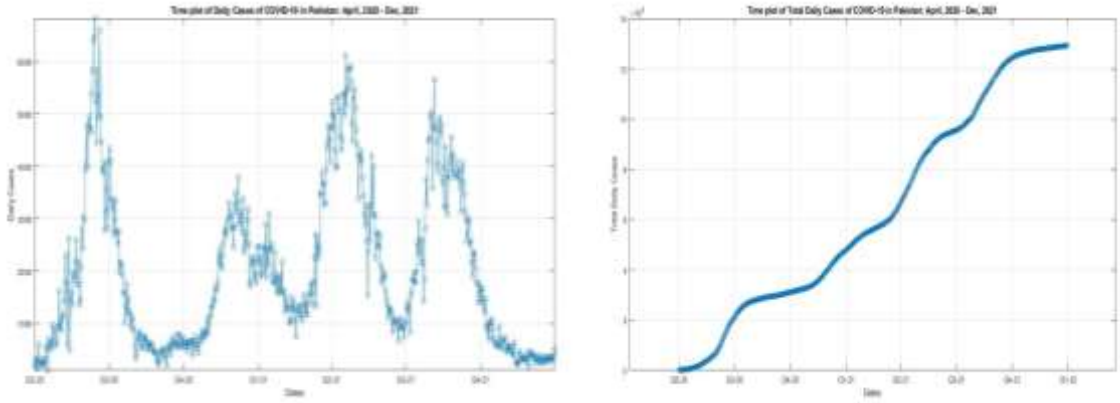
3.1 Pakistan's current status of COVID-19

In this section, the current status of COVID-19 in Pakistan was described using the most widely used measures of central tendency and dispersion as well. These measures include minimum, maximum, mean, variance, skewness and kurtosis. The necessary summary statistics for the variables under study are computed and presented.

Table 2. Summary statistics of COVID-19 in Pakistan: 01-04-2020 to 31-12-2021.

Variables	Min.	Max.	Mean	Variance	Skewness	Kurtosis
Daily Cases	106	6825	2019.889	2348437	0.812682	2.712034
Daily Deaths	1	201	45.25938	1287.372	0.974529	3.651767
Recovered Cases	1	16813	1967.959	4357190	2.871945	15.45198

The minimum values of daily new cases, death cases and recovered cases of COVID-19 in Pakistan are found 106, 1, 1 while as the maximums are 6825, 201, 16813, respectively. The mean of the COVID-19 for these variables are 2019.889, 45.25938, 1967.959, while the variances are 2348437, 1287.372, 4357190, respectively, the large value of variance was found for recovered cases which shows the more variability is present in this variable. Furthermore, all the variables are positively skewed which is an indication that any positively skewed distribution is needed to model these variables. Likewise, the values of kurtosis new cases, deaths and recovered cases are 2.712034, 3.651767, and 15.45198, respectively which further supports the non-normal behaviour of the data. Keeping in view these values, a lognormal distribution (which is positively-skewed distribution) was exploited to model the distribution of these variables.



(a) Daily new cases (b) Total new cases

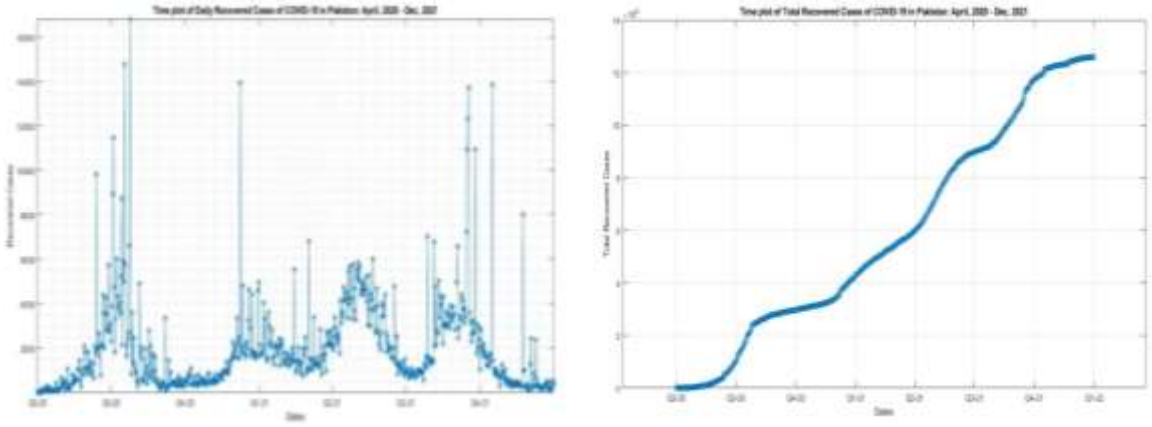
Figure 1. Time Plot of new cases of COVID-19: April, 2020 to Dec, 2021

It can be easily seen from the first panel of the Figure 1 that there has been observed four waves of COVID-19 since its first case of emergence in Pakistan. The first wave was occurred during Q2-20 to Q3-20, in which maximum number new cases was reported to be 6823. Among the observed four waves, the minimum of the maximum cases was observed during the second wave which occurred during Q4-20 to Q1-21. However, it can be clearly seen from the Figure 1a that the reported number of new cases was much higher for the third (6127) and fourth wave (5611) as compared to the second wave. Furthermore, this figure shows that the first wave was the most intensive wave when compared with the remaining waves. So far as Figure 1b is concerned, which shows the cumulative total cases, is can be seen that total daily cases has been increasing exponentially which has already been reported in various literature as a fact that COVID-19 is spreading exponentially (Malik, 2020). Total number of new cases reported was 7048 on 31st December, 2022.

Like the previous figure, Figure 2 is also divided into two subfigures: Figure 4.2a shows the daily deaths cases of COVID-19 in Pakistan while Figure 4.2b shows its cumulative number. Figure 2a clearly reveals that among the four waves of COVID-19, the peak number (201) of death per day was observed during the third wave which was occurred during Q1-21 to Q2-21. Conversely, among the four waves, the minimum of peak deaths (111) was observed during the second wave cases which occurred during this Q3-20 and Q4-20. So far as Figure 2b is concerned, showing that the cumulative of daily deaths has also been increasing exponentially.

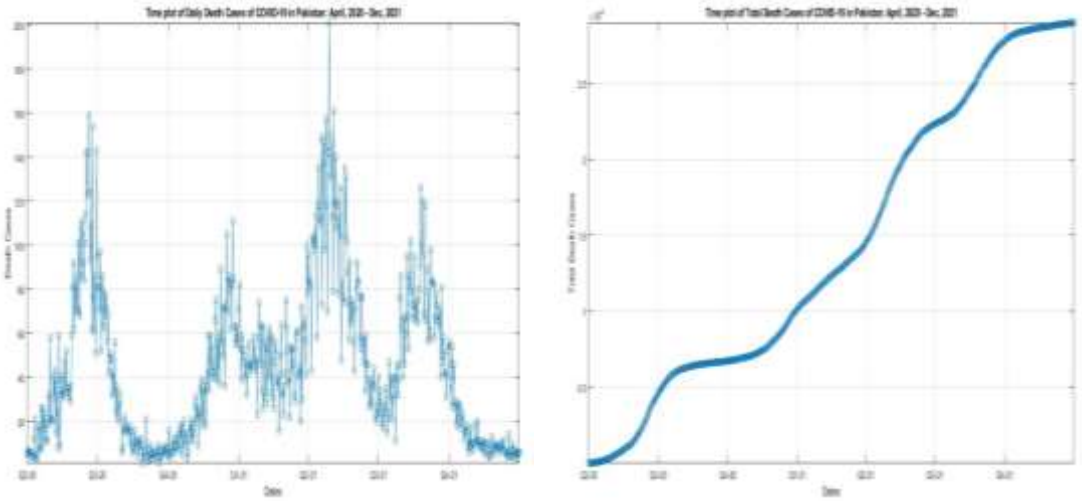
The two panels of the third figure, under the shadow of current status of COVID-19 in Pakistan, show the daily recovered cases and total number of recovered cases, respectively. Among the four waves, the maximum number of daily recovered cases was observed during first wave followed by second and fourth waves, respectively. Likewise, the minimum of this variable was observed during the third wave cases which occurred during this Q3-21 and Q4-21, in which the minimum number new recovered cases were 6021. Just like the second panels of the last two figures, the second panel of this figure depicts the cumulative number of recovered cases and

reveals that this variable is also increasing geometrically (geometric decay is sharp decay as compared to exponential one).



(a) Daily Death Cases (b) Total Death Cases

Figure 2. Time Plot of deaths cases from COVID-19: April, 2020-Dec, 2021



(a) Daily Recovered Cases (b) Total Recovered Cases

Figure 3. Time plot of recovered cases of COVID-19: April, 2020-Dec, 2021

3.2 Estimation and forecasting results

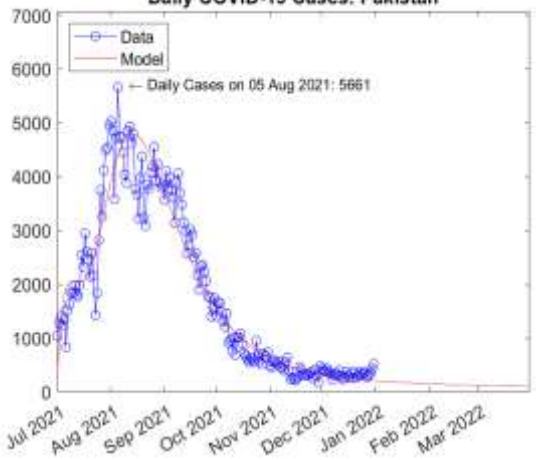
This section describes the results of the estimation and forecasting. The estimation is done using the data starting from 1st July, 2021 to 30th September, 2021 while the forecasts for all the time series were generated every fortnightly starting from 15th October to 31st March, 2022. In this regard, the first figure of this section labelled as Figure 4a is showing the original data (blue points) and the fitted probability density function curve (the red line which is superimposed) using

Bimodal Lognormal distribution. It can be seen that this distribution fits the data well as red curve is passing almost all the data points. According to the fitted model, for the fourth wave, the highest number of daily cases (5661) was estimated for 5th August, 2021. After that, the model shows decreasing trend for this time series.

Figure 4b reports the results of the model prediction for the total cases of COVID-19. It can be seen that the model predicts increasing trend of total number of cases. The highest forecasted value for this time series was found for 31st December, 2021, which was 338267 whereas the actual reported cases at that day were 33.This is a only total number of daily cases the predict current wave of COVID-19 in Pakistan , which is started in 1 July 2021 to 31 December 2021. So far as the results of the forecasted values are concerned for the total daily cases, the model forecasted 337291 for 31st December, 2022 whereas the actual total number of cases reported on this date was 338267. Likewise in terms of out-of-sample forecast, the model forecasted the total COVID-19 cases in Pakistan on 31st March, 2022 as 350115. Precisely, the new cases are going to decrease so as the total number of cases also.

So far the daily recovered cases are concerned, Figure 6a depicts the new recovered cases whereas Figure 4.6b prevailing the total recovered cases of COVID-19 in Pakistan. The peak of daily recovered cases was found on 16th October, 2021 with the number 13848. Once again after that peak, the data showing the decreasing trend for the fourth wave of corona virus in the country. Figure 6b shows the actual and forecast number of total recovered cases in Pakistan. The predicted future trend is strongly increasing which is a very positive sign. The estimate total numbers of COVID-19 recovered cases were reported in 31 December 2021: 362077. The forecasted value by the end of forecast period (31st March, 2022) was 373552.

(a)



(b)

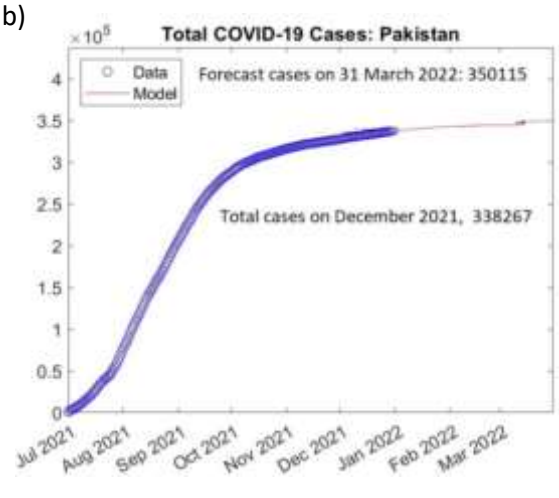


Figure 4. Actual and predicted trends of new cases of COVID-19:July 2021- March 2022

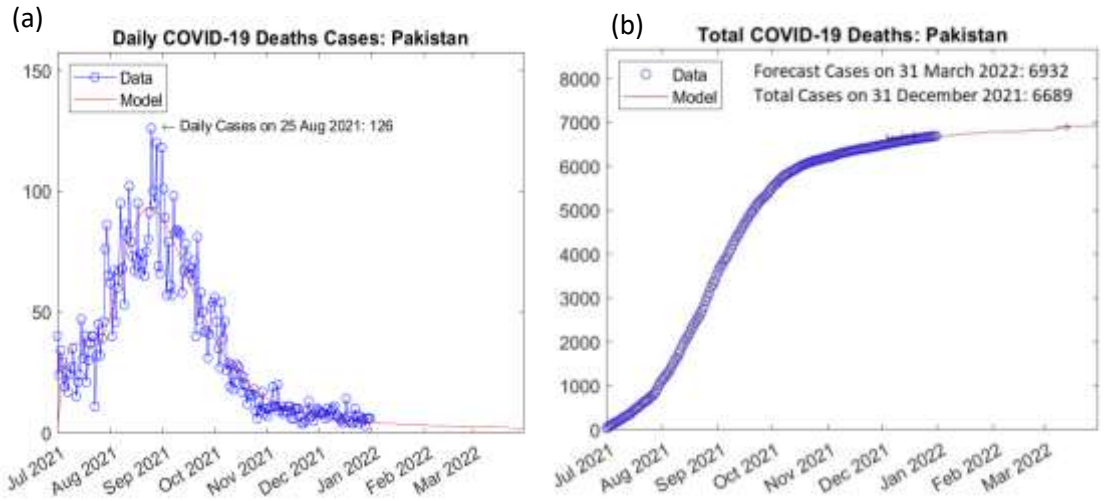


Figure 5. Actual and predicted trends of deaths in Pakistan:July 2021- March 2022

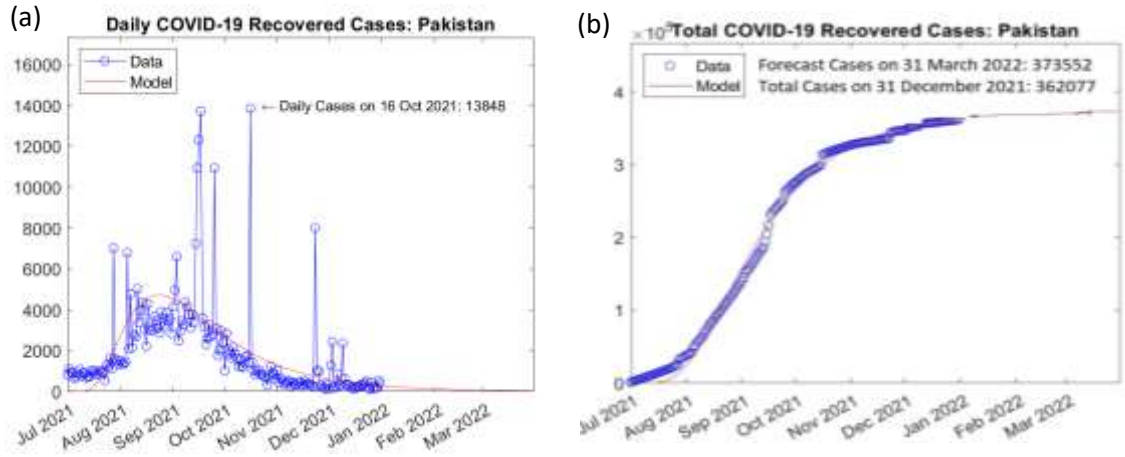


Figure 6. Actual and predicted trends of recovered cases in Pakistan: July, 2021 to March, 2022

3.3 Forecast Accuracy

The forecasting performance of the fitted model was assessed by using out-of-sample forecasting error. Table 2 shows the one-step ahead forecasting errors of the model. It can be seen that the largest percent of error was observed for total recovered cases. Although the values are going to decrease as the time goes far but all these forecast errors are in negative number showing that the fitted model underestimates the scenario. In particular, the three-month forecasts (as of 15 October and 31 October 2021) of total recovered cases, however, carries a marginal weight in the country balance. Relatively large errors can be noted also in the three-month forecasts for total deaths and total cases. Actually, such predictions are computed on a relatively little amount of

data. Three-month forecasts for total deaths and total new cases are the most accurate, thanks to the large amount of data fitted (as of October 2021 the fourth peaks of the epidemic were already overcome). Two-month predictions (as of 15 November 2021, 30 November 2021) and one-month predictions (as of 15 December 2021 and 31december 2021) show absolute errors less than about 10% respectively for all the COVID-19 cases. Based on this, it can be concluded that the model is capable of capturing future trends with reasonable accuracy. Larger errors are expected for Pakistan and COVID-19 cases with little amount of fitted data and where the epidemic is still at early stage.

Table 2. Actual and Forecasted Cases of COVID-19 in Pakistan: 15-10-2021 to 31-12-2022

	15 October		**31 October**		**15 November**		**30 November**		**15 December**		**31 December**	
Variables	A	F	A	F	A	F	A	F	A	F	A	F
Total Deaths	5962 (-2.80%)	6133	6195 (-3.00%)	6393	6358 (-1.10%)	6433	6474 (-0.40%)	6501	6589 (0.20%)	6573	6689 (0.50%)	6654
Total Cases	305291 (-3.90%)	317749	316008 (-2.60%)	324523	322806 (-1.10%)	326600	327770 (-0.80%)	330539	332855 (-0.30%)	333902	338267 (0.20%)	337291
Total Recovered	299611 (-16.60%)	359439	326512 (-12.10%)	371601	333546 (-9.80%)	370141	347297 (-4.20%)	362681	357411 (-1.00%)	361216	362077 (-0.70%)	364765

Where “A” stands for actual values and “F” stands for forecasted values, ( ) shows the percent forecast error

3.4 Model Validation

Figure 7 shows scatter plot of the actual vs. predicted number of total cases, total deaths and total recovered cases of COVID-19. Regression lines with values of intercept, slope, and a coefficient of determination are also shown. The slope values (β’s) for total new cases and total deaths are greater than 1, thus highlighting a tendency of the model to overestimate the actual data; recovered cases has a slope less than 1, corresponding to a tendency to underestimate. So far as total cases and total deaths as concerned, there exists a linear relationship between the actual and forecasted values. The same is not true for total recovered cases, hence the relationship between actual and forecasted total recovered cases follow non-linear relation.

The values of coefficient of determination for total cases and total deaths were 99.09% and 96.19%, respectively. The third panel i.e., panel (c) shows the plot of the actual and predicted values for total recovered cases of COVID-19. This panel clearly shows that linear fit is not a good fit as most of these values are away from linear relation line (orange line). Among the four different fitted models (linear, quadratic, cubic and fourth degree polynomial), the best fitted was found to be the 4th degree polynomial. The same can be confirmed from the models R-square values i.e., the highest value (84.96%) was reported for the 4th degree hence, best fitted the data points.

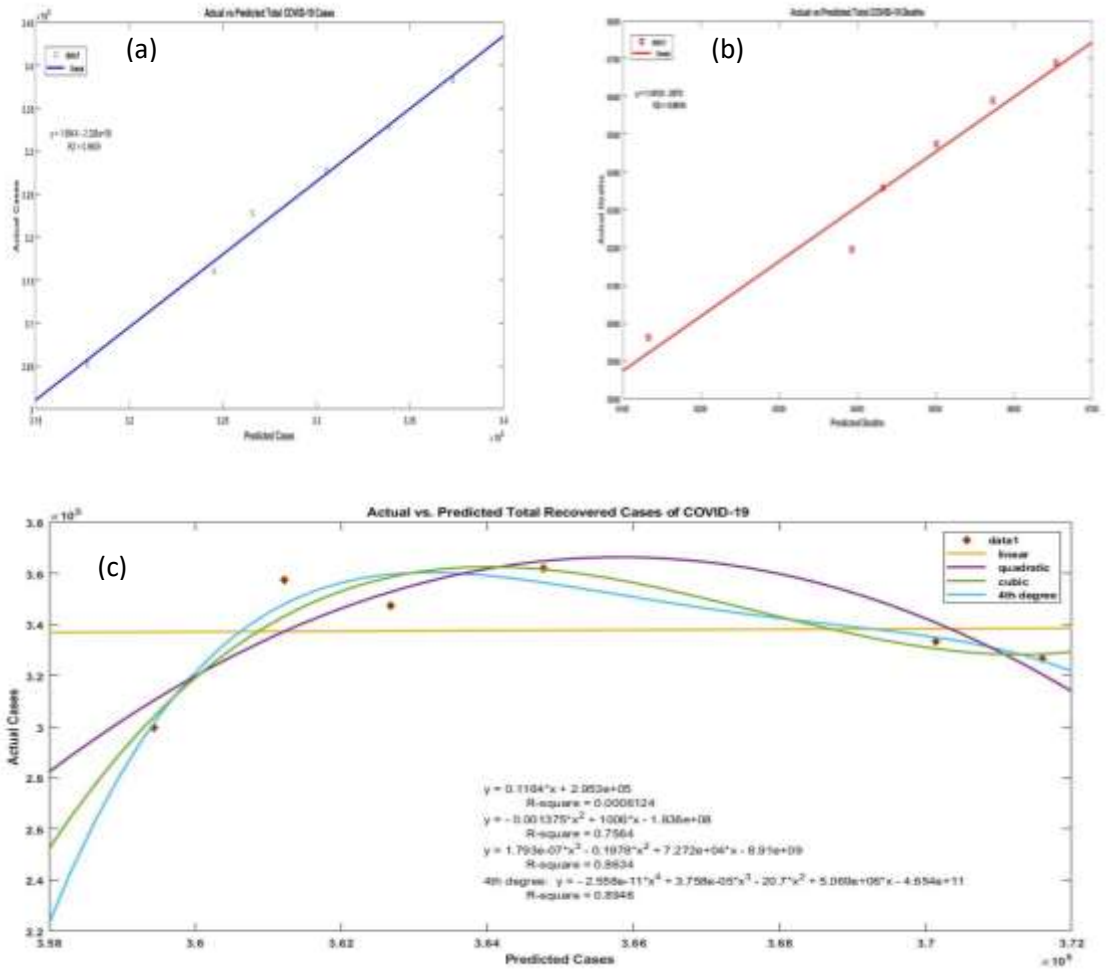


Figure 7. Actual vs. predicted cases: (a) total new cases, (b) total death cases, and (c) total recovered cases

4. Conclusion

Based on the positively skewed behaviour, and minimum and maximum of the variables under study by particularly focusing on the fourth wave, bimodal lognormal distribution was fitted and the forecasts were generated using extrapolation method for the current (fifth) wave in Pakistan. The fitted distribution was performed well as the fitted curve mostly covered all the data points including the peak (maximum) and trough (minimum) of each of the variables under observation. Based on the forecasted values, it is concluded that in case of daily new cases and daily deaths, the negative changed was observed whereas in case of total recovered cases the trend becomes negative showing good results. The plots of actual and predicted values for total new cases and total deaths revealed that forecasting is a linear function whereas in case of total recovered cases, it follows a non-linear (4^{th} degree polynomial) regression model.

4.1 Recommendations

The following recommendations are made base on the results and conclusions.

(1) The present work may be extended by using any multi-model distribution based on the data of the current (fifth) wave or by including the more variables such as number of hospitalizations, population of a country or region, number of test cases etc.
(2) Bases on the non-linear behaviour of the total recovered cases, it is recommended that ARCH/ GARCH models can be exploited to improve the estimation and forecasting for this variable.

References

Ahmad, I., & Muhammad Asad, S. (2020). Predictions of coronavirus COVID-19 distinct cases in Pakistan through an artificial neural network. *Epidemiology and Infection*, *148*, e222–e222. https://doi.org/10.1017/S0950268820002174

Ahsan-ul-Haq, M., Ahmed, M., Zafar, J., & Ramos, P. L. (2022). Modelling of COVID-19 Cases in Pakistan Using Lifetime Probability Distributions. *Annals of Data Science*, *9*(1), 141–152. https://doi.org/10.1007/s40745-021-00338-9

Ali, M., Khan, D. M., Aamir, M., Khalil, U., & Khan, Z. (2020). Forecasting COVID-19 in Pakistan. *PLoS ONE*, *15*(11 November), 1–13. https://doi.org/10.1371/journal.pone.0242762

Arora, S., Jain, R., Harendra, , & Singh, P. (2020). *Epidemiological models of SARS-CoV-2 (COVID-19) to control the transmission based on current evidence: A systematic review*. *2*(July), 1–41. https://doi.org/10.20944/preprints202007.0262

Batista, M. (2020). Estimation of the final size of the COVID-19 epidemic. *MedRxiv*, *Nesteruk*, 1–11.

Chaudhry, R. M., Hanif, A., Chaudhary, M., Minhas, S., Mirza, K., Asif, T., Gilani, S. A., & Kashif, M. (2020). Coronavirus Disease 2019 (COVID-19): Forecast of an Emerging Urgency in Pakistan. *Cureus*, *2019*(5). https://doi.org/10.7759/cureus.8346

Dar, J. G., Ijaz, M., Almanjahie, I. M., Farooq, M., & El-Morshedy, M. (2022). Statistical Analysis of the COVID-19 Mortality Rates with Probability Distributions: The Case of Pakistan and Afghanistan. Computational and mathematical methods in medicine, 2022, 4148801. https://doi.org/10.1155/2022/4148801

Ghanchi, N. K., Nasir, A., Masood, K. I., Abidi, S. H., Mahmood, S. F., Kanji, A., Razzak, S., Khan, W., Shahid, S., Yameen, M., Raza, A., Ashraf, J., Ansar, Z., Dharejo, M. B., Islam, N., Hasan, Z., & Hasan, R. (2021). Higher entropy observed in SARS-CoV-2 genomes from the first COVID-19 wave in Pakistan. *PLoS ONE*, *16*(8 August), 1–17. https://doi.org/10.1371/journal.pone.0256451

Id, F. K., Ali, S., Saeed, A., Kumar, R., & Khan, A. W. (2021). *Forecasting daily new infections , deaths and recovery cases due to COVID-19 in Pakistan by using Bayesian Dynamic Linear Models*. 1–14. https://doi.org/10.1371/journal.pone.0253367

Iftikhar, H., & Rind, M. (2020). *Forecasting daily COVID-19 confirmed, deaths and recovered cases using univariate time series models: A case of Pakistan study*. Cold Spring Harbor Laboratory. https://doi.org/10.1101/2020.09.20.20198150

Mackolil, J. (2020). *Mathematical Modelling of Coronavirus disease (COVID-19) Outbreak in India using Logistic Growth and SIR Models*.

Malik, A. B. (2020). *Analysis and Forecast of COVID-19 Pandemic in Pakistan*. 1–11.

Mohakud, N. K., Yerru, H., Rajguru, M., & Naik, S. S. (2020). An Assumed Vertical Transmission of SARS-CoV-2 During Pregnancy: A Case Report and Review of Literature. *Cureus*, *12*(9), 1–5. https://doi.org/10.7759/cureus.10659

Read, J. M., Bridgen, J. R. E., Cummings, D. A. T., Ho, A., & Jewell, C. P. (2020). *Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions*. Cold Spring Harbor Laboratory. https://doi.org/10.1101/2020.01.23.20018549

Sabir, Q. U. A., Shafqat, A., & Aslam, M. (2023). Prediction of the COVID-19 transmission: a case study of Pakistan. Epidemiology and infection, 151, e89. https://doi.org/10.1017/S0950268823000730

Shakeel, M., Irfan, M., Nisa, Z., Rashid, M., Ansari, S. K., Khan, I. A., Cassmann, E., Rollins, A., Zylich, N. C., Renshaw, R. W., Guarino, C., Wagner, B., Lager, K., Diel, D. G., Bhattacharyya, S., Siddiqui, S., Tyagi, A., Jha, S., Pandey, R., … Veesler, D. (2021). Surveillance of genetic diversity and evolution in locally transmitted SARS-CoV-2 in Pakistan during the first wave of the COVID-19 pandemic. *BioRxiv*.

Silverman, M. (2023). Probability Distribution of SARS-Cov-2 (COVID) Infectivity Following Onset of Symptoms: Analysis from First Principles. Open Journal of Statistics, 13, 233-263. doi: 10.4236/ojs.2023.132013.

Taimoor, M., Ali, S., Shah, I., & Muwanika, F. R. (2022). COVID-19 Pandemic Data Modeling in Pakistan Using Time-Series SIR. Computational and mathematical methods in medicine, 2022, 6001876. https://doi.org/10.1155/2022/6001876

Valvo, P. S. (2020). A Bimodal Lognormal Distribution Model for the Prediction of COVID-19 Deaths. *Applied Sciences*, *10*(23), 8500. https://doi.org/10.3390/app10238500