International Journal on Infectious Disease and Epidemiology

Prediction of Daily COVID-19 Active Cases in Bangladesh

Review Article

Volume 2 Issue 1- 2021

Author Details

*Shipra Banik[1], Mohammad Answer[2] and B M Golam Kibria[3]**

[1]Department of Physical Sciences, Independent University, Bangladesh

[2]Department of Physical Sciences, Independent University, Bangladesh

[3]Department of Mathematics and Statistics, Florida International University, USA

*Corresponding author

BM Golam Kibria, Department of Mathematics and Statistics, Florida International University, USA

Article History

Received: July 09, 2021 Accepted: July 16, 2021 Published: August 03, 2021

Abstract

Recording daily active cases statistics benefits instantly when studying the ongoing COVID-19 dashboard for a country. It would help to illuminate the probable strategy of action for policy makers, as well as planners of a country. The goal of the paper is to forecast daily active COVID-19 cases for Bangladesh using various forecasting models, namely the regression model, the non-linear exponential regression model and some time series models based on the daily positive cases, daily deceased cases, daily recovered cases and some environmentally important variables. Our findings suggest that time series forecasting models have better ability to predict daily active cases as compared to other selected forecasting models.

Keywords: Active cases; COVID-19; prediction; time series modelling; common error measures

Introduction

The worldwide ongoing pandemic of the novel corona virus disease 2019 (COVID-19) is a contagious disease triggered by the severe acute respiratory syndrome coronavirus 2, namely SARS-CoV-2. The first recognized case was from the wet market of Wuhan (a city of China) in December 2019. Since then, this disease has spread, prominently to a continuing pandemic. In the nearby interaction, the germ blowout among people over a small drop out by sneezing, coughing, talking and not moving long distance through air, the dewdrop drops on the surface. The signs of COVID-19 disease are fever, coughing, trouble breathing, tiredness, shortness of breath, chest pain, chills, fatigue, shaking, body aches, sore throat, headache, congestion of runny nose, loss of smell and/or taste, nausea, diarrhea and others. To stop spreading of this deadly virus, the following number of actions carried out, like wearing face mask, personal hygiene, washing hands often with soap, creating social distancing and others.

In order to prevent the spread of the virus, many countries impose shutdown and lockdown. As of June 2021, 181,680,392 cases have been identified in 188 countries and territories, resulting in 3,935,392 deaths (Source: Wikipedia). In March 2020, this deadly disease was first observed to have a blowout in Bangladesh. The first 3 identified instances were stated on March 8 2020, by the Institute of Epidemiology, Disease, Control and Research (IEDCR)[1]. Subsequently, the pandemic has been blowing out gradually over the country. Still, the number of positive cases is growing day by day. From March 23 to May 30 to shelter nationals, this country's republic government affirmed lockdown all over the country. This time, the government of Bangladesh arranged a few essential footsteps, such as reporting daily newspapers, social media and isolation centers at different hospitals in all districts. To control COVID-19 pandemic in Bangladesh, a series of hotline numbers, email address and Facebook page of IEDCR are also given for nationals to interact if one is suspicious about this virus contagion. With this, it is observed that up to the end of March 2020, contaminations persisted low. However, there was a sharp surge that was noticed in the first week of April 2020. In the mid-month of April, positive cases percentage significantly raised, the highest in Asia after Indonesia. On May 6, 2o2o, the positive cases were confirmed in all districts. Just a note, on June 13, 2020 the number of positive cases in Bangladesh surpassed the number of positive cases in China. Bangladesh touched two edge marks of 160000 active cases and 2000 deaths on 5 July, 2020 and 2 days later passed France in terms of the number of new cases. It is observed that on 12 July, 2020, the number of recoveries exceeded the number of active cases, a positive sign

[1]It is a government research institution under the Ministry of Health, responsible for researching epidemiological and communicable diseases in Bangladesh as well as disease control.

for this country. It is notice that Bangladesh was the second most-infected country in South Asia after the first infected country, India. According to the world health organization report, as of June 27, 2021, in Bangladesh, 888,406 cases have been identified. The number of fatalities due to COVID-19 pandemic is 14,172 till date (Source: Wikipedia).

This paper proposes projecting a model for fitting COVID-19 data in Bangladesh to evaluate, prediction and abstract significant evidence that advantages policy makers, researchers, health workers and others. It is believable that daily active cases play a very important role to understand the current scenario of this virus existing condition of the country. Thus, our goal is to predict daily active cases (people affected by this disease are taken into account) based on daily positive cases, daily decreased cases, daily recovered cases and also on some environmental factors such as daily temperature and daily humidity to find a trend the next day's trend in the active cases in Bangladesh. A number of works have been done in literature to study daily active cases. To mention a few: Barria-Sandoval [1], Mahanty et al. [2], Sarkar [3], Puttahonnappa et al. [4], Viswanath [5], Rath et al. [6], Hassan et al. [7] among others. To the best of our knowledge in respect of Bangladesh, no work on prediction of active cases is available in the literature. We thought this was of significance because it would advantage us to explain the possible strategies of exploiting it as well as formulate it. To project the trend of daily active cases, we have used various statistical prediction models, including time series of models since our all considered data sets are time series data. The rest of the paper is organized as follows: A brief description of the considered data sets and descriptive measures are given in section 2. Various prediction models are described in the section 3. Results with discussion are presented in section 4. Finally, some concluding remarks are provided in the final section.

Data Descriptions

The COVID-19 data for Bangladesh (from March 8, 2020 to March 29, 2021) were collected from the IEDCR for the daily positive cases, daily deaths, daily recovered cases and daily active cases. To see the virus effects on the Bangladesh environment, daily temperature and daily humidity are also considered (Source: www.noaa.gov). We have used the following formula to calculate daily humidity,

Relative humidity = $\frac{e_w - N(1+0.00115T_w)(T_d - T_w)}{e_d}$

where

$e_d = 6.112e^{\frac{17.502*T_d}{240.97+T_d}}$, $e_w = 6.112e^{\frac{17.502*T_w}{240.97+T_w}}$, e = 2.718, T_d= Dry bulb temperature (^{0}C), T_w= Wet bulb temperature (^{0}C) and N = 0.6687451584.

To see the impact of daily cases in Bangladesh, daily positive cases, daily deaths, daily recovered cases and daily active cases are plotted against time (day) in the Figure 1 How Bangladeshi people are affecting and recovering w.r.t day, cumulative frequency for each of the considered COVID-19 data sets are plotted in the Figure 2 The summary statistics of the data are given in Table 1.

Table 1: Numerical summary of the considered daily COVID-19 data sets.

Data Sets	Mean	SD	Skewness	Kurtosis
Positive cases	1.55E+03	1.04E+03	0.4973	2.5104
Active cases	7.52E+04	4.33E+04	-0.2401	1.8222
Deaths	23.4599	13.6901	0.1312	2.2001
Recovered cases	1.38E+03	1.26E+03	3.8854	40.5381

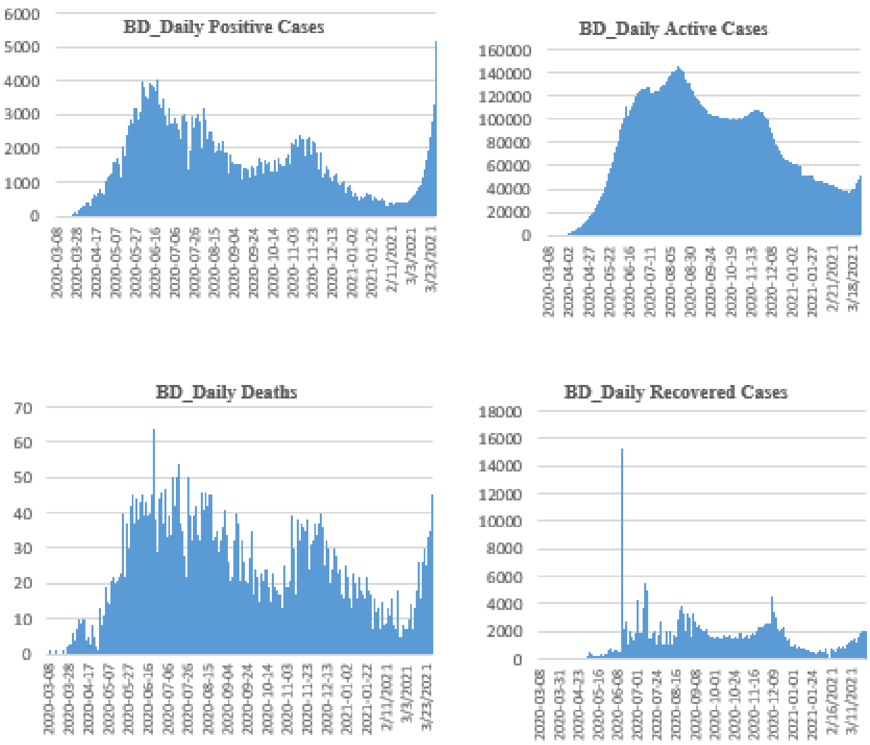

Figure 1: Time plots of the considered COVID-19 data sets.

It is observed from Figure 1 that the BD positive cases sharply increased up to June 19 2020, then there was a declining pattern up to September 2020. The daily decline was also observed for new cases up to January 2021. This is due to the strict lockdown enforced by the Bangladesh government in almost every sector. Again, a daily rise in new cases was observed at the end of March 2021. It happened again after the lockdown due to the lack of consensus of people, the COVID-19 virus has started to increase up. It is noticeable from this Figure 1 that, like daily infected cases, there were almost similar rising and declining patterns observed in daily active cases. Overall plot of daily active cases under our study periods obviously shows a positive scratch for our country because of a declining pattern w.r.t day by day. It gives us an indication the Bangladeshi Government was working hard to recover COVID-19 infected patients. That reflects in the time plot of the daily

Citation: Banik S, Answer M, Kibria BMG. Prediction of Daily COVID-19 Active Cases in Bangladesh. *Int. Jr. Infect Dis & Epidemlgy*. 2021;2(1):25–33. DOI: 10.51626/ijide.2021.02.00006

recovered cases (See the Figure 1), which is an optimistic bulletin for the Bangladeshi people according to the records of daily recovered cases. The daily number of deaths also plotted in the Figure 1 to understand this variable movement w.r.t each day. If we closely notice this plot, the daily rise observed up to the mid-June, 2020. The cause behind this may be that at the beginning of March 2020, the laboratory test was done in a small number of population. It is ensuing that small number of contagion, which increases gradually the number of deaths. After this period, the lab test numbers were increased and also owing to lockdown, we note that the actual number of death statistics came down slowly up to the mid-November. Unexpectedly, then we observe an increasing trend up to the mid-December, which may indicate to us that the population who was identified with COVID positive earlier was likely to die. After this period, the figure of deaths was slowly down up to the mid-March, 2021 because of conciseness of the population. It is noted that at the end of March 2021, daily deaths have risen so far in our considered end of data periods. Hence, increasing daily deaths through this period could mean COVID contamination was deadly. Overall it may be concluded from these plots that there might be a relation between daily active cases with other considered important COVID-19 daily cases. To get an idea about the cumulative number of positive cases, cumulative number of deaths and cumulative number of recovered cases, see the Figure 2.

Since the focus of this paper is to predict daily active cases, it is important to see whether active cases are related to our considered independent variables or not. To observe this, coefficient of correlation values are reported in Table 2. It appears from Table 2 is that our dependent variable, daily active cases are dependent on other considered independent variables. So it is reasonable to apply some modeling tools to assess the comparative influence of active cases due to our daily considered cases. Thus, our next target is to apply some statistical prediction models to forecast daily active cases.

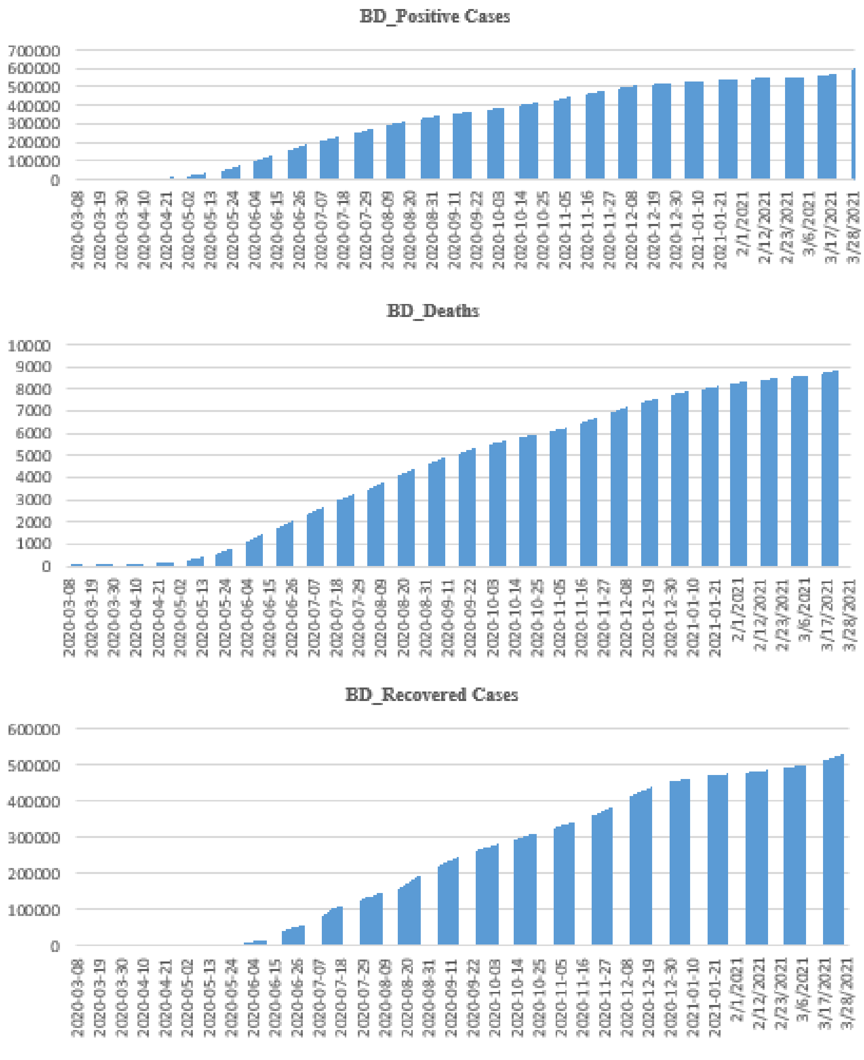

Figure 2: Cumulative frequency of the considered COVID-19 data sets.

Table 2: Correlation matrix of BD COVID-19 selected cases.

	Active	Recovered	Deaths	Positive	Temperature	Humidity
Active	1.0	0.6338	0.7965	0.6518	0.3829	0.6481
Recovered		1.0	0.5524	0.3917	0.5478	0.6256
Deaths			1.0	0.7819	0.3666	0.5943
Positive				1.0	0.1166	0.3904
Temperature					1.0	0.6399
Humidity					0.6399	1.0

Statistical methodology

In this section, we will discuss about some statistical prediction models used to predict daily active cases.

Linear regression (LR) model

First, we will consider a linear two-variable regression model, which is defined as follows:

$Y_t = a + ct + bX_t + e_t$, t =1, 2, … ,n (1)

where Y_t is the dependent variable (daily active cases), X_t is the independent variable (daily positive cases), *a* is the intercept, *c* is the coefficient of time trend to capture deterministic trend, *b* is the regression coefficient Y_t on X_t and e_t is the error term normally distributed with a mean of 0 and variance is 1.

Multiple linear regression (MLR) model

This model is an extension of the linear regression model that uses several independent variables to predict the outcome of a dependent variable. A MLR model can be written as such

$Y_t = \beta_0 + ct + \beta_1 X_{t1} + \beta_2 X_{t2} + \ldots + \beta_p X_{tp} + e_t$, t =1,2, … ,n (2)

where Y_t is the dependent variable (daily active cases), X_t are the independent variables, β_0 is the intercept, c is the coefficient of time trend, β_p are the regression coefficients for each explanatory variable and e_t is the error term. The MLR model is based on the following assumptions: There is a linear relationship between a dependent variable and independent variables, the independent variables are not too highly correlated with each other and residuals should be normally distributed with a mean of 0 and variance 1.

Non-linear regression model

Non-linear regression model is a system of regression analysis in which observational data are modeled by a function, which is a nonlinear combination of the model parameters and depends on one or more independent variables. We have applied the following non-linear models to predict daily active cases.

Exponential regression (ER) model

This model is perhaps one of the simplest nonlinear regression models, generally used to model conditions in which evolution starts gradually. This model accelerates where decline starts and then slows down to grow quicker and quicker to 0. On occasion, the LR model can be used with relationships that are naturally non-linear, but can be made linear after a conversion. In specific, we consider the following exponential model:

$Y_t = ae^{ct+bX_t} + e_t$, t =1, 2, …,n (3)

Taking natural log of both sides of the above equation, we get

$lnY_t = lna + lnct + bX_t + e_t$, t =1, 2, …,n

A model of the above form is referred to as a log-level regression model. If we assume $Z_t = lnY_t$, $a_0 = lna$ and $b=a_1$, then the above model becomes

$Z_t = a_0 + a_1 X_t + u_t$, t =1, 2, …,n

The estimated model becomes

$\hat{Z}_t = \hat{a}_0 + \hat{a}_1 X_t$, where $\hat{a}_0 = e^{a0}$ and b = $\hat{a}_1$

Exponential multiple regression (EMR) model

The EMR model is an extension of the ER model, can be described as follows:

$Y_t = ae^{b_1X_{1t}+b_2X_{2t}+\cdots+b_kX_{kt}} + e_t$, t =1, 2, …, n (4)

where e_t are identically and independently distributed with mean 0 and variance 1. Taking natural log of both sides of the above equation, we get

$lnY_t = lna + b_1X_{1t} + b_2X_{2t} + \cdots + b_kX_{kt} + e_t$, t =1,2,…, n

This is referred as a log-level MLR model. If we assume $Z_t = lnY_t$, $a_0 = lna$ and $b=a_i$, then the above model becomes

$Z_t = a_0 + a_i X_t + u_t$, t =1, 2,…,n

Thus, the estimated model is

$\hat{Z}_t = \hat{a}_0 + \hat{a}_i X_t$, where $\hat{a}_0 = e^{a0}$ and $b = \hat{a}_i$

More on the ER model we refer to Al-Dawsari, et al. [8], Ahmad et al. [9], Joseph, et al. [10] among others.

Time series model

A time series is a structure of discrete-time data taken at consecutive equally spaced points in time. Time series permits us to study major patterns such as trends, seasonality, cyclicity and irregularity. The most common examples of time series are daily closing value of stock indices, ocean tides, rainfall, temperature, humidity and many others. To model a time series, time series forecasting is a statistical tool to forecast the output through a sequence of time. The available procedures forecast future events by studying the trends of the preceding under the hypothesis that future trends will show similar to past trends. The Box-Jenkins ARIMA models are the most widely used approaches in time series forecasting for univariate time series data. Time series analysis is applied for numerous applications such as stock market analysis, pattern recognition, economic and financial forecasting, earthquake prediction, census analysis and so on (See Banik et al [11], Banik and Silvapulle [12], Banik et al [13] and others). Forecasts generally evaluate by using numerical measures such as mean square error, average error, root mean squared error and others. We have used the following widely used time series forecasting models to identify the patterns of COVID-19 active cases in the case of the considered developing country Bangladesh:

Autoregressive exogenous (ARX) model

It is an auto-regressive (AR) model with exogenous input, determined outside of the process we will model. Why is the ARX model different from the standard models? To understand it, consider the following example: Suppose you consider a model which is trying to predict the industrial output. Here we want to include lagged output (the industrial capability is supported from one period to the next) and lagged interest rates (the previous price of money effects to present links). Mutually lagged output and lagged interest rates are endogenous to the structure. What special effects can also mark interest rates? Here we believe that an exogenous variable would be an oil crisis or expected calamity. These measures occurred anyway regardless of the prices of production or interest rates. To understand an ARX modelling structure, the mathematical black box model structure is unconfined. Overall, the system is identified by a black box. The input is only to control the model, and both of the input and output are traced and noted. Though, the system model is concerned with error terms. Usually, inputs and outputs are measurable, whereas the system disturbance is not categorized straight. This has an effect on the outputs. The modelling error is out as the difference between the system output and the model output. The ARX time series model is used for many areas, especially in process dynamics and control analysis. It has also been used since the 1980s in the process industries such as chemical plants and oil refineries. The input-output relationship of an ARX model is given by a linear equation as follows:

$Y_t + a_1Y_{t-1} + \cdots + a_{na}Y_{t-na} = b_0X_{t-d} + \cdots + b_{nb}X_{t-d-nb} + e_t$,
t = 1, 2, …,n (5)

where Y_t is the model output (daily active cases), X is the model input (daily positive cases), a_i is the na^{th} AR parameters, b_i is the nb^{th} exogenous (X) parameters and e_t is the error term. In solving the problem of parameter identification, the adjustable parameters of this structure are expressed as below:

$\theta = [a_1, a_2, \ldots, a_{na}, b_0, b_2, \ldots, b_{nb}]$

Given that $A(q^{-1}) = 1 + a_1 q^{-1} + \cdots + a_{na} q^{-na}$ and $B(q^{-1}) = b_0 + b_1 q^{-1} + \cdots + b_{nb} q^{-nb}$, where q^{-1} is the backward shift operator, the following input-output model structure is informed.

$$A(q^{-1})Y_t = B(q^{-1})X_t + e_t$$

Therefore, $Y_t = \frac{B(q^{-1})}{A(q^{-1})} X_t + \frac{e_t}{A(q^{-1})}$. More on the ARX model we refer to Ashar, et al [14], Haggan-Ozaki, et al [15], Mechaqrane, et al. [16] among others.

VARX model

The VARX model is a statistical vector autoregressive (VAR) model with exogenous variables frequently used in many studies such as solving time series data, e.g., finance, economics, applied econometrics, business and others. This model can enlighten the dynamic performance of the relationship between dependent and independent variables. The VARX is also of attention in situations where one wishes to predict from only a division of a longer time series data set. In marketing, for example, one may practice this model to forecast sales of numerous associated products using their past sales, worth and promotion information. In a VARX(p,s), a series Y_t is modeled in terms of its own past p values and the past s values of an exogenous series of X_t:

$$Y_t = \alpha + \sum_{i=1}^{p} \varphi_i Y_{t-i} + \sum_{j=1}^{s} \beta_j X_{t-j} + e_t, \; t = 1, 2, \ldots, n \quad (5.6)$$

where φ_i are endogeneous lagged coefficient matrices, β_j are exogeneous lagged coefficient matrices and e_t are random error terms mean zero white noise vector time series. The parameter estimates can be obtained by representing the general form of the multivariate linear model Y = Xβ+E,

where $Y = (Y_1, Y_2, \ldots, Y_t)'$, $\beta = (\alpha, \varphi_1, \varphi_2, \ldots, \varphi_p, \beta_1, \beta_2, \ldots, \beta_s)'$, $X_t = (1, Y_t', \ldots, Y_{t-p+1}', X_{t+1}', \ldots, X_{t-s+1}')'$ and $E = (e_1, e_2, \ldots, e_t)'$

The conditional OLS of β can be obtained by using the same technique in a VAR(p) modelling to estimate a VARX model. If this model has different independent variables that correspond to dependent variables, the seeming unrelated regression method is generally used to develop the regression estimates. More on VARX model we refer to Septiani et al. [17], Russel et al. [18], Tuttle et al. [19] among others. The data analyses for all of the models are given in the following section.

Data Analysis

In this paper, MATLAB 2018 Programming languages are used for all computations. For a sample of MATLAB program, see Appendix.

Fitted models

In this section, we will analyze the data and discuss the outcomes of our study based on various considered forecasting models graphically and numerically. The regression analyses for model (1) are presented in Table 3.

Model: $Y_t = a + ct + bX_t + e_t$, t =1, 2, …,387

From Table 3, we can see that the RMSE value is 0.5790, which indicates that the model (1) can predict the daily active cases reasonably. R^2 value is 0.8510, meaning 85% of the variance of the daily active cases is explained by the daily active cases.

The fitted model for the daily active cases is found to be:

$\hat{Y}_t$ = 2.1365 - 0.0049t – 1.0947X_t, t = 1, 2, …, 387

To observe whether the daily active cases can be predicted in a more accurate way or not, two more important variables, daily deaths and daily recovered cases, are added and the regression analyses are tabulated in the Table 4.

Model: $Y_t = \beta_0 + ct + \beta_1 X_{t1} + \beta_2 X_{t2} + \beta_3 X_{t3} + e_t$, t =1,2, … ,387

The RMSE and R^2 values in Table 4 indicated that the model in (2) has better predictive power than the model (1).

Table 3: Linear regression results.

	Estimates	Standard Error	t-Statistic	p-Value
Constant	2.1365	0.1919	11.1310*	4.6482e-25
Time trend	0.0049	0.0002	18.2780*	7.9936e-54
Daily positive cases	1.0947	0.0261	41.8210*	2.8004e-143
Root mean square error	0.5790			
R^2	0.8510			
F-statistic (p-value)	1.07e+03* (1.01e-155)			

Table 4: Multiple linear regression (MLR1) results.

	Estimates	Standard Error	t-Statistic	p-Value
Constant	9.0658	0.2415	37.5280*	2.0241e-119
Time trend	-0.0023	0.0002	-10.1790*	2.8382e-21
Daily positive cases	-0.0563	0.0347	-1.6215	0.1059
Deceased	0.1216	0.0409	2.9674*	0.0032
Recovered cases	0.3881	0.0224	17.3280*	5.994e-48
Root mean square error	0.2330			
R^2	0.8800			
F-statistic (p-value)	217* (5.16e-90)			

The fitted model for the daily active cases is found to be:

$\hat{Y}_t$ = 9.0658 - 0.0023t - 0.0563X_{t1} + 0.1216X_{t2} + 0.3881X_{t3}33t =1,2,…, 387

Since our target is to predict daily active cases in an accurate way as much as possible, some important environmental variables, BD daily average temperature and daily humidity are included as independent variables with previous independent variables. Outcomes are reported in the Table 5

Model: $Y_t = \beta_0 + ct + \beta_1 X_{t1} + \beta_2 X_{t2} + \beta_3 X_{t3} + \beta_4 X_{t4} + \beta_5 X_{t5} + e_t$, t =1,2, … ,387

Table 5: Multiple linear regression (MLR2) results.

	Estimates	Standard Error	t-Statistic	p-Value
Constant	-1.374e+05	8382.9	-16.391*	2.4915e-44
Time trend	264.32	13.449	19.654*	4.5867e-57
Daily positive cases	7.6251	1.4564	19.654*	4.5867e-57
Deceased	927.03	111.09	8.3446*	2.1279e-15
Recovered cases	5.3146	0.79276	6.7038*	9.1049e-11
BD_Tem	4247.3	334.08	12.713*	2.763e-30
BD_ Humidity	137.76	110.44	1.2474	0.21317
Root mean square error	1.52			
R^2	0.6523			
F-statistic (p-value)	415 (3.34e-148)			

It is observed from Table 4 that all the regressors (independent variables) are statistically significant except humidity. If we review the

RMSE and R^2 values, it is clear that this linear predictive model can forecast the daily active cases less accurately as compared to models from Tables 3 & Table 4. The fitted model daily active cases on daily positive cases, daily deaths, daily recovered cases, daily temperature and daily humidity is found to be:

$\hat{Y}_t =$ -1.374e+05 + 264.32t + 7.6251X_{t1} + 927.03X_{t2} + 5.3146 X_{t3} + 4247.3X_{t4} + 137.76X_{t5}, t = 1,2, … , 387

We have also fitted some non-linear models beside linear models, namely ER forecasting models, ARX model and VARX models. Results are tabulated in Tables 6-11.

Model: $Y_t = ae^{ct+bX_t} + e_t$, t =1,2,…,387

In Table 6, we have reported the non-linear ER results, where daily active cases are predicted based on daily positive cases. Table 6 indicates that ER model can forecast the daily active cases less accurately compared to models from Tables 3 & Table 4.

The fitted ER model daily active cases on daily positive cases is found to be:

$\hat{Y}_t = 7.0146\, e^{0.0094t+0.0011X_t}$, t =1, 2,…,387

To increase prediction capability, two more independent variables, daily deaths and daily recovered cases are included and the results are tabulated in Table 7. We found that RMSE value is smaller and R^2 value is higher than the ER model. This suggests that this MER prediction model has better prediction power compared to the ER model.

$Y_t = ae^{ct+b_1X_{1t}+b_2X_{2t}+b_3X_{3t}} + e_t$, t =1, 2,…, 387

The fitted MER model daily active cases on daily positive cases, daily deaths and daily recovered cases is found to be:

$\hat{Y}_t = 6.8196\, e^{0.008t + 0.0005 X_{1t}+0.0503X_{2t}+0.00006X_{3t}}$, t =1,2,…, 387

In the Table 4.6, we have reported the multiple exponential regression prediction model results, where independent variables are daily deaths, daily recovered cases, average temperature and daily humidity to see whether environmental variables have impact on the daily active cases.

Model: $Y_t = ae^{ct+b_1X_{1t}+b_2X_{2t}+b_3X_{3t}+b_4X_{4t}+b_5X_{5t}} + e_t$, t =1,2,…, 387

If we compare the results in Table 5-8, we observe that the model in Table 4.6 has better predictive power than the models that are from Table 5-7 in the sense of smaller RMSE and higher R^2 values.

Finally, we have applied two non-linear time series of models, namely the ARX model and the VARX model and regression analyses are presented in Table 9-11.

Table 6: Exponential regression results.

	Estimates	Standard Error	t-Statistic	p-Value
Constant	7.0146	0.1542	45.4800*	1.0878e-156
Time trend	0.0094	0.0005	18.2450*	5.2419e-54
Daily positive cases	0.0011	5.581e-05	20.7800*	8.2226e-65
Root mean square error	1.12			
R^2	0.63			
F-statistic (p-value)	326* (1.57e-83)			

Table 7: Exponential multiple regression (ER1) results.

	Estimates	Standard Error	t-Statistic	p-Value
Constant	6.8196	0.14414	47.311*	8.1829e-162
Time trend	0.0084416	0.00049899	16.918*	2.5701e-48
Daily positive cases	0.00059086	8.383e-05	7.0483*	8.5188e-12
Deceased	0.050388	0.0068027	7.4071*	8.3463e-13
Recovered cases	6.3067e-05	5.1439e-05	1.2261	0.22093
Root mean square error	1.03			
R^2	0.69			
F-statistic (p-value)	213*(7.05e-96)			

Table 8: Exponential multiple regression (ER2) results.

	Estimates	Standard Error	t-Statistic	p-Value
Constant	-1.374e+05	8382.9	-16.391*	2.4915e-44
Time trend	264.32	13.449	19.654*	4.5867e-57
Daily positive cases	7.6251	1.4564	5.2355*	2.9761e-07
Deceased	927.03	111.09	8.3446*	2.1279e-15
Recovered cases	5.3146	0.79276	6.7038*	9.1049e-11
BD_Tem	4247.3	334.08	12.713*	2.763e-30
BD_Humidity	137.76	110.44	1.2474	0.21317
Root mean square error	0.8945			
R^2	0.885			
F-statistics (p-value)	415* (3.34e-148)			

We have fitted several ARX time series models, where the dependent variable is the daily active cases and the independent variable is the daily positive cases. We have selected one of the best one (ARX(1,1,4)) by using the Akaike Information Criterion (AIC) and presented results in Table 9.

Model: $A(q^{-1})Y_t = B(q^{-1})X_t + e_t$, t = 1, 2, …, 387

The fitted model daily active cases on daily positive cases, daily deaths, daily recovered cases, daily temperature and daily humidity is found to be:

$$\hat{Y}_t = -1.374e05\, e^{264.32t+7.6251\ X_{1t}+927.03X_{2t}+5.3146X_{3t}+4247.3X_{3t}+137.76X_{5t}},\ t=1,2,\ldots,387$$

Table 9: ARX (1,1,4) time series model.

Root Mean Square Error	1.0918
Fit to estimation data (R^2)	97.47%
Final prediction error (FPE)	1.211

In the Table 10 (due to a space problem, results were not possible to tabulate but available upon request), we have reported results of the estimated VARX (4) nonlinear model, where independent variables are daily positive cases, daily deaths and daily recovered cases. The order of the model is detected by the AIC. It is observed that the best estimated model was the order of 4.

$$Y_t = \alpha + \sum_{i=1}^{4} \varphi_i Y_{t-i} + \sum_{j=1}^{3} \beta_j X_{t-j} + e_t,\ t=1,2,\ldots,387$$

This model is also evaluated by two important error measures RMSE and R^2. It is observed that RMSE value (1.673) is higher and R^2 value (0.8231) is lower than the tabulated result of the Table 4. The fitted model daily active cases on daily positive cases is found to be:

$$\hat{Y}_t = \frac{1.174X_t^{-4}}{1-0.9775X_t^{-1}} + \frac{1}{1-0.9775\,X_t^{-1}},\ t = 1, 2, \ldots, 387$$

Table 10: The VARX1(4,3) model.

Root mean square error	1.673
R^2	0.8231

We have reported estimation results of the model with independent variables: daily positive cases, daily deaths, daily recovered cases, daily temperature and daily humidity in the Table 11 (results are not visible due to space problem; however, they are obtainable upon request).

$$Y_t = \alpha + \sum_{i=1}^{4} \varphi_i Y_{t-i} + \sum_{j=1}^{5} \beta_j X_{t-j} + e_t,\ t = 1, 2, \ldots, 387$$

As comparing to the estimation results tabulated in the Table 5, it is found that this time series non-linear prediction model has lower RMSE value (0.67) and higher R^2 value (0.9085), higher prediction ability as comparing to the MLR prediction model.

Table 11: The VARX2(4,5) model.

Root mean square error	0.670
R^2	0.9085

To compare the models performance, we have presented RMSE and R^2 values in Figures 3 & Table 4 respectively.

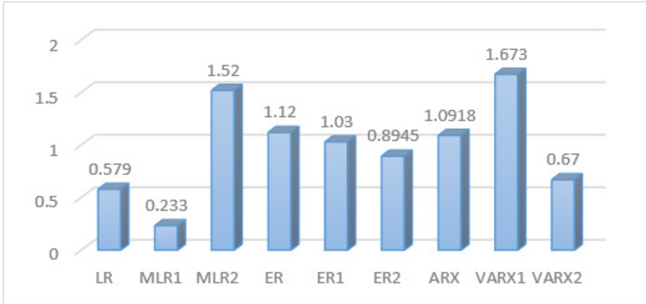

Figure 3: RMSE values of all considered prediction models.

Figure 4: R^2 values of selected prediction models.

It can be observed from the Figure 3 that according to the values of RMSE, the MLR1 model performed the best followed by LR, VARX2 and ER2 where independent variables were daily positive cases, daily deaths and daily recovered cases. According to the Figure 4, the ARX model performed the best followed by VARX2, MLR1, ER2 and LR where daily active cases were predicted based on exogenous variable daily positive cases. If we consider both RMSE and R^2 values, we can see that LR, MLR1, VARX2 are performing better than the rest of the models.

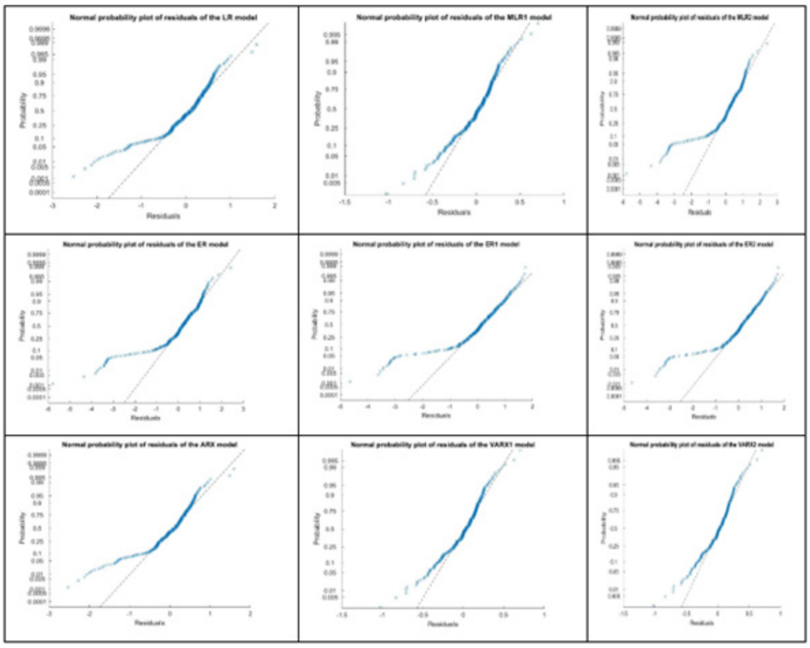

Figure 5: Normal probability plots of the models.

Citation: Banik S, Answer M, Kibria BMG. Prediction of Daily COVID-19 Active Cases in Bangladesh. *Int. Jr. Infect Dis & Epidemlgy*. 2021;2(1):25–33. DOI: 10.51626/ijide.2021.02.00006

Model assumptions

We will discuss the model assumptions in this section. First, we have plotted the normal probability plot of the raw residuals of models in Figure 5. If we review Figure 5, we can see that the MLR1, the VARX1 and the VARX2 models satisfied the normality assumptions.

Next we have plotted the residuals versus the fitted values of our selected models in Figure 6. If we review Figure 6, we can see that the MLR1, VARX1 and the VARX2 models satisfied constant variance assumptions to some extent.

Based on our detailed analysis and findings, it is possible to conclude that the VARX models can be used to forecast BD daily active cases of the ongoing Pandemic COVID-19 deadly virus.

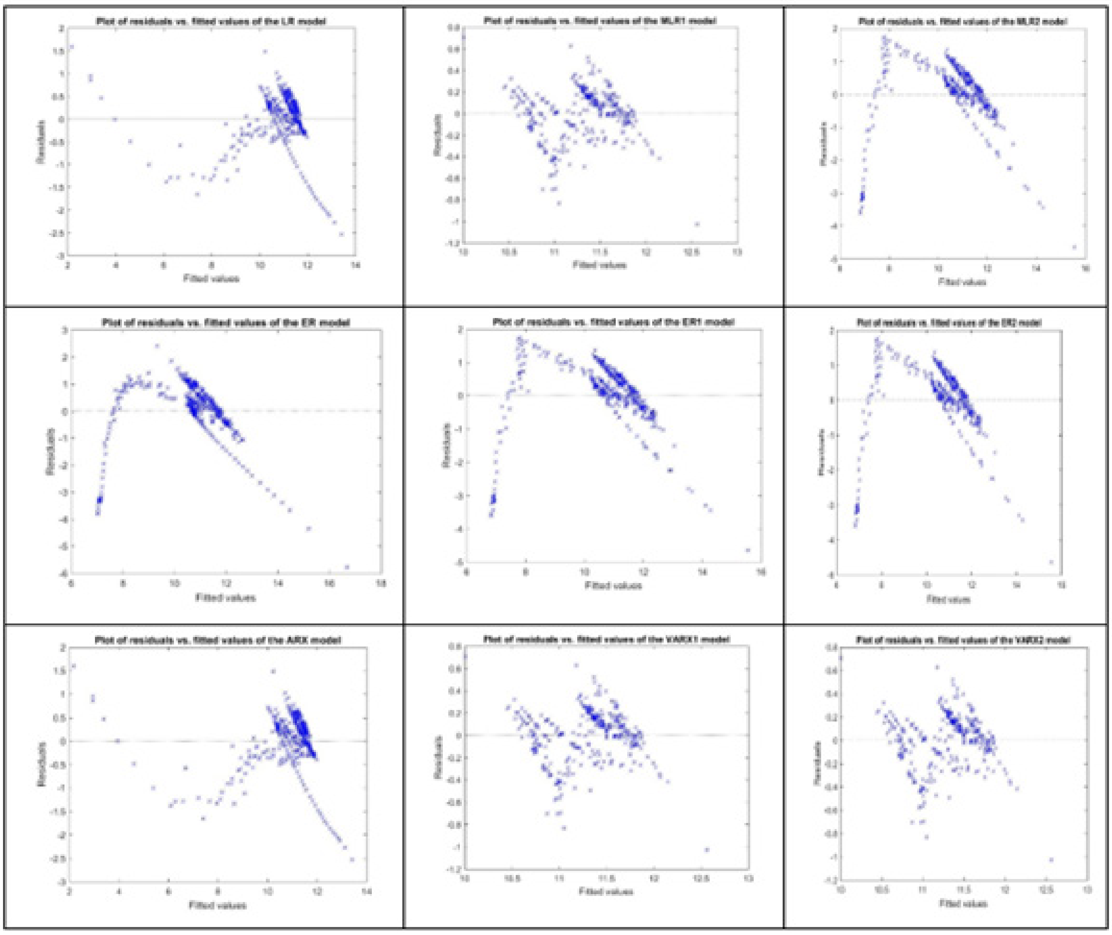

Figure 6: Residual vs fitted values.

Some Concluding Remarks

In judging the COVID-19 circumstances, daily active cases perform to afford a more precise image to recognize the present pandemic disaster. This paper predicts daily active cases for Bangladesh based on some important independent variables such as daily positive cases, daily deaths, daily recovered cases and also some related environmental variables such as average daily temperature and daily humidity. We have used several statistical prediction models: The linear regression model, the exponential regression model, the auto-regressive exogenous model and the vector autoregressive exogenous model to predict daily active cases based on some important considered independent variables. It is observed that the vector auto-regressive exogenous model predicted daily active cases better compared to other models considered in this paper. We believe the findings of our paper will make an important contribution in the literature of the ongoing deadly COVID-19 pandemic.

References

1. Barría-Sandoval C, Ferreira G, Benz-Parra K, López-Flores P (2021) Prediction of confirmed cases of and deaths caused by COVID-19 in Chile through time series techniques: A comparative study. PLoS ONE 16(4): e0245414.
2. Mahanty C, Kumar R, Pashra D B K, Hemanth J, Gupta D (2020) Prediction of COVID-19 active cases using exponential and non-linear growth models.
3. Sarkar D (2020) COVID 19 Pandemic: A real-time forecasts and prediction of confirmed cases, active cases using the ARIMA model and public health in West Bengal, India.
4. Puttahonnappa K, Sharanagouda S S, Prosad P, Gowda S, Hemadri R D, et al. (2021) Prediction of daily and cumulative cases for COVID-19 infection based on reproductive number (R_0) in Karnataka: data-driven analytics. Scientific Reports 11(1): 10075.
5. Viswanath N C (2020) Prediction of COVID-19 active and total cases after a fall and rise of cases.
6. Rath S, Tripathy A, Tripathy A R (2020) Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model. Diabetes and Metabolic Syndrome 14(5): 1467-1474.
7. Hassan R, Dosar A S, Mondol J K, Khan T H, Noman A A, et al. (2020) Prediction of epidemics trend of COVID-19 in Bangladesh. Front. Public Health 8: 559437.
8. Al-Dawsari S R, Sultan K S (2021) Modeling of daily confirmed Saudi COVID-19 cases using inverted exponential regression. Math Biosci Eng 18(3): 2303-2330.

9. Ahmad W M A W, Aleng N A, Halim N A, Dasril Y, Baharum A (2017) Applied exponential growth regression modeling using SAS: An alternative method of Biostatistics. International Journal of Applied Engineering Research 12(18): 7853-7856.

10. Joseph J, Mathew K (2020) Exponential regression via linear modeling using stochastic gradient descent and optimization. International Journal of Science and Research 9(12): 9-11.

11. Banik S, Chanchary F H, Rouf R A, Khan A K M (2012) Modeling chaotic behavior of Dhaka stock market index values using the neuro-fuzzy model. Recent Patents on Computer Science 5(1): 72-77.

12. Banik S, Silvapulle P (1999) Testing for seasonal stability in unemployment series: International evidence. Empirica 26 (2): 123-139.

13. Banik S, Khan A F M, Anwer M (2014) Hybrid machine learning technique for forecasting Dhaka stock market timing decisions. Computational Intelligence and Neuroscience.

14. Ashar N D B K, Yusoff Z M, Ismail N, Hairuddin M A (2020) ARX model identification for the real-time temperature process with Matlab-arduino implementation. ICIC Express Letters 14(2): 103-111.

15. Haggan-Ozaki V, Ozaki T, Toyoda Y (2006) RBF-ARX modeling for prediction and control. 14th IFAC Symposium on System Identification 39(1): 1210-1215.

16. Mechaqrane A, Zouak M (2004) A comparison of linear and neural network ARX models applied to a prediction of the indoor temperature of a building. Neural Computing & Applications 13(1): 32-37.

17. Septiani A, Sumertajaya M, Aidi M N (2019) Vector autoregressive X (VARX) modeling for Indonesian macroeconomic indicators and handling different time variations with cubic spline interpolation. Engineering and Technology 6(1).

18. Warsano, Russel E, Widiarti W, Usman M (2019) Vector autoregressive with exogenous variable model and its application in modeling and forecasting energy data: Case study of PTBA and HRUM energy. International Journal of Energy Economics and Policy 9(2): 390-398.

19. Tuttle J F, Landen D, Klas B, Powell K M (2021) A systematic comparison of machine learning methods for modeling of dynamic processes applied to combustion emission rate modeling. Applied Energy 292: 116886.