International Journal of Applied Sciences: Current and Future Research Trends
(IJASCFRT)
ISSN (Print) , ISSN (Online)

https://ijascfrtjournal.isrra.org/index.php/Applied_Sciences_Journal

Modeling and Forecasting the Spread of Novel Corona Virus Disease (COVID-19) Number of New Cases in Ethiopia

Tesfaye Denano Sorre*

Department of Statistics, Wolaita Sodo University, Ethiopia
Email: Tesfaye.d@wsu.edu.et

Abstract

The novel Corona virus disease (COVID-19) was first found in Wuhan, China in December 2019, but later spread to other parts of the world. The World Health Organization (WHO) defines a pandemic as "the worldwide spread of a new disease". The main aim of this study was to forecast the spread of COVID-19 new cases in Ethiopia. In this study we used the data of confirmed Coronavirus Disease (COVID-19) cases reported daily from March 13, 2020 until April 4, 2021 that were obtained from ministry of health Ethiopia and WHO. ARIMA model was applied to examine the data used to generate 60 days forecast. ARIMA (0, 1, 1) estimates the number of confirmed COVID-19 new cases based on a 95 percent confidence interval between March 13, 2020 and April 4, 2021. In the next two months, the maximum expected new cases per day were 2353, while the minimum estimate was 2042 cases per day. Furthermore, by the end of May 2021, the total number of confirmed COVID-19 predicted cases could reach about 343,563. Although more data are needed to have a more detailed prevision, the spread of the virus seems to be increasing according to time series plot. However the forecasting ARIMA (0, 1, 1) model shows fast increasing within a predicted time period in Ethiopia. According to the study results, the government and all concerned bodies should work to stop the pandemic from spreading through Ethiopia. Increasing the regular test potential for Covid-19 by increasing test sites and implementing successful and efficient behavioral change and communication interventions should be considered throughout the country.

Keywords: COVID-19 new cases; ARIMA model; Forecasting; Ethiopia.

----

* Corresponding author.

1. Introduction

Corona viruses are enveloped single-stranded RNA viruses that infect animals, including humans, and are known for causing symptoms that range from flu-like symptoms to extreme respiratory, enteric, hepatic, and neurological symptoms [31]. In December 2019, the new corona virus (Covid-19) was found for the first time in Wuhan, China. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causing corona virus disease 2019 (COVID -19) and then invades host human cells by binding to the angiotensin - converting enzyme 2 (ACE2) receptor [32]. Since then the world is on a fight with invisible enemy that spreads throughout the countries in a fast speed. WHO stated novel corona virus disease as a world public health emergency by declaring as pandemic on 11 March 2020 by considering its impact on the entire world population [1]. Covid-19 is associated with mild to severe reparatory syndrome. According to CDC the symptoms of corona virus starts with in 2 up to 15 days after a person exposed for a virus [18].Cough, fever and shortness of breath are the most important symptoms to suspect the case has a corona virus [10,24]. Now days emerging findings indicate that the virus has devastating consequences on vital organs like kidney, heart and lung and so far ends in death. The primary way of transmission is from person to person through direct contact with a person infected corona virus within 6 feet. It transmits through reparatory droplets when a person coughs, sneezes or talks. Corona virus also can get the route of entry through touching the surface or object has a virus and then by touching the mouth, nose or eyes. CDC recommends avoid contact with sick people, wash hands thoroughly with soap and water, if you develop cough, use a medical face mask, physical distancing and stay home if there is any symptom of Covid-19 and seek medical care for the additional help [24]. As a new virus, there was a time that countries confused to confront the invasion of it. However through experiences and scientific evidences, the way of transmission and how to prevent the pandemic was recognized to be implemented by governments and every person in every setup. Corona virus has no limit on a specific ethnicity, nation, religion, age and gender and it affects all sides of human being without any separation. Even though the virus affects all age groups, high mortality and severity were recorded in older ages and have underlying comorbidties [9]. The world is facing unprecedented impact of corona virus that all aspects of life including travel, trade, tourism, food supplies, and financial markets affected [18]. Saris covid-2 is a new virus threatening the health of human being which infected more than five point nine million people and with death of more than three hundred sixty five thousand people around the globe until the recent time. The lethality from corona virus is more than 20% in elderly; particularly on those have comorbities like chronic non communicable diseases and HIV/AIDS [19]. USA, European countries like United Kingdom, France, Italy and Spain are the most affected countries by morbidly and mortality due to the pandemic, even if the Latin American country, Brazil, becoming the new epicenter of the corona virus pandemic [12]. Although Africa is the least affected continent by this pandemic, the number of new infections and deaths are increasing from time to time. Currently around one hundred thousand people are infected and three thousand deaths were recorded. WHO and studies done by using mathematical model indicated that the worst time for Africa is yet to come and a millions of people will be infected and a thousands of death will probably be happen. It is the hardest time for the continent that the pandemic affected overall economic and social aspects of the people. Moreover the health system of the continent especially Sub-Saharan Africa is the weakest that sub-continent is still struggling to overcome HIV/AIDS epidemic, TB, Malaria, under nutrition and chronic non communicable diseases. Shortage of medical equipment, intensive care unit (ICU) and adequate human power for treating cases are the worst challenges to confront with corona virus in SSA. Nevertheless, lack of access to potable water and sanitizers and overcrowded living condition increase the chance to get the infection. Reports indicate that sub Saharan countries are not adequately prepared for alarmingly increasing number of Covid-19 cases [3, 17, 21]. The first case of COVID-19 was confirmed in Ethiopia on 13 March 2020 since then the number of cases is being increasing from the day to day. Up to April 16, 2021 the total number of confirmed COVID-19 cases, recovery from the coronavirus diseases and total deaths were 236,554, 175,879 and 3,285 respectively. The real number may be more but it is unknown may be due to shortage of daily tested suspected persons and unable to address the whole community. Ethiopian government is striving to overcome the pandemic by declaring state of emergency for the second time by considering the influence of disease on overall aspects of people living in developed and developing countries without placing a limit for the boundaries. The government stated that the pandemic could have devastating consequence due to the weak socioeconomic and health system in the country [2, 23, 25]. The time that coronavirus disease (covid-19) ends does not really known and the disease remain without drug for treatment even though there are hopeful vaccine and drug trials including the traditional herbal medicines around several countries of the globe. The number of the new cases without contact history continues according to daily reports of Ethiopian ministry of health. Moreover the negligence on prevention methods, social intimacy and overcrowded living conditions in Ethiopia may prone for the increase of the number cases and death in the country [4, 25]. When writing this, the new cases were found overall regions of Ethiopia. Timely information on real situation and forecasting for the spread of the pandemic of COVID-19 cases in Ethiopia is very important for decision making and to be ready to curb it in efficient manner. There for this study will forecast the spread of the cases of Covid-19 in Ethiopia by using limited data.

2. Methods and Materials

2.1. Data

According to the World Health Organization, a pandemic is the worldwide spread of a new virus. The distribution of COVID-19 confirmed cases was taken from the Ministry of Health Ethiopia 2020 and WHO for the validation and review of the proposed sample. For time series forecasting, a minimum sample size of 30 observations is needed [29]. As a result, in this analysis, we used 388 time-series observations from March 13, 2020 to April 4, 2021 to forecast the spread of COVID-19 new cases over the next 60 days with a 95% confidence interval limit. To predict this new pandemic virus, we compared various time series models. We analyzed the data with ARIMA models and predicted 60-day forecasts based on the minimum values of MAPE, MAD, and MSD.

2.2. Time Series Models

2.2.1. Autoregressive Integrated Moving Average (ARIMA) Model

Time series is a set of measurements taken at the same time that can be measured in a continuous or discrete manner [29]. A use of Time series is to look at previous observations and construct a model that can forecast future values. The ARIMA model [11] combines the autoregressive AR(p), moving average MA(q), and ARMA(p,q) models. Order of Autoregression, degree of difference, and order of moving average are represented by the letters p, d, and q, respectively. In an AR(p) model, the current time series value Y_t is expressed as a linear combination of p past observations Y_{t-1} $Y_{t-2}\ldots\ldots Y_{t-p}$ and a random error ε_t, together with a constant term. Similarly, in an MA(q) model, the current time series value Y_t uses past q error terms ε_{t-1} $\varepsilon_{t-2}\ldots\ldots\varepsilon_t$ as the explanatory variables. The general formula for AR(p) and MA(q) models can be expressed as in equation (1) and (2) respectively.

$$Y_t = K + \phi_1 Y_{t-1} + \ldots\ldots + \phi_p Y_{t-p} + \varepsilon_t \quad (1)$$

$$Y_t = \mu + \varphi_1 \varepsilon_{t-1} + \ldots\ldots + \varphi_q \varepsilon_{t-q} + \varepsilon_t \quad (2)$$

Here ϕ_i (i=1, 2...p) and φ_j (j=1, 2...q) are the autoregressive and moving average parameters respectively. Y_t is the observed value at time t and ε_t the random error at time t. K is the constant term, and μ *is* the mean of the series. The random shock is assumed to be a white noise process, that is, a sequence of independent and identically distributed (i.i.d) random variables with mean zero and a constant variance σ^2 [29]. The ARMA(p, q) model is a combination of AR(p) and MA(q) models in which the current time series value Y_t is defined linearly in terms of its past p observations as well as the current and past q random shock, together with a constant term. The general formula of an ARMA(p, q) model can be expressed as in equation (3).

$$Y_t = K + \phi_1 Y_{t-1} + \ldots\ldots + \phi_p Y_{t-p} + \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \ldots\ldots + \varphi_q \varepsilon_{t-q} \quad (3)$$

Where K is a constant and ε_{t-k} (k=1, 2…q) are the values of the previous random shock. Time series analysis requires a stationary time series, that is, the series shows no fluctuation or periodicity with time [22]. In an ARIMA model, a non-stationary time series is made stationary by applying finite differencing to the time series. The differenced stationary time series can be modeled as an ARIMA model to perform an ARIMA forecasting [11, 22].

2.2.2. ARIMA Forecasting Model

Time series forecasting is a multidisciplinary scientific tool used to solve prediction problems. Its implementation is easy and flexible because it only requires historical observations of the necessary variables [7, 15, 26]. ARIMA was first presented by Box and Jenkins in 1976 [15]. The general equation of successive differences at the d^{th} difference of X_t is as follows:

$$\Delta^d x_t = (1 - L)^d x_t \quad (4)$$

where d is the difference order and is usually 1 or 2, and L is the backshift operator.

The successive difference at one-time lag equals to

$$\Delta^1 x_t = (1-L)x_t = x_t - x_{t-1} \quad (5)$$

In this work, the general ARIMA (p, d, q) is briefly expressed as follows [30]:

An ARIMA (p, d, q) process is defined as

$$\begin{aligned} &\phi_p(L)^d y_t = \varphi_q(L)\varepsilon_t \\ &y_t = \left(\frac{1}{\phi_p(L)^d}\right)\varphi_q(L)\varepsilon_t \end{aligned} \quad (6)$$

Where $\phi_p(L)$ and $\varphi_q(L)$ are polynomials in (L) and d is the order of integration or differencing.

Where y_t represents (1) the series of differences, ϖ_i are the coefficients, the y_{t-p} and ε_{t-q} are the lagged predictors for the model [8]. The ARIMA model is a generalization of many sub-models and characterized by three parameters: order of autoregressive observations p, degree of differencing d, and number of moving average terms q in equation (6). A data series is said to be stationary if its error term has zero mean, constant variance and the covariance between any two - time periods depends only on the distance or lag between the two periods and not on the actual time. One of the methodologies that use ARIMA processes and differencing as basis is the Box and Jenkins method [14]. According to the prescription of Box and Jenkins a few steps have to be taken when building an ARIMA model [13]. The first step is to determine if the series is stationary to achieve this one has to take as many differences of the original series as are needed to reduce it to stationary.

2.2.3. Steps of ARIMA modeling

There are four steps involved in the ARIMA modeling, namely, identification, estimation, diagnostic checking, and forecasting. The first step is to check the seasonality and stationarity of the time series data by drawing a time series plot of the observed series with the corresponding time. A time series is considered as stationary if a shift in time doesn't cause a change in the shape of the distribution, that is, the statistical properties such as mean, variance, and autocorrelation are constant over time. The stationarity of time-series data is important as it helps develop powerful techniques to forecast future values [13, 16]. The second step is to construct the autocorrelation (ACF) and the partial autocorrelation (PACF) plots of the stationary time series to determine the order of the AR and MA processes. The ACF is the correlation between the observation at time t and the observation at a different time lag, while PACF is the amount of correlation between the current observation at time t and the observation at lag k that is not explained by the correlation at all lower-order lags (that is, lag<k [8, 14, 28]. The third step involves estimating the parameters of the best fit model, which is done using the performance measure criteria. The ACF plot of residuals as well as the Box Pierce test of white-noise, were determined to evaluate the model goodness of fit. The fourth step involves forecasting future values using a good fit model. Stationary of the data checked by: Phillips and Perron (PP) and Augmented Dickey-Fuller (ADF) test. PP proposes an alternative method of controlling for serial correlation when testing for a unit root

[27] To avoid the pitfall of wrong inferences from the non-stationary regressions, the time series data should be stationary. Hence, before proceeding any further it is very important to ensure that the underlying data generating process are stationary. If the variables have unit root (non-stationary), then the series needs to be differenced to achieve stationarity [16].The foundations of ARIMA lie on the fact that non-stationary time-series can be made stationary by through differencing. We used autocorrelation function ACF and partial autocorrelation function PACF plots for model stationarity check [8, 28]. Stationary of the data checked by: Phillips and Perron and Augmented Dickey-Fuller (ADF) test. This method estimates the non-augmented DF test and modifies the t-ratio of α coefficient so that serial correlation does not affect the asymptotic distribution of the test statistic. Another advantage of using Phillips and Perron test of unit root is that it is robust to general forms of heteroscedasticity in the error term and does not have to specify a lag length for testing the regression. This test constructs a parametric correction for higher order correction by assuming that the series y_t follows an autoregressive or AR (P) process and adding P lagged difference terms of the dependent variable to the right hand side of the test regression [27].

2.3. Exponential Growth Trend Model

Single exponential growth model is one of the best time series model in forecasting data that have growing nature by accounting an exponentially growing (decaying) data.

Suppose a variable y_t is an exponential function of time (t), then the model is given in equation (7):

$$y_t = \beta \exp^{\gamma t} \tag{7}$$

Where, y_t is the number of COVID-19 confirmed cases, β is the initial value of y_t, γ is a positive growth factor and t is the time constant required for y to increase by one factor of γ.

In order to estimate the exponential growth model using OLS, we should to linearize it using a log transformation (8).

$$\log y_t = \log \beta + \gamma t \tag{8}$$

The exponential growth model can be used to estimate the parameters as well as to forecast the future cases of COVID-19. For forecasting part, we can compare it with exponential smoothing (single and double) methods since they are also appropriate to forecast data that have an exponential pattern. Thus, we can compare the forecasting performance of exponential growth model and exponential smoothing methods. Therefore, the single and double exponential smoothing techniques are discussed below.

2.4. Single Exponential Smoothing

Exponential Smoothing is a method of smoothing time series data based on the exponential window function. The exponential functions are used to assign exponentially decreasing weights over time. It is method of data analysis obtained by using some optimal weight generated according to the data estimations with a given specific weight. Forecasts produced using exponential smoothing methods weighted averages of past observations. These methods give decreasing weights to past observations and thus the more recent the observation the higher the associated weight. This framework enables reliable estimates to be produced quickly in most applications. Single Exponential Smoothing method is used when the time series data has no trend and no seasonality.

$$s_t = \alpha y_t + (1-\alpha)s_{t-1} = s_{t-1} + \alpha(y_t - s_{t-1}) \qquad t > 0 \tag{9}$$

Where s_t denotes the current smoothed series obtained by applying simple exponential smoothing series Y. y_t is the current observed value of the time series in period *t*, alpha is the smoothing constant or factor ranging from 0 to 1, $1-\alpha$ estimates the moving average parameter and $t-1$ is the smoothed value at time t. In exponential smoothing technique, $s_t = y_t$. A value of α close to one have less of a smoothing effect and gives a greater weight to recent changes in the data, while values of α closer to zero have a greater smoothing effect and are less responsive to recent changes.

The h-step-head prediction is

$$\hat{Y}_{t+h} = S_t,\, h = 1,2,3,.... \tag{10}$$

The general form of single exponential smoothing forecast function is:

$$\hat{Y}_{t+h/t} = \sum_{j=1}^{T} \alpha(1-\alpha)^{t-j} Y_j + (1-\alpha)^t S_0 \quad ,0 \le \alpha \le 1 \tag{11}$$

Where h is the number of periods in the forecast lead-time and $\hat{Y}_{t+h}$ is the forecast for h periods ahead from origin t.

2.5. Double Exponential Smoothing

Single exponential smoothing cannot smooth well when there is trend in the data. Thus, we should to use a double exponential smoothing (second order exponential smoothing) method when the time series data has a trend but not seasonality component. The smoothing function for any time period t is given by:

$$S_t = \alpha Y_t + (1-\alpha)(S_{t-1} + T_{t-1}) \quad (12)$$

$$T_t = \vartheta(S_t - S_{t-1}) + (1-\vartheta)T_{t-1} \quad (13)$$

Where, equation (12) refers to the level, while the second equation is the trend. S_t Denotes an estimate of the level of the series at time *t*, T_t is the smoothed additive trend at the end of period *t*, α is the data smoothing parameter for the level of the series and ϑ is the smoothing parameter for the trend range between (0,1).

The h-step-head prediction is

$$\hat{Y}_{t+h} = S_t + hT_t,\ h = 1,2,3,... \quad (14)$$

2.6. Best fit model selection

After a model has been developed, it is important to test its fit before predicting future values. By comparing the real values to the expected values, the model's accuracy can be calculated. In this study, we used three performance measures, namely Mean Absolute Error, Mean Absolute Percentage Error, and Root Mean Square Error, to test the forecasting accuracy of a particular model. Mathematically, these measures are expressed as in equation.

$$MAE = \frac{1}{2}\sum_{t=1}^{n}|\varepsilon_t| \quad (15)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\varepsilon^2{}_t} \quad (16)$$

$$MAPE = \frac{1}{2}\sum_{t=1}^{n}\frac{\varepsilon_t}{y_t} * 100 \quad (17)$$

Where, y_t is the actual value at time t, and ε_t is the difference between the actual and the predicted values. Also, n is the number of time points. Lower MAE, RMSE, and MAPE values indicate a model that best fits the data.

3. Results and Discussion

3.1. Results

It is important to test the presence of unit roots in the variables and thus determine their order of integration before estimating models with time series variables. In order to infer a meaningful output from time series models, the variables used in the study must be stationary. **Table 1** shows that the variable is not stationary at level, but become stationary after first differing at a 1% level of significance (ADF and Phillips-Perron Test).

Table 1: Unit Root Test

Phillips-Perron Test			Augmented Dickey-Fuller Test		
Variables	Constant, with Trend		Constant, with Trend		Order of identification
	At level	First difference	At level	First difference	
COVID-19 new cases	-3.2566 (0.0752)	-40.2081 (0.000)*	-1.6827 (0.7573)	-20.7969 (0.000)*	I(1)

Notes: * denotes rejection of the null hypothesis of unit root at 1% level. Critical values at 0.01 are in parenthesis (p-value). (**Source**: author estimation-Eviews version 8)

In figure-1, the time series plot of COVID-19 new case report increases from March 8, 2021. The death graph is fairly consistent. Figure 2 shows the diagnostic plots for COVID-19 cases in Ethiopia that are predicted. ACF is a (complete) auto-correlation function that returns the auto-correlation values and lagged values of any series.

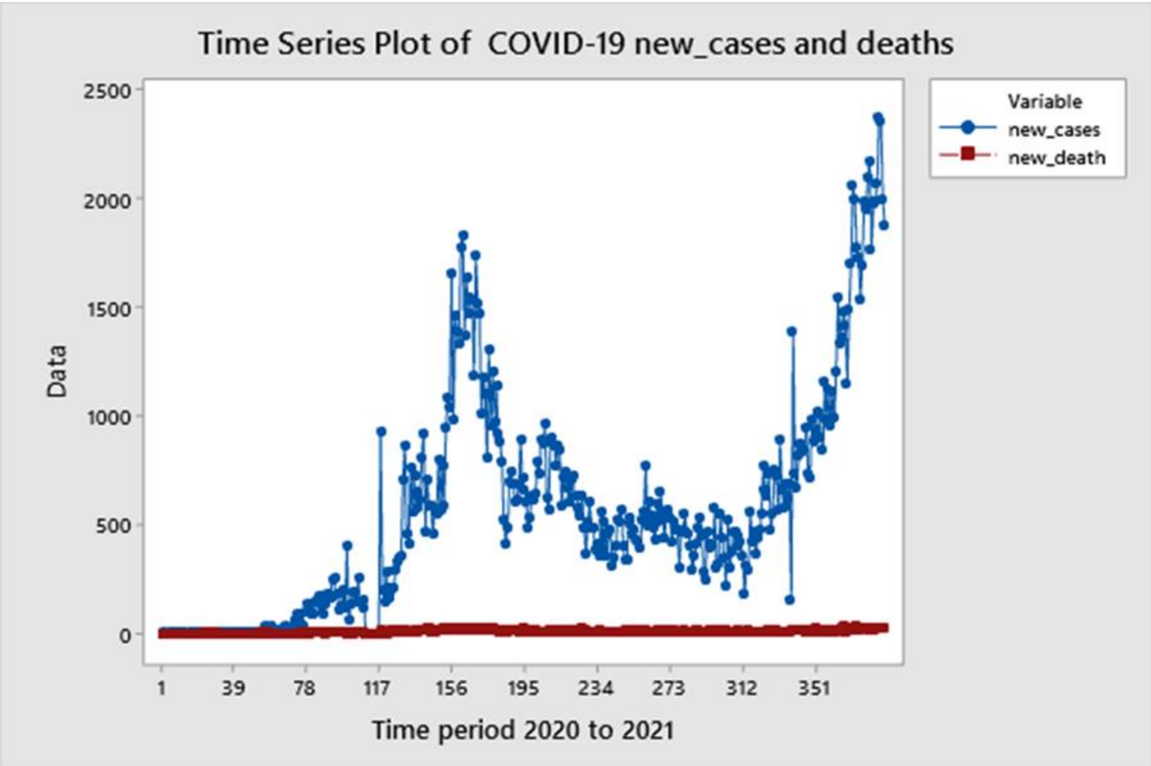

Figure 1: Time series plots for Covid-19 number of cases and deaths in Ethiopia

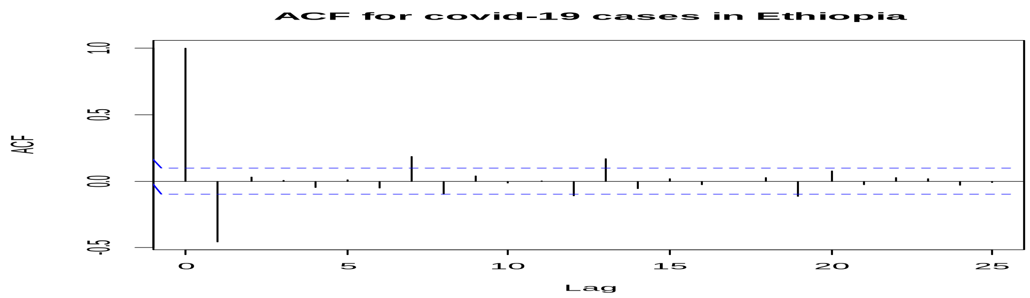

Figure 2: ACF for covid-19 new cases at first difference

In this study, the time series model encompasses to forecast COVID-19 cases in the coming 60 days. The results for the measure of model accuracy for ARIMA, Linear Trend, Quadratic Linear, and S-Curve Trend and Exponential model had displayed in **Table 2**. Look at the mean absolute percent error (MAPE), mean absolute deviation (MAD), and the mean square of deviation (MSD) values suggest that ARIMA(0,1,1) is the most accurate of all for forecasting future values as it possesses the least point for all the measures of the models.

Table 2: The measure of model accuracy of time series of COVID-19 cases, Ethiopia, 2021

Models	MSD	MAPE	MAD
Single exponential smooth	20236.3	27.6	84.2
Linear trend model	259792	742	408
Double exponential smooth	29423	102	115
S-Curve Trend Model	511924	75	491
Quadratic	193383	1067	349
ARIMA(0,1,1)	15249	21.9	45.71

The forecasting of new cases of COVID-19 in Ethiopia: shown in Table 3, with a 95% confidence interval. According to the expected result, the number of confirmed COVID-19 new cases will increase day to day for the next 60 days. This increase is evidenced by the unrestricted response that has been invested in pandemic control in the country. However, the estimated predicted values were high, requiring more effort to minimize the spread of this pandemic across the country. The key problem of the outbreak is the explanation why a few people have not shown any signs of the virus spreading the virus to others without understanding the test.

The outcome of the forecasted maximum and minimum number of new

Covid-19 cases will be 2042 and 2353 in one day. In addition, the cumulative confirmed cases reached 343,563 at the beginning of May, 2021. Basically, it is displayed in Table 3. Thus, more prevention measures and more resources will be introduced by the government; unless the coronavirus relapses and affects the country more.

Table 3: Prediction of COVID-19 new cases with their lower and upper intervals for first two months of 2021 at 95% Confidence Intervals in Ethiopia.

		95% Limits		
Period	**Forecasted new cases**	**Lower**	**Upper**	Within Range?
5/4/2021	2042	1742.68	2342.04	Yes
6/4/2021	2047	1722.85	2372.43	Yes
7/4/2021	2052	1704.83	2401.02	Yes
8/4/2021	2058	1688.27	2428.14	Yes
9/4/2021	2063	1672.93	2454.04	Yes
10/4/2021	2068	1658.62	2478.90	Yes
11/4/2021	2074	1645.21	2502.88	Yes
12/4/2021	2079	1632.58	2526.07	Yes
13/4/2021	2084	1620.64	2548.57	Yes
14/4/2021	2089	1609.32	2570.45	Yes
15/4/2021	2095	1598.56	2591.78	Yes
16/4/2021	2100	1588.29	2612.61	Yes
17/4/2021	2105	1578.48	2632.97	Yes
18/4/2021	2111	1569.10	2652.92	Yes
19/4/2021	2116	1560.10	2672.48	Yes
20/4/2021	2121	1551.45	2691.69	Yes
21/4/2021	2127	1543.14	2710.56	yes
22/4/2021	2132	1535.14	2729.12	Yes
23/4/2021	2137	1527.43	2747.39	Yes
24/4/2021	2142	1519.99	2765.40	Yes
25/4/2021	2148	1512.80	2783.14	Yes
26/4/2021	2153	1505.86	2800.65	Yes
27/4/2021	2158	1499.14	2817.93	Yes
28/4/2021	2163	1492.63	2835.00	Yes
29/4/2021	2169	1486.33	2851.86	Yes
30/4/2021	2174	1480.22	2868.53	Yes
1/5/2021	2179	1474.30	2885.02	Yes
2/5/2021	2184	1468.55	2901.33	Yes
3/5/2021	2190	1462.97	2917.47	Yes
4/5/2021	2195	1457.55	2933.45	Yes
5/5/2021	2200	1452.28	2949.28	Yes
6/5/2021	2206	1447.15	2964.97	Yes
7/5/2021	2211	1442.17	2980.51	Yes
8/5/2021	2216	1437.33	2995.92	Yes
9/5/2021	2221	1432.61	3011.20	Yes
10/5/2021	2227	1428.02	3026.35	Yes
11/5/2021	2232	1423.55	3041.38	Yes
12/5/2021	2237	1419.19	3056.30	Yes
13/5/2021	2243	1414.95	3071.10	Yes
14/5/2021	2248	1410.81	3085.80	Yes
15/5/2021	2253	1406.78	3100.39	Yes
16/5/2021	2258	1402.85	3114.88	Yes
17/5/2021	2264	1399.02	3129.27	Yes
18/5/2021	2269	1395.29	3143.57	Yes
19/5/2021	2274	1391.64	3157.77	Yes
20/5/2021	2279	1388.09	3171.89	Yes
21/5/2021	2285	1384.62	3185.92	Yes

22/5/2021	2290	1381.24	3199.86	Yes
23/5/2021	2295	1377.94	3213.73	Yes
24/5/2021	2301	1374.72	3227.51	Yes
25/5/2021	2306	1371.57	3241.22	Yes
26/5/2021	2311	1368.50	3254.85	Yes
27/5/2021	2316	1365.50	3268.41	Yes
28/5/2021	2322	1362.58	3281.89	Yes
29/5/2021	2327	1359.72	3295.31	Yes
30/5/2021	2332	1356.94	3308.66	Yes
31/5/2021	2338	1354.22	3321.94	Yes
1/6/2021	2343	1351.56	3335.16	Yes
2/6/2021	2348	1348.96	3348.31	yes
3/6/2021	2353	1346.43	3361.41	yes

Sources: author estimation, 2021

4. Conclusion and recommendation

4.1. Conclusion

We have proposed a methodology for the forecasting the spread of coronavirus disease (COVID-19) epidemic in Ethiopia, by considering publicly available data from March 13, 2020 to April 04, 2021. Novel coronavirus diseases (COVID-19) is a new virus strain spreading from person-to-person started in China and today it control all globe, including Ethiopia. In some instances, cases outside of China have been associated with travelers from China. Health experts are concerned because little is known about this new virus and it has the potential to cause severe illness and pneumonia in some people. There are no medications specifically approved for coronavirus. Most people with mild coronavirus illness will recover on their own by drinking plenty of fluids, resting, and taking pain and fever medications. However, some cases develop pneumonia and require medical care or hospitalization. According our objective of the study, the ARIMA model shows that the spread of the diseases is varies time to time. This may be due to shortage of daily tested suspected persons and unable to address the whole community in country level. But the virus was increasing fast from beginning of 2021. Knowing exact number of cases is difficult; this may be the nature of the diseases. To identify all confirmed cases the government must increase the number of laboratories in throughout the country.

4.2. Recommendation

The government and all concerned bodies should strive to curb the spread of the pandemic throughout the Ethiopia. ARIMA model is an effort to predict the future forecast of the distribution of COVID-19, based on current data, so that the institutions have to formulate policies from now on the result of the study. Increasing the number of daily test for Covid-19 and expanding the test sites throughout the country should not be delayed to recognize the real number of the cases. Moreover working on implementation of state of emergency orders should be undertaken by formulating effective strategies by concerned bodies. Besides behavioral change communication activities must be activated all over the urban and rural areas to address the negligence and unawareness among the people. Finally the hotspot regions and sub regions should be identified early and the especial handling should be considered.

What is already known?

i. The new corona virus (Covid-19) is an emerging disease known to cause severe respiratory syndrome and finally end in death

ii. It is a highly contagious disease declared as a global pandemic

What is this study adds?

Provides timely information on real situation and forecasts the spread of the pandemic of COVID-19 in Ethiopia. Indicates to where the country is going and important for taking effective measures. These data are useful as they present facts that drive analytics on COVID-19 cases in Ethiopia. Academic institutions, public health agencies, scientific communities, researchers, students, and self-explorers can use these data and models to analyze COVID-19 cases in Ethiopia. It is also useful for policy development, and decision making in other countries where data is scarce. It also represents an early reference that can be used in the future.

5. Competing interests

The authors declare that they have no competing interests.

6. Author contributions

Tesfaye Denano, considered of the study, performed the statistical analysis. The author reads and approved the final manuscript by himself.

7. Funding source

there was no fund received from any organization or person

Acknowledgment

I would like to thank Ministry of Health of Ethiopia and WHO for their daily report of COVID-19 Cases

8. Appendix A. Supplementary data

Supplementary data to this article, daily updated data from Ministry of Health Ethiopia and WHO.

References

[1]. African union, Impact of corona virus (Covid 19) on the African economy. 2020

[2]. Bacha Kebede Debela. The COVID-19 Pandemic and the Ethiopian Public Administration: Responses and Challenges. 2020

[3]. Bamidele Johnson Alegbeleye and RanaKadhim Mohammed. Challenges of healthcare delivery in the context of COVID-19 pandemic in Sub-Saharan Africa. Iberoamerican journal of medicine. 02 (2020)

100-109

[4]. BirhanuAyenew et.al. (2020) Challenges and opportunities to tackle COVID-19 spread in Ethiopia. Journal of PeerScientist 2(2): e1000014.

[5]. Brockwell PJ, Davis RA (2001). Introduction to Time Series and Forecasting (Second). Springer.

[6]. Chowell, G. 2017. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. Infect. Dis. Model. 2, 379–398.

[7]. Chung, S.S. 2010. Projecting municipal solid waste: The case of Hong Kong SAR. Resour. Conserv. Recycl. 54, 759–768. [CrossRef]

[8]. D.Benvenuto 2020. Application of the ARIMA model on the COVID 2019 epidemic dataset, data in brief https://doi.org/10.1016/j.dib.2020.105340.

[9]. Driggin E, Madhavan MV, BikdeliB , et al. Cardiovascular Considerations for Patients, Health Care Workers, and Health Systems During the Coronavirus Disease 2019 (COVID -19) Pand emic. J Am CollCardiol 2020 Mar 18.

[10]. Francesco Di Gennaro et al. Coronavirus Diseases (COVID-19) Current Status and Future Perspectives: A Narrative Review. Int. J. Environ. Res. Public Health2020, 17, 2690; doi:10.3390/ijerph17082690

[11]. He Z, Tao H (2018). Epidemiology and ARIMA model of positive-rate of influenza viruses among children in Wuhan, China: A nine-year retrospective study. International Journal of Infectious Diseases, 74, 61–70.

[12]. https://www.worldometers.info/coronavirus

[13]. Hyndman, R. J., &Athanasopoulos, 2018, Forecasting: principles and practice.

[14]. J. Fattah et al, 2018. Forecasting of demand using ARIMA model International Journal of Engineering Business Management. 10(2018).

[15]. Khashei, M.; Bijari, M. 2011. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. Appl. Soft Comput. 11, 2664–2675. [CrossRef]

[16]. Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Y. Shin,.1992."Testing the null hypothesis of stationarity against the alternative of a unit root," J. Econom., vol. 54, no. (1-3), pp. 159-178.

[17]. Nuwagira and Muzoora. Is Sub-Saharan Africa prepared for COVID-19? BMC Tropical Medicine and Health. (2020) 48:18 https://doi.org/10.1186/s41182-020-00206-x

[18]. OSHA. Guidance on Preparing Workplaces. 2020

[19]. Sun D, Li H, Lu X, Xiao H, Ren J, Zhang F et al. Clinical features of severe pediatric patients with co - ronavirus disease 2019 in Wuhan: a single center's observational study. World Journal of Pediatrics.2020; 11.

[20]. Tseng YJ, Shih YL (2019). Developing epidemic forecasting models to assist disease surveillance for influenza with electronic health records. International Journal of Computers and Applications, 0(0), 1–6. https://doi.org/10.1080/1206212X.2019.1633762

[21]. United nations economic commission for Africa. COVID-19 in Africa: Protecting Lives and Economies. Addis Ababa, Ethiopia. April 2020

[22]. Wang Y, Shen Z, Jiang Y (2018). Comparison of ARIMA and GM(1,1) models for prediction of hepatitis B in China. PLoS ONE, 13(9), 1–11. https://doi.org/10.1371/journal.pone.0201987

[23]. WHO country office for Ethiopia. Covid-19 response bulletin Ethiopia. MARCH 2020

[24]. www.cdc.gov/COVID19

[25]. www.moh.et

[26]. Xu, L.; Gao, P.; Cui, S.; Liu, C.2013. A hybrid procedure for MSW generation forecasting at multiple time scales in Xiamen City, China. Waste Manag. 33, 1324–1331. [CrossRef]

[27]. Y.-W. Cheung, K. S. Lai, 1995. Lag order and critical values of the augmented Dickey Fuller test. Journal of Business &amp Economic Statistics 13 (1995) 277–280.

[28]. Y.W. Wang, Z.Z. Shen, Y. Jiang, Comparison of ARIMA and GM(1,1) models for prediction of hepatitis B in China. PLoS One 13(2018) e0201987.

[29]. Yaffee RA, McGee M (2000). Introduction To Time Series Analysis and Forecasting with application of SAS and SPSS (1st ed.). Academic Press.

[30]. Younes, M.K.; Nopiah, Z.M.; Basri, N.E.A.; Basri, H. 2014. Medium term municipal solid waste generation prediction by autoregressive integrated moving average. Stat. Oper. Res. Int. Conf. (SORIC 2013) 427–435. [CrossRef]

[31]. Zhu N et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 2020 Feb 20; 382(8): 727-733.

[32]. Zhu N et al. A novel coronavirus from patients with pneumonia in china, 2019. N ENG J Med. 2020 https://doi.org/10.1056/NEJMoa2001017