Open Journal of Public Health

Research Article
Published: 14 Apr, 2022

Time Series Model to Forecasting the New Coronavirus Case in Ethiopia

Kindu Kebede[1], Fasiledes Fetene[2] and Hailemichael Menberu[3]*

[1]Department of Statistics, College of Computing and Informatics, Haramaya University, Ethiopia

[2]Department of Statistics, College of Computational Science, Woldiya University, Ethiopia

[3]Department of Statistics, College of Computational Science, Wolayita Sodo University, Ethiopia

Abstract

Background: Coronaviruses are a large family of respiratory viruses that can cause illness in people. The goal of this study is find an appropriate model and forecast the new cases in Ethiopia.

Methods: ARIMA models are the most general class of models for forecasting a time series data. The ARIMA forecasting equation for a stationary time series is a linear equation in which the predictors consist of lags of the dependent variable or lags of the forecast errors.

Results: In Ethiopia the daily average increment in COVID-19 from March 14th, 2020 to September 13th, 2020 was 347. The maximum value of new cases recorded in Ethiopia was 1829 on August 22nd, 2020. The series is not stationary at level, but the series is stationary at first difference. The selected model in this study was ARIMA (1,1,2) that has small AIC and BIC.

Conclusion: This study focused on find an appropriate model and forecast the new cases in Ethiopia. Based on ARIMA (1,1,2) model the new cases in Ethiopia was 808 people with CI [482,1134] and 532 people with CI [-281,1344] from September 14th, 2020 and September 30th, 2020.

Keywords: Coronavirus; Forecasting; Autoregressive integrative moving average; Stationary; Ethiopia

Introduction

Coronaviruses are a large family of respiratory viruses that can cause illness in people and animals. In rare cases, coronaviruses that circulate among animals can evolve and infect humans. In turn, these infections can easily spread from person-to-person as was the case with Severe Acute Respiratory Syndrome (SARS-CoV) [1,2]. The first novel coronavirus was reported on January 13th, 2020, by the Ministry of Public Health (MoPH) of Thailand. The case was on a 61-year-old Chinese woman living in Wuhan City, Hubei Province, China and WHO declares the virus as pandemic on March 11th, 2020 [3].

No disease in history has seen such rapid development in research. The world's ambition to develop these tools as fast as possible must be matched by its ambition to ensure as many people as possible have access to them. Most people infected with the COVID-19 virus will experience mild to moderate respiratory illness and recover without requiring special treatment. Older people and those with underlying medical problems like cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to develop serious illness [4,5].

The best way to prevent and slow down transmission is be well informed about the COVID-19 virus, the disease it causes and how it spreads. Protect yourself and others from infection by washing your hands or using an alcohol based rub frequently and not touching your face [6].

Africa's first COVID-19 case was recorded in Egypt on February 14th, 2020. Most countries like Ethiopia, Ghana, Kenya, Egypt, Morocco, Tunisia and Nigeria have expanded national testing to multiple labs for testing this pandemic disease. These rapid measures have the opportunity for testing and quarantining those people with positive of the disease [7-9].

In Ethiopia the first case reported was on March 14th, 2020 in Gregorian calendar and March 4th, 2012 in Ethiopian calendar who was citizens of Japan. More than one million laboratory tests have been performed for COVID-19 in Ethiopia so far. The number of positive cases has been increased tremendously with an increasing trend.

OPEN ACCESS

Correspondence:
Kindu Kebede, Department of Statistics, College of Computing and Informatics, Haramaya University, Dire Dawa, Ethiopia,
E-mail: m7.kebede@gmail.com

Received Date: *23 Mar 2022*
Accepted Date: *08 Apr 2022*
Published Date: *14 Apr 2022*

Citation:
Kebede K, Fetene F, Menberu H. Time Series Model to Forecasting the New Coronavirus Case in Ethiopia. Open J Public Health. 2022; 4(2): 1033.

Data and Methods

Source of data

The data for this research was secondary data taken from WHO world COVID dataset and took only Ethiopia data from March 14th, 2020 to September 13th, 2020.

Methods

Time series analysis is a statistical technique that deals with time series data, or trend analysis. Time series data means that data is in a series of particular time periods or intervals. The purpose of time series analysis is generally two fold to understand or model the stochastic mechanism that gives rise to an observed series and to predict or forecast the future values of a series based on the history of that series and, possibly, other related series or factors [10].

Stationary

A stationary process has the property that the first and second moments do not change over time. It is an essential property to define a time series process. Time series, X_t is a finite variance process such that (i) the mean value function, μ_t is constant and does not depend on time t, and the auto covariance function, $\gamma(s, t)$ depends on *s* and *t* only through their difference |*s*-*t*|.

Model specification

The word ARIMA stands for Auto-Regressive Integrated Moving Average. ARIMA models are the most general class of models for forecasting a time series data y_t [11]. The ARIMA forecasting equation for a stationary time series is a linear equation in which the predictors consist of lags of the dependent variable and/or lags of the forecast errors. The general ARIMA (*p,d,q*) model can be written:

$$\varphi(B)(1-B)^d X_t = \theta(B)\varepsilon_t$$

Where,

$$BX_t = X_{t-1}\ ,\ \varphi(B) = (1-\alpha_1 B - \alpha_2 B^2 \ldots \ldots .. \alpha_p B^p)$$

$$\theta(B) = (1+\beta_1 B + \beta_2 B^2\ \ldots \ldots \ldots .\ + \beta_q B^q)$$

Results

Descriptive statistics

From Table 1 in Ethiopia the daily average increment in COVID-19 from March 14th, 2020 to September 13th, 2020 were 347. The maximum value of new cases recorded was 1829 on August 22nd, 2020.

Inferential statistics

Unit root test: The basic assumption of time series data is stationary meaning the mean and variance of the system should be constant through time. This assumption can be checked by either graphically or formally using tests.

From Figure 1 above, it seems that the mean has an increasing upward trend until August 18th, 2020 and then the trend decreases down ward. This implies non stationary of the series. There are also formal tests of stationary. In this paper the ADF tests is used. The choice (with or without trend and intercept) to choose depends on trending behavior of the data. From the first graph the data had an increasing trend with time. So, the choice for testing unit roots include trend.

In the test of stationary the null hypothesis states that the series has unit root means it is not stationary so from result shows the series is not stationary at level because the p-value equal 0.9401 and 0.7486 for intercept and both intercept and trend respectively (p>0.05). But we can reject the null hypothesis implies the series is stationary at first difference. Therefore it is integrated of order one.

Table 1: Descriptive statistics.

	New Cases
Mean	347.2
Median	112.5
Maximum	1829.000
Minimum	0.000000
Std. Dev.	487.5741
Observations	184

Table 2: Parameter estimation.

Variable	Coefficient	Std. Error	t-Statistic	95% CI		Prob.
C	2.8597	8.846	0.3232	-6.753	20.8465	0.7469
AR(1)	0.7338	0.102724	7.1437	0.5325	0.9352	0
MA(1)	-1.5148	0.067273	-22.2929	-1.7232	-1.3064	0
MA(2)	0.6905	0.052029	13.66229	0.5235	0.8576	0

Table 3: Forecasting the new case in Ethiopia.

Date	Point forecast	Lo 95% CI	Hi 95% CI
14/09/2020	808.0217	482.2035	1133.84
15/09/2020	733.9242	400.3814	1067.467
16/09/2020	679.549	328.4535	1030.644
17/09/2020	639.6467	262.5149	1016.778
18/09/2020	610.3651	201.3584	1019.372
19/09/2020	588.8773	144.6258	1033.129
20/09/2020	573.1089	92.06514	1054.153
21/09/2020	561.5375	43.35099	1079.724
22/09/2020	553.046	-1.90349	1107.995
23/09/2020	546.8147	-44.1011	1137.73
24/09/2020	542.2419	-83.6258	1168.11
25/09/2020	538.8863	-120.825	1198.597
26/09/2020	536.4238	-156.002	1228.849
27/09/2020	534.6168	-189.416	1258.65
28/09/2020	533.2907	-221.288	1287.87
29/09/2020	532.3176	-251.804	1316.44
30/09/2020	531.6035	-281.12	1344.327

Model selection

In order to select the appropriate ARIMA model first visualizing the autocorrelations and partial autocorrelation is much more useful to determine how many AR and MA are needed, so the ACF function plots them for each value of *k*. The figures below show the autocorrelation functions and partial autocorrelation for the differenced series. Values above the line are significant. (The height of the dashed line is determined by the amount of data.).

The value of a PACF helps to identify the number of Auto Regression (AR) coefficients and ACF helps to identify Moving Average (MA) coefficients an ARIMA model.

In the original data the autocorrelations are significant for a large

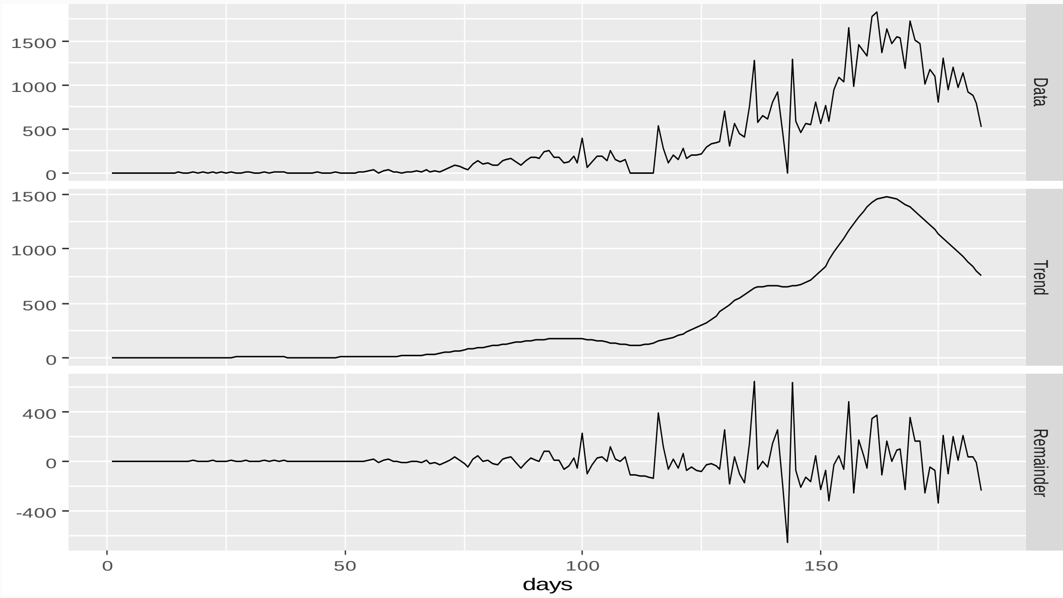

Figure 1: Time series plot of original series of COVID-19 in Ethiopia.

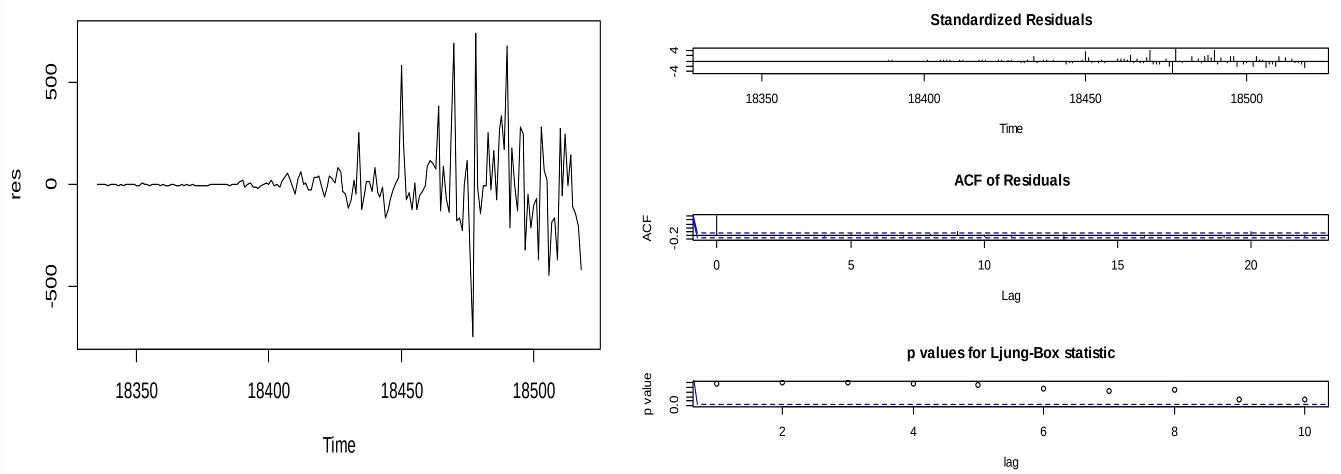

Figure 2: Plot of residuals.

number of lags but perhaps the autocorrelations at lags 2 and above are may be due to the propagation of the autocorrelation at lag 2 this is confirmed by the PACF plot. From the differenced series of the two plot the first candidate model is ARIMA (2,1,1).

There are also other models that can be compared with the candidate model. In E views automatic forecast of ARIMA model select the best model by using minimum value of LogL, AIC, BIC and HQ equal to -1194.343629, 13.107581, 13.195272, and 13.143126 respectively. The model selected by this procedure is ARIMA (1,1,2).

Parameter estimation

From Table 2 the all the parameters are significant except the intercept term (p-value <0.05).

$$(1-\varphi_1 B)(1-B)y_t = C + (1+\theta_1 B + \theta_2 B^2)\varepsilon_t$$

Where φ_1=0.6383, θ_1= -1.4997 and θ_2=0.7108 therefore the model can be written as,

$$\Delta \hat{y}_t = 0.7338\Delta \hat{y}_{t-1} - 1.4997\varepsilon_{t-1} + 0.6905\varepsilon_{t-2}$$

Model diagnostic

After a tentative model has been fit to the data, we must examine its adequacy and, if necessary, suggest potential improvements. This is done through mainly by analyzing the residuals. From Figure 2 plot of the residuals the mean of residuals look randomly distributed around zero and exhibits no clear pattern.

The “residuals” in a time series model are what is left over after fitting a model. The residuals defined as the original value minus fitted value. Residuals are useful in checking whether a model has adequately captured the information in the data. In the above model the residuals should not be correlated between residuals and the residuals have zero mean.

$$\varepsilon_t = y_t - \hat{y}_{t-1}$$

From the above plot of residuals ACF plot of the residuals do not show any significant autocorrelation between the residuals, the p-values for Ljung-Box plot shows that all are large implies no sign of patterns of the residuals. In addition to looking at the ACF plot, we can use formal test for autocorrelation by considering a whole set of values as a group, rather than individual.

Since p-value equal to 0.07454 is large we can conclude that there is no autocorrelation between residuals. Therefore, the selected model is an appropriate model. Forecasting using ARIMA (1,1,2).

The basic idea behind self-projecting time series forecasting models is to find a mathematical formula that will approximately generate the historical patterns in a time series. The goal of this research is to find an appropriate formula so that the residuals are as small as possible and exhibit no pattern (Table 3).

Conclusion

In Ethiopia the daily average increment in COVID-19 from March 14th, 2020 to September 13th, 2020 was 347. The maximum value of new cases recorded in Ethiopia was 1829 on August 22nd, 2020. The series is not stationary at level, but the series is stationary at first difference. The selected model in this study was ARIMA (1,1,2) that has small AIC and BIC. This study focused on find an appropriate model and forecast the new cases in Ethiopia. Based on ARIMA (1,1,2) model the new cases in Ethiopia was 808 people with CI [482,1134] and 532 people with CI [-281,1344] from September 14th, 2020 and September 30th 2020.

Acknowledgment

World Health Organization is gratefully acknowledged for the data they supplied for this research.

References

1. Zu ZY, Jiang MD, Xu PP, Chen W, Ni QQ, Lu GM, et al. Coronavirus disease 2019 (COVID-19): A perspective from China. Radiology. 2020:296(2):200490.
2. Bai Y, Yao L, Wei T, Tian F, Jin DY, Chen L, et al. Presumed asymptomatic carrier transmission of COVID-19. JAMA. 2020;323(14):1406-7.
3. WHO. Coronavirus disease (COVID-19): Weekly epidemiological update. 2020.
4. Gandhi RT, Lynch JB, Rio CD. Mild or moderate COVID-19. N Engl J Med. 2020;383(18):1757-66.
5. Livingston E, Bucher K. Coronavirus disease 2019 (COVID-19) in Italy. JAMA. 2020;323(14):1335.
6. Rajgor DD, Lee MH, Archuleta S, Bagdasarian N, Quek SC. The many estimates of the COVID-19 case fatality rate. Lancet Infect Dis. 2020;20(7):776-7.
7. Ceylan Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. Sci Total Environ. 2020;729:138817.
8. Clerkin KJ, Fried JA, Rajkhelkar J, Sayer G, Griffin JM, Masoumi A, et al. COVID-19 and cardiovascular disease. Circulation. 2020;141(20):1648-55.
9. Yancy CW. COVID-19 and African Americans. JAMA. 2020;323(19):1891-2.
10. Singh RK, Rani M, Bhagavathula AS, Sah R, Rodriguez-Morales AJ, Kalita H, et al. Prediction of the COVID-19 pandemic for the top 15 affected countries: Advanced Autoregressive Integrated Moving Average (ARIMA) model. JMIR Public Health Surveill. 2020;6(2):e19115.
11. Alzahrani SI, Aljamaan IA, Al-Fakih EA. Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. J Infect Public Health. 2020;13(7): 914-9.