International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2020; 5(4): 249-257
© 2020 Stats & Maths
www.mathsjournal.com
Received: 18-05-2020
Accepted: 24-06-2020

Shem Otoi Sam
Research Statistician, University of Nairobi, School of Mathematics, Edwardina Otieno Ndhine, National Commission of Science Technology and Innovation, Nairobi, Kenya

Edwardina Otieno Ndhine
National Commission of Science Technology and Innovation, Nairobi, Kenya

Corresponding Author:
Shem Otoi Sam
Research Statistician, University of Nairobi, School of Mathematics, Edwardina Otieno Ndhine, National Commission of Science Technology and Innovation, Nairobi, Kenya

Prediction and forecasting covid-19 cases, fatalities, and morbidity in Kenya

Shem Otoi Sam and Edwardina Otieno Ndhine

Abstract
In this paper we are looking for the best model to predict COVID-19 cases, fatalities, and active cases. Time series analysis using auto-regressive integrated moving averages is used. The three series are tested for stationarity using Augmented Dickey-Fuller test (ADF) and differenced to obtain stationarity. Both the data and models formulated are tested for autocorrelation using ACF and PACF. In selecting the best ARIMA model Akaike Information Criteria (AIC) that gives the least value is picked. According to AIC selection the best model for cases, fatalities, and active/infected are ARIMA (2, 2, 2), (1, 2, 3), and (0, 2, 1) respectively. It implies that two lags of previous cases have influence on current case occurrence as opposed to one lag of fatalities. Asymptomatic and infected people with clear syndromes, not in quarantine, may move and interact with others leading to more infections, compared to deceased. The residuals of estimated case fatalities and cases models are not autocorrelated as seen from the ACF and PACF. Also, the forecast of case fatality model show that fatalities will stabilize on 9/8/2020 and begin to fall from 22/8/2020.The cases model forecasts that COVID-19 cases stabilize after 17/8/2020 and begin to fall from 23/8/2020, both dates will have 35,779 and 41,517 COVID-19 total cases respectively. From further analysis, the forecast of active cases over time gives additional information; its lower limit forecast begins to fall after 10/8/2020 showing 13669 infected persons and reduces to 10345 on 4/9/2020. Consequently, 10/8/2020 is a possible beginning of Kenyan peak. There is also a statistical possibility that the peaks of cases, fatalities, and infected persons occur at different intervals of time or rotating seasonal peaks. The progress of active cases over time carries the “energies” or “momentum” of COVID-19 pandemic.

Keywords: Kenya COVID-19 forecasts, multiple peaks

Introduction
The acute respiratory disease, COVID-19, was reported in Wuhan China in December 2019 Alangreh *et al.* (2020). Following its rapid spread around the world, the World Health Organization (WHO) declared COVID-19 epidemics as a public health emergency of international concern on 30 January 2020 (Guo *et al.* 2020) [6].

Since declaration of COVID-19 as a global pandemic and subsequent spread to all countries, governments work to predict future dynamics of the infectious disease. In Iran, Dahesh *et al*, (2020) used ARIMA model to forecast confirmed cases in different countries. Shem (2020) [4] scrutinized statistical significance of regional COVID-19 data and concluded that infections among regions were not significantly different and came from the same population. Earlier in March, Tartai and Varalyay (2020) [3] used logistic model to predict future occurrence in China and test the reliability of the model estimations. Foresting is helpful for governments to plan strategies of overcoming impacts of the pandemic on economies Ozili and Arun (2020) [2] and design interventions of coping up with various aspects of mass destruction and economic disruptions.

In this paper, ARIMA model is used to forecast the future patterns of occurrences in Kenya using data from confirmed cases, fatalities, and active cases announced daily by the Ministry of Health covering March 2020 to August 2020. In that regard, total cases, fatalities, and active cases are forecasted to understand the patterns underlying current and future COVID-19 data. The ARIMA model is more robust (AR) one variable predictions. The best model is identified followed by 30 days future predictions. The objective of this paper is to build the best predictive models for cases, fatalities, and active cases to inform planning, budget, preparedness, response and surveillance at the national and county government levels. Lastly, monthly predictions may give pointers to possible periods of infection, and fatality peaks.

Methods

Source of Data

The data in this paper is obtained from the Ministry of Health daily COVID-19 status update covering the period from March to August /2020. The data is then triangulated with what is available on world meter (https://www.worldometers.info/coronavirus/country/kenya/). These are 134 days of COVID-19 infections in the country.

Arima Model

Auto-regressive integrated moving average (ARIMA) is a time series model with (AR), (I) integrated and (MA) moving average components ordered in (P, D, Q) as parameters respectively. First, logarithmic transformation is done to stabilize means and variances of the three series: total cases, total fatalities, and active cases. Examining the time series data for trend, seasonality, cyclic behaviour, and irregularity. And, using Augmented Dickey-Fuller test (ADF) to check stationarity of mean, variance, and covariance to ensure they are time invariant. The moving average and centered moving average were then calculated. Also, model selection was done and an AIC model with the least value chosen. The ARIMA model is estimated and the validity of the model tested. Also, both series and models are tested for autocorrelation and partial autocorrelation using ACF and PACF, respectively. The forecasts for total cases, total case fatality, and active/infected were estimated and followed by validation of the model using Ljung-Box test.

Empirical Review

ARIMA model consists of lags of dependent variable (y_t) which is the (AR) component and or lags from forecast error ($\in_t$) which is the (MA) part. The (I) integrated part describes how many times the series must be differenced to be stationary. Thus, (AR), (I), and (MA) are denoted by P, D, and Q, respectively. The model has the quality of filtering signals from the stochastic errors and backshift in time to n^{th} observation.

The functional form of ARIMA is

$y_t = \pi + \omega_1 y_{t-1} + \cdots + \omega_p y_{t-p} - \varphi_1 \in_{t-1} - \cdots \varphi_q \epsilon_{t-q} \ldots\ldots\ldots\ldots\ldots\ldots\ldots..$ (1) where,
y_t, is the dependent variable
$y_{t-i}, 1 = 1, \ldots, p$ lags of dependent variable.
$\omega_i, i = 1, \ldots, p$ coefficients of lags of dependent variables.
$\epsilon_j, j = 1, \ldots, q$ lags of forecast error

In our method

For cases ARIMA (2, 2, 2), we have two lags of dependent variable, which is differenced twice, and two lags of error. Hence $P = 2, D = 2, and\ Q = 2$.
Solving for parameter D:
$d = 0: y_t = Y_t$
$d = 1: y_t = y_t - Y_t$……………………..(2)
$d = 2: (y_{t-1} - Y_{t-1}) - (y_{t-1} - Y_{t-2})$
$y_t = y_t - Y_{t-1} - Y_{t-1} + Y_{t-2}$
$d: y_t = Y_t - 2Y_{t-1} + Y_{t-2}$………………….(3)
$\pi = 0$ because we do not have a constant, the cases start at 0 on 13/3/2020

The estimated equation is

$y_t = \omega_1 y_{t-1} + \omega_2 y_{t-2} - \varphi_1 \epsilon_{t-1} - \varphi_2 \epsilon_{t-2}$………………………………………… (4)

Considering the signs after calculations,

$\hat{y}_t = \widehat{\omega}_1 \hat{y}_{t-1} + \omega_2 \hat{y}_{t-2} - \varphi_1 \widehat{\in}_{t-1} - \varphi_2 \widehat{\in}_{t-2}$ ………………………………………(5)

For case fatality, ARIMA (1, 2, 3) representing (P, D, Q) respectively.

$y_t = \omega_1 y_{t-1} + \varphi_1 \epsilon_{t-1} - \varphi_2 \epsilon_{t-2} - \varphi_3 \epsilon_{t-3}$……………………………………….. (6)
Considering signs, $y_t = \omega_1 y_{t-1} - \varphi_1 \epsilon_{t-1} - \varphi_2 \epsilon_{t-2} - \varphi_3 \epsilon_{t-3}$……………………..(7)

Assumption

The model works under assumption that data used is stationarity, failure of which data is differenced until the condition is realized.

Results

From stationarity ADF test revealed that the COIVID-19 cases conform to ARIMA (2,2,2) while case fatality data are of ARIMA (2,2,1). The time plots below illustrate the variables before and after differencing to remove trend. The variables movement over time are neither cyclic nor seasonal.

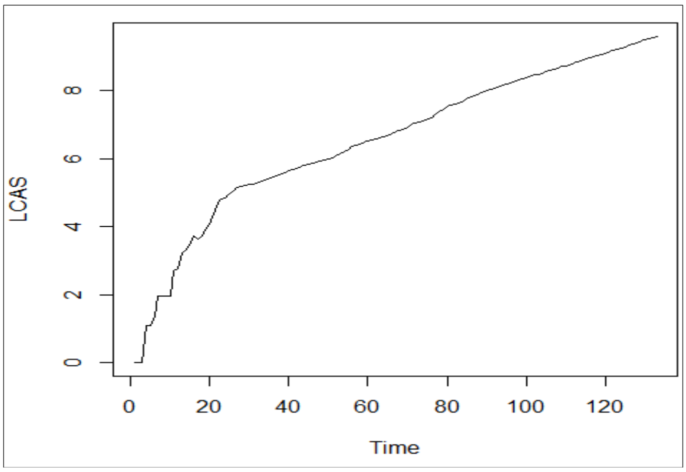

Fig 1: Plot of cases before differencing

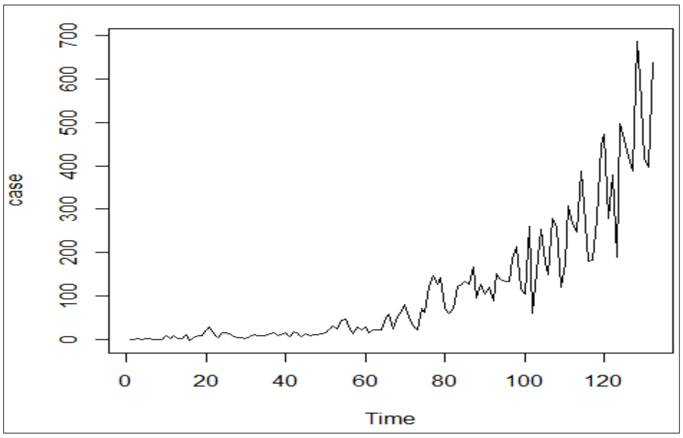

Fig 2: Plot of cases after the first difference

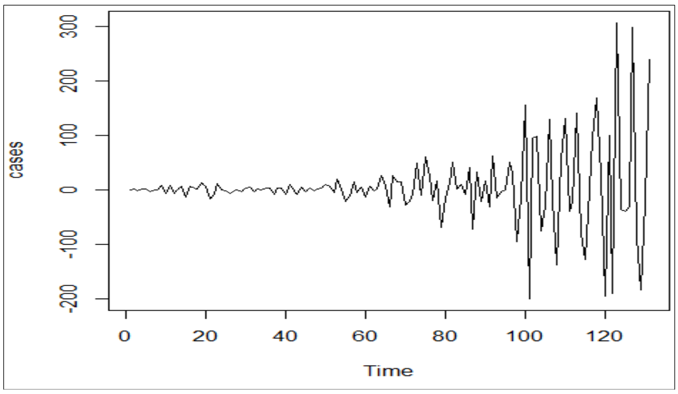

Fig 3: Plot of cases after the second differencing to detrend the series.

Plot of log of Fatality

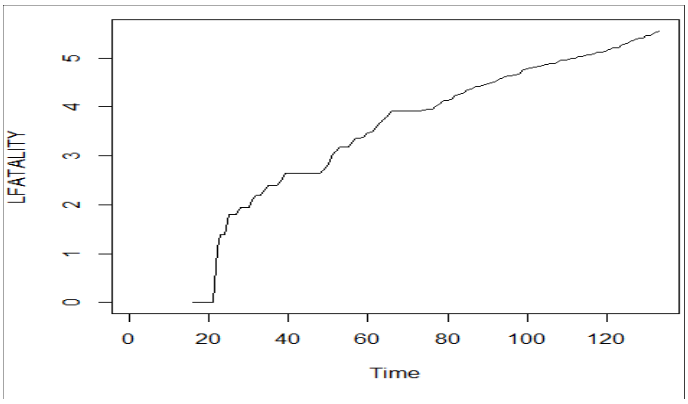

Fig 3: Plot of fatality without detrending.

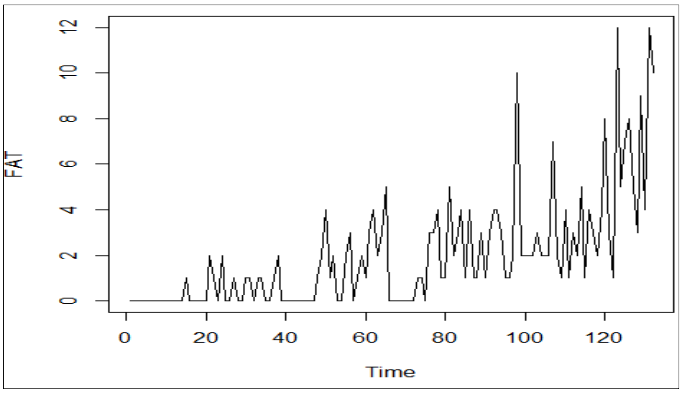

Fig 4: Plot of fatality after first difference, there is still time trend.

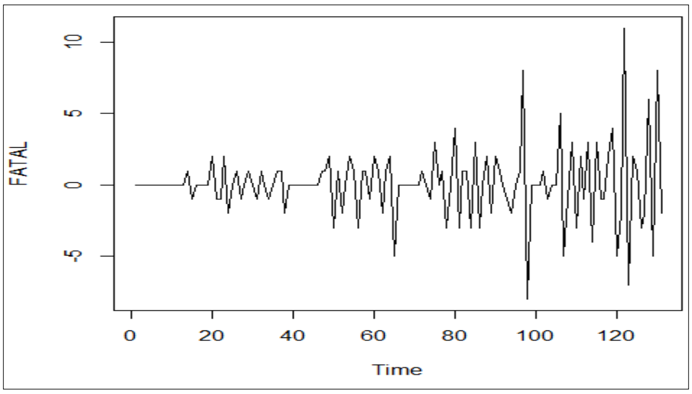

Fig 5: Plot of fatality after second differencing to remove trend

There is element of seasonality or cyclic. However, there is irregularity in CASES and FATALITY.
Testing autocorrelation of both cases and fatality data was done using ACF and PACF and results presented in *figure 6* and *figure 7* below. There is not autocorrelation.

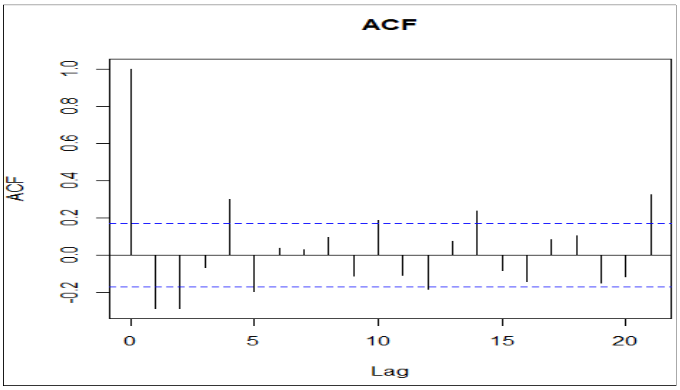

Fig 6: ACF for cases data.

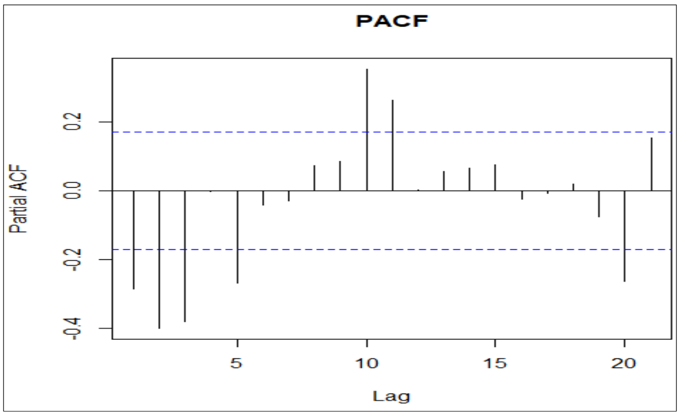

Fig 7: PACF of COVID-19 case data.

ACF and PACF autocorrelation test for case fatality illustrated in *Figure 9* and *Figure 9* which show no autocorrelation.

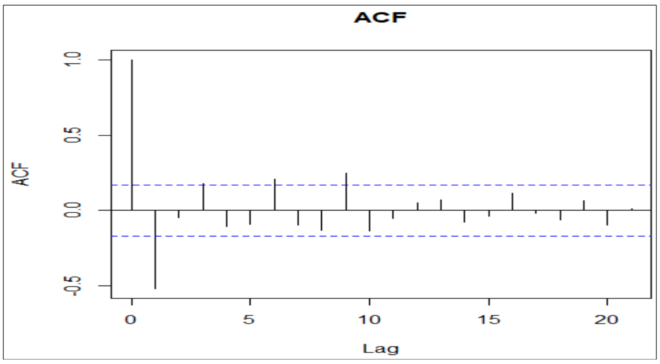

Fig 8: ACF of case fatality

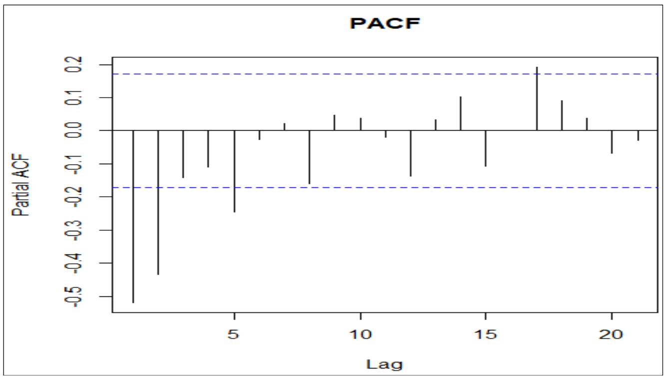

Fig 9: PACF of case fatalities

Choosing Parameters

The parameter for cases is ARIMA (2,2,2). The parameter selection found (P, D, Q) to be (2, 2, 2) respectively. It is decided based on least value of AIC criterion which provides the best model.

Model Estimation

The cases model estimated with two lags of (AR) and (MA) component,

From (5), $\hat{y}_t = \hat{\omega}_1\hat{y}_{t-1} + \omega_2\hat{y}_{t-2} - \varphi_1 \hat{\epsilon}_{t-1} - \varphi_2 \hat{\epsilon}_{t-2}$

Considering signs of coefficients in the output.

$\hat{y}_t = 0.9667_{(0.0906)}y_{t-1} - 0.4437_{(0.0898)}y_{t-2} + 1.7170_{(0.0414)}\epsilon_{t-1} - 0.9271_{(0.0427)}\epsilon_{t-2}$……..(8)

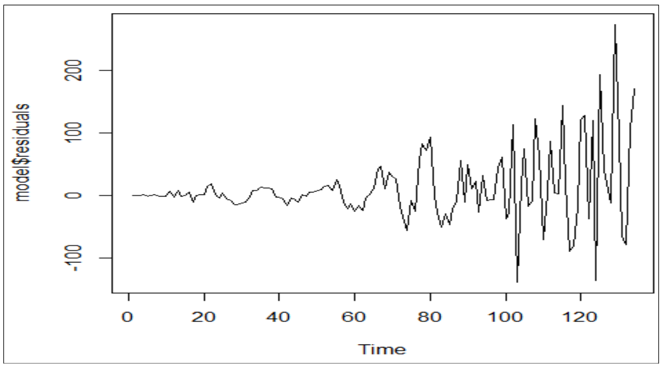

Fig 10: Plot of residuals of fitted ARIMA model for cases.

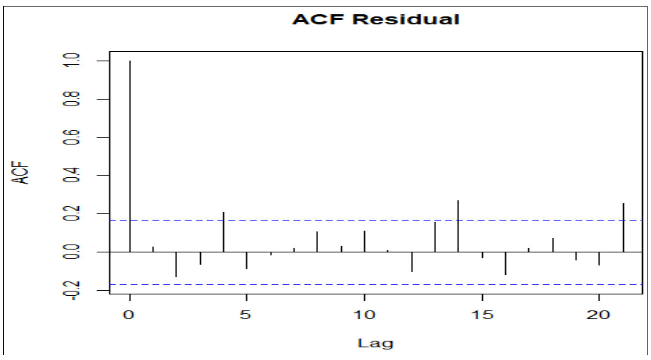

Fig 11: ACF of fitted ARIMA model residuals for cases

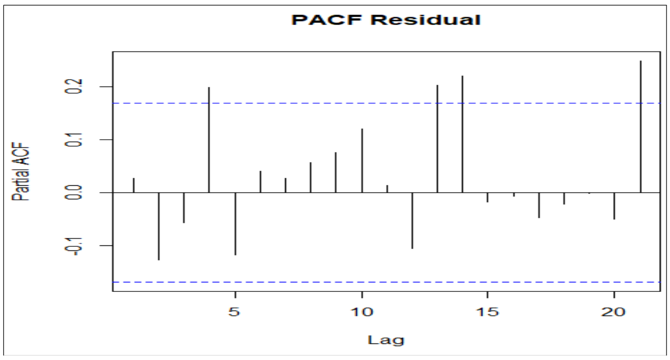

Fig 12: PAFC of the fitted ARIMA model for cases

Forecast for Cases

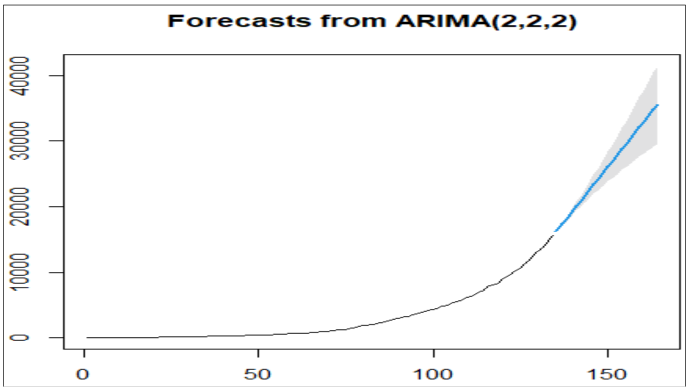

Fig 13: Plot of 30 days forecast for COVID-19 cases.

Table 1: Forecast of COVID-19 cases from 25/7/2020 to 23/8/2020

Data	Forecast	Lowest	Highest
25/7/2020	16248.82	16102.98	16394.65
26/7/2020	16841.16	16607.74	17074.57
27/7/2020	17445.62	17148.78	17742.45
28/7/2020	18086.42	17729.08	18443.76
29/7/2020	18756.97	18324.19	19189.75
30/7/2020	19440.15	18906.90	19973.41
31/7/2020	20122.35	19463.28	20781.42
1/8/2020	20797.99	19993.54	21602.45
2/8/2020	21467.73	20504.49	22430.97
3/8/2020	22134.67	21003.07	23266.26
4/8/2020	22801.52	21493.52	24109.52
5/8/2020	23469.53	21977.27	24961.80
6/8/2020	24138.71	22454.14	25823.27
7/8/2020	24808.48	22923.51	26693.45
8/8/2020	25478.33	23385.04	27571.63
9/8/2020	26147.98	23838.77	28457.20
10/8/2020	26817.41	24285.04	29349.77
11/8/2020	27486.70	24724.28	30249.12
12/8/2020	28155.97	25156.84	31155.11
13/8/2020	28825.28	25582.96	32067.60
14/8/2020	29494.63	26002.82	32986.43
15/8/2020	30164.00	26416.55	33911.45
16/8/2020	30833.38	26824.26	34842.50
17/8/2020	31502.76	27226.10	35779.42
18/8/2020	32172.13	27622.18	36722.07
19/8/2020	32841.49	28012.63	37670.35
20/8/2020	33510.85	28397.57	38624.12
21/8/2020	34180.21	28777.11	39583.31
22/8/2020	34849.57	29151.35	40547.79
23/8/2020	35518.94	29520.39	41517.49

Suitable fatality model and forecast

Case Fatality parameters were found to be ARIMA (1, 2, 3) for (P, D, Q) respectively. They were equally selected based on least AIC value depicting the best model.

Case fatality model estimation

The estimated case fatality ARIMA model is,

$$y_t = \omega_1 y_{t-1} - \varphi_1 \epsilon_{t-1} \varphi_2 \epsilon_{t-2} - \varphi_3 \epsilon_{t-3}$$

$$\hat{y}_t = -0.7684_{(0.117)} y_{t-2} + 0.1389_{(0.1263)} \in_1 + 0.7099_{(0.0794)} \epsilon_2 - 0.2502_{(0.0859)} \in_3 \text{ ……… (9)}$$

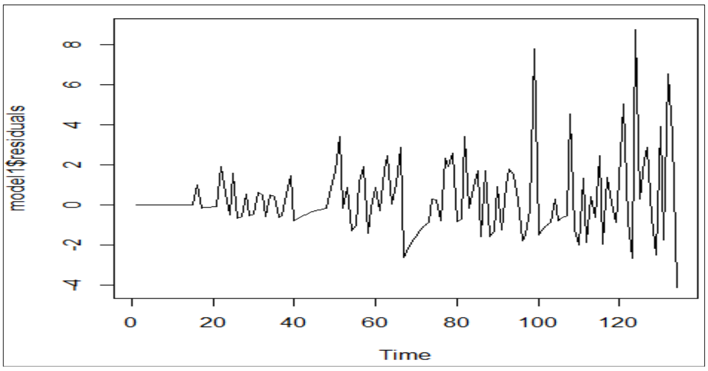

Fig 14: Plot of residuals of fitted case fatality fitted ARIMA model.

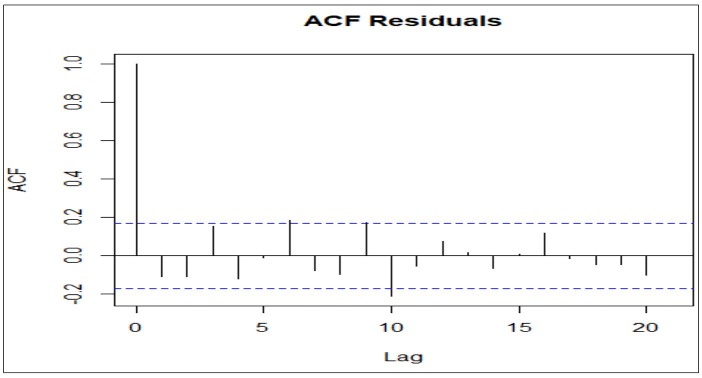

Fig 15: ACF of residuals of case fatality fitted ARIMA model

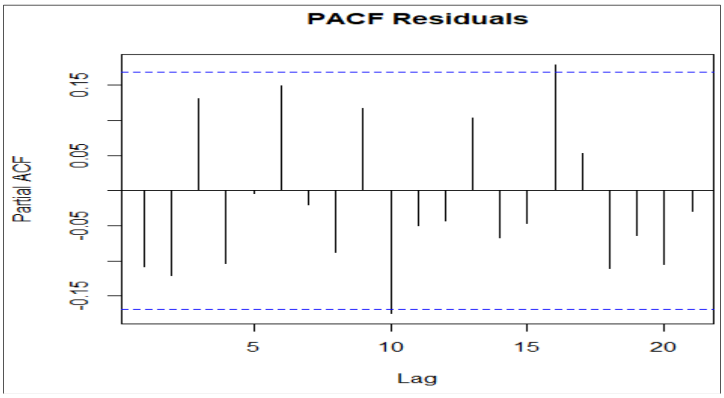

Fig 16: PACF for fitted case fatality ARIMA model

Forecast of Case Fatality

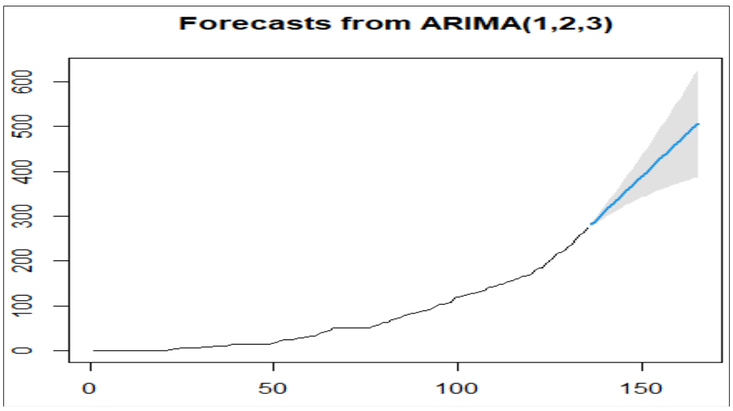

Fig 17: Plot of 30 days forecast for case fatality ARIMA model

Table 2: Forecast of COVID-19 case fatalities from 25/7/2020 to 23/8/2020

Date	Forecast	Lowest	Highest
25/7/2020	281.0975	276.2535	285.9414
26/7/2020	288.7979	281.6231	295.9728
27/7/2020	296.6420	287.4908	305.7932
28/7/2020	304.3757	292.6526	316.0987
29/7/2020	312.1942	297.9981	326.3902
30/7/2020	319.9475	302.9055	336.9895
31/7/2020	327.7509	307.8484	347.6534
1/8/2020	335.5159	312.4945	358.5372
2/8/2020	343.3103	317.1031	369.5176
3/8/2020	351.0821	321.4936	380.6707
4/8/2020	358.8714	325.8108	391.9320
5/8/2020	366.6472	329.9555	403.3388
6/8/2020	374.4333	334.0103	414.8564
7/8/2020	382.2115	337.9204	426.5026
8/8/2020	389.9958	341.7337	438.2580
9/8/2020	397.7754	345.4201	450.1308
10/8/2020	405.5587	349.0078	462.1096
11/8/2020	413.3391	352.4806	474.1976
12/8/2020	421.1217	355.8555	486.3879
13/8/2020	428.9026	359.1240	498.6813
14/8/2020	436.6848	362.2965	511.0731
15/8/2020	444.4661	365.3691	523.5630
16/8/2020	452.2480	368.3483	536.1478
17/8/2020	460.0294	371.2325	548.8264
18/8/2020	467.8113	374.0260	561.5965
19/8/2020	475.5928	376.7286	574.4569
20/8/2020	483.3745	379.3434	587.4057
21/8/2020	491.1561	381.8706	600.4416
22/8/2020	498.9378	384.3126	613.5631
23/9/2020	506.7194	386.6700	626.7688

Active Cases

From the analysis active cases follow ARIMA (0, 2, 1)

$y_t = \omega_1 y_{t-1} + \varphi_1 \epsilon_t + \varphi_2 \epsilon_{t-1}$

Where, $\omega_1 \; and \; \varphi_1 = 0$

$y_t = \varphi_2 \epsilon_{t-1}$

The estimated model, $\hat{y}_t = 0.5506 \in_{t-1}$

Table 3: Forecast for infected(active) persons in Kenya from 6/8/2020 to 6/9/2020

Date	Forecast	Low	High
6/8/2020	13712	13490	13935
7/8/2020	13950.84	13559.68	14342.01
8/8/2020	14188.77	13613.50	14764.03
9/8/2020	14426.69	13650.13	15203.25
10/8/2020	*14664.61*	*13669.98*	*15659.24*
11/8/2020	14902.53	13673.88	16131.18
12/8/2020	15140.45	13662.66	16618.24
13/8/2020	15378.37	13637.09	17119.66
14/8/2020	15616.30	13597.85	17634.75
15/8/2020	15854.22	13545.54	18162.90
16/8/2020	16092.14	13480.70	18703.58
17/8/2020	16330.06	13403.83	19256.30
18/8/2020	16567.98	13315.34	19820.63
19/8/2020	16805.91	13215.63	20396.18
20/8/2020	17043.83	13105.06	20982.60
21/8/2020	17281.75	12983.94	21579.56
22/8/2020	17519.67	12852.58	22186.76
23/8/2020	17757.59	12711.25	22803.94
24/8/2020	17995.51	12560.20	23430.83
25/8/2020	18233.44	12399.67	24067.20
26/8/2020	18471.36	12229.88	24712.84
27/8/2020	18709.28	12051.03	25367.53
28/8/2020	18947.20	11863.31	26031.09
29/8/2020	19185.12	11666.91	26703.34
30/8/2020	19423.05	11461.99	27384.10
31/8/2020	19660.97	11248.72	28073.22

1/9/2020	19898.89	11027.23	28770.54
2/9/2020	20136.81	10797.69	29475.93
3/9/2020	20374.73	10560.23	30189.24
4/9/2020	20612.66	10314.96	30910.35

Discussion

From the analysis both cases and case fatality variables require differencing to obtain stationarity. The ACF and PACF of both variable data show no autocorrelation. While cases models are ARIMA (2, 2, 2), case fatalities are ARIMA (1, 2, 3). The active cases are (0, 2, 1). Cases are lagged twice while case fatality is lagged once. It implies that previous infections have influence on current case occurrence as opposed to previous deaths. Asymptomatic and infected people with clear syndromes, not in quarantine, nor isolation, may move and interact with others leading to more infections, as opposed to deceased. The residuals of estimated case fatality and cases models are not autocorrelated as seen from the ACF and PACF. Also, the forecast of case fatality model in Table 2 and Figure 17 show that fatalities will stabilize on 9/8/2020 and begin to fall from 22/8/2020, the mentioned dates are forecasted to have 450 and 613 total fatalities respectively. The cases model forecasts that COVID-19 cases will stabilize after 17/8/2020 and begin to fall from 23/8/2020, both dates will have 35,779 and 41,517 COVID-19 total cases respectively.

Further, the forecast of active cases over time gives additional information. Whereas forecasts for cases and fatalities increase in both upper and lower bounds or limits, the lower limit of active cases forecast begins to reduce after 10/8/2020 showing 13669 infected persons. After this date the number of infected persons reduces to 10345 on 4/9/2020. Consequently, 10/8/2020 is possible Kenyan low peak. There is also a statistical possibility that the peaks of cases, fatalities, and infected persons occur at different intervals of time or rotating seasonal peaks. The progress of active cases over time carries the “energies”, “force” or “momentum” of COVID-19 pandemic.

Conclusion

Kenya reported the first case of the novel coronavirus disease in early March 2020 and instituted stringent measures to contain the pandemic in Kenya. The purpose of these public health measures were to reduce the spread of COVID-19 pandemic causing illnesses and mass deaths in the population The measures taken to control the spread and severity of COVID-19 from 7/7/2020 should be maintained to heighten saving lives during the pandemic of COVID-19 its magnitude and effect on individuals and local communities involved. It is expected to have a low peak in terms of fatalities in late August and early September 2020. Necessary actions should be done to accommodate more COVID-19 patients in referral and county health facilities. The COVID-19 burden on county health facilities and its severity could be reduced to real-time planning of resource allocation and improved access to home-based care, preceded by clear, accurate, and precise public education to increase awareness and continuum of informed decisions in the setting of various diseases. Appropriate resource planning for intense surveillance and response should be reviewed to accommodate the health security needs for the low peak. Public education should be considered to guard against stigmatization which is already evident from affected and infected individuals and local communities. Directives to mitigate emerging ethical, legal and economic hardships should be issued to strengthen the country’s commitment to promote and implement public health policies as well as reduce inequalities and cushion the vulnerable groups against depression and mental illness. Phased opening of learning may also be considered after September 2020.

Reference

1. Tania Dehesh HA, Mardani-Fard, Paria Dehesh. Forcasting of COVID-19 Confirmed Cases in Different Countries with ARIMA Models, http://doi.org/10.1101/2020.03.1320035345.
2. Ozili, Peterson K, COVID-19 in Africa: Socioeconomic Impact, Policy Response and Opportunities (May 2020). International Journal of Sociology and Social Policy, 2020, Available at SSRN: https://ssrn.com/abstract=3574767 or http://dx.doi.org/10.2139/ssrn.3574767
3. David Tartai and Zoltan Varalyay, COVID-19 epidemic outcome predictions based on logistic fitting and estimation of its reliability. 2020; Xiv:2003-14160.
4. Shem Otoi Sam. Exploring the statistical significance of Africa COVID-19 Data, ISSN: 2456-1452 Maths 2020; 5(4): 34-42 © 2020 Stats & Maths www.mathsjournal.com http://www.mathsjournal.com/pdf/2020/vol5issue4/PartA/5-3-17-711.pdf
5. Alanagreh L, Alzoughool F, Atoum M. The human coronavirus disease COVID-19: Its origin, characteristics, and insights into potential drugs and its mechanisms. Pathogens. 2020; 29(5):E331.
6. Guo Y, Cao Q, Hong Z *et al*. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak – an update on the status. Military Med Res, 2020, 7(11). https://doi.org/10.1186/s40779-020-00240-0