Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2022, 2022:17
https://doi.org/10.28919/cmbn/6888
ISSN: 2052-2541

FORECASTING CONFIRMED AND RECOVERED COVID-19 CASES AND DEATHS IN EGYPT AFTER THE GENETIC MUTATION OF THE VIRUS: ARIMA BOX-JENKINS APPROACH

MOHAMED R. ABONAZEL[1,*], NESMA M. DARWISH[2]

[1]Department of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research, Cairo University, Giza, Egypt

[2]Management and Information System Department, Higher Institute of Computer and Information Technology, El-Shrouk Academy, Cairo, Egypt

Abstract: Coronavirus disease 2019 (COVID-19) is spreading disease all over the world. It is a real test for all health authorities all over the world including Egypt. After the eruption of severe acute respiratory syndrome (SARS-2002/2003) and Middle East respiratory syndrome (MERS-2012/2014) in the world, new public health crisis, called new coronavirus disease (COVID-19). The coronavirus epidemic has spread over the world, affecting practically every country. As a result, it's become critical to comprehend disease trends in order to limit the consequences. The aim of this study, use the suitable statistical prediction models to meaningful in forecasting and controlling this global pandemic threat, especially after the genetic mutation of the virus in 2021. For this purpose, the Autoregressive Integrated Moving Average (ARIMA) model based on the Box-Jenkins approach was used to predict the confirmed, recovered cases and deaths of COVID-19 in Egypt. The most recent data available to determine the best prediction

*Corresponding author
E-mail address: mabonazel@cu.edu.eg
Received October 14, 2021

models for daily cases and death in Egypt, and to forecast them up to April 2021. The COVID-19 confirmed, recovered, and death cases were collected on a daily basis from the official Ministry of Health. According to our results, ARIMA models with ideally selected variables are excellent tools for monitoring and predicting trends of COVID-19 cases in Egypt. The results indicated that the estimated ARIMA models have a high ability to predict the number of confirmed cases, recovered COVID-19 cases and death in Egypt. The four stages of Box-Jenkins approach are conducted to obtain an appropriate ARIMA model for the number of confirmed cases, recovered COVID-19 cases and death in Egypt. According to the goodness-of-fit measures, the best model is ARIMA (1, 1, 1) for confirmed cases, ARIMA (1, 0, 1) for recovered cases, and ARIMA (1, 0, 0) for death. Moreover, we used these models to forecast the number of confirmed cases, recovered COVID-19 cases and death for the next twenty days. The results will enable us to provide a suitable advice to help in taking decision in Egypt on how to avoid negative effects for this epidemic.

Keywords: autoregressive integrated moving average (ARIMA) model; Box-Jenkins approach; COVID-19; forecasting; time series analysis.

2010 AMS Subject Classification: 37M10, 62P10.

1. INTRODUCTION

On 11 March 2020, the World Health Organization (WHO) declared COVID-19 a pandemic. Since it was discovered in Wuhan, China, on 31 December 2019, this pathogen has rapidly spread and infected more than 4 million people globally, with over 300,000 deaths as of March 2020. This virus is highly contagious with a basic reproductive. The eruption started on 12 December 2019 continues to spread worldwide and results in fatality. Since there is no approved treatment for COVID-19 currently prevention and preparation in healthcare services are crucial. Egypt is a country in Africa, located in the northeast corner of the continent, with a population of around 100 million people and a land area of 1,010,408 km2. It is surrounded on the north by the Mediterranean Sea, on the east by the Red Sea, on the west by Libya, and on the south by Sudan. The Egyptian Constitution declares health to be a fundamental human right. On January 26, 2020, Egypt will ban all flights from China to Egypt. The first confirmed case of a Chinese individual in Egypt was officially declared on February 14, 2020. The patient was taken to a quarantine facility.

The Egyptian cabinet formally denied any proven Egyptian cases on February 28, 2020. The second COVID-19 case was formally revealed on March 1, 2020. The first Egyptian case of COVID-19 death was reported on March 20, 2020, twenty days after the first case of COVID-19 death in Egypt (for a German person). Since the 7th of March through the 21st of March, all schools/universities and mosques have been closed. The Egyptian government has instituted various sorts of lockdowns in order to mitigate the epidemic's impact. Declining in the spring or summer or becoming the world's greatest pandemic. Most of Egyptian hospitals have been prepared to face COVID-19 cases to overcome shortages of medical supplies and staff.

The importance of projecting the pandemic's likely trajectory is underscored by the fact that, as far as we know, studies predicting cases and deaths in Egypt are still conservative and have not been revised to reflect the most recent scenario. Zaki et al. [1] talked about that coronavirus was confined from the sputum of a 60-year-old man who displayed with intense pneumonia and ensuing renal disappointment with a deadly result in the Kingdom of Saudi Arabia (KSA). Using statistical and artificial intelligence methods such as Autoregressive Integrated Moving Average (ARIMA) and Nonlinear Autoregressive Artificial Neural Networks, Saba and Elsheikh [2] predicted the likely number of cases in Egypt. The study used reported instances from March 1, 2020, to May 9, 2020, to forecast one month ahead until June 8, 2020. In May 2020, a 280 percent increase in cases was predicted.

El-Ghitany [3] used data from February 14 to April 18, 2020 to make a short-term projection for the pandemic scenario in Egypt. From April 19 to June 6, the researcher used an exponential growth rate model to forecast the daily cases. According to the findings, infections are likely to reach more than 20,000 by late May, after which they will begin to fall.

Elmousalami and Hassanien [4] used different time series models such as moving average (MA), weighted moving average, and single exponential smoothing to offer daily forecasting models of COVID-19 cases. According to the findings, the number of confirmed cases in Egypt would jump to four times in April. That instance, they claim that the number of coronavirus cases in Egypt is increasing dramatically.

Anwar and AbdelHafez [5] also predicted the expected timing of the coronavirus peak in Egypt. They used the Susceptible, Exposed, Infectious, and Removed model with the epidemic online calculator tool. The daily reports for the period 14 February to 11 May 2020 were used in their research. According to the findings, the number of hospitalized cases is expected to peak at 20,126 in mid-June. A total of 12,303 deaths were predicted. The research also claimed that the quarantine restrictions should be kept until the end of June 2020.

Nosier and Salah [6] presented a predictive study of COVID-19 infection and mortality in Egypt, a three-part study that attempts to discover the best prediction models for daily cases and deaths in Egypt and forecast them up to November 7, 2020, using the most recent available data. Second, using Google Community Movement Reports (GCMR) to evaluate the results of easing lockdown limitations, investigate the impact of mobility on pandemic incidence. Finally, they provided some recommendations that may help lessen the spread of the virus and eradicate new deaths as possible. Likewise, The Tachy Health team used a three-compartmental Susceptible-Infected-Recovered/Removed time-dependent epidemiological model to study the epidemic curve in four Middle East/North Africa countries: Egypt, the United Arab Emirates, KSA, and Algeria. The results for Egypt predicted a likely peak on June 6, 2020, assuming 5% of the population is vulnerable, a reproduction number of 2.6, and a 1/14 recovery rate. Another two scenarios were to reach the highest number of infections at 18 June or 16 July based on different assumptions.

Benvenuto et al. [7] proposed an econometric model to predict the spread of COVID-2019. They introduced ARIMA model prediction on the Johns Hopkins epidemiological data to predict the epidemiological trend of the prevalence and incidence of COVID-2019. Wu et al. [8] estimated the overall symptomatic case fatality risk (the probability of dying after developing symptoms) of COVID-19 in Wuhan.

Recently, Awwad et al. [9] used the ARIMA and spatial time ARIMA models to study the effect of the curfew on the spread of COVID-19 in the KSA during the period from May 31 to October 11, 2020. Their results indicated that the Spatial Time-Autoregressive Integrated Moving Average (STARIMA) model is very reliable in predicting cases of COVID-19.

Recently, there are several papers that have used ARIMA models to study the behavior of the COVID-19 pandemic in different countries, such as [10, 11, 12, 13].

2. Statistical Methodology

2.1 Data

Data for this study were obtained from 1 February 2021 to 31 March 2021. The Egyptian daily COVID-19 confirmed, recovered and death cases were sourced from Egyptian Ministry of Health and Population reports. This study applied ARIMA model to forecast the Egyptian daily COVID-19 confirmed, recovered and death cases. Minitab statistical software version 16 was used to conduct the analyses. Thus, the appropriate ARIMA model and then use it to forecast the confirmed, recovered cases and death for the next twenty days.

2.2 Time Series Models

Time series analysis has been used in a variety of medical, engineering, and economic disciplines. Monitoring the responses of a phenomenon through time and anticipating future responses is a significant subject that can aid decision-makers in developing future policies and plans to address various issues that humans confront. In the literature, both statistical and artificial intelligence-based approaches for forecasting time series problems have been documented. Statistical-based methods like ARIMA are commonly used to assess this type of data and estimate future reactions as a function of time. The ARIMA model combines three processes: the Autoregressive (AR) process, the Integration (I) process (by taking the difference), and the Moving-Average (MA) process. These processes are referred to the primary univariate time series models in statistical literature, and they are widely employed in a variety of applications.

1- Autoregressive (AR) Model

Abraham and Ledolter [14] defined an AR model forecasts future behavior depending on the past behavior. It’s used for forecasting when there is some correlation between values in a time series and the values that precede and succeed them. You only use past data to model the behavior, hence the name autoregressive (the Greek prefix auto–means “self.”). The process is basically a linear regression of the data in the current series against one or more past values in the same series. The

AR model is known in statistical literature as main univariate time series model, and is commonly used in many applications.

An AR (p) model is an autoregressive model in which the predictor variables are specific lagged values of y_t are used as predictor variables. Lags exist when the outcomes of one time period have an impact on subsequent time periods. The AR (p) model is defined by the equation:

$$y_t = c + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p} + \varepsilon_t, t = 1,2,.,T, \quad (1)$$

Where ε_t is a white noise, a sequence of independently and identically distributed (iid) random variables with $E(\varepsilon_t) = 0$ and $var(\varepsilon_t) = \sigma^2$; i.e. $\varepsilon_t \sim iidN(0, \sigma^2)$.

2- Moving-Average (MA) model

Abraham and Ledolter [14] described that, rather than using past values of the forecast variable in a regression, a moving average (MA) model uses past forecast errors in a regression model. The notation MA(q) refers to the moving average model of order q:

$$y_t = c - \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}. \quad (2)$$

3- Autoregressive moving-average model

An Autoregressive Moving-Average (ARMA) is a model-based time series fitting model that provide a parsimonious description of a (weakly) stationary stochastic process in terms of two polynomials, one for the auto-regression and the second for the moving-average [14]. We can write an $\mathrm{ARMA(p, q)}$ as a mixture of AR(p) and MA(q) models:

$$y_t = c + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}. \quad (3)$$

4- ARIMA Model

ARIMA, which stands for "Autoregressive Integrated Moving Average," is a class of models that "explains" a given time series based on its own previous values, i.e., its own lags and lagged prediction errors, so that equation can be used to forecast future values. ARIMA models can be used to model any ‘non-seasonal' time series and it is not random white noise. An ARIMA model is one where the time series was differenced at least once to make it stationary and you combine the AR and the MA terms [14]. For example, if y_t is non-stationary series, we will take a first-difference of y_t so that Δy_t becomes stationary, then the ARIMA (p, 1, q) model is:

$$\Delta y_t = c + \alpha_1 \Delta y_{t-1} + \alpha_2 \Delta y_{t-2} + \cdots + \alpha_p \Delta y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}, \quad (4)$$

where $\Delta y_t = y_t - y_{t-1}$. But if $p = q = 0$ in equation (4), then the model becomes a random walk model which classified as ARIMA (0, 1, 0).

2.3 Box-Jenkins Approach

The Box-Jenkins [15] model is a mathematical model designed to forecast data ranges based on inputs from a specified time series. The Box-Jenkins model can analyze several different types of time series data for forecasting purposes [16]. Its methodology uses differences between data points to determine outcomes. The methodology allows the model to identify trends using autoregresssion, moving averages, and seasonal differencing to generate forecasts. Figure 1 shows the four iterative stages of modeling according this approach. The four stages modeling in the Box-Jenkins iterative approach [17]:

Step 1 (Model identification): determining the order of the model required (p, d, and q) in order to capture the salient dynamic features of the data. This usually results in the employment of graphical approaches (plotting the series, the Autocorrelation function (ACF), partial autocorrelation function (PACF), etc).

Step 2 (Model estimate and selection): estimating the parameters of various models (using step 1) and then making a first model selection (using information criteria). The most common methods for estimation are Maximum Likelihood Estimation (MLE) or non-linear least-squares estimation.

Step 3 (Model checking): determining if the provided and estimated model(s) are sufficient. One, in particular, employs residual diagnostics. The residuals should be independent of one another and consistent in mean and variance over time; displaying the residuals' ACF and PACF can assist reveal misspecification. If the estimation isn't good enough, we'll have to go back to step one and try again. Furthermore, the estimated model should be compared to different ARIMA models in order to determine which model is appropriate for the data. The three common criteria used in model selection: Mean Squared Error (MSE), Mean Absolute Deviation (MAD), and Mean

Absolute Percentage Error (MAPE).

Step 4 (Forecasting): when the selected ARIMA model conforms to the specifications of a stationary univariate process, then we can use this model for forecasting.

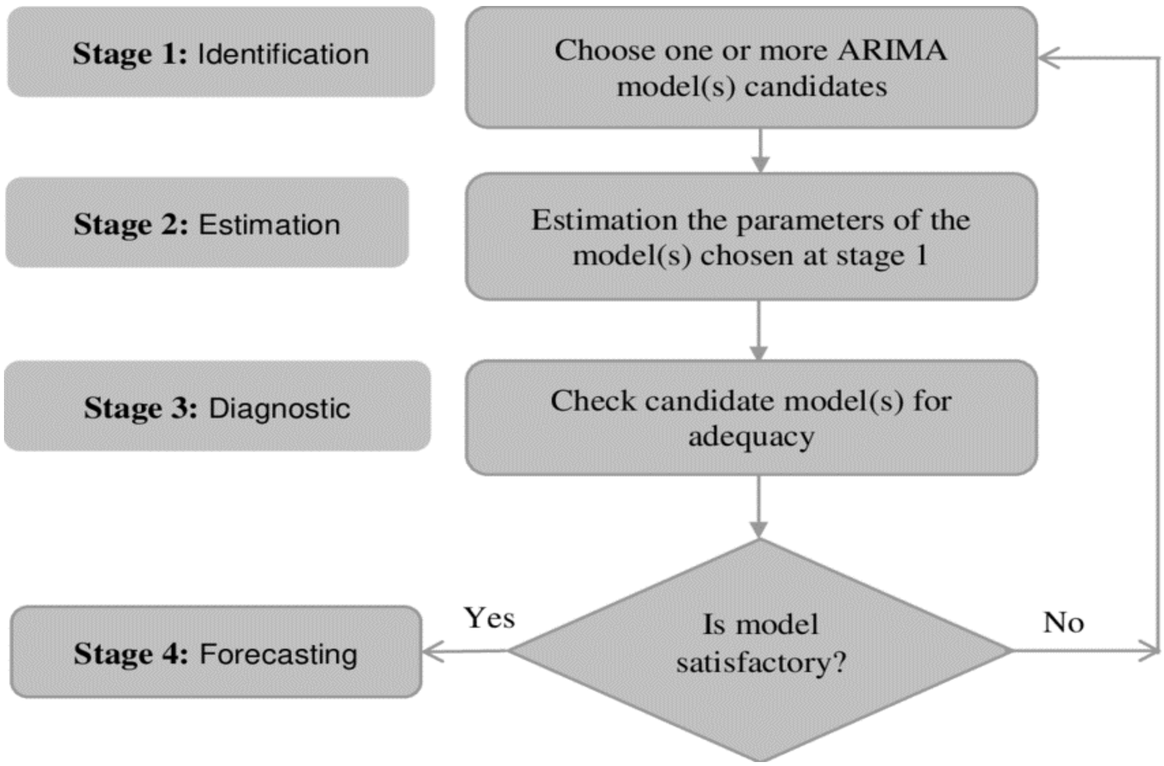

Figure 1: Stages in the Box-Jenkins iterative approach [17]

3. RESULTS

3.1 Estimated models

Time series analysis was made for the number of confirmed, recovered cases and deaths in Egypt. When the time series graphs are examined, the trend is seen, after taking first differences as Figure 2. The ACF and PACF graphics were used to see this more clearly and determine it's stationary. In the ACF graph, it is seen that the series of confirmed case is not stationary since many delays exceed the confidence limits. In this case, the first order difference was applied to the series of confirmed cases and it was ensured to become stationary. Figure 2 and the ACF and PACF graphs (in Figures 3, 4, and 5) are confirmed that, after the first difference procedure to the confirmed cases, the series became stationary.

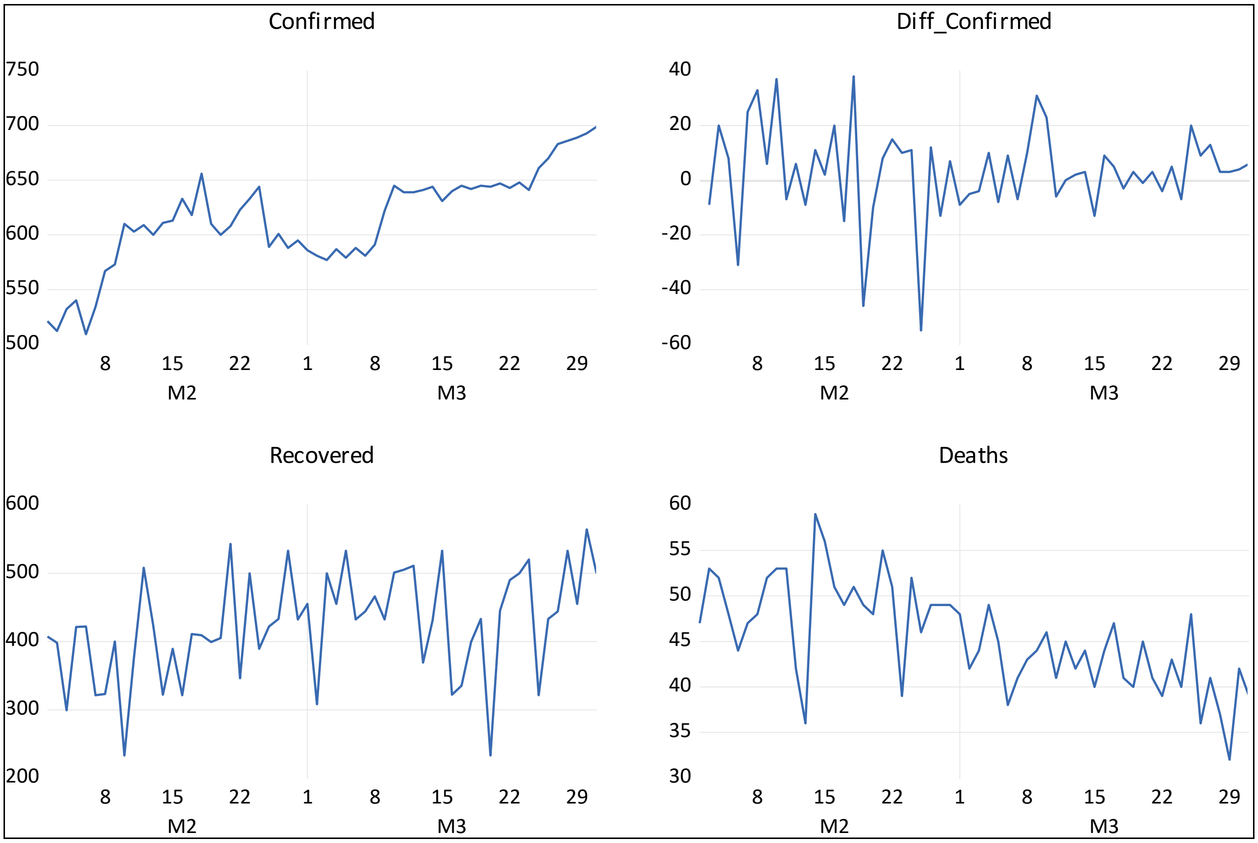

Figure 2: Time series plots for confirmed, difference confirmed, recovered, and death cases

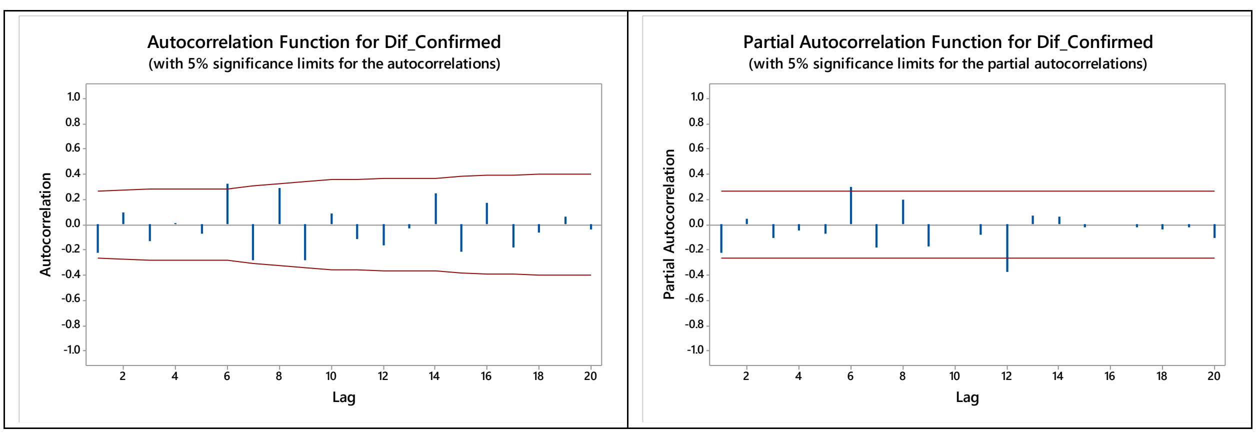

Figure 3: ACF and PACF plots for confirmed cases (after the first difference)

Autocorrelation function (ACF) and partial autocorrelation (PACF) graphs were used to determine the order of the ARIMA model. The best ARIMA models for the three series are shown in Table 1: the ARIMA (1, 1, 1) model for confirmed cases, ARIMA (1, 0, 1) model for recovered cases, and

ARIMA (1, 0, 0) model for deaths. The above models were compared with different ARIMA models to select the best model for the data using goodness-of-fit measures (MSE, MAD, MAPE), as in Table 1.

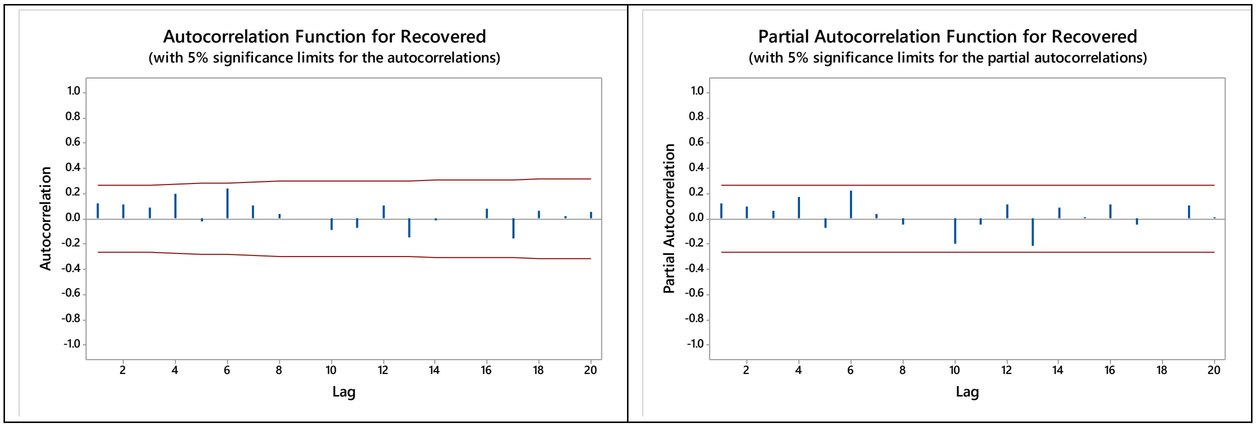

Figure 4: ACF and PACF plots for recovered cases

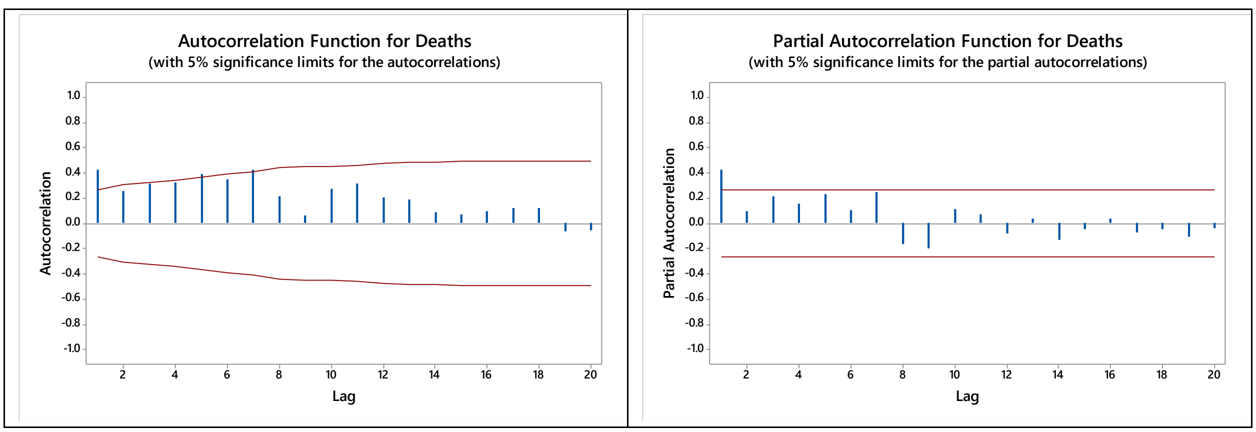

Figure 5: ACF and PACF plots for death cases

1- Confirmed Cases Model

The properly model for the confirmed cases is ARIMA (1, 1, 1). Modeling results of an ARIMA (1, 1, 1) process have been estimated by MLE and are presented in the Table 2. The coefficient estimate of AR (1) and MA (1) are statistically significant at 1% level of significance and the model overall are statistically significant at 5% level of significance.

Table 1: Evaluation of various ARIMA models

Data	**Model**	**MSE**	**MAD**	**MAPE**
Confirmed cases	ARIMA (2,1,0)	275.1	192.787	21.539
	ARIMA (2,1,1)	264.0	212.654	11.787
	ARIMA (1,1,1)	**260.2**	**164.3**	**0.10**
Recovered cases	ARIMA (2,0,0)	6046	121.32	14.41
	ARIMA (1,0,0)	6011	76.56	12.32
	ARIMA (1,0,1)	**5842**	**54.51**	**0.10**
Death cases	ARIMA (1,0,1)	26.62	28.238	6.21
	ARIMA (2,0,0)	25.81	26.82	3.2383
	ARIMA (1,0,0)	**25.67**	**3.2384**	**0.251**

Table 2: Estimated ARIMA (1, 1, 1) model for confirmed cases

Variable	**Estimate**	**Std. Error**	**t-statistic**	**p-value**
AR(1)	-0.9752	0.0544	-17.93	0.000
MA(1)	-0.8851	0.1052	-8.41	0.000
Constant	6.226	4.028	1.55	0.128
Diagnostic tests				
F-statistic	3.537294	**Box-Pierce test Chi-Square**		40.5
p-value of F	0.036141	**p-value of Chi-Square**		0.173

According to Box-Jenkins approach, the diagnostic tests of the model are checking; the normality and the stationary of the residuals. Figure 6 shows the residuals are stationary. Also, Figure 7 shows the values of residuals are distributed normally it is obvious from histogram and Anderson-Darling Normality test, as p-value for the series of confirmed is (0.074) greater than 0.05 which means the data is normal with 95% confidence interval. According to the results in Table 1, the best model is ARIMA (1, 1, 1), because have the minimum values of MSE, MAD and MAPE. Moreover, the predictive power of the model is very high as indicated by the small difference between actual and fitted values as presented in Figure 7.

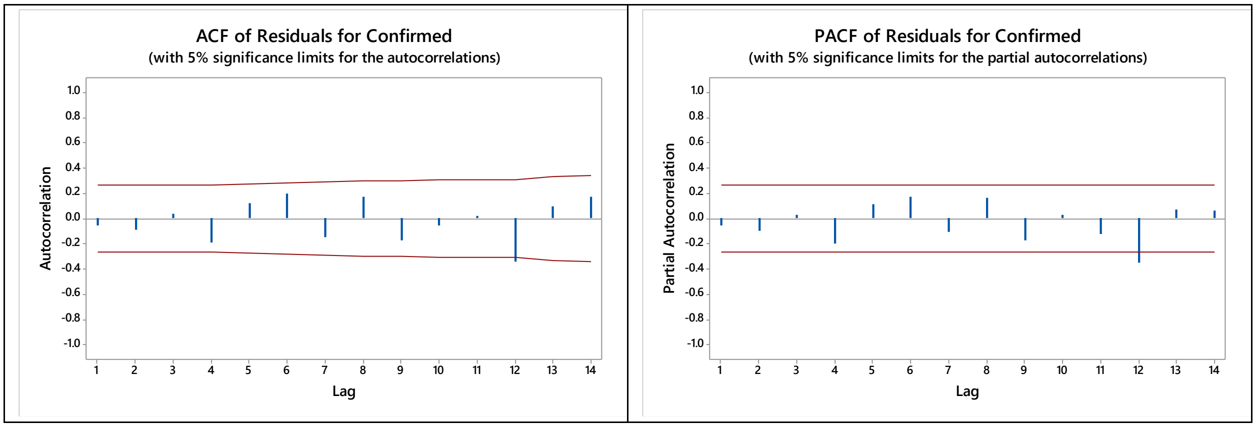

Figure 6: ACF and PACF plots of the residuals of the ARIMA (1, 1, 1) model

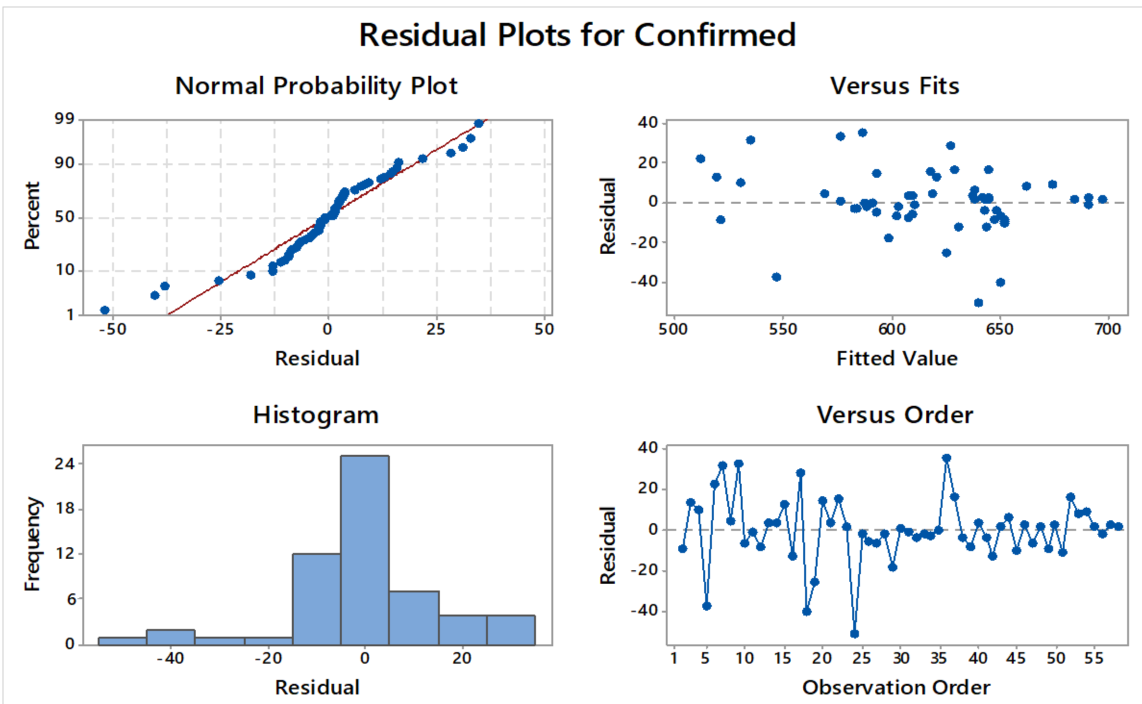

Figure 7: Residuals analysis plots of ARIMA (1, 1, 1) model

2- Recovered Cases Model

The properly model for the recovered cases is ARIMA (1, 0, 1), see Table 1. Modeling results of an ARIMA (1, 0, 1) process have been estimated by MLE and are presented in the Table 3. The coefficient estimate of AR (1) and MA (1) are statistically significant at 1% level of significance and the model overall are statistically significant at 1% level of significance.

Table 3: Estimated ARIMA (1, 0, 1) model for recovered cases

Variable	**Estimate**	**Std. Error**	**t-statistic**	**P-value**
AR(1)	0.9180	0.1561	5.88	0.000
MA(1)	0.7967	0.2190	3.64	0.001
Constant	34.844	2.218	15.71	0.000
Diagnostic tests				
F-statistic	5.519797	**Box-Pierce test Chi-Square**		26.2
p-value of F	0.006589	**p-value of Chi-Square**		0.793

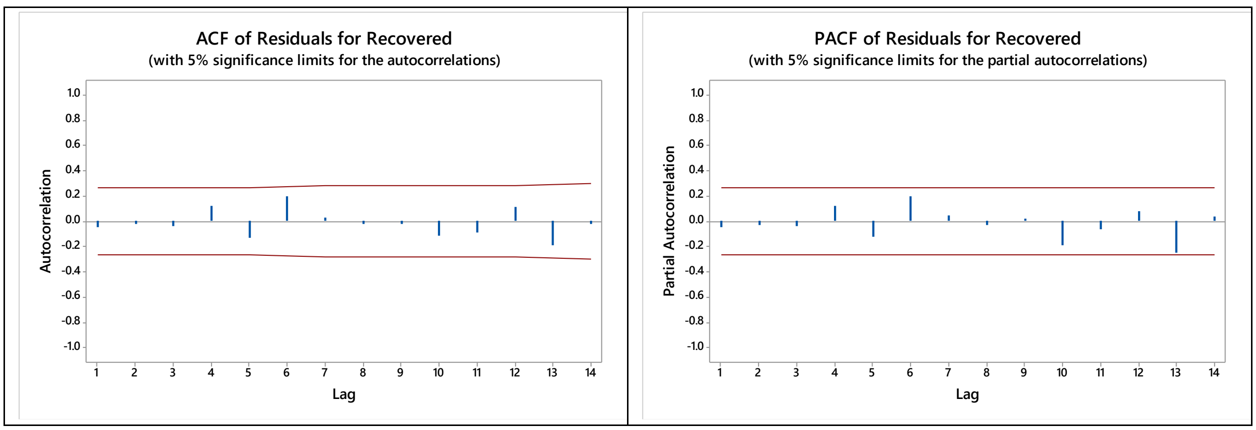

Figure 8: ACF and PACF plots of the residuals of the ARIMA (1, 0, 1) model

According to Box-Jenkins approach, the diagnostic tests of the model are checking the normality and the stationary of the residuals. Figure 8 shows the residuals are stationary. Also, Figure 9 shows the values of residuals are distributed normally it is obvious from histogram and Anderson-Darling Normality test, as p-value for the series of confirmed is (0.084) greater than 0.05 which means the data is normal with 95% confidence interval.

3- Death Cases Model

The properly model for the death cases is ARIMA (1, 0, 0), see Table 1. Modeling results of an ARIMA (1, 0, 0) process have been estimated by MLE and are presented in the Table 4. The coefficient estimate of AR (1) is statistically significant at 1% level of significance and the model overall are statistically significant at 1% level of significance.

According to Box-Jenkins approach, the diagnostic tests of the model are checking the normality and the stationary of the residuals. Figure 10 shows the residuals are stationary. Also, Figure 11 shows the values of residuals are distributed normally it is obvious from histogram and Anderson-Darling normality test, as p-value for the series of confirmed is (0.657) greater than 0.05 which means the data is normal with 95% confidence interval.

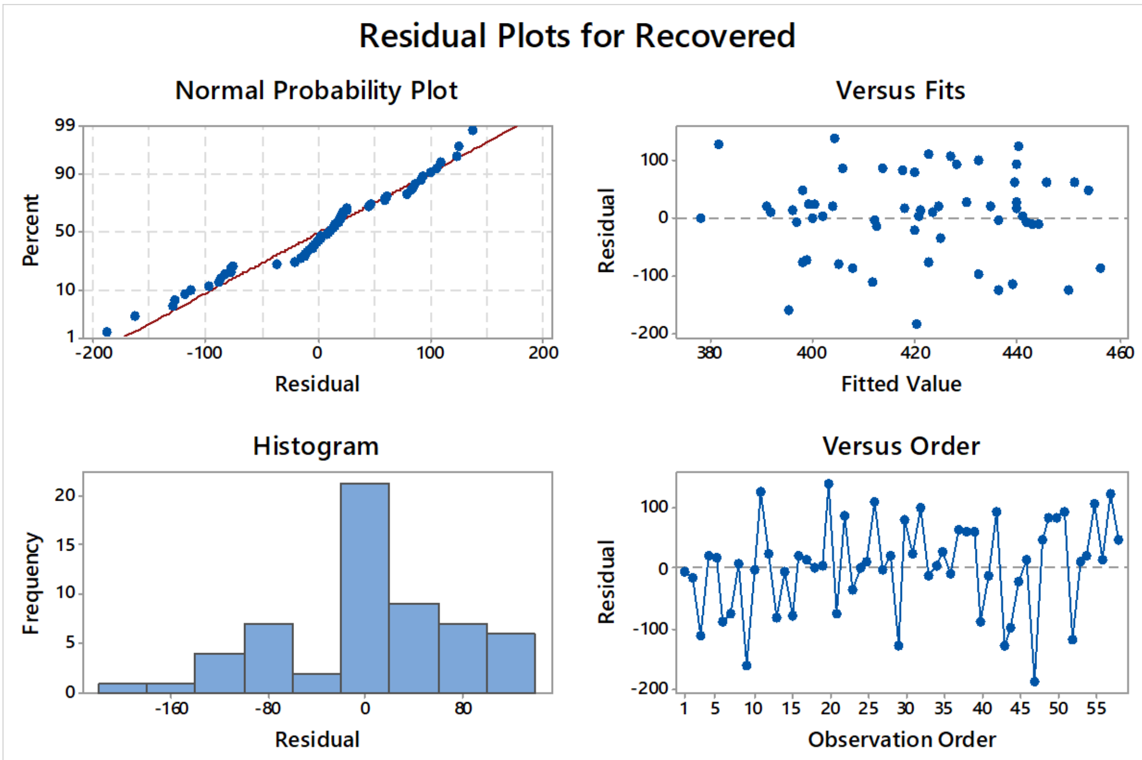

Figure 9: Residuals analysis plots of ARIMA (1, 0, 1) model

Table 4: Estimated ARIMA (1, 0, 0) model for death cases

Variable	**Estimate**	**Std. Error**	**t-statistic**	**P-value**
AR(1)	0.4356	0.1221	3.57	0.001
Constant	25.6040	0.6656	38.47	0.000
Diagnostic tests				
F-statistic	12.49723	**Box-Pierce test Chi-Square**		28.2
p-value of F	0.000835	**p-value of Chi-Square**		0.171

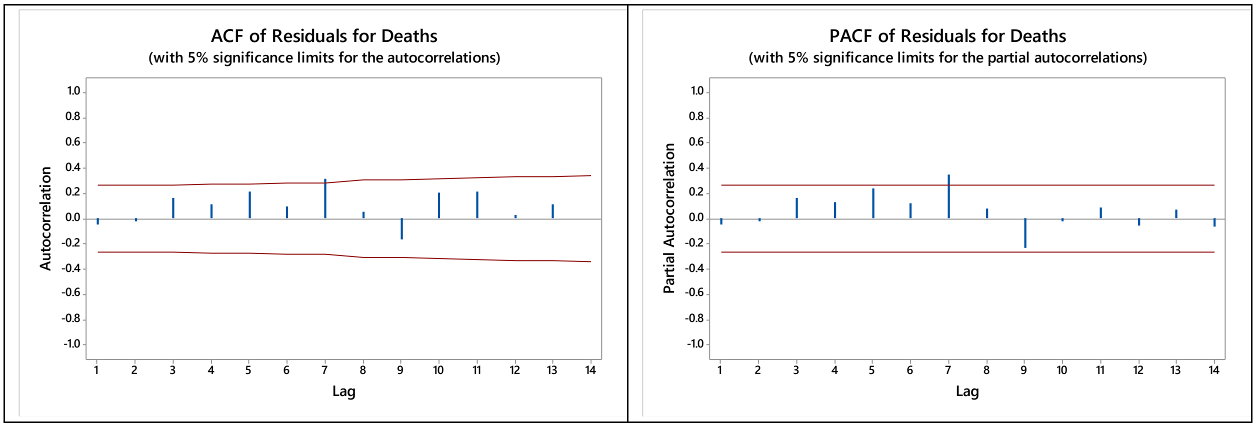

Figure 10: ACF and PACF plots of the residuals of the ARIMA (1, 0, 0) model

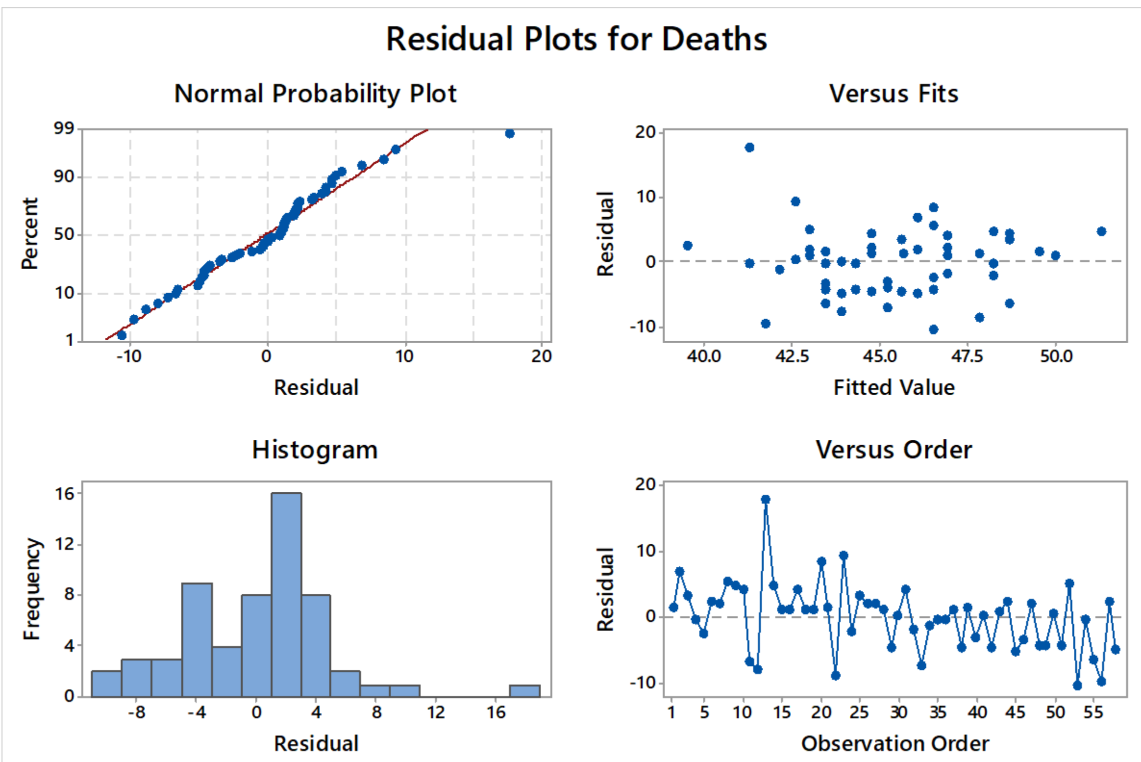

Figure 11: Residuals analysis plots of ARIMA (1, 0, 0) model

3.2 Forecasting

The forecasting results of the number of confirmed cases, recovered cases and death of the Egypt are given in Tables 5, 6, and 7. As ARIMA (1, 1, 1) model is fit the number of confirmed cases, then, forecasts the confirmed cases values with 95% confidence limits for the next twenty days. The forecasted values of the number of confirmed cases are given in Table 5, also ARIMA (1, 0, 1) model is fit the number of recovered cases, therefore we forecast recovered values for the next twenty days. The forecasted values of the numbers of recovered cases are given in Table 6, finally ARIMA (1, 0, 0) model is fit the death cases, therefore we forecast death values for the next twenty days. The forecasted values of the death cases are given in Table 7. Figures 12, 13, and 14 present the trend of the actual and the forecasted confirmed, recovered and death values with their 95% confidence limits.

Table 5: Forecasted values of daily confirmed cases

Date	**Forecast**	**95% Confidence Interval**		**Actual**
		Lower	**Upper**	
4/1/2021	202832	202800	202863	202843
4/2/2021	203537	203469	203605	203546
4/3/2021	204244	204130	204359	204256
4/4/2021	204956	204789	205122	204965
4/5/2021	205669	205444	205895	205732
4/6/2021	206387	206098	206676	206510
4/7/2021	207107	206748	207466	207293
4/8/2021	207831	207398	208264	208082
4/9/2021	208557	208046	209069	208876
4/10/2021	209288	208693	209882	209677
4/11/2021	210020	209339	210701	210489
4/12/2021	210757	209985	211528	211307
4/13/2021	211495	210629	212362	212130
4/14/2021	212238	211275	213202	212961
4/15/2021	212984	211918	214049	213798
4/16/2021	213733	212563	214902	214639
4/17/2021	214484	213206	215762	215484
4/18/2021	215240	213851	216628	216334
4/19/2021	215997	214495	217500	217186
4/20/2021	216759	215140	218379	218041

Table 6: Forecasted values of daily recovered cases

Date	**Forecast**	**95% Confidence Interval**		**Actual**
		Lower	**Upper**	
4/1/2021	155151	155002	155301	155016
4/2/2021	155606	155381	155831	155448
4/3/2021	156058	155767	156350	155886
4/4/2021	156508	156155	156862	156219
4/5/2021	156956	156543	157369	156574
4/6/2021	157402	156931	157874	157006
4/7/2021	157847	157318	158375	157450
4/8/2021	158290	157705	158874	157889
4/9/2021	158731	158091	159371	158454
4/10/2021	159171	158477	159865	159054
4/11/2021	159610	158862	160357	159499
4/12/2021	160047	159247	160847	159999
4/13/2021	160484	159632	161335	160431
4/14/2021	160919	160017	161822	161031
4/15/2021	161354	160402	162307	161470
4/16/2021	161788	160787	162790	162170
4/17/2021	162221	161171	163271	162714
4/18/2021	162654	161556	163752	163479
4/19/2021	163086	161941	164230	163812
4/20/2021	163517	162327	164708	164368

Table 7: Forecasted values of daily death cases

Date	**Forecast**	**95% Confidence Interval**		**Actual**
		Lower	**Upper**	
4/1/2021	12037.6	12027.7	12047.5	12041
4/2/2021	12081.8	12064.4	12099.1	12084
4/3/2021	12126.6	12102.9	12150.3	12123
4/4/2021	12171.7	12142.6	12200.9	12163
4/5/2021	12217.0	12183.1	12250.9	12210
4/6/2021	12262.3	12224.2	12300.5	12253
4/7/2021	12307.7	12265.7	12349.7	12290
4/8/2021	12353.0	12307.5	12398.6	12323
4/9/2021	12398.4	12349.6	12447.2	12362
4/10/2021	12443.8	12391.9	12495.6	12405
4/11/2021	12489.1	12434.3	12543.9	12445
4/12/2021	12534.5	12477.0	12592.0	12487
4/13/2021	12579.9	12519.7	12640.0	12526
4/14/2021	12625.2	12562.5	12687.9	12570
4/15/2021	12670.6	12605.5	12735.7	12611
4/16/2021	12716.0	12648.5	12783.4	12653
4/17/2021	12761.3	12691.6	12831.1	12694
4/18/2021	12806.7	12734.8	12878.6	12738
4/19/2021	12852.1	12778.1	12926.1	12778
4/20/2021	12897.4	12821.4	12973.5	12820

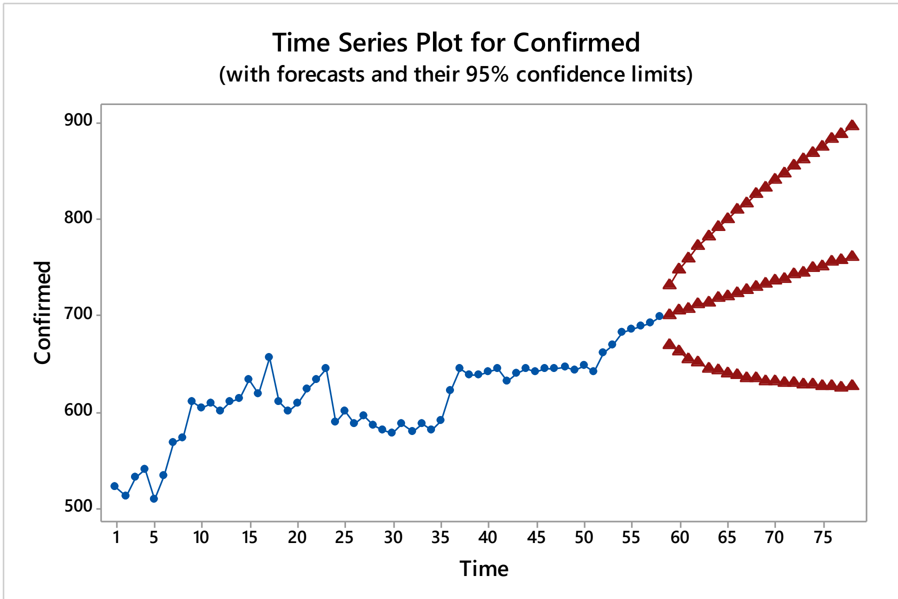

Figure 12: Time series plot for actual and forecasted for confirmed cases

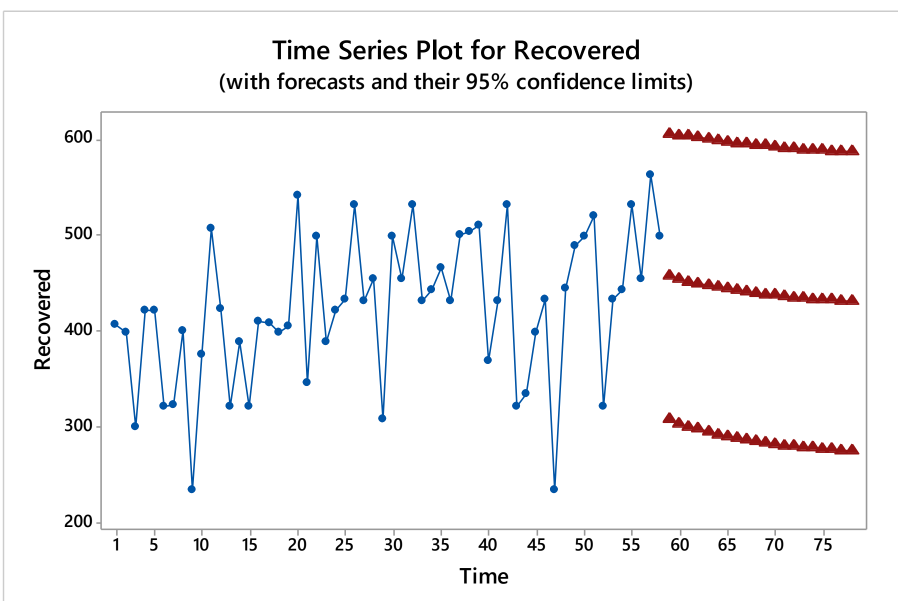

Figure 13: Time series plot for actual and forecasted for recovered cases

Figure 14: Time series plot for actual and forecasted for death cases

4. Discussion

The aim of this study is proposed three ARIMA models to meaningful in forecasting and controlling COVID-19 cases in Egypt, especially after the genetic mutation of this virus in 2021. Based on the recently data from the official Ministry of Health up to April 2021, we conduct three ARIMA models of COVID-19 confirmed, recovered, and death cases. The Box-Jenkins approach has been used to get more efficient ARIMA models for the data. We can summarize the resulting models in the following equations:

- The estimated regression equation of ARIMA (1, 1, 1) model for confirmed cases is:

$$\Delta Y_t = 6.226 - 0.9752\ \Delta Y_{t-1} - 0.8851\ \varepsilon_{t-1} + \varepsilon_t,$$

- The estimated regression equation of ARIMA (1, 0, 1) model for recovered cases is:

$$Y_t = 34.844 + 0.9180\ Y_{t-1} + 0.7967\varepsilon_{t-1} + \varepsilon_t,$$

- The estimated regression equation of ARIMA (1, 0, 0) model for death cases is:

$$Y_t = 25.60 + 0.4356 Y_{t-1} + \varepsilon_t,$$

Since the aim of this study is forecast the COVID-19 cases in Egypt, we used the three proposed ARIMA models to make a forecast of the COVID-19 confirmed, recovered, and death cases in the next twenty days. The forecasted values indicated that the forecasted confirmed and death cases would increase, while the cases of recovered would decrease. So we can say that the ARIMA model is a good statistical model for predicting COVID-19 cases in Egypt even if the virus is genetically modified. Our findings will enable us to provide appropriate advice to assist decision-making in Egypt on how to avoid the negative effects of this pandemic.

5. CONCLUSION

The aim of the study was to model and forecast the number of confirmed cases, recovered COVID-19 cases and death using the Box-Jenkins approach based on data were obtained from 1 February 2021 to 31 March 2021. The four stages of Box-Jenkins approach are conducted to obtain an appropriate ARIMA model for the number of confirmed cases, recovered COVID-19 cases and death in Egypt, and we used this model to forecast the number of confirmed cases, recovered COVID-19 cases and death for the next twenty days. Time series plots were used for testing the stationarity of the data. Also, the MLE was used for estimating the model. Using the different goodness-of-fit measures (MSE, MAD, and MAPE), the various ARIMA models with different order of autoregressive and moving-average terms were compared. According to MSE, MAD, and MAPE measures, the best model is ARIMA (1, 1, 1) for confirmed cases, ARIMA (1, 0, 1) for recovered cases, and ARIMA (1, 0, 0) for death.

In Future work, we can use one of the multivariate time series models to predict COVID-19 cases in Egypt such as an autoregressive distributed lag model [18, 19, 20] or a vector autoregressive model [21].

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

[1] A.M. Zaki, S. van Boheemen, T.M. Bestebroer, A.D.M.E. Osterhaus, R.A.M. Fouchier, Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia, N. Engl. J. Med. 367 (2012), 1814–1820.

[2] A.I. Saba, A.H. Elsheikh, Forecasting the prevalence of COVID-19 outbreak in Egypt using nonlinear autoregressive artificial neural networks, Process Safe. Environ. Protect. 141 (2020), 1–8.

[3] E. El-Ghitany, A short-term forecast scenario for COVID-19 epidemic and allocated hospital readiness in Egypt, (2020). https://doi.org/10.20944/preprints202004.0473.v1.

[4] H.H. Elmousalami, A.E. Hassanien, Day level forecasting for coronavirus disease (COVID-19) spread: analysis, modeling and recommendations, ArXiv:2003.07778 [q-Bio, Stat]. (2020). http://arxiv.org/abs/2003.07778.

[5] W.A. Anwar, A. Mokhtar, Forecasting the peak of novel coronavirus disease in Egypt using current confirmed cases and deaths, Epidemiology, (2020). https://doi.org/10.1101/2020.05.31.20118182.

[6] S. Nosier, R. Salah, Forecasting Covid-19 infections and deaths horizon in Egypt, Epidemiology, (2020). https://doi.org/10.1101/2020.09.28.20202911.

[7] D. Benvenuto, M. Giovanetti, L. Vassallo, S. Angeletti, M. Ciccozzi, Application of the ARIMA model on the COVID-2019 epidemic dataset, Data Brief. 29 (2020), 105340.

[8] J.T. Wu, K. Leung, M. Bushman, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China, Nat. Med. 26 (2020), 506–510.

[9] F.A. Awwad, M.A. Mohamoud, M.R. Abonazel, Estimating COVID-19 cases in Makkah region of Saudi Arabia: Space-time ARIMA modeling, PLoS ONE. 16 (2021), e0250149.

[10] H.B. Kibria, O. Jyoti, A. Matin, Forecasting the spread of the third wave of COVID-19 pandemic using time series analysis in Bangladesh, Inform. Med. Unlocked. 28 (2022), 100815.

[11] R. Yagoub, H. Eledum, Modeling of the COVID-19 cases in gulf cooperation council countries using ARIMA and MA-ARIMA models, J. Probab. Stat. 2021 (2021), 1623441.

[12] D. Talukdar, V. Tripathi, COVID-19 forecast for 13 Caribbean countries using ARIMA modeling for confirmed, death, and recovered cases, F1000Res. 10 (2021), 1068.

[13] M.A. Rguibi, N. Moussa, A. Madani, A. Aaroud, K. Zine-dine, Forecasting Covid-19 transmission with ARIMA and LSTM techniques in Morocco, SN Comput. Sci. 3 (2022), 133.

[14] B. Abraham, J. Ledolter, Statistical methods for forecasting, Wiley, New York, (1983).

[15] G.E.P. Box, G.M. Jenkins, Times series analysis forecasting and control, Holden-Day, San Francisco, (1970).

[16] P.J. Brockwell, R.A. Davis, Introduction to time series and forecasting, 2nd. ed., Springer, New York, (2002).

[17] M.R. Abonazel, A.I. Abd-Elftah, Forecasting Egyptian GDP using ARIMA models, Rep. Econ. Finance, 5(2019), 35-47.

[18] M.R. Abonazel, N. Elnabawy, Using the ARDL bound testing approach to study the inflation rate in Egypt, Econ. Consult. 31(2020), 24-41.

[19] A.A. El-Sheikh, F.A. Alteer, M.R. Abonazel, Four imputation methods for handling missing values in the ARDL model: An application on Libyan FDI. J. Appl. Probab. Stat. In Press.

[20] M.R. Abonazel, F.A. Awwad, K. Nwuju, et al. Long-run determinants of Nigerian inflation rate: ARDL bounds testing approach, WSEAS Trans. Bus. Econ. 18(2021), 1370-1379.

[21] K. Rajab, F. Kamalov, A.K. Cherukuri, Forecasting COVID-19: Vector autoregression-based model, Arab J. Sci. Eng. (2022). https://doi.org/10.1007/s13369-021-06526-2.