COVID-19 IN SAUDI ARABIA: *DEMYSTIFYING REALITY*

[1]Abdullah M. Almarashi, [2]Diyab H.Alyami, [3]Saeed Baradwan, [4]Adnan A. Alsuwat, [5]Mohammed I.Alnasser, [6]Ahmad M. Ghulman, [7]Gahad Mathai Tmiahai, [8]Mana N. Alzamanan, [9]Fahad Z.Almutairi, [10]Abdullah o al-Ansari, [11]Ahmad M. Alkhoshi, [12]Majed M Alshehri, [13]Ahmed A.Alasmari, [14]Abdullah Abdrabuh Aloufi, [15]Amer S.Alruwaili, [16]Ahmad M Alzmanan, [17]Mohammed Q Alotaibie, [18]Naser S. Alhareth , [19]Ahmed A Almurashi, [20]Khushnoor Khan*

[1,20] Department of Statistics, Faculty of Science, King Abdulaziz University, 21589, Jeddah, Saudi Arabia
[2,18]Department of Health Services and Hospital, Faculty of Economics and Administration, King Abdulaziz University, Jeddah. Saudi Arabia
[19]School of Mathematical Sciences, Faculty of Science and Technology, University of Kebangsaan Malaysia.

**Email: Khushnoorkhan64@gmail.com*

ABSTRACT: *Present study focuses on a two-pronged analytical strategy -exploratory analysis and time series analysis, on COVID-19 data of Saudi Arabia. The data was collected by students of Executive Masters' Program HSA613 studying in the first semester of 2021. Data on five study variables were collected covering a period of 220 days (15th March 2020 – 20th October 2020). In an exploratory analysis, simple line graphs were used to study the behavioral pattern of three variables (confirmed cases, recovered cases, death cases) individually. The outcome of exploratory results revealed a cyclic pattern of seven months for the three study variables. For exploratory analysis, statistical software R was employed. For Time series analysis the autoregressive integrated moving average (ARIMA) model and Autoregressive Conditional Heteroscedasticity (ARCH) models were used. Augmented Dickey Fuller (ADF) stationary test (unit root test) showed that daily confirmed cases, death cases, and daily recovered cases of COVID-19 are stationary after the first differences. The future daily forecast of confirmed cases of COVID-19 using the ARCH(1) model was statistically significant and therefore suitable for future forecasting. Results of the 30-days forecasts exhibited exponential growth of COVID-19 which then started declining in the middle of the 7-month cycle. Results of the present study will have a dual effect both on the civil administration and health providing personnel on when to focus and exert more during the COVID-19 if it resurges again. Complete eradication of the present pandemic is still a far cry but a step in the right direction at the outset of a pandemic minimizes its effect on the public in general and on health-providing personnel in particular*.

Keyword: Confirmed cases, Death cases, Recovery cases, Lockdown, ARIMA ADF test, ARCH Model.

1. INTRODUCTION/ LITERATURE REVIEW

The focus of this study is to demystify some aspects of the data on the COVID-19 pandemic. It is important to note that this rapidly spread coronavirus was the cause of respiratory disease in a group of people in the city of Wuhan, Hubei, China on 31st December 2019 [11] . A few days later, the causative agent of this mysterious pneumonia was identified as a novel coronavirus. This causative virus has been temporarily named as severe acute respiratory syndrome coronavirus 2 and the relevant infected disease has been named as coronavirus disease 2019 (COVID-19) by the WHO. The emergence of this terrible disease has caused fear and great concern among the nations of the world. The Ministry of Health (MOH) of Saudi Arabia announced on 02 March 2020 the first case of coronavirus infection in a citizen who returned from Iran via the Kingdom of Bahrain Saudi Arabia, which led the Government of Saudi Arabia to take a number of measures, including a complete lockdown of the entire country, to contain the spread of the coronary virus pandemic.

In Saudi Arabia, the emergence of covid-19 is part of the 2019 global coronavirus pandemic (COVID-19) caused by severe acute coronavirus respiration syndrome 2 [1]. The Kingdom's first case was confirmed in the following months by the Ministry of Health [2], whereas cases were strongly confirmed in the following months in the Persian Gulf Arab States.

The Kingdom announced on 21 March that it would halt all domestic and international travel and resume domestic travel on 21 May. The number of confirmed daily cases decreased dramatically when curfews and lockouts were installed at several administrative levels, with the exception of Mecca, and all curfews were lifted by a 3-phase national programme, until 21 June. [3] has identified several economic stimulus measures, including a decrease of 5 % to 15% in a value-added tax effective July 1 and a decrease of 100 billion riyals ($26.6 million). As of 7 September, the Kingdom has seen 320,000 confirmed cases, with over 4,000 viral deaths and international air travel continues. Saudi Arabia announced on 27 February 2020 that Muslim pilgrims visiting Saudi Arabia would be temporarily stopped from visiting the country [4].

A plethora of research has been conducted on COVID-19 encompassing its origin, causes, symptoms, treatment and after effects but the focused domain of such research has been Western countries or some countries in the East like China and Iran that were more affected by the pandemic. For a wholesome reference material see [13].

Authors have found very little research material that focused on the MENA region vis-e-viz COVID-19. Most of the research that we come across is of medical nature but non has focused on the behavioural patterns emerging from the COVID-19 data. Hence the rationale for the present research emerges.

For the present study data were collected on five study variables (1) Confirmed cases, (2) Recovered cases, and (3) Death cases due to COVID-19 along with the minimum and maximum temperatures and whether there was lockdown or not. [11] has recommended that at least 50 observations in order to apply time series tools to a set of data. Many others have recommended at least 100 observations but in the current study, the day was collected for 220 days.

The present research is a continuation of other research work that has been done on COVID-19, but from a different perspective. A two-pronged analytical strategy was adopted to study COVID-19 data, firstly -exploratory analysis, and secondly time series analysis. In the exploratory analysis the pattern of confirmed cases, death cases, and recovery cases were plotted in order to observe the behavioural pattern of each study variable during the period under review (*15th March 2020-20th October 2020*). In time series analysis two most popular prediction techniques were carried out ARIMA and ARCH.

1.1. FORMAT OF THE PAPER

The rest of the paper proceeds as follows: In Section 2 methods and materials and statistical framework of Exploratory analysis, ARIMA and ARCH techniques are elaborated: results and discussion are presented in Section 3; Section 4 and 5 briefly conclude the study with some future implications followed by acknowledgments.

2. METHODOLOGY

2.1. Data Analysis and Interpretation

The data for the current research was collected from [14] and covers the period from *15 March 2020 to 22 October 2020* a total of 222 days. Time Series data for the three study variables were analyzed using R software coupled with E-view version 11.0

2.2. Exploratory Analysis: In the exploratory analysis only the visual representation of three study variables is discussed coupled with some emerging patterns and is depicted in Figure1. Panel (a) of Figure 1 shows the original time series graphs for the three study variables. The first thing that came to the limelight is that COVID-19's first cycle is about 220 days starting from somewhere in the middle of March 2020 and the confirmed cases reached their highest on June 16, 2020 (4507) subsided to a minimal level somewhere in the middle of October 2020. Since the death cases cannot be read accurately in panel (a) therefore, original variables were transformed using Natural logarithm transformation to get a clearer view of the data dynamics.

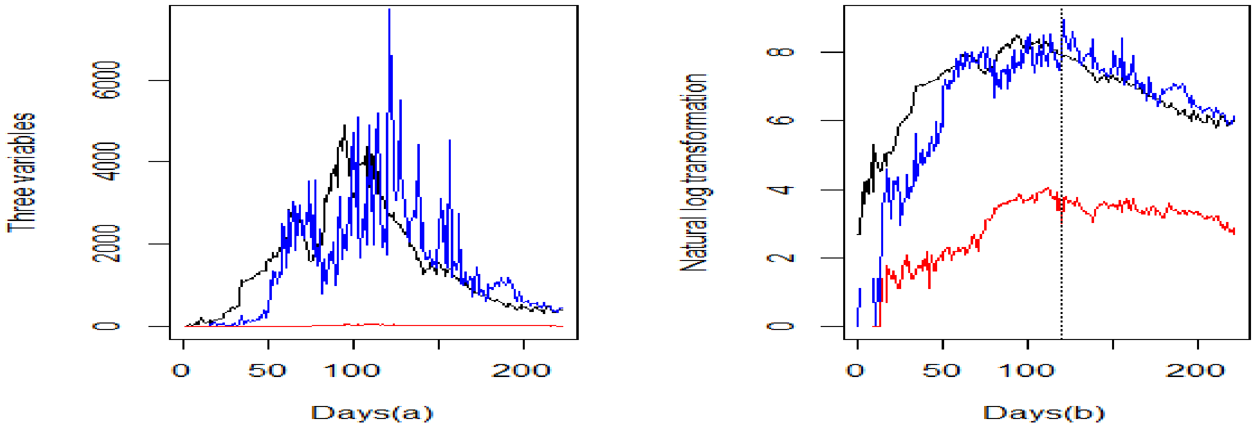

Figure 1: Panel(a) Simple Line Plots of the three variables and panel(b) shows the Natural Log Transformation of the three study variables.(*Black=Confirmed cases, Blue= Recovered cases, and Red= Death cases*)

Panel (b) of Figure 1 shows that though the confirmed cases and the recovered cases grew exponentially throughout March, April, May, and June but on July 6, 2020, the recovered cases magically surpassed the confirmed cases and till the last week of October the recovered cases were more than the confirmed cases. Death cases also rose till the end of June but started subsidizing from the first week of July. As compared to recovered cases the decline rate of death cases is quite slow.

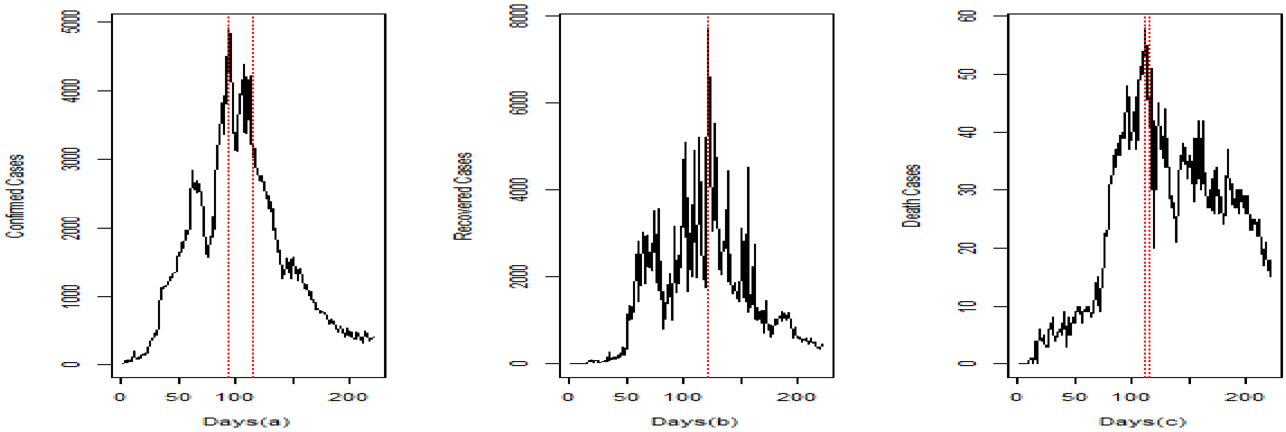

Figure 2: Panel(a) Confirmed Cases Panel(b) Recovered Cases Panel(c) Death Cases

Figure 2 with three panels shows the time series of three study variables individually. Red vertical lines in panel (a) in Figure 2 depict the day on which maximum confirmed cases were reported (June 16, 2020) and the day when the confirmed cases started subsidizing (July 6, 2020). Red vertical lines in panel (b) in Figure 2 depict the day on which maximum recovered cases were reported (July 13, 2020). Red vertical lines in panel(c) in Figure 2 depict the day on which maximum death cases were reported (July 04, 2020) and the day when the death cases started subsidizing (July 10, 2020). Therefore, in a cycle of 220 days of COVID-19, the most critical days ranged from 94th day (June 16) to 104th day (June 26).

2.3. Time Series Analysis(ARIMA): An ARIMA model is a statistical analysis used to understand time series data and to predict future values. The model uses autoregression and moving average terms to predict future values based on past data and past error terms. Since the data COVID-19 are recorded chronologically therefore it is called time series data and an appropriate statistical tool used for analyzing such a series data is Time series analysis. Forecasting involves making estimates of the future values of variables of interest using past and current information. The most common classes of time series forecasting methods are the autoregressive integrated moving average (ARIMA) models [5]. The conceptual framework and the data have been adopted by [6] where the authors have used the ARIMA model for the prediction of stock prices for the Flying cement industry. For ARIMA modeling and all relevant graphs and tables, an online statistical software [8] was used.

2.3.1. ARIMA for Confirmed Cases:

An ARIMA time series analysis was conducted to forecast future values over time for Confirmed Cases. The assumption of stationarity was assessed by creating a line plot for the scale variable, Confirmed Cases, after differencing over time [7]. For the assumption of stationarity to be met, there shouldn't be any trends (line plot is increasing/decreasing over time), or seasonality (changes that occur at a set interval) [7]. Strong trends or seasonal data can cause the results of the ARIMA model to be unreliable. Figure 3 presents the line plot over time for Date.

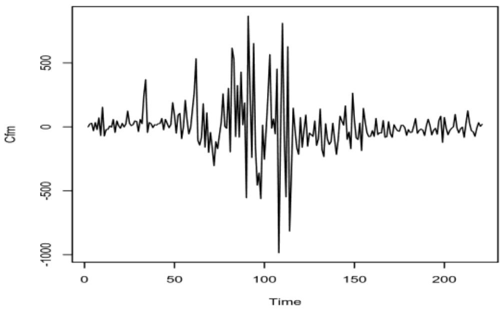

Figure 3: Line plot of Confirmed cases against time to assess stationarity

Autocorrelation plot function (ACF). The ACF plot was examined to determine the q (MA order) and seasonal P (SAR) parameters for the ARIMA model [7]. The plot should exhibit strong correlations until the lag point, where the correlation should decline rapidly. This lag point suggests the optimal value for the q parameter. If the autocorrelation value has a positive spike at each point of the seasonal period, then this suggests including the P parameter in the model (P > 1 is rare and can cause feedback problems in the model). On the other hand, if there are negative autocorrelation values that spike at the beginning of each season, then this indicates including an SMA term to the model (Q). The [7]algorithm indicates optimal values of q = 3 and P = 0. The ACF plot is presented in Figure 4.

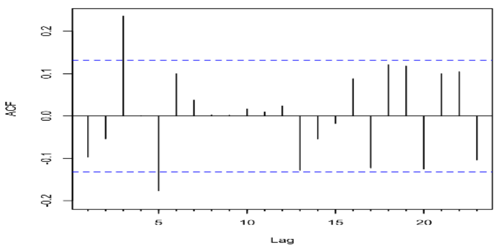

Figure 4 : ACF plot of Confirmed cases

Partial ACF (PACF). The PACF plot was conducted to examine the p (AR order) and Q (SMA) parameters for the ARIMA model [7] . The plot should have a strong correlation until the lag point, where the correlation will drop off. This lag point will indicate the optimal value for the p parameter. If the partial autocorrelation value has spikes for each seasonal period, then this suggests including the Q parameter in the model (Q > 1 is rare and can cause feedback problems in the model). The suggested value from the [7] algorithm is p = 2 and Q = 0. Figure 5 shows the PACF plot.

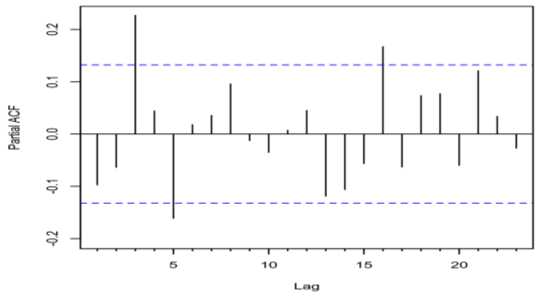

Figure 5:PACF plot of Confirmed cases

Residuals. The residuals were examined by plotting a line plot of the model residuals over time and by plotting an autocorrelation function (ACF) plot [7]. According to [7], the model residuals should not be correlated and should have a mean of zero. Forecasts with correlated residual terms signify that there is information being left out of the model that should be captured, and if the residuals do not have a mean of zero then the model results have a bias. The correlations of the residuals can be seen in the ACF plot of the residuals. This plot should have small correlations for each time point if there is no correlation among the residual terms. The mean of the residuals can be examined by looking at the line plot of the residuals. If the residuals are randomly distributed around zero, then the mean is roughly zero for the residuals. [7] also adds that it is useful if the residuals are normally distributed and have a constant variance. The normality of the residuals can be determined from a histogram of the residuals, where the residuals should follow a bell-shaped curve if the residuals are normal. The variance can be checked with the line plot of the residuals. The variance between the points should remain practically constant over time for the assumption to be met. Figure 6 presents the residuals line plot over time. The ACF plot of the results is shown in Figure 7. Figure 8 shows the histogram of the model residuals with a normal distribution overlaying the plot for comparison.

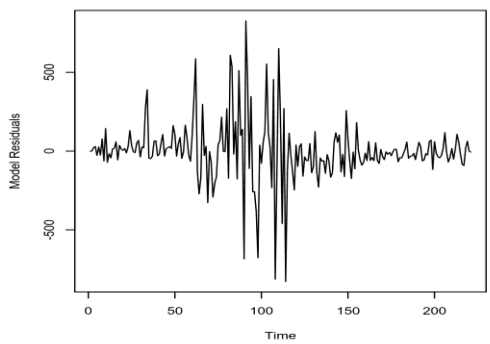

Figure 6:Plot of model residuals to assess model validity

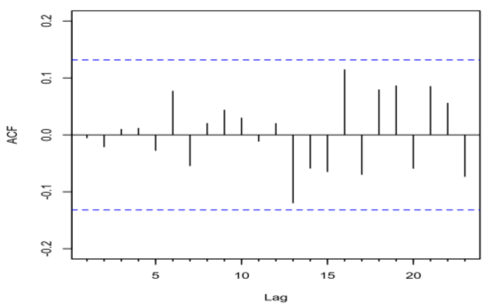

Figure 7:ACF plot of the residuals to examine model fit

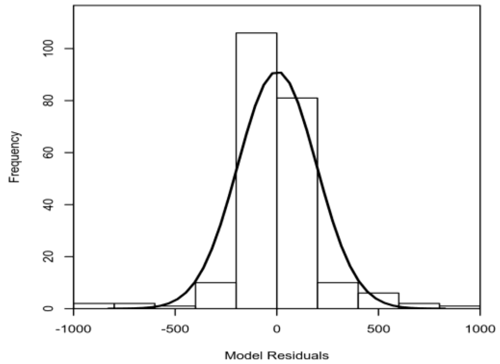

Figure 8: Histogram of model residuals with a normal distribution overlayed

*Residuals***.** The residuals were examined by conducting a Ljung-Box test. The residuals of the time series model should be randomly distributed with now trends or patterns. The results of the Ljung-Box test were not significant based on an alpha value of 0.05, $\chi^2 = 3.10$, $p = .685$. This result suggests that the residuals do not contain trends or patterns, and the model is reliable.

Forecast. The forecast for the ARIMA model was compared to the observed data starting at the time 2020-10-22 to check the validity of the model. The results comparing the predictions against the observed data can be seen in Table 1. Figure9 shows 30 days ahead of forecasted values.

Table 1:ARIMA Model Accuracy Comparing Observed and Forecasted Values.

ME	RMSE	MAE	MPE	MAPE
-9.79	9.79	9.79	-2.44	2.44

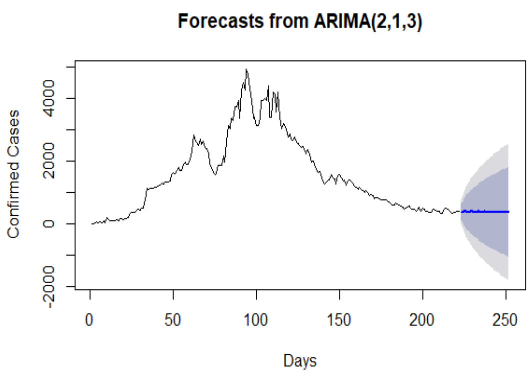

Figure 9: Forecasted values for 30 days ahead

2.3.2. ARIMA for Recovered Cases:

An ARIMA time series analysis was conducted to forecast future values over time for Recovered cases. The assumption of stationarity was assessed by creating a line plot for the scale variable, Rec, after differencing over time[7]. For the assumption of stationarity to be met, there shouldn't be any trends (line plot is increasing/decreasing over time), or seasonality (changes that occur at a set interval) [7]. Strong trends or seasonal data can cause the results of the ARIMA model to be unreliable. Figure 10 presents the line plot over time for Date.

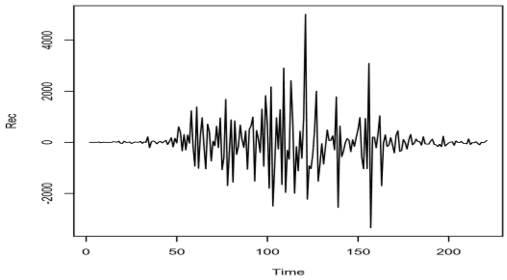

Figure 10: Line plot of Recovery against time to assess stationarity

Autocorrelation plot function (ACF). The ACF plot was examined to determine the *q* (MA order) and seasonal *P* (SAR) parameters for the ARIMA model [7]. The plot should exhibit strong correlations until the lag point, where the correlation should decline rapidly. This lag point suggests the optimal value for the *q* parameter. If the autocorrelation value has a positive spike at each point of the seasonal period, then this suggests including the *P* parameter in the model ($P > 1$ is rare and can cause feedback problems in the model). On the other hand, if there are negative autocorrelation values that spike at the beginning of each season, then this indicates including an SMA term to the model (*Q*). The [7] algorithm indicates optimal values of $q = 2$ and $P = 0$. The ACF plot is presented in Figure 11.

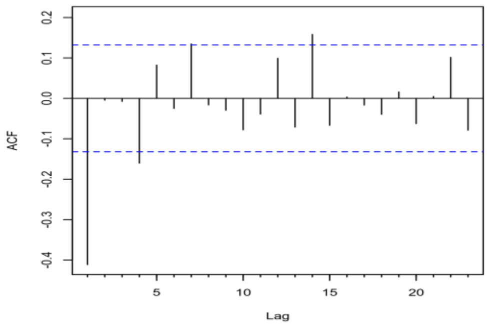

Figure 11: ACF of Recovery to determine model validity

Partial ACF (PACF). The PACF plot was conducted to examine the *p* (AR order) and *Q* (SMA) parameters for the ARIMA model [7]. The plot should have a strong correlation until the lag point, where the correlation will drop off. This lag point will indicate the optimal value for the *p* parameter. If the partial autocorrelation value has spikes for each seasonal period, then this suggests including the *Q* parameter in the model ($Q > 1$ is rare and can cause feedback problems in the model). The suggested value from the [7]algorithm is $p = 0$ and $Q = 0$. Figure 12 shows the PACF plot.

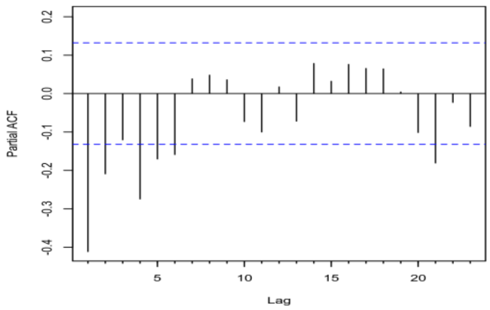

Figure 12: PACF plot of Recovery cases to determine model validity

Residuals. The residuals were examined by plotting a line plot of the model residuals over time and by plotting an autocorrelation function (ACF) plot [7] According to [7], the model residuals should not be correlated and should have a mean of zero. Forecasts with correlated residual terms signify that there is information being left out of the model that should be captured, and if the residuals do not have a mean of zero then the model results have a bias. The correlations of the residuals can be seen in the ACF plot of the residuals. This plot should have small correlations for each time point if there is no correlation among the residual terms. The mean of the residuals can be examined by looking at the line plot of the residuals. If the residuals are randomly distributed around zero, then the mean is roughly zero for the residuals. [7] also adds that it is useful if the residuals are normally distributed and have a constant variance. The normality of the residuals can be determined from a histogram of the residuals, where the residuals should follow a bell-shaped curve if the residuals are normal. The variance can be checked with the line plot of the residuals. The variance between the points should remain practically constant over time for the assumption to be met. Figure 13 presents the residuals line plot over time. The ACF plot is shown in Figure 14. Figure 15 shows the histogram of the model residuals with a normal distribution overlaying the plot for comparison.

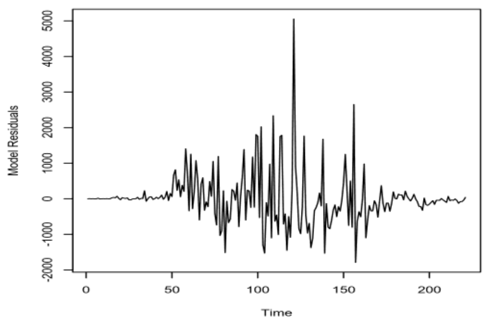

Figure 13: Plot of model residuals to assess model validity

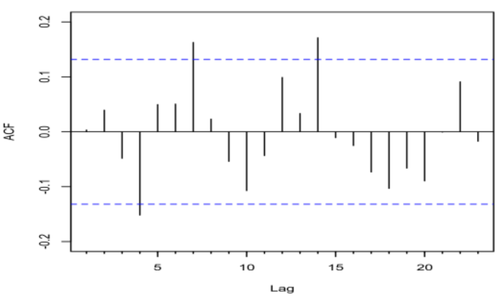

Figure 14: ACF plot of the residuals to examine model fit

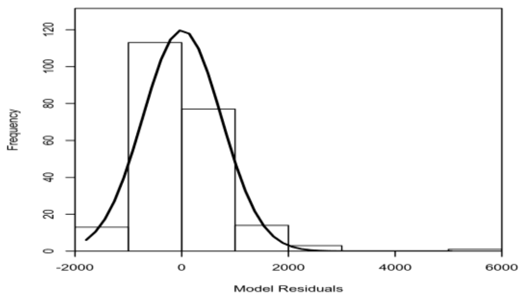

Figure 15: Histogram of model residuals with a normal distribution overlayed

Residuals. The residuals were examined by conducting a Ljung-Box test. The residuals of the time series model should be randomly distributed with now trends or patterns. The results of the Ljung-Box test were significant based on an alpha value of 0.05, $\chi^2 = 16.85$, $p = .032$. This result suggests that the residuals contain trends or patterns, and the model is may be unreliable.

Forecast. The forecast for the ARIMA model was compared to the observed data starting at the time 2020-10-22 to check the validity of the model. The results comparing the predictions against the observed data can be seen in Table 2. Figure 16 shows the observed and predicted values for the time series analysis.

Table 2:ARIMA Model Accuracy Comparing Observed and Forecasted Values.

ME	RMSE	MAE	MPE	MAPE
40.48	40.48	40.48	8.69	8.69

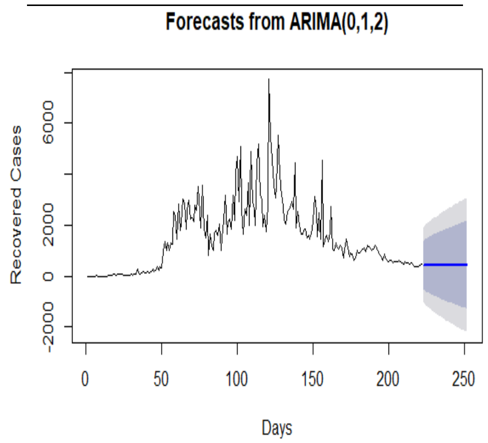

Figure 16: Forecasted values for Recovery cases

2.3.3. ARIMA for Death cases:

An ARIMA time series analysis was conducted to forecast future values over time for Death. The assumption of stationarity was assessed by creating a line plot for the scale variable, Death, after differencing over time [7]. For the assumption of stationarity to be met, there shouldn't be any trends (line plot is increasing/decreasing over time), or seasonality (changes that occur at a set interval) [7] Strong trends or the seasonal data can cause the results of the ARIMA model to be unreliable. Figure 17 presents the line plot over time for Date.

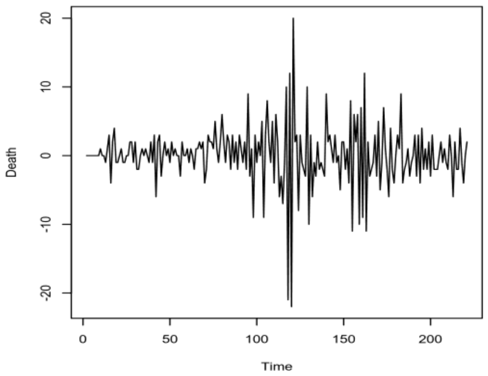

Figure 17: Line plot of Death against time to assess stationarity

Autocorrelation plot function (ACF). The ACF plot was examined to determine the *q* (MA order) and seasonal *p* (SAR) parameters for the ARIMA model [7]. The plot should exhibit strong correlations until the lag point, where the correlation should decline rapidly. This lag point suggests the optimal value for the *q* parameter. If the autocorrelation value has a positive spike at each point of the seasonal period, then this suggests including the *P* parameter in the model ($P > 1$ is rare and can cause feedback problems in the model). On the other hand, if there are negative autocorrelation values that spike at the beginning of each season, then this indicates including an SMA term to the model (*Q*). The [7]algorithm indicates optimal values of $q = 0$ and $P = 0$. The ACF plot is presented in Figure 18.

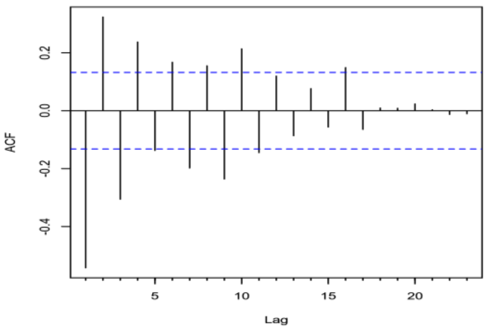

Figure 18: ACF plot of Death to determine model validity

Partial ACF (PACF). The PACF plot was conducted to examine the *p* (AR order) and *Q* (SMA) parameters for the ARIMA model [7]. The plot should have a strong correlation until the lag point, where the correlation will drop off. This lag point will indicate the optimal value for the *p* parameter. If the partial autocorrelation value has spikes for each seasonal period, then this suggests including the *Q* parameter in the model ($Q > 1$ is rare and can cause feedback problems in the model). The suggested value from the [7] algorithm is $p = 1$ and $Q = 0$. Figure 19 shows the PACF plot.

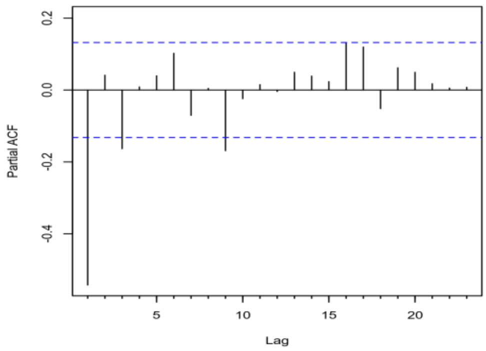

Figure 19: PACF plot of Death to determine model validity

Residuals. The residuals were examined by plotting a line plot of the model residuals over time and by plotting an autocorrelation function (ACF) plot [7]. According to [7] the model residuals should not be correlated and should have a mean of zero. Forecasts with correlated residual terms signify that there is information being left out of the model that should be captured, and if the residuals do not have a mean of zero then the model results have a bias. The correlations of the residuals can be seen in the ACF plot of the residuals. This plot should have small correlations for each time point if there is no correlation among the residual terms. The mean of the residuals can be examined by looking at the line plot of the residuals. If the residuals are randomly distributed around zero, then the mean is roughly zero for the residuals. [7] also adds that it is useful if the residuals are normally distributed and have a constant variance. The normality of the residuals can be determined from a histogram of the residuals, where the residuals should follow a bell-shaped curve if the residuals are normal. The variance can be checked with the line plot of the residuals. The variance between the points should remain practically constant over time for the assumption to be met. Figure 20 presents the residuals line plot over time. The ACF plot is shown in Figure 21. Figure 22 shows the histogram of the model residuals with a normal distribution overlaying the plot for comparison.

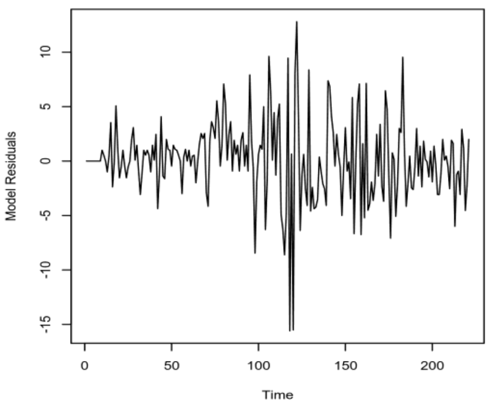

Figure 20: Plot of model residuals to assess model validity

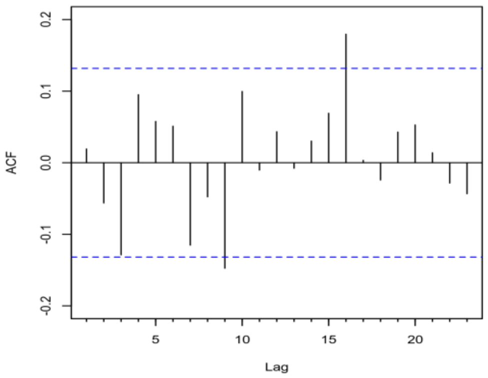

Figure 21: ACF plot of the residuals to examine model fit

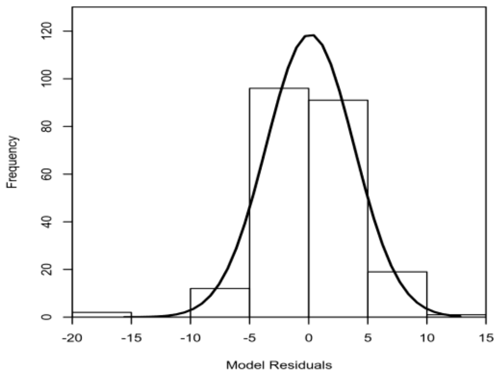

Figure 22: Histogram of model residuals with a normal distribution overlayed

Residuals. The residuals were examined by conducting a Ljung-Box test. The residuals of the time series model should be randomly distributed with now trends or patterns. The results of the Ljung-Box test were significant based on an alpha value of 0.05, $\chi^2 = 18.91$, $p = .026$. This result suggests that the residuals contain trends or patterns, and the model is may be unreliable.

Forecast. The forecast for the ARIMA model was compared to the observed data starting at the time 2020-10-22 to check the validity of the model. The results comparing the predictions against the observed data can be seen in Table 3. Figure 23 shows 30 days forecasted values.

Table 3:ARIMA Model Accuracy Comparing Observed and Forecasted Values.

ME	RMSE	MAE	MPE	MAPE
-1.92	1.92	1.92	-12.80	12.80

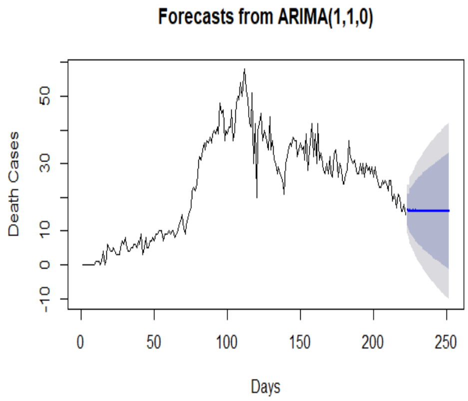

Figure 23: Forecasted values for Death cases

2.3.4. Forecast for the Study Variables: 30 days forecast for the three study variables are shown in Table 4

Table 4: 30 Days Forecast for the Study variables

	Confirmed Cases	Recovered Cases	Death Cases
1	383	437	16
2	379	433	16
3	390	433	16
4	391	433	16
5	384	433	16
6	383	433	16
7	389	433	16
8	388	433	16
9	385	433	16
10	388	433	16
11	387	433	16
12	385	433	16
13	386	433	16
14	384	433	16
15	387	433	16
16	386	433	16
17	387	433	16
18	387	433	16
19	386	433	16
20	387	433	16
21	387	433	16
22	386	433	16
23	386	433	16
24	387	433	16
25	387	433	16
26	386	433	16
27	387	433	16
28	386	433	16
29	386	433	16
30	386	433	16

2.4. ARCH Model: Econometrics describes a statistic model for time series data that describes the variance in present error term or innovation according to the current size of the preceding error period terms [9]Autoregressive Conditional Heteroscedasticity (ARCH) model. ARCH models are commonly used in the modeling of financial time series that exhibit time-varying volatility and volatility clustering, i.e., swing periods interwoven with periods of relative calm. ARCH-type models are sometimes considered to be in the stochastic volatility model family, but this is quite wrong since volatility (deterministic) is sometimes completely pre-determined with previous values.

Stationarity testing is necessary because it is a strong fundamental underlying assumption that must be checked on every time series data before going further in the application of the time series. The unit root testing must be applied to test whether a time series variable is non-stationary and has a unit root [10]. In general, the null hypothesis is defined as the presence of a unit root (meaning not stationary) and there is no unit root (series are stationary) in the alternative hypothesis

Table 5:Unit Root Test (Stationarity Test)

Variable	*T-statistic*	*F-statistic*	*Sig.*
Confirmed Cases	-6.8603	88.34	0.000
Death Cases	-26.2553	689.34	0.000
Recovered Cases	-10.5589	115.83	0.000

p < 0.01 (1%), significant level

The test of stationarity (*Unit Root test*) as revealed above using Augmented Dickey Fuller test (ADF) in Table 5 shows that the p-values for daily confirmed cases, daily death cases, and recovered cases of covid-19 respectively are less than or equal to a significant level (that is p-value ≤ α) and that implies that we reject the null hypothesis which says there is a presence of unit root and then concludes that there is no unit root which implies that the three above mentioned variables are stationary, hence further analysis can be applied to the series. This also suggests that a useful forecast can be applied accordingly.

The ARCH (I) MODEL can be expressed as $h_t = b_0 + b_1 U_{t-1}^2$. The null hypothesis states that the effect of ARCH is not statistically significant, while the alternative hypothesis states that the effect of ARCH is statistically significant. The mean equation was worked out using *EVIEWS* and stated as follows:

$$h_t = 78443.83 + 1.075341 U_{t-1}^2 \quad (1)$$

3. RESULTS

3.1 Exploratory Analysis: Combined view of the plots shown in Figure 1 and Figure 2 is that dates between July 6 and July 12 are pivotal and can very rightly be captioned as a "***MIRACLE WEEK***" in this particular week the tables were turned. Though the cycle of COVID-19 was approximately 7 months a decrease in the number of confirmed cases and an increase in recovered cases resulted in the "***MIRACLE WEEK".*** But the improvement in the situation is largely due to the initiation and drastic execution of the safety protocols by the Saudi government to contain the COVID-19 pandemic.

3.2 ARIMA Results

3.2.1 Confirmed Cases: The optimal model found using the [7] algorithm was ARIMA(2, 1, 3), which was used for the results. The AR(2) coefficient was significant, $p < .001$, indicating a lag of 2 for the autoregressive term fits the data well. The MA(2) coefficient was significant, $p = .055$, indicating a lag of 3 for the moving average term does not fit the data well. The estimated variance (volatility/white noise/σ^2) in the model was 38260.76. The table of coefficients for the ARIMA model is presented in Table 6.

Table 6:Coefficient Table for the ARIMA model.

Variable	*B*	*SE*	*t*	*p*
AR1	-0.10	0.14	-0.70	.485
AR2	-0.69	0.11	-6.30	< .001
MA1	0.00	0.15	0.03	.975
MA2	0.74	0.13	5.64	< .001
MA3	0.17	0.09	1.93	.055

Note: df = 216;

3.2.2 Recovered Cases: The optimal model found using the [7] algorithm was ARIMA(0, 1, 2), which was used for the results. The MA(0) coefficient was significant, $p = .101$, indicating a lag of 2 for the moving average term does not fit the data well. The estimated variance (volatility/white noise/σ^2) in the model was 546530.46. The table of coefficients for the ARIMA model is presented in Table 7.

Table 7:Coefficient Table for the ARIMA model.

Variable	*B*	*SE*	*t*	*p*
MA1	-0.62	0.06	-9.55	< .001
MA2	-0.10	0.06	-1.64	.101

Note: df = 219;

3.2.3 Death Cases: The optimal model found using the [7] algorithm was ARIMA(1, 1, 0), which was used for the results. The AR(1) coefficient was significant, $p < .001$, indicating a lag of 1 for the autoregressive term fits the data well. The estimated variance (volatility/white noise/σ^2) in the model was 13.88. The table of coefficients for the ARIMA model is presented in Table 8.

Table 8: Coefficient Table for the ARIMA model.

Variable	*B*	*SE*	*t*	*p*
AR1	-0.54	0.06	-9.57	< .001

Note: df = 220;

3.3. ARCH Model: Estimated ARCH (1) model reveals that the model is statistically significant as the p-value of the residual is 0.0000 which is less than the significant level of 5% which implies a rejection of the null hypothesis and accepting the alternative hypothesis of the significance of ARCH(1) MODEL and is useful for future forecasting. The static forecast of confirmed-cases of COVID-19 Figure 24 shows that the confirmed cases of covid-19 fall between the two red lines which is the confidence interval. The lower panel of Figure 24 is the forecast of the variance and the pattern shown that COVID-19 will drastically fall and may end. The dynamic forecast of the confirmed cases of COVID-19 as depicted in Figure 25 shows that the blue forecasted line falls between the two-confidence interval (red lines) and as we can see above that it is constant over time. This is evident that Saudi Arabia kingdom has nothing to fear as regards the future occurrence of confirmed cases of covid-19 provided all protective safety protocols are strictly adhered to.

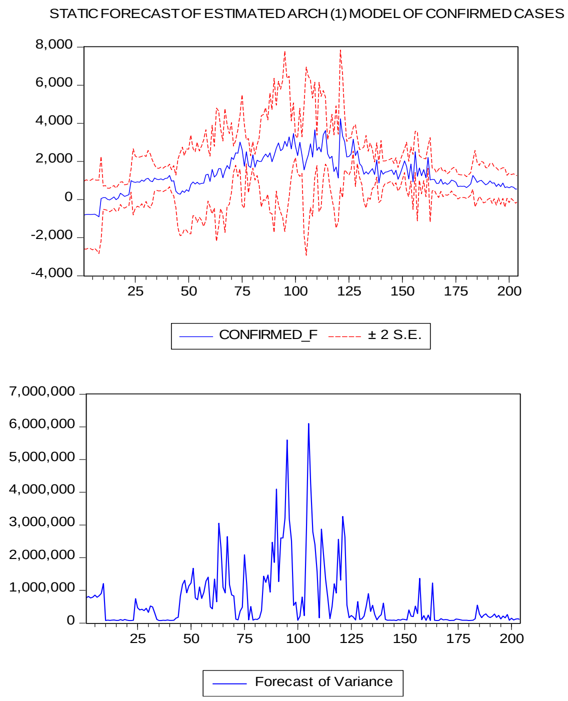

Figure 24: Staticforecast of Estimated ARCH(1) Confirmed Cases Forecast of Variance

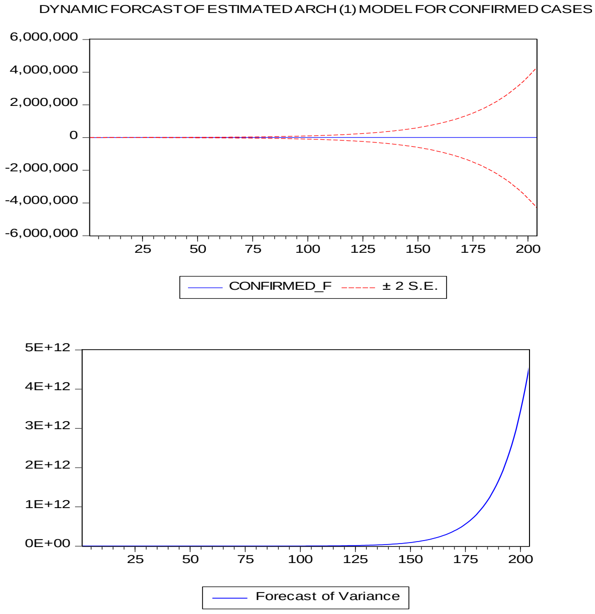

Figure 25: Dynamic Forecast of Estimated ARCH(1) with Forecast of Variance

4. CONCLUSION

The present study focused on the analysis of COVID-19 data collected by the students of the first semester of Executive Master's program, Department of Health Services and Hospital, Faculty of Economics and Administration, King Abdulaziz University, Jeddah as part of their statistical project. Primary three study variables were selected for analysis (Confirmed Cases, Recovered Cases, and Death Cases). The study had a two-pronged analytical strategy (a) exploratory analysis (b) time series analysis. The results of the exploratory analysis revealed very informative behaviour of COVID-19 in Saudi Arabia. The results of the time series analysis both ARIMA and ARCH(1) depicted suitable modelling for the three study variables and forecasts for 30 future days were also produced.

The virus has been well contained in Saudi Arabia and this is also supported by the future forecast of the cases of the COVID-19 pandemic. Based on the values of the forecasts, it is important to note that the government should not relent in constantly reminding the people that COVID-19 is still a global surge and hence the need to always bear in mind all the safety protocols like washing hands regularly with soap and water, using face masks, avoiding crowded places as well as keeping a social distance (i.e. strictly executing safety protocols).

5. LIMITATIONS AND FUTURE IMPLICATIONS

1. Due to paucity of time data was collected by the students for a specific period ranging from 15th March 2020 to 20th October 2020. But for an in-depth apprehension of the problem more data should be collected in the future.
2. Time Series techniques of Cointegration among Confirmed cases and Death cases be carried out to study the short term and long term relationship between the study variables.

ACKNOWLEDGEMENT: The authors would like to extend their gratitude to ***Dr. Mohammed K. Al-Hanawi,*** Head of Health Services and Hospital Administration Department, Faculty of Economics & Business Administration, King Abdulaziz University for his able guidance and encouragement throughout the completion of the study and also for providing the necessary wherewithal and conducive environment for carrying out the present study.

CONFLICT OF INTEREST: There exists no conflict of interest among the authors of the present study.

REFERENCES

[1] Reynold, Matt (2020). " What is Coronavirus and how close is it to becoming a pandemic?". ISSN 1357-0978. Archived retrieved 5 March 2020.
[2] Ministry of Health Saudi Arabia(2020). "MOH Reports First Case of Coronavirus Infection", https://www.moh.gov.sa/en/Ministry/MediaCenter/News/Pages/News-2020-03-02-002.aspx.
[3] Amlot, Matthew (2020). "Jobs, health priority in coronavirus spending cuts, VAT hike" : Saudi finance minister.
[4] Abueish, Tamara (2020). "Coronavirus: Mosques in Saudi Arabia". Mecca Set to re-open on June 21.
[5] Saud M. Al-Fattah (2006), Time Series Modeling for U.S. Natural Gas Forecasting, E-Journal of Petroleum Management and Economics, ISSN: 1718 – 7559.
[6] Abbasi, U., Almarashi, A., Khan, K. and Hanif, S (2017) Forecasting cement stock prices using ARIMA model: A case study of flying cement, *Sci. Int.*, 29(5), 1039-1043.
[7] Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts.
[8] Intellectus Statistics [Online computer software]. (2020). Intellectus Statistics. https://analyze.intellectusstatistics.com/
[9] Hacker, R.S ; Hatemi J, A (2003). " A test for multivariate ARCH effects". Applied Economic letters. PP 12(7). 411-417
[10] Dickey, D.A ; Fuller, W.A (1979). "Distribution of the estimators for autoregressive time series with a unit root".
[11] World Health Organization "Timeline of WHOs response to COVID-19". https://www.who.int/news/item/29-06-2020-covidtimeline
[12] Chatfield, C. (1996). The analysis of time series, 5th ed, Chapman & Hall, New York, NY
[13] Council of the European Union "Reading References Council Library". Rue de la Loi/Wetstraat 175 - B-1048 Bruxelles/Brussel - Belgique/België http://www.consilium.europa.eu/en/library-blog/-
[14] Weqaya (2020) "Daily Updates". Saudi Center for Disease Prevention and Control. https://covid19.cdc.gov.sa/daily-updates/