

On Statistical Analysis of Forecasting COVID-19 for the Upcoming Months in the Kingdom of Saudi Arabia

Bachioua Lahcene*

Department of Basic Sciences Prep, Year, P.O. Box 2440, University of Ha'il, Ha'il, Saudi Arabia

Abstract: This paper presents a statistical analysis using fitted prediction models that revealed a high exponential growth in the number of confirmed cases, deaths, and treated case processes based on our model predictions and the results of experimental COVID-19 predictions. The studies aimed to build inductive statistical models using the automatic integrated mean regression model methodology, and its preferred method for tracking data that represent the spread of the epidemic and then effectively predicting its numbers over the next six months, in addition to the number of deaths and cases that responded to recovery treatment using ARIMA.

Moreover, the number of infected cases per day is expected to stabilize less than 500, daily deaths are less than 15, and this situation will continue until the largest number of people are vaccinated in order to obtain herd immunity, and control the causes of the spread of the epidemic such as human gatherings and friction. Among individuals, in addition to obtaining the appropriate vaccine in the future, especially since the Kingdom of Saudi Arabia is waiting for this year's pilgrims from inside and outside the Kingdom, the results of this work will be useful for practitioners in various fields of theoretical and applied sciences.

Keywords: Time Series, Traffic Accidents, Automatic Integrated Average Regression Model, Forecasting, Best Practices, COVID-19, Kingdom of Saudi Arabia.

1. INTRODUCTION

Today, many complex random phenomena and problems are sweeping the world that causes many human and economic losses and causes and lags behind many negative social, economic and psychological phenomena, the most important of which are epidemics of all kinds, especially the COVID-19 epidemic, whose daily consequences have become frightening and terrifying.

Statistical science always and throughout history provide effective models for managing and tracking the direction of data, and perhaps the Corona pandemic today has summoned scientists in all disciplines allowance of efforts and cooperation among them in order to build models capable of making many efforts to combat the COVID-19, such as creating interactive information panels and analyzes of epidemiological models and suggesting the best compounds to help access treatments for the virus [17].

One of the fundamental differences between the epidemics that previously attacked the world, such as the Spanish flu a hundred years ago, and what the world is facing today since the beginning of the year 2020 with the emerging COVID-19, which is accelerating frighteningly, is the amount of huge data that flows from official reports of each country, and

scientific studies related to virology and epidemics that include the family of this series of COVID-19, and the viruses that emanate from them [24].

Statistics, mathematics, and knowledge related to nature and how it is applied constitute hyper-thinking and data analysis, simulating the human mind and its ways of working, such as its ability to think, discover and benefit from past experiences, machine learning to collect, monitor and analyze data, and the system's ability to correctly interpret data, learn from it, and use knowledge. [3]. The field of data science is one of the most in-demand fields in the labor market in the world during the past five years, and according to the report of Glassdoor, it is the most sought-after job in 2018 in the US market, which is also the view of the annual LinkedIn report for 2017. In the mid of December, a viral infection called coronavirus disease 2019 (COVID-19) was initially identified in Wuhan City of China [18].

At present (as of March 13, 2021) the infected patient population worldwide is recorded as 119,341,245 (119.3 million) with 2,645,270 deaths and 94,949,328 recoveries, and it is expected that these statistics is increasing exponentially in the upcoming days [2]. In the Kingdom of Saudi Arabia, statistics indicate that the number of infected cases is (381,708), the number of deaths recorded due to this epidemic is (6,556), and as for the recovered cases, it amounted to (372,217) [23].

This study is important due to its subject matter and also in order to complete the studies that dealt with this

*Address correspondence to this author at the Department of Basic Sciences Prep, Year, P.O. Box 2440, University of Ha'il, Ha'il, Saudi Arabia; Tel: 00966538592706; E-mail: bachiouaia@hotmail.com

topic, whether at the national or Arab level, and because the problem is complex and its consequences are costly, and for this reason, diagnosing the causes and consequences through foresight allows leaders to set protection plans and avoid as many pests and health waste as possible, and reduce their economic consequences and social.

This study also seeks to enrich Arab libraries in the field of foreseeing health safety studies, managing the problem of the spread of the epidemic, and addressing some of the problems that arise from it by presenting the role that can be used to avoid what can be done through proactive action and leadership of future management in such sensitive and relevant topics priority in anticipating future expectations.

2. LITERATURE REVIEW

Time series modeling has been an important area of research, time series models have attracted researchers from various disciplines. Research in time series analysis is being conducted in two domains; theoretical and applied. The classical theoretical development in time series analysis is the discovery of ARIMA (p, d, q) models discussed by Box and Jenkins [7]. Granger and Joyeux [13], and Hosking [15] have extended the ARIMA (p, d, q) model to handle the long memory process which is known as Autoregressive Fractionally Integrated Moving Average [7]. The ARCH model provides an explicit link between the risk (conditional volatility) and the best forecast of a time series, and the GARCH model provides the best fitting, Bollerslev [6] are useful tools to model the variability or volatility in a time-series data [6]. Zakoian [10] has extended these models to power ARCH by Higgins and to threshold ARCH. Several other versions of the GARCH models are available to capture various situations [14].

The multivariate ARIMA models have been derived and discussed by Lutkepohl, H. (2005) and the multivariate GARCH models have been proposed and discussed by Engle (2002) and by Silvennoinen and Terasvirta (2009) [21].

The Bayesian time series models have emerged as a useful class of model and are used when the parameters are assumed random. The Bayesian time series models are discussed in detail by Barber *et al.* (2011) [5]. Xiao and Wu (2017) [17] have proposed some new tests for testing changes in volatility. Various authors have proposed various modifications of the above-discussed time series models.

3. STUDY METHODOLOGY

The time series method is considered one of the statistical methods that attracted the attention of those interested in the study of prediction, as it constitutes the period, and it is the main method for planning and forecasting. If the period is very short or equal to zero, then we do not need planning, but if the period is long, and planning must be considered very important, and in this area, we need to forecast so that we know when the accident will occur to take appropriate action towards it, in addition to the statistical tools that it requires [4].

Time series is an important statistical topic for tracking random behaviors and for explaining life accidents that occur during an identifiable and controllable time, which has been proven by some special models. The time series method is one of the most important methods used to predict and anticipate the paths of expected results according to successive times through which the future path is envisioned, as it has been classified within the modern scientific methods and methods which are used in such problems using time series analysis.

The science that depends on predicting future changes of the variable, depending on its behavior in the past, made the ARIMA method one of the effective advanced methods of forecasting and drawing predictive paths, and for this we target the problem with this methodology for the sake of foresight.

Forecasting has been a major area of study by researchers in statistics as well as in economics. Several advanced models have been developed from time to time to forecast the available time series. The current study aimed at proposing some modified models for modeling various data. A brief about various time series models is given next.

3.1. Univariate Models

The univariate models will be used to build a forecast model for an individual county. The study will start with building the Box–Jenkins ARIMA (p, d, q) model of the form [7]:

$$\phi(B^p)(1-B)^d(Y_t - \mathbf{x}'_t\boldsymbol{\beta}) = \theta(B^q)\varepsilon_t, \quad (1)$$

where "d" is an integer and

$$\begin{aligned} \phi(B^p) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p ; B^p Y_t = Y_{t-p} \\ \theta(B^q) &= 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q \\ (1-B)^d &= \sum_{j=0}^d (-1)^j \binom{d}{j} B^j \end{aligned}$$

The ARIMA (p, d, q) model has a short memory. The GARCH (m, k) model of Engle (1982) [11] is useful model to model the volatility of the data. The GARCH (m, k) model is a collection of two models; one for mean and one for variance. The GARCH (m, k) model is given as

$$\left. \begin{aligned} \Delta^d (Y_t - \mathbf{x}_t' \boldsymbol{\beta}) &= ARMA(p, q) + \varepsilon_t && \text{(Mean Eq.)} \\ \sigma_t^2 &= \gamma_0 + \sum_{i=1}^m \gamma_i \varepsilon_{t-i}^2 + \sum_{i=1}^k \delta_i \sigma_{t-i}^2 && \text{(Variance Eq.)} \end{aligned} \right\} (2)$$

3.2. Multivariate Models

The multivariate time series models are useful when information on several time series is available. Lütkepohl (2005) [16] as gives the multivariate autoregressive model, known as the VAR model

$$\mathbf{y}_t = \mathbf{A}\mathbf{y}_{t-1} + \mathbf{B}_0\mathbf{x}_t + \mathbf{u}_t, \tag{3}$$

Where

\mathbf{y}_t is a (k x 1) vector of endogenous variables

\mathbf{x}_t is a (M x 1) vector of exogenous variables

\mathbf{A} is a (k x kp) matrix of autoregressive coefficients.

Another multivariate model that will be fitted is the Multivariate GARCH model given by

Silvennoinen and Terasvirta (2009). The model is given as

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \boldsymbol{\varepsilon}_t \text{ with } \boldsymbol{\varepsilon}_t = \mathbf{H}^{1/2} \mathbf{v}_t, \tag{4}$$

Where $\mathbf{H}^{1/2}$ is the Chelsey factor of time varying coefficients, the VAR (p) are used to model several time series when all the series are stationary. Since the data under consideration may contain non-stationary series, the model is given as [13]:

$$\Delta \mathbf{y}_t = \mathbf{A}\mathbf{y}_{t-1} + \sum_{i=1}^p \boldsymbol{\Gamma}_i \Delta \mathbf{y}_{t-i} + \mathbf{u}_t. \tag{5}$$

These models are useful to the model of ARIMA (p, d, q) the co-integrated time series simultaneously. The

VARMA (p, q) is a relatively extended model that combines the effects of VAR (p) and multivariate MA (q) models. Lütkepohl (2005) describes the model as [16]:

$$\mathbf{y}_t = \mathbf{v} + \sum_{i=1}^p \mathbf{A}_i \mathbf{y}_{t-i} + \sum_{i=0}^q \mathbf{M}_i \mathbf{u}_{t-i}, \tag{6}$$

where

\mathbf{A}_i are autoregressive parameter matrices

\mathbf{M}_i are moving average parameter matrices.

The VARMA (p, q) models can also be fitted to differenced series giving rise to VARIMA (p, d, q) models. The VARMA and/or VARIMA (p, d, q) models are be fitted for joint modelling of several time series. In case of joint volatility, the Multivariate GARCH models will be fitted to accomplish the volatility in the data. The model selection will be done by using various criteria including AIC, SBC, HQC etc. The models will be fitted by using available softwares and languages like EViews 8, R and/or SAS etc.

4. DATA SOURCE

Data are based on the stock of the previous private data bank based on the data obtained from the publications of the Ministry of Health in the Kingdom of Saudi Arabia during the period 01/01/2020 until 01/04/2021 [23].

Due to the scarcity of available data, its limitations and the accelerated impact of the fluctuations experienced by the Kingdom, which caused severe changes that appear during the daily follow-up of the announced results, especially in the first six months of the previous year, the researcher was forced during the statistical treatment to use the cumulative data to predict the number of cases declared per month in The Kingdom of Saudi Arabia, and with this, after treatment, the results of the cumulative number of COVID-19 cases of confirmed cases, deaths, and recoveries appeared in an increasing exponential manner, which can be described as a generalized exponential model over time [3].

5. STATISTICAL MODEL USED FOR THE ANALYSIS

After tracking the data for 14 months and ensuring the requirements of the time series, the researcher chose the simple time series methods of the automatic regression integrated moving average model (ARIMA)

[8] for the prediction of the number of confirmed cases, the number of deaths and cases that have been cured, which are models that allow the deduction. Recoveries for the next month have a high degree of reliability [22].

The ARIMA model is distinguished according to what is known in such cases with higher synthesis and prediction accuracy than exponential smoothing, which requires additional variables to control the accuracy of operations because it captures seasonal and non-seasonal prediction trends as mentioned in the reference [10].

Given the limited amount of data available to the researcher from the time of the outbreak of this pandemic, the researcher proposes to adopt non-seasonal models to describe the pattern in which the phenomena targeted for the study change concerning time, by adjusting the hypothesis that the pattern of current cases will accelerate at any moment shortly (half a month at least), This is after performing all statistical analyzes using the R software and making equivalent calculations confirming the success of the model building operations, which are compatible with the ARIMA model [19].

According to specialized references, the ARIMA model consists of a combination of automatic regression (AR) and moving average (MA), so it provides a good prediction for short time-series data and fits in good consistency with the nature of the available data, and therefore it can provide high-rudder prediction periods.

Until mid-year 2021 by estimating the parameters of the adopted model, in order to increase the reliability ratio and evaluate the suitability of the model parameters, the parameters (p, d, q) are determined by the partial automatic correlation function (PACF); Autocorrelation function (ACF) and ARIMA results (p, d, q) are based on Information Standard (AIC) which is a good test for the fitness test, as this model with minimum AIC are the best, where p denotes is the index of the autoregressive parameter, D is the divergence ranking parameter indicator and q is the moving averages parameter indicator.

6. RESULTS AND DISCUSSION

The adaptation and prediction models used allowed the provision of a future analytical environment that saves the time and effort to do the necessary simulation and prediction to select the best-preferred practices for future adoption.

The Normal state has been achieved to track the epidemic infections through the natural spread of the epidemic after the community obtains mass immunity and life returns to its normal, and thus normal distribution models provide the best tool for exploring cases and reaching the total immunity of the community members, which has been studied through the use of normal distribution [1].

The results that came in the subject of the study confirmed many extrapolations and recommendations for security and health safety management, and therefore we recommend completing the study by introducing other variables that affect the construction of predictive models targeted to build models for managing the future of the Corona pandemic in the Kingdom of Saudi Arabia while studying the qualitative details according to age and type The spread of Corona and the results achieved by approved pharmaceutical vaccines to choose the best and localize their industry locally and make a vaccination map for two years which lives can be saved and societal immunity to such diseases, and the best ways to manage hospitals and human gatherings under the slogan of life with zero death due to the Coronavirus.

Since the Kingdom of Saudi Arabia is a rich country, it has sufficient medical facilities, which affects the stability of the health situation in the number of corona injuries in the second year, after achieving a more broad level in prevention and adherence to health laws and standards adopted by gatherings of individuals in the community. Public health officials and the government must make difficult decisions to preserve the positive results achieved in the number of COVID-19 infections, and the general public must maintain social distancing and use precautions to ensure their safety and control the disease from spreading further, and all cases of infection from the outbreak of the epidemic until the beginning of March 2021 Shown in Figure 1.

Given the health policies adopted by public health officials and the government, represented in making difficult decisions to maintain the positive results achieved, and increasing vaccination for all members of society, with priority given to the elderly and those suffering from diseases that have complications when contracting the COVID-19 epidemic, and people preserving social distancing and the use of precautions to ensure their safety and control of the disease from the spread of the number of deaths due to the Corona epidemic and its consequences decreased and

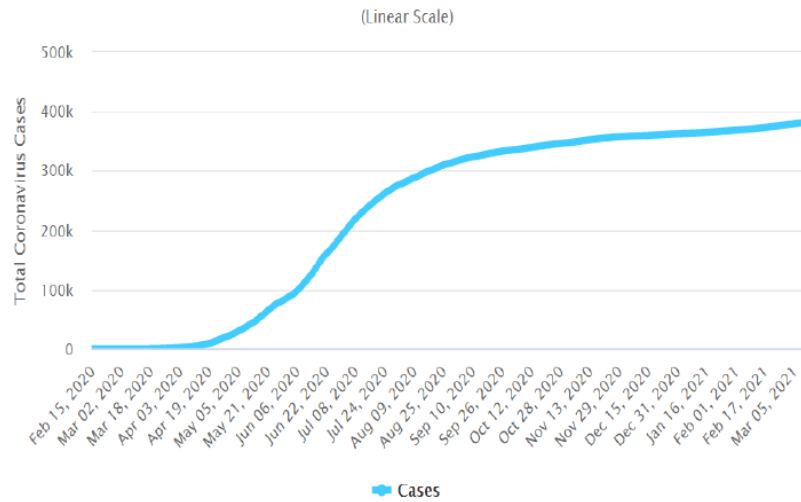


Figure 1: Total Coronavirus Cases in Saudi Arabia [https://covid19.moh.gov.sa/].

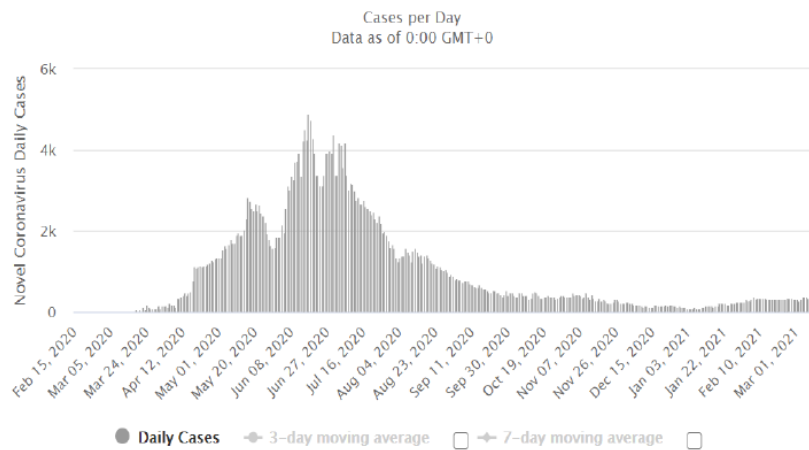


Figure 2: Daily New Cases in Saudi Arabia [https://covid19.moh.gov.sa/].

stabilized at an average number of fewer than 30 cases per day, and Figure 2 shows the numbers of the

number of deaths since the outbreak of the epidemic until the beginning of March 2021.

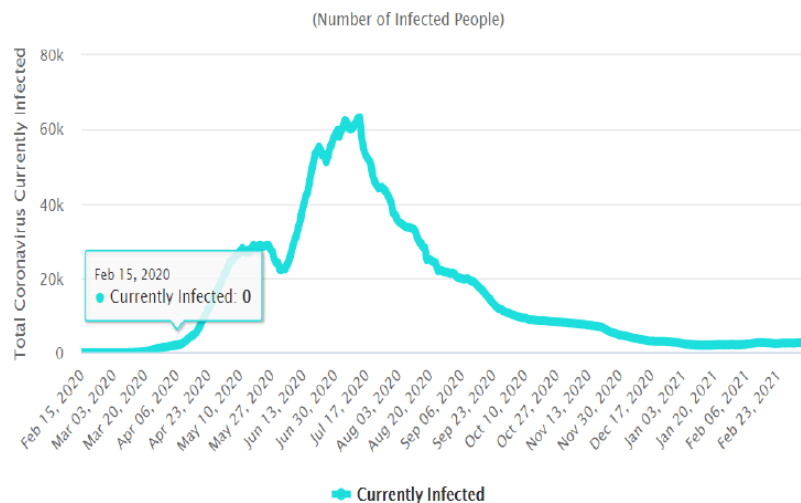


Figure 3: Active Cases in Saudi Arabia [https://covid19.moh.gov.sa/].

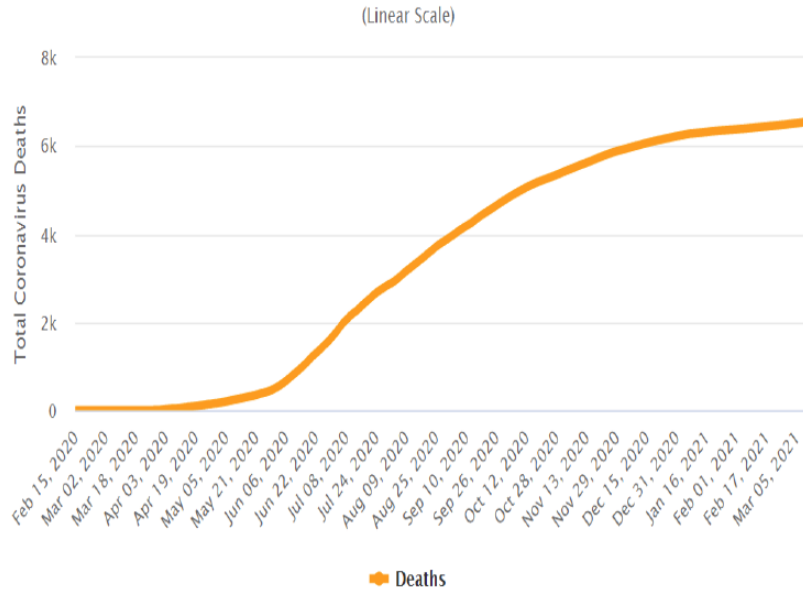


Figure 4: Total Coronavirus Deaths in Saudi Arabia [<https://covid19.moh.gov.sa/>].

The numbers of those recovering from infection with Corona disease have improved daily, and almost the number of recovery cases is close to the number of cases of infection, which has achieved a kind of positive stability and shows the achieved results of the preventive policies adopted since the spread of the epidemic, and all these results are shown in Figure 3.

The collective immunity of individuals improved, the incidence of Corona disease stabilized daily, and almost the number of cases was constant, approaching 400 per day, which achieved a kind of positive stability and demonstrated the results of the preventive policies adopted since the outbreak of the epidemic, and all these results are shown in Figure 4.

During 14 months, the cases of infections became stable and appear naturally, and in exceptional times in the form of mixtures distributions with a different level of modes, which prevents the stability of the condition for the normal distribution in one mode, these cases are shown in Figure 5, and in Figure 6 It shows the cumulative distributions of cases of the Coronavirus outbreak 14 months after the outbreak of the epidemic.

7. CONCLUSION AND RECOMMENDATIONS

We note that this study is not official to make short-term forecasts about confirmed cases of COVID-19 in addition to the number of deaths and related recoveries, but rather it is a study that relies on what

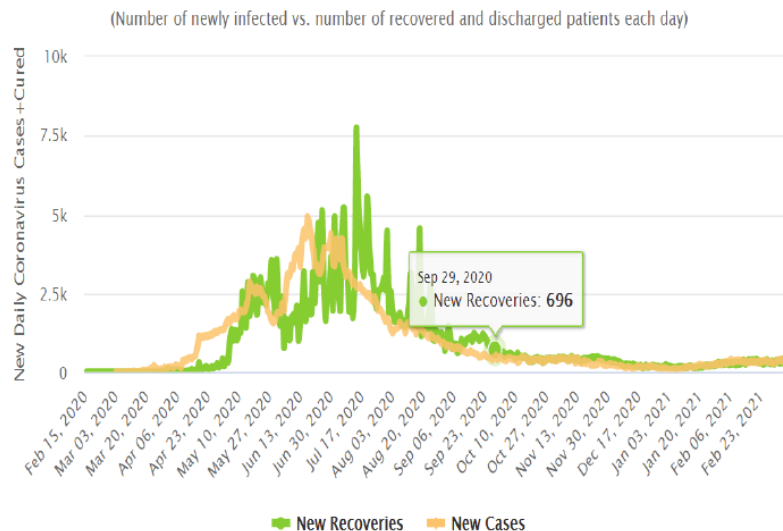


Figure 5: Newly Infected vs. Newly Recovered in Saudi Arabia [<https://covid19.moh.gov.sa/>].

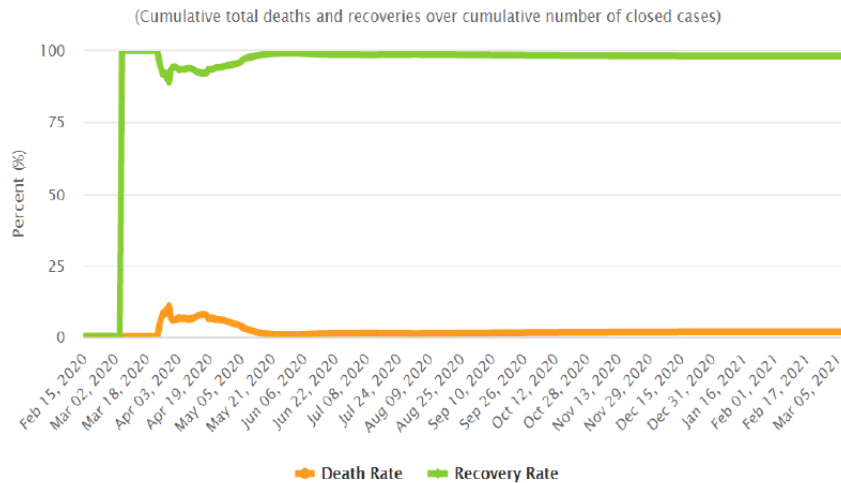


Figure 6: Outcome of Cases (Recovery or Death) in Saudi Arabia [<https://covid19.moh.gov.sa/>].

has been done before and is looking for predictions after a series of stability.

The results of this study revealed that there could be an inconsistent increase in the number of confirmed cases in the middle of each month starting from April, according to what is confirmed by the predictions of the current trend of the time series that represent the data of this pandemic if it continues with the same approved stability during the first quarter of this year. Currently, the statistics indicate that the number of deaths is very low, as it stabilized at the number less than five, and it is expected to decrease further after most of the age group older than fifty years is vaccinated, and the group that suffers from complications of other diseases that affect negatively, and increases the complications after the patient is infected with Corona.

The results of the adopted model for tracking cases of injuries and the expected results confirm that it will stabilize significantly and its prediction model shows that there can be a constant amount of fewer than 400 infections per day with a decrease in deaths and near-zero deaths due to Corona after the results of vaccination and the results of the culture of distancing appear between Individuals, which reduces the number of deaths. However, the number of confirmed cases is decreasing in Saudi Arabia at a lower rate compared to the number of cured cases, as the disease spreads to a smaller area of the country, thus reducing the foci in which there is an increase in vaccinations and adherence to approved health security standards.

The study recommends continuing vaccination and adopting standards of health prevention and social dependence until the data of community-based

prevention reaches the level at which the complete immunity of society is achieved.

REFERENCES

- [1] Bachioua L. On Extended Normal Distribution Model with Application in Health Care. *International Journal of Statistics in Medical Research* 2018; 7: 88-95. <https://doi.org/10.6000/1929-6029.2018.07.03.4>
- [2] Bachioua L. On Extended Exponential Distribution: Properties and Applications In Tracking the Pandemic Covid-19. *SunText Review of Medical and Clinical Research* 2020; 1(2): 1-8. <https://doi.org/10.51737/2766-4813.2020.009>
- [3] Bachioua L. A New Extended- Gamma Family of Distributions: Properties and Applications. *Journal of Applied Mathematics and Computation* 2021; 5(1): 9-17. <https://doi.org/10.26855/jamc.2021.03.002>
- [4] Bachioua L. Foreseeing the Path of Traffic Accidents over the Next Five Years in Algeria Using the Box-Jenkins Method. *Journal of alfikr.alshurti* 2021 ; 116: 134-175.
- [5] Barber D, Cemgil AT, Chiappa S *Bayesian Time Series Models*. Cambridge University Press, New York 2011.
- [6] Bollerslev T. Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics* 1986; 31: 307-327. [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)
- [7] Box GEP, Jenkins GM. *Time Series Analysis: Forecasting and Control*. 4th ed. Hoboken, NJ: Wiley 1970.
- [8] Box GEP, Jenkins GM, Reinsel GC. *Time Series Analysis: Forecasting and Control*. 4th ed. Hoboken, NJ: Wiley 2008. <https://doi.org/10.1002/9781118619193>
- [9] Chatfield C. *The Analysis of Time Series: an Introduction* 1984. <https://doi.org/10.1007/978-1-4899-2921-1>
- [10] Chen P, Yuan H, Shu X. Forecasting Crime Using the ARIMA Model. *Proceedings - 5th international conference on fuzzy systems and knowledge discovery. FSKD; 2008; Vol(5): pp. 627-630.* <https://doi.org/10.1109/FSKD.2008.222>
- [11] Engle RF. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* 1982; 50: 987-1007. <https://doi.org/10.2307/1912773>

- [12] Engle RF. Dynamic conditional correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business & Economic Statistics* 2002; 20: 339-350.
<https://doi.org/10.1198/073500102288618487>
- [13] Granger CWJ, Joyeux R. An Introduction to Long-memory Time Series Models and Fractional Differencing. *Journal of Time Series Analysis* 1980; 1: 15-29.
<https://doi.org/10.1111/j.1467-9892.1980.tb00297.x>
- [14] Higgins ML, Bera AK. A Class of Nonlinear ARCH Models. *International Economic Review* 1992; 33: 137-158.
<https://doi.org/10.2307/2526988>
- [15] Hosking JRM. Fractional Differencing. *Biometrika* 1981; 68: 165-176.
<https://doi.org/10.1093/biomet/68.1.165>
- [16] Lütkepohl H. *New Introduction to Multiple Time Series Analysis*. Springer Verlag, Heidelberg, Germany 2005.
<https://doi.org/10.1007/978-3-540-27752-1>
- [17] Yousaf M, Zahir S, Riaz M, Hussain SM, Shahc K. Statistical Analysis of Forecasting COVID-19 for Upcoming Month in Pakistan. Elsevier Public Health Emergency Collection, Public Health Emergency COVID-19 Initiative, Published online 2020 May 25.
<https://doi.org/10.1016/j.chaos.2020.109926>
- [18] Paules CI, Marston HD, Fauci AS. Coronavirus Infections—more Than Just the Common Cold. *JAMA* 2020; 323(8): 707-708.
<https://doi.org/10.1001/jama.2020.0757>
- [19] Quick-R. Time Series 2020. <https://www.statmethods.net/advstats/timeseries.html>. Accessed April 14.
- [20] Reinsel GC. *Elements of Multivariate Time Series Analysis*. Springer Verlag, 2nd ed, USA 2003.
<https://doi.org/10.1002/0471667196.ess0691>
- [21] Silvennoinen A, Terasvirta T. Multivariate GARCH Models. In *Handbook of Financial Time Series*, Andersen TG, Davis RA, Kreiß J-P, Mikosch T, Ed. New York: Springer 2009; pp. 201-229.
https://doi.org/10.1007/978-3-540-71297-8_9
- [22] Sultana N, Hasan M. Forecasting Temperature in the Coastal Area of Bay of Bengal—An Application of Box-Jenkins Seasonal ARIMA Model. *Civil and Environmental Research* 2015; 7(8): 149-155.
- [23] Worldometers. Covid-19 Coronavirus Pandemic 2021. <https://www.worldometers.info/coronavirus/?fbclid=IwAR0mzmTnHKYnHzwEnJ3jxhUjaW2xKTRK0LOBRvKID0ibVzCyQR YvTD35Pi4>.
- [24] Zhu Z, Lian X, Su X, Wu W, Marraro GA, Zeng Y. From SARS and MERS to COVID-19: a Brief Summary and Comparison of Severe Acute Respiratory Infections Caused by Three Highly Pathogenic Human Coronaviruses. *Respiratory Research* 2020; 21: 224.
<https://doi.org/10.1186/s12931-020-01479-w>

Received on 15-01-2021

Accepted on 25-03-2021

Published on 16-04-2021

<https://doi.org/10.6000/1929-6029.2021.10.03>

© 2021 Bachioua Lahcene; Licensee Lifescience Global.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.