Prediction of Covid-19 Cases in Egypt Using ARIMA Models

ا.د.عبد الوهاب حجاج

استاذ مساعد –رئيس قسم الإحصاء

د.محمد عبدالسلام عجمي

مدرس بقسم الإحصاء

كلية التجارة – جامعة الأزهر

Prediction of Covid-19 Cases in Egypt Using ARIMA Models

د/ عبد الوهاب السيد حجاج
رئيس قسم الإحصاء
MAIL:Wahabstat@yahoo.com
Mobile: 0020-01011604005

د/ محمد عبد السلام عجمي
مدرس بقسم الإحصاء
MAIL:dr.muhammad_alagamy@azhar.edu.eg
Mobile: 0020-01024255565

كلية التجارة - جامعة الأزهر

الملخص

يعد فيروس كورونا (COVID-19) أحد التحديات الهائلة التي واجهها العالم منذ نهاية الحرب العالمية الثانية. ولقد كانت مصر واحدة من الدول التي تأثرت بشكل خطير بـCOVID-19 ، ونتيجة لذلك، اضطرت السلطات المصرية والحكومة إلى اتخاذ العديد من الإجراءات الاحترازية والوقائية لحماية المواطنين المصريين.

يهدف هذا البحث إلى نمذجة الحالات المؤكدة لإصابات Covid-19 في مصر من (٤ مارس ٢٠٢٠) إلى (٢٩ يونيو ٢٠٢١) كفترة تقديرية للنموذج، في حين أن البيانات من (٣٠ يونيو ٢٠٢١) إلى (١٤ يوليو ٢٠٢١) تم استخدامها كفترة تنبؤ. وقد ثبت أن نموذج ARIMA (4،1،5) هو النموذج المناسب والفعال لتمثيل بيانات السلاسل الزمنية، بأقل متوسط مربعات خطأMSE ، وذلك بعد مقارنة الاختلافات الأولى لعدد من نماذج ARIMA.

كما خلص أيضًا إلى أن عدد المصابين سيزداد. لذلك من الضروري أن تتخذ السلطات المصرية إجراءات احترازية عاجلة للحد من هذه الزيادة الوبائية.

الكلمات المفتاحية: Covid-19، السلاسل الزمنية ، نماذج ARIMA ، التنبؤ ، طريقة.Box-Jenkins

Abstract

COVID-19 is one of the massive challenges that has faced the world since the end of World War II. Egypt was one of the countries that were seriously affected by COVID-19. As a result, the Egyptian authorities, and government had to take many precautionary and preventive procedures to protect Egyptian citizens.

This research aims to model confirmed cases of Covid-19 infections in Egypt from (March 4th, 2020) to (June 29th, 2021) as an estimation period of the model, while the data from (June 30th (2021) to (July 14th, 2021) was used as a prediction period. It has been proved that The ARIMA (4,1,5) model is the appropriate and efficient one for representing the time series data, with the least mean squares of MSE errors after comparing several ARIMA models at the first differences.

It was, also, concluded that the number of infected people is going to increase. So, it is necessary for the Egyptian authorities to take urgent precautionary measures to limit this pandemic increase.

Keywords: Covid-19, time series, ARIMA models, forecasting, Box-Jenkins method.

Introduction:

In December 2019, COVID- 19 Virus broke out in Hubei Province, China, especially in Wuhan City (Yang, 2020, p.2). This virus belongs to the Coronavirus family, which is transmitted from animals to humans, and it has been confirmed that this virus has an animal source. (Elias, 2020, p. 2). Being contagious means that it has a rapidly spreading dynamic structure (Roosa, 2020, pp. 256-263). This resulted in the rapid transmission of infection from one person to another. The main symptoms of this disease are coughing and respiratory diseases, as well as difficulty breathing. It appears in a form that is similar to the common cold (Pandey, 2020, p.3). On February 11th, 2020, the World Health Organization called this disease (2019 Coronavirus disease) Consequently, COVID-19 was considered as a global pandemic (Tran, 2020, p.2). Covid-19 has spread at all countries of the world , is not merely in China . United States of America, Italy, Spain, Iran, and France were among the most affected countries (Chintalapudi, 2020, p.4). The repercussions of Covid 19 were not only limited to human health, but it also caused a global economic crisis as well. Unlike what happened with the generations of Corona in previous decades, this is the reason why Covid 19 epidemic was considered as the largest global crisis after World War II.

This is due to the total impact on almost all countries of the world. (Boccaletti, 2020, p.6).

Egypt, like the rest of the world, was not apart from this dangerous epidemic. The Egyptian Ministry of Health announced the first case at Cairo International Airport, related to a Chinese citizen, on February 14th.

In view of the high cost of precautionary measures, it was necessary to focus on using quantitative methods to model daily confirmed cases of Coronavirus. ARIMA time series models are among the methods that have contributed greatly to representing many of the economic, health and social phenomena, by giving a mathematical function that helps explain these phenomena and predict its future values. Prediction is an important part of decision-making, and many of our decisions are based on predictions of unknown future events.

In time series, in particular, prediction is defined as "using the history of the past series, Z_t, (where t expresses time), in estimating its values in the future."

Research objectives:

This research aims to provide a model to predict the number of Covid 19 confirmed infected cases in Egypt.

Importance of Research:

In view of the changes the world has experienced since the outbreak of COVID-19 and its spread in all countries, its economic and social ramifications seem to have a significant and influential impact at global and national levels. The outbreak of this epidemic has led to a series of preventive measures, including "isolation, keeping social distance, travel ban/ banning, complete closure of all national institutions such as schools, universities, companies, factories, entertainment places and tourism companies", which has a negative impact on the economies of all countries in the whole world. It plunged the world order into recession and had a profound impact on Egypt's economic and social systems and all countries in the world as well. It is certain that any crisis experienced by the world society will affect Egypt. Leaving their effects on everyone, although they vary from one country to another, depending on the economic and political situation of each one.

In order to adopt an accurate strategy to overcome this disease and equip health services, resources, and other services to face it, it was useful to pay attention to the issue of prediction in such cases, because prediction models are one of the ways that help in detecting the trend of this disease (Moftakhar, 2020, pp.92). -100).

Research Scope and Limits

The research is concerned with prediction using ARIMA models to predict the number of confirmed affected cases of COVID-19 in Egypt during the period from June 30th, 2021 to July 14th, 2021.

Research Methodology and Tools

This research combines the analytical descriptive approach theoretically with the case study method / approach in application / practically.

Therefore, this research is divided into two aspects, one of which is the theoretical aspect, which simplifies the theoretical basis of the time series model from the general form, model construction stage, estimation and prediction methods and other aspects. Regarding the application, an applied study (case study) was conducted based on the real data of the number of people infected with COVID-19.

The last part includes the most important conclusions, recommendations, appendices, and sources.

The tool used is Minitab statistical software

Literature Review

(Moftakhar 2020), This study aimed to predict the number of new daily cases of COVID-19 in Iran, and this in the period from February 19 to March 30, 2020. And to compare the performance of ARIMA models and ANN artificial neural networks in the prediction process.

(Bayyurt 2020), The aim of this study was to build a model to predict the spread of COVID-19 in Italy, Spain, and Turkey Fr this purpose ARIMA models were used on the data of the European Center to predict the number of infected cases and deaths with Covid-19. The results showed that it is expected that cases will decrease In Italy and Spain, this is as of July, while in Turkey, the decline will be from September. As for deaths, in Italy and Spain, they may reach their lowest level in July, while in Turkey they may reach their highest level in July.

(Ceylan, 2020), This study aimed to find solutions to monitor and predict the spread of Covid-19 and to control it more effectively. Time series models were used to predict the impact of the outbreak of this pandemic.

(2020Dehesh), Study titled Predicting Confirmed Cases of COVID-19 in Different Countries (Italy, China, South Korea, Iran, Thailand) Using ARIMA Models.

(Huang 2020), In this study, a multi-input CNN neural network used to analyze and predict confirmed cases of COVID-19 in China.

(Bastos 2020), This study aimed to use recent data of Brazil to predict the early development of Covid-19 infection.

(Boudrioua 2020), Study entitled: “Predicting the COVID-19 Epidemic in Algeria Using SIR Models”.

(Duangchimkarn, K., Puncheng, W., Wiwatanadat, P., and Shuvatut, P. 2022), Study entitled: SARIMA Model for Predicting the Performance of Daily COVID-19 Statistics in Thailand.

(Ibrahim Demir, and Murat Kresci 2022), This study aimed to use time series, specifically the hybrid SARIMA-NNAR model, to predict the number of cases of COVID-19.

١- Theoretical aspect:

Box-Jenkins presented one of the most important methods that dealt with time series analysis and predicting future values of various phenomena. This method includes a special and unique family/ unit of random models, which are known in the time series literature as Autoregressive Integrated Moving Average models (ARIMA). These models include many special cases, including the autoregressive model and the Mixed Moving Averages model (ARMA).

Therefore, in this part of the research, we will discuss the study of time series models and their construction stages, following the Box & Jenkens method .

(١.١)Time Series

Box&Jenkins' definition of time series is the most important definition given by statisticians, as it defines the time series as "a set of time-sequential observations". If the observations were taken at equal intervals of time previously determined, it is called a “discrete time series”, but if the observations are generated at all points in a certain period of time, it is called a “continuous time series”. Box & Jenkins believe that it is possible to obtain discontinuous time series from continuous time series by taking the readings at equal intervals of time, and it is called “Instantaneously Series”. Observations can also be accumulated or collected for a specific period of time, so, in this case, it is called the “Accumulated Series”. The following are some of the statistical criteria that were used to describe the quality of the time series and facilitate its modeling.

(2-1) Box-Jenkins model for time series analysis:

The (Box-Jenkins) method is one of the most used methods for analyzing time series, and when we are exposed to the (Box-Jenkins) model, we are talking about a unit of models called autoregressive models and Integral Moving Averages. These models can be divided into three basic categories:

(a): Autoregressive (AR) models:

Time series data is said to be generated based on an autoregressive process if the current observation of the series can be expressed as a linear function of the previous observation, plus the random error can be denoted by (in) where (a_t) where (a_t)

$$a_t \sim N(0 , \sigma_a^2)$$

$$y_t = \zeta + \phi_1 \, y_{t-1} - \ldots\ldots - \phi_p y_{t-p} + a_t \quad (1)$$

whereas:

(y_t): the actual, stable observations of the time series of interest.

(ξ): a constant quantity where:

$$\xi = \mu \, (1- \phi_1 - \phi_2 \, \ldots\ldots \, -\phi_P)$$

(μ): the series mean.

(ϕ_i): autoregressive coefficient. Where: i = 1,2, ……,P

(a_t): is a random variable that follows a normal distribution.

(b) Moving Average Models (MA):

In this type of model, the current observation is expressed as a function of the current random error limit and the previous random changes and is indicated by the symbol MA(q), where (q) is the model rank.

The general formula for moving averages is:

$$y_t = \mu + a_t - \theta_1 a_{t-1} - .... - \theta_q a_{t-q} \quad \textbf{(2)}$$

Where:

(y_t) **:** actual observations of the time series of interest.

μ : the series mean.

θ_i: the parameters of the moving average models.

a_i: random variable that follows a normal distribution with an expected value of zero.

And by using the (MA) model, future observations of the series can be predicted by random changes.

(c) Autoregressive and Mixed Moving Average Models (ARMA)

This process includes both autoregressive and moving average processes, and these models are characterized as leading to a joint treatment taking advantage of the advantages of each of the autoregressive and moving averages methods.
The general form of ARMA is:

$$y_t = \xi + \sum_{i=1}^{p} \phi_i y_{t-i} + a_t - \sum_{j=1}^{q} \theta_j a_{t-j} \quad \textbf{(3)}$$

Where:
y_t: actual observations of the time series of interest
ξ : a constant quantity = $\mu\left(1 - \sum_{i=1}^{p} \phi_i\right)$.□
ϕ_i: parameters of self-regression.□
θ_j: Parameters of moving averages.

a_i: random variable that follows the normal distribution.

ARIMA models can be defined as ARMA models with the difference element (d) added, which expresses the process by which an unstable series is converted into a stable series. Suppose that the series is unstable, so it is converted into a stable series by taking the first differences (y_t - y_{t-1}) or the second differences (Δy_t -Δy_{t-1}).

(١-٦) Stages of Model Building

The Box & Jenkins methodology relies on a systematic study of time series based on its specifications, in order to identify it within the family of ARIMA models and to determine the appropriate model for the studied phenomenon, and it goes through four stages: (box and Jenkins, 1976: 243):

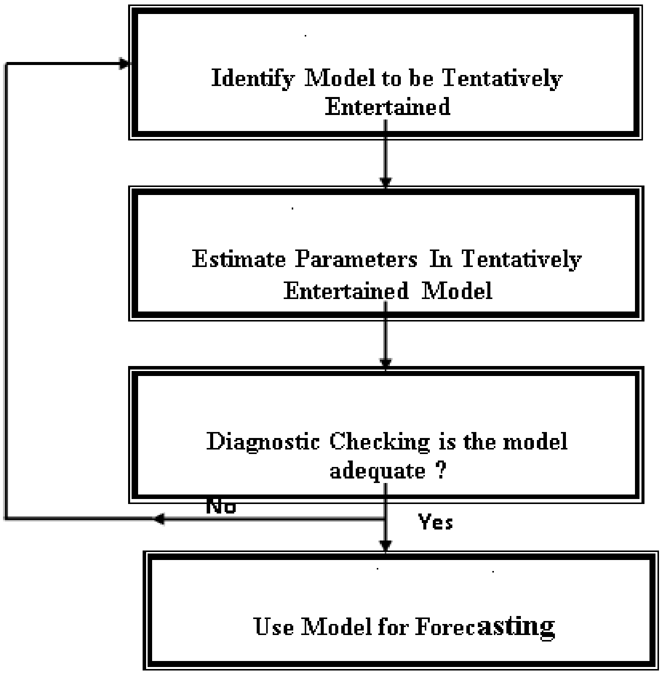


Figure (1) shows the four stages of time series analysis

First Stage: Model Identification

After achieving stability in the time series, the process of determining the appropriate model to represent the series and its degree begins using the autocorrelation functions (ACF) and partial autocorrelation (PACF). of the time series with the theoretical behavior of the autocorrelation and partial autocorrelation functions shown in the following table Pankratz(1983;122).

process	PACF	pacf
AR(p)	Tails off toward zero (exponential decay or damped sine wave)	Cuts off to zero (after lag p)
MA(q)	Cuts off to zero (after lag q)	Tails off toward zero (exponential decay or damped sine wave)
ARMA (pq)	Tails off toward zero	Tails off toward zero

Table (1) Nature of the model according to autocorrelation and partial autocorrelation

Second Stage: Estimation of Model Parameters:

After selecting the appropriate model, its parameters are estimated using one of the complete or approximate estimation methods, which differ according to the model used, which is one of the following two models:

(a) Exact Maximum Likelihood Method (EML)

(b) Non-Linear Least Square Method (NLS)

Third Stage: Diagnostic Checking of Model

After estimating the model, it is importing to test the model applicability or validity of the to represent seasonal time series data, and there are several methods for this, including (Wegman, 2000: 443):

A- The coefficients of the model must be statistically significant, that is, they are significantly different from zero, and for this we use the student's (t) test. If it is not significant, one of the AR or MA ranks must be excluded.

B- Residual analysis. The following tests are used for this:

1- Confidence Interval Checking: To test whether the autocorrelation of errors at seasonal displacement (rs(a))s is significantly different from zero or not, its value must fall between the confidence limits. with probability (0.95). and whereas:

$$Z_t = \frac{t_s(a) - o}{\frac{1}{\sqrt{n}}}$$

So:

$$\Pr\left\{-1.96\left(\frac{1}{\sqrt{n}}\right) < r_s(a) \le +1.96\left(\frac{1}{\sqrt{n}}\right)\right\} = 0.95 \qquad (٤)$$

If this is achieved, then this means that the errors (residuals) are distributed randomly, that the model is good and appropriate (efficient), and can be used in prediction, and that the autocorrelations of the residuals are independent and distributed normally with an arithmetic mean of zero and a variance of $\left(\frac{1}{n}\right)$, meaning that:

$$r_s(a) \sim \text{NID}\left(0, \frac{1}{n}\right)$$

2- Portmanteau Test

One of the most common tests for checking model fitness, is the Q statistic (Pierce & Box statistic), which is used to test the statistical significance of the autocorrelations of the residuals according to the following formula (Box & Price, 1970: 1509-1525):

$$Q = n\sum_{k=1}^{L} r_k^2(a) \sim \chi^2_{((L-m),\,\alpha)} \qquad (٥)$$

If the value of Q is smaller than the tabular value of $\chi 2$, we accept the null hypothesis H0 and conclude that the autocorrelations are not significant, which indicates that the residuals are random and distributed independently, which confirms that the model fit is good and appropriate. This formula has been modified and developed by “Lung and Box” to take the following formula:

$$Q^{*} = n\left(n+2\right) \sum_{k=1}^{L} \frac{r_k^2\left(a\right)}{n-k} \qquad (٦)$$

And this statistic also follows the distribution $\chi^2_{((L-m),\alpha)}$

The fourth stage – Using the model in prediction and/or control:

After confirming the validity of the model, it is used to predict future observations of the phenomenon, and for the model to be a good prediction, the predictions must have: (Minimum Mean Squares Error Forecasts)

Box & Jenkins has proposed the idea of updating forecasts, which is based on the fact that it is possible to use the actual results of the forecast year to update the forecast for the next year, and this is done using two methods:

***Sequential Updating Forecasting Method:**

Through this method, the model parameters are re-estimated, and the fitness of the model is checked whenever we get a new observation, and then the model is used in prediction.

***Adaptive Forecasting Method:**

In this method, (n+1) is used instead of (n) as a starting point to obtain the new predictions, using the new observation value (y_{n+1}) without returning to the estimation of the model's parameters.

2- Application

(2-1) Data description: The research is concerned with using ARIMA models in predicting the number of COVID-19 cases in Egypt using daily 2021 data during the period from March 4, 2020, to July 14, 2021. Obtained from an Exploratory Study on Coronavirus in Egypt, Third Edition, Central Agency for Public Mobilization and Statistics, Egypt, August. Where the data was used from (March 4, 2020) to (June 29, 2021) as an estimation period for the model, data were used from June 30, 2021, to (14 July 2021) as a forecast period.

(٢-٢)Time Series Analysis:

(١-٢-٢)Drawing the time series: Before starting the time series analysis, the time series data was drawn to identify its initial characteristics, as shown in (Figure 1)

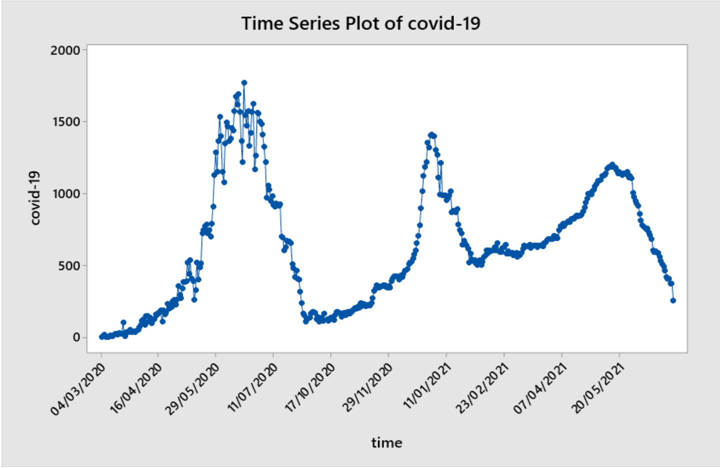


Figure (1): The time series of the number of COVID-19 cases.

By drawing the general direction of the series, we get the following figure:

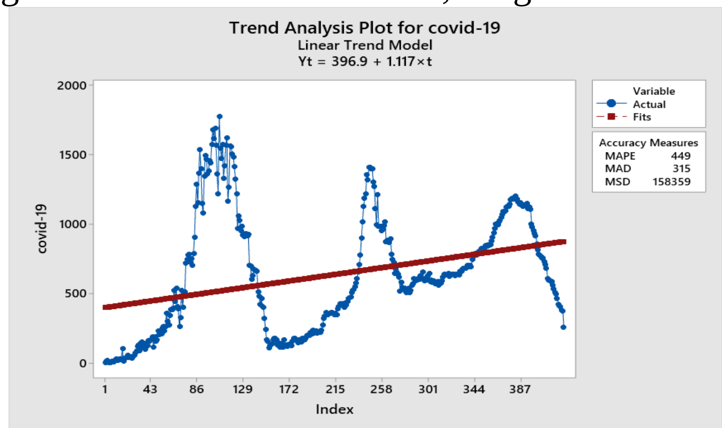


Figure (2): The general trend of the time series for covid-19 cases number.

It is noted from the figure that there is a general trend increasing with time, as well as the presence of oscillations represented in concavities and protrusions, and these oscillations are repeated regularly and at the same pace. These changes indicate the existence of a “Secular Trend”.

From the figure, we find that the direction of the chain is upward.

(2-2-2): Testing the stability of the time series:

To obtain the stability of the variance, the data were processed by taking the first differences again, and figure (3) shows the relative stability of the difference’s series.

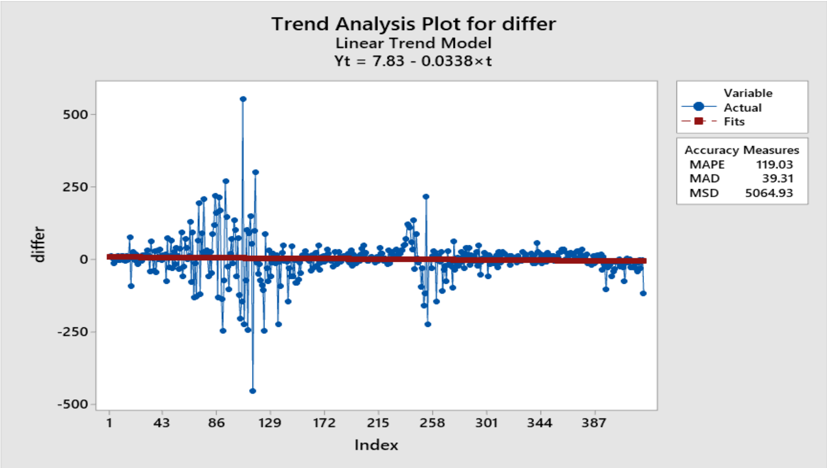


Figure (3): The general trend of the time series for the number of cases after taking the first differences

Accordingly, it can be said that the time series has become relatively stable and capable of conducting the four stages of time series analysis:

The First Stage: Model Characterization:

At this stage, the appropriate model was chosen for the available data, and since the first series of differences are stable, we resort to drawing the two functions:

Autocorrelation Function (ACF), and the Partial Auto-Correlation Function (PACF), as in figures (4) and (5).

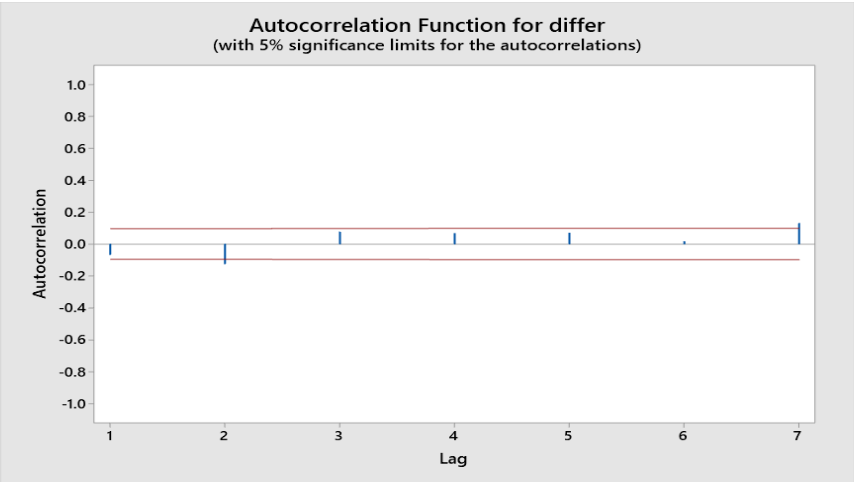


Figure (4): The autocorrelation function of the number of COVID-19 cases

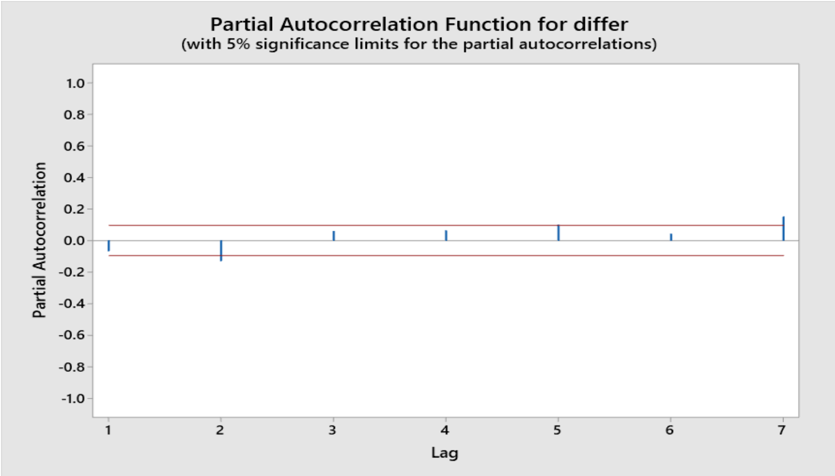


Figure (5): The partial autocorrelation function of the number of COVID-19 cases

By observing the behavior of the autocorrelation and partial autocorrelation functions in Figures (4) and (5), we find that they gradually decrease, following an exponential behavior or the behavior of the sine function (gradually fading) (Decays Exponentially).

Therefore, and for further investigation, the model is determined by taking all possible probabilities to estimate all models' first differences (d = 1)) and parameters (p,q = 0,1,.....,5) according to the measure of the mean sum of squares error MSE using MINITAB program as in the following table .

P,d,q	**MSE**	**P,d,q**	**MSE**	**P,d,q**	**MSE**	**P,d,q**	**MSE**
(0,1,1)	٥٠٧٤.٣٨	(0,1,4)	4968.86	(3,1,1)	4857.63	(4,1,4)	4750.37
(1,1,1)	5071.06	(0,1,5)	4928.54	(3,1,2)	4854,79	(4,1,5)	4688.87
(1,1,2)	4963.03	(1,1,0)	5082.13	(3,1,3)	4806.73	(5,1,0)	4952.25
(1,1,3)	4974.79	(2,1,1)	5049,39	(3,1,4)		(5,1,1)	4862.53
(1,1,4)	٤٩٦٩.٩٩	(2,1,2)	5012.93	(3,1,5)		(5,1,2)	4807.96
(1,1,5)	٤٨٥١.٣٠	(2,1,3)	4920.97	(4,1,0)	4991.34	(5,1,3)	4771.72
(2,1,0)	5006.24	(2,1,4)	4810.76	(4,1,1)	4859.97	(5,1,4)	4781.20
(0,1,2)	5029.74	(2,1,5)		(4,1,2)	4862.99	(5,1,5)	
(0,1,3)	4976.12	(3,1,0)	4999,88	(4,1,3)	4817.57		

Table No. (2) MSE value for ARIMA models at first differences

We find from the previous table that the best model is the ARIMA (4,1,5) model with least mean square errors MSE = 4688.87, noting that the models ARIMA(2,1,5), ARIMA(5,1,1) , ARIMA(5 ,1,2) ARIMA(5,1,5) was (non-stationary) or non-invertible so the program could not complete its calculation.

The Second Stage: Model Estimation:

After examining the possible models, we concluded that ARIMA (4, 1,5) is the proper model, and by applying the nonlinear least squares method (NLS) to the time series data under study, and using the Minitab program, the following results were obtained:

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	-0.766	0.121	-6.31	0.000
AR 2	0.0547	0.0779	0.70	0.483
AR 3	0.8666	0.0781	11.10	0.000
AR 4	0.548	0.119	4.60	0.000
MA 1	-0.690	0.120	-5.76	0.000
MA 2	0.2721	0.0914	2.98	0.003
MA 3	0.8915	0.0774	11.52	0.000
MA 4	0.306	0.125	2.45	0.015
MA 5	-0.2985	0.0493	-6.05	0.000
Constant	0.01	1.73	0.01	0.993

Table (3) Results of ARIMA Parameter Estimations(١,٥ ,٤)

We note from the previous results that the parameters are statistically significant (significantly different from zero

The Third Stage: Testing the Model Validity

It is necessary to ensure that the model is valid and effective, then to determine its grade and estimate, and all this can be done after diagnosing the model. This can be implemented by the following means:

A- Testing the autocorrelation coefficients of residues.

The autocorrelation coefficients and partial autocorrelation coefficients were extracted and plotted for the residuals (errors) of the model as in the following figures (6) and

(7)

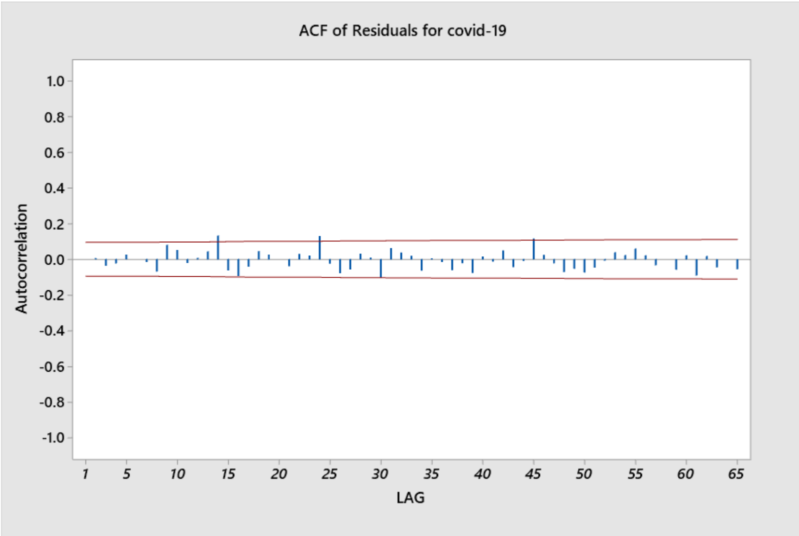
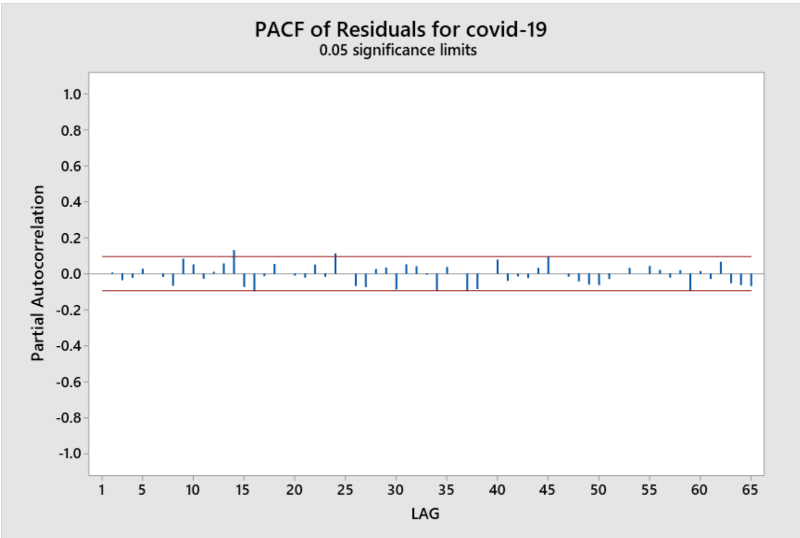


Figure (6): The autocorrelation function of the estimated model residuals

Figure



(7): Partial autocorrelation function for the estimated model residuals

It is noted from the figures that all the values of the autocorrelation coefficients and the partial residual values fall within the confidence limits, and this confirms the randomness of the series of residual values as well as the quality and suitability of the model used.

b Test that the residuals of the model follow a normal distribution:

To calculate the limits of predictive confidence and to ensure the effectiveness of (t) tests on the parameters, it is necessary to ensure that the errors shown in Figure (8) follow the normal distribution:

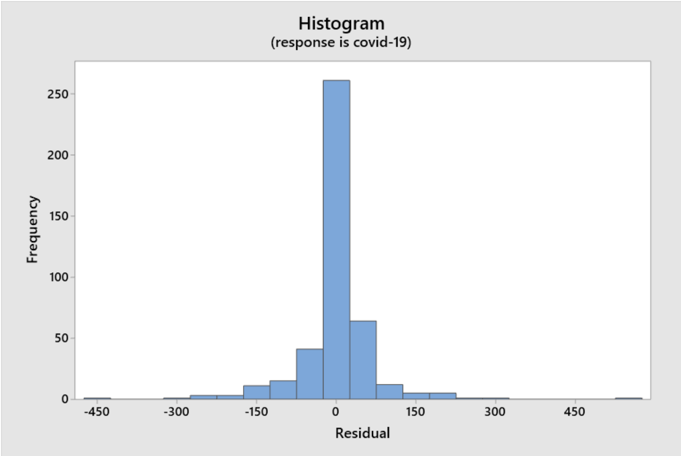


Figure (8): The normal distribution of the estimated model residuals

By testing the hypotheses of symmetry, and normal flatness of the residual series it is shown that the series of the residual exhibits the properties of a normal distribution. Based on the previous tests, it can be said that they lead to the acceptance of the model statistically, and therefore it can be used for prediction.

Fourth Stage: Prediction

Using the prediction model obtained in the estimation stage, the number of cases infected with COVID-19 during the period from (June 30, 2021) to (July 14, 2021) was predicted. The following table shows these predictions with confidence intervals.

Period	95% Limits		Forecast
	Upper	Actual	
30/06/2021	387.977	119.500	253.739
01/07/2021	445.088	79.509	262.299
02/07/2021	447.658	28.358	238.008
03/07/2021	462.904	-15.479	223.713
04/07/2021	479.918	-62.762	208.578
05/07/2021	506.032	-99.953	203.040
06/07/2021	510.263	-148.737	180.763
07/07/2021	539.454	-186.299	176.578
08/07/2021	557.153	-226.190	165.481
09/07/2021	571.270	-268.427	151.422
10/07/2021	596.730	-305.221	145.755
11/07/2021	616.931	-342.074	137.428
12/07/2021	632.582	-382.101	125.241
13/07/2021	658.914	-415.889	121.512
14/07/2021	678.403	-451.618	113.393

Table (4) predictions of the proposed model

Conclusions and Recommendations:

First: Conclusions:

From the above-mentioned issues, we conclude the following results:

1- In view of the high cost of precautionary measures, it is necessary to concentrate on using quantitative methods to model daily confirmed cases of the Coronavirus, as time series models are among the ones that contributed significantly to the modeling of many "economic, health and social phenomena", by giving a mathematical function that explains these phenomena and predicts its future values.

2- The time series is considered as the most accurate method in prediction, especially in the absence of causal relationships between the variables or the lack of sufficient information about the explanatory variables.

3- Statistical tests showed that the time series is unstable and that there is a clear general trend in the series, and to provide stability conditions in the series, we first modified them by fixing the variance and removing the general trend using first-order differences.

4- Using the comparison criteria (the least variance of the model and the least value of residual sum of squares), the best tested model among the ones that were used, the validity and fittings of the proposed model was examined statistically through tests (the significance of the estimated parameters, the analysis of the autocorrelation function of the residuals, and the normal distribution of the residuals).

5- It was found that the appropriate and efficient model to represent the time series data is the ARIMA (4, 1,5) model.

6- Using the prediction model was obtained in the estimation stage, the number of COVID-19 cases in Egypt was predicted during the period from (June 30th, 2021) to (July 14th, 2021).

Second: Recommendations:

Based on the results achieved, we recommend the concerned authorities to take into account the results of this research and the approved formula for prediction, due to its adoption of the appropriate scientific method in this field.

References:

1) Aashish Jain, Rohit Khokher, Ram Chandra Singh, Rahul Kumar(2020), “Forecasting of COVID-19 Cases Using SARIMA Model in India”, Solid State Technology, Volume: 63 Issue: 2s (2020)3516 archives. Available@ www.solidstatetechnology.us

2) Al-Ghannam, Ahmed bin Abdullah, (2003), “Analysis of the time series of the stock price index in the Kingdom of Saudi Arabia using the Box-Jenkins methodology,” King Abdulaziz University Journal, Issue Two, pp.(٢٥-٣) .

3) Abdelkader Sahed, Hassan Kahwi(2020), “Prediction of Confirmed Cases of COVID-19 in Algeria Using ARIMA Models”, Shuaa Journal of Economic Studies Volume 04 / Issue: 02-2020.

4) Abonazel, Mohamed Reda, and Ahmed Ibrahim Abd-Elftah(2019). "Forecasting Egyptian GDP Using ARIMA Models." Reports on Economics and Finance 5.1 .2019, PP 35 - 47.

5) Akaike, H.(1974), "A new look at the statistical model Identification " , IEEE Transactione on Automatic Control,Vol.19 , No.6 ,PP.716-723 .

6) Anderson, O.D. (1974), 'Time series analysis and forecasting " , Butter worths , London and Boston .

7) An Exploratory Study on Coronavirus in Egypt(2021), Third Edition, Central Agency for Public Mobilization and Statistics, Egypt, August 2021.

8) Attia, Abdel-Qader Mohamed, (2000), “The Way of Measuring Economic Relations,” Alexandria, the Egyptian Universities House.

9) Bastos, S. B., &Cajueiro, D. O.(2020), “Modeling and forecasting the Covid-19 pandemic in Brazil”. arXiv preprint arXiv:2020.

10) Bayyurt, L., &Bayyurt, B.(2020), “Forecasting of COVID-19 Cases and Deaths Using ARIMA Models”. medRxiv.2020.

11) Boccaletti, S., Ditto, W., Mindlin, G., &Atangana, A.(2020), “Modeling and forecasting of epidemic spreading: The case of Covid-19 and beyond. Chaos, Solitons, and Fractals. 2020.

12) Boudrioua, M. S., &Boudrioua, A. (2020), “Predicting the COVID-19 epidemic in Algeria using the SIR model”. medRxiv. 2020.

13) Box G., E. , P. , and Jenkins G. , M. , T. (1976) , "Time Series Analysis Forecasting and conteol " , San Farncisco , Holden-Day , U.S.A .

14) Box , G. E. and Price , D. A. (1970) , "Distribution of Residual Autocorrelations in Autoregressive – Integrated Moving Average Time Series models" , JASA ,Vol.55 , NO.332 , PP .1509-1525 .

15) Brock well , P.J. and Davis , R.A. (1991) , " Time Series Theory and Methods , 2nd ed , Springer Verlag York lac ,   New York .

16) Ceylan, Z(2020). "Estimation of COVID-19 prevalence in Italy", Spain, and France. Science of The Total Environment, 2020.

17) Chintalapudi, N., Battineni, G., &Amenta, F. (2020), "COVID-19 disease outbreak forecasting of registered and recovered cases after sixty-day lockdown in Italy: A data driven model approach". Journal of Microbiology, Immunology, and Infection. 2020.

18) Dehesh, T., Mardani-Fard, H. A., &Dehesh, P. (2020), "Forecasting of covid-19 confirmed cases in different countries with arima models". medRxiv. 2020.Ilyas, M., Rehman, H., &Nait-ali, A.

19) Muhammad Ilyas, Hina Rehman, and Amine Nait-ali (2020), Detection of Covid-19 From Chest X-ray Images Using Artificial Intelligence: An Early Review. arXiv preprint arXiv:2020. p2

20) Duangchaemkarn. K, Boonchieng.W, Wiwatanadate, P.,Chouvatut, V.(2022), "SARIMA Model Forecasting Performance of the COVID-19 Daily Statistics in Thailand during the Omicron Variant Epidemic. Healthcare 2022, 10, 1310. https://doi.org/10.3390/healthcare10071310

21) Hashman, Mouloud, (1998), "Short-Term Forecasting Models," Algeria, University Press.

22) Huang, C. J., Chen, Y. H., Ma, Y., &Kuo, P. H.(2020), Multiple-Input DeepConvolutional Neural Network Model for COVID-19 Forecasting in China. medRxiv. 2020.

23) Lawrence , A.K. and Paul , I.N. (1978) , " on Conditional Least Square Estimation For Stochesses " Ann .of Stat . , vol.6 , No.3 ,PP.629-642 .

24) Makridakis , S. , Wheel Wrifht S. , C. , and McGee (1983) , "forecasting Method and Application " 2nd ed , John wily and Sons. Inc. , U.S.A.

25) Moftakhar, L., Mozhgan, S. E. I. F., & Safe, M. S. (2020), "Exponentially Increasing Trend of Infected Patients with COVID-19 in Iran: A Comparison of Neural Network and ARIMA Forecasting Models". Iranian Journal of Public Health, 49, Ceylan, 2020, pp 92-100

26) Pandey, G., Chaudhary, P., Gupta, R., & Pal, S. (2020), "SEIR and Regression Model based COVID-19 outbreak predictions in India". arXiv preprint arXiv:2020.

27) **Pankratz , Alan (1983).** " Forecasting With Univariate Box-Jenkins Models" , Publishers Limited , Market Place , , Dedication Oxford Ox 545 E, Great

28) Tan, C.V.; Singh, S.; Lai, C.H.; Zamri, A.S.S.M.; Dass, S.C.; Aris, T.B.; Ibrahim, H.M.; Gill, B.S., (2022), "Forecasting COVID-19 Case Trends Using SARIMA Models during the Third Wave of COVID-19 in Malaysia". Int. J. Environ. Res. Public Health 2022, 19, 1504. https://doi.org/10.3390/ijerph19031504.

29) Tran, T. T., Pham, L. T., & Ngo, Q. X.(2020), "Forecasting epidemic spread of SARS-CoV-2 using ARIMA model (Case study: Iran)", Global J. Environ. Sci. Manage, 6, 4.2020. P2

30) Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J. M., ...&Chowell, G. (2020), Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. Infectious Disease Modelling, 2020, pp 256-263.

31) Rosmelina D. Satrisna, Aniq A. Rohmawati, Siti Sa'adah (2021), Forecasting the Covid-19 Increment Rate in DKI Jakarta Using Non-Robust STL Decomposition and SARIMA Model, Intl. Journal on ICT, Vol. 7, No. 1, Jun 2021. pp. 21-30 doi: 10.34818/ijoict.v7i1.554, OPEN ACCESS ISSN 2460-9056, http://socj.telkomuniversity.ac.id/ijoict/

32) Samir Shaarawy, (2005). "Introduction to the modern analysis of time series", Center for Academic Publishing, King Abdulaziz University, Saudi Arabia.

33) Wegman , E.J. (2000) , " Time Series Analysis – Theory , Data Analysis and Computation " ,Addison-Wesley Publishing Company.

34) Wei , W.S (1990) , Time Series Analysis : Univariate and Multivariate Methods " , Addison –Wesley Publishing Company Inc . , U.S.A.

35) Yang, Z., Zeng, Z., Wang, K., Wong, S. S., Liang, W., Zanin, M., ...& Liang, J.(2020), "Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions". Journal of Thoracic Disease, 2020, 12(3) .P2