

Employment of GARCH Model and L.S.M.E Method in Time Series with Application to COVID-19 Virus

Aaedah A. Mohammad ^(a), Nihad S. Khalaf Aljboori ^(b)

^{(a),(b)} College of Education for Women, Mathematic department, Tikrit University, Iraq

Email's

^a ayedamohammed@st.tu.edu.iq , ^b nihad.shreef16@tu.edu.iq

Received: 2022 March 15; **Revised:** 2022 April 20; **Accepted:** 2022 May 10

Abstract

This paper is concerned with approximating the short-term spread of the COVID-19 virus at the country from January 1, 2021 to July 25, 2022, extracted from the approved website of the Ministry of Health in Iraq using GARCH models and L.S.M.E method. In addition, we will try to determine the hypothetical inflection point and the final volume of cases of COVID-19.

Where the analysis is from two parts firstly by applying the conditions for stability to GARCH models, and secondly part in the application of the Long-Short terms memory network , and we have found that the best model according to AIC and BIC information criteria is GARCH (1,0) this is regarding data analysis using GARCH Models, while the analysis using L.S.M.E according to the same AIC and BIC criteria. The results are of a standard value less than GARCH models due to the lack of high fluctuation in the severity of the contrast, where the GARCH models are used for states of high fluctuation and insecure in the variation of error, due to the adaptive nature of the network models Nervousness made it preferred models to describe these data and get the lowest medium for errors.

Keywords : ARCH, GARCH, stationarity, Leverage effect, Box-Jenkins, AIC, BIC, residuals, L.S.M.E method.

1. Introduction

A time series is simply a sequence of different data points that occurred in a sequential order for a certain period of time. As it usually depends on the assumptions of stability, linear and natural, and these three characteristics are very important in deciding and building time series models, so the time series must include these assumptions and how to address the instability of the studied series and build an appropriate mathematical model for it in terms of linearity or otherwise, and was it a normal distribution or not.

In this paper, we will use GARCH models on the dynamics of COVID-19 spread in Iraq to illustrate the simplicity and transparency of these models as well as their high predictability. Whereas, the Autoregressive Conditional Heteroscedasticity Variance model known as shortly by (ARCH), which was proposed by researcher Engle in 1982[1], It is a primitive model of the model developed and proposed by Bollerslev in 1986, where (GARCH) model

is a generalized model of (ARCH) model where GARCH(P,Q) model describes P from the squares of past errors while Q represents the variances of conditional errors.[2]

Many models as it suggested that the GARCH model expand. For instance exponential GARCH (EGARCH) by Nelson in 1991 [5], Glosten-Jagannathan-Rankle in 1993 (GJR-GARCH) [4], and TARARCH by Zakoian et al in 1994 [3].

2.Prelimiaries

Robert Engle's 1982 suggested ARCH model, which will have the formula :

$$y_t = \sigma_t \epsilon_t \text{ where } \epsilon_t \sim iid N(0,1)$$

$$\sigma_t^2 = \omega + \sum_{j=1}^Q \varphi_j y_{t-j}^2 \quad \dots(2.1)$$

Where σ_t^2 is the conditional variance, while $\omega, \sum_{j=1}^Q \varphi_j$ They are the parameters of the model.

This model, which is based on the martingale difference,

$$E(Y_{t+1}^2 / F_t) = \sigma_t^2 \quad \dots(2.2)$$

where F_t is a σ -field of a random variables called a sometimes filter $(y_{t-1}, y_{t-2}, \dots, y_{t-Q})$. [3]

The generalized autoregressive conditional heteroscedasticity model, or GARCH(Q,P), developed by Bollerslev [2] in 1986 as an extension of the ARCH model, has the following equations:

$$y_t = \sigma_t \epsilon_t \text{ where } \epsilon_t \sim iid N(0,1)$$

$$\sigma_t^2 = \omega + \sum_{j=1}^Q \varphi_j y_{t-j}^2 + \sum_{i=1}^P \gamma_i \sigma_{t-i}^2 \quad \dots(2.3)$$

where ω is a constant and φ_j, γ_i a represent the ARCH and GARCH parameters, respectively. Whereas, Q represents the number of conditional variances of the variance parameters, while P represents the number of variances of conditional errors.

And by imposing some restrictions regarding the parameters in the conditional variance equation so that it must be $\omega > 0, \varphi_j \geq 0 \forall j, j = 1, 2, \dots, Q, \gamma_i \geq 0 \forall i, i = 1, 2, \dots, P$, and $\sum_{j=1}^Q \varphi_j + \sum_{i=1}^P \gamma_i < 1$ then the conditional variance is

$$\sigma_t^2 = \frac{\omega}{1 - (\sum_{j=1}^Q \varphi_j + \sum_{i=1}^P \gamma_i)} \quad \dots(2.4)$$

This is precisely the condition for the stability of the model.

Despite the advantages that GARCH models possess, they also have weaknesses, which were indicated by Black in 1976, which is the failure to obtain the Leverage effect.[8]

In order for the GARCH model to be a stationarity model, the conditional variance must converge to the value of the unconditional variance. This means that $\sigma_t \rightarrow \sigma_y$. [6],[7]

2.1 Box-Jenkins methodology

This methodology was developed by Box and Jenkins in 1970 and then developed during the period (1970-1996) to include working on more models as needed, as this methodology includes three stages main to reach the best predictive model and these stages are identification (Data preparation, Model selection), Estimation and testing (Estimation, Diagnostics) and Application (forecasting).[9]

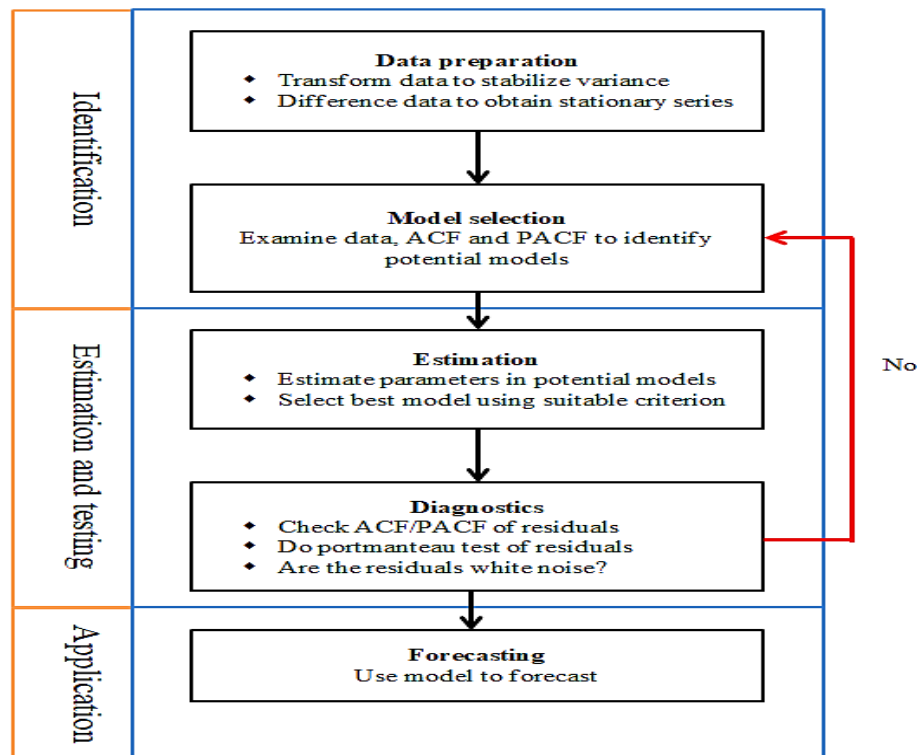


Figure (2.1) Box-Jenkins methodology [10]

2.2 Some Criteria of Model order Selection

2.2.1 Akaike information criterion (AIC)

It is the most commonly used and is given by the following relationship

$$AIC(p, q) = -2 \log(\text{maximum likelihood}) + 2k \quad \dots(2.5)$$

Where k is number of parameters.[11]

2.2.2 bayesian information criterion (BIC)

This criterion was proposed in a similar way to the Akiake criterion in order to achieve more convergent properties and is given by the following relationship

$$BIC = -2l(\text{maximum likelihood}) + k \ln n \quad \dots(2.6)$$

Where m number of observation, and k is number of parameters.[12]

2.3 Long short-term memory (LSTM) network

LSTM networks are well-suited to classifying, processing and making predictions based on time series data, since there can be lags of unknown duration between important events in a time series. LSTMs were developed to deal with the vanishing gradient problem.

It is an artificial neural network used in the fields of artificial intelligence and deep learning. Such a recurrent neural network can process not only single data points, but also entire sequences of data.[13]

3. implementation

3.1 Data descriptions

The data collected in the current study refer to the number of confirmed cases of COVID-19 for the period from (1 January 2021- 25 July 2022) by 570 observation and extracted from the approved website of the Ministry of Health in Iraq (<https://moh.gov.iq/?page=46>)

So that we will apply the conditions of stability of the GARCH model and Long-short terms Network method to this data to get the best predictive model.

3.2 Modeling and creating a GARCH model

We will apply the GARCH model to the data series and note and verify that the expected conditional variance is approaching of unconditional variance, as well as steps to detect changes, form the form, estimate parameters, check appropriate, then predict the police contrast.

The program used in the process of creating and programming the time series is the Matlab R2020a program. So that we will review several models of different orders and see the best model obtained.

3.3 Data modeling and analyzing

Beginning to generate a time series, we enter data and draw it. Figure (3.1) represents a time-series plot of historical data for COVID-19 infections for the period from (1 January 2021- 25 July 2022)

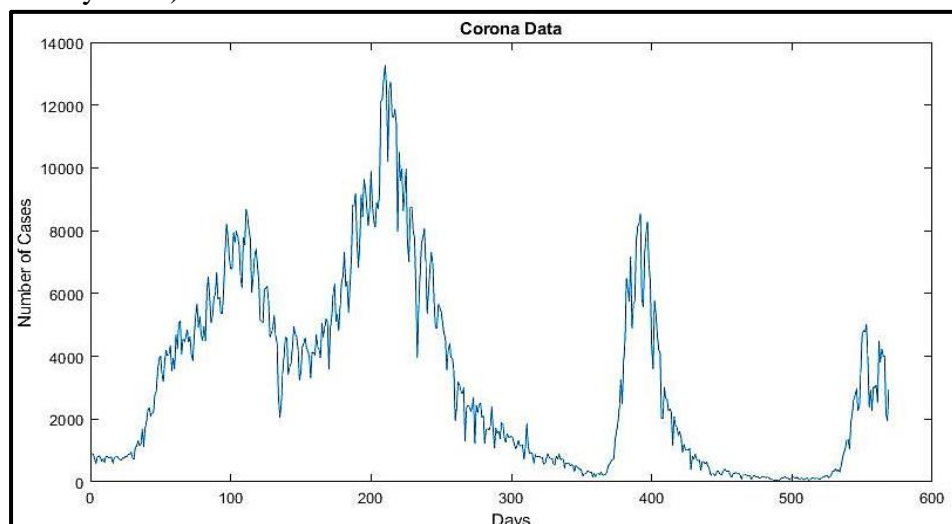


Figure (3.1) represents a time-series plot of historical data for COVID-19 for the period from(1 January 2021- 25 July 2022)

The next step includes several stages, which is the conversion of the original series to the series of returns as well as the conversion of the series of returns to the series of square errors, and draw the Autocorrelation functions of the GARCH model with different orders. Where Figure (3.2) represents a shows the autocorrelation functions of these models.

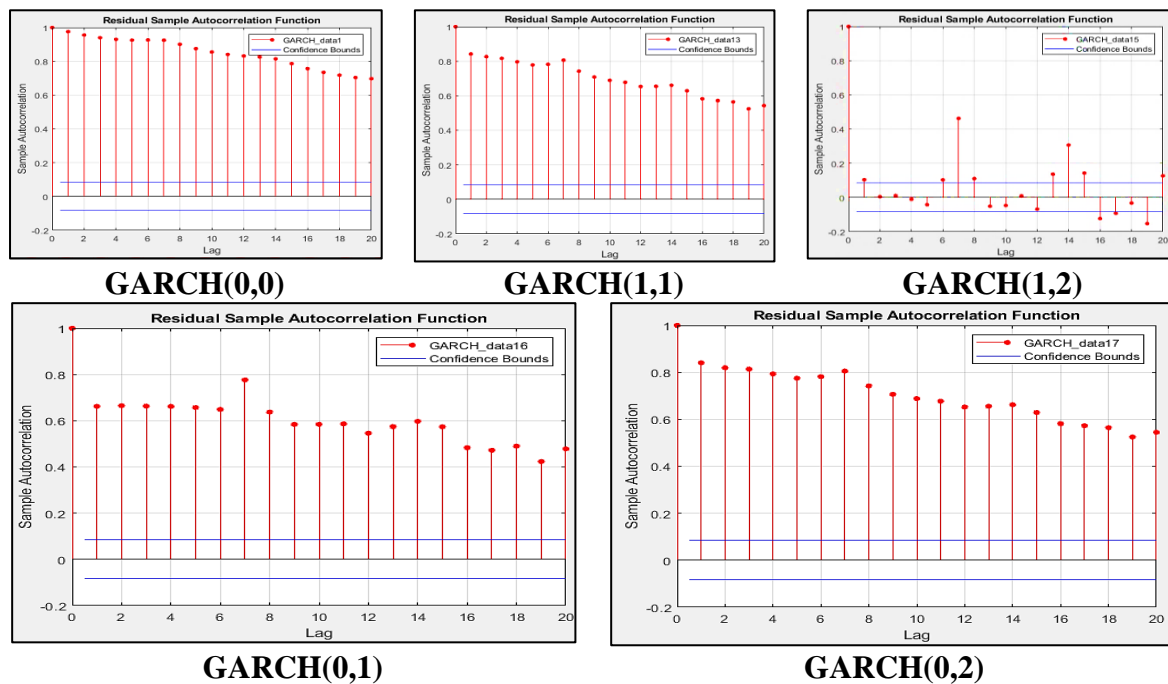


Figure (3.2) the Autocorrelation functions of the GARCH model with different orders

And then Conducting the Ljung-Box test to detect the effect of heteroscedasticity for a number of lags, at the conclusion of this stage we convert the square error series to a series of residuals, where Figure (3.3) represents a drawing of a number of residual functions of the GARCH model with different orders, while Figure (3.4) shows the autocorrelation functions of these models.

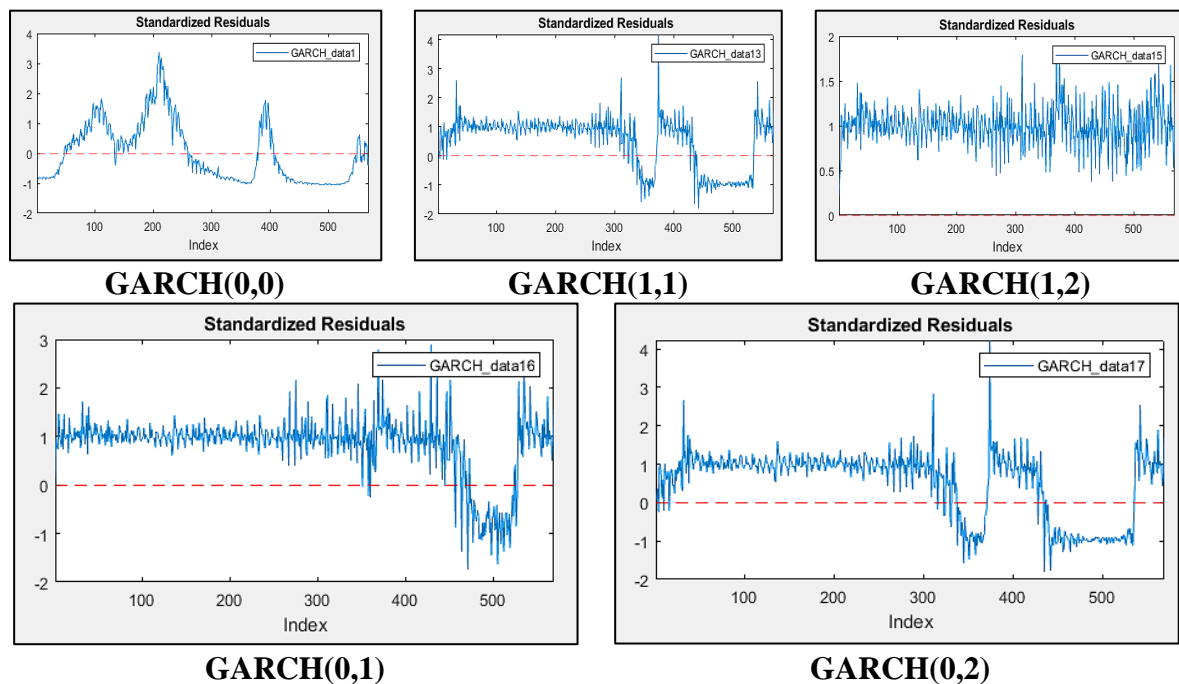


Figure (3.3) the Residual functions of the GARCH model with different orders

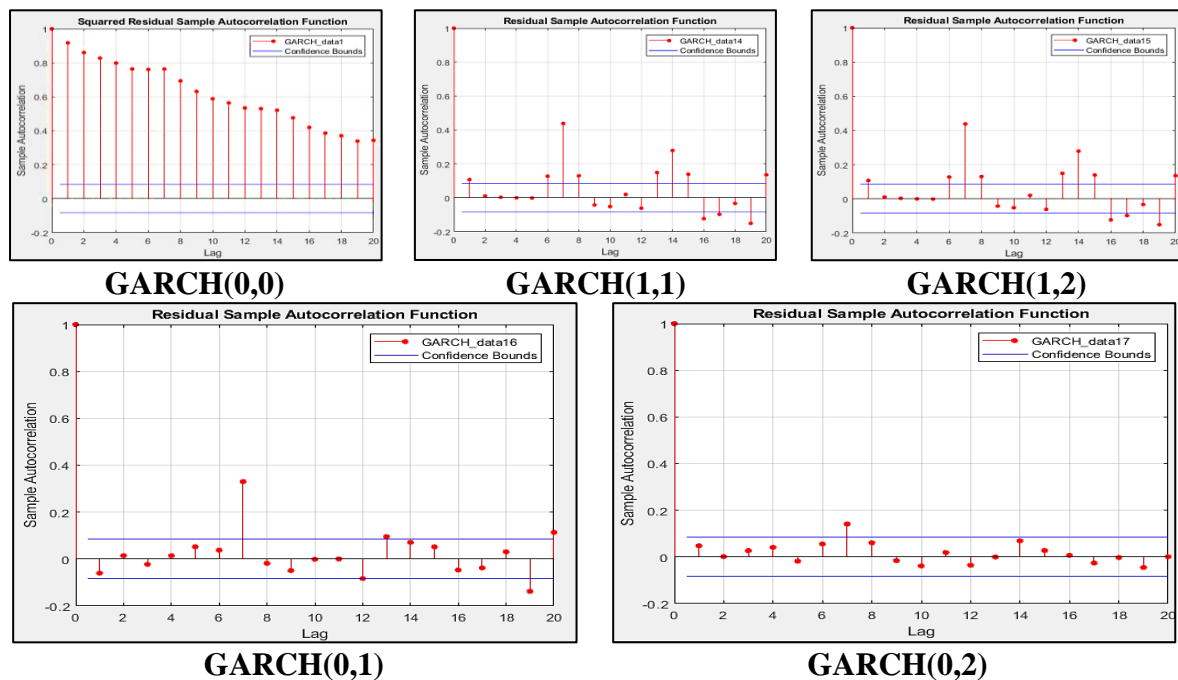


Figure (3.4) the Autocorrelation functions of the GARCH model with different orders

It is clear from the light figures, that the series are formed in different shapes and the Autocorrelation function ACF for the previous templates and for 20 lags that the best model for controlling volatility is GARCH(0,1).

Now we review the quantum and quantum function, known for short as (QQ Plot), which is a visual evaluation to determine whether the series takes the form of a normal distribution or not for the model and the frequency distribution curve of the model GARCH (0,1)? for the Residual series.

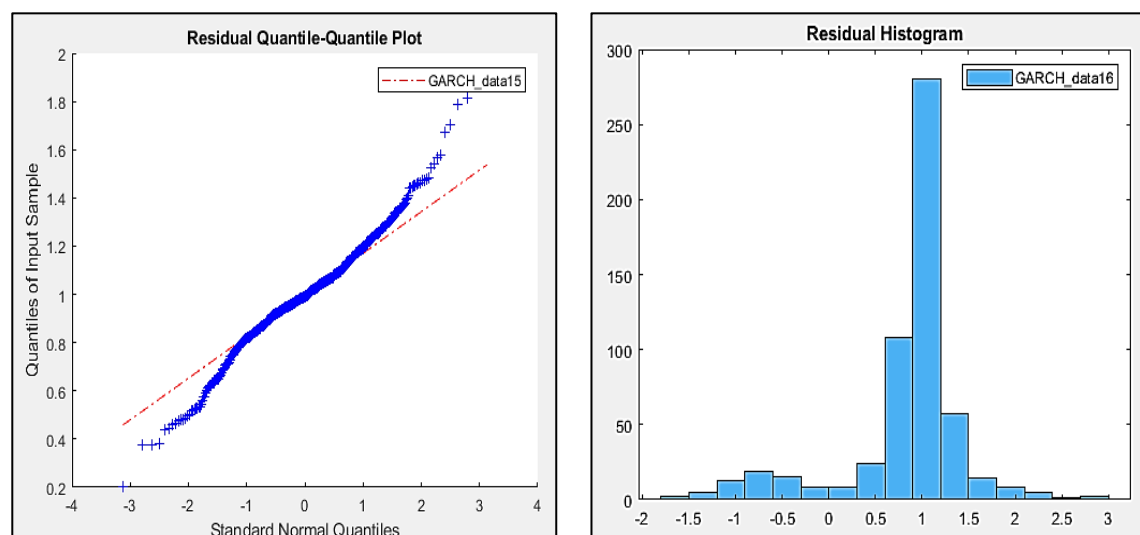


Figure (3.5) the QQ Plot functions and distribution curve of the GARCH(0,1) model for Residual series

In addition to the previously presented illustrations that showed the general trend of the series, we must ensure the suitability of the model that we will rely on to give future

predictions. We will present two tables, the first representing the estimated parameters for each model, while the second table shows the results of the AIC and BIC criteria to choose the best model.

GARCH(P,Q)	Constant	ARCH(1)	ARCH(2)	GARCH(1)	GARCH(2)
GARCH(0,0)	8.997e+06	0	0	0	0
GARCH(1,1)	1.7080e+04	0.8293	0	0.1706	0
GARCH(1,2)	2.1760e+04	0.7956	0.2044	0	0
GARCH(0,1)	3.0490e+03	1.0000	0	0	0
GARCH(0,2)	2.1761e+04	0.7956	0.2044	0	0

Table (3.1) the value of parameters model of deferent rank from GARCH model

GARCH(P,Q)	AIC	BIC
GARCH(0,0)	1.0730e+04	1.739e+04
GARCH(1,1)	9.8993e+03	9.9166e+03
GARCH(1,2)	1.0085e+04	1.0098e+04
GARCH(0,1)	9.8482e+03	9.8612e+03
GARCH(0,2)	9.8975e+03	9.9149e+03

Table (3.2) the value of AIC and BIC of deferent rank from GARCH model

Therefore, the value of the conditional variance of the studied models is shown in Figure (3.6)

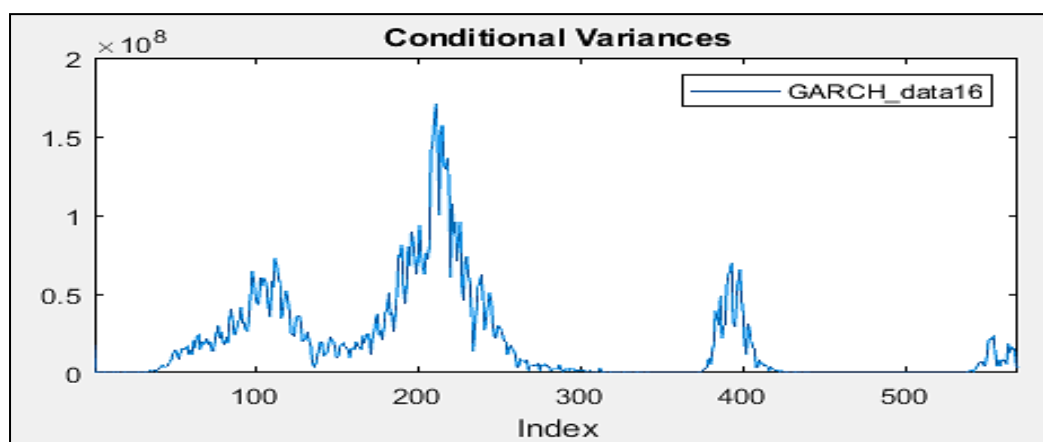


Figure (3.6) the conditional variance of the GARCH(0,1) model

Through all the previous steps of the data analysis stages shown in (graphs, tables), it becomes clear to us that the best model that can be relied upon and used for forecasting is the

GARCH(1,0) model, but now we only have to check the stability of the chosen model by applying the stability conditions of the model and the value of the conditional variance shown in the equation (2.4).

$$\varphi_1 = 1.0000$$

By applying the stability conditions to the model, we find that the model is stable. The value of the unconditional variance of the GARCH model is

$$\begin{aligned}\sigma^2 &= \frac{\omega}{1 - (\varphi_1 + \sum_{i=1}^2 \gamma_i)} \\ &= \frac{3.0490e+03}{1-1} = \text{unknown quantity}\end{aligned}$$

Therefore, the general equation of the GARCH(1,0) model In order to perform the prediction of unconditional variance values on it is in the form:

$$\begin{aligned}y_t &= \sigma_t \epsilon_t \text{ where } \epsilon_t \sim iid N(0,1) \\ \sigma_t^2 &= \omega + \varphi_1 y_{t-1}^2\end{aligned} \quad \dots(3.1)$$

4. Modeling and creating by using (LSTM)

We will use the L.S.T.M method, another method of time series modeling and analysis used. So that we will employ the method of the neural network that uses Feed-Backward method by using a binary algorithm, first we draw the time series where Figure (4.1) represents the time series

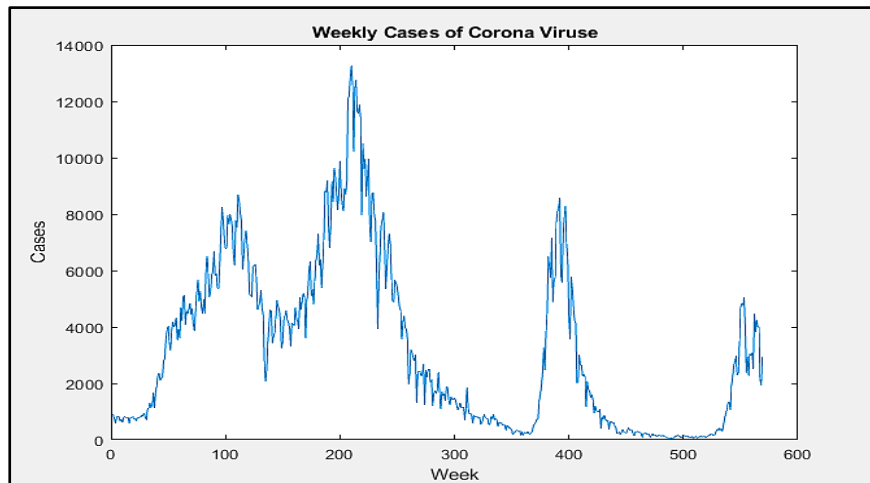


Figure (4.1) represents a time-series plot of historical data for COVID-19 for the period from(1 January 2021- 25 July 2022)

The previous graph displays the time series of the daily casualty data, which shows clear fluctuations, as it appears in the figure and a decline followed by another decline, then a decline and an unordered rise.

In this part, the data was divided into two groups, the first group constitutes the training group for the network, which constitutes 80% of the observations, and the second group is the test group, which constitutes 20% of the data, and the data is converted into values confined to the period [.550,1]. Given that the upper limit of the number of hidden layers is 50 and the

number of inputs is one variable with one time difference, and by its iteration of 250 iterations. And by exporting the daily data of the COVID-19 virus to MATLAB R2020a.

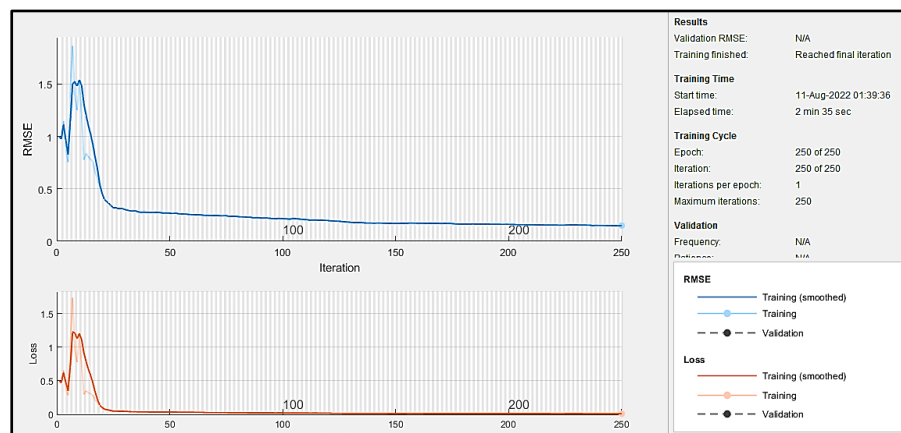


Figure (4.2) Training progress

The previous figure represents the values of the training set and the minimum mean of the squares error for the training set was (0.0359). While the following figure shows the expected and true values of the test set.

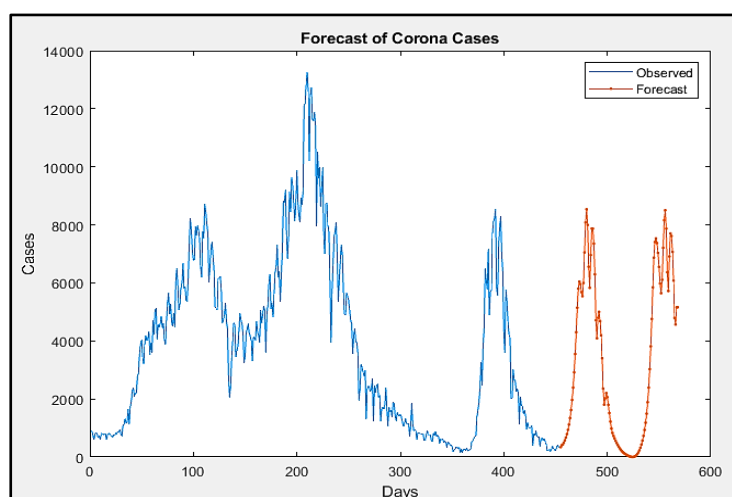


Figure (4.3) Predicted values for the test set

And that the value $AIC = 1.1915e + 03$

$BIC = 1.1956e + 03$

The estimated values for the training set were:

Weeks	Day(1)	Day(2)	Day(3)	Day(4)	Day(5)	Day(6)	Day(7)
Week(1)	1.0e+04	0.0638	0.0864	0.0868	0.0942	0.0976	0.0923
Week(2)	0.0928	0.1093	0.1113	0.1097	0.1066	0.0963	0.0749
Week(3)	0.0645	0.0697	0.0827	0.0876	0.0925	0.0870	0.0677
Week(4)	0.0556	0.0717	0.0855	0.0891	0.0912	0.0915	0.0788
Week(5)	0.0770	0.0906	0.0990	0.1114	0.1218	0.1344	0.1215
Week(6)	0.1142	0.1520	0.1662	0.1869	0.2077	0.2361	0.2394
Week(7)	0.2191	0.2413	0.2816	0.3214	0.3432	0.3907	0.3980

Week(8)	0.3410	0.3272	0.3688	0.4079	0.4298	0.4406	0.4288
Week(9)	0.3670	0.3366	0.3994	0.4174	0.4574	0.4570	0.4724
Week(10)	0.4275	0.3911	0.4518	0.4843	0.4853	0.4835	0.4502
Week(11)	0.4218	0.4151	0.4489	0.5073	0.5523	0.5757	0.5203
Week(12)	0.4666	0.4646	0.5137	0.5586	0.5398	0.5842	0.6028
Week(13)	0.5359	0.4910	0.5548	0.6284	0.6498	0.6623	0.5915
Week(14)	0.5397	0.5496	0.6273	0.7160	0.7954	0.8363	0.7670
Week(15)	0.6388	0.6654	0.7732	0.8326	0.8152	0.8028	0.7612
Week(16)	0.7046	0.6857	0.7511	0.8121	0.8166	0.8584	0.8216
Week(17)	0.7513	0.6125	0.6797	0.7346	0.7303	0.7100	0.6422
Week(18)	0.5528	0.4739	0.5544	0.5809	0.6112	0.5987	0.5667
Week(19)	0.4740	0.4244	0.4997	0.5175	0.5003	0.4192	0.3430
Week(20)	0.2410	0.2673	0.3264	0.3642	0.4089	0.4728	0.4232
Week(21)	0.3106	0.3797	0.4380	0.4410	0.4588	0.4571	0.3994
Week(22)	0.3659	0.3937	0.4229	0.4467	0.4540	0.4654	0.4174
Week(23)	0.3684	0.4100	0.4327	0.4581	0.4575	0.4482	0.4382
Week(24)	0.3912	0.4157	0.4541	0.5075	0.4989	0.4910	0.4673
Week(25)	0.4482	0.4357	0.5053	0.5508	0.5895	0.5970	0.5058
Week(26)	0.4778	0.5236	0.6073	0.6721	0.6830	0.6817	0.5767
Week(27)	0.5625	0.6294	0.7317	0.8025	0.8936	0.8725	0.8046
Week(28)	0.6632	0.7566	0.8355	0.8916	0.8709	0.9134	0.8960
Week(29)	0.7628	0.8677	0.9641	0.9368	0.8641	0.8265	0.8601
Week(30)	0.9024	0.9066	1.1498	1.2555	1.2569	1.3310	1.2892
Week(31)	1.0271	1.2475	1.2548	1.1840	1.1462	1.1766	1.1715
Week(32)	0.7933	1.0461	0.9863	0.9726	0.8941	0.9006	1.0141
Week(33)	0.7817	0.6888	0.8647	0.8791	0.8213	0.7630	0.6424
Week(34)	0.4011	0.5292	0.6882	0.7384	0.7771	0.8143	0.7118
Week(35)	0.5419	0.6066	0.6968	0.7072	0.6864	0.6100	0.4801
Week(36)	0.4828	0.5448	0.5707	0.5534	0.5154	0.4683	0.3859
Week(37)	0.3751	0.4272	0.4576	0.4473	0.3915	0.3417	0.2289
Week(38)	0.2071	0.2864	0.3133	0.3000	0.3057	0.2606	0.1946
Week(39)	0.1941	0.2544	0.2463	0.2292	0.2366	0.2128	0.1760
Week(40)	0.1748	0.2349	0.2242	0.2237	0.2466	0.2114	0.1605
Week(41)	0.1627	0.1938	0.1789	0.1678	0.1914	0.2025	0.1403
Week(42)	0.1340	0.1785	0.1603	0.1498	0.1613	0.1663	0.1294
Week(43)	0.1349	0.1576	0.1474	0.1298	0.1430	0.1400	0.1039
Week(44)	0.1066	0.1276	0.1159	0.1096	0.1187	0.1145	0.0845
Week(45)	0.0755	0.1050	0.1343	0.1168	0.1113	0.0949	0.0661
Week(46)	0.0631	0.0834	0.0761	0.0706	0.0801	0.0728	0.0529
Week(47)	0.0548	0.0692	0.0672	0.0711	0.0798	0.0736	0.0564
Week(48)	0.0588	0.0719	0.0758	0.0717	0.0843	0.0758	0.0578
Week(49)	0.0641	0.0741	0.0711	0.0624	0.0644	0.0532	0.0337
Week(50)	0.0347	0.0458	0.0501	0.0463	0.0477	0.0411	0.0244
Week(51)	0.0190	0.0310	0.0319	0.0308	0.0357	0.0322	0.0218
Week(52)	0.0199	0.0351	0.0364	0.0340	0.0351	0.0296	0.0216
Week(53)	0.0231	0.0367	0.0454	0.0468	0.0580	0.0647	0.0661
Week(54)	0.0782	0.1051	0.1540	0.1976	0.2396	0.2861	0.3478

Week(55)	0.3657	0.4549	0.5353	0.6488	0.6736	0.6145	0.5726
Week(56)	0.5075	0.5983	0.7032	0.8271	0.8401	0.8563	0.6032
Week(57)	0.5250	0.7383	0.8231	0.8045	0.7081	0.6457	0.4524
Week(58)	0.3647	0.5113	0.5632	0.5128	0.4336	0.4068	0.2488
Week(59)	0.1712	0.2652	0.2891	0.2418	0.2557	0.2502	0.1680
Week(60)	0.1652	0.1918	0.1958	0.1570	0.1658	0.1628	0.1186
Week(61)	0.1141	0.1357	0.1248	0.0940	0.1033	0.1074	0.0784
Week(62)	0.0544	0.0822	0.0750	0.0651	0.0759	0.0791	0.0551
Week(63)	0.0411	0.0640	0.0615	0.0460	0.0561	0.0564	0.0358
Week(64)	0.0252	0.0364	0.0291	0.0126	0.0190	0.0271	0.0162
Week(65)	0.0256						

5. Conclusions

- ❖ When drawing the data series used, it turns out that it is an unstable series and there is no clear pattern for it, in addition to the fact that the distribution is close to the exponential function
- ❖ The largest number of recorded infections with the COVID-19 virus during observation was 268 from the data series, where the number of infections reached 13,259.
- ❖ There is no high fluctuation in the intensity of the variance that is known in the time series models with volatility.
- ❖ The time series adopts a pattern with a gradual increase in the number of cases of infection and then a decrease, which made GARCH models less efficient in predicting the number of future cases and instability in the variance of the error term during data analysis using GARCH models.
- ❖ It is noticeable that during the estimation of the parameters of the GARCH models of different orders, it was found that the models GARCH(0,1), GARCH(0,2) and GARCH(1,2) all have the sum of their parameters one, and this gives an undefined (unrealistic) amount of variance.
- ❖ The best model obtained according to AIC and BIC informatics criteria is the GARCH(0,1) model, which we relied on as a predictive model for the time series.
- ❖ Because of the gradual time-series nature and the iterative adaptive state of neural network models, I made them preferred models for describing these data.
- ❖ The least mean square error was obtained using the neural network method.
- ❖ The best way to analyze the COVID-19 time series as a comparison between the parametric method (GARCH models method) and the non-parametric method (Neural network method) is to use the neural network method according to AIC and BIC Informatics criteria.

References

- [1] Engle, R.F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates variance of United Kingdom Inflation. *journal of Econometrica*, **50 (4)**: 987–1008.
- [2] Wiphatthanananthakula, C. and Sriboonchittab, S. (2010). The Comparison among ARMA-GARCH, -EGARCH, -GJR. *The Thailand Econometrics Society journal*, **2 (2)**:140–148.

- [3] Francq, C. and Zakoian, J. M. (2019). GARCH Models: Structure. Statistical Inference and Financial Applications. 2nd edn., John Wiley and Ltd Publication: 492 pp.
- [4] GLOSTEN, L.R. ; JAGANNATHAN, R. and RUNKLE, D.E. (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance*, **48**(5): 1779–1801.
- [5] Nelson, D.B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach”, *Econometrica*, **59** (2): 347-370.
- [6] Mohammad, A.A. and Salim, A.J. (1996). The Analysis and Modeling of the time series of annual mean temprature in Mosul City. *Rafidain Journal of Science*, **7** (1): 37-48.
- [7] Yuan Chen, M. (2013). Time series Analysis: Conditional Volatility models. Dep. of Finance. National chung Hsing University: 1-42 pp.
- [8] Black, F. , (1976) , " Studies of Stock Price Volatility Changes " , In Proceeding of the Meeting of the Business and Economic Statistics Section , American Statistical Association , Washington DC , PP (177-181).
- [9] Makridakis, S. , Wheelwright, S. C. and Hyndman, R. J. , (1998) , " Forecasting Method and Applications " , 3rd ED. , John Wiley and Sons, Inc.
- [10] Pankratz, Alan. Forecasting with univariate Box-Jenkins models: Concepts and cases. John Wiley & Sons, 2009.
- [11] Akaike, H. , (1974) , " A New Look of Statistical Model Identification " , IEEE Transaction on Automatic Control , A C-19 , PP(716-722).
- [12] Fan, J. and Yao, Q. , (2003), " Nonlinear Time Series: Nonparametric and Parametric Methods " , Sippingr-Verlag New York, Inc.
- [13] Graves, Alex, et al. "A novel connectionist system for unconstrained handwriting recognition." *IEEE transactions on pattern analysis and machine intelligence* 31.5 (2008): 855-868.