IRAQI STATISTICIANS JOURNAL

https://isj.edu.iq/index.php/rjes

ISSN: 3007-1658 (Online)

Comparison and Forecasting of Some Nonlinear Models for Time Series Numbers of Baghdad Covid_19 Infections

Jinan Abdullah Anber [1], Mohammed Qadoury Abed [2], Wadhah S. Ibrahim [3,*]

[1]Baghdad Technical College of Management, Deb. Of Information Techniques, Middle Technical University, Baghdad, Iraq.
[2]A l-Mansour University College Dept. of Accounting & Banking Science, Baghdad, Iraq.
[3]College of Management and Economics, Deb. Of Statistic, Mustansiriyah University, Baghdad, Iraq.
[1]E-mail: jinanaa69@mtu.edu.iq , [2]E-mail: mohammed.qadaury@muc.edu.iq ,[3]E-mail: dr_wadhah_stat@uomustansiriyah.edu.iq

ARTICLE INFO

Article history:

Received	20 August 2024
Revised	22 August 2024
Accepted	04 September 2024
Available online	06 September 2024

Keywords:
Time series
logistic model
Chapman-Richard model
Von Bertalanf model
AIC,BIC

ABSTRACT

In this research paper, three nonlinear models of time series models will be addressed, represented in the Logistic model, Chapman Richard model and Von Bertalanff model, which is used to predict the number of people infected with the Covid_19 virus for Baghdad Governorate for the period from (1/10/2021) to (31/12/2021) and compare these models using the criteria (AIC, BIC, H-Q), and that One of the most important conclusions reached was that the Logistic model is the best model to represent the time series of the number of people infected with the Covid_19 virus through statistical criteria (AIC, BIC, H-Q), and there is a great convergence between the actual and estimated numbers of people infected with the Covid_19 virus.

1. Introduction

Previous years witnessed great interest in the subject of forecasting and the emergence of many methods of forecasting, including time series models of linear and nonlinear types, many models have been studied for both types [8], including (The moving average model and the autoregressive mode) [1].

That covid_19 is one of the infectious and dangerous viruses that threaten human life[12][16], and that the covid_19 virus is a health problem and a global pandemic that included all parts of the world without exception[4][19], and Baghdad is one of the cities affected by this pandemic, and the number of infections from the beginning of the pandemic until (31/12/2022) reached approximately (Total cases: 2,465,107 infected, and the number of deaths was approximately (25,373) deaths, and the total number of hospitalizations was approximately (Total Recovered: 2,439,079) cases, as it was discovered for the first time in the Chinese city of Wuhan in December 2019[3]. This virus has turned into a global pandemic affecting many countries, as this virus has spread in many Arab and foreign countries, and the first infection was recorded in Iraq on 24-2-2020, specifically in the Al-Najaf Governorate [2], while in the Baghdad Governorate, the first infection with this was recorded. The virus on 2/27/2020[14], and that the importance of the research was due to the increase in a certain period of the spread of the pandemic of the

* Corresponding author E-mail address: dr_wadhah_stat@uomustansiriyah.edu.iq
https://doi.org/10.62933/302mze77

covid-19 virus in the city of Baghdad, and that the human and material losses caused by this virus, as well as the economic, health and social problems that included all groups of society.

The research problem is that the Covid-19 pandemic has swept the entire world, including the city of Baghdad, and the different numbers between one city and another, as well as the difference between one country and another, have led statisticians to conduct several studies to determine the pattern and behaviour of this virus, and models have been built to predict this phenomenon and to present the behaviour of this virus to decision-makers as a form of future study to confront similar dangers in the future.

From this point, the aim of the research was to study the number of daily infections in Baghdad for the period from (1/10/2021) to (30/12/2021) by developing an appropriate forecast model through a comparison between three Logistic models, Chapman Richard model and Von Bertalanff model.

2. Properties the time series

It is known that the characteristics of the time series must be confirmed in terms of detecting the stability of the time series by the Dickey Fuller test, the linearity of the time series test, as well as the homogeneity test by the standard normal homogeneity test [8].

3. Nonlinear time series models

In this topic, the nonlinear general trend models will be addressed, as some researchers noticed that some data or cases that follow the nonlinear trend cannot use linear models or methods to describe the changes of the time series, as well as it is not possible to find the general trend of the series, and this requires that we search for other models fit this data [17].

4. Logistic Model

This model was used for the first time by the Belgian scientist Verhuist in 1845[10]. The most important characteristic of the simple logistic model is that it is symmetric about the inflection point, and this feature indicates that the process that will occur after the inflection point is a mirror image of the process that occurred after it[11], and expresses the model logistic in the following form [9]:

$$Y_t = \frac{a}{1+\beta e^{-kt}} + e_t \quad \dots (1)$$

Since: Y_t:represents the growth rate (infections), t: the independent variables, (α,β,k): represent the parameters of the model, e_t:the random error.

To estimate the parameters of the logistic model using the greatest possibility method (MLE), as follows [15][18]:

$$L(Y_t) = \frac{\alpha^n}{1+\beta^n e^{-k\sum t}} + \sum e_t \quad \dots (2)$$

$$Ln[L(Y_t)] = nLn(\alpha) - Ln\left[1 + \beta^n e^{-k\sum t}\right] + \sum Ln(e_t) \quad \dots (3)$$

The model function is derived in equation (3) with respect to the parameters (α, β, k), and the derivative is equal to zero, as follows [18]:

$$\frac{\partial Ln[L(Y_t)]}{\partial(\alpha,\beta,k)} = 0 \quad \dots (4)$$

After the derivation, we get a set of equations according to the number of parameters in the model, and since these equations are non-serious equations that are difficult to estimate by the usual methods, therefore we resort to using one of the iterative methods to estimate these parameters as the Newton-Raphson iterative method [7].

5. Chapman-Richard model

The Chapman-Richard model is a generalization of the rest of the nonlinear models such as the logistic model and the von Bertlanvy model. The function of these models is to provide an accurate and more realistic description of the various phenomena. The Chapman-Richard model has been used in many studies, such as studying the factors that control the growth of animals, such as the growth of fish, sheep, cows, horses, and others. This model was chosen to describe the law of spread of infectious diseases and to study the factors that control and affect the spread of the covid-19 virus [22]. This form is expressed in the following form:

$$Y_t = \alpha(1 - \beta e^{-kt})^\theta + e_t \quad \dots (5)$$

Since: Y_t represents the average number of infected individuals. : $(\alpha, \beta, k, \theta)$ represent the parameters of the model.

To estimate the parameters of the Chapman-Richard model using the greatest possibility method (MLE), as follows [23][24]:

$$L(Y_t) = \alpha^n \sum(1-\beta e^{-kt})^\theta + \sum e_t \quad \dots(6)$$

$$Ln[L(Y_t)] = nLn(\alpha) - \theta \sum Ln(1-\beta e^{-kt}) + \sum Ln(e_t) \quad \dots(7)$$

Since: n: maximize (MLE).

We derive the model function in equation (7) for the coefficients, and the derivative is equal to zero, as follows:

$$\frac{\partial Ln[L(Y_t)]}{\partial(\alpha,\beta,k,\theta)} = 0 \quad \dots(8)$$

By the number of derivations, we get a set of equations according to the number of parameters in the model. Since these equations are nonlinear equations, which are difficult to estimate in the usual way, so we resort to using one of the iterative methods to estimate these parameters, such as the Newton-Raphson iterative method, which was relied upon in this research [7].

6. The Von Bertlanffy Model

The Von Bertlanffy model is often used as a growth model. It is mainly used to study the factors that control and influence growth. The development of infectious diseases and the emergence of viruses such as the Corona virus is similar to the growth of individuals and populations. This model was chosen to describe the law of the spread of infectious diseases and to study the factors that control and affect the spread of the covid-19 virus [5]. The model is expressed in the following equation [13]:

$$Y_t = \alpha\left(1-e^{-\beta t}\right) + e_t \quad \dots(9)$$

As:Y_t : Variable response (infections). (α,β): represents the parameters of the model.e_t: random error.

And to estimate the parameters of the Von Bertlanffy model using the greatest possibility method (MLE), as follows [21]:

$$L(Y_t) = a^n\left(1-e^{-\beta t}\right)! + \sum e_t \quad \dots(10)$$

$$Ln[L(Y_t)] = nLn(a) + \sum Ln\left(1-e^{-\beta t}\right) + \sum Ln(e_t) \quad \dots(11)$$

We derive the model function in equation (7) with respect to the parameters (α,β,k,θ), and the derivative is equal to zero, as follows:

$$\frac{\partial Ln[L(Y_t)]}{\partial(\alpha,\beta)} = 0 \quad \dots(12)$$

By the number of derivations, we get a set of equations according to the number of parameters (α,β) in the model. Since these equations are nonlinear, the Newton-Raphson iterative method is used to estimate the model parameters [7].

7. Box-Jenkis methodology

The Box-Jenkis methodology is the process of building time series models and analyzing them to best represent the data. It is represented by several stages first: the diagnostic stage, which studies the shape of the data whether it has a general trend or contains seasonal or cyclical effects, as well as Characteristics of the time series (stability, linearity, and homogeneity). Second: the estimation stage, which is the stage of estimating the parameters of the models that were studied using the Maximum Likelihood method. Third: the stage of selecting the best model. This is done by comparing the models with the test criteria, including (Akanke, Hannan-Quinn, Bayesian Information), Fourth: The stage of testing the validity of the model, where the quality of the selected model is examined through a set of statistical tests, and finally forecasting, which is the main and primary goal of the model estimation process [6],[20].

8. The applied side

For the applied side of the research, a time series of 92 observations was applied, representing the number of daily infections with the Covid-19 virus in Baghdad Governorate, from (1/10/2021) to (31/12/2021). The data were obtained from the Ministry of Health / Baghdad Health Department. The tests, measures and statistical standards mentioned were applied to the time series data of the aforementioned phenomenon, as these data were analyzed and all tests were applied, as well as estimating the parameters of the models used.

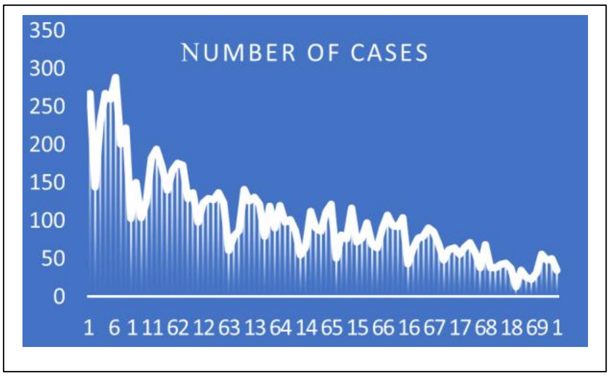

Figure (1) shows the graphic representation of the time series of the numbers of people infected with the Covid-19 virus.

Table(1): Show Augmented Dickey-fuller test

	Lag	ADF	P-value
1	0	-1.364	0.234
2	1	-1.320	0.335
3	2	-0.811	0.432
4	3	-0.577	0.483
5	4	-0.365	0.562

It is clear from Table (1) that the series of numbers of people infected with the Covid_19 virus is unstable, as the P-value values are greater than the significance level (0.05).

From the results reached, the P-value of the Mann-Kendal test is equal to (6.642 E-8), and that the value is smaller than the level of significance (0.05), so the null hypothesis that says that the direction of the time series is linear, that is, there is a trend in the time series and that the time series is non-linear, as well as the P-value of the Van Neumann test is equal to (4.322E-10), which is less than the level of significance (0.05), so it turns out that the data are heterogeneous.

Now the Maximum Likelihood method is used to estimate the parameters of the nonlinear models that were mentioned in the theoretical aspect, and the results were as follows:

Table (2): Show the estimating parameters of Non-Linear Models.

	Parameter	Estimator	Std.Error	t-value	P-value
Logistic model	α	204.341	8.42353	21.668	2.56E-12
	β	22.2404	14.6786	2.5664	1.89E-04
	K	0.10910	0.02272	5.4362	3.07E-03
Chapman Richard model	α	211.435	10.4523	21.567	2.03E-13
	β	0.06456	6.44627	0.1028	8.23E-02
	K	0.08435	0.41285	2.0129	7.45E-02
	θ	131.867	20.4537	0.0782	6.99E-01
Van Bertalanffy model	α	211.984	14.3244	15.023	2.33E-15
	β	0.05678	0.00668	4.6572	5.63E-06

Table (2) can build estimation forms, the formulas of which are as follows:

$$\hat{Y}_{t(Log)} = \frac{204.341}{1 + 22.2404\, e^{-0.1091\, t}} \quad \ldots (13)$$

$$\hat{Y}_{t(CR)} = 211.435\,(1 - 0.06456\, e^{-0.08435\, t})^{131.867} \quad \ldots (14)$$

$$\hat{Y}_{t(VB)} = 211.984\,(1 - e^{-0.05678\, t}) \quad \ldots (15)$$

It is noted from Table (2) that the p_value values of the estimated features for all models are less than the level of significance (0.05), so we reject the null hypothesis and that the parameters of all models are statistically significant.

Using Forms (13), (14) and (15), the estimated values of the infected population shown in Table (3) are reached.

Table (3): Show the estimated values for the models Logistic, Chapman-Richard and Von Bertlanffy

T	$\hat{Y}_{t(Log)}$	$\hat{Y}_{t(CR)}$	$\hat{Y}_{t(VB)}$	t	$\hat{Y}_{t(Log)}$	$\hat{Y}_{t(CR)}$	$\hat{Y}_{t(VB)}$	T	$\hat{Y}_{t(Log)}$	$\hat{Y}_{t(CR)}$	$\hat{Y}_{t(VB)}$
1	204.144	210.669	210.842	32	198.707	201.201	205.345	63	111.385	107.162	173.389
2	204.122	210.602	210.775	33	198.077	200.325	204.957	64	105.835	100.924	171.135
3	204.096	210.528	210.705	34	197.380	199.375	204.547	65	100.263	94.546	168.748
4	204.068	210.449	210.630	35	196.608	198.347	204.112	66	94.703	88.059	166.222
5	204.037	210.362	210.551	36	195.754	197.234	203.652	67	89.187	81.503	163.549
6	204.002	210.268	210.467	37	194.810	196.031	203.166	68	83.747	74.920	160.719
7	203.963	210.166	210.379	38	193.768	194.729	202.650	69	78.412	68.355	157.724
8	203.919	210.054	210.285	39	192.619	193.323	202.105	70	73.211	61.860	154.554
9	203.871	209.933	210.185	40	191.354	191.805	201.528	71	68.169	55.487	151.199
10	203.817	209.802	210.080	41	189.962	190.167	200.917	72	63.306	49.292	147.647
11	203.757	209.658	209.969	42	188.434	188.400	200.271	73	58.640	43.329	143.889
12	203.689	209.503	209.851	43	186.758	186.496	199.586	74	54.186	37.653	139.910
13	203.615	209.334	209.727	44	184.923	184.446	198.862	75	49.955	32.313	135.699
14	203.531	209.150	209.595	45	182.920	182.242	198.095	76	45.953	27.354	131.243
15	203.438	208.950	209.455	46	180.736	179.873	197.284	77	42.184	22.813	126.526
16	203.335	208.732	209.308	47	178.361	177.330	196.425	78	38.648	18.720	121.533
17	203.219	208.496	209.151	48	175.784	174.604	195.516	79	35.345	15.090	116.248
18	203.091	208.239	208.986	49	172.997	171.685	194.554	80	32.268	11.931	110.655
19	202.948	207.961	208.811	50	169.992	168.564	193.536	81	29.413	9.236	104.735
20	202.788	207.657	208.625	51	166.760	165.232	192.458	82	26.771	6.987	98.469
21	202.611	207.328	208.429	52	163.298	161.682	191.317	83	24.334	5.153	91.838
22	202.413	206.971	208.221	53	159.603	157.905	190.110	84	22.090	3.698	84.818
23	202.193	206.582	208.001	54	155.674	153.895	188.832	85	20.031	2.574	77.389
24	201.949	206.160	207.769	55	151.514	149.647	187.479	86	18.145	1.734	69.525
25	201.677	205.703	207.523	56	147.130	145.157	186.048	87	16.420	1.127	61.203
26	201.374	205.205	207.262	57	142.530	140.423	184.532	88	14.846	0.704	52.394
27	201.037	204.666	206.986	58	137.727	135.447	182.928	89	13.412	0.421	43.070
28	200.663	204.081	206.694	59	132.739	130.232	181.231	90	12.108	0.240	33.201
29	200.248	203.446	206.385	60	127.586	124.784	179.434	91	10.924	0.130	22.756
30	199.787	202.757	206.058	61	122.291	119.113	177.533	92	9.849	0.066	11.701
31	199.275	202.011	205.712	62	116.881	113.233	175.520				

Table (3) shows that the values estimated by the method of the greatest possibility Logistic and Chapman Richard models are closer to the real values than the Von Bertalanffy model and can be seen from the graph of the estimated values of the non-models and the values of the dependent variable Y as in Figure (2).

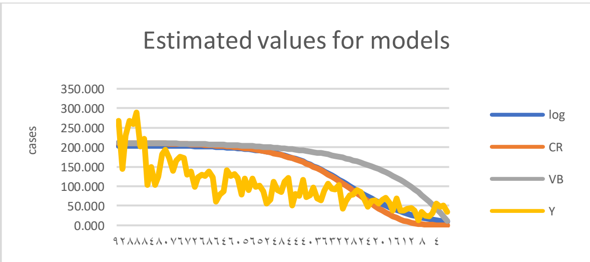

Figure (2) shows the estimated number of infected people for the three models with the original values of the number of infected with the Covid_19 virus

The stage of selecting the best model among three models of the series data for the number of people infected with the Covid_19 virus is done through statistical criteria (AIC, BIC, H-Q), which showed that the Logistic model was better than such as the time series of the number of infected and then the Chapman Richard model, which comes in second place as shown in Table (4).

Table (4): Show the AIC, BIC and H-Q of the models.

	Models		
	Log.	C. R.	V. B.
AIC	1012.163	1016.056	1030.345
BIC	1022.643	1030.345	1039.045
H-Q	1016.723	1023.321	1033.678

* Corresponding author E-mail address: dr_wadhah_stat@uomustansiriyah.edu.iq
https://doi.org/10.62933/302mze77

9. Conclusions:

1- The series of numbers of people infected with the Covid_19 virus represents an unstable, non-linear and heterogeneous chain according to the drawing of the chain and some tests.

2- The Logistic model is the best model to represent the time series of the number of people infected with the Covid_19 virus through statistical standards (AIC, BIC, H-Q).

3- There is a great convergence between the abacus and the estimated number of people infected with the Covid_19 virus, obtained from the Logistic and Chapman Richard models.

References:

[1] Al-Nassir, A.-M. H., Abed, M. Q. ., & Ibraheem, W. S. (2014). Estimating the Spectral Power Density Function of Non-Gaussian Second Order Autoregressive Model. *Journal of Asian Scientific Research*, 4(9), 513–521.

[2] Al-Saffar, R. S., Rasheed, A. A. R., & Sadik, N. J. (2021). Nonlinear modeling of the factors affecting the disease Covid-19 by adopting the method of sampling by reference. *International Journal of Agricultural and Statistical Sciences*, 2301-2307.

[3] Abbas Gulmurad Beg Murad, Sozan Saber Haider Ali, (2023) ANALYSIS OF PROBABILITY DISTRIBUTION FOR CORONA VIRUS MORTALITY IN IRAQ BEFORE AND AFTER THE DISCOVERY OF THE EPIDEMIOLOGICAL VACCINES. *Journal of Northeastern University*, 26(1), 514-533.

[4] Aljabri, M., Aljameel, S. S., Khan, I. U., Aslam, N., Charouf, S. M. B., & Alzahrani, N. (2022). Machine learning model for sentiment analysis of COVID-19 tweets. *Int J Adv Sci Eng Inf Technol*, *12*(3), 1206-14.

[5] Amaral, M. T. R., Conceição, K. S., de ANDRADE, M. G., & Padovani, C. R. (2020). Generalized growth curve model for covid-19 in brazilian states. *Brazilian Journal of Biometrics, 38(2), 125-146.*

[6] Bozdogan, H.(2000), Akaike's Information Criterion and Recent Developments in Information Complexity., *Journal of Mathematical Psychology, Vo. 44, Issue 1, P.p 62-91*

[7] Clark, A. J., Lake, L. W., & Patzek, T. W. (2011, October). Production forecasting with logistic growth models. *In SPE Annual Technical Conference and Exhibition? (pp. SPE-144790). SPE.*

[8] Ibraheem, W. S. (2011). Adaptive Filtering Method using of Non-Gaussian Moving Average model from First order (simulation study). *Journal of Al-Rafidain University College For Sciences (Print ISSN: 1681-6870, Online ISSN: 2790-2293)*, (2), 79-93.

[9] Ibrahim, W. S. (2020). To determine the most important factors affecting pancreatic cancer in Iraq using the logistic regression model. *Periodicals of Engineering and Natural Sciences (PEN), 8(1), 277-283.*

[10] Ibrahim, W. S., & Hajar, F. T. (2019). Use the logistic regression model to find the most important factors affecting lung cancer in Iraq in 2017. *Journal of Administration and Economics, (121).*

[11] Ibrahim, W. S., Majeed, G. H., & Anber, J. A. (2022). Determine the most important factors affecting breast cancer in Iraq. *Periodicals of Engineering and Natural Sciences (PEN), 10(2), 260-266.*

[12] Islam, L., Rahman, M., Ahmad, N., Sharmin, T., & Sorna, J. F. (2021). A Crowdsourced Contact Tracing Model to Detect COVID-19 Patients using Smartphones. *arXiv preprint arXiv:2112.01244.,, Vol. 11 (2021).No.5,P:1818-1824, DOI:10.18517 /ijaseit.11.5.14644*

[13] Katsanevakis, S., & Maravelias, C. D. (2008). Modelling fish growth: multi-model inference as a better alternative to a priori using von Bertalanffy equation. *Fish and fisheries, 9(2), 178-187.*

[14] Jbarah, A., & AL-Saffar, R. S. (2021). DETERMINE THE MOST IMPORTANT FACTORS AFFECTING THE CORONA VIRUS TO BUILD A MODEL USING THE MAIN COMPOUNDS. *International Journal of Agricultural & Statistical Sciences, 17.*

[15] Mahdi, N. N., Abed, H. T., & Sadik, N. J. (2020). The discriminant analysis in the evaluation of cancers diseases in Iraq. *Int J Adv Sci Eng Inf Technol*, *10*(5), 2170-2176.

[16] Martínez, F., Martínez, F., & Jacinto, E. (2020). Performance evaluation of the NASNet convolutional network in the automatic identification of COVID-19. *International Journal on Advanced Science, Engineering and Information Technology*, *10*(2), 662.

[17] Mohammad, A. J. (2021). Using the penalized splines regression method to estimate the logistic model and using a practical application. *journal of the college of basic education, 27(113).*

[18] Mohammed, A. J., & Ibrahim, W. S. (2020). Study of the probability of change in share prices based on the value of the Babylon Hotel using logistic regression. *Journal of Administration and Economics, (123).*

[19] Na'am, J., Pranata, F. S., Hidayat, R., Adif, A. M., & Ellyzarti, E. (2021). Automated Identification Model of Ground-Glass Opacity in CT-Scan Image by COVID-19. *International Journal on Advanced Science, Engineering and Information Technology*, *11*(2), 595-602.

[20] Rasheed, A. A. R., Abd, H. T., & Sadik, N. J. (2022). Forecasting the Number of (Infected, Recovered, Death) For COVID-19 in Iraq by Using Neural Network Models. *Mathematical Statistician and Engineering Applications*, *71*(4), 4984-4996.

[21] Ratkowsky, D. A. (1986). Statistical properties of alternative parameterizations of the von Bertalanffy growth curve. *Canadian Journal of Fisheries and Aquatic Sciences, 43(4), 742-747.*

[22] Reddy, T., Shkedy, Z., Janse van Rensburg, C., Mwambi, H., Debba, P., Zuma, K., & Manda, S. (2021). Short-term real-time prediction of total number of reported COVID-19 cases and deaths in South Africa: a data driven approach. *BMC medical research methodology, 21, 1-11.*

[23] Trappey, C. V., & Wu, H. Y. (2008). An evaluation of the time-varying extended logistic, simple logistic, and Gompertz models for forecasting short product lifecycles, *Advanced Engineering informatics, 22(4). 421-430.*

[24] Wang, X. S., Wu, J., & Yang, Y. (2012). Richards model revisited: Validation by and application to infection dynamics. *Journal of theoretical biology, 313, 12-19.*