Ban Ghanim Al-Ani*

Statistical modeling of the novel COVID-19 epidemic in Iraq

https://doi.org/10.1515/em-2020-0025
Received July 24, 2020; accepted June 24, 2021; published online July 22, 2021

Abstract

Objectives: This study aimed to apply three of the most important nonlinear growth models (Gompertz, Richards, and Weibull) to study the daily cumulative number of COVID-19 cases in Iraq during the period from 13th of March, 2020 to 22nd of July, 2020.
Methods: Using the nonlinear least squares method, the three growth models were estimated in addition to calculating some related measures in this study using the "nonlinear regression" tool available in Minitab-17, and the initial values of the parameters were deduced from the transformation to the simple linear regression equation. Comparison of these models was made using some statistics (*F*-test, AIC, BIC, AIC_c and WIC).
Results: The results indicate that the Weibull model is the best adequate model for studying the cumulative daily number of COVID-19 cases in Iraq according to some criteria such as having the highest F and lowest values for RMSE, bias, MAE, AIC, BIC, AIC_c and WIC with no any violations of the assumptions for the model's residuals (independent, normal distribution and homogeneity variance). The overall model test and tests of the estimated parameters showed that the Weibull model was statistically significant for describing the study data.
Conclusions: From the Weibull model predictions, the number of cumulative confirmed cases of novel coronavirus in Iraq will increase by a range of 101,396 (95% PI: 99,989 to 102,923) to 114,907 (95% PI: 112,251 to 117,566) in the next 24 days (23rd of July to 15th of August 15, 2020). From the inflection points in the Weibull curve, the peak date when the growth rate will be maximum, is 7th of July, 2020, and at this time the daily cumulative cases become 67,338. Using the nonlinear least squares method, the models were estimated and some related measures were calculated in this study using the "nonlinear regression" tool available in Minitab-17, and the initial values of the parameters were obtained from the transformation to the simple linear regression model.

Keywords: COVID-19; Gompertz model; growth models; non-linear regression; Richards' model; Weibull model.

Introduction

The coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-COV-2) which is one of the biggest public health crises the world has ever faced. In this context, it is important to have effective models to describe the different stages of the epidemic's evolution in order to guide the authorities in taking appropriate measures to fight the disease. Generally, there are three kinds of methods to study the infectious of diseases. (i) Dynamic model establishing of infectious diseases; (ii)

***Corresponding author**: **Ban Ghanim Al-Ani**, Department of Statistics and Informatics, University of Mosul, Mosul, Iraq,
E-mail: drbanalani@uomosul.edu.iq

statistical modeling building based on random process with analyzing of time series and other statistical methods; (iii) using data mining methodology to obtain the information in the data and then find the epidemic law of infectious diseases (Jiang, Zhao, and Cao 2020).

The researchers have sought understanding of (COVID-19), and many of them undertaken statistical models. And because the disease started to spread in China, so the first studies in this field was carried out in China. A Markov Chain Monte Carlo (MCMC) stochastic process is used to evaluate the coronavirus transmissibility in China with using the logistic model (Shen 2020). Majumder and Mandl (2020) studied the incidence decreasing of COVID-19 using exponential adjustment model in Wuhan. Some of the researchers adopted the exponential growth model for (SARS) using data-driven analysis in the early phase of the outbreak in China (Zhao et al. 2020). Generation of short-term forecasts for cumulative number of COVID-19 cases by using some of the nonlinear regression models in China (Roosa et al. 2020a). One of the statistical models has developed a "susceptible, un-quarantined infected, quarantined infected and confirmed infected" (SUQC) model in order to characterize the dynamics of outbreaks (Zhao and Chen 2020). Forecasts of the COVID-19 epidemic in Guangdong and Zhejiang, in China were generated using Richards' growth and a sub-epidemic wave models (Roosa et al. 2020b).

We prefer to use such a growth models over other epidemiological models like SIR due to its simplicity and for other many reasons like, firstly, the SIR is a Compartmental Model and the data related to each part not available in such a country like Iraq due to absence of strategic and scientific planning in many governmental sectors. Secondly, the SIR model assumes homogeneous mixing of the population, meaning that all individuals in the population are assumed to have an equal probability of coming in contact with one another. This does not reflect human social structures, in which the majority of contact occurs within limited networks. The SIR model also assumes a closed population with no migration, births, or deaths from causes other than the epidemic (Tolles and Luong 2020).

The use of models in public health decision making has become increasingly important in the study of the spread of disease, designing interventions to control and prevent further outbreaks, and limiting their devastating effects on a population. Iraq today reported over 101,258 cases with 4,122 deaths since the start of the COVID-19 outbreak in the country in February 22nd, 2020. The main contribution of this work is that it is very important for health authorities to know future expectations of the numbers of disease cases in order to use the available capabilities that prevent the worsening of the pandemic, and this work can be considered as the basis to comprehensive studies of this disease that deals with deaths, the necessary laboratory tests, and building and equipping Hospitals for this purpose. In addition, the researcher did not find any work dealing with mathematical and statistical modeling for Corona virus infections in Iraq.

The objective of this study is to describing of well-known growth models to a large extent with application to the daily cumulative numbers of confirmed cases of infection by the novel coronavirus for the interval from March 13th, 2020 to July 22nd, 2020.

Materials and methods

Statistical models

In many study fields, the growth models had played significant role, where many researchers have contributed in developing relevant models. There are several common models such as Gompertz, Weibull, negative exponential, Richards, logistic, monomolecular, Brody, Mitcherlich, von Bertalanffy, S-shaped curves, etc. There are about 77 equations with the associated parameter meanings, these models (or curves) referred to as Sigmoidal Growth Models which arise in various applications including diseases epidemic, bioassay, agriculture, engineering field, tree diameter, height distribution in forestry (Dagogo, Nduka, and Ogoke 2020). Nonlinear statistical models have been used to describe growth behavior, as it varies in time. The type of model needed in a specific area and specific situation depends on the type of growth that occurs. A nonlinear model is one in which at least one of the parameters appears nonlinearly. Three of the above models often used especially for the study of growth curves in diseases epidemic and outbreak studies were analyzed: Gompertz, Richards, and Weibull. The formulas of these models are showed in Table 1. In all concerned models, y_t stands for COVID-19 cumulative cases recorded at time t, t stands for the time index

Table 1: Nonlinear growth models presented in the study.

Model	Equation	
Gompertz	$y_t = \theta_1 e^{-\theta_2 e^{-\theta_3 t}} + \varepsilon_t$	(1)
Richards'	$y_t = \frac{\theta_1}{\left(1+\theta_2 e^{-\theta_3 t}\right)^{1/\theta_4}} + \varepsilon_t$	(2)
Weibull	$y_t = \theta_1 - \theta_2 e^{-\theta_3 t^{\theta_4}} + \varepsilon_t$	(3)

$(t = 1, 2, 3, \ldots, \mathrm{n})$, θ_1 represent maximum value (asymptotic value) of y_t when time (t) approaches $+\infty$, θ_2 is the scale parameter, θ_3 is the shape parameter that is the intrinsic growth rate representing growth rate, θ_4 is the inflection parameter, which determines the function shape, and ε_t is a random error term such that $\varepsilon_t \sim \mathrm{NID}(0, \sigma_\varepsilon^2)$. Note that all the parameters of these models are of positive values.

Gompertz model: The Gompertz model is a type of mathematical model for a time series, named after Benjamin Gompertz (1779–1865). Gompertz function describes growth that starts and ends slow of a given time period. The right-hand value (future value) asymptotic of the function is approached much more gradually by the curve than the left-hand valued (lower value) asymptotic. It is a special case of the generalized logistic model (GLM) (Draper and Smith 1981). For t=0, the initial value of y_t is $y_0 = \theta_1 e^{-\theta_2}$, and as $t \to +\infty$, $y_t \to \theta_1$ (the upper limit to growth).

Richards' model: The Richards' model also known as generalized logistic model, sometimes named a "Richards' curve" after F. J. Richards, who proposed the general form for the family of models in 1959 (Archontoulis and Miguez 2015). For t=0, the initial value of y_t is $y_0 = \frac{\theta_1}{\sqrt[\theta_4]{1+\theta_2}} e^{-\theta_2}$, and as $t \to +\infty$, $y_t \to \theta_1$ (the upper limit to growth).

Weibull model: The Weibull model was first introduced by Waloddi Weibull (1951), which was initially described as a statistical distribution. It has many applications in population growth, agricultural growth and is also used to describe survival in cases of injury or disease or in population dynamic studies (Mahanta and Borah 2014). For t=0, the initial value of y_t is y_0=θ_1− θ_2, and as $t \to +\infty$, $y_t \to \theta_1$ (the upper limit to growth). The source of Eq. (3) is an extension of the Weibull cumulative distribution function:

$$F\left(t; \theta_2, \theta_4\right) = 1 - e^{-(t/\theta_2)^{\theta_4}} \quad (4)$$

as a less restrictive upper limit to growth, "1" is replaced by θ_1; that is, $\lim_{t\to\infty} y_t = \theta_1$ hence θ_1 is termed the limit to growth parameter.

The complete derivation of the above three models is given by (Tran 2017).

Inflection points

The mathematical definition of inflection point is: inflection point of a continuous function $f(t)$ is a point t=a, on an open interval containing point t=a where the second derivative $f''(t) < 0$ on one side and $f''(t) > 0$ on the other side of t=a, and $f''(a)$ is either 0 or does not exist. In practice, inflection point is the point at which the rate of growth gets maximum value. There are some interesting applications and practical uses of inflection points in areas including demography, economics, computer science, diseases epidemic, animal science, plant science, forestry and biology (Goshu and Koya 2013). The derivation of inflection points for above three models are shown in Appendix A1.

Model assumptions

In nonlinear models as in linear models, three main assumptions related to the model errors must be tested: errors are independent normally distribution with common variance. Deviations from the assumptions could result in bias (inaccurate estimates), distorted standard errors, or both (Ritz and Streibig 2008). Violations of these assumptions can be detected from analysis of the residuals by graphical procedures and statistical tests.

Practically, to test whether the errors are follow normal distribution, (p–p) plot procedure can be used. The probability–probability (p–p) plot is a graph of the model residuals values plotted against the normal CDF values. It is used to determine how well a normal distribution fits to the residuals. This draw will be approximately linear if the normal distribution is the correct model. The standardized residual plot is commonly applied (Pinheiro and Bates 2000). The extreme values or outliers are common causes for deviations from normality. Also one of the frequent tests can be used, such as Anderson–Darling (AD) procedure uses the cumulative distribution function to test if a data set comes from a specified distribution or not by the following formula:

$$\mathrm{AD} = -\mathrm{n} - \frac{1}{\mathrm{n}} \sum_{i=1}^{\mathrm{n}} (2i-1)\left[\ln F(x_i) + \ln\left(1 - F\left(x_{\mathrm{n}-i+1}\right)\right)\right] \tag{5}$$

where $F(x)$ is the cumulative distribution function for the specified distribution and i=the ith sample when the data are sorted in ascending order. p-Value was given when running the software that compared with 5% level of significance (Miller, Vandome, and McBrewster 2011). It is important to refer that in dynamic models especially when the data is of the type counts, the more appropriate distribution of residuals is Poisson and some previous studies have incorporated a Poisson error structure via parametric bootstrapping (Chowell 2017).

Homogeneity of variance can be detected by looking at the plot of the explanatory variable over the standardized residuals, when there is a trend (e.g., increasing variability as the explanatory variable increases), this means that the residuals variance is a function of the explanatory variable. If variance heterogeneity is ignored, the parameter estimates might not be influenced much, but this may result in severely misleading confidence and prediction intervals (Carroll and Ruppert 1988).

The residuals are assumed to be independent, and when this assumption is violated it is visually evident in a plot of correlations of residuals against "lag" (or units of separation in time or space), or by using the Ljung–Box Q test (sometimes called the Portmanteau test) is used to test whether or not errors over time are random and independent. The test statistic given by:

$$Q_k = \mathrm{n}(\mathrm{n}+2) \sum_{j=1}^{k} \frac{\hat{\rho}_j^2}{\mathrm{n}-j} \tag{6}$$

where $\hat{\rho}_j$ is the estimated autocorrelation of the series at lag j among fitted model residuals $e_1, e_2, \ldots, e_{\mathrm{n}}$, such that:

$$\hat{\rho}_j = \frac{\left(\frac{1}{\mathrm{n}-j}\right) \sum_{i=1}^{\mathrm{n}-j} e_i e_{i+1}}{\left(\frac{1}{\mathrm{n}}\right) \sum_{i=j}^{\mathrm{n}} e_i^2}, \quad j = 1, 2, \ldots, k$$

$Q_k \sim \chi^2_{k-\mathrm{p}}$ and k is the number of lags being tested such that $k \leq 0.5\mathrm{n}$, usually $k \approx 20$. Some packages will give the Q_k statistic for several different values of k (Brockwell and Davis 2016).

Model estimation

The frequently methods for the estimation of nonlinear models parameters are nonlinear ordinary least squares (OLS) which minimizes the sum of squared error of estimated model and the maximum likelihood method (ML), which searches to find the probability distribution that makes the actual data most likely. These methods are used in many statistical package software like: MatLab, GenStat, SAS, Minitab, R, JMP, Sigmaplot, OriginLab, and SPSS. In general, when the response variable data are not follows a normal distribution then the estimation results from (OLS) and (ML) methods will be different. While when the data are normally distributed, the estimates are approximately identical (Myung 2003). The main algorithms that implemented in estimation methods belong to local optimization, like the Nelder–Mead, Gauss–Newton, and Newton–Raphson algorithms. Local optimization algorithms are sensitive to the initial values of the model parameters. The convergence often failed due to wrong choosing of initial values (Archontoulis and Miguez 2015).

The general form of the growth models or nonlinear models is:

$$y_t = f(t; \boldsymbol{\theta}) + \varepsilon_t, \quad t = 1, 2, \ldots, \mathrm{n} \tag{7}$$

where:

y_t is the dependent or response variable,
t is the time (independent variable),
$\boldsymbol{\theta}$ is the vector of unknown p-parameters such that $\boldsymbol{\theta} = \left(\theta_1, \theta_2, \ldots, \theta_{\mathrm{p}}\right)'$,
ε_t is a random error term and $\varepsilon_t \sim \mathrm{NID}(0, \sigma_\varepsilon^2)$.

The nature relation between y_t and t is not linear, and the goal is to estimate $\theta_j's$ by nonlinear (OLS) which minimizing the sum of squares residual ($\mathrm{SS_{Res}}$) function:

$$\mathrm{SS_{Res}} = \sum_{i=1}^{\mathrm{n}} \left[y_t - f(t; \boldsymbol{\theta})\right]^2 \tag{8}$$

When the values of $\boldsymbol{\theta}$ estimates are substituted into Eq. (8) this makes the SS_{Res} a minimum, and then $\boldsymbol{\theta}$ can be founded by $\frac{\partial SS_{Res}}{\partial \theta} = 0$, this provides the p-normal equations that must be solved for $\hat{\boldsymbol{\theta}}$. The estimation steps are shown in Appendix A2.

Initial values of parameters

The most difficult problems encountered in estimating parameters of nonlinear models is the starting or initial value specification (Fekedulegn, Mac Siurtain, and Colbert 1999). However, the problem of specifying initial values of parameters can be solved with proper understanding of the definition of the parameters in the context of the phenomenon being modelled. Wrong starting values may be led to non-convergence of the parameters and SS_{Res}. Regarding of the selection the initial values of parameters, there are some practical methods (Archontoulis and Miguez 2015):

Use information from the literature when the model has parameters with meaning related to the studied phenomena.
Use graphical representation of the data.
Nonlinear model transformation to linear model.
Use pre-specified algorithms.

All the iterative methods like Levenberg–Marquardt method requires that starting or an initial value for each parameter be estimated of θ_1, θ_2, θ_3 and θ_4. In presented growth models, the parameter θ_1, which is simply to determine, is defined as the maximum possible value of the dependent variable. Therefore, in modelling of the COVID-19 epidemic, θ_1 was specified as the maximum value of COVID-19 cumulative cases. The derivation of parameters initial values is shown in Appendix A3.

Model selection criteria

When we are fitting several models to certain sample data and the aim is to select the preferable model among these models, we use *F*-test such that:

F-test:

$$F = \frac{\mathrm{MSR}}{\mathrm{MSE}} = \frac{SS_{Reg}/\mathrm{p}}{SS_{Res}/\mathrm{n-p-1}} \tag{9}$$

where SS_{Reg} is the sum of squared regression and SS_{Res} is the sum of squared residuals (errors), such that:

$$SS_{Reg} = \sum_{t=1}^{\mathrm{n}} (y_t - \bar{y})^2 - \sum_{t=1}^{\mathrm{n}} (y_t - \hat{y}_t)^2 \tag{10}$$

$$SS_{Res} = \sum_{t=1}^{\mathrm{n}} e_t^2 = \sum_{t=1}^{\mathrm{n}} (y_t - \hat{y}_t)^2 \tag{11}$$

A significant and larger value of *F* indicates a preferable model.

In order to obtain a more complete evaluation of the performance of the models, three additional criteria based on the information theory were applied to compare the models: the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), such that (Teleken, Galvão, and Robazza 2017):

$$\mathrm{AIC} = \mathrm{n} \ln (SS_{Res}/\mathrm{n}) + 2\mathrm{p} \tag{12}$$

$$\mathrm{BIC} = \mathrm{n} \ln (SS_{Res}/\mathrm{n}) + \mathrm{n} \ln (\mathrm{n}) \tag{13}$$

A smaller value of AIC and BIC criteria indicate a preferable model, and if $\mathrm{n/p} < 40$ then the AIC might not be accurate, therefore the corrected AIC ($\mathrm{AIC_c}$) was used, such that (Burnham and Anderson 2002):

$$\mathrm{AIC_c} = \mathrm{AIC} + \frac{2\mathrm{p}(\mathrm{p}+1)}{\mathrm{n-p-1}} \tag{14}$$

The weighted average information criterion (WIC) (Rinke and Sibbertsen 2016):

$$\mathrm{WIC} = \mathrm{n} \ln (\mathrm{MSE}) + \frac{[2\mathrm{n}(\mathrm{p}+1)/(\mathrm{n-p-2})]^2 + [\mathrm{p} \ln (\mathrm{n})]^2}{[2\mathrm{n}(\mathrm{p}+1)/(\mathrm{n-p-2})] + \mathrm{p} \ln (\mathrm{n})} \tag{15}$$

Goodness of fit

There is no criterion or single method to best assess the goodness of fit, but there are many different methods (graphical and numerical) that highlight different features of the data and the model. Graphical comparison provides a quick visual assessment of the goodness of fit. Numerical statistical indices like: bias, mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), concordance correlation, and others. In our study, we use:

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} = \sqrt{\frac{\mathrm{SS_{Res}}}{\mathrm{n-p-1}}} \tag{16}$$

$$\mathrm{Bias} = \frac{1}{\mathrm{n}}\sum_{t=1}^{\mathrm{n}} e_t \tag{17}$$

$$\mathrm{MAE} = \frac{1}{\mathrm{n}}\sum_{t=1}^{\mathrm{n}} |e_t| \tag{18}$$

In nonlinear models analysis, it is important to test the hypotheses about the models' parameters by evaluating the 95%$(1-\alpha)$ C.I. of these parameters. This approach is completely different from linear models analysis. Our hypothesis $H_0: \theta_j=0, j=1, 2, \ldots, \mathrm{p}$ was rejected when the C.I. of θ_j does not include zero, in this case the parameter estimator of the fitted model are statistical significant at 5% level (Fekedulegn, Mac Siurtain, and Colbert 1999).

Data

The daily confirmed number of cumulative COVID-19 cases for four months ago starts from March 13, 2020 (which the first COVID-19 case is recorded in Iraq) to July 22, 2020 were taken from the website of Public Health Directorate (PHD) at Iraqi Ministry of Health http://phd.iq/CMS.php?CMS_P=293. The cumulative number of confirmed cases has reached 101,258 in Iraq as shown in Table 2.

In addition, Figure 1 shows the confirmed daily cases and tests for COVID-19 in Iraq during April, 6th to July, 22nd, 2020, where the daily rate of cases reached (949), with an increasing rate of about (20) cases/day, while the daily average of tests reached (7,958), with an increasing rate of about (150) tests/day.

Table 2: Confirmed cumulative daily cases of COVID-19 in Iraq for Mar. 13 to Jul. 22, 2020.

Date	Cases	Date	Cases	Date	Cases	Date	Cases	Date	Cases	Date	Cases
13/3	15	04/4	857	26/4	1743	18/5	3,507	09/6	15,310	01/7	53,604
14/3	29	05/4	927	27/4	1824	19/5	3,620	10/6	16,571	02/7	55,916
15/3	35	06/4	1,018	28/4	1899	20/5	3,773	11/6	17,666	03/7	58,250
16/3	56	07/4	1,098	29/4	1981	21/5	3,860	12/6	18,846	04/7	60,375
17/3	62	08/4	1,128	30/4	2049	22/5	4,168	13/6	20,105	05/7	62,171
18/3	73	09/4	1,175	01/5	2,155	23/5	4,365	14/6	21,211	06/7	64,597
19/3	88	10/4	1,214	02/5	2,192	24/5	4,528	15/6	22,596	07/7	67,338
20/3	109	11/4	1,248	03/5	2,242	25/5	4,744	16/6	24,150	08/7	69,508
21/3	128	12/4	1,274	04/5	2,327	26/5	5,031	17/6	25,613	09/7	72,356
22/3	161	13/4	1,296	05/5	2,376	27/5	5,353	18/6	27,248	10/7	75,090
23/3	211	14/4	1,311	06/5	2,439	28/5	5,769	19/6	29,118	11/7	77,402
24/3	241	15/4	1,330	07/5	2,499	29/5	6,075	20/6	30,764	12/7	79,631
25/3	277	16/4	1,378	08/5	2,575	30/5	6,335	21/6	32,572	13/7	81,653
26/3	353	17/4	1,409	09/5	2,663	31/5	6,764	22/6	34,398	14/7	83,863
27/3	401	18/4	1,435	10/5	2,714	01/6	7,283	23/6	36,598	15/7	86,144
28/3	442	19/4	1,470	11/5	2,809	02/6	8,064	24/6	39,035	16/7	88,167
29/3	525	20/4	1,498	12/5	2,928	03/6	8,736	25/6	41,089	17/7	90,216
30/3	590	21/4	1,527	13/5	3,039	04/6	9,742	26/6	43,158	18/7	92,526
31/3	624	22/4	1,573	14/5	3,089	05/6	10,994	27/6	45,298	19/7	94,689
01/4	668	23/4	1,604	15/5	3,156	06/6	12,262	28/6	47,047	20/7	96,803
02/4	716	24/4	1,659	16/5	3,300	07/6	13,377	29/6	49,005	21/7	99,000
03/4	774	25/4	1716	17/5	3,450	08/6	14,164	30/6	51,420	22/7	101,258

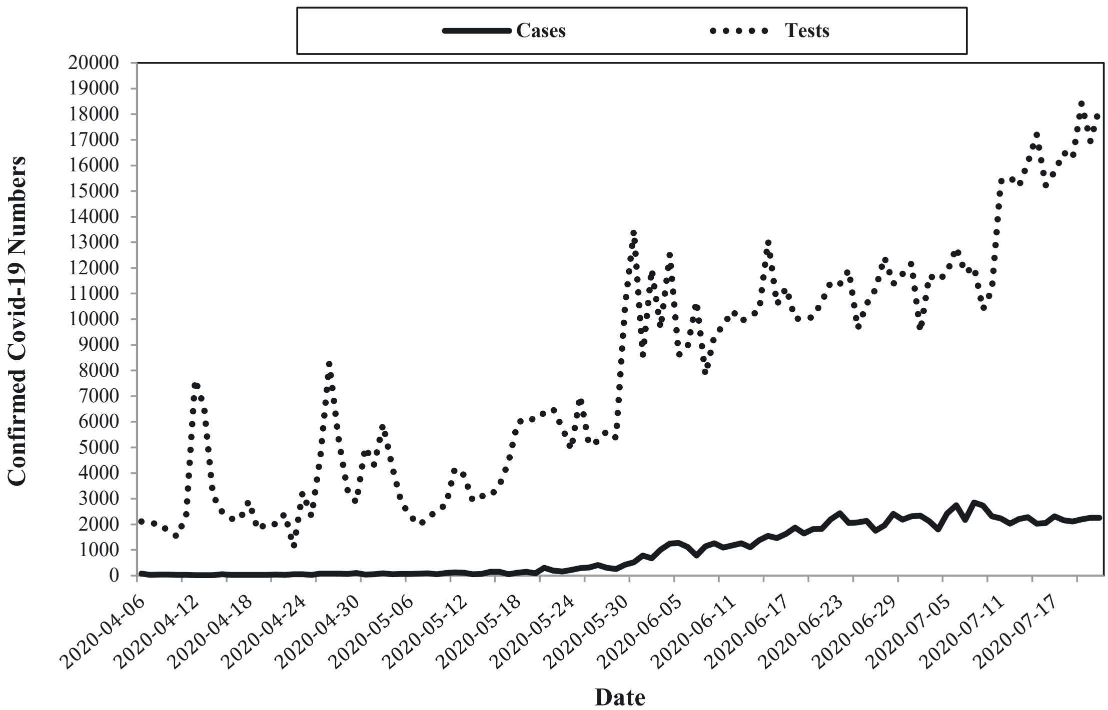

Figure 1: Confirmed daily cases and tests of COVID-19 in Iraq for Apr. 6 to Jul. 22, 2020.

Results and discussion

Starting values of parameters

For Gompertz model, as we explain above, we choose θ_1 as the maximum of daily cumulative COVID-19 cases, so $\hat{\theta}_1^{(0)} = 101,259$, and estimation of Eq. (30) yields $\hat{y}_t^* = 2.694 - 0.031t$, so $\hat{\theta}_2^{(0)} = \mathrm{e}^{2.694} = 14.791$ and $\hat{\theta}_3^{(0)} = 0.031$.

For Richards' model, we choose θ_1 as in Gompertz model, $\hat{\theta}_1^{(0)} = 101,259$, and $\hat{\theta}_4^{(0)} = 1.00$, then estimation of Eq. (30) yields $\hat{y}_t^* = 7.260 - 0.069t$, so $\hat{\theta}_2^{(0)} = \mathrm{e}^{7.260} = 1,422.257$ and $\hat{\theta}_3^{(0)} = 0.069$.

For Weibull model, we choose θ_1 as in other models, $\hat{\theta}_1^{(0)} = \hat{\theta}_2^{(0)} = 101,259$, and estimation of Eq. (31) yields $\hat{y}_t^* = -11.675 + 2.233t$, so $\hat{\theta}_3^{(0)} = \mathrm{e}^{-11.675} = 0.000009$ and $\hat{\theta}_4^{(0)} = 2.233$.

Table 3 summarizes the estimation values of studied models.

Models estimation

The statistical package Minitab-17 was used to estimate the models to the daily cumulative COVID-19 cases data and estimate the parameters. The Levenberg–Marquardt iterative method was chosen as it represents a

Table 3: Starting values for studied growth models.

	Gompertz			Richards'				Weibull			
Parameter	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4
Initial	101,259	14.791	0.031	101,259	1,422.257	0.069	1.000	101,259	101,259	0.000009	2.233

compromise between the linearization Gauss–Newton method and the steepest descent method and appears to combine the best features of both while avoiding their most serious limitations. Using the above initial values of parameters, the estimation of studied models are:

Gompertz estimated model:

$$\hat{y}_{\mathrm{t}} = 249,427.694\mathrm{e}^{-28.975\mathrm{e}^{-0.026t}}$$

Richards' estimated model:

$$\hat{y}_t = \frac{131,204.395}{\left(1+4,727.22\mathrm{e}^{-0.073t}\right)^{-\frac{1}{0.969}}}$$

Weibull estimated model:

$$\hat{y}_t = 115,108.058 - 114,114.844\mathrm{e}^{-(5.516\times10^{-15})t^{6.867}}$$

Figure 2 shows that plotting of predicated of cumulative COVID-19 cases with actual cumulative cases for all above three models.

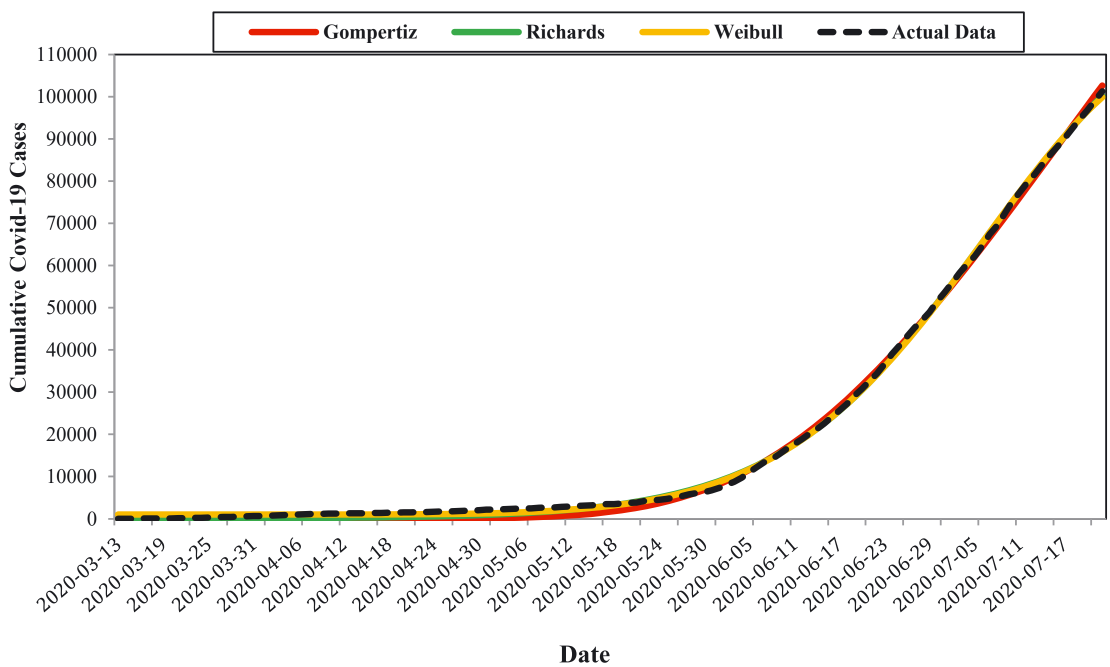

Figure 2: Fitting growth models to the daily cumulative COVID-19 cases.

Estimated parameters, standard error and its 95% confidence lower and upper bounds values are showed in the following table (Table 4):

It is shown that all parameters' estimations of Weibull model are statistically significant at 5% level, since the confidence interval of all model estimators does not include zero. While some parameters' estimation of other models are insignificant. This suggests that the Weibull model is the better than other models for representation of daily cumulative COVID-19 cases in Iraq.

Model selection criteria

Table 5 presents the analysis of variance results of the three models. F-ratio values indicate that all three models are statistically significant at α=1% level. When we compare the results of three models, we see that the Weibull model have largest F-ratio, this explained that Weibull model is preferred to the data.

Table 4: Estimation results of studied growth models.

Model	Parameter	Estimator	Std. error	95% confidence interval	
				Lower bound	Upper bound
Gompertz	θ_1	249,427.694	12,392.964	225,563.268	279,408.115
	θ_2	28.975	1.778	25.166	33.789
	θ_3	0.026	0.001	0.003	0.069
Richards'	θ_1	131,204.395	4,043.747	124,534.456	139,425.209
	θ_2	4,727.220	2,993.641	1,522.801	845,267,028.689
	θ_3	0.073	0.005	0.019	0.276
	θ_4	0.969	0.097	1.418	1.965
Weibull	θ_1	115,108.058	1,221.911	112,690.298	117,525.819
	θ_2	114,114.844	1,246.365	111,648.699	226,580.990
	θ_3	5.516×10^{-15}	3.216×10^{-15}	2.559×10^{-15}	8.472×10^{-15}
	θ_4	6.867	0.060	6.748	6.987

Table 5: ANOVA results of studied growth models.

Model	Source	Sum of squares	*df*	Mean squares	*F*-ratio	p-Value
Gompertz	Regression	110,294,188,700.215	3	36,764,729,560.507	26,914.612**	0.000
	Residual	176,210,976.086	129	1,365,976.559		
Richards'	Regression	110,393,839,700.214	4	27,598,459,930.519	46,141.135**	0.000
	Residual	76,559,935.925	128	598,124.499		
Weibull	Regression	163,337,624,202.681	4	40,847,817,114.231	97,393.322**	0.000
	Residual	53,684,590.321	128	419,410.862		
	Uncorrected total	163,391,308,793.000	132			
	Corrected total	110,470,399,653.295	131			

**Statistically significant at 1% level.

The above result can be confirmed through other criteria AIC, AIC_c, BIC and WIC as explain in Table 6. The results showed that the AIC value ranged from 1,711.894 to 1,880.88 and the Weibull model was ranked first in term of the lowest AIC value. The AIC_c, BIC and WIC of the three models have the same changes trend as AIC.

Table 6: Goodness of fit and criterion results of studied growth models.

Criteria	Gompertz model	Richards' model	Weibull model
RMSE	1,168.750	773.385	647.619
Bias	605.602	284.401	0.000
MAE	955.205	616.621	531.753
AIC	1,880.880	1,772.010	1,711.894
AIC_c	1,881.066	1,772.322	1,712.207
BIC	1,889.551	1,783.582	1,723.425
WIC	1,877.155	1,772.173	1,725.274

Goodness of fit

From Table 6, it was observed that the Weibull model provided the best fit since the model gives lowest value of RMSE 647.619 which about 80% less than Gompertz model, and about 19% less than Richards' model. In other hand, Weibull model had lowest value of bias that is zero. This result reflects that the predicted of cumulative COVID-19 cases by Weibull model is very close (in mean) to actual cumulative cases. The MAE of the three models have the same changes trend as RMSE.

Evaluation of model assumptions

We see form the results of Tables 5 and 6 that the Weibull model is more suitable than the Gompertz and Richards' models to describe the growth of daily cumulative cases of COVID-19 in Iraq. The important question here is: can this model generate efficient forecasts of the cumulative COVID-19 cases? Any model fits the data and give efficient and reliable forecasts when it had acceptable criteria as in Tables 5 and 6 and must consider the following assumptions: residuals must be independent, have the same variance and have the normal distribution. The estimates may be biased and estimation of errors may be overestimated or underestimated when these assumptions are not considered (Table 7).

Table 7: Normally and independent tests for residuals (Weibull model).

Test name	Test value	p-Value
Anderson-Darling	0.580	0.129
Ljung–Box *Q*	38.064	0.203

When absolute values of Weibull model residuals are plotted against time as in Figure 3, we see that no pattern relation between time and residuals, that's mean the variance of residuals is homogeneous, in other words, the residuals have the same variance. Also, Figure 4 and Table 8 indicates that the residuals are normally distributed since the p–p normal plot of residuals shows that the points lie in the straight line and

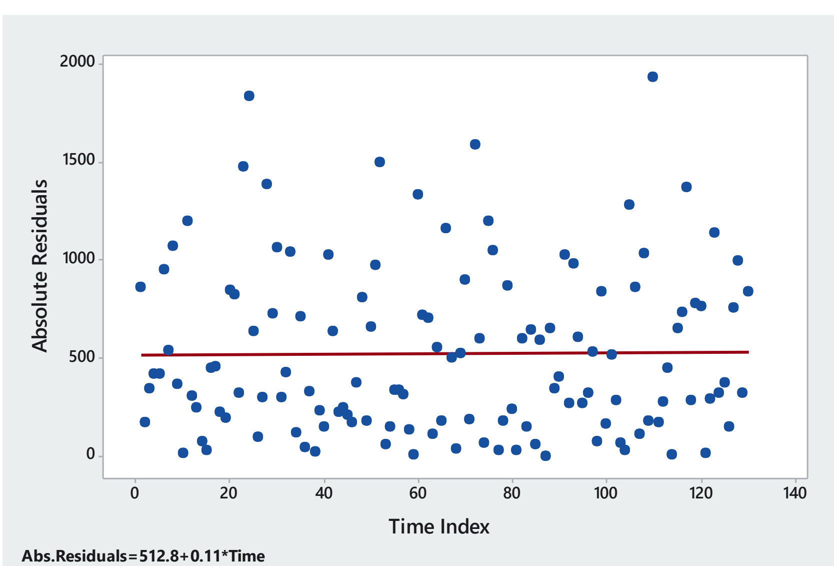

Figure 3: Homoscedasticity of Weibull model residuals.

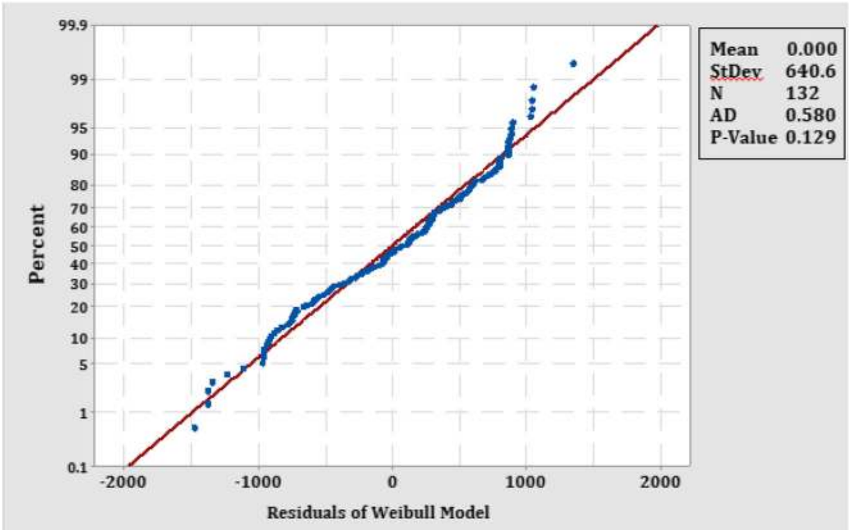

Figure 4: Normal distribution of Weibull model residuals.

Table 8: Absolute distance criterion for three models based on "out-sample" data.

Index, t	Date	y_t	$\lvert y_t-\hat{y}_{t,G}\rvert$	$\lvert y_t-\hat{y}_{t,R}\rvert$	$\lvert y_t-\hat{y}_{t,W}\rvert$
133	July 23, 2020	103,743	36,908	66,500	2,287
134	July 24, 2020	106,228	36,785	61,257	3,325
135	July 25, 2020	109,090	36,254	55,832	4,838
136	July 26, 2020	111,549	36,095	50,992	6,047
137	July 27, 2020	114,102	35,810	46,226	7,449
138	July 28, 2020	116,849	35,298	41,423	9,141
139	July 29, 2020	119,817	34,532	36,544	11,148
140	July 30, 2020	122,780	33,737	31,806	13,242
141	July 31, 2020	126,126	32,526	26,810	15,806
142	Aug. 01, 2020	128,221	32,532	23,182	17,203
143	Aug. 02, 2020	130,668	32,151	19,310	19,030
144	Aug. 03, 2020	133,403	31,447	15,251	21,219
145	Aug. 04, 2020	136,239	30,607	11,184	23,578
146	Aug. 05, 2020	139,073	29,735	7,207	25,997
147	Aug. 06, 2020	142,120	28,615	3,097	28,687
148	Aug. 07, 2020	145,581	27,045	1,352	31,843
149	Aug. 08, 2020	148,906	25,577	5,595	34,909
150	Aug. 09, 2020	151,632	24,673	9,175	37,418
151	Aug. 10, 2020	155,116	22,976	13,452	40,721
152	Aug. 11, 2020	158,512	21,333	17,585	43,968
153	Aug. 12, 2020	161,953	19,610	21,711	47,287
154	Aug. 13, 2020	165,794	17,453	26,189	51,029
155	Aug. 14, 2020	169,807	15,090	30,794	54,963
156	Aug. 15, 2020	174,100	12,413	35,637	59,193

y_t: Observed cumulative daily cases; $\hat{y}_{t,G}$: Predictions of Gompertiz model; $\hat{y}_{t,R}$: Predictions of Richards' model; $\hat{y}_{t,W}$: Predictions of Weibull model.

the p-value of A−D test is greater than 5%. Moreover, the Weibull model residuals are independent since the p-value of Ljung−Box Q test with 30 lags is greater than 5%.

"Out-sample" data models' performance

In order to verify the performance of the models based on "outside the sample" data, 24 subsequent observations for the period from July 23rd to August 15th were collected from the source of the study data. These values have been compared with the corresponding predictions from the three previously estimated models by using mean of absolute distance (MAD) criterion as in Table 8.

The mean absolute distance for the models Gompertz, Richards' and Weibull are respectively 28,717, 27,421 and 25,430, so we conclude that the Weibull model is still the more appropriate than other models to describing the confirmed daily cumulative cases of COVID-9 in Iraq.

Forecasting

The prevalence of epidemics is usually accompanied by many chance variables that cannot be measured or controlled, so the process of predicting the number of infected people resulting from these epidemics will not be 100% accurate and cannot produce the same actual values, the model only approximates the number of people infected. Table 8 and Figure 5 presents the predictions of confirmed daily cumulative COVID-19 cases in Iraq according to Weibull growth model. We believe that the number of confirmed cumulative cases of novel coronavirus in Iraq will rise with range from 101,396 to 114,907 cases in coming 24 days (form July 23rd to August 15th, 2020). So, this will provide reference value for all levels of departments and hospitals in the next few days to implement effective intervention and prevention of the spread of novel coronavirus.

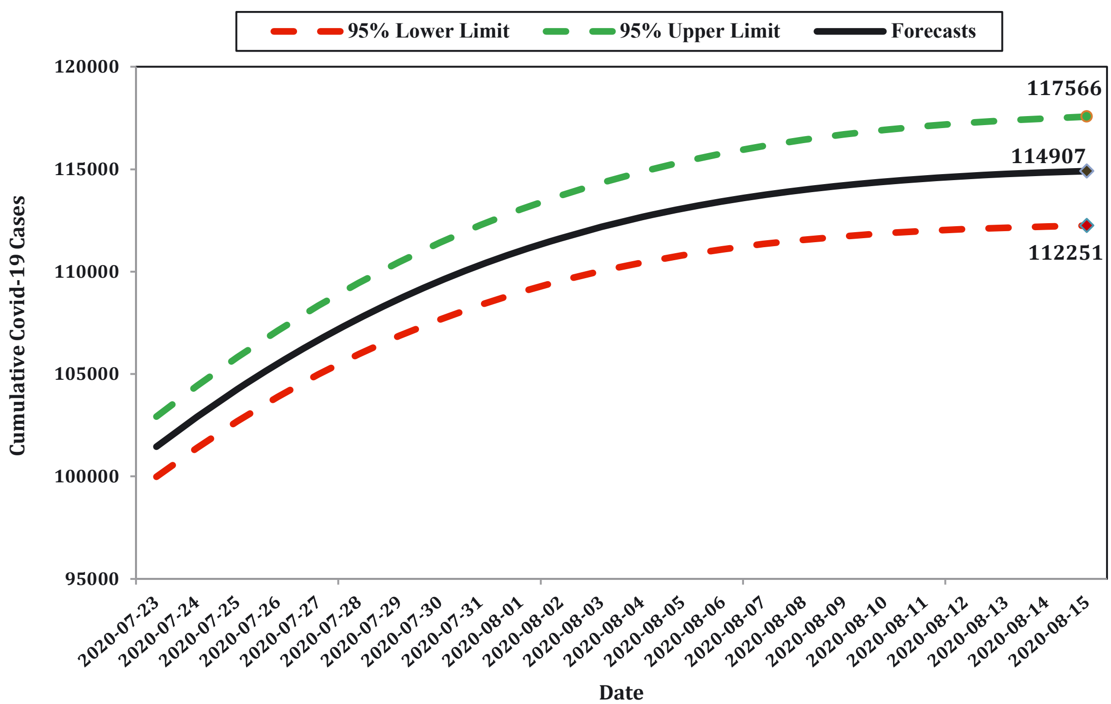

Figure 5: Forecasts of cumulative COVID-19 cases by Weibull model.

Regarding the physical interpretation of the inflection point of the Weibull model that was found to be the most suitable for the present data than other models, we apply Eq. (21) then we get: $\left(t_{\text{inf.}}, \hat{y}_{\text{inf.}}\right) = (11, 766, 546)$, the time index ($t$=117) gives the position of the point of inflection, i.e. the time when the growth rate is maximum, and at this date the peak growth rate will be (66,546) cases. And since $\hat{y}_{\text{inf.}} = 66; 546$ is close to $y_{117} = 67, 338$. Therefore, we conclude that the incidence number of novel COVID-19 will be maximum at date (Jul. 7, 2020).

Conclusion

The Weibull model proved to be very effective in describing epidemic curve of Covid-19 and estimating important epidemiological parameters, such as the time of the peak of the curve for daily cumulative cases, thus allowing a practical and efficient monitoring of the epidemic evolution.

In this study, the Weibull model gives the best results with zero bias, the lowest RMSE = 647.619 and WIC = 1725.274 compared to the other applied growth models Gompertz and Richards. On this basis, the daily cumulative of COVID-19 cases in Iraq can reach 114,907 at 15th of August, 2020 with 95% prediction interval (from 112,251 to 117,566). The inflection point of the Weibull curve indicates that the peak time when the growth rate is maximum, is 7th of July, 2020, and at this time the daily cumulative incidence is 67,338 cases. The fitting models presented and some measures in this study were performed with the “Nonlinear Regression” tool available in the Minitab-17 software.

Research funding: The research is supported by College of Computer Sciences and Mathematics, University of Mosul, Mosul, Republic of Iraq.
Author contribution: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.
Competing interests: The author declares that there are no conflicts of interest regarding the publication of this paper.
Informed consent: Not applicable.
Ethical approval: Not applicable.

Appendix

A1 Inflection points of growth models

Gompertz model

To ascertain the shape of the Gompertz function, we first look to derivatives of Eq. (1):

$$\frac{\mathrm{d}y_t}{\mathrm{d}t} = \theta_1\theta_2\theta_3 \mathrm{e}^{-\theta_3 t} \mathrm{e}^{-\theta_2 e^{-\theta_3 t}} = \theta_2\theta_3 \mathrm{e}^{-\theta_3 t} y_t$$

$$\frac{\mathrm{d}^2 y_t}{\mathrm{d}t^2} = \theta_2\theta_3^2 \mathrm{e}^{-\theta_3 t}\left(\theta_2 \mathrm{e}^{-\theta_3 t} - 1\right) y_t$$

Therefore, when $\frac{\mathrm{d}^2 y_t}{\mathrm{d}t^2} = 0$ we get $t = \frac{\ln\,\theta_2}{\theta_3}$ substitution in Eq. (1) gives y_t=0.367888θ_1 thus, the Gompertz model has a point of inflection at:

$$(t_{\mathrm{inf.}}, y_{\mathrm{inf.}}) = \left(\frac{\ln\,\theta_2}{\theta_3}, 0.367888\theta_1\right) \tag{19}$$

and since $\theta_1 \approx \max(y_t)$ that’s mean the ordinate y_t at the point of inflection is approximately, when 37% of the final growth has been reached.

Richards’ model

The inflection point of the Richards’ model can be founded as follows:

$$\frac{\mathrm{d}y_t}{\mathrm{d}t} = \frac{\theta_2\theta_3}{\theta_4} \mathrm{e}^{-\theta_3 t}\left(1 + \theta_2 \mathrm{e}^{-\theta_3 t}\right)^{-1} y_t$$

$$\frac{\mathrm{d}^2 y_t}{\mathrm{d}t^2} = \frac{\theta_2\theta_3^2}{\theta_4} \mathrm{e}^{-\theta_3 t} y_t \left(1 + \theta_2 \mathrm{e}^{-\theta_3 t}\right)^{-1} \left\{\theta_3\left(1 + \frac{1}{\theta_4}\right) \mathrm{e}^{-\theta_3 t}\left(1 + \theta_2 \mathrm{e}^{-\theta_3 t}\right)^{-1} + 1\right\}$$

Therefore, when $\frac{d^2y_t}{dt^2}=0$ we get $t=-\frac{1}{\theta_3}\ln\left(\theta_4/\theta_2\right)$ substitution in Eq. (2) gives $y_t=\frac{\theta_1}{\sqrt[\theta_4]{\theta_4+1}}$ thus, the Richards' model has a point of inflection at:

$$(t_{\mathrm{inf.}},y_{\mathrm{inf.}})=\left(-\frac{1}{\theta_3}\ln\left(\frac{\theta_4}{\theta_2}\right),\frac{\theta_1}{\sqrt[\theta_4]{\theta_4+1}}\right) \tag{20}$$

Weibull model

To find the inflection point of the Weibull model, we have:

$$\frac{dy_t}{dt}=\theta_3\theta_4t^{\theta_4-1}\left(\theta_1-y_t\right)$$

$$\frac{d^2y_t}{dt^2}=\theta_3\theta_4t^{\theta_4-1}\left\{\left(\theta_4-1\right)t^{-1}\left(\theta_1-y_t\right)-\theta_3\theta_4t^{\theta_4-1}\left(\theta_1-y_t\right)\right\}$$

Therefore, when $\frac{d^2y_t}{dt^2}=0$ we get $t=\left[\left(\theta_4-1\right)/\theta_3\theta_4\right]^{1/\theta_4}$ substitution in Eq. (3) gives $y_t=\theta_1-\theta_2e^{\left[-\left(\theta_4-1\right)/\theta_4\right]}$ thus, the Weibull model has a point of inflection at:

$$(t_{\mathrm{inf.}},y_{\mathrm{inf.}})=\left(\left[\frac{\theta_4-1}{\theta_3\theta_4}\right]^{\frac{1}{\theta_4}},\theta_1-\theta_2e^{-\left(\frac{\theta_4-1}{\theta_4}\right)}\right) \tag{21}$$

A2 Estimation method of growth models

These normal equations take the form:

$$\sum_{i=1}^{n}\left[y_t-f\left(t;\boldsymbol{\theta}\right)\right]\left[\frac{\partial f\left(t;\boldsymbol{\theta}\right)}{\partial\theta_j}\right]=0,\quad j=1,2,\ldots,p \tag{22}$$

For the nonlinear models like Gompertz, Richards' and Weibull it is very difficult to solve Eq. (22) to obtain the vector $\hat{\boldsymbol{\theta}}$ of p-parameters, in this case we used the iterative methods (Draper and Smith 1981). The Gauss–Newton modified method is one of the important and frequently method that used in the case of growth models. According to this method, firstly we write $f\left(t;\boldsymbol{\theta}\right)$ in terms of Taylor's expansion formula (Bates and Watts 2007):

$$f\left(t;\boldsymbol{\theta}\right)=f\left(t;\boldsymbol{\theta}^{(0)}\right)+\left(\theta_1-\theta_1^{(0)}\right)\left[\frac{\partial f\left(t;\boldsymbol{\theta}\right)}{\partial\theta_1}\right]+\left(\theta_2-\theta_2^{(0)}\right)\left[\frac{\partial f\left(t;\boldsymbol{\theta}\right)}{\partial\theta_2}\right]+\cdots+\left(\theta_p-\theta_p^{(0)}\right)\left[\frac{\partial f\left(t;\boldsymbol{\theta}\right)}{\partial\theta_p}\right] \tag{23}$$

$$\Rightarrow f\left(t;\boldsymbol{\theta}\right)-f\left(t;\boldsymbol{\theta}^{(0)}\right)=\gamma_1w_{1t}+\gamma_2w_{2t}+\cdots+\gamma_pw_{pt} \tag{24}$$

where

$$w_{jt}=\frac{\partial f\left(t;\boldsymbol{\theta}\right)}{\partial\theta_j}\quad\text{and}\quad\gamma_j=\theta_j-\theta_j^{(0)},\quad t=1,2,\ldots,n;\quad j=1,2,\ldots,p \tag{25}$$

such that $\boldsymbol{\theta}^{(0)}$ is the vector of initial values of parameter, i.e. $\boldsymbol{\theta}^{(0)}=\left(\theta_1^{(0)},\theta_2^{(0)},\ldots,\theta_p^{(0)}\right)'$ substitute Eq. (7) in Eq. (24), we get:

$$y_t-f\left(t;\boldsymbol{\theta}^{(0)}\right)=\gamma_1w_{1t}+\gamma_2w_{2t}+\cdots+\gamma_pw_{pt}+\varepsilon_t \tag{26}$$

by using ordinary least squares method to estimate the linear model in Eq. (22) to get the estimates $\hat{\gamma}_{1,}\hat{\gamma}_{2,}\ldots\hat{\gamma}_p$, and since the initial values of the parameters are known then form Eq. (22) we have:

$$\hat{\theta}_j=\hat{\gamma}_j-\hat{\theta}_j^{(0)},\quad j=1,2,\ldots,p \tag{27}$$

Now, we return to Eq. (8) to use the estimators from Eq. (27) to get new the vector of new estimators $\hat{\boldsymbol{\theta}}^{(1)}$ (first iteration). This process will be repeated until the convergence in estimators was hold, i.e. at $(s-1)$ and (s) iterations we have:

$$\left|\hat{\boldsymbol{\theta}}^{(s)} - \hat{\boldsymbol{\theta}}^{(s-1)}\right| \leq \in \text{ for some small positive} \in \text{error}$$

This means also that SS_{Res} will be stabilized after (s) iteration, and then we have the lowest value of SS_{Res} and the final parameters estimation vector will be $\hat{\boldsymbol{\theta}}^{(s)} = \left(\hat{\theta}_1^{(s)}, \hat{\theta}_2^{(s)}, \ldots, \hat{\theta}_{\text{p}}^{(s)}\right)'$ which is computed from (Ratkowsky 1983):

$$\begin{aligned}\hat{\boldsymbol{\theta}}^{(s)} &= \hat{\boldsymbol{\theta}}^{(s-1)} + \left(\boldsymbol{w}^{(s-1)\prime}\boldsymbol{w}^{(s-1)}\right)^{-1}\boldsymbol{w}^{(s-1)\prime}\left[\boldsymbol{y} - f\left(t; \hat{\boldsymbol{\theta}}^{(s-1)}\right)\right] \\ &= \hat{\boldsymbol{\theta}}^{(s-1)} + \left(\boldsymbol{w}^{(s-1)\prime}\boldsymbol{w}^{(s-1)}\right)^{-1}\boldsymbol{w}^{(s-1)\prime}\boldsymbol{y}^{(s-1)}\end{aligned} \tag{28}$$

where $\boldsymbol{w}^{(s-1)}$ is a $(\text{n} \times \text{p})$ of partial derivatives matrix, i.e.:

$$\boldsymbol{w}^{(s-1)} = \left[\frac{\partial f\left(t; \hat{\boldsymbol{\theta}}\right)}{\partial \hat{\boldsymbol{\theta}}}\right]_{\hat{\theta} = \hat{\theta}^{(s-1)}} \tag{29}$$

A3 Initial values of growth models parameters

Gompertz model

In order to estimate the Gompertz model by above estimation method, we need the starting values of Eq. (1). We can transform Eq. (1) to the following linear form:

$$y_t^* = \theta_2^* + \theta_3^* t + \varepsilon_t \tag{30}$$

where:

$$y_t^* = \ln\left[\ln\left(\frac{\theta_1}{y_t}\right)\right], \quad \theta_2^* = \ln(\theta_2), \quad \theta_3^* = -\theta_3$$

The initial values of θ_1 is $\hat{\theta}_1^{(0)} \geq \max(y_t)$, and estimation of Eq. (30) yields other initial values, such that:

$$\hat{\theta}_2^{(0)} = e^{\hat{\theta}_2^*}, \quad \hat{\theta}_3^{(0)} = -\hat{\theta}_3^*$$

Richards' model

To find the starting values of the Eq. (2), we can transform it to the linear form as in Eq. (20), where:

$$y_t^* = \ln\left[\left(\frac{\theta_1}{y_t}\right)^{\theta_4} - 1\right], \quad \theta_2^* = \ln(\theta_2), \quad \theta_3^* = -\theta_3$$

The initial values of θ_1 is $\hat{\theta}_1^{(0)} \geq \max(y_t)$, and of θ_4 is $\hat{\theta}_4^{(0)}$ which can be take it according to the inflection point such that $\theta_2 \geq \theta_4 \geq 1$, to make the calculation easy, we can takes $\hat{\theta}_4^{(0)} = 1$. Estimation of Eq. (30) yields other initial values, such that:

$$\hat{\theta}_2^{(0)} = \text{e}^{\hat{\theta}_2^*}, \quad \hat{\theta}_3^{(0)} = -\hat{\theta}_3^*$$

Weibull model

To find the starting values of the Eq. (3), we can transform it to the following linear form:

$$y_t^* = \theta_3^* + \theta_4^* t^* + \varepsilon_t \tag{31}$$

where:

$$y_t^* = \ln\left[-\ln\left(\frac{\theta_1 - y_t}{\theta_2}\right)\right], \quad t^* = \ln(t), \quad \theta_3^* = \ln(\theta_3), \quad \theta_4^* = \theta_4$$

The initial values of θ_1 is $\hat{\theta}_1^{(0)} \geq \max(y_t)$, and since $y_0=\theta_1 - \theta_2$ so $\theta_1=\theta_2$, therefore $\hat{\theta}_1^{(0)} = \hat{\theta}_2^{(0)}$. Estimation of Eq. (31) yields other initial values, such that:

$$\hat{\theta}_3^{(0)} = e^{\hat{\theta}_3^*}, \quad \hat{\theta}_4^{(0)} = \hat{\theta}_4^*$$

References

Archontoulis, S. V., and F. E. Miguez. 2015. "Nonlinear Regression Models and Applications in Agricultural Research." *Agronomy Journal* 107 (2): 786–98.

Bates, D. M., and D. J. Watts. 2007. *Nonlinear Regression Analysis and its Applications*. New York: John Wiley & Sons.

Brockwell, P. J., and R. A. Davis. 2016. *Introduction to Time Series and Forecasting*, 3rd ed. Switzerland: Springer-Verlag.

Burnham, K. P., and D. R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York: Springer-Verlag.

Carroll, R. J., and D. Ruppert. 1988. *Transformations and Weighting in Regression*. New York: Chapman & Hall.

Chowell, G. 2017. "Fitting Dynamic Models to Epidemic Outbreaks with Quantified Uncertainty: A Primer for Parameter Uncertainty, Identifiability, and Forecasts." *Infectious Disease Modelling* 2: 379–98.

Dagogo, J., E. C. Nduka, and U. P. Ogoke. 2020. "Comparative Analysis of Richards', Gompertz and Weibull Models." *IOSR Journal of Mathematics* 16 (1): 15–20.

Draper, N. R., and H. Smith. 1981. *Applied Regression Analysis*, 2nd ed. New York: John Wiley & Sons.

Fekedulegn, D., M. P. Mac Siurtain, and J. J. Colbert. 1999. "Parameter Estimation of Nonlinear Growth Models in Forestry." *Silva Fennica* 33 (4): 327–36.

Goshu, A. T., and P. R. Koya. 2013. "Derivation of Inflection Points of Nonlinear Regression Curves-Implications to Statistics." *American Journal of Theoretical and Applied Statistics* 2 (6): 268–72.

Jiang, X., B. Zhao, and J. Cao. 2020. "Statistical Analysis on COVID-19." *Biomedical Journal of Scientific & Technical Research* 26 (2): 19716–27.

Mahanta, D. Y., and B. Borah. 2014. "Parameter Estimation of Weibull Growth Models in Forestry." *International Journal of Mathematics Trends and Technology* 8 (3): 157–63.

Majumder, M., and K. D. Mandl. 2020. "Early Transmissibility Assessment of a Novel Coronavirus in Wuhan, China." *SSRN Electronic Journal* 3524675. National Health Commission of the People's Republic of China. https://doi.org/10.2139/ssrn.3524675.

Miller, F. P., A. F. Vandome, and J. McBrewster. 2011. *Anderson-Darling Test*. England: International Book Marketing Service Limited.

Myung, I. J. 2003. "Tutorial on Maximum Likelihood Estimation." *Journal of Mathematical Psychology* 47 (1): 90–100.

Pinheiro, J. C., and D. M. Bates. 2000. "Mixed-Effects Models in S and S-PLUS." In *Statistics and Computing Series*. New York: Springer-Verlag.

Ratkowsky, D. A. 1983. *Nonlinear Regression Modeling*. New York: Marcel Dekker.

Rinke, S., and P. Sibbertsen. 2016. "Information Criteria for Nonlinear Time Series Models." *Studies in Nonlinear Dynamics & Econometrics* 20 (3): 325–41.

Ritz, C., and J. C. Streibig. 2008. *Nonlinear Regression with R*. New York: Springer.

Roosa, K., Y. Lee, R. Luo, A. Kirpich, R. Rothenberg, J. M. Hyman, P. Yan, and G. Chowell. 2020a. "Real-Time Forecasts of the COVID-19 Epidemic in China from February 5th to February 24th, 2020." *Infectious Disease Modelling* (5): 256–63. https://doi.org/10.1016/j.idm.2020.02.002.

Roosa, K., Y. Lee, R. Luo, A. Kirpich, R. Rothenberg, J. M. Hyman, P. Yan, and G. Chowell. 2020b. "Short-term Forecasts of the COVID-19 Epidemic in Guangdong and Zhejiang." *Journal of Clinical Medicine* 9 (596): 1–9.

Shen, C. Y. 2020. "Logistic Growth Modeling of COVID-19 Proliferation in China and its International Implications." *International Journal of Infectious Diseases* (96): 582–9. https://doi.org/10.1016/j.ijid.2020.04.085.

Teleken, J. T., A. C. Galvão, and W. S. Robazza. 2017. "Comparing Non-linear Mathematical Models to Describe Growth of Different Animals." *Acta Scientiarum. Animal Sciences* 39 (1): 73–81.

Tolles, J., and T. Luong. 2020. "Modeling Epidemics with Compartmental Models." *Journal of the American Medical Association* 323 (4): 2515–16.

Tran, D. 2017. "Modeling and Forecasting Stock Markets Prices with Sigmoidal Curves." MSc diss., Los Angeles: Faculty of the Department of Mathematics, California State University.

Zhao, S., Q. Lin, J. Ran, S. S. Musa, G. Yang, W. Wang, Y. Lou, D. Gao, L. Yang, D. He, and M. H. Wang. 2020. "Preliminary Estimation of the Basic Reproduction Number of Novel Coronavirus (2019-nCoV) in China, from 2019 to 2020: A Data-Driven Analysis in the Early Phase of the Outbreak." *International Journal of Infectious Diseases* (92): 214–17. https://doi.org/10.1016/j.ijid.2020.01.050.

Zhao, S., and H. Chen. 2020. "Modeling the Epidemic Dynamics and Control of COVID-19 Outbreak in China." *Quantitative Biology* 1–9.