ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH
DOI: 10.5336/medsci.2020-78094

A Study on Determination of the Growth Models Using COVID-19 Cases Between March 17 and July 12 2020 in Turkey: Cross-sectional Study

Türkiye'de 17 Mart ile 12 Temmuz 2020 Arasında COVID-19 Vakaları Kullanılarak Büyüme Modellerinin Belirlenmesi Üzerine Bir Çalışma: Kesitsel Çalışma

Levent ÖZBEK[a]

[a]Department of Statistics, Ankara University Faculty of Sciences, Ankara, TURKEY

ABSTRACT Objective: We use the most frequently used growth functions in the literature for estimating Turkey's cumulative number of confirmed coronavirus disease-2019 (COVID-19) cases. We analyze Brody, Bertalanffy, Logistic, Generalized Logistic, Gompertz, Richards, Negative Exponential, Stevens, and Tanaka models for determining the appropriate model. **Material and Methods:** The number of cases of COVID-19 in Turkey between March 17, 2020 and July 12, 2020 were included in the study. The data used in the study were obtained from Johns Hopkins University. We used the most frequently used Non-linear growth models in the literature for estimating Turkey's cumulative number of confirmed COVID-19 cases. Matlab software was used to estimate the parameters in the models studied. non-linear least squares regression function is used in Matlab software. **Results:** According to the estimation results, the best fitting model is the Richards model in terms of both the mean squared error and R^2 (coefficient of determination). **Conclusion:** We recommend to use the Richards model in modeling the cumulative COVID-19 cases. For estimating the future cases, it will be appropriate to use the Richards model. In addition, estimates of the number of daily cases obtained from the Richards model are not compatible with the actual number of daily cases. This may occur because estimations on these growth models can sometimes lead to erroneous results.

Keywords: Turkey; COVID-19; pandemic; growth function

ÖZET Amaç: Türkiye'nin teyit edilmiş koronavirüs hastalığı-2019 [coronavirus disease-2019 (COVID-19)] vakalarının kümülatif sayısını tahmin etmek amacıyla literatürde en sık kullanılan büyüme fonksiyonları göz önünde bulundurulmuştur. Bu çalışmada, literatürde mevcut olan toplam 9 model araştırılmıştır. Bu modeler; Brody, Bertalanffy, Lojistik, Genelleştirilmiş Lojistik, Gompertz, Richards, Negatif Üstel, Stevens ve Tanaka'dır. Bu modellerden hangisinin daha iyi olduğu tespit edilmeye çalışılmıştır. **Gereç ve Yöntemler:** Türkiye'de 17 Mart 2020 ve 12 Temmuz 2020 tarihleri arasında ortaya çıkan COVID-19 vaka sayıları çalışmaya dâhil edilmiştir. Çalışmada kullanılan veriler Johns Hopkins Üniversitesinden alınmıştır. Literatürde çok sık kullanılan lineer olmayan büyüme modellerinden bazıları ele alınmış ve hangi modelin Türkiye COVID-19 verilerine daha iyi uyum sağladığı araştırılmıştır. Ele alınan modellerdeki parametreleri tahmin etmek için Matlab yazılımı kullanılmıştır. Matlab yazılımında doğrusal olmayan en küçük kareler regresyonu fonksiyonu kullanılmıştır. **Bulgular:** Çalışmadaki tahmin sonucunda, toplam vaka sayılarını modellemede en iyi analizin Richards modeli tarafından, hem ortalama hata karesi hem de R^2 (belirleme katsayısı) açısından yapıldığı görülmüştür. **Sonuç:** Richard modelinin kümülatif COVID-19 vaka sayısını modellemede kullanılması tavsiye edilmektedir. Gelecek sürecin tahmininde, Richard modeli kullanılarak tahminlerin yapılmasının uygun olacağı düşünülmektedir. Bununla birlikte, Richard modelinden elde edilen günlük vaka sayısı tahminleri gerçek günlük vaka sayısıyla uyumlu değildir. Bu büyüme modelleri üzerinde tahmin yapmanın bazen hatalı sonuçlara yol açabileceği düşünülmektedir.

Anahtar Kelimeler: Türkiye; COVID-19; pandemic; büyüme fonksiyonu

In December 2019, a new coronavirus disease characterized as a viral infection with a high level of transmission emerged in Wuhan, China. Coronavirus disease-2019 (COVID-19) is caused by the virus known as severe acute respiratory syndrome-coronavirus-2 established by the International Committee

Correspondence: Levent ÖZBEK
Department of Statistics, Ankara University Faculty of Sciences, Ankara, TURKEY/TÜRKİYE
E-mail: ozbek@science.ankara.edu.tr

Peer review under responsibility of Turkiye Klinikleri Journal of Medical Sciences.

Received: 14 Jul 2020 ***Received in revised form:*** 02 May 2021 ***Accepted:*** 11 May 2021 ***Available online***: 26 May 2021

on Taxonomy of Viruses.[1-3] Gompertz and Logistic models have been used in China, South Korea, Italy and Singapore by Jia et al. and Castorina et al. Roosa et al. have used Generalized Logistic Growth Model (GLM) for the data between February 5 and February 24, 2020 for China.[4-6] Roosa et al. have used the GLM and Richard model for the data between February 13 and February 20, 2020 for China.[7] Munayco et al. have used the Generalized Growth Model for the dates February 29 and March 30, 2020 for Peru.[8] Gompertz, logistic and artificial neural network models were applied to carry out the modeling of the number of cases of COVID-19.[9] Zuzana et al. used the model trajectory of the number of infections for the USA by the Gompertz curve.[10] Cata et al. employed the Gompertz function to analyze the dynamics.[11] Petropoulos et al. used time series methods.[12]

The term "growth" is used to describe various biological phenomena. Growth of populations involves the reproduction of animals; growth of bodies involves increase in the number of cells or growth in the size of cells; growth of cells involves molecular replication. In a more broad sense, growth indicates a development in mass due to an increase in size and number of living things. The growth of organic structures is distinct from the growth of crystal structures in inorganic substances. Growth in organic structures, or more precisely living things, is the result of an increase in the height and weight of a single organism, a tissue or an organ due to a combination of biological and biochemical events, or it is the result of an increase in the number of members in a population formed by organisms.

Among the notable methods in modelling studies made on COVID-19, growth curves are primary. It is observed that the growth curves are commonly used for modelling the progress of the virus since it is the field of study for microbiologists. Growth curves are used for modelling the increase in the number of plants, bacteria or viruses in an environment. Expressing the growth of an organism or the increase in the number of viruses temporally is called "growth". Identification of the complex growth process is aimed using the growth curves. There are different models for the growth curves.

The Gompertz model is well known and widely used in many sub-fields of biology. Researchers have adapted the Gompertz model to various fields such as plant growth, bird growth, fish growth, and growth of other animals, as well as tumour growth and bacterial growth.[13,14]

The Gompertz is a special case among the four-parameter Richards model, and thus belongs to the Richards family of three-parameter sigmoidal growth models, along with familiar models such as the negative exponential (including the Brody), the logistic, and the von Bertalanffy (or only Bertalanffy). Numerous parametrisations and re-parametrisations of the Gompertz model may be found in the literature.[15-27]

This study emphasizes the growth models to estimate of the COVID-19 in Turkey. We consider most frequently used growth functions. These 9 models are Brody, Bertalanffy, Logistic, Generalized Logistic, Gompertz, Richards, Negative Exponential, Stevens, and Tanaka. However, as far as is known, there is no published article for Turkey that uses all of these models and determines which one fits better. The data used in the study is taken from Johns Hopkins University.[28]

MATERIAL AND METHODS

This research is managed in accordance with the tenets of Helsinki Declaration. Ethics committee approval is not required for this study.

Names of the used models and their mathematical equations are given in Table 1.[24] In all the models,, is the logarithmic cumulative cases which are encountered until t day. α, β, k, m are the parameters to be estimated and α represents asymptotic (final) value of $y(t)$, β is the change in $y(t)$ when t passes from 0 to infinity and k represents the factor by which the deviation of $y(t)$ from its asymptotic value is reduced every time, is widely used to represent the growth of an organism when it is approaching the maturity. ε is the error term.

A non-linear model may be written as Eq.1

$$y(t)=f(t,\alpha,\beta,k,m)+\varepsilon_i \qquad (1)$$

Here, $y(t)$ is the logarithmic cumulative cases and α, β, k, m are the parameters to be estimated in the model. ε is the error term. Estimators of α, β, k, m parameters are found by minimizing the Mean Squares Error (MSE), which is given by Eq.2

$$\frac{1}{n}\sum_{t=1}^{n}(y(t)\text{-}f(t,\,\alpha,\,\beta,\,k,\,m))^2 \qquad (2)$$

where, n is the observation number. Since $y(t)$ values are known during t=1,2..., n times, the value which sets this function to minimum will depend on parameters. In order to estimate α, β, k, m parameters, derivative is calculated with regard to α, β, k, m parameters in Eq. 2 and equaled to zero; and estimation of α, β, k, m parameters are found.

Bilgin et al. gave partial derivatives of some of these models and more detailed information. These estimators are called Least Squares Estimators (LSE).[27]

Matlab software is used for estimating the parameters in the models. The parameters were estimated using the nlinfit (non-linear least-squares regression) function in the matlab software. This function estimates the coefficients of a nonlinear regression function, using LSE.

RESULTS

In this study, MSE and R^2 statistics are used for making comparison between the models to determine which one is better. MSE and R^2 statistics for all models are given in Table 2. The values of the parameters estimated are shown in Table 3. As seen in Table 2, R^2 values are very close to each other. Nevertheless, it may be stated that Richards model are the most consistent in terms of both R^2 and MSE. The best analysis has been made by Richards model is terms of both R^2 and MSE. Actual values and estimated values using the Richard model is given in Figure 1. As a result of estimation with the data between March 17 and July 12, 2020, the best analysis for modelling the total case counts is made by Richards model in terms of both MSE and R^2. It has been determined that the use of this model is appropriate.

Actual cumulative cases and estimations which have been made using the Richards model are given in Figure 1. The number of daily cases can be easily calculated with $i_t = e^{yt}\text{-}e^{yt\text{-}1}$ to show the total number of cases up to y_t t days. Since we have the estimates of y_t, we can easily find the estimations of i_t with $i_t = e^{\hat{y}}{}_t\text{-}e^{\hat{y}}{}_{t\text{-}1}$. Daily cases and estimations are given in Figure 2. It should be noted that, as can be seen from Figure 2, the actual daily cases and the daily cases estimate obtained from the Richards model are not compatible.

TABLE 1: Non-linear models and their mathematical notations.[24]

Model name	**Statistical model**
Brody	$y(t;\alpha,\beta,k)=\alpha(1-\beta\exp(-kt))+\varepsilon$
Bertalanffy	$y(t;\alpha,\beta,k,m)=(\alpha^{1-m}-\beta\exp(-kt))^{1/(1-m)}+\varepsilon$
Logistic	$y(t;\alpha,\ \beta,k)=\alpha/(1+\beta\exp(\text{-}kt))+\varepsilon$
Generalized logistic	$y(t;\beta,k,m)=\alpha/((1+\beta\exp(\text{-}kmt)^{1/m})+\varepsilon$
Richards	$y(t;\alpha,k,m)=\ \alpha(1-\exp(-kt))^{1/m}+\varepsilon$
Negative exponential	$y(t;\alpha,k)=\alpha(1-\exp(-kt))+\varepsilon$
Stevens	$y(t;\ \alpha,\beta,p)=\alpha-\beta(k^t)+\varepsilon$
Tanaka	$y(t;\alpha,\ \beta,k,m)=(1/\sqrt{\beta})\ln\lvert 2\beta.(t\text{-}m)+2\sqrt{(k^2(t\text{-}m)^2+\alpha\beta)}\ \rvert+\varepsilon$
Gompertz	$Y(t)=\alpha\ \exp(-\beta\ \exp(\text{-}kt))+\varepsilon$

TABLE 2: Mean squares error and R^2 values regarding the models.

Model name	**MSE**	**R^2**
Brody	0.0084	0.9953
Bertalanffy	0.0084	0.9953
Logistic	0.0277	0.9651
Generalized logistic	0.0170	0.9815
Gompertz	0.0167	0.9817
Richards	0.0061	0.9999
Negative exponential	0.3358	0.5221
Stevens	0.0084	0.9953
Tanaka	0.1283	0.9997

MSE: Mean squares error.

TABLE 3: Estimated model parameters.

Model	α	β	k	m
Bordy	12.09	0.63	0.07	
Bertalanffy	12.09	7.71	0.93	0.30
Logistic	12.03	1.33	0.10	
Gen. logistic	12.05	0.009	8.46	0.01
Gompertz	12.05	0.91	0.08	
Richards	12.13	-0.86	0.05	-1.72
Negative exp.	11.87	0.12		
Stevens	12.09	7.71	0.93	
Tanaka	-842.48	0.69	27.71	1.76

Actual cumulative cases and estimations and estimated error which have been made using the all other models are given in Figure 3, 4, 5, 6, 7, 8, 9, 10.

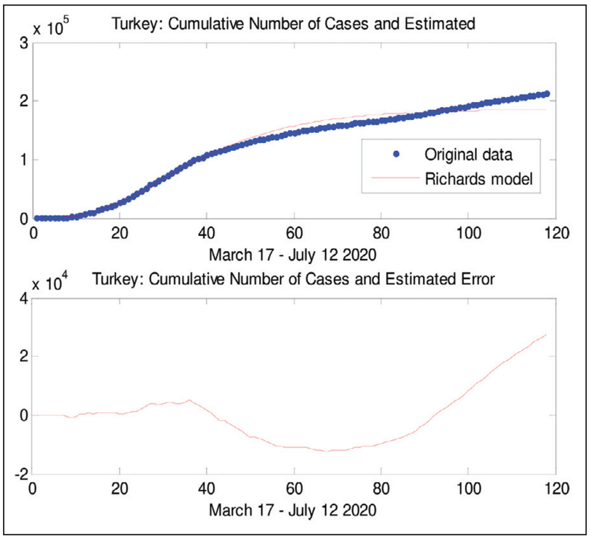

FIGURE 1: Cumulative cases and estimated values and estimated error.

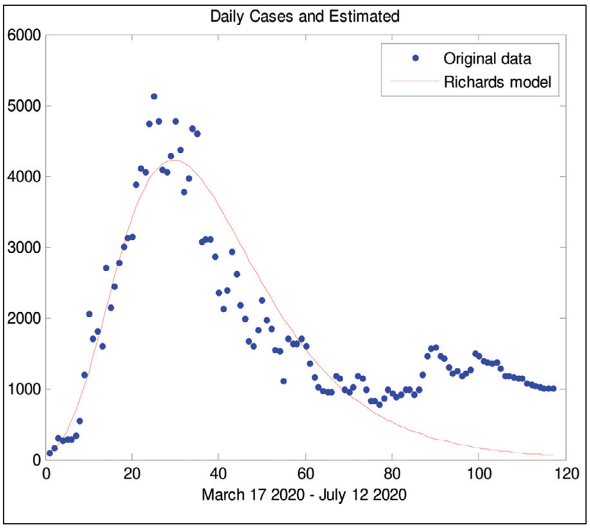

FIGURE 2: Daily cases, actual and estimated values.

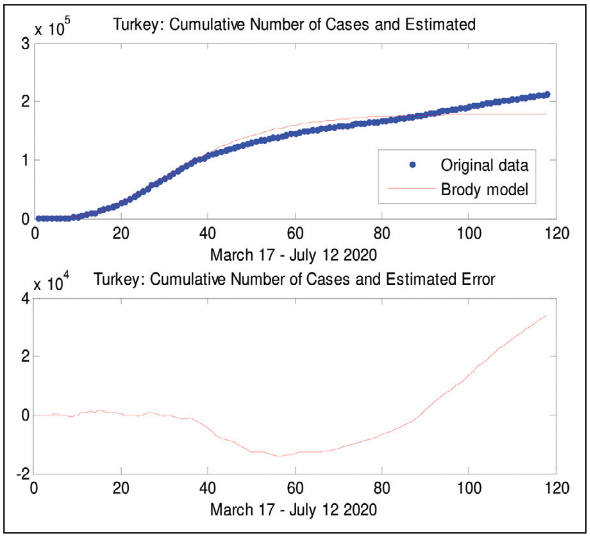

FIGURE 3: Cumulative cases and estimated values and estimated error.

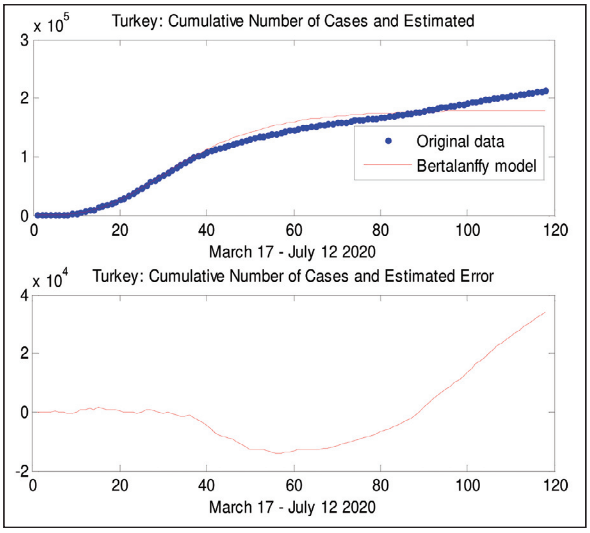

FIGURE 4: Cumulative cases and estimated values and estimated error.

DISCUSSION

In this work, we applied nine growth models to estimate the COVID-19 cases. The results of the modeling show a good fit between the estimated and the observed data on total confirmed cases. The nine non-linear growth models have been explored using the total COVID-19 case counts between March 17-July 12, 2020 in Turkey. These models are Brody, Bertalanffy, Logistic, Generalized Logistic, Gompertz, Richards, Negative Exponential, Stevens, and Tanaka. It may be stated that Richards model is the most consistent in terms of both R^2 and MSE. As a result of estimation with the data between March 17 and July 12, 2020, the best fitting results for modelling the total case counts belong to Richards model in terms of both MSE and R^2. It is proposed that the use of this model is appropriate. Among the studies made on COVID-19 pandemic, modelling the disease progress is remarked primarily. Modelling the disease progress is substantial for the precautions which will be taken by countries, interventions, and administations of treatments. In estimation of the forthcoming process, it will be appropriate to make estimations using Richards model. It should be noted that, the actual daily cases and

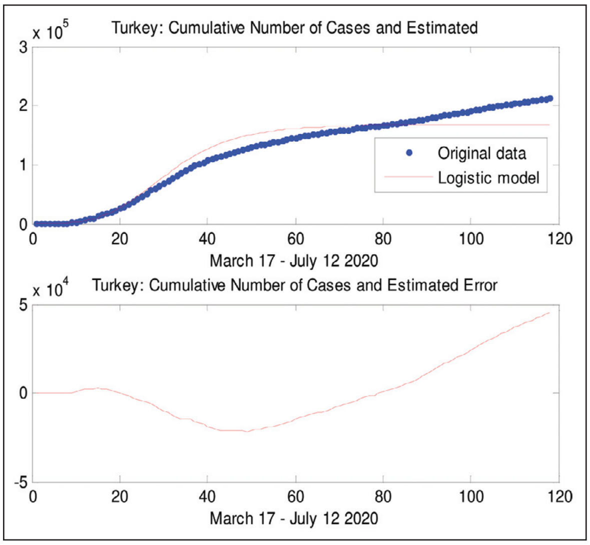

FIGURE 5: Cumulative cases and estimated values and estimated error.

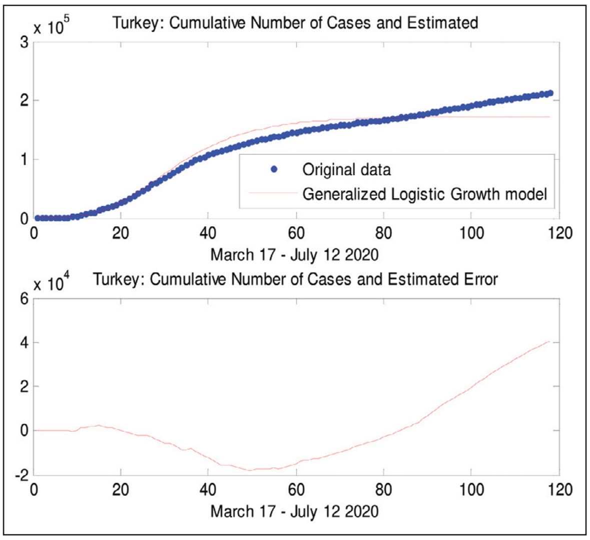

FIGURE 6: Cumulative cases and estimated values and estimated error.

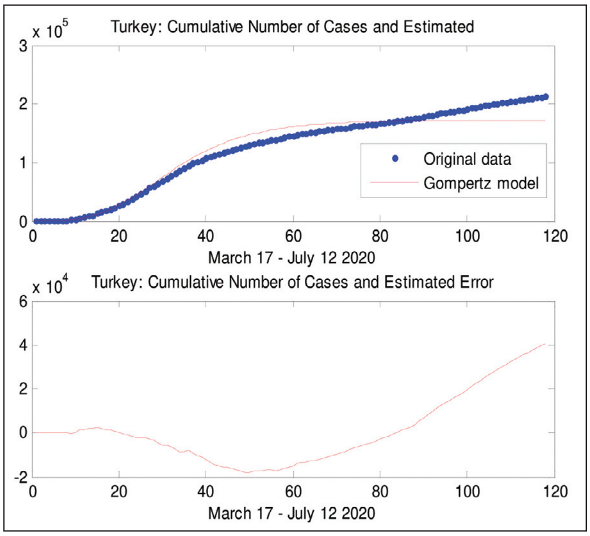

FIGURE 7: Cumulative cases and estimated values and estimated error.

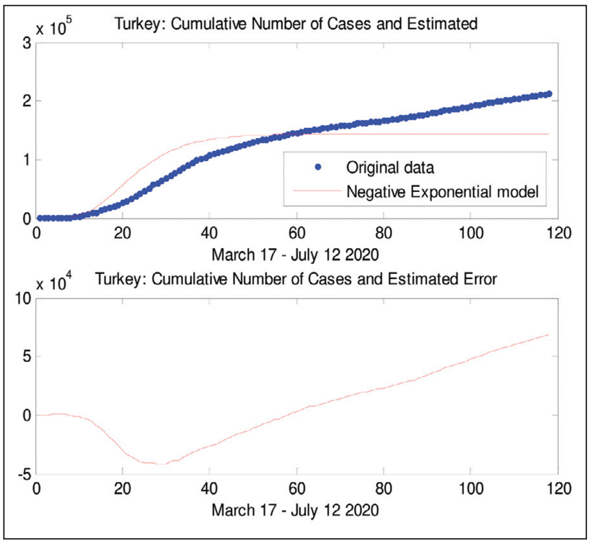

FIGURE 8: Cumulative cases and estimated values and estimated error.

the daily case estimate obtained from the Richards model are not compatible. Forecasting on these growth models can sometimes lead to erroneous results.

In other previous studies that we have cited above on this subject, not all models in this study have been addressed. Only a few of the models we have considered have been studied. The estimation results obtained from the models differ greatly from data to data. Therefore, the data of each country reveals its own parameter estimation. Due to this difficulty, the parameters were not compared. It can be treated as a subject of a separate study.

CONCLUSION

It is recommended to use the Richards model in modeling the cumulative COVID-19 case counts. In estimating the future process, it will be appropriate to make estimations using the Richards model. In addition, estimates of the number of daily cases obtained from the Richards model are not compatible with the actual number of daily cases. This may occur because

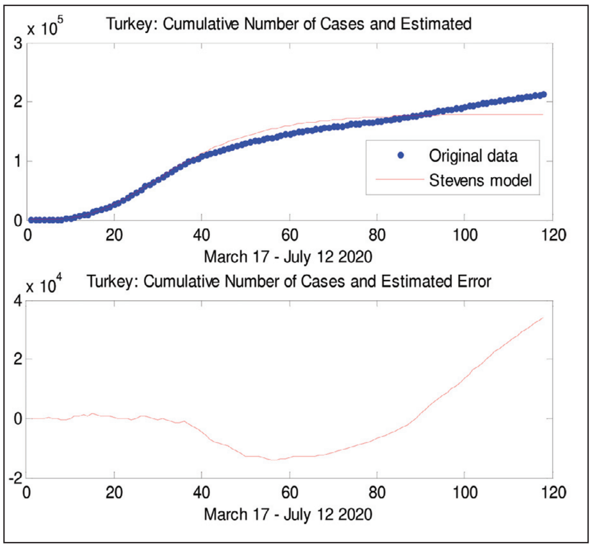

FIGURE 9: Cumulative cases and estimated values and estimated error.

estimations on these growth models can sometimes lead to erroneous results.

We think that it will be more beneficial and realistic if the model's parameter estimations are evaluated by the expert health authorities dealing with COVID-19. The expert health authorities can decide how to use these results in an appropriate way. In this study, we analyzed the growth models that are widely used in the literature and try to determine the most fitting growth model for COVID-19 data.

Other studies in the literature have not covered all the models in our article. The estimation results of the growth models vary due to the data set. Therefore, the data of each country reveals its own parameter estimation. Due to this problem, the parameters are not compared. This situation can be a subject of another study. The aim of this study is to examine the models related with COVID-19 data. In further studies, co-studying with the researchers dealing with various sub-fields of COVID-19 is planned to recognize and analyze the COVID-19 process on more specific and general issues.

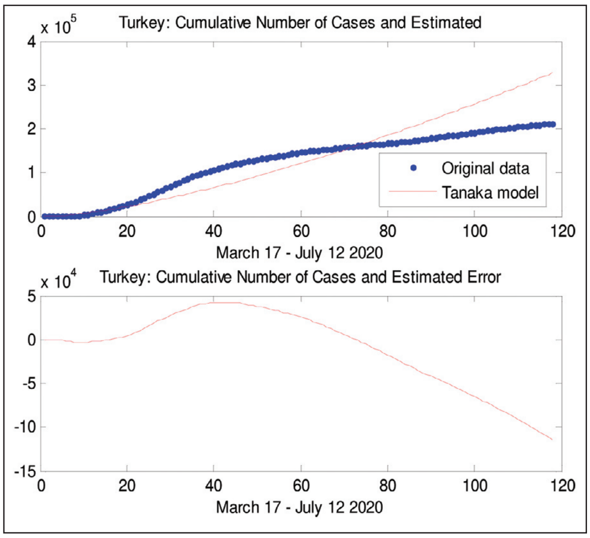

FIGURE 10: Cumulative cases and estimated values and estimated error.

Acknowledgements

This study was carried out without any support from any institution, organization and person. I would like to thank my wife Nuran Taş for her contributions throughout the study.

Source of Finance

During this study, no financial or spiritual support was received neither from any pharmaceutical company that has a direct connection with the research subject, nor from a company that provides or produces medical instruments and materials which may negatively affect the evaluation process of this study.

Conflict of Interest

No conflicts of interest between the authors and / or family members of the scientific and medical committee members or members of the potential conflicts of interest, counseling, expertise, working conditions, share holding and similar situations in any firm.

Authorship Contributions

This study is entirely author's own work and no other author contribution.

REFERENCES

1. Gorbalenya AE, Baker SC, Baric RS, et al. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol. 2020;5:536-44. [Crossref] [Pubmed] [PMC]
2. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med. 2020;382(13):1199-207. [Pubmed] [PMC]
3. World Health Organization [Internet]. Novel coronavirus (2019-nCoV) situation reports. 2020. [Link]
4. Jia L, Li K, Jiang Y, Guo X, Zhao T. Prediction and analysis of coronavirus disease 2019. arXiv. 2020. [Link]
5. Castorina P, Iorio A, Lanteri D. Data analysis on coronavirus spreading by macroscopic growth laws. arXiv. 2020. [Crossref]
6. Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman JM, et al. Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. Infect Dis Model. 2020;5:256-63. [Crossref] [Pubmed] [PMC]
7. Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman JM, et al. Short-term forecasts of the COVID-19 epidemic in Guangdong and Zhejiang, China: february 13-23, 2020. J Clin Med. 2020;9(2):596. [Crossref] [Pubmed] [PMC]
8. Munayco CV, Tariq A, Rothenberg R, Soto-Cabezas GG, Reyes MF, Valle A, et al; Peru COVID-19 working group. Early transmission dynamics of COVID-19 in a southern hemisphere setting: Lima-Peru: February 29th-March 30th, 2020. Infect Dis Model. 2020;5:338-45. [Crossref] [Pubmed] [PMC]
9. Torrealba-Rodriguez O, Conde-Gutiérrez RA, Hernández-Javier AL. Modeling and prediction of COVID-19 in Mexico applying mathematical and computational models. Chaos Solitons Fractals. 2020;138:109946. [Crossref] [Pubmed] [PMC]
10. Mazurek J, Nenickova Z. Predicting the number of total COVID-19 cases in the USA by a Gompertz curve. 2020. [Link]
11. Català M, Alonso S, Alvarez-Lacalle E, López D, Cardona PJ, Prats C. Empiric model for short-time prediction of COVID-19 spreading. medRxiv. 2020. [Crossref]
12. Petropoulos F, Makridakis S. Forecasting the novel coronavirus COVID-19. PLoS One. 2020;15(3):e0231236. [Crossref] [Pubmed] [PMC]
13. Zwietering MH, Jongenburger I, Rombouts FM, van 't Riet K. Modeling of the bacterial growth curve. Appl Environ Microbiol. 1990;56(6):1875-81. [Crossref] [Pubmed] [PMC]
14. Gerlee P. The model muddle: in search of tumor growth laws. Cancer Research. 2013;73(8):2407-11. [Crossref] [Pubmed]
15. Tjørve KMC, Tjørve E. The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the Unified-Richards family. PLoS One. 2017;12(6): e0178691. [Crossref] [Pubmed] [PMC]
16. Dennis B, Ponciano JM, Subhash R, Taper ML, Staples DF. Estimating density dependence, process noise and observation erros. Ecological Monographs. 2006;76(3):323-41. [Crossref]
17. Dennis B, Munholland PL, Scott JM. Estimation of growth and extinction parameters for endangered species. Ecological Monographs 1991;61(2):115-43. [Crossref]
18. Dennis B, Taper ML. Density dependence in time series observations of natural populations: estimation and testing. Ecological Monographs. 1994;64(2):205-24. [Crossref]
19. Berryman AA. Principles of Population Dynamics and Their Application. 1st ed. UK: Taylor & Francis Group; 1999. [Link]
20. Turchin P. Complex Population Dynamics: A Theoretical/Empirical Synthesis. 1st ed. Princeton, New Jersey, USA: Princeton University Press; 2003. [Link]
21. Dennis B, Desharnais RA, Cushing JM, Costantino RF. Nonlinear demographic dynamics: mathematical models, statistical methods, and biological experiments. Ecological Monographs. 1995;65(3):261-81. [Crossref]
22. Hilborn R, Mangel M. The Ecological Detective: Confronting Models with Data. 1st ed. Princeton, New Jersey, USA: Princeton University Press; 1997. [Link]
23. Ives AR, Dennis B, Cottingham KL, Carpenter SR. Estimating community stability and ecological interactions from time-series data. Ecological Monographs. 2003;73(2):301-30. [Crossref]
24. Vincek D, Kralik G, Kušec G, Sabo K, Scitovski R. Application of growth functions in the prediction of live weight of domestic animals. CEJOR. 2012;20(4):719-33. [Crossref]
25. Kshirsagar AM, Smith WB. Introduction. Growth Curves. 1st ed. United State of America: CRC Press; 1995. p.1-57. [Link]
26. Öztürk F, Özbek L. Matematiksel Modelleme ve Simülasyon. 3. Baskı. Ankara: Pigeon Yayıncılık; 2016. [Link]
27. Bilgin ÖC, Esenbuğa N. Doğrusal olmayan büyüme modellerinde parameter tahmini [Parameter estimation in nonlinear growth models]. Hayvansal Üretim. 2003;44(2):81-90. [Link]
28. CSSEGISandData [Internet]. © 2021 GitHub, Inc. COVID-19 data repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. Cited: 1.7.2020. Available from: [Link]