Predicting COVID-19 Peaks Around the World

Constantino Tsallis [1,2,3] and Ugur Tirnakli [4]*

[1] Centro Brasileiro de Pesquisas Fisicas, National Institute of Science and Technology for Complex Systems, Rio de Janeiro, Brazil, [2] Santa Fe Institute, Santa Fe, NM, United States, [3] Complexity Science Hub Vienna, Vienna, Austria, [4] Department of Physics, Faculty of Science, Ege University, Izmir, Turkey

The official data for the time evolution of active cases of COVID-19 pandemics around the world are available online. For all countries, a peak has been either observed (China and South Korea) or is expected in the near future. The approximate dates and heights of those peaks have important epidemiological implications. Inspired by similar complex behavior of volumes of transactions of stocks at the NYSE and NASDAQ, we propose a q-statistical functional form that appears to describe satisfactorily the available data for all countries. Consistently, predictions of the dates and heights of those peaks in severely affected countries become possible unless efficient treatments or vaccines, or sensible modifications of the adopted epidemiological strategies, emerge.

Keywords: COVID-19, pandemics, complex systems, non-extensive statistical mechanics, epidemiology

It is possible to predict the thermostatistical properties of uncountable physical systems at thermal equilibrium through the one-body distribution $p(\epsilon) \propto \omega(\epsilon)e^{-\beta\epsilon}$, where $\omega(\epsilon)$ is the density of states as a function of the energy ϵ, multiplied by the celebrated Boltzmann factor $e^{-\beta\epsilon}$, β being the inverse temperature. The function $\omega(\epsilon)$ comes from mechanical considerations (classical, quantum, relativistic) related with the number of degrees of freedom and does not depend on the temperature; the exponential weight comes instead from standard statistical-mechanical considerations. In many cases it is, either exactly or approximatively, $w(\epsilon) \propto \epsilon^{\alpha}$ ($\alpha \in R$). For the thermostatistical properties of the stationary- or quasi-stationary-state of wide classes of complex systems, the Boltzmann factor is to be generalized into the q-exponential factor $e_q^{-\beta_q \epsilon} \equiv [1-(1-q)\beta_q \epsilon]^{1/(1-q)}$ ($q \in R$; $e_1^z = e^z$) [1–3]. This procedure yielded quite satisfactory results for high-frequency stock-markets, such as the NYSE, NASDAQ, and others [4–6].

Let us focus now on the data available for the COVID-19 pandemics. Soon after the beginning of the pandemics, several studies analyzing the available data and employing different models and candidate functions started to appear in the literature [7–12]. Most of them are interested in the behavior of total cases and fatality curves. We will concentrate here on the analysis of the active cases and deaths *per day*. Inspection of the public data[1] (updated on a daily basis) and, in particular, of the time evolution of the number N of active cases (surely a lower bound of the unknown actual numbers) showed a rather intriguing similarity with the distributions of volumes of stocks. Along this line, we adopt the following functional form for each country or region:

$$N = C(t-t_0)^{\alpha} e_q^{-\beta(t-t_0)^{\gamma}} = \frac{C(t-t_0)^{\alpha}}{[1+(q-1)\beta(t-t_0)^{\gamma}]^{1/(q-1)}}, \qquad (1)$$

---

[1] Available online at: https://www.worldometers.info/coronavirus/#countries (accessed May 08, 2020).

FIGURE 1 | Two possible fits for the evolution of active COVID-19 cases in **(A)** China and **(B)** South Korea: the dots are the data available at 28 April 2020, and the lines are fits using Equation (1). For China, a strange kink is present in the ascending part of the data curves, which means that it is not possible to make a single fit that would satisfactorily account for both the increasing and decreasing parts. We here present two log-log representations of different curves that describe either the increasing part or the decreasing part but not both. Inset: linear-linear representation of the same data.

with $C > 0, \alpha > 0, \beta > 0, \gamma > 0, q > 1$, and $t_0 \geq 0$. The constant t_0 indicates the first day of appearance of the epidemic in that particular region; it is conventionally chosen to be zero for China; for the other countries, it is the number of days elapsed between the appearance of the first case in China and the first case in that country. The normalizing constant C reflects the total population of that particular country. For $\alpha = 0$, if $\gamma = 1$, we recover the standard q-exponential expression; if $\gamma = 2$, it is currently referred to in the literature as q-Gaussian; for other values of γ, it is referred to as stretched q-exponential. Through the inspection of the roles played by the four non-trivial parameters, namely $(\alpha, \beta, \gamma, q)$, it became clear that (α, β) depend strongly on the epidemiological strategy implemented in that region in addition to the biological behavior of the coronavirus in that geographical climate. In contrast, the parameters (γ, q) appear to be more universal, mainly depending on the coronavirus. Therefore, we investigate several countries that have not reached their peaks yet, with the basic assumption

that these two parameters would not change much from one country to another, and we fixed these values at the values that we determine for China, since this country has already had nearly the full evolution of the pandemic. This assumption seems to be working. For other countries whose peaks have already been reached, we use the same functional form (1) but adjusting all parameters for a better fit. The results for China and South Korea are given in **Figure 1**. It is evident that, although the functional form (1) does yield satisfactory results for both China and South Korea, the (γ, q) parameter values differ somewhat for each of these countries. On the other hand, as can easily be seen from **Figure 2**, our assumption is corroborated by several countries that we have numerically analyzed. In **Table 1**, we present the forecasted dates and heights of the peaks, as well as the values of the fitting parameters using the data accumulated until April 28, 2020.

We also test our formula (1) and assumption for the evolution of deaths per day. Again, we fixed the (γ, q) parameter values at China's results, which can be seen in **Figure 3**, and tried to fit the data of the same six countries. It is quite surprising to see that the results given in **Figure 4** seem to suggest that it is also possible to fit the evolution of deaths per day without changing these two parameters.

We may summarize as follows. The death curve of South Korea is atypical in the sense that it sensibly differs from all of those that we analyzed in the present work. Because of that, we have not included it here. We remind the reader that the values for (γ, q) used for the South Korea evolution curve of active cases also differ from those used for all the other countries, which reinforces the fact that some sort of exceptionality exists there for reasons that are unknown to us. For all countries that have not reached their peak values yet, we have adopted the values of (γ, q) obtained from the inspection of the entire curve of China. In all cases, we have dismissed the form of the short initial transient after the appearance of the first active case. The extrapolation procedure is tested in various countries, as indicated in **Figure 5**. We have indicated, for four typical countries, how the predicted day and height of the peak evolves with time while gradually incorporating the newly available data (which not only add recent information but also modify old public information, even the day of first appearance of a Covid-19 case in a particular country). As we can see, the peak date is more robustly predicted than the peak height. Indeed, 1 and even 2 months before, it has been possible to correctly predict the date within the span of 1 week. The prediction for the height is more sensitive to the new information and can easily fluctuate between simple and double depending on the country and its pandemic health strategy (or lack of it), in particular as concerns population mobility.

We verify that the present work appears to belong to the realm of complex systems, which includes not only, as mentioned above, high-frequency financial transactions (with $\alpha > 0$) [4, 5] but also anthropological issues, such as medieval trading networks and biotech intercorporate networks (with $\alpha < 0$) [13], and relaxation in spin-glasses [14], as well as q-Weibull distribution-like systems [15–18], which correspond to the particular case $\alpha = \gamma - 1$. It remains open as a highly desirable
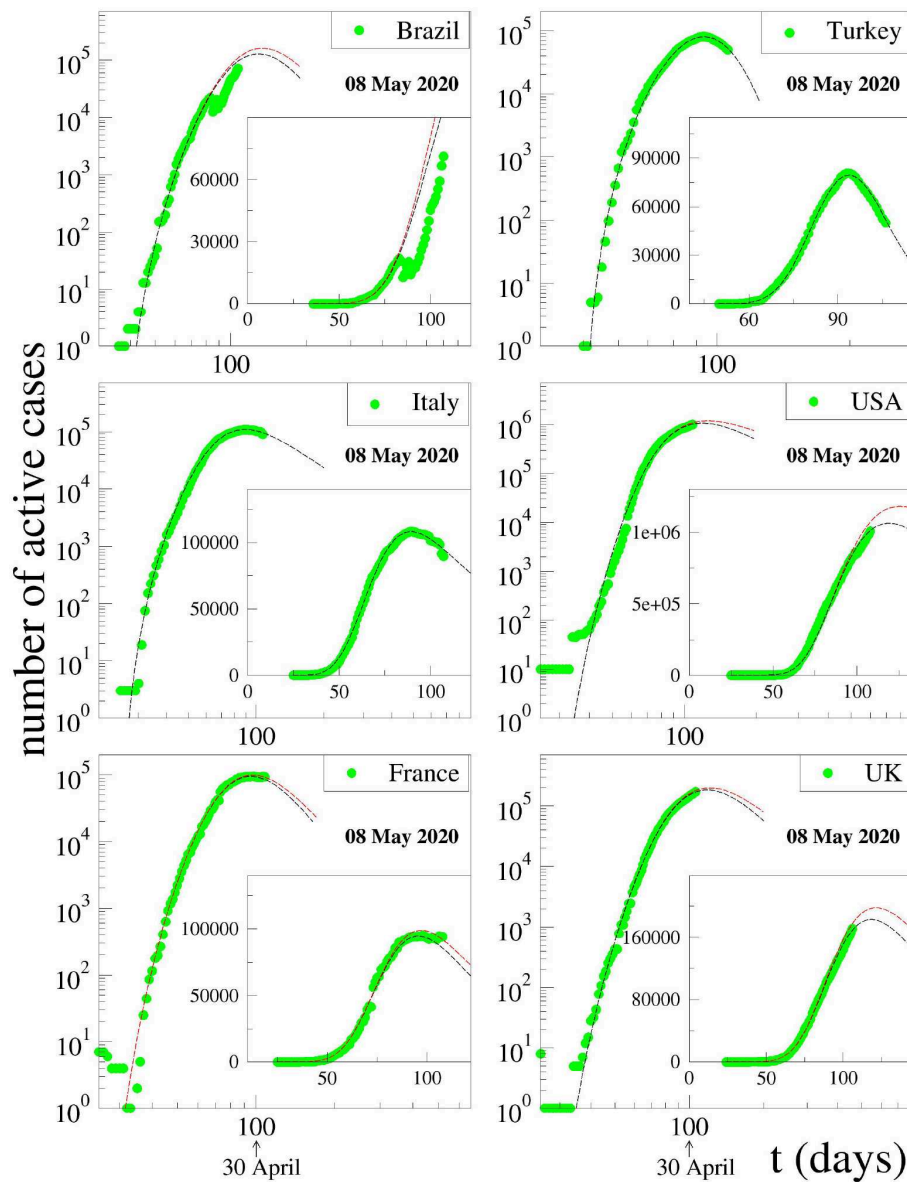
FIGURE 2 | Fits of the data for active cases available on 08 May 2020 for various severely affected countries around the world with Equation (1). The fitting parameters γ and q are fixed at the values for China for all countries that have not yet reached their peak values. Notice that, in the case of Brazil, a disruption occurred in the publicly available data in mid-April. We do not know the cause of this. Coincidentally, however, precisely at that time, the President of Brazil decided to change his Minister for Health.
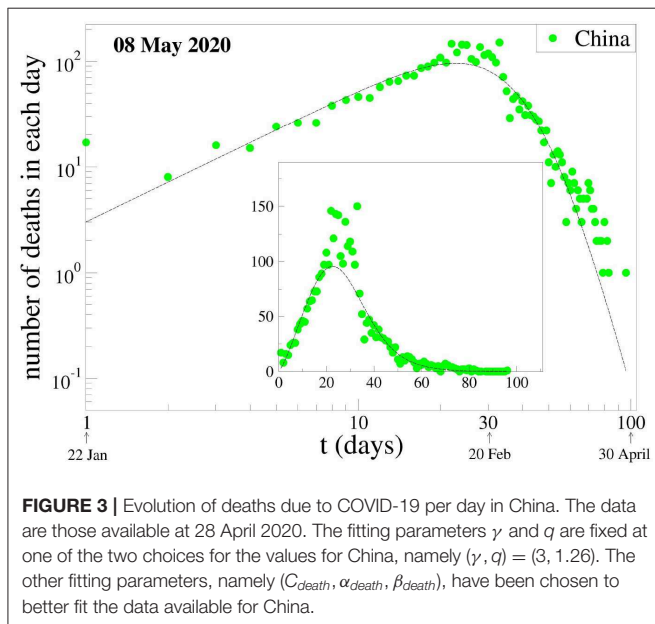
goal to formulate a model which, along lines somewhat similar to epidemiological models, such as the SIR one, would predict an evolution curve, such as the present Equation (1). To be more precise, a great variety of SIR-like epidemiological models have been proposed in the literature. They typically yield an increase before the peak quite that is similar to the decrease after the peak. The behavior in our present Equation (1) is at variance with such characteristics, since it provides a power-law increase (with a typically large positive exponent α as illustrated in the table) and a quite different power-law decrease (with negative exponent given by $[\alpha - \gamma/(q-1)]$, the absolute value of which can independently

be either larger or smaller than α) after the peak. We emphasize that such increase-decrease quantitative behaviors appear to satisfactorily conform to reality, in contrast with the logistic-like growth behavior typical of most SIR-like models. An important issue remains to be clarified, namely the conditions under which the values of (γ, q) could indeed be (strictly or nearly) universal and essentially determined by the biology of the infecting agent, such as the present COVID-19. Let us also mention that, quite obviously, there is nearly everywhere a severe under-notification of the publicly available data for active cases (and even deaths). The real number could easily be 10 times larger, depending on
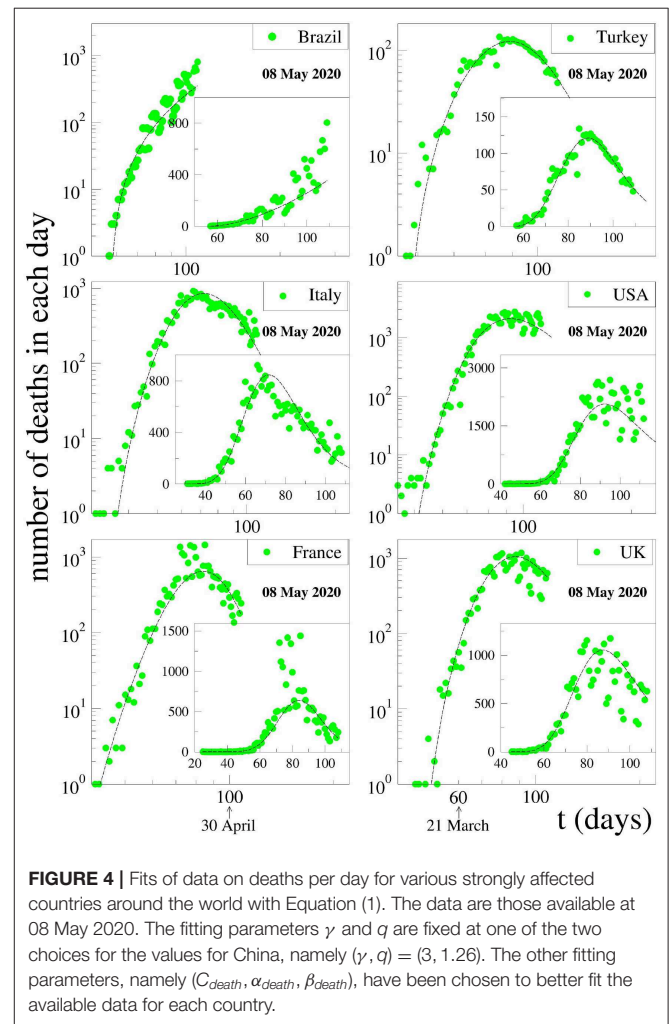
TABLE 1 | The values of the parameters and predictions for the maximum number of active cases and for the day at which the maximum will be achieved.

Country	t_0	C	α	β	γ	q	Prediction interval of peak height	Prediction interval of peak day
				γ and q parameters are fixed at China's values if peak has not yet been achieved				
Brazil	34	1.55×10^{-5}	5.58	4.00×10^{-6}	3	1.26	[145112, 159580]	[132, 134]
				3.30×10^{-6}	3.037	1.28	[0.00069,0.00076]	[1 June 1, June 3]
Turkey	48	8.0×10^{-3}	4.625	1.52×10^{-5}	3.025	1.02	Achieved at 80808	Achieved on day 93 April 23
Italy	24	4.0×10^{-3}	4.68	1.10×10^{-5}	3.083	1.43	Achieved at 106100	Achieved on day 96 April 26
USA	24	3.4×10^{-9}	8.60	1.65×10^{-5}	3	1.26	[1061270, 1178556]	[119, 126]
				1.46×10^{-5}	3.037	1.28	[0.0032,0.0036]	[May 19, May 26]
France	24	1.6×10^{-6}	6.56	9.20×10^{-6}	3	1.26	Achieved at 95365	Achieved on day 98 April 28
				1.19×10^{-5}	3.037	1.28		
UK	24	1.6×10^{-8}	7.50	8.50×10^{-6}	3	1.26	[182659, 197775]	[118, 121]
				8.00×10^{-6}	3.037	1.28	[0.0027,0.0030]	[May 18, May 21]

The second line in the column for the prediction interval of the peak height of active cases shows the same values divided by the total population of the country.

FIGURE 3 | Evolution of deaths due to COVID-19 per day in China. The data are those available at 28 April 2020. The fitting parameters γ and q are fixed at one of the two choices for the values for China, namely $(\gamma, q) = (3, 1.26)$. The other fitting parameters, namely $(C_{death}, \alpha_{death}, \beta_{death})$, have been chosen to better fit the data available for China.

the particular region. However, the consequences of this lack of important information onto the real number of deaths are somewhat mitigated by making use of the *case fatality rate*, which is relatively stable throughout recent weeks for a given country and can be found at websites, such as[2,3]. Finally, the present prediction algorithm could, in principle, be included within an

---

[2] Available online at: https://www.cebm.net/covid-19/global-covid-19-case-fatality-rates/ (accessed May 08, 2020).

[3] Available online at: https://coronavirus.jhu.edu/data/mortality (accessed May 08, 2020).

FIGURE 4 | Fits of data on deaths per day for various strongly affected countries around the world with Equation (1). The data are those available at 08 May 2020. The fitting parameters γ and q are fixed at one of the two choices for the values for China, namely $(\gamma, q) = (3, 1.26)$. The other fitting parameters, namely $(C_{death}, \alpha_{death}, \beta_{death})$, have been chosen to better fit the available data for each country.
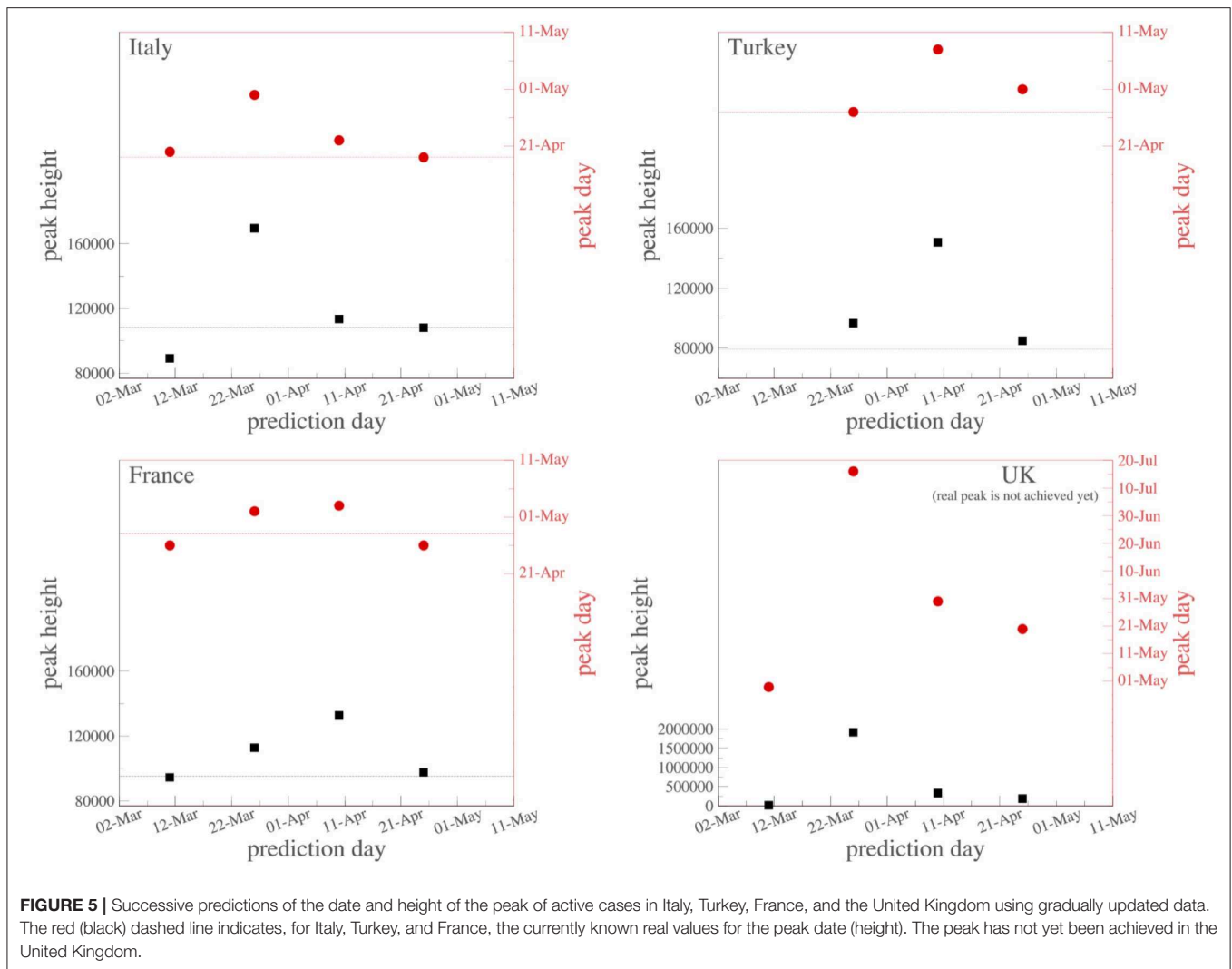
FIGURE 5 | Successive predictions of the date and height of the peak of active cases in Italy, Turkey, France, and the United Kingdom using gradually updated data. The red (black) dashed line indicates, for Italy, Turkey, and France, the currently known real values for the peak date (height). The peak has not yet been achieved in the United Kingdom.

internet *app*, which could access the data publicly available at a given website and automatically update the predicted dates and heights of the disease peaks of epidemics, such as the present Covid-19 one. Any initiative along these lines would be highly welcome.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.worldometers.info/coronavirus/#countries.

REFERENCES

1. Tsallis C. Possible generalization of Boltzmann-Gibbs statistics. *J Stat Phys.* (1988) **52**:479. doi: 10.1007/BF01016429
2. Gell-Mann M, Tsallis C. *Nonextensive Entropy–Interdisciplinary Applications.* New York, NY: Oxford University Press (2004).
3. Tsallis C. *Introduction to Nonextensive Statistical Mechanics–Approaching a Complex World.* New York, NY: Springer (2009).

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

4. Tsallis C, Anteneodo C, Borland L, Osorio R. Nonextensive statistical mechanics and economics. *Phys A.* (2003) **324**:89. doi: 10.1016/S0378-4371(03)00042-6
5. Osorio R, Borland L, Tsallis C. Distributions of high-frequency stock-market observables. In: Gell-Mann M, Tsallis C, editors. *Nonextensive Entropy–Interdisciplinary Applications.* New York, NY: Oxford University Press (2004). p. 321.

6. Ruiz G, de Marcos AF. Evidence for criticality in financial data. *Eur Phys J B.* (2018) **91**:1. doi: 10.1140/epjb/e2017-80535-3

7. Zlatic V, Barjašić I, Kadović A, Štefančić H, Gabrielli A. Bi-stability of SUDR+K model of epidemics and test kits applied to COVID-19. (2020) *arXiv [Preprint]*. arXiv:2003.08479.

8. Vasconcelos GL, Macêdo AMS, Ospina R, Almeida FAG, Duarte-Filho GC, Brum AA, et al. Modelling fatality curves of COVID-19 and the effectiveness of intervention strategies. *medRxiv.* (2020). doi: 10.1101/2020.04.02.20051557

9. Piccolomini EL, Zama F. Preliminary analysis of COVID-19 spread in Italy with an adaptive SEIRD model. (2020) *arXiv [Preprint]*. arXiv:2003.09909.

10. Perc M, Gorišek Miksić N, Slavinec M, Stožer A. Forecasting COVID-19. *Front Phys.* (2020) **8**:127. doi: 10.3389/fphy.2020.00127

11. Pluchino A, Biondo AE, Giuffrida N, Inturri G, Latora V, Le Moli R, et al. A novel methodology for epidemic risk assessment: the case of COVID-19 outbreak in Italy. (2020) *arXiv [Preprint]*. arXiv:2004.02739.

12. Bastos SB, Cajueiro DO. Modeling and forecasting the early evolution of the Covid-19 pandemic in Brazil. (2020) *arXiv [Preprint]*. arXiv:2003.14288v2.

13. White DR, Kejzar N, Tsallis C, Farmer D, White S. A generative model for feedback network. *Phys Rev E.* (2006) **73**:016119. doi: 10.1103/PhysRevE.73.016119

14. Pickup RM, Cywinski R, Pappas C, Farago B, Fouquet P. Generalized spin glass relaxation. *Phys Rev Lett.* (2009) **102**:097202. doi: 10.1103/PhysRevLett.102.097202

15. Picoli S Jr, Mendes RS, Malacarne LC. *q*-exponential, Weibull, and *q*-Weibull distributions: an empirical analysis. *Phys A.* (2003) **324**:678. doi: 10.1016/S0378-4371(03)00071-2

16. Assis EM, Borges EP, Vieira de Melo SAB. Generalized *q*-Weibull model and the bathtub curve. *Int J Qual Reliab Manage.* (2013) **30**:720. doi: 10.1108/IJQRM-Oct-2011-0143

17. Podesta TSV, Rosembach TV, Dos Santos AA, Martins ML. Anomalous diffusion and *q*-Weibull velocity distributions in epithelial cell migration. *PLoS ONE.* (2017) **12**:e0180777. doi: 10.1371/journal.pone.0180777

18. Xu M. *Using the q-Weibull Distribution for Reliability Engineering Modeling and Applications.* College Park, MD: University of Maryland (2019).