Quantifying the unexpected: A scientific approach to Black Swans

Giordano De Marzo [1,4,5] Andrea Gabrielli,[1,2,3] Andrea Zaccaria [2,1,*] and Luciano Pietronero[1,2,4]

[1]*Centro Ricerche Enrico Fermi, Piazza del Viminale, 1, I-00184 Rome, Italy*
[2]*Istituto dei Sistemi Complessi (ISC) - CNR, UoS Sapienza, P. le A. Moro, 2, I-00185 Rome, Italy*
[3]*Dipartimento di Ingegneria, Università degli Studi Roma Tre, Via Vito Volterra 62, I-00146 Rome, Italy*
[4]*Dipartimento di Fisica Università "Sapienza," P. le A. Moro, 2, I-00185 Rome, Italy*
[5]*Sapienza School for Advanced Studies, "Sapienza," P. le A. Moro, 2, I-00185 Rome, Italy*

Many natural and socioeconomic systems are characterized by power-law distributions that make the occurrence of extreme events not negligible. Such events are sometimes referred to as Black Swans, but a quantitative definition of a Black Swan is still lacking. Here, by leveraging on the properties of Zipf-Mandelbrot law, we investigate the relations between such extreme events and the dynamics of the upper cutoff of the inherent distribution. This approach permits a quantification of extreme events and allows to classify them as White, Gray, or Black Swans. Our criterion is in accordance with some previous findings but also allows us to spot new examples of Black Swans, such as Lionel Messi and the Turkish Airlines Flight 981 disaster. The systematic and quantitative methodology we developed allows a scientific and immediate categorization of rare events, also providing insight into the generative mechanism behind Black Swans.

I. INTRODUCTION

During the last decade the complexity paradigm [1–3] has been successfully applied not only to the area of physical systems but also to many other phenomena, including socioeconomic systems [4,5]. One of the most ubiquitous features in the complexity science is the emergence of extreme events, orders of magnitude larger than typical ones. Severe financial crisis, devastating earthquakes, or deadly wars all can be interpreted in terms of complex inherent structures which give rise to power-law distributed phenomena [6]. Mandelbrot and Taleb have been among the first [7,8] to investigate these phenomena and to stress the limits of standard statistical techniques in such frameworks, the latter with the introduction of the celebrated Black Swan metaphor. According to Taleb [8], a Black Swan event.

    (a) is unexpected and unpredictable;

    (b) has a great impact, both positive or negative; and

    (c) makes people try to explain its occurrence once it has been observed.

If extreme events are found to follow a power law, their effects are in some way mitigated because they cease being totally unexpected. In this case Taleb speaks of Gray Swans [8], which are unpredictable but not unexpected. However, as we are going to show, even if we know that the underlying distribution is power-law-like, Black Swans can still occur due to the possible dynamics of the upper cutoff.

The First World War and 9/11 terrorist attacks are often used as examples of Black Swans [8–11]; however, up to now a scientific and quantitative assessment of Black Swans has been lacking, and this sometimes allows policy makers to improperly use this terminology. For instance, during the last months many governments, financial institutions, and journals identified the COVID-19 pandemic as a Black Swan, even though the possibility of the spread of a newborn virus had been already pointed out by several studies and even by Taleb himself [8]. This arbitrariness is favored by the fact that after the efforts of Taleb and Mandelbrot, most of the discussion about Black Swans is still conducted on a qualitative philosophical and/or psychological level.

In the following we show, on a quantitative basis, that Black Swans are related to a nonstationarity in the inherent power-law distribution, generated by a jump dynamics of the upper cutoff. This implies that when performing risk analysis, the assumption of a time-independent probability distribution may result in severe underestimations of extreme events. A central property of Black Swans is that they cannot be predicted starting from data about the system under consideration; however, an analysis that also includes the environment the system is coupled to permits, in some cases, the jumps of the upper cutoff to be perceived and so to foresee Black Swans. This perspective allows us to introduce a quantitative measure of the surprise associated with large events, that we call Blackness, which can be used to scientifically classify them into three distinct sets: White Swans, Gray Swans, and Black Swans. Our quantitative criterion partly confirms what was previously stated by Taleb using qualitative arguments, namely, that the First World War and 9/11 terrorist attacks

*Corresponding author: andrea.zaccaria@cnr.it

were Black Swans with respect to the number of casualties, while the 1987 Black Monday was only a Gray Swan. New examples of Black Swans we have found are the amount of goals scored by Lionel Messi in LaLiga and the number of casualties of the Turkish Airlines Flight 981 disaster, since both events drastically overcame the previous estimate of the upper cutoff.

II. RESULTS

A. Zipf-Mandelbrot law

Our scientific framework to analyze Black Swans is given by Zipf's law and, more generally, power laws. Note that one can observe extreme events also in connection with other probability distributions and not only as a result of power laws. However, since many extreme phenomena, such as wars, earthquakes, disasters, or pandemics, are power-law distributed, in this paper we consider only this class of probability distributions. Zipf's law [12] is a ubiquitous scaling law found in many natural and socioeconomic systems [6,13–16]. Given a system composed of N objects and denoting by $S(k)$ the size of the kth largest one, Zipf's law reads

$$S(k) = \frac{S(1)}{k^\gamma}, \qquad (1)$$

where k is the rank, γ is the Zipf's exponent, and $S(1)$ is the empirical maximum that is the largest element or event in the system under consideration. Zipf's law is generally visualized by the so-called rank-size plot, obtained by plotting the ordered sequence of the sizes as a function of their position in the sequence; a straight line in loglog scale is thus expected. However, it is common to observe deviations from Zipf's law at low ranks [17–19], which can be quantified by introducing a parameter Q in the so-called Zipf-Mandelbrot law [20]:

$$S(k) = \frac{\bar{S}}{(k+Q)^\gamma}. \qquad (2)$$

The parameter $Q > 0$ will play a crucial role in the analysis of Black Swans. As shown in [17], the Zipf-Mandelbrot law is observed whenever the inherent distribution is a power law, and its parameters are related to those of the probability distribution by the following relations:

$$\begin{cases} \gamma = \frac{1}{\alpha-1} \\ \bar{S} = N^\gamma s_m \\ Q = N(*)\frac{s_m}{s_M}^{1/\gamma} \end{cases}, \qquad (3)$$

where α is the exponent of the inherent power-law distribution, that is, $P(S) \sim S^{-\alpha}$. Note that here and in the following we always refer as α to the exponent of the probability density function, not to be confused with the exponent of the cumulative. s_m and s_M are the lower and upper cutoffs of the distribution, possibly corresponding to physical limits, that is, $P(S) = 0$ for $S < s_m$ and $S > s_M$, and N is the number of elements in the system or records in the catalog.

The deviation parameter Q is a measure of the level of sampling [17] (see Methods section for details): for $Q \approx 0$ the underlying distribution is undersampled and the upper cutoff cannot be inferred, while for $Q \gg 1$ it is completely sampled and s_M coincides with the empirical maximum. Indeed, by combining Eqs. (2) and (3) we obtain an expression relating the upper cutoff of the distribution to the measurable parameters Q, $S(1)$, and γ:

$$s_M = S(1)\left(\frac{Q+1}{Q}\right)^\gamma. \qquad (4)$$

In the limit $Q \to 0$ the upper cutoff diverges, meaning that data do not provide a sufficient level of sampling for inferring it. This last expression thus allows, when Q is sufficiently large, to compute the upper cutoff of a power-law distribution and will be used in the following to study real systems.

B. The Blackness

Let us consider a random number generator extracting values from a power-law distribution with unknown parameters α, s_m, and s_M. If we look at the first $N \ll (s_M/s_m)^{(1/\gamma)}$ numbers, being $Q \ll 1$, the corresponding rank-size plot will be straight [17]. Moreover, our set undersamples the inherent distribution, and so there would be no surprise if the next draw returns a value much larger than those previously observed, because it is not possible to infer the upper cutoff s_M.

If we keep drawing numbers, sooner or later we will completely sample the inherent distribution, this occurring for $N \gg (s_M/s_m)^{(1/\gamma)}$. In this way the empirical maximum becomes very close to the upper cutoff, that can thus be inferred using Eq. (4), and nothing unexpected should occur. In this situation Black Swans are absent, but what happens if the upper cutoff increases to $s'_M \gg s_M$? (Note that this actually happens in various social and biological systems, for instance, as a consequence of the coupling with an external system or a technological change.) The answer is simple. Our apparently perfect knowledge of the underlying distribution becomes problematic as soon as a number close to s'_M is extracted. Without knowing if and when the upper cutoff jumps, there is no way not only to predict, but also to expect such an event. This kind of event is what we will classify as a Black Swan.

This simple example we sketched captures some points that are crucial for understanding the phenomenology of Black Swans (and possibly mitigate their utmost effects):

1. An arbitrary large event can be or not be a Black Swan, depending on the level of sampling of the distribution, so depending on the value of Q. A system showing exact Zipf's law ($Q \approx 0$) never gives rise to Black Swans. The idea is that when the level of sampling is low there is no way to characterize the upper limit and so to realize that $s_M \to s'_M$.

2. An event can be classified as a Black Swan if and only if it is (much) larger than any event previously observed *and* the deviation parameter Q of the system is large. Indeed, this implies that it is beyond the estimate of the upper limit.

3. A stationary power law only gives rise to Gray Swans, since when the level of sampling is low there is no surprise if an event much larger than those previously observed occurs, while when the level of sampling is high, no event much larger than the observed maximum can occur. Note that for a power law without upper cutoff, the level of sampling is always low, since the support of the distribution is not finite. As a consequence, also in this case, no Black Swan can be observed.
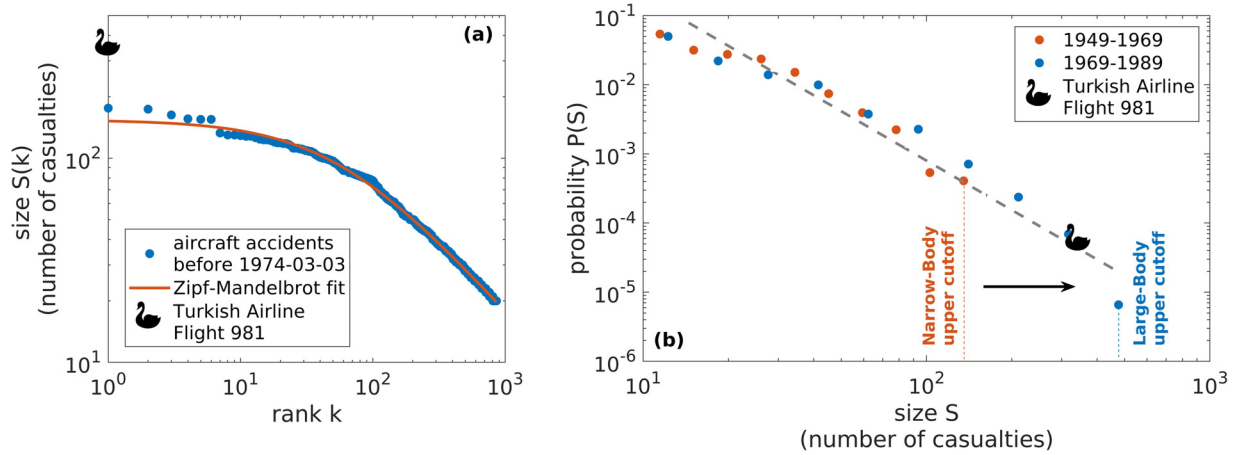
FIG. 1. Black Swans arise from jumps in the upper cutoff. (a) Rank-size plot of airline disaster casualties occurring before 1974-03-03 (blue dots) with the corresponding fit to Zipf-Mandelbrot law (red line). The presence of strong deviations from Zipf's law would have suggested that this set completely sampled the inherent distribution, meaning that the maximal possible number of casualties (i.e., the upper cutoff) was approximately 200. The stylized black swan corresponds to the Turkish Airlines Flight 981 disaster. 346 people died in that circumstance, subverting the estimate of the upper cutoff. This accident, which has been the first to involve a large-body aircraft, can be then considered a Black Swan. (b) Probability distribution of airline disaster casualties. The red dots represent the distribution of accidents occurring between 1949 and 1969, while blue ones show the distribution of those occurring between 1969 and 1989. Approximately the same number of accidents occurred in these two periods (798 and 781, respectively) and so if the distribution would not have changed, then red and blue dots would have spanned the same range of casualties. However, the upper cutoff approximately doubles, making the Black Swan possible.

In order to better clarify these points, we consider a real system showing this kind of dynamics: aircraft accidents. Here the size of the event is given by the number of casualties. In Fig. 1(a) we show the corresponding rank-size plot computed by considering all events until the February 3, 1974, the day before the crash of Turkish Airlines Flight 981, which is represented by the Black Swan. Strong deviations from Zipf's law are present, and Q is relatively large. On the basis of the previous considerations, one could have concluded that the upper cutoff of the distribution had already been reached, and so no surprise should have been expected. However, the following day Flight 981 crashed, provoking the death of 346 people, a number approximately twice as large as the previous most severe accident. This increase was possible because of the introduction in the late 1960s of wide-body aircrafts that can carry approximately twice as many passengers as the older narrow-body models. This implied a sudden increase of the upper cutoff of the casualty distribution, which we depict in Fig. 1(b). In this sense we can conclude that the Flight 981 was indeed a Black Swan. Note that before and after the jump, the scaling exponent remained the same. In summary, usually an empirical power law is characterized by three parameters, the exponent and the two cutoffs, and it must be nonstationary in order to give rise to Black Swans. Now, the lower cutoff clearly does not play any role regarding large events, while a variation of the exponent influences the frequency of extreme events but not their size. As a consequence, the only form of nonstationarity that may produce Black Swans is a jump of the upper cutoff. Note that we use the term jump, meaning that the typical time of the increase of the upper cutoff must be much smaller than the typical time of the sampling. Also, a continuous growth of the upper cutoff can produce Black Swans, provided that this condition is satisfied.

Once the phenomenology of extreme events has been discussed, it is simple to derive a quantitative criterion to classify them in Black, Gray, and White Swans: (i) an event smaller than the maximum of those which previously occurred is a White Swan; (ii) a Gray Swan is an event whose size is larger than those previously observed, but that is not unexpected in the sense that it is lower than the estimated upper cutoff; and (iii) a Black Swan is an event whose size is larger than those previously observed and that is also unexpected, being larger than the expected upper cutoff. We can then define the *Blackness* β of a new event of size S_{new} as

$$\beta = \frac{S_{\mathrm{new}} - S(1)}{s_M - S(1)}, \qquad (5)$$

where $S(1)$ is the empirical maximum and s_M is the best estimate of the upper cutoff we can infer from the data, which is given by Eq. (4). Formalizing the three definitions just mentioned, β can be used to determine the "color" of an extreme event:

White Swan: $S_{\mathrm{new}} < S(1) \rightarrow \beta < 0$;
Gray Swan: $S(1) < S_{\mathrm{new}} < s_M \rightarrow 0 < \beta < 1$;
Black Swan: $S_{\mathrm{new}} > s_M \rightarrow \beta > 1$.

As shown in the Methods section, β can be expressed directly in terms of only empirical quantities

$$\beta = \frac{\mathcal{R} - 1}{\left(\frac{Q+1}{Q}\right)^{\gamma} - 1}, \qquad (6)$$

where $\mathcal{R} = S_{\mathrm{new}}/S(1)$ is the relative size of the new event with respect to the empirical maximum. This last expression shows that β well summarizes the two crucial points we stressed above. Namely, $\mathcal{R} > 1$ is a necessary but not sufficient condition for an event to be a Black Swan, since also the deviation parameter Q (or, equivalently, the level of sampling) must be
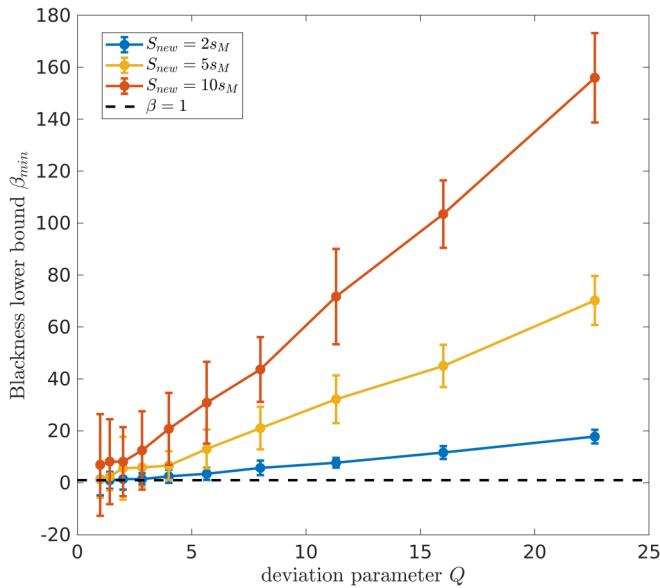
FIG. 2. Synthetic Black Swans. Lower bound of the Blackness $\beta_{\min}$ for events with size $S_{\text{new}} = 2 \times 10^3, 5 \times 10^3, 10 \times 10^3$ as a function of the level of sampling Q obtained generating samples of varying numerosity from a power law with parameters $\alpha = 2$, $s_m = 1$, and $s_M = 10^3$. For each value of Q and S_{new} we generated 100 samples in order to compute the mean value and the uncertainty of $\beta_{\min}$. Starting from Q 5 the events are classified as Black Swans, while if the deviation parameter Q is around 1, even an event 10 times larger than the upper cutoff is often classified only as a Gray Swan. The black dashed lines correspond to $\beta_{min} = 1$ and thus divide Black and Gray Swans.

considered. Indeed, if $Q = 0$, that is, for a perfect Zipf's law, the Blackness goes to zero independently of the relative size. This because in the case of a pure Zipf's law the upper cutoff cannot be estimated.

Finally, in order to understand how the Blackness is influenced by the level of sampling, we applied Eq. (6) to synthetic data. More precisely, we generated samples from a power law with $\alpha = 2$, $s_m = 1$, and $s_M = 10^3$ and with varying levels of sampling Q. For each sample we considered an event with size $S_{\text{new}} = 2s_M, 5s_M, 10s_M$ larger than the upper cutoff (and thus not compatible with the power-law distribution), and we computed the lower bound $\beta_{\min}$ of its blackness (see the Appendix for details on how we compute this lower bound) using a 90% confidence interval. For each value of the level of sampling Q and of S_{new} we repeated the procedure 100 times so to compute both the mean value of $\beta_{\min}$ and its uncertainty. Results are reported in Fig. 2, as it is possible to see the events are always correctly classified as Black Swans as soon as the deviation parameter is $Q \sim 5$, while when the level of sampling is smaller we obtain a Blackness smaller than 1. As we expected, it is thus necessary to have a sufficiently large level of sampling in order to observe Black Swans.

C. Applications to real systems

1. Classification of past events

The Blackness concept allows to search for Black or Gray swans in any system showing an inherent power-law distribu-

tion, obtaining an objective and scientific-based measure of the "surprise" associated to an event. By performing fits to the rank-size plots (see the Methods section for details), and by using Eq. (6), we computed the Blackness of a number of extreme events that previous works have found to be power-law distributed [7,21–25] and that range from sport to natural disasters and finance; the results are displayed in Fig. 3. The "Blackest" Black Swan among the events we considered is the Turkish Airlines disaster, followed by WWI; the size of the latter has been defined as the number of casualties normalized by the world population. For what concerns WWI, the growing globalization of the world and the extensive usage of new deadly weapons may be responsible for the jump of the upper cutoff. Surprisingly, WWII is only a Gray Swan. Intuitively, the occurrence of the Great War, a conflict much more severe than any previous interstate war, already proved that the upper cutoff of the distribution had increased, making the second world conflict relatively less surprising in terms of casualties. Another Swan that presents a Blackness lower than 1 (considering uncertainty) is the Black Monday of 1987. Here the size is given by the module of Dow Jones index daily return; during Black Monday this index lost about 23% of its value, the worst fall in its history. Also other "officially unexpected" events in finance and economics are found to be Gray Swans as for instance the oil market volatility due to the spreading of COVID-19. The event we analyzed is the largest monthly fluctuation of the West Texas Intermediate (WTI) index (+75%, occurring between March and April 2020) and we considered pre-COVID-19 data (1986–2019) for computing its Blackness. Another example is the growth of China (the size is defined as the percentage variation of GDPppp). Here we considered 35 years of returns of countries between the periods 1865–1900, 1900–1935, and 1950–1985. Countries that grew thanks to oil or natural resources were excluded, but analogous results were found considering all the countries. It is interesting to note that according to the common wisdom the growth of China is an incredible outlier [26], while in our framework its growth is a remarkable but not unexpected event, confirming recent analysis based on the methods of analogs [5,27]. Conversely, the soccer top star Lionel Messi is a Black Swan if measured by the number of goals he scored in LaLiga (455). We can argue that the growth of the upper cutoff is due to the increase of the number of teams taking part in the championships, which, in turn, increased the number of games played in a season and the number of potential goals. Indeed, the previous record holder was Telmo Zarra, who scored 268 goals playing between 1949 and 1957; even if, as Lionel Messi, he scored an average of 0.9 goals per match and played for the same number of seasons (16), the number of teams involved in LaLiga was between 12 and 16, against the 20 of today. Using the same criterion, Cristiano Ronaldo, with his 311 goals, is only a Gray Swan. The September 11 attacks, often used as an archetype of Black Swan events, is characterized by a Blackness much larger than 1, confirming the blackness of this terrorist attack in terms of casualties, even if in this case a clear and objective motivation for the jump of the upper cutoff is hard to find. Finally, the Burj Khalifa, which is the tallest building in the world, is only a Gray Swan, despite a height approximately 70% larger than Taipei 101, the previous record holder. Setting $\beta = 1$ in
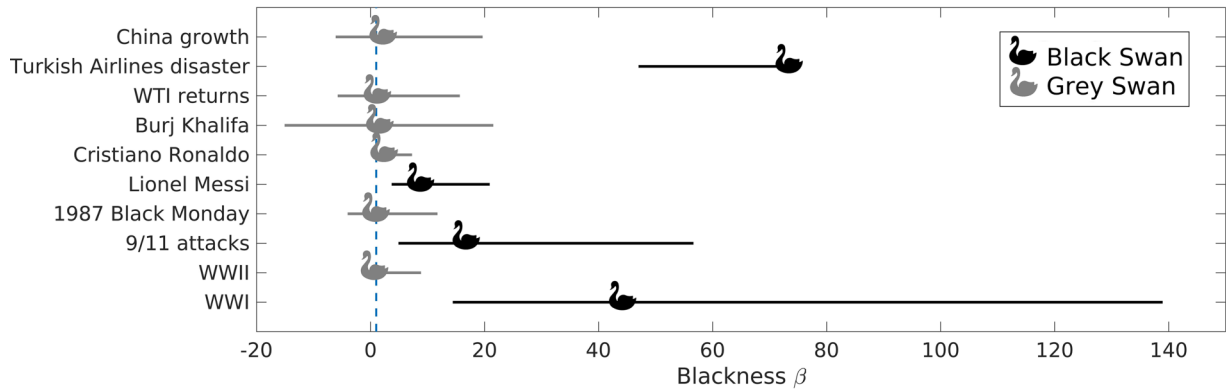
FIG. 3. Blackness of ten events with 90% confidence bounds. Here the threshold for an event to be considered a Black Swan is $\beta > 1$; only four events out of the ten we considered are Black Swans (Turkish Airlines disaster, Lionel Messi, 9/11 terrorist attacks, and World War I), while the remaining six are Gray Swans. These results have been obtained by means of Eq. (6); see Methods for details.

Eq. (6), we can obtain a threshold value $\mathcal{R}_{th}$ for the relative size dividing Black from Gray Swans:

$$\mathcal{R}_{th} = \left(\frac{Q+1}{Q} \right)^{\gamma} = \frac{s_M}{S(1)}, \quad (7)$$

which, once again, can be expressed in terms of only empirical quantities and coincides with the ratio between the estimate of the upper cutoff and the empirical maximum. In this way we can now define the *Swan's plane* $Q - \mathcal{R}^{1/\gamma}$ (Fig. 4), since only these two quantities are needed to determine the nature of an extreme event. Three areas can be identified: the White Swan's region [$\mathcal{R}^{1/\gamma} < 1$, if the event has a size less than the already-seen maximum $S(1)$], the Gray Swan's region [$1 < \mathcal{R} < \frac{Q+1}{Q}$, a size higher than the maximum, but not surprising given the upper cutoff estimation], and the Black Swan's region [$\mathcal{R} > \frac{Q+1}{Q}$, higher than both the maximum and the upper cutoff and so, in this sense, surprising].

Note that the trivial White Swan region is not represented in the figure. As discussed above, stationary power laws are not expected to produce Black Swans, as such events are related to an abrupt nonstationarity of the upper cutoff. This is confirmed by Fig. 4, where we exploited the Swan's plane to visualize the outcomes of a stationary truncated power law. More precisely, we generated $N_i = 100$ numbers from a stationary truncated power law with parameters $\alpha = 1.5$, $s_m = 1$, and $s_M = 10^6$, so that the initial deviation parameter is $Q_i = 0.1$, and then we extracted other $N_f = 10^5$ numbers so as to arrive to a final deviation parameter $Q_f = 100$. At each draw we checked whether the newcomer was larger than the previously observed numbers. If this was the case, we fitted the sample available before the new extraction occurred with Zipf-Mandelbrot law, determining the value of Q; the full procedure was repeated 10^3 times. Brighter colors correspond to a larger frequency, and, as expected, frequencies go to zero out of the Gray Swan region delimited by the white curve
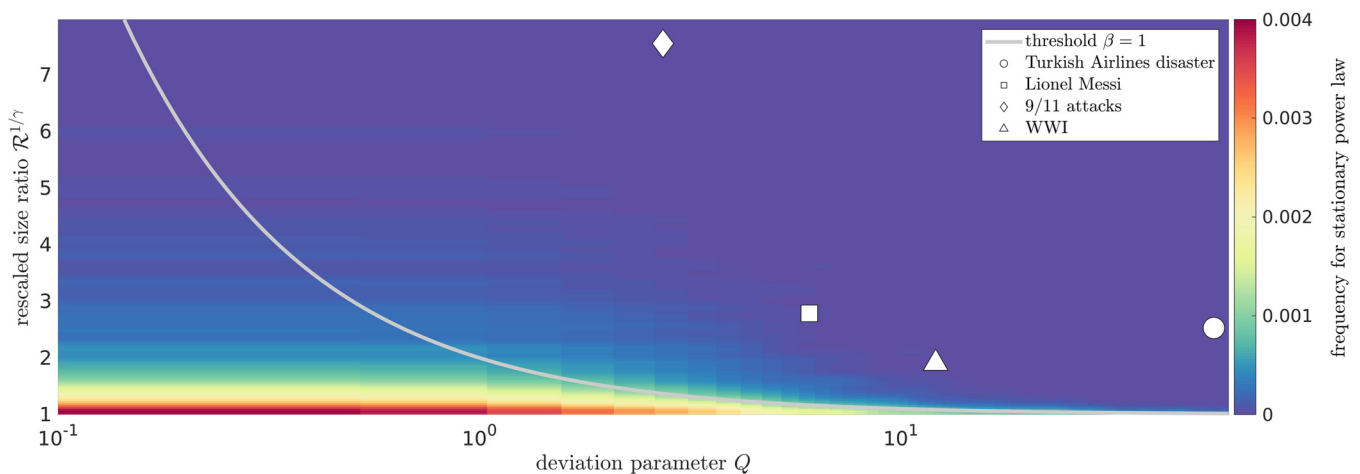

FIG. 4. Visualization in the Swans plane of the outcomes of a truncated stationary power law with $\alpha = 1.5$, $s_m = 1$, and $s_M = 10^6$ (color map) and of the four Black Swans we identified (white dots). Concerning the stationary power law, the figure shows the frequency of events with $\mathcal{R} > 1$, obtained by exploiting a binning procedure in the Swans plane; lighter colors correspond to higher frequencies. The white curve corresponds to the threshold $\beta = 1$ and divides Gray Swans and Black Swans. As expected, the frequency goes to zero above the curve, proving that a stationary power law only produces Gray Swans. Note that as expected, all four Black Swans reside in a region of the Swans Plane where the frequency of power-law events is vanishing, so they cannot be explained by a stationary power law.

TABLE I. Future Black Swans. Upper cutoff s_M and relative size $\mathcal{R}_{th}$ of five events that could be considered Black Swans with a confidence of 90%. Systems highlighted in red are characterized by a low level of sampling, while those highlighted in green present a high level of sampling.

System	$\mathcal{R}_{th}$	s_M
Pandemics	42.29	5.81×10^9 deaths
Alberta wildfires	1.54	1.55×10^6 ha
Interstate wars	∞	∞
Dow Jones index return	∞	∞
Italian earthquake	1.23	7.46M

corresponding to $\beta = 1$. In the same figure we also plotted the four Black Swans we identified, as it is possible to see they all reside in a region of the Swan Plane where the stationary power law produces no events and so they are not compatible with a stationary power law.

2. Future Black Swans

By using our estimation of the upper cutoff, given by Eq. (4), we can estimate the size of a newcomer event to be a Black Swan. We computed the upper cutoff of five social and natural systems and, in turn, the minimum size for new events to be Black Swans. We show our results in Table I. It shows that a pandemic should kill more than 5 billion people, i.e., approximately 40 times the Black Death, the worst pandemic ever with its 200 millions deaths, to be considered a Black Swan. This stems from the fact that only a few pandemics have been recorded during human history, and so the inherent distribution is not much sampled. Conversely, the distributions of wildfires which occur in Alberta (Canada) and Italian earthquakes are characterized by a high level of sampling. Indeed, a wildfire should be just 1.54 times larger than the observed maximum[1] for being classified as a Black Swan, while an earthquake should release only 1.22 times the energy of the worst Italian earthquake ever.[2] Finally, the distribution of Interstate Wars and Dow Jones index daily returns are completely undersampled, and this is reflected in the fact that in these systems any event, no matter how large, can be at most a Gray Swan. More details about these databases are reported in the Methods section.

III. DISCUSSION

In this work we exploit the connections among the Zipf-Mandelbrot law, the inherent power law, and the level of sampling, discussed in [17], to obtain a relation to derive an empirical estimation of the upper cutoff of a power-law distribution. We define Black Swans as events whose size is both extreme (i.e., larger than the observed maximum) and unexpected (i.e., larger than the estimated cutoff). We show that three ingredients are needed to produce Black Swans: an inherent truncated power-law distribution, a high level of

---

[1]August 1981 wildfire, 10^6 ha burned.

[2]1693 Val di Noto earthquake, 7.32 Mw.

TABLE II. Number of elements considered and p value for the systems we analyzed. All the p values are above the threshold of 0.1, meaning that all these systems are well described by a power-law distribution.

System	Number of elements N	p value
Wars before WWI	38	0.95
Wars before WWII	55	0.30
Terrorist attacks	150	0.48
Dow Jones returns before 1987	250	0.24
LaLiga top scorers	150	0.59
Tallest buildings	500	0.52
WTI monthly returns	75	0.66
Airplanes disasters	859	0.68
Countries growth	70	0.18
Pandemics	35	0.89
Alberta wildfires	250	0.39
Wars	80	0.82
Dow Jones returns	250	0.67
Italian earthquakes	500	0.19

sampling, and a jump dynamics of the upper cutoff. Turkish Airlines Flight 981 is a clear-cut example where such a dynamics is particularly evident, since the jump of the upper cutoff has been provoked by a well-defined technological advance—the introduction of large-body aircrafts. As a consequence, looking only at data regarding a certain system may result in ignoring the jump of the upper cutoff, not allowing one to foresee the possibility of a Black Swan. When performing risk assessment it is instead crucial to also analyze the context the systems live in, since jumps can be provoked by eternal factors rather than by the laws governing the system itself, and not to rely on the assumption of a stationary probability distribution. It is also important to stress that in our definition of Black Swan, the level of sampling plays a crucial role, since in order to understand if an event is an outlier and not compatible with the upper cutof,f we need an estimate of such a cutoff. This implies that nontruncated power-law distribution, for which the level of sampling is always null, cannot produce Black Swans. We also introduced a quantitative and scientific measure of very large events, the Blackness, which allows them to be divided into White, Gray, and Black Swans. Using this parameter, which depends on empirical quantities only, we checked that stationary power laws only produce Gray Swans, and we analyzed several empirical systems in search of Black Swans. We found out that our criterion is in line with most of the qualitative findings of Taleb, since it correctly classifies World War I and the 9/11 terrorist attacks as Black Swans, while the 1987 Black Monday is classified only as a Gray Swan. We also spotted new examples of Black Swans, such as the Turkish Airlines disaster and the number of goals scored by Lionel Messi. The latter is probably connected to an increase of the number of teams involved in the LaLiga Championship. Finally, we determined how large various events shouldbe in order to be classified as Black Swans by estimating the upper cutoff of their inherent power-law distributions. For instance, in the case of Italian earthquakes the distribution is highly sampled, and so an earthquake releasing just 1.23 times the energy of

the largest earthquake ever recorded would be a Black Swan. Conversely, the distribution of Dow Jones index daily returns is undersampled, and so any fluctuation, no matter how large, would be only an unsurprising Gray Swan. Note that in this work we considered only examples of temporal phenomena, but in principle one could apply all these techniques also to nontemporal systems where the level of sampling varies due to other mechanisms. As an example, in a previous study [16] we considered clusters of galaxies, which show a high level of sampling and for which we can compute the upper cutoff. Such a system is clearly stationary on human timescales, but new data can become available thanks to the creation of new and more powerful telescopes. A Black Swan would occur if we observed clusters much larger than the upper cutoff we estimated and would imply that the physics governing such structures is different in different regions of the universe.

We believe that the introduction of a quantitative criterion to classify extreme events in Black Swans (or not) can be extremely useful not only from a scientific point of view, but also to scientifically ground the discussion among public opinion, academia, and policy makers.

IV. METHODS

A. Analytical results

As done in [17], let us consider a truncated power-law distribution of sizes, $P(S)$, that is

$$P(S) = \begin{cases} 0 \text{ for } S < s_m \\ \frac{c}{S^\alpha} \text{ for } s_m \leqslant S \leqslant s_M \\ 0 \text{ for } S > s_M \end{cases}, \tag{8}$$

where c is the normalization constant, and s_m and s_M respectively correspond to the natural lower and upper cutoffs, always present in real systems. These cutoffs are connected to c by the normalization condition

$$c \int_{s_m}^{s_M} \frac{ds}{s^\alpha} = 1 \ \rightarrow \ c = \frac{\alpha - 1}{s_m^{1-\alpha} - s_M^{1-\alpha}}. \tag{9}$$

We can then express the rank-size relation as a function of the probability density function (PDF) parameters using the fact that given the PDF $P(S)$ of a continuous variable S, the values of its cumulative distribution function (CDF) $C(S)$, associated to the different values of S, are approximately equiprobable. In fact, if $P(s)$ is the PDF of the variable S defined in the interval $[s_m, s_M]$, then $C(S) = \int_{s_m}^{S} ds' P(s')$. By performing the change of variables from S to $C = C(S)$, and calling $f(C)$ its PDF, we get, by definition of PDF and CDF, $f(C) = \frac{dS(C)}{dC} P(S)|_{S=S(C)} = 1$ for $0 \leqslant C \leqslant 1$. This implies that given N values of S independently extracted from $P(S)$, with good approximation they can be taken as uniformly spaced in the corresponding variable C. Thus the kth size ranked value $S(k)$ approximately corresponds to the CDF value $\frac{N+1-k}{N+1}$. In formulas

$$\int_{s_m}^{S(k)} P(S)dS = c \int_{s_m}^{S(k)} \frac{ds}{s^\alpha} \simeq \frac{N+1-k}{N+1},$$

which, together with Eq. (9), gives

$$\frac{S(k)^{1-\alpha} - s_m^{1-\alpha}}{s_M^{1-\alpha} - s_m^{1-\alpha}} \simeq \frac{N+1-k}{N+1}.$$

By assuming $N + 1 \approx N$, $s_M \gg s_m$, and introducing $\gamma = \frac{1}{\alpha-1}$, we end up with the final rank-size formula:

$$S(k) = \left[\frac{N s_m^{\frac{1}{\gamma}} s_M^{\frac{1}{\gamma}}}{N s_m^{\frac{1}{\gamma}} + k s_M^{\frac{1}{\gamma}}} \right]^\gamma = \frac{N^\gamma s_m}{\left[k + N \left(\frac{s_m}{s_M} \right)^{\frac{1}{\gamma}} \right]^\gamma}.$$

By comparing this expression with Zipf-Mandelbrot law, that is,

$$S(k) = \frac{\bar{S}}{(k+Q)^\gamma}, \tag{10}$$

we obtain

$$\begin{cases} \gamma = \frac{1}{\alpha-1} \\ \bar{S} = N^\gamma s_m \\ Q = N(*) \frac{s_m}{s_M}^{\alpha-1} \end{cases} . \tag{11}$$

These expressions relate the number of values or objects and the parameters of the PDF $P(S)$ on one side and the Zipf-Mandelbrot parameters on the other. Note that Q not only quantifies deviations from Zipf's law but also quantifies the level of sampling of the inherent distribution. Indeed, Q is

1. the larger the wider is the statistical sample, so the larger is the numerosity of the sample N;

2. the smaller the wider is the extension of the truncated power law, given by the ratio between the upper cutoff and the lower one.

It is possible to derive an expression connecting the upper cutoff of the distribution to the deviation parameter Q, Zipf's exponent γ, and the empirical maximum $S(1)$. Combining Eqs. (10) and (11) we obtain

$$S(1) = \frac{N^\gamma s_m}{(Q+1)^\gamma} = s_M \frac{N^\gamma \frac{s_m}{s_M}}{(Q+1)^\gamma} = s_M \left(\frac{Q}{Q+1} \right)^\gamma,$$

which yields

$$s_M = S(1) \left(\frac{Q+1}{Q} \right)^\gamma. \tag{12}$$

As expected for $Q \to \infty$, the empirical maximum $S(1)$ coincides with the upper cutoff and consequently, any event larger than $S(1)$ is a surprise, while for $Q \to 0$ no event can be surprising since the upper cutoff is diverging. Again, we see that Q plays the role of level of sampling quantifier, since the truncation point (the upper cutoff) can be appreciated only if Q is sufficiently large. Exploiting Eq. (12), we can rewrite the Blackness $\beta = (S_{\text{new}} - S(1))/(s_M - S(1))$ of an event with size S_{new} as

$$\beta = \beta(\mathcal{R}, Q, \gamma) = \frac{S_{\text{new}} - S(1)}{S(1)\left[\left(\frac{Q+1}{Q} \right)^\gamma - 1 \right]}$$

$$= \frac{\mathcal{R} - 1}{\left(\frac{Q+1}{Q} \right)^\gamma - 1}, \tag{13}$$

where $\mathcal{R} = S_{\text{new}}/S(1)$ is the ratio between the size of the new event and the empirical maximum. Since the new event is a Black Swan if its size is larger than the estimate of the upper cutoff, the threshold $\mathcal{R}_{th}$ dividing Gray and Black Swans is

$$\mathcal{R}_{th} = \frac{S_{\text{new}}}{s_M} = \left(\frac{Q+1}{Q} \right)^\gamma.$$

In these terms we can restate the criterion introduced in the main text terms of three empirical quantities: the deviation parameter Q, Zipf's exponent γ, and the ratio $\mathcal{R} = S(1)/s_M$. More precisely,

$$\begin{cases} \text{White Swan} \iff \mathcal{R} \leqslant 1 \\ \text{Gray Swan} \iff 1 < \mathcal{R} \leqslant \mathcal{R}_{th} \\ \text{Black Swan} \iff \mathcal{R} > \mathcal{R}_{th} \end{cases}, \quad (14)$$

Note that the threshold ratio $\mathcal{R}_{th} = \mathcal{R}_{th}(Q, \gamma)$ (i) diverges if Q goes to zero, (ii) tends to 1 for Q going to infinity, and (iii) is increasing with $\gamma = 1/(\alpha - 1)$, so it is the larger the fatter is the inherent distribution.

B. Fitting procedure and confidence bounds estimation

We adopted a standard nonlinear least-squares (NLS) fitting procedure to determine the parameters of Zipf-Mandelbrot law. The accuracy of this technique, when applied to the rank-size plot or to the complementary cumulative distribution, is comparable to maximum likelihood estimates [28] while being much simpler (in the case of an unknown upper cutoff). In particular, we used Eq. (2) partially linearized through logarithms,

$$\log S(k) = -\gamma \ln(*)k + Q + c,$$

where Q, γ, and c are free parameters. All the systems we considered have already been widely studied, and there is strong evidence regarding the presence of an underlying power-law distribution; nevertheless, we exploited the p value to check the goodness of our fits. We followed the procedure described in [21]:

(1) We compute the parameter Q and γ of the empirical data with the NLS.

(2) We use Eq. (3) to determine the parameters of the underlying power-law distribution. Note that N and s_m are given, respectively, by the number of elements in the sample and by the size of the smallest object.

(3) We compute the Kolmogorov-Smirnov distance between the empirical data and the power law.

(4) We generate $M = 1000$ Monte Carlo samples (each with the same numerosity N of the empirical sample) from the inferred power-law distribution and, for each of them, we repeat steps 1–3 to determine the statistics of the Kolmogorov-Smirnov distance.

(5) The p value is defined as the fraction of the Monte Carlo samples whose Kolmogorov-Smirnov distance is larger than that of the empirical data.

Following [21], the power-law hypothesis is rejected if the p value is smaller than 0.1.

All the p values for the systems we analyzed are reported in Table II, as it is possible to see all values are above the threshold 0.1, meaning that, as previously noticed, all the systems considered are well described by an underlying power-law distribution. Note, however, that in some systems, as in the case of pandemics, only a limited number of elements is available, and so the parameters estimated by the fitting procedure may not be very precise. In order to take into account this fact, as explained below, we perform a parametric bootstrap, which allows us to include in the

uncertainty over the parameters also the effect of the low numerosity.

Once the parameters Q and γ have been obtained, the Blackness β can be computed by exploiting Eq. (6); however, determining the uncertainty of such a quantity is not trivial. Naively one could propagate on β the uncertainty on Q, and γ, σ_Q, and σ_γ returned by the NLS fitting method. However, this method does not take into account statistical fluctuations that are encountered considering different samples generated by the same power-law distribution. For this reason we exploited a parametric bootstrap so to obtain a more realistic measure of uncertainty. In particular, the procedure we adopted is the following:

1. We compute the parameter Q and γ of the empirical data with the NLS, and we use them to obtain the Blackness β.

2. We use Eq. (3) to determine the parameters of the underlying power-law distribution.

3. We generate $M = 1000$ Monte Carlo samples with numerosity N as the empirical sample using the power-law distribution obtained in the previous step.

4. Each synthetic sample m is fitted with the NLS technique in order to obtain the parameters Q^m and γ^m and their standard deviation σ_{Q^m} and σ_{γ^m}. These quantities are used to determine the Blackness of the event under analysis with respect to the synthetic sample, β^m, whose uncertainty is obtained by propagating σ_{Q^m} and σ_{γ^m}:

$$\sigma_{\beta^m} = \sqrt{\left(\frac{d\beta}{dQ}\sigma_{Q^m}\right)^2 + \left(\frac{d\beta}{d\gamma}\sigma_{\gamma^m}\right)^2}.$$

5. The distribution of β, $P(\beta)$, is obtained as a mixture of M Gaussians with parameters β^m and σ_{β^m},

$$P(\beta) = \frac{1}{M}\sum_{m=1}^{M}\mathcal{N}(*)\beta^m, \sigma_{\beta^m},$$

where $\mathcal{N}(*)x, y$ is a Gaussian with mean x and variance y^2.

6. Starting from the probability distribution $P(\beta)$, the confidence bound for β is easily obtained using the cumulative distribution and determining the interval containing 90% of the probability.

C. Databases

All the databases we used in our analysis are freely accessible on the web, details can be found below:

GDPppp of countries. In order to study the growth of countries so to determine if the growth of China has been a Black Swan, we used the Maddison database [29], which provides GDP PPP of countries from 1 AD to 2008. We integrated it with IMF data [30] to obtain a database which ranges from 1900 to 2020.

Airplane disasters. The analysis of airplane disasters has been performed exploiting a dataset from Kaggle.com con-

taining data of airplane accidents involving civil, commercial, and military transport worldwide from 1908-09-17 to 2009-06-08 [31].

WTI returns. The monthly returns of West Texas Intermediate Oil can be accessed from different sources; we used a dataset ranging from 1986 to 2020 [32].

Tallest buildings. The list of the tallest buildings before Burj Khalifa, so before 2010, was generated from [33].

LaLiga top scorers. The list of LaLiga top scorers was retrieved from [34]. We considered only seasons between 1928/1929 and 2004/2005, so before the blowup of Messi and Ronaldo.

Down Jones index returns. Historical daily returns of Dow Jones index from 1986 to 2016 have been downloaded from Quandl.com [35]. Data from 2016 to present days have been retrieved from [36].

Terrorist attacks. The analysis of casualties provoked by terrorist attacks has been performed using the RAND Database of Worldwide Terrorism Incidents (RDWTI) [37].

Interstate wars. Data on Inter-State Wars have been taken from The Correlates of War (COW) Project [38].

Pandemics. Regarding pandemics, we exploited the average estimate of the number of casualties described in [23].

Wildfires. The analysis of wildfires is based on Alberta Wildfire datasets [39]. More precisely, we merged the four databases, obtaining a dataset spanning spanning the period 1961–2018.

Italian earthquakes. For our analysis of the Italian earthquakes we used the INGV Parametric Catalogue of Italian Earthquakes, which "provides homogeneous macroseismic and instrumental data and parameters for Italian earthquakes with maximum intensity $\geqslant 5$ or magnitude $\geqslant 4.0$ in the period 1000-2017" [40].

[1] P. W. Anderson, More is different, Science **177**, 393 (1972).

[2] M. Gell-Mann, *The Quark and the Jaguar: Adventures in the Simple and the Complex* (Macmillan, New York, 1995).

[3] L. Pietronero, Complexity ideas from condensed matter and statistical physics, Europhys. News **39**, 26 (2008).

[4] C. Castellano, S. Fortunato, and V. Loreto, Statistical physics of social dynamics, Rev. Mod. Phys. **81**, 591 (2009).

[5] A. Tacchella, D. Mazzilli, and L. Pietronero, A dynamical systems approach to gross domestic product forecasting, Nat. Phys. **14**, 861 (2018).

[6] M. E. J. Newman, Power laws, Pareto distributions and Zipf's law, Contemp. Phys. **46**, 323 (2005).

[7] B. B. Mandelbrot, Scaling in financial prices: I. Tails and dependence, Quant. Finance **1**, 113 (2001).

[8] N. N. Taleb, *The Black Swan: The Impact of the Highly Improbable* (Random House, New York City, 2007), Vol. 2.

[9] A. M. Nafday, Strategies for managing the consequences of black swan events, Leadersh. Manage. Eng. **9**, 191 (2009).

[10] S. Hajikazemi, A. Ekambaram, B. Andersen, and Y. J. T. Zidane, The black swan–Knowing the unknown in projects, Procedia Social and Behavioral Sciences **226**, 184 (2016).

[11] N. Ferguson, Black swans, dragon kings and gray rhinos: The world war of 1914-1918 and the pandemic of 2020-? *History Working Paper 2020-1* (Stanford University, Stanford, CA, 2020).

[12] G. K. Zipf, *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology* (Ravenio Books, 2016).

[13] B. Blasius and R. Tönjes, Zipf's Law in the Popularity Distribution of Chess Openings, Phys. Rev. Lett. **103**, 218701 (2009).

[14] C. Furusawa and K. Kaneko, Zipf's Law in Gene Expression, Phys. Rev. Lett. **90**, 088102 (2003).

[15] T. Maillart, D. Sornette, S. Spaeth, and G. von Krogh, Empirical Tests of Zipf's Law Mechanism in Open Source Linux Distribution, Phys. Rev. Lett. **101**, 218701 (2008).

[16] De Marzo, Giordano, Sylos Labini, Francesco, and Pietronero, Luciano, Zipf's law for cosmic structures: How large are the greatest structures in the universe? Astron. Astrophys. **651**, A114 (2021).

[17] G. De Marzo, A. Gabrielli, A. Zaccaria, and L. Pietronero, Dynamical approach to Zipf's law, Phys. Rev. Research **3**, 013084 (2021).

[18] M. Cristelli, M. Batty, and L. Pietronero, There is more than a power law in Zipf, Sci. Rep. **2**, 812 (2012).

[19] S. M. Burroughs and S. F. Tebbens, Upper-truncated power laws in natural systems, Pure Appl. Geophys. **158**, 741 (2001).

[20] B. B. Mandelbrot, An informational theory of the statistical structure of language, in *Communication Theory*, edited by W. Jackson (Academic Press, Princeton, 1953), pp. 486–502.

[21] A. Clauset, C. R. Shalizi, and M. E. J. Newman, Power-law distributions in empirical data, SIAM Rev. **51**, 661 (2009).

[22] L. C. Malacarne and R. S. Mendes, Regularities in football goal distributions, Physica A **286**, 391 (2000).

[23] P. Cirillo and N. N. Taleb, Tail risk of contagious diseases, Nat. Phys. **16**, 606 (2020).

[24] V. A. Golyk, Self-organized criticality, *Phase Transitions* (World Scientific, 2004), pp. 113–128.

[25] A. Chatterjee and B. K. Chakrabarti, Fat tailed distributions for deaths in conflicts and disasters, Rep. Adv. Phys. Sci. **01**, 1740007 (2017).

[26] L. Pritchett and L. H. Summers, Asiaphoria Meets Regression to the Mean, Tech. Rep., National Bureau of Economic Research, 2014.

[27] M. Cristelli, A. Tacchella, and L. Pietronero, The heterogeneous dynamics of economic complexity, PLoS One **10**, e0117174 (2015).

[28] E. P. White, B. J. Enquist, and J. L. Green, On estimating the exponent of power-law frequency distributions, Ecology **89**, 905 (2008).

[29] A. Maddison, The maddison-project, Vol. 1 (2013), p. 14, http://www.ggdc.net/maddison/maddisonproject/home.htm.

[30] IMF World Economic Outlook, https://www.imf.org/en/Publications/WEO/weo-database/2022/April.

[31] Airplane Crashes Since 1908, https://www.kaggle.com/datasets/saurograndi/airplane-crashes-since-1908.

[32] Brent and WTI Spot Prices, https://datahub.io/core/oil-prices.

[33] skyscrapercenter.com.

[34] transfermarkt.com.

[35] Dow Jones Industrial Average historical data, https://data.nasdaq.com/data/BCB/UDJIAD1-dow-jones-industrial-average.

[36] finance.yahoo.com.

[37] The RAND Database of Worldwide Terrorism Incidents (RDWTI), https://www.rand.org/nsrd/projects/terrorism-incidents/download.html.

[38] M. R. Sarkees and F. W. Wayman, *Resort to War: A Data Guide to Inter-State, Extra-State, Intra-State, and Non-State Wars, 1816-2007* (Cq Pr, 2010).

[39] Historical Wildfire Database. Alberta wildfire records, https://wildfire.alberta.ca/resources/historical-data/historical-wildfire-database.aspx.

[40] A. N. Rovida, M. Locati, R. D. Camassi, B. Lolli, and P. Gasperini, CPTI15, The 2015 version of the Parametric Catalogue of Italian Earthquakes (2016).