Tail Risk of Contagious Diseases

Pasquale Cirillo* and Nassim Nicholas Taleb†
*Applied Probability Group, Delft University of Technology
†Tandon School of Engineering, New York University (Corresponding author)

***Abstract*—Using methods from extreme value theory, we examine the distribution of fatalities from major pandemics in history, and build a statistical picture of their tail properties.**

Applying the dual distribution approach developed by the authors for violent conflicts and operational risk for banks, we provide rough estimates for quantities not immediately observable in the data. We also stress the data to gauge the effect of the reliability of sources.

Epidemics and pandemics are extremely fat-tailed, with a potential existential risk for humanity. This property should markedly increase vigilance and override conclusions derived from local epidemiological models in what relates to tail events.

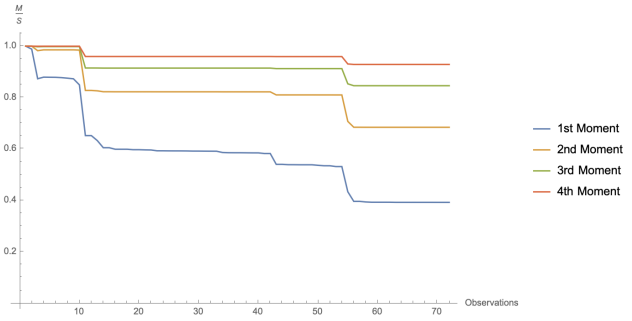

Fig. 1: Maximum to Sum plot (MS plot) of the average death numbers in pandemic events in history, as per Table 1.

I. Introduction

In the wake of the current Covid-19 pandemic, epidemiological models such as SIR [9] are presented and discussed, with no attention for model risk or parameters' uncertainty. Logistic curves and lines of all sort are fitted and interpolated, even with very few and unreliable data points, often ignoring basic problems like under-reporting or censoring. Even more, they are continuously referred to by politicians as a justification for mild policies and to criticize people's "paranoïa"–even if, as we shall discover, such a paranoïa can be sound risk management [14] in front of a certain class of potentially destructive events–as if pandemics were thin-tailed phenomena, with irrelevant tail risk. It should be evident that one cannot compare death from multiplicative infectious diseases (fat-tailed, like a Pareto) to car accidents (thin-tailed, like a Gaussian), yet this is a common (and costly) error in policy making and in the journalistic literature.

Using tools from Extreme Value Theory (EVT) we show that the distribution of the victims of infectious diseases is extremely fat-tailed[1], more than what one could be led to believe from the outset. For the best of our knowledge, only war casualties and operational risk losses for banks show a comparable behavior [3], [4], and they are both phenomena very difficult to model.

Figure 1 gives an idea of the problem, presenting the so-called Maximum-to-Sum plot [6] of the number of pandemic fatalities in history (data in Table 1). Such a plot relies on a simple consequence of the law of large numbers: for a sequence $X_1, X_2, ..., X_n$ of nonnegative i.i.d. random variables, if $E[X^p] < \infty$ for $p = 1, 2, 3...$, then $R_n^p = M_n^p / S_n^p \to^{a.s.} 0$ as $n \to \infty$, where $S_n^p = \sum_{i=1}^n X_i^p$ is the partial sum of order p, and $M_n^p = \max(X_1^p, ..., X_n^p)$ the corresponding partial maximum. Figure 1 clearly shows that no finite moment is likely to exist for the number of victims in pandemics, as the R_n ratio does not converge to 0 for $p = 1, 2, 3, 4$, no matter how many data points we use. Such a behavior hints that the victims distribution has such a fat right tail that not even the first theoretical moment is finite. We are looking at a phenomenon for which quantities like the naive sample average or the standard deviation are therefore meaningless for inference.

However, Figure 1 (or a naive use of extreme value statistics) does not imply that pandemic risk is infinite and there is nothing we can do or model. Using the methodology we developed to study war casualties [3], [15], we are in fact able to extract useful information from the data, quantifying the large yet finite risk of pandemic diseases.

We note that the more "fat-tailed" the distribution, the more "the tail wags the dog", that is, the more statistical information resides in the extemes and the less information there is in the "bulk" (that is events of high frequency), where it becomes almost noise, making EVT the most effective approach and our sample of extremes very highly sufficient and informative for risk management purposes. More technically, because the law of large numbers works slowly for such a class of distribution, the body becomes increasingly dominated by noise and averages and higher moments, even when they exist, become uninformative, though extremes are rich in information [14].

On the basis of our findings and the above property of the tails, the message is clear: focus on the tails of pandemics and avoid toy models like SIR for policy making. Similar tools can only be useful ex-post, when the pandemics is over, to try to understand how it actually developed, but, given their assumptions and nature [9], they should not be used

March 23, 2020

[1] A non-negative continuous random variable X is fat-tailed if its survival function $S(x) = P(X \geq x)$ decays as a power law $x^{-\frac{1}{\xi}}$, the more we move into the tail, that its for x growing towards the right endpoint of X. For a more technical definition involving regularly varying functions, [5], [6].

for forecasting and risk management. It is the (non-naïve) precautionary principle that should be the leading driver for decisions [11].

II. Data and descriptive statistics

We investigate the distribution of deaths from the major epidemic and pandemic diseases of history, from 429 BC until now. The data are available in Table 1, together with the indication of the sources, and only refer to events with more than 1K estimated victims, for a total of 72 observations. As a consequence, potentially highly risky diseases, like the Middle East Respiratory Syndrome (MERS), do not appear in our collection[2]. All diseases whose end year is 2020 are to be taken as still occurring worldwide, as for the running COVID-19 pandemic.

Three estimates of the reported cumulative death toll have been used: the minimum, the average and the maximum numbers. When the three numbers coincide in Table 1, our sources simply do not provide intervals for the estimates. Since we are well aware of the volatility and possible unreliability of historical data [13], [15], in Section IV we deal with such an issue by perturbing and omitting observations.

Since one could be interested in comparing fatalities with respect to the coeval population, by looking at the relative impact of pandemics, in column *Rescaled* of Table 1 we provide the rescaled version of column *Avg Est*, using the information in column *Population*[3] [8], [10], [16]. For example, the Antonine plague of 165-180 killed an average of 7.5M people, that is to say 3.7% of the coeval world population of 202M people. With respect to today's population, such a number would correspond to about 283M deaths, a terrible hecatomb, killing more people than WW2.

For the sake of space, in the rest of the comment, we restrict our attention to the actual average estimates in Table 1, but all our findings and conclusions hold true for the lower, the upper and the rescaled estimates as well.

Figure 2 shows the histogram of the actual average numbers of deaths in the 72 large contagious events. The distributions appears highly skewed and possibly fat-tailed. The sample numbers are as follows: the sample average is 4.9M, while the median is 60K, compatibly with skewness observable in Figure 2. The 90% quantile is 6.5M and the 99% quantile is 137.5M. The sample standard deviation is 19M.

Using common graphical tools for fat tails [6], in Figure 3 we show the log log plot (also known as Zipf plot) of the empirical survival functions for the average victims over the diverse contagious events. In such a plot possible fat tails can be identified in the presence of a linearly decreasing behavior of the plotted curve. To improve interpretability a naive linear fit is also proposed. Figure 3 suggests the presence of fat tails.

The Zipf plot shows a necessary but not sufficient condition for fat-tails [2]. Therefore, in Figure 4 we complement the

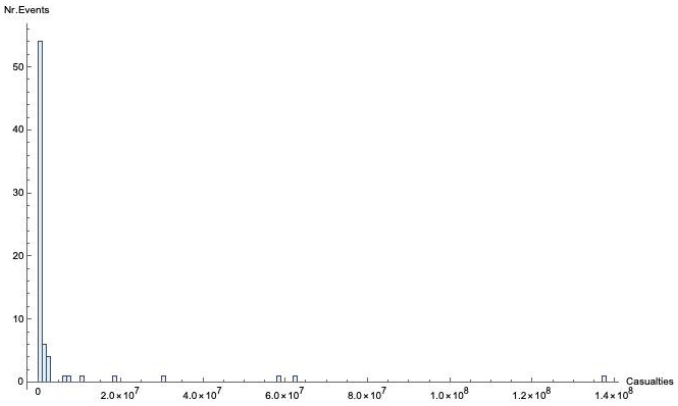

Fig. 2: Histogram of the average number of deaths in the 72 contagious diseases of Table 1.

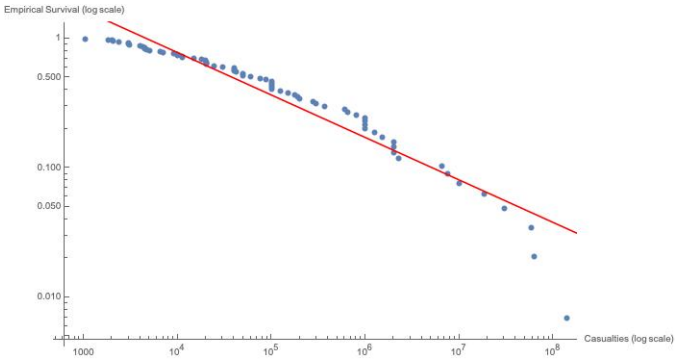

Fig. 3: Log log plot of the empirical survival function (Zipf plot) of the actual average death numbers in Table 1. The red line represents a naive linear fit of the decaying tail.

analysis with a mean excess function plot, or meplot. If a random variable X is possibly fat-tailed, its mean excess function $e_X(u) = E[X - u|X \geq u]$ should grow linearly in the threshold u, at least above a certain value identifying the actual power law tail [6]. In a meplot, where the empirical $e_X(u)$ is plotted against the different values of u, one thus looks for some (more or less) linearly increasing trend, as the one we observe in Figure 4.

Other graphical tools (in Figure 1, we have already shown the Maximum-to-Sum plot) could be used and they would all confirm the point: we are in the presence of fat tails in the distribution of the victims of pandemic diseases. Even more, a distribution with possibly no finite moment.

A. The dual distribution

As we observed for war casualties [3], the non-existence of moments for the distribution of pandemic victims is questionable. Since the distribution of victims is naturally bounded by the coeval world population, no disease can kill more people than those living on the planet at a given time time. We are indeed looking at an *apparently* infinite-mean phenomenon, like in the case of war casualties [3], [15] and operational risk [4].

Let $[L, H]$ be the support of the distribution of pandemic victims today, with $L >> 0$ to ignore small events not

[2]Up to the present, MERS has killed 858 people as reported in https://www.who.int/emergencies/mers-cov/en. For SARS the death toll is at 916 victims until now.

[3]Population estimates are by definitions estimates, and different sources can give different results (most of the times differences are minor), especially for the past. However our methodology is robust to this type of variability, as we stress later in the paper.

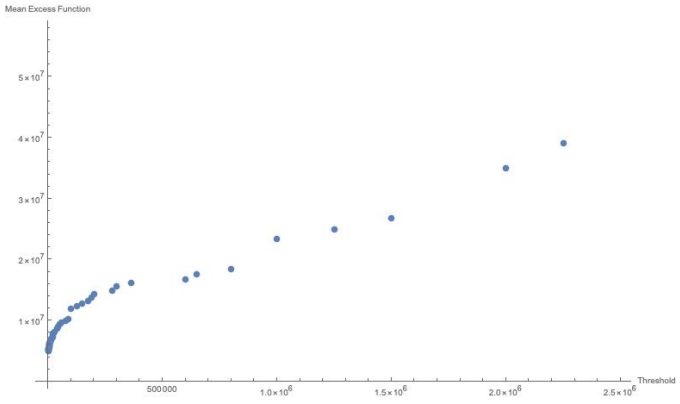

Fig. 4: Mean excess function plot (meplot) of the average death numbers in Table 1. The plot excludes 3 points on the top right corner, consistently with the suggestions in [6] about the exclusion of the more volatile observations.

officially definable as pandemic [19]. For what concerns H, its value cannot be larger than the world population, i.e. 7.7 billion people in 2020[4]. Evidently H is so large that the probability of observing values in its vicinity is in practice zero, and we always end up finding observations below a given $M << H < \infty$ (something like150M deaths for us). Thus one could be fooled by data into ignoring H and taking it as infinite, up to the point of believing in an infinite mean phenomenon, as Figure 1 suggests. However notice that a finite upper bound H is not compatible with infinite moments, hence Figure 1 risks to be dangerously misleading.

In Figure 5, the real tail of the random variable Y with remote upper bound H is represented by the dashed line. However, if one only observes values up to $M << H$, and more or less consciously ignores the existence of H, one could be fooled by the data into believing that the tail is actually the continuous one, the so-called apparent tail [4]. The tails are indeed indistinguishable for most cases, virtually in all finite samples, as the divergence is only clear in the vicinity of H. A bounded tail with very large upper limit is therefore mistakenly taken for an unbounded one, and no model will be able to see the difference, even if epistemologically we are in two extremely different situations. This is the typical case in which critical reasoning, and the a priori analysis of the characteristics of the phenomenon under scrutiny, should precede any instinctive and uncritical fitting of the data.

A solution is the approach of [3], [4], which introduces the concept of dual data via a special log-transformation [5]. The basic idea is to find a way of matching naive extrapolations (apparently infinite moments) with correct modelling.

Let L and H be respectively the finite lower and upper

[4]Today's world population [16] can be safely taken as the upper bound also for the past.

[5]Other log-transformations have been proposed in the literature, but they are all meant to thin the tails, without actually taking care of the upper bound problem: the number of victims can still be infinite. The rationale behind those transformations is given by the observation that if X is a random variable whose distribution funtion is in the domain of attraction of a FrÃl'chet, the family of fat-tailed distributions, then the log X is in the domain of attraction of a Gumbel, the more reassuring family of normals and exponentials [6].

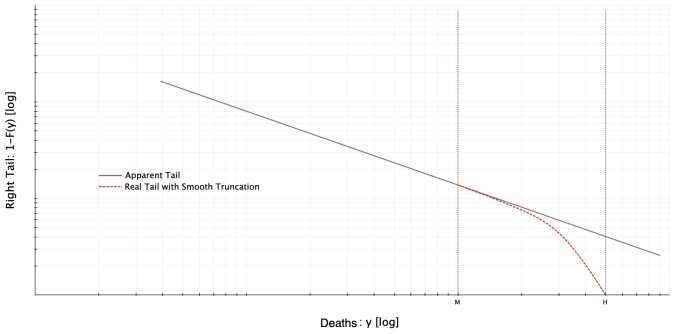

Fig. 5: Graphical representation (log-log plot) of what may happen if one ignores the existence of the finite upper bound H, since only M is observed.

bounds of a random variable Y, and define the function

$$\varphi(Y) = L - H \log\left(\frac{H-Y}{H-L}\right). \tag{1}$$

We can easily check that

1) $\varphi \in C^\infty$,
2) $\varphi^{-1}(\infty) = H$,
3) $\varphi^{-1}(L) = \varphi(L) = L$.

Then $Z = \varphi(Y)$ defines a new random variable with lower bound L and an infinite upper bound. Notice that the transformation induced by $\varphi(\cdot)$ does not depend on any of the parameters of the distribution of Y, and that $\varphi(\cdot)$ is monotone. From now on, we call the distributions of Y and Z, respectively the real and the dual distribution. It is easy to verify that for values smaller than $M << H$, Y and Z are in practice undistinguishable (and do are their quantiles [4]).

As per [3], [4], we take the observations in the column "Avg Est" of Table 1, our Y's, and transform them into their dual Z's. We then study the actually unbounded duals using EVT (see Section 5), to find out that the naive observation of infinite moments can makes sense in such a framework (but not for the bounded world population!). Finally, by reverting to the real distribution, we compute the so-called *shadow* means [4] of pandemics, equal to

$$E[Y] = (H-L)e^{\frac{1}{\xi}\frac{\sigma}{H}}\left(\frac{\sigma}{H\xi}\right)^{\frac{1}{\xi}}\Gamma\left(1-\frac{1}{\xi},\frac{\sigma}{H\xi}\right)+L, \tag{2}$$

where $\Gamma(\cdot,\cdot)$ is the gamma function.

Notice that the random quantity Y is defined above L, therefore its expectation corresponds to a tail expectation with respect to the random variable Z, an expected shortfall in the financial jargon, being only valid in the tail above μ [3]. All moments of the random variable Y are called shadow moments in [4], as they are not immediately visible from the data, but from plug-in estimation.

III. The dual tail via EVT and the shadow mean

Take the dual random variable Z whose distribution function G is unknown, and let $z_G = \sup\{z \in \mathbb{R} : G(z) < 1\}$ be its right-end point, which can be finite or infinite. Given a

threshold $u < z_G$, we can define the exceedance distribution of Z as

$$G_u(z) = P(Z \leq z | Z > u) = \frac{G(z) - G(u)}{1 - G(u)}, \tag{3}$$

for $z \geq u$.

For a large class of distributions G, and high thresholds $u \to z_G$, G_u can be approximated by a Generalized Pareto distribution (GPD) [5], i.e.

$$G_u(z) \approx GPD(z; \xi, \beta, u) = \begin{cases} 1 - (1 + \xi \frac{z-u}{\beta})^{-1/\xi} & \xi \neq 0 \\ 1 - e^{-\frac{z-u}{\beta}} & \xi = 0 \end{cases}, \tag{4}$$

where $z \geq u$ for $\xi \geq 0$, $u \leq z \leq u - \beta/\xi$ for $\xi < 0$, $u \in \mathbb{R}$, $\xi \in \mathbb{R}$ and $\beta > 0$. The shape parameter ξ is a pivotal quantity[6], as it governs the behavior of tails and the existence of moments. The p-th moment of a GPD exists and is finite, if and only if $\xi < 1/p$ [6].

Let us just consider $\xi > 0$, being $\xi = 0$ not relevant for fat tails. From equation (3), we see that $G(z) = (1 - G(u))G_u(z) + G(u)$, hence we obtain

$$\begin{aligned} G(z) &\approx (1 - G(u))GPD(z; \xi, \beta, u) + G(u) \\ &= 1 - \bar{G}(u) \left(1 + \xi \frac{z-u}{\beta}\right)^{-1/\xi}, \end{aligned}$$

with $\bar{G}(x) = 1 - G(x)$. The tail of Z is therefore

$$\bar{G}(z) = \bar{G}(u) \left(1 + \xi \frac{z-u}{\beta}\right)^{-1/\xi}. \tag{5}$$

Equation (5) is known as the tail estimator of $G(z)$ for $z \geq u$. Given that G is in principle unknown, one usually substitutes $G(u)$ with its empirical estimator n_u/n, where n is the total number of observations in the sample, and n_u is the number of exceedances above u.

Equation (5) then changes into

$$\bar{G}(z) = \frac{n_u}{n} \left(1 + \xi \frac{z-u}{\beta}\right)^{-1/\xi}, \tag{6}$$

and this tells us that

$$\bar{G}(z) \approx 1 - GPD(z^*; \xi, \sigma, \mu), \;\; z^* \geq \mu, \tag{7}$$

where $\sigma = \beta \left(\frac{n_u}{n}\right)^{\xi}$, $\mu = u - \frac{\beta}{\xi}\left(1 - \left(\frac{n_u}{n}\right)^{\xi}\right)$ and z^* is an auxiliary random variable. Both σ and β can be estimated semi-parametrically, starting from the estimates of ξ and β in equation (4), and then inputing the empirical n_u/n. If $\xi > -1/2$, the preferred estimation method is maximum likelihood [5], while for $\xi \leq -1/2$ other approaches are better used [6]. For both the exceedances distribution and the recovered tail, the parameter ξ is the same.

One can thus study the tail of Z without caring too much about the rest of the distribution, i.e. the part below u. All in all, the most destructive risks come from the right tail, and not from the first quantiles or even the bulk of the distribution. The identification of the correct u is a relevant question in extreme value statistics [5], [6]. One can rely on heuristic graphical tools [2], like the Zipf plot and the meplot we have seen before, or on statistical tests for extreme value conditions [7] and GPD goodness-of-fit [1].

What is important to stress–once again–is that the GPD fit needs to be performed on the dual quantities, to be statistically and epistemologically correct. One could in fact work with the raw observation directly, without the log-transformation of Equation (1), surely ending up with $\xi > 1$, in line with Figure 1. But a similar approach would be wrong and naive, because only the dual observations are actually unbounded.

Working with the dual observations, we find out that the best GPD fit threshold is around 200K victims, with 34.7% of the observations lying above. For what concerns the GPD parameters, we estimate $\xi = 1.62$ (s.e. 0.52), and $\beta = 1'174.7K$ (s.e. $536.5K$). As expected $\xi > 1$ once again supporting the idea of an infinite first moment. Visual inspections and statistical tests [1], [7] support the goodness-of-fit for the exceedance distribution and the tail.

Given ξ and β, we can use Equations (2) and (7) to compute the shadow means of the numbers of victims in pandemics. For actual data we get a shadow mean of 20.1M, which is definitely larger (almost 1.5 times) than the corresponding sample tail mean of 13.9M (this is the mean of all the actual numbers above the 200K threshold.). Combining the shadow mean with the sample mean below the $200K$ threshold, we get an overall mean of 7M instead of the naive 4.9M we have computed initially. It is therefore important to stress that a naive use of the sample mean would induce an underestimation of risk, and would also be statistically incorrect.

IV. Data reliability issues

As observed in [13], [3], [15] for war casualties, but the same reasoning applies to pandemics of the past, the estimates of the number of victims are not at all unique and precise numbers. Figures are very often anecdotal, based on citations and vague reports, and usually dependent on the source of the estimate. In Table 1, it is evident that some events vary considerably in estimates.

Natural questions thus arise: are the tail risk estimates of Section 5 robust? What happens if some of the casualties estimates change? What is the impact of ignoring some events in our collection? The use of extreme value statistics in studying tail risk already guarantees the robustness of our estimates to changes in the underlying data, when these lie below the threshold u. However, to verify robustness more rigorously and thoroughly, we have decided to stress the data, to study how the tails potentially vary.

First of all, we have generated 10K distorted copies of our dual data. Each copy contains exactly the same number of observations as per Table 1, but every data point has been allowed to vary between 80% and 120% of its recorded value before imposing the log-transformation of Equation (1). In other words, each of the 10K new samples contains 72 observations, and each observation is a (dual) perturbation ($\pm 20\%$) of the corresponding observation in Table 1.

Figure 6 contains the histogram of the ξ parameter over the 10K distorted copies of the dual numbers. The values are

[6]The parameter ξ is also linked to the α parameter commonly found in other power law representations of fat tails [6], given that $\alpha = 1/\xi$.

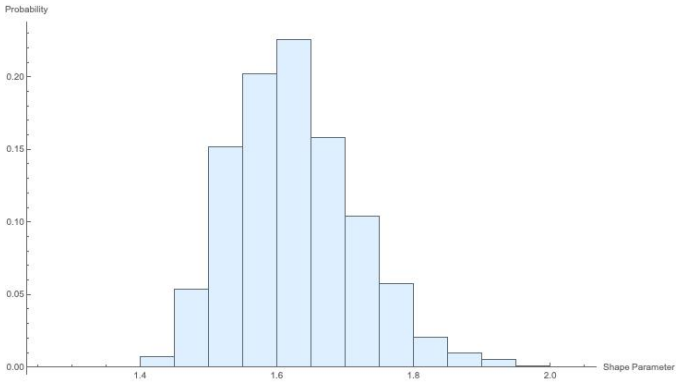

Fig. 6: Values of the shape parameter ξ over 10,000 distorted copies of the the dual versions of the average deaths in Table 1, allowing for a random variation of $\pm 20\%$ for each single observation. The ξ parameter consistently indicates an apparently infinite-mean phenomenon.

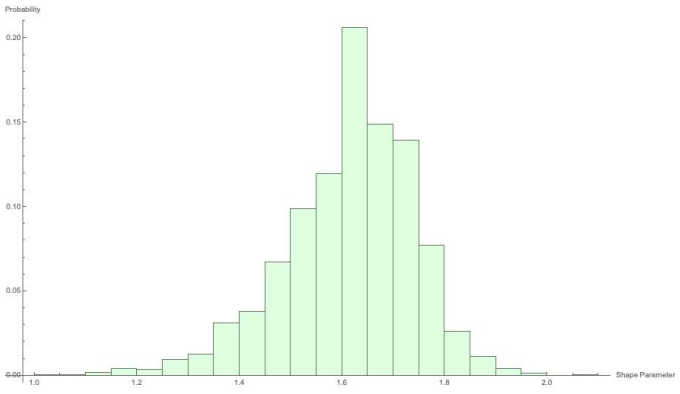

Fig. 7: Values of the shape parameter ξ over 10,000 jackknifed versions of the dual versions of the actual average numbers in Table 1, when allowing at least 1% and up to 10% of the observations to be missing. The ξ parameter consistently indicates an apparently infinite-mean phenomenon.

always above 1, indicating an apparently infinite mean, and the average value is 1.62 (standard deviation 0.10), in line with our previous findings. Our tail estimates are thus robust to imprecise observations. Consistent results hold for the β parameter.

But it also true that our data set is likely to be incomplete, not containing all epidemics and pandemics with more than 1K victims, or that some of the events we have collected are too biased to be reliable and should be discarded anyway. To account for this, we have once again generated 10K copies of our sample via jackknife. Each new dual sample is obtained by removing from 1 to 6 observations at random, so that one sample could not contain the Spanish flu, while another could ignore the Yellow Fever and AIDS. In Figure 7 we show the impact of such a procedure on the ξ parameter. Once again, the main message of this work remains unchanged.

REFERENCES

[1] M. Arshad , M.T. Rasool, M.I. Ahmad (2003). Anderson Darling and Modified Anderson Darling Tests for Generalized Pareto Distribution. Journal of Applied Sciences 3, 85-88.

[2] P. Cirillo (2013). Are your data really Pareto distributed? *Physica A: Statistical Mechanics and its Applications* 392, 5947-5962.

[3] P. Cirillo, N.N. Taleb (2016). On the statistical properties and tail risk of violent conflicts. Physica A: Statistical Mechanics and its Applications 452, 29-45.

[4] P. Cirillo, N.N. Taleb (2016). Expected shortfall estimation for apparently infinite-mean models of operational risk. Quantitative Finance 16, 1485-1494.

[5] L. de Haan, A. Ferreira (2006). *Extreme Value Theory: An Introduction*. Springer.

[6] P. Embrechts, C. Klüppelberg, T. Mikosch (2003). *Modelling Extremal Events*. Springer.

[7] M. Falk, J. Hüsler J, R. D. Reiss R-D (2004). *Laws of small numbers: extremes and rare events*, Birkhäuser.

[8] K. Goldewijk, K. Beusen, P. Janssen (2010). Long term dynamic modeling of global population and built-up area in a spatially explicit way, hyde 3.1. *The Holocene* 20, 565-573.

[9] HW Hethcote (2000). The mathematics of infectious diseases. *SIAM review* 42, 599-653.

[10] K. Klein Goldewijk, G. van Drecht (2006). HYDE 3.1: Current and historical population and land cover. In A. F. Bouwman, T. Kram, K. Klein Goldewijk. *Integrated modelling of global environmental change. An overview of IMAGE 2.4*. Netherlands Environmental Assessment Agency.

[11] J.Norman, Y. Bar-Yam, N.N. Taleb (2020). Systemic Risk of Pandemic via Novel Pathogens - Coronavirus: A Note, New England Complex Systems Institute.

[12] S. Scasciamacchia, L. Serrecchia, L. Giangrossi, G. Garofolo, A. Balestrucci, G. Sammartino (2012). Plague Epidemic in the Kingdom of Naples, 1656âĂŞ1658. Emerging Infectious Diseases 18, 186-188.

[13] T. B. Seybolt, J. D. Aronson, B. Fischhoff, eds. (2013). *Counting Civilian Casualties, An Introduction to Recording and Estimating Nonmilitary Deaths in Conflict*. Oxford University Press.

[14] N.N. Taleb (2020). *Statistical Consequences of Fat Tails*. STEM Academic Press.

[15] N.N. Taleb, P. Cirillo (2019). The Decline of Violent Conflict: What do the data really say? In A. Toje, N.V.S. Bård, eds. The Causes of Peace: What We Know Now. Nobel Symposium Proceedings. Norwegian Nobel Institute, 57-85.

[16] United Nations - Department of Economic and Social Affairs (2015). *2015 Revision of World Population Prospects*. UN Press.

[17] Ancient History Encyclopedia, retrieved on March 30, 2020: https://www.ancient.eu/article/1528/plague-in-the-ancient--medieval-world/.

[18] Visual Capitalist, Visualizing the History of Pandemics, retrieved on March 30, 2020: https://www.visualcapitalist.com/history-of-pandemics-deadliest/.

[19] World Health Organization, Covid-19 page, retrieved on March 31, 2020: https://www.who.int/emergencies/diseases/novel-coronavirus-2019

[20] Wikipedia List of epidemics, retrieved on March 30, 2020: https://en.wikipedia.org/wiki/List_of_epidemics.

[21] Wikipedia Great Northern War plague outbreak, retrieved on March 30, 2020: https://en.wikipedia.org/wiki/Great_Northern_War_plague_outbreak.

[22] Weblist of Epidemics Compared to Coronavirus, retrieved on March 30, 2020: https://listfist.com/list-of-epidemics-compared-to-coronavirus-covid-19.

Name	Start Year	End Year	Lower Est (1k)	Avg Est (1k)	Upper Est (1k)	Rescaled Avg Est (1k)	Population (1mio)	Source
Plague of Athens	-429	-426	75	88	100	13376	50	[20]
Antonine Plague	165	180	5000	7500	10000	283355	202	[20]
Plague of Cyprian	250	266	1000	1000	1000	37227	205	[20]
Plague of Justinian	541	542	25000	62500	100000	2246550	213	[20]
Plague of Amida	562	562	30	30	30	1078	213	[17]
Roman Plague of 590	590	590	10	20	30	719	213	[20]
Plague of Sheroe	627	628	100	100	100	3594	213	[22]
Plague of the British Isles	664	689	150	175	200	6290	213	[20]
Plague of Basra	688	689	200	200	200	7189	213	[17]
Japanese smallpox epidemic	735	737	2000	2000	2000	67690	226	[20]
Black Death	1331	1353	75000	137500	200000	2678283	392	[20]
Sweating sickness	1485	1551	10	10	10	166	461	[20]
Smallpox Epidemic in Mexico	1520	1520	5000	6500	8000	107684	461	[20]
Cocoliztli Epidemic of 1545–1548	1545	1548	5000	10000	15000	165668	461	[20]
1563 London plague	1562	1564	20	20	20	277	554	[20]
Cocoliztli epidemic of 1576	1576	1580	2000	2250	2500	31045	554	[20]
1592–93 London plague	1592	1593	20	20	20	275	554	[20]
Malta plague epidemic	1592	1593	3	3	3	41	554	[20]
Plague in Spain	1596	1602	600	650	700	8969	554	[20]
New Engliand epidemic	1616	1620	7	7	7	97	554	[20]
Italian plague of 1629–1631	1629	1631	280	280	280	3863	554	[20]
Great Plague of Sevilla	1647	1652	150	150	150	2070	554	[20]
Plague in Kingdom of Naples	1656	1658	1250	1250	1250	15840	603	[12]
Plague in the Netherlands	1663	1664	24	24	24	306	603	[20]
Great Plague of London	1665	1666	100	100	100	1267	603	[20]
Plague in France	1668	1668	40	40	40	507	603	[20]
Malta plague epidemic	1675	1676	11	11	11	143	603	[20]
Great Plague of Vienna	1679	1679	76	76	76	963	603	[20]
Great Northern War plague outbreak	1700	1721	176	192	208	2427	603	[21]
Great Smallpox Epidemic in Iceland	1707	1709	18	18	18	228	603	[20]
Great Plague of Marseille	1720	1722	100	100	100	1267	603	[20]
Great Plague of 1738	1738	1738	50	50	50	470	814	[20]
Russian plague of 1770–1772	1770	1772	50	50	50	470	814	[20]
Persian Plague	1772	1772	2000	2000	2000	15444	990	[20]
Ottoman Plague Epidemic	1812	1819	300	300	300	2317	990	[22]
Caragea's plague	1813	1813	60	60	60	463	990	[22]
Malta plague epidemic	1813	1814	5	5	5	35	990	[20]
First cholera pandemic	1816	1826	100	100	100	772	990	[20]
Second cholera pandemic	1829	1851	100	100	100	772	990	[20]
Typhus epidemic in Canada	1847	1848	20	20	20	154	990	[20]
Third cholera pandemic	1852	1860	1000	1000	1000	6053	1263	[20]
Cholera epidemic of Copenhagen	1853	1853	5	5	5	29	1263	[22]
Third plague pandemic	1855	1960	15000	18500	22000	111986	1263	[20],[22]
Smallpox in British Columbia	1862	1863	3	3	3	18	1263	[22]
Fourth cholera pandemic	1863	1875	600	600	600	3632	1263	[22]
Fiji Measles outbreak	1875	1875	40	40	40	242	1263	[22]
Yellow Fever	1880	1900	100	125	150	757	1263	[18]
Fifth cholera pandemic	1881	1896	9	9	9	42	1654	[20]
Smallpox in Montreal	1885	1885	3	3	3	14	1654	[20]
Russian flu	1889	1890	1000	1000	1000	4620	1654	[20]
Sixth cholera pandemic	1899	1923	800	800	800	3696	1654	[22]
China plague	1910	1912	40	40	40	185	1654	[20]
Encephalitis lethargica pandemic	1915	1926	1500	1500	1500	6930	1654	[20]
American polio epidemic	1916	1916	6	7	7	30	1654	[20]
Spanish flu	1918	1920	17000	58500	100000	193789	2307	[20]
HIV/AIDS pandemic	1920	2020	25000	30000	35000	61768	3712	[18]
Poliomyelitis in USA	1946	1946	2	2	2	5	2948	[20]
Asian flu	1957	1958	2000	2000	2000	5186	2948	[20]
Hong Kong flu	1968	1969	1000	1000	1000	2102	3637	[20]
London flu	1972	1973	1	1	1	2	3866	[20]
Smallpox epidemic of India	1974	1974	15	15	15	29	4016	[20]
Zimbabwean cholera outbreak	2008	2009	4	4	4	5	6788	[20]
Swine Flu	2009	2009	152	364	575	409	6788	[20]
Haiti cholera outbreak	2010	2020	10	10	10	11	7253	[20]
Measles in D.R. Congo	2011	1018	5	5	5	5	7253	[20]
Ebola in West Africa	2013	2016	11	11	11	12	7176	[20]
Indian swine flu outbreak	2015	2015	2	2	2	2	7253	[20]
Yemen cholera outbreak	2016	2020	4	4	4	4	7643	[20]
2018–19 Kivu Ebola epidemic	2018	2020	2	2	3	2	7643	[20]
2019-20 COVID-19 Pandemic	2019	2020	33	50	66	50	7643	[22]
Measles in D.R. Congo	2019	2020	5	5	5	5	7643	[22]
Dengue fever	2019	2020	2	2	2	2	7643	[20]

Table 1: The data set used for the analysis. All estimates in thousands, apart from coeval population, which is expressed in millions. For Covid-19, the upper estimate includes the supposed number of Chinese victims (42K) for some Western media.