# What are the chances of a third world war?

Pasquale Cirillo and Nassim Nicholas Taleb

**To appear in *Significance*.**

IN [a recent] issue of *Significance* Mr. Peter McIntyre asked what the chances are that World War III will occur this century.

Prof. Michael Spagat wrote that nobody knows, nobody can really answer–and we totally agree with him on this. Then he adds that "a really huge war is possible but, in my view, extremely unlikely." To support his statement, Prof. Spagat relies partly on the popular science work of Prof. Steven Pinker, expressed in The Better Angels of our Nature and journalistic venues. Prof. Pinker claims that the world has experienced a long-term decline in violence, suggesting a structural change in the level of belligerence of humanity.

It is unfortunate that Prof. Spagat, in his answer, refers to our paper [1], which is part of a more ambitious project we are working on related to fat-tailed variables.

What characterizes fat tailed variables? They have their properties (such as the mean) dominated by extreme events, those "in the tails". The most popularly known version is the "Pareto 80/20".

We show that, simply, data do not support the idea of a structural change in human belligerence. So Prof. Spagat's first error is to misread our claim: we are making neither pessimistic nor optimistic declarations: we just believe that statisticians should abide by the foundations of statistical theory and avoid telling data what to say.

Let us go back to first principles.

## Foundational Principles

Fundamentally, statistics is about ensuring people do not build scientific theories from hot air, that is without significant departure from random. Otherwise, it is patently "fooled by randomness".

Further, for fat tailed variables, the conventional mechanism of the law of large numbers is considerably slower and significance requires more data and longer periods. Ironically, there are claims that can be done on little data: inference is asymmetric under fat-tailed domains. ***We require more data to assert that there are no black swans than to assert that there are black swans*** hence we would need much more data to claim a drop in violence than to claim a rise in it [2].

Finally, statements that are not deemed statistically significant –and shown to be so –should never be used to construct scientific theories.

These foundational principles are often missed because, typically, social scientists' statistical training is limited to mechanistic tools from thin tailed domains [2]. In physics, one can often claim evidence from small data sets, bypassing standard statistical methodologies, simply because the variance for these variables is low. The higher the variance, the more data one needs to make statistical claims. For fat-tails, the variance is typically high and underestimated in past data.

The second –more serious –error Spagat and Pinker made is to believe that tail events and the mean are somehow different animals, not realizing that the mean includes these tail events. ***For fat-tailed variables, the mean is almost entirely determined by extremes. If you are uncertain about the tails, then you are uncertain about the mean***. It is thus incoherent to say that violence has dropped but maybe not the risk of tail events; it would be like saying that someone is "extremely virtuous except during the school shooting episode when he killed 30 students".

## Robustness

Our study tried to draw the most robust statistical picture of violence, relying on methods from extreme value theory and statistical methods adapted to fat tails. We also put robustness checks to deal with the imperfection of data collected some thousand years ago: our results need to hold even if a third (or more) of the data were wrong.

## Inter-arrival times

We show that the inter-arrival times among major conflicts are extremely long, and consistent with a homogenous Poisson process: therefore no specific trend can be established: we as humans can not be deemed as less belligerent than usual. For a conflict generating at least 10 million casualties, an event less bloody than WW1 or WW2, the waiting time is on average 136 years, with a mean absolute deviation of 267 (or 52 years and 61 deviations for data rescaled to today's population). The seventy years of what is called the "Long Peace" are clearly not enough to state much about the possibility of WW3 in the near future.

## Underestimation of the mean

We also found that the average violence observed in the past underestimates the true statistical average by at least half. Why? Consider that about 90-97% of the observations fall below the mean, which requires some corrections with the help of extreme value theory. (Under extreme fat tails, the statistical mean can be closer to the past maximum observation than sample average.)

## A common mistake

Similar mistakes have been made in the past. In 1860, one H.T. Buckle used the same unstatistical reasoning as Pinker and Spagat.

That this barbarous pursuit is, in the progress of society, steadily declining, must be evident, even to the most hasty reader of European history. If we compare one country with another, we shall find that for a very long period wars have been becoming less frequent; and now so clearly is the movement marked, that, until the late commencement of hostilities, we had remained at peace for nearly forty years: a circumstance unparalleled (...) The question arises, as to what share our moral feelings have had in bringing about this great improvement.

Moral feelings or not, the century following Mr. Buckle's prose turned out to be the most murderous in human history.

We conclude by saying that we find it fitting –and are honored –to expose fundamental statistical mistakes in a journal called *Significance*, as the problem is precisely about significance and conveying notions of statistical rigor to the general public.

### REFERENCES

[1] Cirillo P., Taleb N.N. (2016), On the statistical properties and tail risk of violent conflicts. *Physica A: Statistical Mechanics and Its Applications*.

[2] Taleb, N.N. (2007), *The Black Swan: The Impact of the Highly Improbable*, Penguin.

[3] Buckle, H.T. (1858) *History of Civilization in England, Vol. 1*, London: John W. Parker and Son.

# On the statistical properties and tail risk of violent conflicts

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

*Abstract*—We examine statistical pictures of violent conflicts over the last 2000 years, finding techniques for dealing with incompleteness and unreliability of historical data.

We introduce a novel approach to apply extreme value theory to fat-tailed variables that have a remote, but nonetheless finite upper bound, by defining a corresponding unbounded dual distribution (given that potential war casualties are bounded by the world population).

We apply methods from extreme value theory on the dual distribution and derive its tail properties. The dual method allows us to calculate the real mean of war casualties, which proves to be considerably larger than the sample mean, meaning severe underestimation of the tail risks of conflicts from naive observation. We analyze the robustness of our results to errors in historical reports, taking into account the unreliability of accounts by historians and absence of critical data.

We study inter-arrival times between tail events and find that no particular trend can be asserted.

All the statistical pictures obtained are at variance with the prevailing claims about "long peace", namely that violence has been declining over time.

## I. Introduction

Since the middle of last century, there has been a multidisciplinary interest in wars and armed conflicts (quantified in terms of casualties), see for example [5], [15], [24], [30], [31], [32], [44] and [45]. Studies have also covered the statistics of terrorism, for instance [8], [35], and the special issue of *Risk Analysis* on terrorism [33]. From a statistical point of view, recent contributions have attempted to show that the distribution of war casualties (or terrorist attacks' victims) tends to have heavy tails, characterized by a power law decay [8] and [14]. Often, the analysis of armed conflicts falls within the broader study of violence [6], [30], with the aim to understand whether we as human are more or less violent and aggressive than in the past and what role institutions played in that respect. Accordingly, the public intellectual arena has witnessed active debates, such as the one between Steven Pinker on one side, and John Gray on the other concerning the hypothesis that the long peace was a statistically established phenomenon or a mere statistical sampling error that is characteristic of heavy-tailed processes, [16] and [27] –the latter of which is corroborated by this paper.

Using a more recent data set containing 565 armed conflicts with more than 3000 casualties over the period 1-2015 AD, we confirm that the distribution of war casualties exhibits a very heavy right-tail. The tail is so heavy that — at first glance — war casualties could represent an infinite-mean phenomenon,

as defined by [26]. But should the distribution of war casualties have an infinite mean, the annihilation of the human species would be just a matter of time, and the sample properties we can compute from data have no meaning at all in terms of statistical inference. In reality, a simple argument allows us to rule out the infiniteness of the mean: no event or series of events can kill more than the whole world population. The support of the distribution of war casualties is necessarily bounded, and the true mean cannot be infinite.

Let $[L, H]$ be the support of the distribution of war casualties today. $L$ cannot be smaller than 0, and we can safely fix it at some level $L^* >> 0$ to ignore those small events that are not readily definable as armed conflict [44]. As to $H$, its value cannot be larger than the world population, i.e. 7.2 billion people in 2015[1] [40].

If $Y$ is the random variable representing war casualties, its range of variation is very large and equal to $H - L$. Studying the distribution of $Y$ can be difficult, given its bounded but extremely wide support. Finding the right parametric model among the family of possible ones is hard, while nonparametric approaches are more difficult to interpret from a risk analysis point of view.

Our approach is to transform the data in order to apply the powerful tools of extreme value theory. Since $H < \infty$ we suggest a log-transformation of the data. This allows to use tools such as [9]. Theoretical results such as the Pickands, Balkema and de Haan's Theorem allow us to simplify the choice of the model to fit to our data.

Let $L$ and $H$ be respectively the lower and the upper bound of a random variable Y, and define the function

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

It is easy to verify that

1) $\varphi$ is "smooth": $\varphi \in C^\infty$,
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 $Y$. In what follows, we will call the distributions of $Y$ and $Z$, respectively the real and the dual distribution.

---

[1]Today's world population can be taken as the upper bound, as never before humanity reached similar numbers.
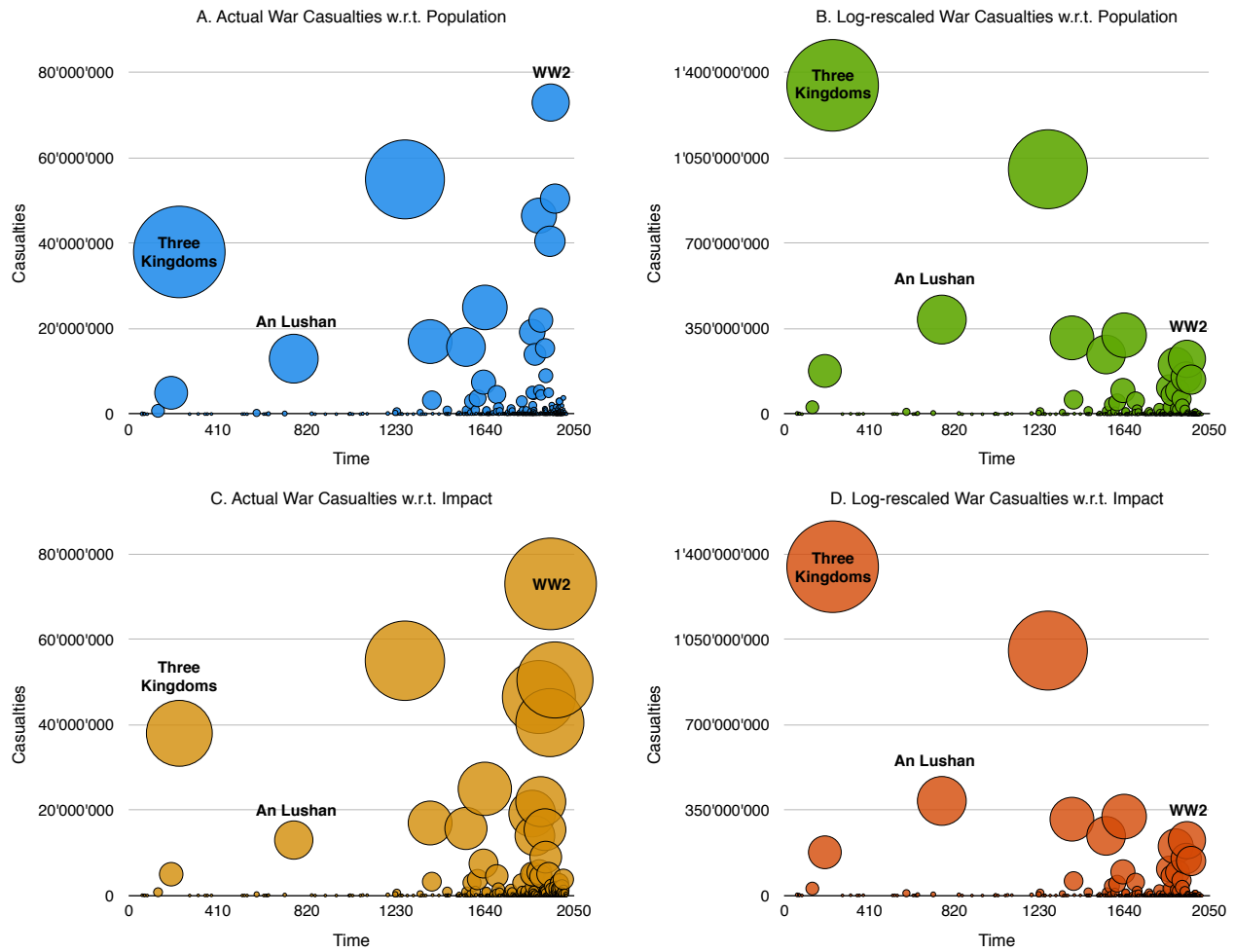
Fig. 1: War casualties over time, using raw (A,C) and dual (B,D) data. The size of each bubble represents the size of each event with respect to today's world population (A,B) and with respect to the total casualties (raw: C, rescaled: D) in the data set.

By studying the tail properties of the dual distribution (the one with an infinite upper bound), using extreme value theory, we will be able to obtain, by reverting to the real distribution, what we call the *shadow* mean of war casualties. We will show that this mean is at least 1.5 times larger than the sample mean, but nevertheless finite.

We assume that many observations are missing from the dataset (from under-reported conflicts), and we base on analysis on the fact that war casualties are just imprecise estimates [37], on which historians often have disputes, without anyone's ability to verify the assessments using period sources. For instance, an event that took place in the eighth century, the An Lushan rebellion, is thought to have killed 13 million people, but there no precise or reliable methodology to allow us to trust that number –which could be conceivably one order of magnitude smaller.[2]. Using resampling techniques, we show that our tail estimates are robust to changes in the quality and

reliability of data. Our results and conclusions will replicate even we missed a third of the data. When focusing on the more reliable set covering the last 500 years of data, one cannot observe any specific trend in the number of conflicts, as large wars appear to follow a homogeneous Poisson process. This memorylessness in the data conflicts with the idea that war violence has declined over time, as proposed by [15] or [30].

The paper is organized as follows: Section I describes our data set and analyses the most significant problems with the quality of observations; Section III is a descriptive analysis of data; it shows some basic result, which are already sufficient to refute the thesis [15] that we are living in a more peaceful world on the basis of statistical observations; Section IV contains our investigations about the upper tail of the dual distribution of war casualties as well as discussions of tail risk; Section V deals with the estimation of the shadow mean; in Section VI we discuss the robustness of our results to imprecision and errors in the data; finally, Section VII looks at the number of conflicts over time, showing no visible trend in the last 500 years.

---

[2]For a long time, an assessment of the drop in population in China was made on the basis of tax census, which might be attributable to a drop in the aftermath of the rebellion in surveyors and tax collectors.[4]

## II. ABOUT THE DATA

The data set contains 565 events over the period 1-2015 AD; an excerpt is shown in Table I. Events are generally armed conflicts[3], such as interstate wars and civil wars, with a few exceptions represented by violence against citizens perpetrated by the bloodiest dictatorships, such as Stalin's and Mao Zedong's regimes. These were included in order to be consistent with previous works about war victims and violence, e.g. [30] and [32].

We had to deal with the problem of inconsistency and lack of uniformity in the attribution of casualties by historians. Some events such as the siege of Jerusalem include death from famine, while for other wars only direct military victims are counted. It might be difficult to disentangle death from direct violence from those arising from such side effects as contagious diseases, hunger, rise in crime, etc. Nevertheless, as we show in our robustness analysis these inconsistencies do not affect out results —in contrast with analyses done on thin-tailed data (or data perceived to be so) where conclusions can be reversed on the basis of a few observations.

The different sources for the data are: [5], [18], [25], [29], [30], [32], [44], [45] and [46]. For websites like Necrometrics [25], data have been double-checked against the cited references. The first observation in our data set is the Boudicca's Revolt of 60-61 AD, while the last one is the international armed conflict against the Islamic State of Iraq and the Levant, still open.

We include all events with more than 3000 casualties (soldiers and civilians) in absolute terms, without any rescaling with respect to the coeval world population. More details about rescaling in Subsection II-D. The choice of the 3000 victims threshold was motivated by three main observations:

- Conflicts with a high number of victims are more likely to be registered and studied by historians. While it is easier today to have reliable numbers about minor conflicts — although this point has been challenged by [36] and [37] — it is highly improbable to ferret out all smaller events that took place in the fifth century AD. In particular, a historiographical bias prevents us from accounting for much of the conflicts that took place in the Americas and Australia, before their discovery by European conquerors.
- A higher threshold gives us a higher confidence about the estimated number of casualties, thanks to the larger number of sources, even if, for the bloodiest events of the far past, we must be careful about the possible exaggeration of numbers.
- The object of our concern is tail risk. The extreme value techniques we use to study the right tail of the distribution of war casualties imposition of thresholds, actually even larger than 3000 casualties.

To rescale conflicts and expressing casualties in terms of today's world population (more details in Subsection II-D),

we relied on the population estimates of [22] for the period 1-1599 AD, and of [39] and [40] for the period 1650-2015 AD. We used:

- Century estimates from 1 AD to 1599.
- Half-century estimates from 1600 to 1899.
- Decennial estimates from 1900 to 1949.
- Yearly estimates from 1950 until today.

### A. Data problems

Accounts of war casualties are often anecdotal, spreading via citations, and based on vague estimates, without anyone's ability to verify the assessments using period sources. For instance, the independence war of Algeria has various estimates, some from the French Army, others from the rebels, and nothing scientifically obtained [19].

This can lead to several estimates for the same conflict, some more conservative and some less. Table I, shows different estimates for most conflicts. In case of several estimates, we present the minimum, the average and the maximum one. Interestingly, as we show later, choosing one of the three as the "true" estimate does not affect our results –thanks to the scaling properties of power laws.

Conflicts, such as the Mongolian Invasions, which we refer to as "named" conflicts, need to be treated with care from a statistical point of view. Named conflicts are in fact artificial tags created by historians to aggregate events that share important historical, geographical and political characteristics, but that may have never really existed as a single event. Under the portmanteau Mongolian Invasions (or Conquests), historians collect all conflicts related to the expansion of the Mongol empire during the thirteenth and fourteenth centuries. Another example is the so-called Hundred-Years' War in the period 1337-1453. Aggregating all these events necessarily brings to the creation of very large fictitious conflicts accounting for hundreds of thousands or million casualties. The fact that, for historical and historiographical reasons, these events tends to be more present in antiquity and the Middle Ages could bring to a naive overestimation of the severity of wars in the past. Notice that named conflicts like the Mongolian Invasions are different from those like WW1 or WW2, which naturally also involved several tens of battles in very different locations, but which took place in a much shorter time period, with no major time separation among conflicts.

A straightforward technique could be to set a cutoff-point of 25 years for any single event –an 80 years event would be divided into three minor ones. However it remains that the length of the window remains arbitrary. Why 25 years? Why not 17?

As we show in Section IV, a solution to all these problems with data is to consider each single observation as an imprecise estimate, a fuzzy number in the definition of [43]. Using Monte Carlo methods we have shown that, if we assume that the real number of casualties in a conflict is uniformly distributed between the minimum and the maximum in the available data, the tail exponent $\xi$ is not really affected (apart from the

---

[3]We refer to the definition of [44], according to which "An armed conflict is a contested incompatibility which concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths". This definition is also compatible with the Geneva Conventions of 1949.

TABLE I: Excerpt of the data set of war casualties. The original data set contains 565 events in the period 1-2015 AD. For some events more than one estimate is available for casualties and we provide: the minimum (Min), the maximum (Max), and an intermediate one (Mid) according to historical sources. Casualties and world population estimates (Pop) in 10000.

| Event | Start | End | Min | Mid | Max | Pop |
|---|---|---|---|---|---|---|
| Boudicca's Revolt | 60 | 61 | 70 | 7.52 | 8.04 | 19506 |
| Three Kingdoms | 220 | 280 | 3600 | 3800 | 4032 | 20231 |
| An Lushan's Reb. | 755 | 763 | 800 | 1300 | 3600 | 24049 |
| Sicilian Vespers | 1282 | 1302 | - | 0.41 | 0.80 | 39240 |
| WW1 | 1914 | 1918 | 1466 | 1544 | 1841 | 177718 |
| WW2 | 1939 | 1945 | 4823 | 7300 | 8500 | 230735 |

obvious differences in the smaller decimals).[4] Similarly, our results remain invariant if we remove/add a proportion of the observations by bootstrapping our sample and generating new data sets.

We believe that this approach to data as imprecise observations is one of the novelties of our work, which makes our conclusions more robust to scrutiny about the quality of data. We refer to Section IV for more details.

### B. Missing events

We are confident that there are many conflicts that are not part of our sample. For example we miss all conflicts among native populations in the American continent before its European discovery, as no source of information is actually available. Similarly we may miss some conflicts of antiquity in Europe, or in China in the sixteenth century. However, we can assume that the great part of these conflicts is not in the very tail of the distribution of casualties, say in the top 10 or 20%. It is in fact not really plausible to assume that historians have not reported a conflict with 1 or 2 million casualties, or that such an event is not present in [25] and [29], at least for what concerns Europe, Asia and Africa.

Nevertheless, in Subsection VI, we deal with the problem of missing data in more details.

### C. The conflict generator process

We are not assuming the existence of a unique conflict generator process. It is clear that all conflicts of humanity do not share the same set of causes. Conflicts belonging to different centuries and continents are likely to be not only independent, as already underlined in [32] and [18], but also to have different origins. We thus avoid performing time series analysis beyond the straightforward investigation of the existence of trends. While it could make sense to subject the data to specific tests, such as whether an increase/decrease in war casualties is caused by the lethality of weapons, it would be unrigorous and unjustified to look for an autoregressive component. How could the An Lushan rebellion in China (755 AD) depend on the Siege of Constantinople by Arabs (717 AD), or affect the Viking Raids in Ireland (from 795 AD on)? But this does not mean that all conflicts are independent: during WW2, the attack on Pearl Harbor and the Battle of France

were not independent, despite the time and spatial separation, and that's why historians merge them into one single event. And while we can accept that most of the causes of WW2 are related to WW1, when studying numerical casualties, we avoid translating the dependence into the magnitude of the events: it would be naive to believe that the number of victims in 1944 depended on the death toll in 1917. How could the magnitude of WW2 depend on WW1?

Critically, when related conflicts are aggregated, as in the case of WW2 or for the Hundred Years' war (1337-1453), our data show that the number of conflicts over time — if we focused on very destructive events in the last 500/600 years and we believed in the existence of a conflict generator process — is likely to follow a homogeneous Poisson process, as already observed by [18] and [32], thus supporting the idea that wars are randomly distributed accidents over time, not following any particular trend. We refer to Section IV for more details.

### D. Data transformations

We used three different types of data in our analyses: raw data, rescaled data and dual data.

*1) Raw data:* Presented as collected from the different sources, and shown in Table I. Let $X_t$ be the number of casualties in a given conflict at time $t$, define the triplet $\{X_t, X_t^l, X_t^u\}$, where $X_t^l$ and $X_t^u$ represent the lower and upper bound, if available.

*2) Rescaled data:* The data is rescaled with respect to the current world population (7.2 billion[5]). Define the triplet $\{Y_t = X_t \frac{P_{2015}}{P_t}, Y_t^l = X_t^l \frac{P_{2015}}{P_t}, Y_t^u = X_t^u \frac{P_{2015}}{P_t}\}$, where $P_{2015}$ is the world population in 2015 and $P_t$ is the population at time $t = 1, ..., 2015$.[6] Rescaled data was used by Steven Pinker [30] to account for the relative impact of a conflict. This rescaling tends to inflate past conflicts, and can lead to the naive statement that violence, if defined in terms of casualties, has declined over time. But we agree with those scholars like Robert Epstein [12] stating "[...] why should we be content with only a relative decrease? By this logic, when we reach a world population of nine billion in 2050, Pinker will conceivably be satisfied if a mere two million people are killed in war that year". In this paper we used rescaled data

---

[4]Our results do not depend on choice of the uniform distribution, selected for simplification. All other bounded distributions produce the same results in the limit.

[5]According to the United Nations Department of Economic and Social Affairs [40]

[6]If for a given year, say 788 AD, we do not have a population estimate, we use the closest one.

for comparability reasons, and to show that even with rescaled data we cannot really state statistically that war casualties have declined over time (and thus violence).

*3) Dual data:* those we obtain using the log-rescaling of equation (1), the triplet $\{Z_t = \varphi(Y_t), Z_t^l = \varphi(Y_t^l), Z_t^u = \varphi(Y_t^u)\}$. The input for the function $\varphi(\cdot)$ are the rescaled values $Y_t$.

Removing the upper bound allows us to combine practical convenience with theoretical rigor. Since we want to apply the tools of extreme value theory, the finiteness of the upper bound is fundamental to decide whether a heavy-tailed distribution falls in the Weibull (finite upper bound) or in the Fréchet class (infinite upper bound). Given their finite upper bound, war casualties are necessarily in the Weibull class. From a theoretical point of view the difference is large [9], especially for the existence of moments. But from a practical point of view, heavy-tailed phenomena are better modeled within the Fréchet class. Power laws belong to this class. And as observed by [11], when a fat-tailed random variable $Y$ has a very large finite upper bound $H$, unlikely to be ever touched (and thus observed in the data), its tail, which from a theoretical point of view falls in the so-called Weibull class, can be modeled as if belonging to the Fréchet one. In other words, from a practical point of view, when $H$ is very large, it is impossible to distinguish the two classes; and in most of the cases this is indeed not a problem.

Problems arise when the tail looks so thick that even the first moment appears to be infinite, when we know that this is not possible. And this is exactly the case with out data, as we show in the next Section.
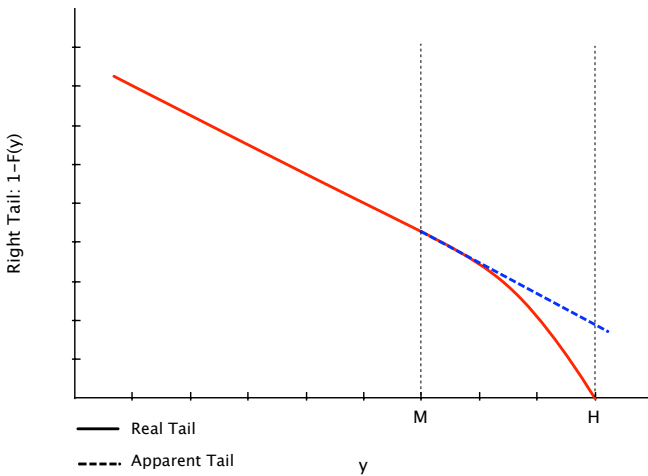


Fig. 2: Graphical representation (Log-Log Plot) of what may happen if one ignores the existence of the finite upper bound $H$, since only $M$ is observed.

Figure 2 shows illustrates the need to separate real and dual data. For a random variable $Y$ with remote upper bound $H$, the real tail is represented by the continuous line. However, if we only observe values up to $M$, and we ignore the existence of $H$, which is unlikely to be reached and therefore hardly observable, we could be inclined to believe the the tail is the dotted one, the apparent one. The two tails are indeed

essentially indistinguishable for most cases, but the divergence is evident when we approach $H$. Hence our transformation. Notice that $\varphi(y) \approx y$ for very large values of $H$. This means that for a very large upper bound, unlikely to be reached as the world population, the results we get for the tail of $Y$ and $Z = \varphi(Y)$ are essentially the same most of the times. This is exactly what happens with our data, considered that no armed conflict has ever killed more than 19% of the world population (the Three Kingdoms, 184-280 AD). But while $Y$ is bounded, $Z$ is not.

Therefore we can safely model $Z$ as belonging to the Fréchet class, study its tail behavior, and then come back to $Y$ when we are interested in the first moments, which under $Z$ could not exist.

### III. DESCRIPTIVE DATA ANALYSIS

Figure 1 showing casualties over time, is composed of four subfigures, two related to raw data and two related to dual data[7]. For each type of data, using the radius of the different bubbles, we show the relative size of each event in terms of victims with respect to the world population and within our data set (what we call Impact). Note that the choice of the type of data (or of the definition of size) may lead to different interpretation of trends and patterns. From rescaled and dual data, one could be lead to superficially infer a decrease in the number of casualties over time.

As to the number of armed conflicts, Figure 1 seems to suggest an increase over time, as we see that most events are concentrated in the last 500 years or so, an apparent illusionary increase most likely due to a reporting bias for the conflicts of antiquity and early Middle Ages.

Figure 3 shows the histogram of war casualties, when dealing with raw data. Similar results are obtained using rescaled and dual casualties. The graphs suggests the presence of a long right tail in the distribution of victims, and the maximum is represented by WW2, with an estimated amount of about 73 million victims.

The presence of a Paretian tail is also supported by the quantile-quantile plot in Figure 4, where we use the exponential distribution as a benchmark. The concave behavior we can observe in the plot is considered a good signal of fat-tailed distribution [11], [21].

In order to investigate the presence of a right fat tail, we can make use of another graphical tool: the *meplot*, that is, mean excess function plot.

Let $X$ be a random variable with distribution $F$ and right endpoint $x_F$ (i.e. $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$). The function

$$e(u) = E[X-u|X > u] = \frac{\int_u^\infty (t-u)\mathrm{d}F(t)}{\int_u^\infty \mathrm{d}F(t)}, \qquad 0 < u < x_F,$$
(2)

is called mean excess function of $X$ (mef). The empirical mef of a sample $X_1, X_2,..., X_n$ is easily computed as

$$e_n(u) = \frac{\sum_{i=1}^n (X_i - u)}{\sum_{i=1}^n 1_{\{X_i > u\}}},$$
(3)

[7] Given Equation (1), rescaled and dual data are approximately the same, hence there is no need to show further pictures.
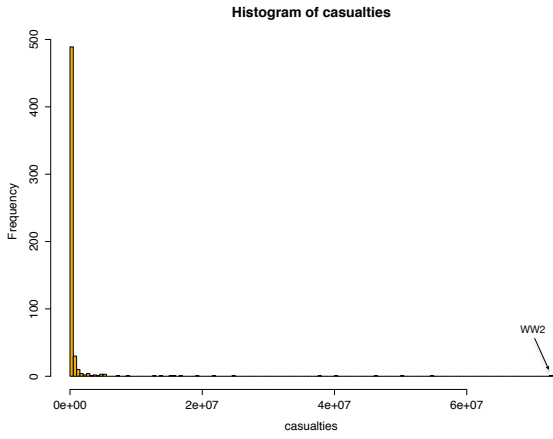
Fig. 3: Histogram of war casualties, using raw data. A long fat right tail is clearly visible.

that is the sum of the exceedances over the threshold $u$ divided by the number of such data points. Interestingly, the mean excess function is a way of characterizing distributions within the class of continuous distributions [21]. For example, power law distributions are characterized by the van der Wijk's law [41], that is, by a mean excess function linearly increasing in the threshold $u$, while an exponential distribution of parameter $\lambda$ would show a constant mean excess function with value $\lambda$.

Figure 5 shows the the meplot of war casualties, again using raw data. The graph is easily obtained by plotting the pairs $\{(X_{i:n}, e_n(X_{i:n})) : i = 1, ..., n\}$, where $X_{i:n}$ is the $i$−th order statistic, used as a threshold. The upward trend in Figure 5 is a further signal of fat-tailed data, as discussed for example in [7] and [11].



Fig. 5: Mean Excess Function Plot (meplot) for war casualties, using raw data. A steep upward trend is visible, suggesting the presence of a Paretian right tail.

variance is not finite, as is often the case with financial data, it becomes more problematic to build confidence intervals for the mean. Similarly, if the third moment (skewness) is not defined, it is risky to build confidence intervals for the variance.

An interesting graphical tool showing the behavior of moments in a given sample, is the *Maximum-to-Sum plot*, or MS Plot. The MS Plot relies on simple consequence of the law of large numbers [11]. For a sequence $X_1, X_2, ..., X_n$ of nonnegative i.i.d. random variables, if for $p = 1, 2, 3...$, $E[X^p] < \infty$, 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, and $M_n^p = \max(X_1^p, ..., X_n^p)$ the partial maximum.



Fig. 4: Exponential qq-plot of war casualties, using raw data. The clear concave behavior suggest the presence of a right fat tail.

A characteristic of the power law class is the non-existence (infiniteness) of moments, at least the higher-order ones, with important consequences in terms of statistical inference. If the
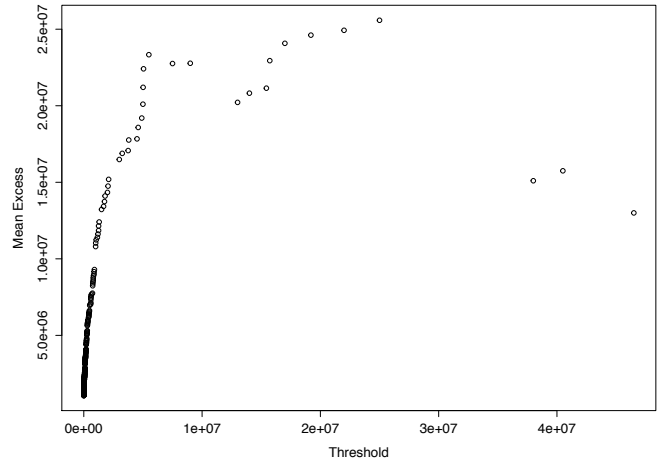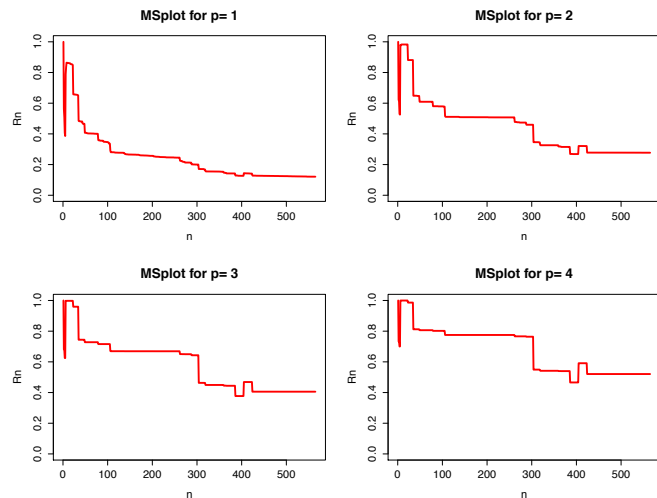


Fig. 6: Maximum to Sum plot to verify the existence of the first four moments of the distribution of war casualties, using raw data. No clear converge to zero is observed, for $p = 2, 3, 4$, suggesting that the moments of order 2 or higher may not be finite.

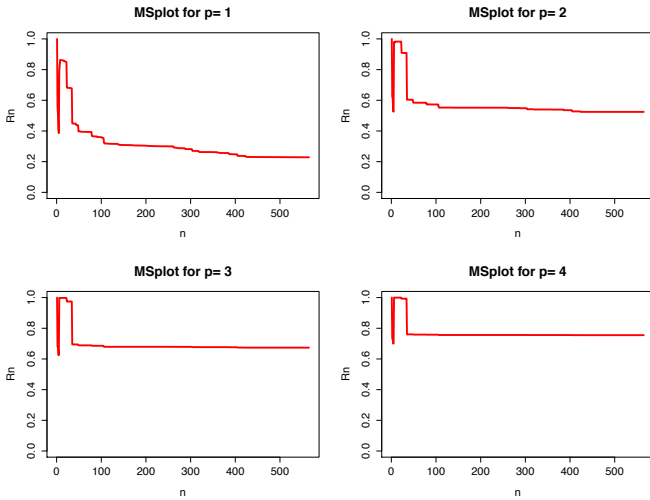Figures 6 and **??** show the MS Plots of war casualties for

Fig. 7: Maximum to Sum plot to verify the existence of the first four moments of the distribution of war casualties, using rescaled data. No clear converge to zero is observed, suggesting that all the first four moments may not be finite.

raw and rescaled data respectively. No finite moment appears to exist, no matter how much data is used: it is in fact clear that the $R_n$ ratio does not converge to 0 for $p = 1, 2, 3, 4$, thus suggesting that the distribution of war casualties has such a fat right tail that not even the first moment exists. This is particularly evident for rescaled data.

To compare to thin tailed situations, 8 shows the MS Plot of a Pareto ($\alpha = 1.5$): notice that the first moment exists, while the other moments are not defined, as the shape parameter equals $1.5$. For thin tailed distributions such as the Normal, the MS Plot rapidly converges to 0 for all $p$.
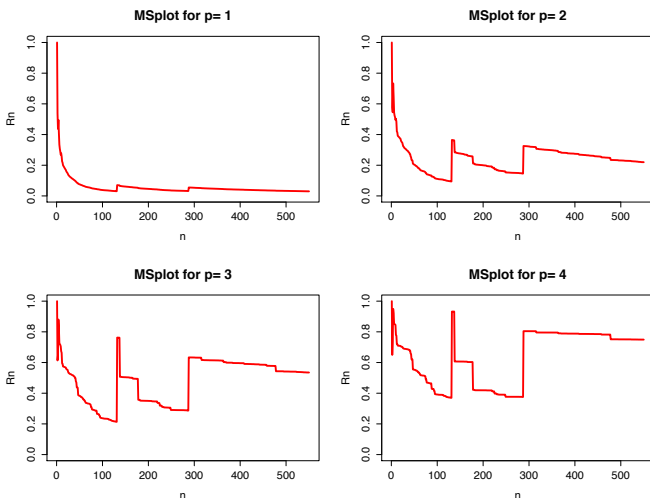


Fig. 8: Maximum to Sum plot of a Pareto(1.5): as expected, the first moment is finite ($R_n \to 0$), while higher moments do not exists ($R_n^p$ is erratic and bounded from 0 for $p = 2, 3, 4$).

Table II provides information about armed conflicts and their occurrence over time. Note that we have chosen very large thresholds to minimize the risk of under-reporting. We

TABLE II: Average inter-arrival times (in years) and their mean absolute deviation for events with more than 0.5, 1, 2, 5, 10, 20 and 50 million casualties, using raw (Ra) and rescaled data (Re).

| Thresh. | Avg Ra | MAD Ra | Avg Re | MAD Re |
|---------|--------|--------|--------|--------|
| 0.5     | 23.71  | 35.20  | 9.63   | 15.91  |
| 1       | 34.08  | 47.73  | 13.28  | 20.53  |
| 2       | 56.44  | 72.84  | 20.20  | 28.65  |
| 5       | 93.03  | 113.57 | 34.88  | 46.85  |
| 10      | 133.08 | 136.88 | 52.23  | 63.91  |
| 20      | 247.14 | 261.31 | 73.12  | 86.19  |
| 50      | 345.50 | 325.50 | 103.97 | 114.25 |

compute the distance, in terms of years, between two time-contiguous conflicts, and use for measure of dispersion the mean absolute deviation (from the mean). Table II shows the average inter-arrival times between armed conflicts with at least 0.5, 1, 2, 5, 10, 20 and 50 million casualties. For a conflict of at least 500k casualties, we need to wait on average 23.71 years using raw data, and 9.63 years using rescaled or dual data (which, as we saw, tend to inflate the events of antiquity). For conflicts with at least 5 million casualties, the time delay is on average 93.03 or 34.88, depending on rescaling. Clearly, the bloodier the conflict, the longer the inter-arrival time. The results essentially do not change if we use the mid or the ending year of armed conflicts.

The consequence of this analysis is that the absence of a conflict generating more than, say, 5 million casualties in the last sixty years highly insufficient to state that their probability has decreased over time, given that the average inter-arrival time is 93.03 years, with a mean absolute deviation of 113.57 years! Unfortunately, we need to wait for more time to assert whether we are really living in a more peaceful era,: present data are not in favor (nor against) a structural change in violence, when we deal with war casualties.

Section I asserted that our data set do not form a proper time series, in the sense that no real time dependence is present (apart from minor local exceptions).[8] Such consideration is further supported by the so-called record plot in Figure 9. This graph is used to check the i.i.d. nature of data, and it relies on the fact that, if data were i.i.d., then records over time should follow a logarithmic pattern [11]. Given a sequence $X_1, X_2, ...$, and defined the partial maximum $M_n$, an observation $x_n$ is called a record if $X_n = M_n = x_n$.

For the sake of completeness, if we focus our attention on shorter periods, like the 50 years following WW2, a reduction in the number and the size of the conflict can probably be observed — the so-called Long Peace of [15], [24] and [30]; but recent studies suggest that this trend could already have changed in the last years [20], [27]. However, in our view, this type of analysis is not meaningful, once we account for the

---

[8]Even if redundant, given the nature of our data and the descriptive analysis performed so far, the absence of a relevant time dependence can be checked by performing a time series analysis of war casualties. No significant trend, season or temporal dependence can be observed over the entire time window.
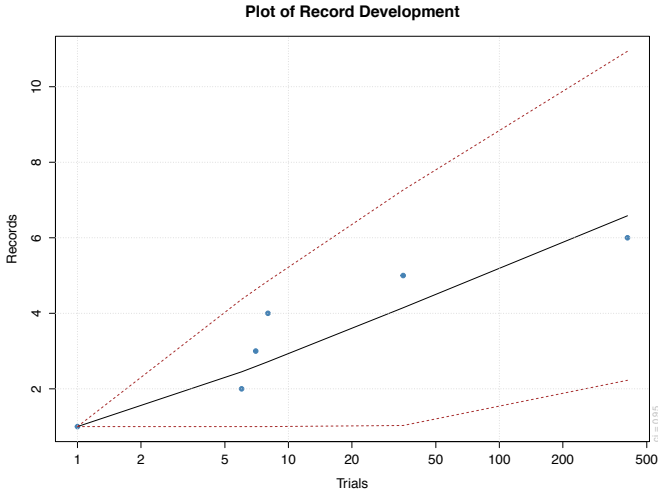
Fig. 9: Record plot for testing the plausibility of the i.i.d. hypothesis, which seems to be supported.

extreme nature of armed conflicts, and the long inter-arrival times between them, as shown in Section III.

## IV. TAIL RISK OF ARMED CONFLICTS

Our data analysis in Section III suggesting a heavy right tail for the distribution of war casualties, both for raw and rescaled data is consistent with the existing literature, e.g. [8], [14], [35] and their references. Using extreme value theory, or EVT [2], [3], [9], [11], [17], [38], we use the Generalized Pareto Distribution (GPD). The choice is due to the Pickands, Balkema and de Haan's Theorem [1], [28].

Consider a random variable $X$ with distribution function $F$, and call $F_u$ the conditional df of $X$ above a given threshold $u$. We can then define the r.v. $W$, representing the rescaled excesses of $X$ over the threshold $u$, so that

$$F_u(w) = P(X - u \le w | X > u) = \frac{F(u + w) - F(u)}{1 - F(u)},$$

for $0 \le w \le x_F - u$, where $x_F$ is the right endpoint of the underlying distribution $F$.

Pickands, Balkema and de Haan [1], [28] have showed that for a large class of underlying distribution functions $F$, and a large $u$, $F_u$ can be approximated by a Generalized Pareto distribution (hence GPD), i.e. $F_u(w) \to G(w)$, as $u \to \infty$ where

$$G(w) = \begin{cases} 1 - (1 + \xi \frac{w}{\sigma})^{-1/\xi} & if \, \xi \ne 0 \\ 1 - e^{-\frac{w}{\sigma}} & if \, \xi = 0 \end{cases}, \, w \ge 0. \quad (4)$$

The parameter $\xi$, known as the shape parameter, is the central piece of information for our analysis: $\xi$ governs the fatness of the tails, and thus the existence of moments. The moment of order $p$ of a Generalized Pareto distributed random variable only exists if and only if $\xi < 1/p$ [11].

The Pickands, Balkema and de Haan's Theorem thus allows us to focus on the right tail of the distribution without caring too much about what happens below the threshold $u$. One powerful property of the generalized Pareto is the tail stability

with respect to threshold [11]. Formally, if $W \sim \text{GPD}(\xi, \sigma_1)$, for $W \ge u_1$, then $W \sim \text{GPD}(\xi, \sigma_2)$ for $W \ge u_2 > u_1$. In other words: increasing the threshold does not affect the shape parameter. What changes is only the scale parameter.

We start by identifying the threshold $u$ above which the GPD approximation may hold. Different heuristic tools can be used for this purpose, from Zipf plots, i.e. log log plots of the survival function, to mean excess function plots (as the one in Figure 5), where one looks for the threshold above which it is possible to visualize — if present — the linearity that is typical of fat-tailed phenomena. Other possibilities like the Hill plot are discussed for example in [9].

Our investigations suggest that the best threshold for the fitness of GPD is $25,000$ casualties, when using raw data, well above the $3,000$ minimum we have set in collecting our observations. A total of 331 armed conflicts lie above this threshold (58.6% of all our data).

The use of rescaled and dual data requires us to rescale the threshold; the value thus becomes 145k casualties. It is worth noticing that, nevertheless, for rescaled data the threshold could be lowered to 50k, and still we would have a satisfactory GPD fitting.

If $\xi > -1/2$, the parameters of a GPD can be easily estimated using Maximum Likelihood Estimation (MLE) [9], while for $\xi \le -1/2$ MLE estimates are not consistent and other methods are better used [11].

In our case, given the fat-tailed behavior, $\xi$ is likely to be positive. This is clearly visible in Figure 10, showing the Pickands plot, based on the Pickands' estimator for $\xi$, defined as

$$\hat{\xi}_{k,n}^{(P)} = \frac{1}{\log 2} \log \frac{X_{k,n} - X_{2k,n}}{X_{2k,n} - X_{4k,n}},$$

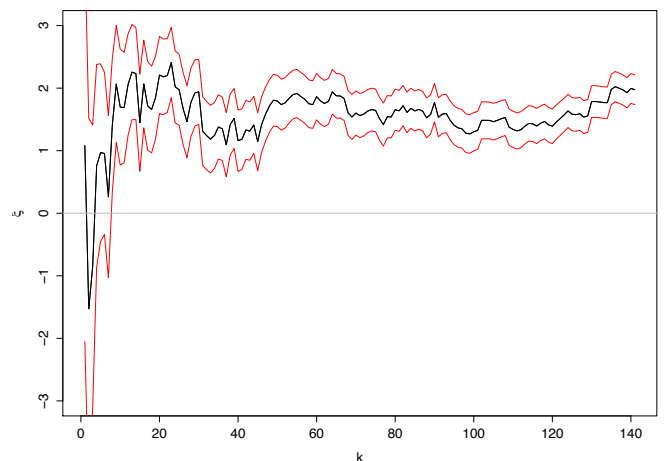where $X_{k,n}$ is the $k$-th upper order statistics out of a sample of $n$ observations.



Fig. 10: Pickands' estimator for the $\xi$ shape parameter for war casualties (raw data). The value 1.5 appears to be a good educated guess.

In the Pickands plot, $\hat{\xi}_{k,n}^{(P)}$ is computed for different values of $k$, and the "optimal" estimate of $\xi$ is obtained by looking

TABLE III: Maximum likelihood estimates (and standard errors) of the Generalized Pareto Distribution parameters for casualties over different thresholds, using raw, rescaled and dual data. We also provide the number of events lying above the thresholds, the total number of observations being equal to 565.

| **Data** | Threshold | $\xi$ | $\sigma$ |
|---|---|---|---|
| Raw Data | 25k | 1.4985 | 90620 |
| | | (0.1233) | (2016) |
| Rescaled Data | 145k | 1.5868 | 497436 |
| | | (0.1265) | (2097) |
| Dual Data | 145k | 1.5915 | 496668 |
| | | (0.1268) | (1483) |

at a more or less stable value for $\hat{\xi}_{k,n}^{(P)}$ in the plot. In Figure 10, the value $1.5$ seems to be a good educated guess for raw data, and similar results hold for rescaled and dual amounts. In any case, the important message is that $\xi > 0$, therefore we can safely use MLE.

Table III presents our estimates of $\xi$ and $\sigma$ for raw, rescaled and dual data. In all cases, $\xi$ is significantly greater than 1, and around $1.5$ as we guessed by looking at Figure 10. This means that in all cases the mean of the distribution of casualties seems to be infinite, consistently with the descriptive analysis of Section III.

Further, Table III shows the similarity between rescaled and dual data. Looking at the standard errors, no test would reject the null hypothesis that $\xi_{rescaled} = \xi_{dual}$, or that $\sigma_{rescaled} = \sigma_{dual}$.

Figure 11 shows the goodness of our GPD fit for dual data, which is also supported by goodness-of-fit tests for the Generalized Pareto Distribution [42] (p-value: 0.37). As usual, similar results do hold for raw and rescaled data. Figure 12 shows the residuals of the GPD fit which are, as can be expected (see [11]), exponentially distributed.

## V. ESTIMATING THE SHADOW MEAN

Given the finite upper bound $H$, we known that $E[Y]$ must be finite as well. A simple idea is to estimate it by computing the sample mean, or the conditional mean above a given threshold[9], if we are more interested in the tail behavior of war casualties, as in this paper. For a minimum threshold $L$ equal to 145k, this value is $1.77 \times 10^7$ (remember that $Y$ represents the rescaled data).

However, in spite of its boundedness, the support of $Y$ is so large that $Y$ behaves like a very heavy-tailed distribution, at least until we do not approach the upper bound $H$. This makes the (conditional) sample mean not really reliable, or robust: one single extreme observation can make it jump, given that the (conditional) sample mean has a breakdown point equal to 0 [23].

A more robust way of computing what we define the conditional shadow mean of $Y$ (the true conditional mean not

[9]Notice that the sample mean above a minimum threshold $L$ corresponds to the concept of expected shortfall used in risk management.
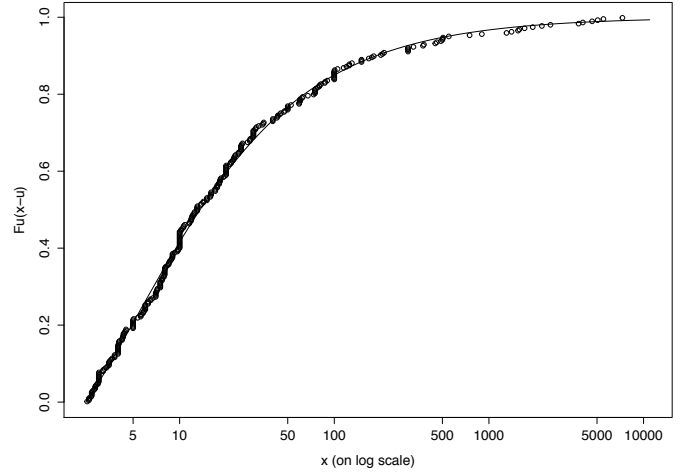


(a) Distribution of the excesses (ecfd and theoretical).
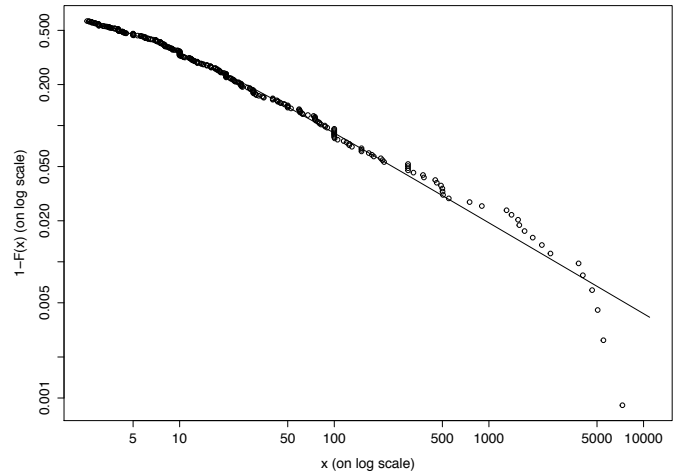


(b) Right Tail (empirical and theoretical).

Fig. 11: GPD fitting of war casualties exceeding the 145k threshold, using dual data.

visible from data) is therefore to use the log-transformation of Equation (1).

Let $F$ and $G$ be respectively the distribution of $Y$ and $Z$. With $f$ and $g$, we indicate the densities.

In the previous section we have seen that, when using the dual version of rescaled data, $G \approx \text{GPD}(\xi, \sigma)$ for $Z \geq L = 145000$ (even 50k gives good results, but 145k is comparable to the 25k threshold of raw data). Moreover we know that $Z = \varphi(Y)$, so that $Y = \varphi^{-1}(Z) = (L - H)e^{\frac{L-Z}{H}} + H$.
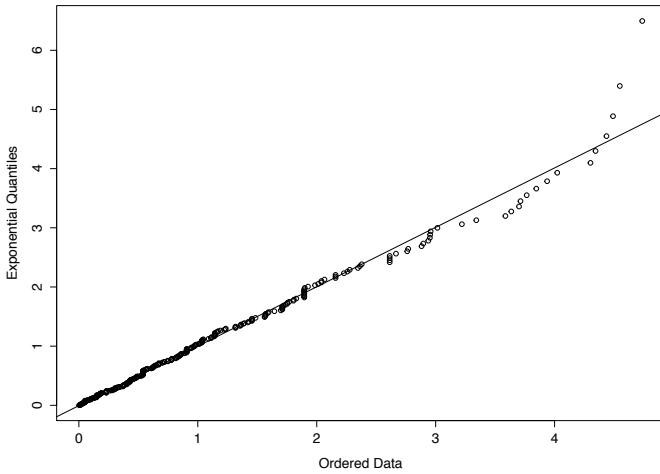
This implies that the distribution of $Y$, for $Y \geq L$, can be obtained from the distribution of $G$.

First we have

$$\int_L^\infty g(z)\,\mathrm{d}z = \int_L^{\varphi^{-1}(\infty)} f(y)\,\mathrm{d}y. \tag{5}$$
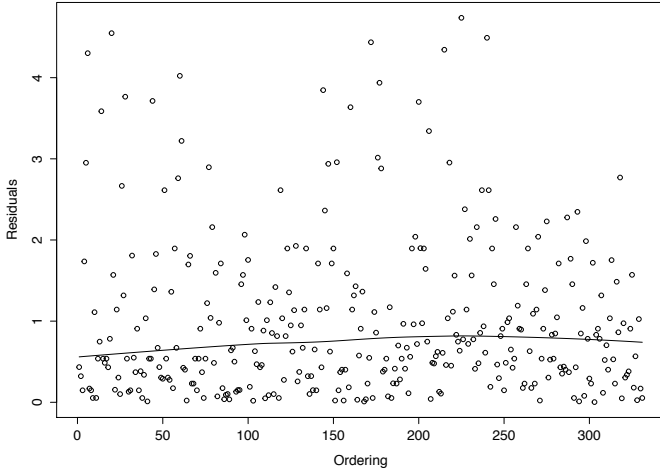
(a) Scatter plot of residuals and their smoothed average.



(b) QQ-plot of residuals against exponential quantiles.

Fig. 12: Residuals of the GPD fitting: looking for exponentiality.

And we know that

$$g(z; \xi, \sigma) = \frac{1}{\sigma} \left( 1 + \frac{\xi z}{\sigma} \right)^{-\frac{1}{\xi} - 1}, \qquad z \in [L, \infty). \quad (6)$$

This implies, setting $\alpha = \xi^{-1}$ and $k = \sigma/H$,

$$f(y; \alpha, k) = \frac{\left( 1 - \frac{\log(H-y) - \log(H-L)}{\alpha k} \right)^{-\alpha - 1}}{(H - y)k}, \qquad y \in [L, H]. \quad (7)$$

We can then derive the shadow mean of $Y$ as

$$E[Y] = \int_L^H y \, f(y) \, \mathrm{d}y, \quad (8)$$

obtaining

$$E[Y] = L + (H - L)e^{\alpha k}(\alpha k)^\alpha \Gamma(1 - \alpha, \alpha k). \quad (9)$$

Table IV provides the shadow mean, the sample mean and

their ratio for different minimum thresholds $L$, using rescaled data. As already said, our GPD fit is already satisfactory for $L = 50000$, that is why we start from that threshold.

The ratios in Table IV show how the sample mean underestimates the shadow mean, especially for lower thresholds, for which the shadow mean is almost 3 times larger than the sample mean. For these reasons, a "journalistic" reliance on sample mean would be highly misleading, when considering all conflict together, without setting large thresholds. There would be the serious risk of underestimating the real lethality of armed conflicts.

TABLE IV: Conditional shadow mean, conditional sample mean and their ratio for different minimum thresholds. In bold the values for the 145k threshold. Rescaled data.

| Thresh. $\times 10^3$ | Shadow $\times 10^7$ | Sample $\times 10^7$ | Ratio |
|---|---|---|---|
| 50 | 3.6790 | 1.2753 | 2.88 |
| 100 | 3.6840 | 1.5171 | 2.43 |
| **145** | **3.6885** | **1.7710** | **2.08** |
| 300 | 4.4089 | 2.2639 | 1.95 |
| 500 | 5.6542 | 2.8776 | 1.96 |
| 1000 | 6.5511 | 3.9481 | 1.85 |

A similar procedure can be applied to raw data. Fixing $H$ equal to 7.2 billion and $L = 25000$, we can define $V = \varphi(X)$ and study its tail. The tail coefficient $\xi_V$ is 1.4970. Even in this case the sample mean underestimates the shadow mean by a factor of at least 1.6 (up to a maximum of 7.5!).

These results tell us something already underlined by [12]: we should not underestimate the risk associated to armed conflicts, especially today. While data do not support any significant reduction in human belligerence, we now have both the connectivity and the technologies to annihilate the entire world population (finally touching that value $H$ we have not touched so far). In the antiquity, it was highly improbable for a conflict involving the Romans to spread all over the world, thus also affecting the populations of the Americas. A conflict in ancient Italy could surely influence the populations in Gaul, but it could have no impact on people living in Australia at that time. Now, a conflict in the Middle East could in principle start the next World War, and no one could feel completely safe.

It is important to stress that changes in the value of $H$ do not effect the conclusions we can draw from data, when variations do not modify the order of magnitude of $H$. In other terms, the estimates we get for $H = 7.2$ billion are qualitatively consistent to what we get for all values of $H$ in the range $[6, 9]$ billion. For example, for $H = 9$ billion, the conditional shadow mean of rescaled data, for $L = 145000$, becomes $4.0384 \times 10^7$. This value is surely larger than $3.6885 \times 10^7$, the one we find in Table IV, but it is still in the vicinity of $4 \times 10^7$, so that, from a qualitative point of view, our conclusions are still completely valid.

*Other shadow moments*

As we did for the mean, we can use the transformed distribution to compute all moments, given that, since the support of $Y$ is finite, all moments must be finite. However, because of the wide range of variation of $Y$, and its heavy-tailed behavior, one needs to be careful in the interpretation of these moments. As discussed in [38], the simple standard deviation (or the variance) becomes unreliable when the support of $Y$ allows for extremely fat tails (and one should prefer other measures of variability, like the mean absolute deviation). It goes without saying that for higher moments it is even worse.

Deriving these moments can be cumbersome, both analytically and numerically, but it is nevertheless possible. For example, if we are interested in the standard deviation of $Y$ for $L = 145000$, we get $3.08 \times 10^8$. This value is very large, three times the sample standard deviation, but finite. Other values are shown in Table V. In all cases, the shadow standard deviation is larger than the sample ones, as expected when dealing with fat-tails [13]. As we stressed that care must be used in interpreting these quantities.

Note that our dual approach can be generalized to all those cases in which an event can manifest very extreme values, without going to infinity because of some physical or economical constraint.

TABLE V: Shadow standard deviation, sample standard deviation and their ratio for different minimum thresholds. Rescaled data.

| Thresh.$\times 10^3$ | Shadow$\times 10^8$ | Sample$\times 10^8$ | Ratio |
|---|---|---|---|
| 50 | 2.4388 | 0.8574 | 2.8444 |
| 100 | 2.6934 | 0.9338 | 2.8843 |
| 145 | 3.0788 | 1.0120 | 3.0422 |
| 300 | 3.3911 | 1.1363 | 2.9843 |
| 500 | 3.8848 | 1.2774 | 3.0412 |
| 1000 | 4.6639 | 1.4885 | 3.1332 |

## VI. DEALING WITH MISSING AND IMPRECISE DATA

An effective way of checking the robustness of our estimates to the "quality and reliability" of data is to use resampling techniques, which allow us to deal with non-precise [43] and possibly missing data. We have performed three different experiments:

- Using the jackknife, we have created 100k samples, by randomly removing up to 10% of the observations lying above the 25k thresholds. In more than 99% of cases $\xi >> 1$. As can be expected, the shape parameter $\xi$ goes below the value 1 only if most of the observations we remove belong to the upper tail, like WW1 or WW2.
- Using the bootstrap with replacement, we have generated another set of 100k samples. As visible in Figure 13, $\xi \leq 1$ in a tiny minority of cases, less than 0.5%. All other estimates nicely distribute around the values of Table III. This is true for raw, rescaled and dual data.
- We tested for the effect of changes in such large events as WW2 owing to the inconsistencies we mentioned earlier

(by selecting across the various estimates shown in Table I) —the effect does not go beyond the smaller decimals.

We can conclude that our estimates are remarkably robust to imprecisions and missing values in the data[10].
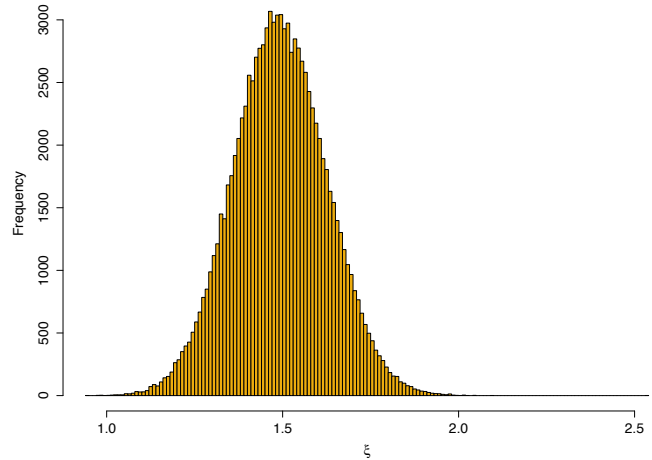


(a) Raw data.
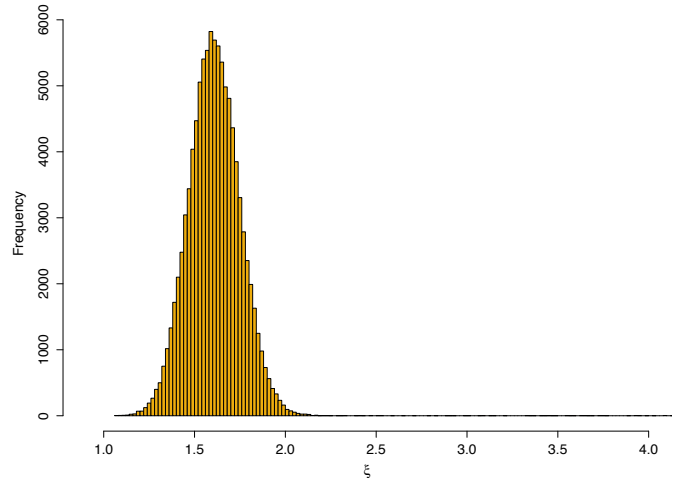


(b) Rescaled data.

Fig. 13: Distribution of the $\xi$ estimate (thresholds 25k and 145k casualties for raw and rescaled data respectively) according to 100k bootstrap samples.

## VII. FREQUENCY OF ARMED CONFLICTS

Can we say something about the frequency of armed conflicts over time? Can we observe some trend? In this section, we show that our data tend to support the findings of [18] and [32], contra [24], [30], that armed conflict are likely to follow a homogeneous Poisson process, especially if we focus on events with large casualties.

---

[10]That's why, when dealing with fat tails, one should always prefer $\xi$ to other other quantities like the sample mean [38]. And also use $\xi$ in corrections, as we propose here.

The good GPD approximation allows us to use a well-known model of extreme value theory: the Peaks-over-Threshold, or POT. According to this approach, excesses over a high threshold follow a Generalized Pareto distribution, as stated by the Pickands, Balkema and de Haan's Theorem [1], [28], and the number of excesses over time follows a homogeneous Poisson process [13]. If the last statement were verified for large armed conflicts, it would mean that no particular trend can be observed, i.e. that the propensity of humanity to generate big wars has neither decreased nor increased over time.

In order to avoid problems with the armed conflicts of antiquity and possible missing data, we here restrict our attention on all events who took place after 1500 AD, i.e. in the last 515 years. As we have shown in Section IV, missing data are unlikely to influence our estimates of the shape parameter $\xi$, but they surely have an impact on the number of observations in a given period. We do not want to state that we live in a more violent era, simply because we miss observations from the past.

If large events, those above the 25k threshold for raw data (or the 145k one for rescaled amounts), follow a homogeneous Poisson process, in the period 1500-2015AD, their inter-arrival times need to be exponentially distributed. Moreover, no time dependence should be identified among inter-arrival times, for example when plotting an autocorrelogram (ACF).

Figure 14 shows that both of these characteristics are satisfactorily observable in our data set.[11] This is clearly visible in the QQ-plot of Subfigure 14a, where most inter-arrival times tend to cluster on the diagonal.
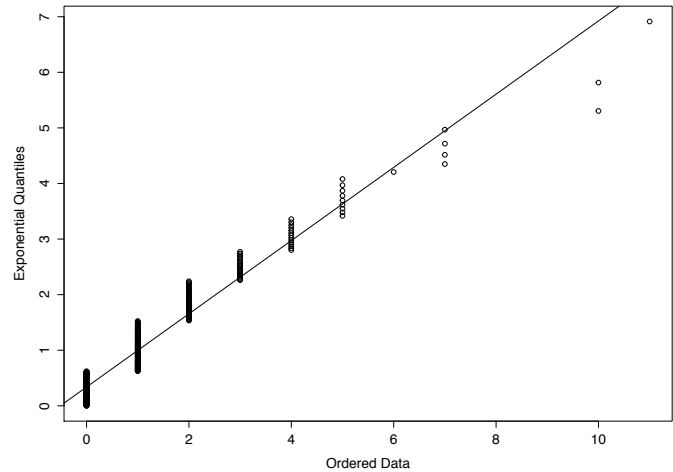
Another way to test if large armed conflicts follow a homogeneous Poisson process is by considering the number of events in non-overlapping intervals. For a Poisson process, given a certain number of events in a time interval, the numbers of events in non-overlapping subintervals follow a Multinomial distribution. If the Poisson process is homogeneous, then the Multinomial distribution is characterized by the same probability of falling in any of the sub-intervals, that is an equiprobability. It is not difficult to verify this with our data, and to see that we cannot reject the null hypothesis of equiprobable Multinomial distribution, over the period 1500-2015, in which 504 events took place, choosing a confidence level of 10%.

Once again, the homogeneous Poisson behavior is verified for raw, rescaled and dual data. Regarding the estimates of $\xi$, if we restrict our attention on the period 1500-2015, we find $1.4093$ for raw data, and $1.4653$ for rescaled amounts.

To conclude this section, no particular trend in the number of armed conflicts can be traced. We are not saying that this is not possible over smaller time windows, but at least in the last 500 years humanity has shown to be as violent as usual.

To conclude our paper, one may perhaps produce a convincing theory about better, more peaceful days ahead, but this cannot be stated on the basis of statistical analysis —this is not what the data allows us to say. Not very good news, we

[11] When considering inter-arrival times, we deal with integers, as we only record the year of each event.

have to admit.



(a) Exponential QQ-plot of gaps (notice that gaps are expressed in years, so that they are discrete quantities).
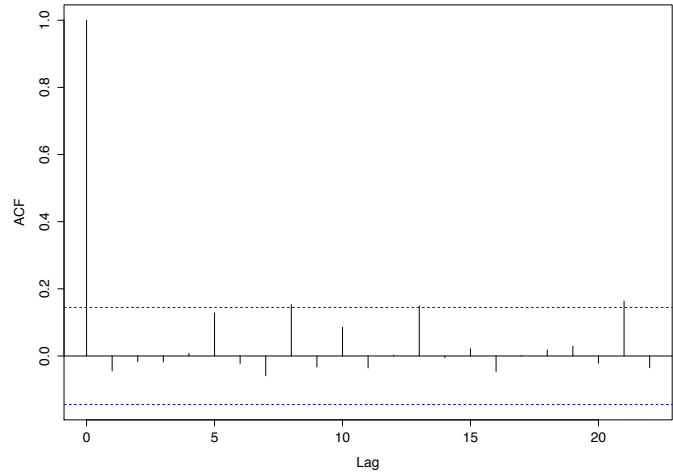


(b) ACF of gaps. Notice that the first lag has order 0.

Fig. 14: Two plots to verify whether armed conflicts (raw data) over the 25k casualties threshold follow a homogeneous Poisson Process

## References

[1] A. A. Balkema, L. De Haan (1974). Residual life time at great age. *Annals of Probability* 5, 792-804.

[2] A. A. Balkema, L. De Haan (1978). Limit distributions for order statistics I. *Theory of Probability & Its Applications* 23:1, 77-92.

[3] A. A. Balkema, L. De Haan (1979). Limit distributions for order statistics II. *Theory of Probability & Its Applications* 23:2, 341-358.

[4] BBC (2012). In our times: The An Lushan Rebellion. http://www.bbc.co.uk/programmes/b01by8ms.

[5] D. Berlinski (2009). *The Devil's Delusion: Atheism and its Scientific Pretensions*. Basic Books.

[6] L.-E. Cederman (2003). Modeling the size of wars: from billiard balls to sandpiles. *The American Political Science Review* 97, 135-150.

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

[8] A. Clauset, R. Woodard (2013). Estimating the historical and future probabilities of large terrorist events. *Annals of Applied Statistics* 7, 1838-1865.

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

[10] U. Einmahl, D. Li (2005). Some results on two-sided LIL behavior. *Annals of Probability* 33, 1601-1624.

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

[12] R. Epstein (2011). Book Review: The Better Angels of Our Nature: Why Violence Has Declined. *Scientific American*, October 7.

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

[14] J. A. Friedman (2014). Using power laws to estimate conflict size. *Journal of Conflict Resolution*, 1-26.

[15] J.L. Goldstein (2011). *Winning the war on war: the decline of armed conflict worldwide*. Penguin.

[16] J. Gray (2015). Steven Pinker is wrong about violence and war. *The Guardian*, Friday 13 March. http://www.theguardian.com/books/2015/mar/13/john-_gray-_steven-_pinker-_wrong-_violence-_war-_declining

[17] E. J. Gumbel (1958). *Statistics of Extremes*. Cambridge University Press.

[18] B. Hayek (2002). Statistics of Deadly Quarrels. *American Scientist* 90, 10-14.

[19] A. Horne (2006). *A Savage War of Peace*. New York Review Books Classics

[20] N. Inkester, ed. (2015). *Armed Conflict Survey 2015*. IISS.

[21] C. Kleiber, S. Kotz (2003). *Statistical Size Distribution in Economics and Actuarial Sciences*. Wiley.

[22] 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.

[23] R. Maronna, R.D. Martin, V. Yohai (2006). *Robust Statistics - Theory and Methods*. Wiley.

[24] J. E. Mueller (1989). *Retreat from Doomsday: the obsolescence of major war*. Basic Books.

[25] Necrometrics (2015). http://necrometrics.com.

[26] J. Nešlehová, P. Embrechts, V. Chavez-Demoulin (2006). Infinite mean models and the LDA for operational risk. *Journal of Operational Risk* 1, 3-25.

[27] R. Norton-Taylor (2015). Global armed conflicts becoming more deadly, major study finds. *The Guardian*, Wednesday 20 May. http://www.theguardian.com/world/2015/may/20/armed-_conflict-_deaths-_increase-_syria-_iraq-_afghanistan-_yemen

[28] J. Pickands (1975). Statistical inference using extreme order statistics. *the Annals of Statistics*, 119-131.

[29] C. Phillips, A. Axelrod (2004). *Encyclopedia of Wars*, 3-Volume Set. Infobase Pub.

[30] S. Pinker (2011). *The Better Angels of Our Nature: Why Violence Has Declined*. Viking Adult.

[31] L. F. Richardson (1948). Variation of the Frequency of Fatal Quarrels with Magnitude. *Journal of the American Statistical Association* 43, 523-46.

[32] L. F. Richardson (1960). *Statistics of Deadly Quarrels*. Chicago Quadrangle Book.

[33] L. A. Cox, M. R. Greenberg, eds. (2011). Special Issue: Advances in Terrorism Risk Analysis. *Risk Analysis*, online ISSN 1539-6924.

[34] C.Robert, G. Casella (2005). Monte Carlo Statistical Methods. Springer.

[35] A. Scharpf, G. Schneider, A. Noh, A. Clauset (2014). Forecasting of the risk of extreme massacres in Syria. *European Review of International Studies* 1, 50-68.

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

[37] M. Spagat, A. Mack, T. Cooper, J. Kreutz (2009). Estimating War Deaths: An Arena of Contestation. *Journal of Conflict Resolution* 53, 934-50.

[38] N. N. Taleb (2015). *Silent Risk: Lectures on Probability*, Vol 1. SSRN.

[39] United Nations - Department of Economic and Social Affairs (1999). *The World at Six Billion*. UN Press.

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

[41] J. van der Wijk (1939). Inkomens- en Vermogensverdeling. *Publication of thÂŐe Nederlandsch Economisch Instituut* 26.

[42] J.-A. Villasenor-Alva, E. Gonzalez-Estrada (2009). A bootstrap goodness of fit test for the generalized Pareto distribution. *Computational Statistics and Data Analysis* 53, 3835-3841.

[43] R. Viertl (1995). *Statistical Methods for Non-Precise Data*. CRC Press.

[44] P. Wallensteen, M. Sollenberg (2001). Armed Conflict 1989-2000. *Journal of Peace Research* 38, 629-644.

[45] M. White (2011). *The Great Big Book of Horrible Things*. W.W. Norton and Company.

[46] M. White (2013). *Atrocities*. W.W. Norton and Company.