# Catastrophic Risk, Contagion and Public Policy: Could Data Have Revealed the Risk of COVID-19?

Hamid Mohtadi[a] and John A. Schwendel[b]

## Abstract

When data include extremes, expectations, and thus rational policymaking, may fail. Applying Extreme Value Theory to *pre-COVID* mortality data on respiratory contagions we ask: Did the US fail to predict a COVID-19-scale event because its catastrophic scale "surprised" policymakers where prior data did not help, or because policymakers ignored the predictions of that data? The answer depends on how the US and World data are related: Comparing baseline predictions with counterfactuals where US data randomly admits global contagions, estimated distributions–robust to nonparametric Hill estimators–produce infinite *and* finite expectations under different conditions. The latter accurately predicts current COVID-era mortality.

Keywords: Extreme Value Theory, Hill estimator, COVID-19, bootstrap, public policy, health policy
JEL Classification Codes: D81, H12, C46, I18

[a]Corresponding author. Professor of Economics, University of Wisconsin-Milwaukee, mohtadi@uwm.edu and Affiliate Research Professor of Applied Economics, University of Minnesota, mohtadi@umn.edu (ORCID ID: https://orcid.org/0000-0002-9497-787X)
[b]Lecturer of Economics, Indiana University, joschwen@iu.edu (ORCID ID: https://orcid.org/0000-0003-2399-3589)

# Catastrophic Risk, Contagion and Public Policy: Could Data Have Revealed the Risk of COVID-19?

"We've got to be more willing to consider observations made in other countries dealing with [the coronavirus]"[1]

## 1 Introduction

The striking contrast between the tame US mortality rates from global respiratory contagions over the past 40 years relative to the rest of the world (Table 1) and its dramatic experience with COVID-19, where mortality has surpassed all other countries at the time of this writing and exceed its own 1918 flu pandemic (Figure 1), point to the classic case of a rare and extreme event. The US public's response to this event remained muted for some time after it began. A recent paper by [Pike, et al. (2020)] analyzed such public under-reaction to pandemics in a fascinating natural experiment in the context of the 2014 Eboli outbreak. The public's attitude towards pandemics remained surprisingly unchanged before and after the event. Similarly, [Viscusi 2009] finds that an age old dramatic event like 9/11 is more highly valued by the public than a deadly pandemic or a catastrophe due to climate change[2] The study by [Pike, et al. (2020)] contained a warning to policymakers: "The present COVID-19 pandemic drives home the importance of these behavioral results and the *necessity of taking stronger preventive measures* than the public might consider necessary." (ibid; italics added for emphasis).

This is where our query begins, i.e., policymakers' response to catastrophic events in general and health epidemics in particular. Here the question arises: if the public's reaction, based on behavioral patterns rooted in fear or other departures from "rationality," leads to anomalies as described above, does some form of myopia also characterize the policymakers'

[1] James E.K. Hildreth, chief executive of Meharry Medical College. Quoted in Washington Post, August 4, 2021

[2] [Viscusi 2009] found that the public was twice as concerned about terrorism risk than about other forms of risk.

early under-reaction to a COVID-19 scale event, or their failure to anticipate an event of such scale despite some anecdotal evidence of early warnings[3]? While we do not know the direct answer to this question, if we assume that government decisions are "rational" and driven by data, then understanding the nature and pattern of the data available to policymakers becomes crucial. In this paper, we will study the distributional properties of pre-COVID-19 mortality data from all respiratory infections both in the US and globally, to assess whether policymakers could have predicted a coming COVID-19 scale event.[4] Our answer is surprisingly "yes" under some reasonable assumptions and "no" under others.[5]

In general, because extreme events belong to the tail of heavy tailed distributions, the key question for policymaking is whether the past offers any help in predicting such extreme events for rational policymaking. Statistically, this is equivalent to asking whether the policymakers' early under-reaction to the COVID-19 pandemic was because an event of this scale was so extreme as to defy the existence of a finite mean or expectation and thus remain unpredictable, amounting to a surprise, or whether expectations from the distributional fit to the data did exist (despite heavy tails) and were well behaved (had finite variance), in which case data could have allowed policymakers to have been better prepared for a COVID-19 scale outlier.

We answer this question by comparing, contrasting, and lastly synthesizing US and global mortality data on respiratory contagions pre-COVID-19. In the synthesis stage, we develop a counterfactual framework that accounts for outcomes on the scale of other nations, helping to

[3]See for example Bill Gates speach at TED talk in 2015. https://www.ted.com/talks/bill_gates_the_next_outbreak_we_re_not_ready?language=en
and President Obama speeach, in 2014:
https://www.cnn.com/videos/politics/2020/04/10/barack-obama-2014-pandemic-comments-sot-ctn-vpx.cnn

[4]In an interesting paper, [Viscusi 2020] uses the value of statistical life (VSL) to show that the economic cost morbidity from COVID-19 is more costly than from mortality, partly due to the overrepresentation of the old in COVID-related mortality. Since our focus in the policymakers ability to predict a COVID-19 event *before* it actually happened, such a considerations, while important, is not as closely related to the question at hand.

[5]While this answer may point to other (non-data) related factors (e.g., political, behavioral), for the lack of an early government response to COVID-19, studying these issues is certainly impotrant but beyond the scope of our investigation.

form expectations in the face of random global transmission of unknown contagions. When we do this, the predicted US mortality becomes far higher than the US historical average and, when expectation do exist, it approximates US mortality from COVID-19 at the time of this writing, offering affirmation of the above quote. Our paper provides a clear path for how data can be properly utilized to both diagnose certain types of policy failure and inform policymaking of how extreme events may be better anticipated.

To do this, we adopt methods from the statistics of Extreme Value Theory (EVT) to examine what they imply for policymaking. Making the determination of whether expectations exist and are well defined is important in assessing the "rationality" of public policy based on the ability to form proper expectations. While there have been many investigations of heavy tailed distributions in the cases of both natural and social phenomena (section 2), we are aware of a only handful of papers within two lines of research that explore heavy tails explicitly in relation to public policy: one on climate change ([Weitzman, 2009]; [Weitzman, 2011]; [Pindyck 2011];[Costello et al. (2010)]), and the other on terrorism ([Mohtadi & Weber 2021]). As for pandemics and heavy tailed distributions, to our knowledge only two other recent papers have studied the heavy tail properties of global contagions; one has a historical focus covering ancient times up to the present ([Cirillo & Taleb (2020)]) and the other focuses on superspreaders of COVD-19 ([Wong & Collins (2020)]). Neither paper explicitly explores the *policy* implications of fat tail properties of contagions. We aim to fill this gap.

Returning to the data and the context of respiratory contagions, the question of whether policymakers can form proper expectations to predict the emergence of an extreme event, or whether such events are too extreme to allow for finite expectations to be found, depends on the structure of the probability density functions which in turn of course depends on the structure of the data. The two datasets that we use for our comparison, contrast, and synthesis, as previously stated, are the US domestic data from the Centers for Disease Control (CDC) (1980-2017) and the global data from Global Burden of Diseases, Injuries,

and Risk Factors Study (GBD) by the Institute for Health Metrics (IHME) at the University of Washington covering 195 countries(1990-2017)[6]. This excludes COVID-19 both because it is still in progress at the time of this writing, and also intentionally, as it permits "out of sample predictions," allowing us to see if our prediction (should expectations exist) produces mortality rates in the neighborhood of what we actually observe from the COVID-19 event. A glimpse of our results indicates predictions of mortality rates that are very close to the present mortality from COVID-19 under some conditions.

We first examine the distributional properties of each dataset separately and explore the possible existence of fat tails for each, as detailed in section 3.1. One may ask, if the goal is to estimate US expected mortality rates, assuming they exist (yet to be established), why consider global mortality rates at all? How would using global data inform US policy? One answer is that using only the US' own past history may ignore potential US vulnerability from a highly fatal respiratory contagion that occurred elsewhere. For example, all four instances of unexpected and severe respiratory contagions over the past two decades, SARS, N1H1, MERS and COVID-19, began elsewhere before being transmitted to the US. But while the first three led to US mortality rates much lower than some other countries, the fourth has not. In this case, by using only US data, policymakers would have been lulled into a false sense of security due to their inattention to tail behavior represented by the extremes in the distribution of global mortality rates, leaving them unprepared to deal with COVID-19. Naturally, the converse is also unsatisfactory: Using only global mortality rates as the basis for forming expectations in the US, ignores US specificity and context.[7]

[6]Below are the links to the public data sources used in this paper. For CDC data variables were extracted from the following source: https://www.cdc.gov/nchs/hus/contents2018.htm#Table_005

For GBD data, the general source is:
http://ghdx.healthdata.org/gbd-2017 and the specific link is:
https://gbd2017.healthdata.org/gbd-search?params=gbd-api-2017-permalink/86b42a8f452fedd2d8e9d859cee17000

GBD data covers 205 countries. Our covering of 195 is necessitated by the lack of population and some other data for some small countries and Island nations that make up the remaining 10 countries.

[7]While the US and the global mortality data will both include instances of severe illnesses regardless of their origin, the additional variation introduced by the global data will shed light on the potential instances of extreme events which, as we will see, would be missed out otherwise. Section 3.2 provides additional

One logical solution is to synthesize the two and explore the tail properties of the resulting density function. But how should this be done? Since US policymakers cannot know when global mortality rates may be relevant to the US for any given year, we develop a framework in which we *randomly* allow global mortality rates to enter into the US data. This is akin to asking the following counterfactual: "What if for a given year, or a number of years, the US was as unprepared for a pandemic as other countries were for any new or known yearly infection?" Given US COVID-19 mortality rates relative to the rest of the world this question is not just a hypothetical exercise, but one that may inform future policy preparation and planning. *How* this synthesis is to be done, the technical issues that it raises, and what it means for tail behavior, expectation formation and policymaking, are at the core of this paper and presented in subsequent pages.

In what follows, section 2 presents the theory and method, including discussion of the nonparametric Hill estimator of the tail and the EVT distributions that are used for the parametric estimates of the tail, section 3 presents the actual tail estimates, predicts mortalities, and discusses the policy challenge when a finite mean does not exist, proposing a path forward, section 4 provides concluding remarks.

# 2 Theory and Method

To examine tail behavior and estimate the tail of a distribution, one approach is to calculate the non-parametric Hill estimator ([Hill (1975)]; [Embrechts, Klüppelberg & Mikosch (2003)]) which is agnostic to the choice of distributions. Another is to parametrically estimate the tail by applying any of three classes of distributions; simple Pareto (and the related Power Law distributions), Generalized Pareto distribution (GPD), and the family of Generalized Extreme Value Distributions (EVD). We will follow *both* the parametric and the non-parametric approaches, using each approach as a robustness check on the other. For

discussion of these important considerations.

greater robustness, we use all three of the parametric distributions to estimate the tail and, by implication, the expected mortality rates. For parametric distributions, however, since mathematical linkages between the three classes of distributions exist, with implications for our results, all three classes will be described below prior to the estimation.

Examples of heavy-tailed distributions abound. They have included natural disasters (e.g., earthquakes) ([Newman 2005 ]), wealth ([Jones 2015]), finance ([Bali 2007]) and terrorism ([Mohtadi 2006], [Mohtadi 2009], [Mohtadi & Murshid 2009]) as well as events characterized by a subset of heavy-tailed distributions, Power Law and Pareto distributions, that are scale-independent and exhibit fractal structures. These range from biology ([West, Brown & Enquist 1997], [Ravasz et al. 2002]) to firm structure ([Stanley et al. 1996]), city size ([Mori, Smith & Wen-Tai 2020], [Sarabia & Prieto 2009]), and wars and terrorism ([Spagat, Johnson and Weezel 2018]).

## 2.1 Heavy-tailed distribution classes and their links

In the Pareto distribution which has the form of a power law, a continuous random variable $X$ is 'fat-tailed' if its survival function $P(X \geq x)$ decays according to a power law $x^{-\alpha}$ as $x$ increases. The size of $\alpha$ also determines how fat the tail is. This in turn determines the existence of moments. In particular, the first moment exists, $E[X] < \infty$, if and only if $\alpha > 1$, and the second moment exists, $E[X^2] < \infty$, if and only if $\alpha > 2$. When $\alpha \leq 1$, the tail is so heavy and the decay rate so slow that no finite mean exists and when $\alpha \leq 2$, it is slow enough that the mean is infinitely imprecisely estimated. The classical Pareto distribution has a cumulative distribution function of the form,

$$F_X(x) = 1 - (\frac{x}{\sigma})^{-a} \quad x \geq \sigma > 0 \tag{1}$$

where $\sigma$ is a known scale parameter, representing the lowest value of $x$. While the fractal property and the scale independence of the Power Law and Pareto distribution are well

known, thus making them attractive for estimation of potentially heavy tailed phenomena, the two other classes of heavy-tailed distributions, GPD and GEV, capture other features of data and each are appropriate in certain data environments. GPD is related to the classical Pareto distribution in that the latter is known to be the asymptotic form of GPD ([Gnedenko (1943)]). GPD includes a shape parameter, $\xi$, which is the inverse of Pareto's decay rate of $\alpha$ and is critical for deciding the existence of fat tails (see below). It also adds a location parameter, $\mu$, that captures threshold values above which tail behavior is expected, as well as a scale parameter, $\sigma$, that indicates roughly the expected range of the data. The cumulative distribution function for GPD is of the form,

$$F_{X,\mu,\sigma,\xi}(x) = \left\{ \begin{array}{c} 1 - [1 + \frac{\xi(x-\mu)}{\sigma}]^{-\frac{1}{\xi}} \text{ for } \xi \neq 0 \\ 1 - e^{(-\frac{x-\mu}{\sigma})} \text{ for } \xi = 0 \end{array} \right\} \tag{2}$$

A second representation of GPD is in the form of the Peak over Threshold model (POT). In particular, for a random variable $X$, the distribution of the standardized excesses over threshold approaches GPD in the limit. To see this, suppose $x_F$ is defined as the right end point of the distribution, whether finite or infinite. Then, as the threshold approaches this value, the probability of excess approaches GPD:

$$P\left(\frac{X-\mu}{\sigma} < x | X > x\right) \rightarrow F_{X,\xi}(x) \text{ as } \mu \rightarrow x_F \tag{3}$$

where,

$$F_{X,\xi}(x) = \left\{ \begin{array}{c} 1 - (1+\xi]^{-\frac{1}{\xi}} \text{ for } \xi \neq 0 \\ 1 - e^{-x} \text{ for } \xi = 0 \end{array} \right\} \tag{4}$$

The most general class of fat tailed distributions is GEV. This distribution is on one hand related to the GPD distribution (see below) and, on the other, to an important theorem, the Fisher-Tippet Theory ([Fisher & Tippett (1928)]) of Extrema, otherwise known as Extreme Value Theory (EVT). To understand this, let $X_1$, $X_2$,...$X_n$ be independently and identically

distributed (iid) random variables whose distribution is $F$. Let $M_n = \max(X_1, X_2,...X_n)$. Then, *if* there exists a sequence of constants $\{a_n > 0\}$ and $\{b_n\}$ such that,

$$P\left(\frac{M_n - b_n}{a_n} < x\right) \rightarrow G(x) \text{ as } n \rightarrow \infty \tag{5}$$

for a non-degenerate limiting distribution, $G$, it follows that the distribution $F$ is in the (maximum) domain of attraction of $G$ where $G$ is a GEV distribution. By the Fisher-Tippet Theory, the GEV distribution is given by ([Coles et al. (2001)])[8]

$$G_{X,\xi,\mu,\sigma}(x) = \begin{Bmatrix} e^{-(1+\xi\frac{x-\mu}{\sigma})^{-\frac{1}{\xi}}} \text{ for } \xi \neq 0 \\ -e^{-\frac{x-\mu}{\sigma}} \text{ for } \xi = 0 \end{Bmatrix} \tag{6}$$

This class of heavy tailed distributions is particularly suited to the block maxima structure of our global data as we will see shortly. It is also the most general of the three, and embodies several different distributions; the extremely heavy-tail Frechet distribution ($\xi > 0$), the light-tail Gumble distribution ($\xi = 0$) and the bounded-tail Weibull distribution ($\xi < 0$). Importantly the value of shape parameter, $\xi$, is the same in both GPD and GEV distributions when limits exist ([Németh & Zempléni (2020)]). It turns out that this distribution also represents some forms of Gamma and Cauchy distributions.[9]

In what follows, we will estimate these distributions for each of our datasets as well as for their synthesis. The latter is critical to our analysis. It was briefly described in the Introduction but will be described much more fully later. As previously stated, for robustness, we will also employ a nonparametric method for the various order statistics of the data to estimate the tail of the distributions. This latter approach will be described as we proceed with the estimation procedure.

[8]From above, $P(M_n < x) \rightarrow G(\frac{z-b_n}{a_n}) \equiv G^*(z)$ where $G^*$ is another distrbution also in the GEV family. In practice, since $\mu$ and $\sigma$ are estimated, the difference in parameters of $G$ and $G^*$ make little difference. See Coles et al. (2001, p. 48-49 for greater detail).

[9]Note that for classic Pareto distribution the mean exists only when $\alpha > 1$ while for GPD and GEV distributions, mean exists only when $\xi < 1$. This because a faster (slower) decay for Pareto implies a lighter (heavier) tail for GPD and GEV distributions. In fact $\alpha = 1/\xi$ (e.g., see [Cirillo & Taleb (2020)]).

# 3 Estimation

## 3.1 US and the World Data Separated

In this subsection, we first consider a world where the US estimates of mortality from respiratory contagions are based solely on its own history and go about estimating the tail of the corresponding distribution. For comparison, we will also estimate the tail of the distributions associated with the global data. In subsection 3.2 we will then integrate the two datasets, following a somewhat elaborate randomization approach that is aimed at mimicking rational policymakers' view of how to prepare for an unknown event of unknown severity. These types of considerations are especially salient given the potential for global infections which arrive unexpectedly and bring with them highly variable outcomes. In each subsection, we will conduct both a parametric and a nonparametric estimate of the tail.

### 3.1.1 Parametric Estimates of the Tail

Table 1 presents the US data for two different periods, 1980-2017 and 1990-2017, the latter to conform to the world data which is only available from 1990 to 2017. The world data is presented for three different measures; country average per year, world uniform average per year, and world's most affected country per year, all per 100,000 inhabitants.[10] A simple glance at the data confirms that US mortality rate is far more modest that any of the three global measures. Yet, when fitted with any of the three parametric distributions discussed, not only does the US data not exhibit heavy tails, but even the considerably higher values from the world data (in any of its measures) fail to produce heavy tails (Table 2). This can be seen either from the high decay rate of the Pareto distribution exceeding 1, or the shape

[10] Specifically, let, $N_{it}$ and $P_{it}$ represent the number of deaths and total population, respectively, per country $i$ in year $t$ . For country level average mortality we calculate, $100,000 \times \sum_{i=1}^{195} N_{it}/P_{it}$ . For the uniform average mortality, we calculate $100,000 \times (\sum_{i=1}^{195} N_{it})/P_{wt}$ where $P_{wt}$ is the world (sum of 195 countries) population for year $t$. For the most affected country we calculate, $100,000 \times (N/P)_{\mu t}$ where $(N/P)_{\mu t} = \max[(N/P)_{1t,}(N/P)_{2t}...(N/P)_{195,t}]$ captures the highest fatality ratio for year $t$ belonging to country $\mu$.

parameter estimates of the GEV and GPD distributions far below 1, with the confidence intervals crossing zero into negative range. For GEV, this indicates a light tail associated with the Gumble distribution when $\xi = 0$, or a bounded tail associated with the Weibull distribution when $\xi < 0$. Note that the distributions for the US mortality data show markedly smaller *location* parameters than those for the global mortality rates. This is especially so in the case of the "most affected country per year" from the global data where the mean value of the location parameter is 171.7, as compared to mean US mortality rates of 18.7 or 24.8, depending on which sample period is considered. Later, we learn of large differences in the implied US mortality rates when the global data is randomly combined with the CDC data as will be described in detail**.**

This absence of heavy tails for either the US or global data, together with the vastly different estimates of the location parameters between the US and globally most affected country suggests that when the data are separated, distinct *clusters* in the distribution of the mortality data exist so that, statistically speaking, the behavior of the US data does not extrapolate to the global data and hence no heavy tails emerge. In fact, at a first glance one might argue that mortality data from many of the past global contagions such as the SARS family *are* already present in the US data and hence there is no need to consider global mortality data. But this approach may lull policymakers into a false sense of security, given the tame US mortality data form the past contagions. The COVID-19 event has demonstrated the need to model contagions based on unexpectedly large mortality. As we will show, accounting for such outcomes is only possible when we integrate the US and the global data. *How* this is done and *what* it implies for policy myopia or rational policymaking is the core of our paper. This will be studied in Section 3.2.

### 3.1.2 Nonparametric Estimate of the Tail: The Hill Estimator

For robustness, we now employ a nonparametric approach of approximating the potential tail of the data using the Hill estimator ([Hill (1975)]; [Embrechts, Klüppelberg & Mikosch

(2003)]). Calculating the Hill estimator requires order statistics. Tailored to the structure of our data, we define two sets of order statistics. The first set is defined for a single observation per year. For the US data this is obvious. For the world data single observation per year consists of any of the three different measures discussed above and in footnote 4. By contrast, the second set has a panel format which we will discussed shortly.

**Single yearly observations:** For a set of randomly distributed variable, $X_1..., X_T$ over the sample period $T$ corresponding to mortality per year, let $X_T^{(k)} \leq ... \leq X_T^{(1)}$ be order statistic of order $k$. Then the Hill estimator of the tail for $k$ order statistic is given by,

$$\hat{\xi}_k = \frac{1}{k} \sum_{i=1}^{k} \ln(X_T^{(i)}) - \ln(X_T^{(k)}),\ 1 < k \leq T \tag{7}$$

Figures 2-6 show the Hill estimators of the tail as a function of variations in order statistics $k$: Figure 2 and 3 represent US mortality rates starting either from 1980 or from 1990, the latter for compatibility with the global data as previously mentioned; Figures 4-6 present the Hill estimators for the world data consisting of world mortality rate averaged over all countries (Figure 4), uniform world average mortality, i.e., total world mortality relative to total world population (Figure 6), and world maximally affected country (Figure 6). None of the figures exhibit a heavy tail pattern as seen by the fact that the Hill estimators of the tail fall far below 1. This corroborates the findings of Table 1.

**Panel approach – k observations per year:** We now calculate a Hill estimator of the tail associated with the second type of order statistics which exploits the *panel* nature of the global mortality data. In this form, instead of a single observation for each year $t$ corresponding to either global average or global most affected country, as we have done above, we map observations from the $k^{th}$ most affected country across the world (highest mortality rate), for each year $t$, and calculate a Hill estimator over the sample period, $T$. To formalize let, $X_{1,1}, ..X_{n,t}.., X_{N,T}$ be a set of randomly distributed variables representing

mortality across $N$ countries over $T$ periods. Then for each given period $t$ we can write:

$$X_{N,t}^{(k)} \leq ... \leq X_{N,t}^{(1)} \; | t = 1...T \tag{8}$$

where $X_{N,t}^{(k)}$ is the $k^{th}$ order across the $N$ countries in year $t$. Note the contrast with the previous definition: The generalization, from a single value associated with either the most affected or the average of all affected countries each year, to now $k^{th}$ order of the same, implies that each year $t$ now contributes $k$ observations to the calculation of Hill estimator instead of just 1. This introduces a second summation to the traditional Hill estimator:

$$\hat{\xi}_k = \frac{1}{kT} \sum_{t=1}^{T} \sum_{i=1}^{k} \ln(X_{N,t}^{(i)}) - \ln(X_{NT}^{(k)}), \; 1 < k \leq N \tag{9}$$

where $X_{NT}^{(k)}$ denotes the $k^{th}$ order among a panel of $NT$ observations[11]. Figure 7 which presents the Hill estimator for the panel data finds that for values of $k \gtrapprox 125$, the tail becomes so heavy ($\xi \gtrapprox 1$) that the underlying mean fails to exist. Intuitively, when we include highly affected countries (low $k$ values) with countries with very low mortality rates ($k \gtrapprox 125$), the variations in mortality rates are so spread out that any resulting distribution which embodies *all* such cases is "flattened out," leading to a decay rate so slow that a finite mean fails to exist. Most interestingly, this result holds up for the (parametric) Pareto distribution fit (Figure 8) with precisely the same support as above. Here, at $k \gtrapprox 125$ the value of the Pareto's $\alpha$ parameter cross 1 downwardly ($\alpha \lessapprox 1$), implying really fat tails and infinite means. To see if it is the time variations or the variations across countries that are responsible for this finding, a yearly $k$ order statistics is also calculated for each year of the sample. Results, reported in the appendix, confirm that cross country variations are the source of flattening and of the resulting fat tails since in all years of the sample one still finds the value of Pareto's $\alpha$ to cross-over from 1 at $k \simeq 125$.

[11] To elaborate note that $X_{N,t}^{(k)}$ corresponds to the $k^{th}$ lowest mortality rate in a given year for a given $k$ order statistic. Since there are $T$ observations added for each iteration of the order, $X_{NT}^{(k)}$ is the minimum of all those values for some year $t_o \in T$

## 3.2 Combining US and World Data: Creating Counterfactuals

We stated in the Introduction that the additional variations from the global data will shed light on the potential instances of extreme events. But there is also a deeper reason for combining the two datasets. The current COVID-19 pandemic and previous infectious diseases such as SARS, MERS and H1N1 have demonstrated the importance of planning for such outbreaks. But by definition severe outbreaks are unexpected. If US policymakers focus on annual US data, their expectations for mortalities, costs, and infrastructure preparedness will be based on diseases for which we are already well prepared, for example the seasonal flu. While outcomes related to the flu do vary year to year, our experience with this level of variation has tuned our infrastructure to respond well to such changes. If policymakers were to plan for an unexpected event, one strategy might be to act *as if* we did not have this preparation in place. This is, of course, a counterfactual not widely available in US data. It is this fact that motivates the need to look to global data in some form. However, while a focus on global data which includes high mortality rates for many poor nations with poor infrastructure and preparedness, *could* be illustrative of a lack of preparedness in the US for a future extreme event, pure focus on global data may cause a cluster at the extremes which, without considering the mild US mortality cases, would actually underestimate the large variations that are necessary, if not sufficient, for fat tails and are at the root of such statistical behavior.

Thus, we must look for a way to combine the data. We thus opt for creating a counterfactual vector of observations that effectively asks; "what if randomly, for some number of years, the US faced a contagion which its infrastructure and policy was as unprepared for as other high-impacted nations are today for existing outbreaks?" We answer this question by randomly combining the US mortality data with the global mortality data and then estimating the tail of the data. We consider this to be a reasonable way to model a pandemic because no matter how well suited our strategies and infrastructure may be for known diseases, pandemics have shown that these same tools may not be well suited for unknown

events. As such the best way to estimate the outcomes of unknown events for which we are unprepared is to include some data from countries which are unprepared for known events.

We approach this in two ways, parametrically, by fitting one of our candidate distributions to the data; and non-parametrically, by calculating the Hill estimator of $\hat{\xi}_k$ for the order distribution of the combined data. Both approaches involve randomization but with minor differences. We begin with the parametric approach which is simpler in both description and execution. We then move to describe the nonparametric approach.

### 3.2.1 Parametric tail estimates: randomized data mixing

Let $R =$ the number of randomly selected years for which the US mortality rate is replaced with the mortality rate of the most affected country for *that* year. We begin with randomly selecting a single year, i.e., $R = 1$. This is effectively swapping a value from columns 5-7 in table 1 into column 4, supplanting that observation. We then fit our three candidate distributions for this "single-year replaced" dataset. We next simulate the process 1000 times corresponding to other randomly chosen single years and record the median value for the parameters of the estimated distribution from the 1000 runs. In the second step, we choose $R = 2$ corresponding to *two* random years (not necessarily consecutive) to perform the same exercise. We continue this approach by increasing $R$. While the choice of 1000 runs may seem overdone for a low number of replacement years, it addresses the challenge of the large number permutations that arises (in tens of millions) when the number of years replaced increases. Figures 9-11 depict variations in the shape parameter as a function of the number of years of replacement. Note that Pareto decay passes downwardly crossing $\alpha = 1$, at about $R = 7$, so that $\alpha \precsim 1$ when about 7+ years of global data randomly replace US data (Fig. 9). The shape parameters for GPD and GEV indicate a *mid-range* of replacement years in which $\xi \succsim 1$.This occurs, for the case of the GPD distribution, from about of 4 to about 12 years of data replacement, and for the case of the GEV distribution, from about 2 to about 15 years of data replacement (Figs. 10 and 11). While this mid range occurrence

of infinite means requires some explanation (below), the common feature of all the three distributions is that for some level of data integration that indicates the global severity of a contagion, expectations fail. This means that a surprise will occur for which policymakers will be left unprepared. Later, we will calculate the predicted US death tolls from various random mixtures for the various years of replacement in the range where means still exit.

An intuition for the fact that GPD and GEV distributions have their very heavy tails occurring in the mid-range of the number of replacement years, $R$, may be that the mid-range of $R$ produces the highest spread in the data by mixing sufficient observations from a historically low mortality country (the US) with the much higher mortality rate associated with the globally most affected country. The GPD and GEV distributions can better capture this feature. This higher sensitivity of these distributions to the variations in the data may also be at play for their wide confidence band of $\hat{\xi}$ estimates in the mid range of R (figs. 9 and 10): In this range, the immense number of permutations of the mixed dataset is in its highest range, leading to a large variance in our limited random sample of 1000.

We may also ask if there is a relation between the fat tails of the global panel data from the nonparametric Hill estimator for $k = 125+$ (Fig. 7) and the fat tails of the parametric estimates of the integrated US-global data. The answer lies in the source of the fat tails. Specifically, the high spread that produces the fat tails in the integrated panel data is due to combining a low-mortality country (the US) with a very high-mortality country (the most affected country). Similarly , the high spread that produces the fat tails in the parametric estimates is caused by mixing the low mortality group ($k \geq 125$) with the high mortality group ($k < 125$).

### 3.2.2 Nonparametric Hill estimator: Generalizing randomization scheme

In this approach, for $k = 1 = R = 1$ in the combined data, random replacement amounts to replacing the US mortality rate for any single randomly selected year, with the mortality rate of the most affected country for that year, as described above. However, Instead of estimating

a parametric distribution, the non-parametric approach involves simply calculating the Hill estimators to examine any potential tails in the data. As in the above scheme, the Hill estimator is calculated for 1000 bootstraps. The values of the Hill estimator and its standard errors are then averaged across the bootstraps. Once $R$ and $k$ exceed one, variations in two distinct dimensions have significant effects on the results as we will describe shortly: However, when we consider the fact that replacement years need not be consecutive and can be spread in numerous different patterns over 28 years, the number of permutations in this approach increases dramatically. While this fact limits the consideration of *all* the possible configurations, we can still carry out 1000 random draws from the much larger set of possibilities. Then the values of the Hill estimator and the standard errors are averaged across those thousand estimations.

For economy of space, results are reported in Figures 12-15 for the entire range of order statistics $k = 1 - 28$, but select values of $R = 2, 3, 13$ and 14 (though all are available from the authors.) Note that in all cases, the Hill estimator of the tail rises precisely at the point where US' own most affected year first starts combining with the "globally most affected country" for each of the $R$ random years. For example, if $R = 2$ and $k = 3$, the Hill estimator is calculated using 2 observations from the "globally most affected county" and a single observation from the US' own highest mortality year. As Figure 12 shows, $k = 3$ is the point at which the Hill estimator of tail spikes, exceeding 1. Similarly, when $R = 3$, and $k = 4$ the Hill estimator is calculated using 3 observations from the "globally most affected county" and a single observation from the US' own highest mortality year. Once again, one observes a spike in the Hill estimator at $k = 4$ (Fig. 13).

The above result is intertwined with the key message of the paper and therefore it needs to be well understood: The main cause of this phenomenon is that once data from the US most affected year, or years (if $R > 1$), is combined with the global most affected country, the combination of a historically low mortality case (the US) with high mortality cases from the global observations produces a wide spread in the observations, causing a heavy tail and

unbounded mean behavior. Apart from the spikes, one also observes *upward* shifts in the graphs as the number of replacement years, $R$, increases. For sufficiently large $R$, the shift is large enough that the Hill estimators of the tail remain above $\xi = 1$ for all $k$, even after the values subside past the spike point. This indicates a very heavy tail and the absence of a finite mean. This is best seen by a comparison of say Figure 12 for $k = 2$, against Figures 14 and 15 for $k = 13$ and 14 respectively. These results are consistent with the parametric results, at least in terms of the Pareto distribution. In short, while the US remains a low risk country in general, the possibility of a US contagion on par with observed global occurrences raises the risks significantly. These results are in fact consistent with the expected US mortality rates as will be discussed below.

### 3.3 Predicting expected US death tolls from a global contagion

Predicting US mortality from a global contagion depends on the policymakers' choice of the number years that US data could be randomly swapped with global data (Table 3). It also depends on what heavy tail distribution one chooses to fit to that data. Regardless of these variations, the range of values is wide but in some instances quite close to the actual US mortality rate from COVID-19. For example, under the Pareto distribution, randomly allowing 6 years of US mortality data to mimic global data raises total expected US death toll, caused by a respiratory contagion, from about 76,000 based on purely US data for 1990-2017 period (Table 1), to near 790,000. Under GPD and EVT distributions, randomly allowing only 2 years of data to mimic global data raises the death tolls to about 110,000 and 431,000 respectively. These latter values fare much worse if US mortality is allowed to be as likely as the most affected country for all of 28 years. In that case, the death toll rises to between 588,000 and 631,000. These last set of numbers are surprisingly close to the actual US death toll from COVID-19 at the time of this writing. Note that the cases where no finite mean exists in Table 3 correspond precisely to the regions of space in Figures 9-11 where the shape parameters indicate nonexistence of finite means.

The above results suggest that scant attention to the global mortality rates, as indicated by just a few years of replacement of domestic data with global data, would have led to predicting mortality rates that far exceed the 76,000 value obtained from CDC data alone. Both the finite means (shown) and the infinite means (implied) suggest that heavy tails are driven by what appears–from the perspective of the domestic data–to be extreme outliers. This is sometimes expected as in the case of finite means, other times unexpected as in the case of infinite mean (precluding the formation of expectations). This is possible because of the slower decay rate of the Pareto distribution parameter when extrapolating from the domestic cluster to the more distant global cluster. Below, we will discuss what this means for policy formation.

## 3.4 Rational Policy Challenge and "infinite mean" phenomenon

The idea of an infinite mean is unrealistic and only a mathematical abstract. An infinite mean results because statistical inference from the extreme tail of the actual data extrapolates to an unbounded upper value. To remedy this problem [Cirillo & Taleb (2020)] develop an equivalent "dual" distribution that incorporates a finite upper bound. A similar point is raised regarding the second moment: Here, [Costello et al. (2010)] argue against a point which was initially raised in [Weitzman, 2009]; [Weitzman, 2011] regarding the infinite variance and thus uncertainty of the long-run global distribution of temperature, and the consequent failure of a rational, expectation-based environmental policy, by imposing and upper bound on the extent of that uncertainty.

Yet, when viewed *ex anti*, the idea that a finite mean fails to exist, and therefore that expectations cannot be formed, is the very definition of a surprise, i.e., an unexpected or unpredictable event that agents were not prepared for. Two recent examples come to mind: 9/11 and COVID-19. The question is: does this absolve policymakers because they could not have prepared for a surprise? We argue that the answer is no. This is because preparing for large potentially catastrophic events, *ex anti*, is the very task of numerous risk management

agencies of any government (e.g., DHS and CDC in the US). No one prepares for an infinite event but one *can* prepare for a large catastrophic event. One way to do this is to impose a large upper bound on the largest conceivable value in the data as was done by [Costello et al. (2010)]. Our approach is different but equivalent in its effect. Here, rather than imposing an upper bound on the number of fatalities, we impose equivalently an upper on the *extent of the mixing* of the two datasets, US and global, up to the point to still achieve a finite mean (Table 3). Our predicted fatality rates are much larger than the historical average from the US-only data as was discussed above. In fact, our fatality rates when data mixing is allowed are surprisingly close the total US death toll from COVID-19 at the time of this writing when the event has nearly run its course, though the emergence of the Delta variant may change this picture. The key message in the case of COVID-19 is that policymakers could have been much better prepared based solely on a statistical analysis that would have taken global contagions into account. We leave the implication for other extreme events including natural disasters to future research.

# 4 Concluding Remarks and Further Observations

Using methods from Extreme Value Theory to estimate heavy tails of the distribution of mortalities from upper and lower respiratory contagions and allowing for global mortality rates to randomly influence US mortality[12], we are able to predict US mortality rates that are very close to the US mortality rate from COVID-19 and are far higher than historical US average of about 76,000 for the 1990-2017 period (Table 1). Only under extreme circumstances does a finite mean fail to exist, indicative of a very fat tail and amounting to a surprise and failure of rational expectations. Yet the fact that estimates of statistical means do exist for a large number of circumstances suggests that risk considerations and rational policymaking could predict most large events and that policymakers can only rarely claim

[12]This can signify for example a way to produce a US surprise by mimicking other countries' lack of preparedness to the poorer infrastructure.

to be truly surprised. When attention is only paid to mortality rates for illnesses for which we have well-developed strategies in place, the ability to plan for the unknown is vastly diminished.

Yet, underestimating risk by policymakers (as distinct from the scientific community) may have characterized the US health care policy for some time. For example, it is generally well known that the US has faced major challenges in public health over the years and has simultaneously experienced a general decline in related expenditures [(Meit, et al. 2013)] with the outcome that it ranks well behind major industrial economies in a large number of health indices [United Health Foundation, 2019]. A 2014 commonwealth Fund Report by [Davis et al. (2014)] ranked the US last among 34 industrialized countries in terms of the quality of care, access to care and equity for the fifth time over a 10 year period. Moreover, it has been suggested that the poor health outcomes are not limited to low income or minority groups but are found across all incomes ages, races and education levels.[13] While the current COVD-19 pandemic has revealed possible inadequacies in public health policy in the US, the origin for this trend may date back at least to the 1990's, if not earlier [World Health Organization Report, 2000].

[13]See American Public Health Association. https://www.apha.org/topics-and-issues/health-rankings

Table 1. Mortality from upper and lower respiratory contagions*

| | | | Per 100,000 inhabitants | | | | Per total inhabitants (2017 equivalent population)[2] | |
|---|---|---|---|---|---|---|---|---|
| Year | US | Year | US | Most affected country[1] | Uniform global average[1] | Country-wise global average[1] | US (1980-1989) | US (1990-2017) |
| 1980 | 31.4 | 1990 | 36.8 | 315.34 | 65.39 | 72.31 | 102,081.4 | 119,636.8 |
| 1981 | 30.0 | 1991 | 34.7 | 306.00 | 63.85 | 70.41 | 97,530.0 | 112,809.7 |
| 1982 | 26.5 | 1992 | 32.8 | 296.51 | 62.32 | 69.58 | 86,151.5 | 106,632.8 |
| 1983 | 29.8 | 1993 | 35 | 285.67 | 60.70 | 68.65 | 96,879.8 | 113,785.0 |
| 1984 | 30.6 | 1994 | 33.6 | 273.21 | 58.97 | 67.33 | 99,480.6 | 109,233.6 |
| 1985 | 34.5 | 1995 | 33.4 | 262.02 | 57.39 | 65.56 | 112,159.5 | 108,583.4 |
| 1986 | 34.8 | 1996 | 32.9 | 249.20 | 55.65 | 64.05 | 113,134.8 | 106,957.9 |
| 1987 | 33.8 | 1997 | 33.3 | 243.17 | 54.23 | 62.53 | 109,883.8 | 108,258.3 |
| 1988 | 37.3 | 1998 | 34.6 | 232.66 | 52.66 | 60.83 | 121,262.3 | 112,484.6 |
| 1989 | 35.9 | 1999 | 23.5 | 214.31 | 50.92 | 59.02 | 116,710.9 | 76,398.5 |
| | | 2000 | 23.7 | 203.96 | 49.29 | 57.06 | | 77,048.7 |
| | | 2001 | 22.2 | 200.67 | 47.59 | 54.95 | | 72,172.2 |
| | | 2002 | 23.2 | 197.20 | 46.11 | 53.23 | | 75,423.2 |
| | | 2003 | 22.6 | 192.66 | 44.80 | 51.79 | | 73,472.6 |
| | | 2004 | 20.4 | 184.69 | 43.42 | 50.29 | | 66,320.4 |
| | | 2005 | 21 | 175.14 | 42.39 | 49.02 | | 68,271.0 |
| | | 2006 | 18.4 | 173.18 | 41.29 | 47.72 | | 59,818.4 |
| | | 2007 | 16.8 | 172.36 | 40.35 | 46.56 | | 54,616.8 |
| | | 2008 | 17.6 | 165.98 | 39.57 | 45.37 | | 57,217.6 |
| | | 2009 | 16.5 | 161.92 | 38.79 | 44.30 | | 53,641.5 |
| | | 2010 | 15.1 | 158.17 | 38.08 | 43.30 | | 49,090.1 |
| | | 2011 | 15.7 | 152.19 | 37.54 | 42.46 | | 51,040.7 |
| | | 2012 | 14.4 | 146.46 | 36.90 | 41.22 | | 46,814.4 |
| | | 2013 | 15.9 | 143.95 | 36.28 | 40.64 | | 51,690.9 |
| | | 2014 | 15.1 | 138.56 | 35.72 | 40.07 | | 49,090.1 |
| | | 2015 | 15.2 | 136.36 | 35.34 | 39.66 | | 49,415.2 |
| | | 2016 | 13.5 | 133.26 | 34.88 | 39.12 | | 43,888.5 |
| | | 2017 | 14.3 | 129.46 | 34.29 | 38.59 | | 46,489.3 |
| Average over sample period | US: 1980-2017: 25.71 | | US: 1990-2017: 23.29 | 201.58 | 46.60 | 53.06 | 105,527.5 | 75,725.1 |
| Minimum over sample period | 13.5 | | 13.5 | 129.46 | 34.29 | 38.59 | 86151.5 | 43,888.5 |

Notes

*Data are from Centers for Disease Control (CDC) and Institute for Health Metrics (IHME) at the University of Washington.

1. Let, $N_{it}$ and $P_{it}$ = total deaths and population per country $i$ in year $t$. Then, uniform average mortality $= 100,000 \times (\sum_{i=1}^{195} N_{it})/P_{wt}$ where $P_{wt}$ is the world population for year $t$; country level average mortality $= 100,000 \times \sum_{i=1}^{195} N_{it}/P_{it}$; most affected country mortality $= 100,000 \times (N/P)_{\mu t}$ where $(N/P)_{\mu t} = \max[(N/P)_{1t,}(N/P)_{2t}...(N/P)_{195,t}]$ is maximum fatality ratio for year $t$ belonging to country $\mu$.

2. Numbers reflect what would have been the equivalent total death for 2017 US population of 325,100,000.

Table 2. Maximum likelihood parameter estimates of the heavy tailed distributions*

| Distribution types estimated X= death per 100,000: | Parameter estimates | | |
|---|---|---|---|
| | shape ξ (GEV & GPD) or decay rate α (Pareto) | scale (σ) | location μ (GEV) or threshold μ (GPD) |
| | **data type & measure: US 1980-2017** | | |
| **PARETO** | 1.699 | 13.500 | |
| SE | 0.209 | 0.585 | |
| 95% CI | (1.29, 2.11) | (12.35, 14.64) | |
| **GEV** | -0.75 | 9.58 | 24.81 |
| SE | 0.14 | 1.57 | 1.69 |
| 95% CI | (-1.03, -0.47) | (6.50, 12.66) | (21.50, 28.11) |
| **GPD** | -1.06 | 25.19 | 25.71 |
| SE | 0.01 | NaN | |
| 95% CI | (1.07, -1.04) | - | |
| | **data type & measure: US 1990-2017** | | |
| **PARETO** | 2.053 | 13.500 | |
| SE | 0.370 | 0.582 | |
| 95% CI | (1.33, 2.78) | (12.36, 14.64) | |
| **GEV** | 0.45 | 4.88 | 18.17 |
| SE | 0.34 | 1.20 | 1.31 |
| 95% CI | (-0.22, 1.12) | (2.52, 7.24) | (15.61, 20.73) |
| **GPD** | -0.96 | 22.34 | 23.29 |
| SE | 0.00 | 0.18 | |
| 95% CI | (-0.96, -.96) | (21.9802, 22.7030) | |
| | **uniform world average per year Global data: 1990-2017** | | |
| **PARETO** | 3.50 | 34.29 | |
| SE | 0.56 | 0.55 | |
| 95% CI | (2.41, 4.60) | (33.20, 35.38) | |
| **GEV** | 0.22 | 6.84 | 41.12 |
| SE | 0.33 | 1.56 | 1.83 |
| 95% CI | (-0.43, 0.88) | (3.78, 9.90) | (37.53, 44.71) |
| **GPD** | -0.78 | 24.84 | 46.60 |
| SE | 0.24 | 6.65 | |
| 95% CI | ( -1.25, -0.31) | (11.81, 37.86) | |
| | **cross country average per year: global data 1990-2017** | | |
| **PARETO** | 3.36 | 38.59 | |
| SE | 0.49 | 0.54 | |
| 95% CI | (2.41, 4.32) | (37.52, 39.65) | |
| **GEV** | -0.05 | 9.22 | 48.05 |
| SE | 0.37 | 2.17 | 2.57 |
| 95% CI | (-0.77, 0.66) | (4.97, 13.48) | (43.01, 53.09) |
| **GPD** | -0.95 | 32.10 | 53.06 |
| SE | 0.01 | 0.44 | |
| 95% CI | (-0.96, -0.93) | (31.24, 32.96) | |
| | **most affected country per year: global data 1990-2017** | | |
| **PARETO** | 2.47 | 129.46 | |
| SE | 0.37 | 3.87 | |
| 95% CI | (1.74, 3.19) | (121.88, 137.03) | |
| **GEV** | 0.17 | 39.87 | 171.69 |
| SE | 0.26 | 7.96 | 9.71 |
| 95% CI | (-0.34, 0.69) | (24.26, 55.48) | (152.65, 190.73) |
| **GPD** | -0.68 | 131.89 | 201.58 |
| SE | 0.24 | 36.62 | |
| 95% CI | (-1.15, -0.20) | (60.11, 203.66) | |

*The location parameter for the GEV distribution is estimated as its other parameters. For the GPD distribution, the threshold value is assumed based on the smallest observation (minimum mortality rate) over the sample period as is seen in Table 1. See the text for the detailed of the estimation methods.

Table 3. Expected US mortality from respiratory contagions when considering global risk*

| # of years replaced | Expected US mortality per 100,000[1] | | | Expected total US Mortality[2] | | |
|---|---|---|---|---|---|---|
| | Pareto Distribution | Generalized Pareto Distribution | Extreme Value Distribution | Pareto Distribution | Generalized Pareto Distribution | Extreme Value Distribution |
| 0 | 26.32 | 24.92 | 24.87 | 85,566 | 80,999 | 80,850 |
| | (26.32,26.32) | (24.92,24.92) | (24.87,24.87) | | | |
| 1 | 31.09 | 29.09 | 41.20 | 101,066 | 94,559 | 133,926 |
| | (29.26,31.47) | (27.68,30.87) | (31.14,44.62) | | | |
| 2 | 37.74 | 37.63 | 148.60 | 122,690 | 122,349 | 483,101 |
| | (35.26,38.54) | (33.20,47.03) | (43.65,inf) | | | |
| 3 | 48.15 | 56.15 | . | 156,537 | 182,538 | . |
| | (42.50,50.06) | (41.14,161.61) | . | | | . |
| 4 | 66.47 | 123.67 | . | 216,097 | 402,042 | . |
| | (54.84,70.85) | (51.94,inf) | . | | . | . |
| 5 | 106.67 | . | . | 346,786 | . | . |
| | (76.88,121.02) | . | . | | . | . |
| 6 | 271.68 | . | . | 883,241 | . | . |
| | (129.59,408.04) | . | . | . | . | . |
| 7 | . | . | . | . | . | . |
| 8 | . | . | . | . | . | . |
| 9 | . | . | . | . | . | . |
| 10 | . | . | . | . | . | . |
| 11 | . | . | . | . | . | . |
| 12 | . | 245.79 | . | . | 799,070 | . |
| | . | (94.12,inf) | . | . | | . |
| 13 | . | 113.74 | . | . | 369,764 | . |
| | . | (99.77,inf) | . | . | | . |
| 14 | . | 115.87 | . | . | 376,698 | . |
| | . | (105.40,126.07) | . | . | | . |
| 15 | . | 120.98 | . | . | 393,319 | . |
| | . | (111.2,137.38) | . | . | | . |
| 16 | . | 130.00 | 127.61 | . | 422,635 | 414,866 |
| | . | (118.14,147.43) | (115.43,inf) | . | | |
| 17 | . | 137.24 | 130.47 | . | 446,170 | 424,144 |
| | . | (119.79,157.47) | (121.62,140.30) | . | | |
| 18 | . | 141.12 | 137.49 | . | 458,786 | 446,994 |
| | . | (122.12,162.16) | (128.51,146.31) | . | | |
| 19 | . | 148.57 | 144.22 | . | 483,005 | 468,855 |
| | . | (127.06,164.09) | (135.32,152.77) | . | | |
| 20 | . | 153.81 | 150.70 | . | 500,022 | 489,932 |
| | . | (130.14,167.78) | (141.97,158.64) | . | | |
| 21 | . | 157.99 | 157.23 | . | 513,626 | 511,148 |
| | . | (141.64,169.97) | (148.61,164.94) | . | | |
| 22 | . | 160.10 | 163.72 | . | 520,475 | 532,261 |
| | . | (145.08,170.44) | (155.27,170.79) | . | | |
| 23 | . | 161.53 | 170.10 | . | 525,134 | 552,990 |
| | . | (148.55,172.31) | (162.42,176.50) | . | | |
| 24 | . | 162.66 | 176.52 | . | 528,802 | 573,852 |
| | . | (148.97,172.57) | (169.13,182.01) | . | | |
| 25 | . | 165.07 | 182.58 | . | 536,639 | 593,551 |
| | . | (151.79,173.62) | (176.31,187.03) | . | | |
| 26 | . | 165.08 | 189.11 | . | 536,673 | 614,783 |
| | . | (152.01,177.96) | (183.49,192.88) | . | | |
| 27 | . | 170.86 | 195.70 | . | 555,459 | 636,228 |
| | . | (160.73,177.97) | (191.73,197.44) | . | | |
| 28 | 217.80 | 207.96 | 202.78 | 708,064 | 676,070 | 659,237 |
| | (217.80,217.80) | (207.96,207.96) | (202.78,202.78) | | | |

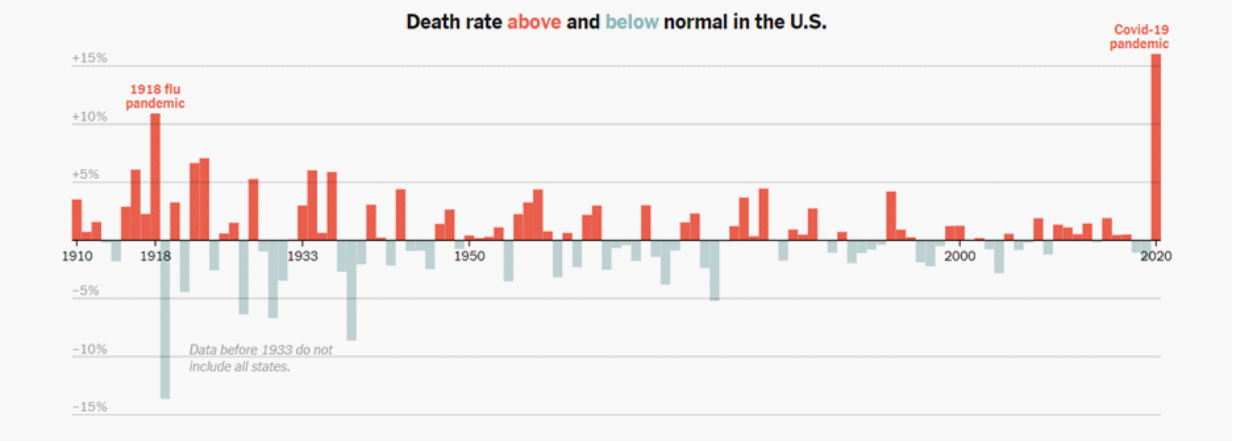
Notes

*Numbers in parenthesis are 95% simulated confidence intervals from 1000 runs for each distribution fit for each replacement day. Dots (.) indicate that the expected mortality rate is not finite.

1. Reported values are the medians of expected mortality from 1000 runs for each distribution. To construct 95% confidence intervals, we drop the smallest and largest 2.5%.

2. Expected total US death is calculated based on 2017 US population of 325,100,000.

Figure 1. US historical excess death rates: 1910-2020




Source: New York Times, April 23, 2021

# Figures 2-3. The Hill estimators of the tail: US

Fig. 2 US 1980-2017

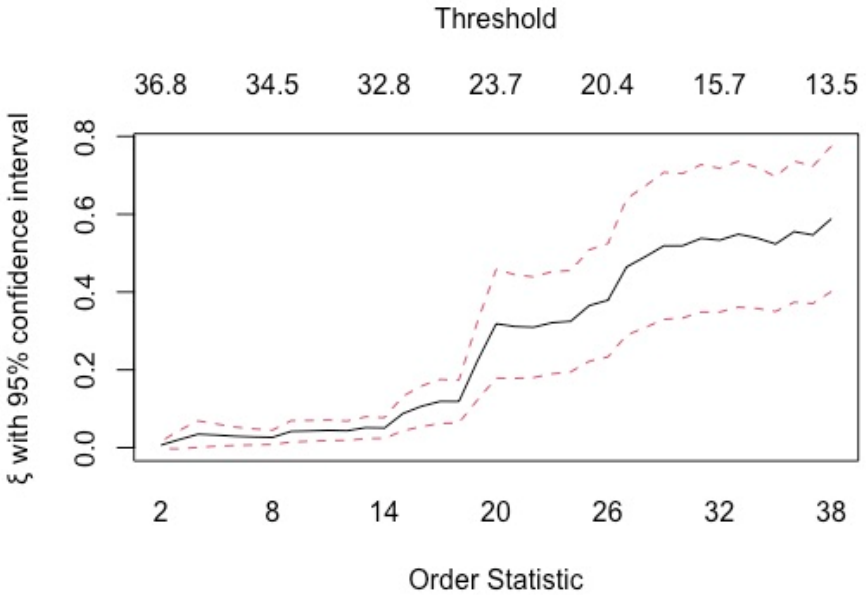
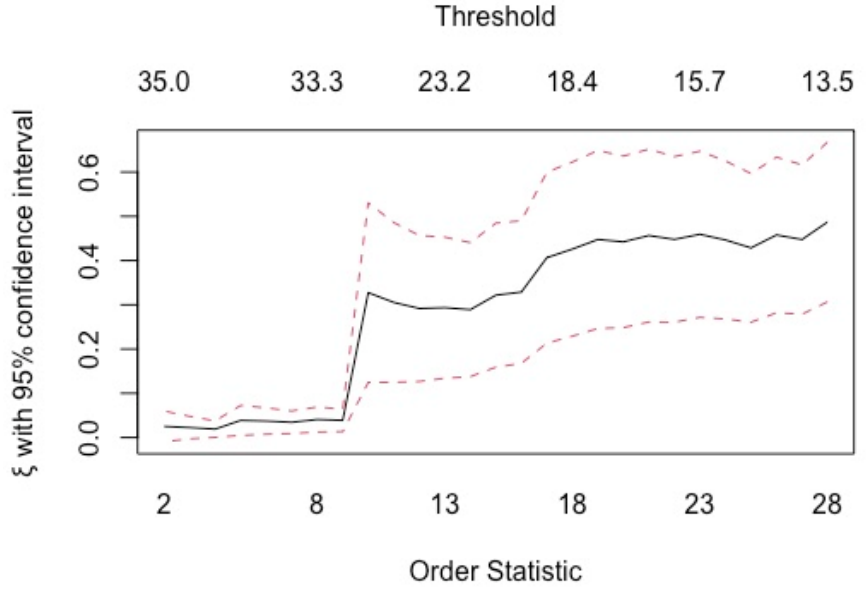



Fig 3. US 1990-2017




Note. The threshold values in Figures 2 and 3 are implied by the order statistic and thus move inversely with the latter: Given a vector of 28 observations, ordered from the highest to the lowest, the order statistic of 2 means that threshold is the second highest value and the order statistic of 3 means that it is the third highest value, etc.

Figures 4-6. The Hill estimators of the tail: the world

Figures 4 & 5. world country average and world uniform average, 1990-2017

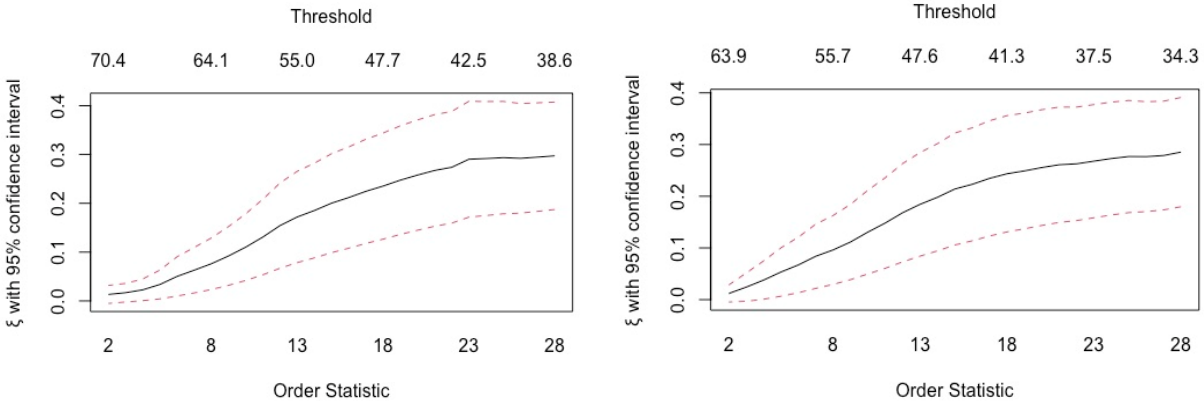
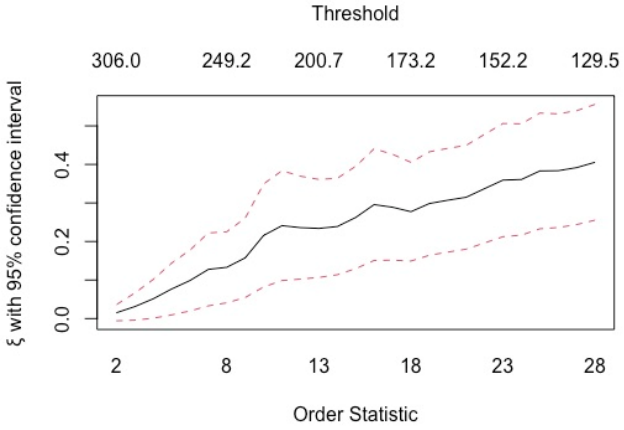



Figure 6. world's most affected country each year, 1990-2017




Notes.
1. For the definitions of the three measures used in these figures, see footnote 6 in the main text.
2. Regarding the threshold values, see note on figures 2 and 3.

Figure 7-8. Panel 1990-2017 and 195 countries. The Hill and Pareto estimates

Figure 7. The Hill estimators of the tail

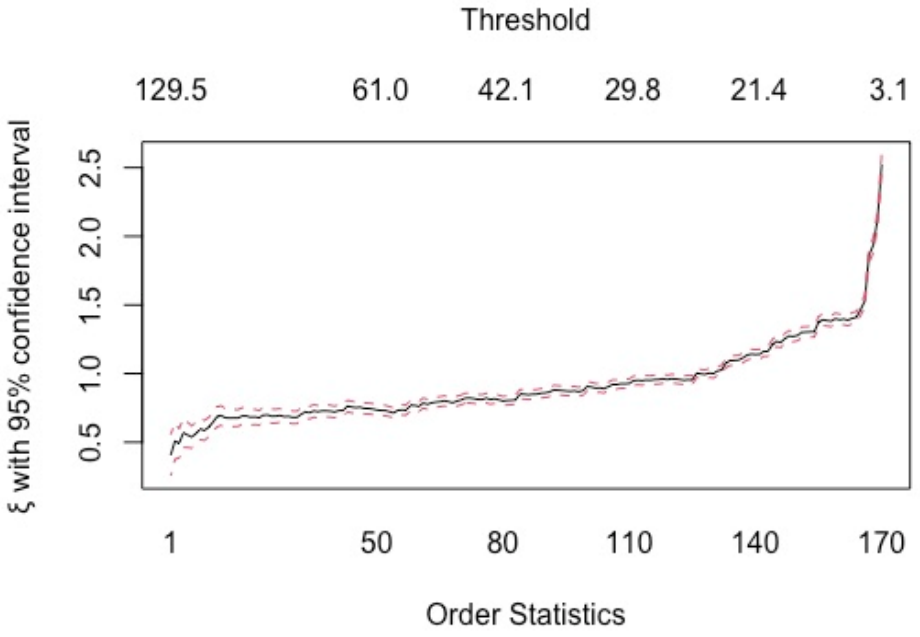
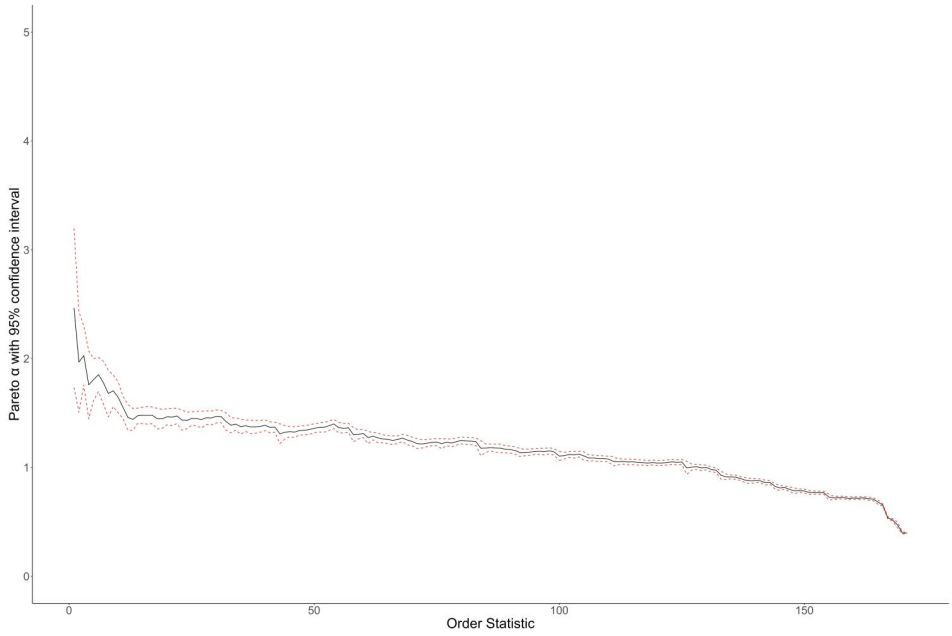



Figure 8. Pareto shape parameter (decay rate)




Note on figure 7. For threshold values, see note on figures 2.
Note on Figure 8. The order statistics for Pareto distribution for the panel data works similarly to the Hill estimator: For example, for $k = 10$ we use the 10 most affected per year for a total of 28 years, yielding a vector of 280 observations. We then estimate the Pareto distribution for that vector.

# Figures 9-11. Estimates of shape parameter for counterfactuals as a function of years of replacing US data with the world's most affected country

(see notes following figure 11)

Figure 9. Pareto distribution (shape or decay rate)

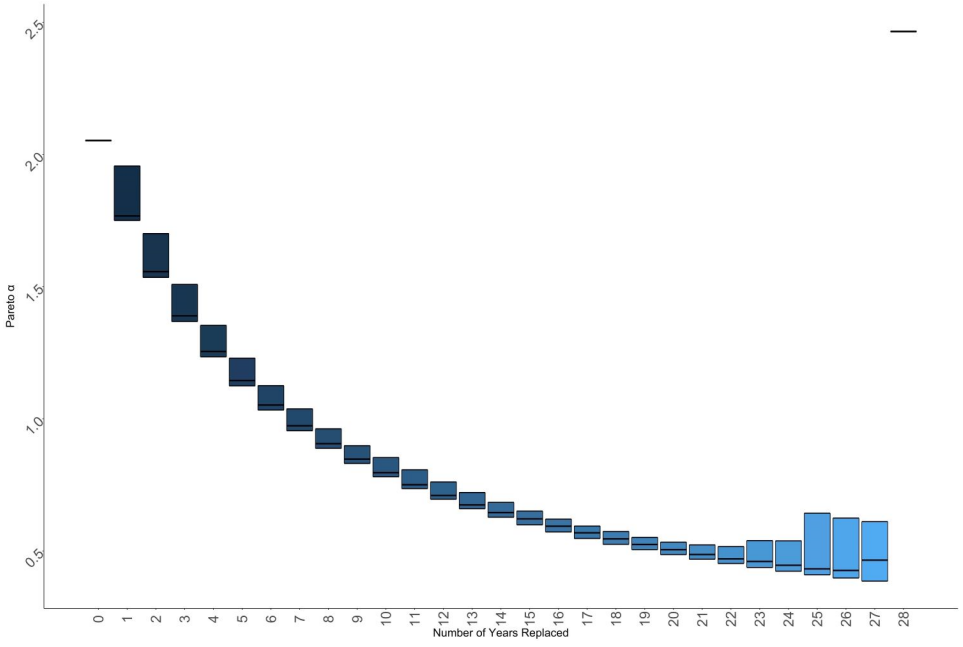


Figure 10. Generalized Pareto distribution

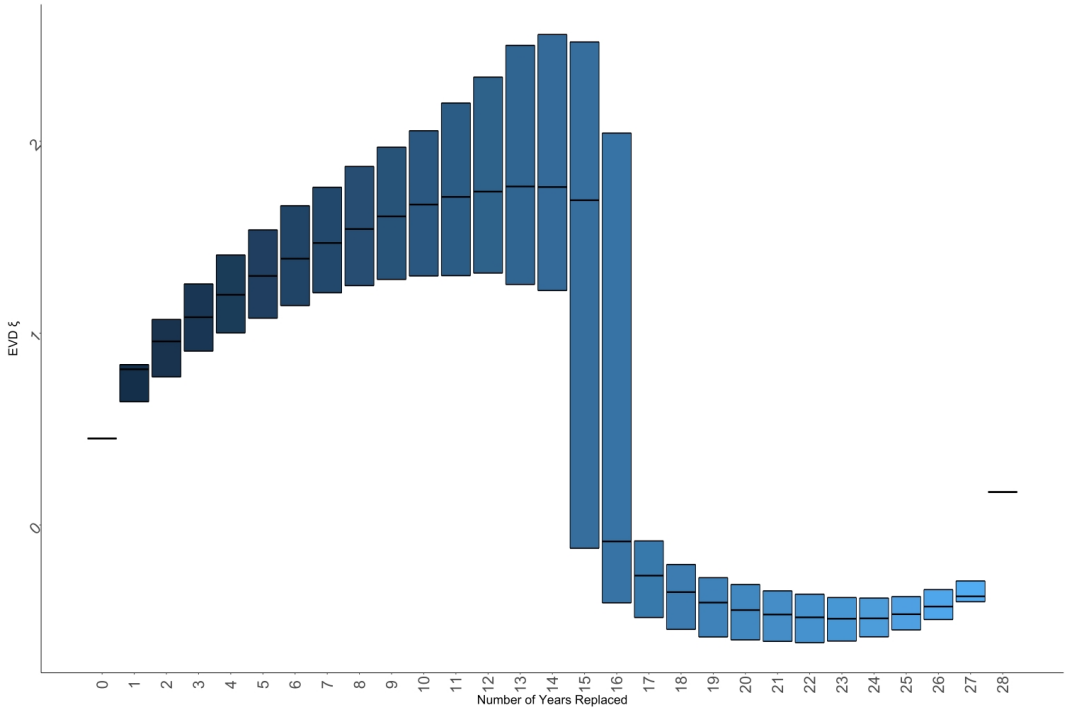
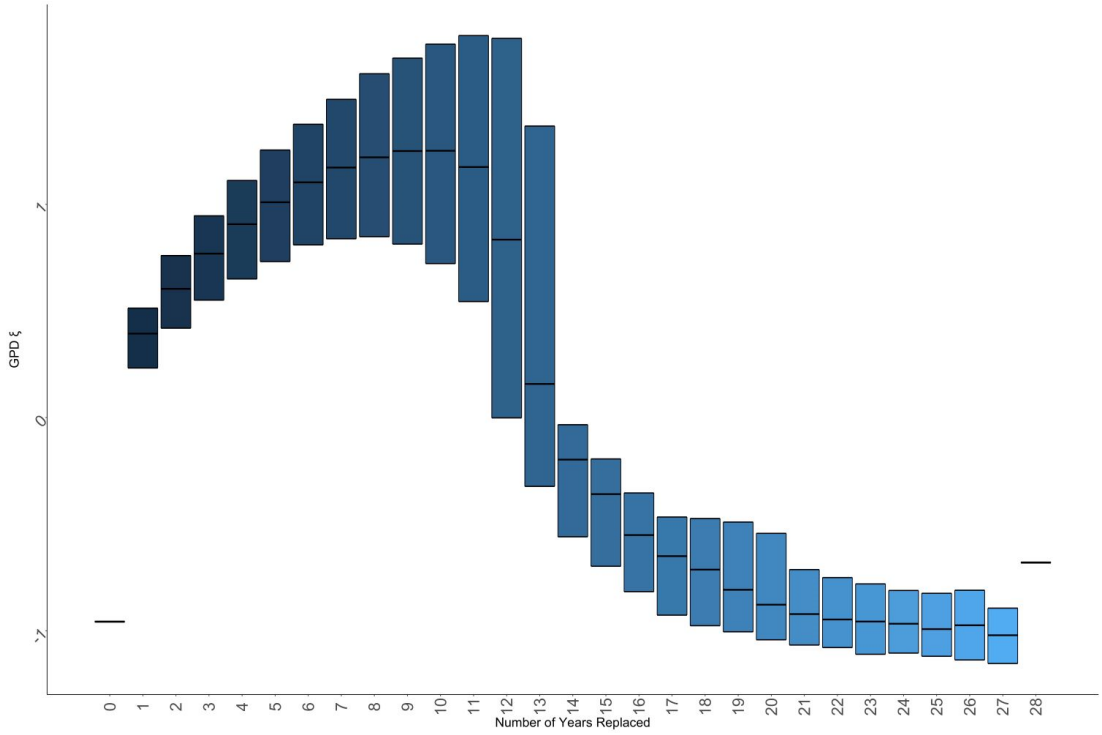


Figure 11. Generalized Extreme Value distribution




Notes on Figures 9-11. Estimates were obtained by taking 28 years of US CDC data on mortalities per 100,000 and replacing data points with the highest mortality rate for that year from the global data. In the case of 0 years replaced, the dataset is just the US data. In the case of 28 years replaced the dataset is just global annual most affected country rates. In between these extremes, say for 3 years replaced, 3 year's values from the global dataset are randomly selected. These three values then are used to replace US data for that same year, creating a counterfactual vector of observations that effectively shows, what if for those three random years, the US had a mortality rate equal to the highest mortality rate observed globally in that year. This is done 1000 times, because when the number of replacement years increase, the permutations of possible replacements grows extremely large. The graphs show the median estimate of the shape parameter, with bands for the estimated 95% of simulated cases (the middle 950 shape parameters). Estimates come from having different years replaced, and thus represent the variety of point estimates seen.

# Figures 12-15.

## Hill estimator of counterfactuals for # of random years (R) of letting US mortality to mirror world's most affected country, as a function of order statistics (see notes)

Figure 12. R=2, k=1-27

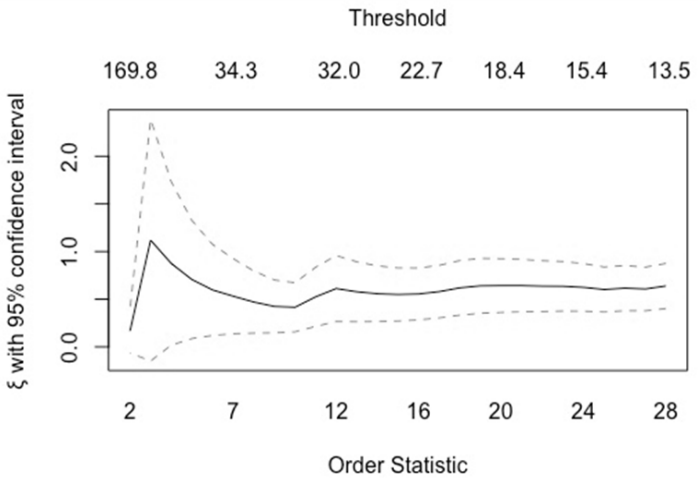



Figure 13. R=3, k=1-27

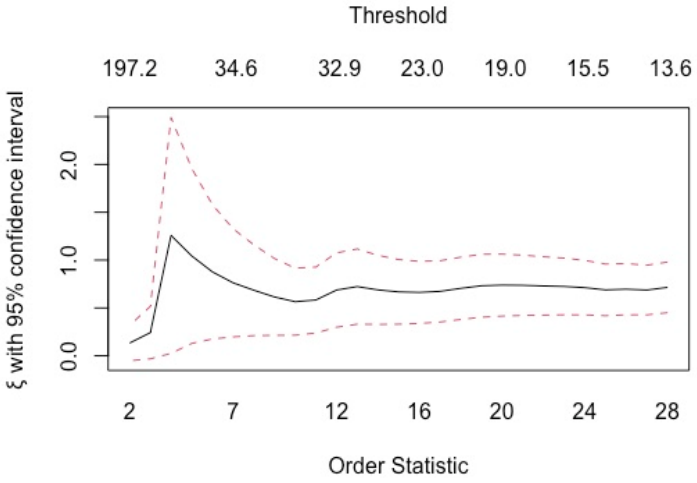



Figure 14. R=13, k=1-27

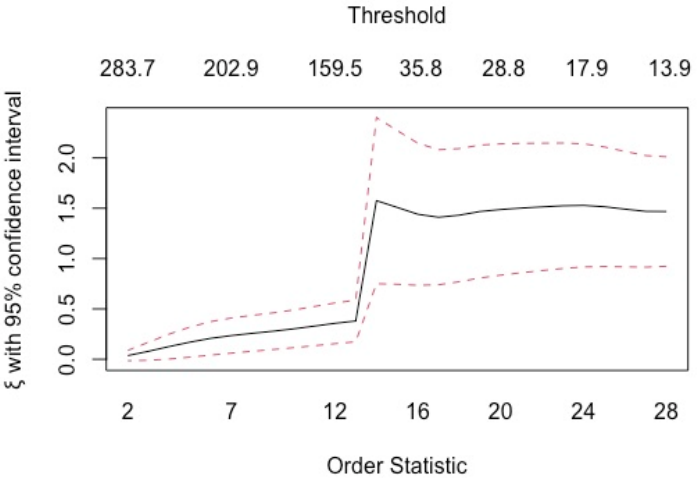
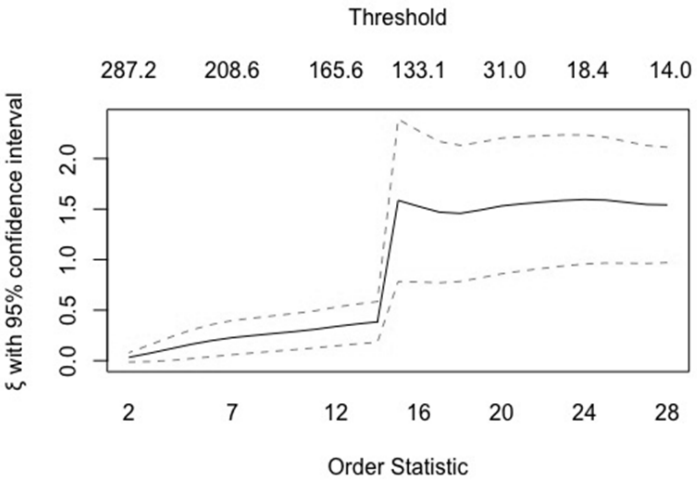



Figure 15. R=14, k=1-27




Note on figures 12-15. For a given $k$, and $R$, if $k > R$, we replace $R$ years of US data with the most affected countries for each of the $k$ years. Naturally, in the resulting order statistics the remaining $k - R$ years comes from US's own mortality rates. Specifically, when $k = R + 1$, the entry of US's own most affected year when combined with $R - k$ most affected country years produces a wide spread in the data causing a jump in the tail estimate passed tail value of 1. See also the text.

# 5 Disclosures and Declarations

## 5.1 Funding and/or Conflicts of interests/Competing interests

Funding: none

Conflicts of Interest or Competing Interests: Neither author has any conflicting or competing interests to report.

# References


Bali T (2007) A Generalized Extreme Value Approach to Financial Risk Measurement. *Journal of Money, Credit, and Banking* 39(7):1611-1647.

Coles S, Bawa J, Trenner L, Dorazio, P (2001) An introduction to statistical modeling of extreme values. London: Springer.

Costello C, Neubert M, Polasky S, Solow A (2010) Bounded Uncertainty and Climate Change Economics. *Proceedings of the National Academy of Sciences* 107(18):8108–8110.

Cirillo P, Taleb N (2020) Tail risk of contagious diseases. *Nature Physics* 16: 606-613.

Davis K, Stremikis K, Squires D, Schoen C (2014) Mirror, Mirror on the Wall: How the Performance of the U.S. Health Care System Compares Internationally. *The Commonwealth Fund.* (https://www.commonwealthfund.org/sites/default/files/documents/___media_files_publications_fund_report_2014_jun_1755_davis_mirror_mirror_2014.pdf)

Embrechts P, Klüppelberg C, Mikosch T (2003) Modelling Extremal Events (Springer).

Fisher R & Tippett L (1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society* 24(02):180-190.

Gnedenko B (1943) Sur la distribution limit´ee du terme maximum d'une s´erie al´eatoire, *Annals of Mathamatics* (44):423–453.

Hill B (1975) A simple general approach to inference about the tail of a distribution. The *Annals of Statistics* 3(5):1163–1174.

Jones C (2015) Pareto and Piketty: The Macroeconomics of Top Income and Wealth Inequality. *Journal of Economic Perspectives* 29(1):29–46.

Lin J (1991) Divergence Measures Based on the Shannon Entropy *IEEE Transactions on Information Theory* 37(1):145-151.

Kullback S Leibler R (1951) On Information and Sufficiency. *The Annals of Mathematical Statistics* 22(1):79-86.

Meit M, Knudson A, Dickman I, Brown A, Hernandez N, and Kronstadt, J. An Examination of Public Health Financing in the United States. (Prepared by NORC at the University of Chicago.) Washington, DC: The Office of the Assistant Secretary for Planning and Evaluation. March 2013.

Meyer S, Held L (2014) Power-law models for infectious disease spread. The Annals of Applied Statistics 8(3):1612-1639.

Mohtadi H, Murshid A (2006) Analyzing Catastrophic Terrorist Events with Applications to the Food Industry in The Economic Costs and Consequences of Terrorism, Harry Richardson (ed), Edward Elgar 2006, Ch 10. (London, UK), ISBN #1845427343

Mohtadi H (2009) Developing Risk Metrics to Estimate Risks of Catastrophic Biological and Bioterrorist Events: Applications To The Food Industry. in *Handbook of Science and Technology for Homeland Security*, John G. Voeller (ed), John Wiley & Sons: Hoboken, NJ.

Mohtadi H, Murshid A (2009) Risk of Catastrophic Terrorism: An Extreme Value Approach. *Journal of Applied Econometrics* 24(4):537-559.

Mohtadi H, Weber B (2021) Catastrophe and Rational Policy: The Case of National Security. *Economic Inquiry* 59(1): 140-161

Mori M, Smith T E, Wen-Tai Hsu (2020) Common Power Laws for Cities and Spatial Fractal Structures. *Proceedings of the National Academy of Sciences* 117(12):6469-6475.

Németh L , Zempléni A (2020) Regression Estimator for the Tail Index. *Journal of Statistical Theory and Practice* 14:48

Newman M (2005) Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* 46(5):23-351.

Ravasz E, Somera A, Mongru D, Oltvai Z, Barabási A (2002) A Hierarchical organization of modularity in metabolic networks. *Science* 297(5586): 1551–1555.

Pike J, Shogren J F, Aadland D, Viscusi W K, Finnoff D, Skiba A, Daszak P (2020) Catastrophic Risk: Waking Up to the Reality of a Pandemic? EcoHealth 17(2):217-221.

Pindyck R (2011) Fat Tails, Thin Tails, and Climate Change Policy. *Review of Environmental Economics and Policy* 5(2):258–274

Sarabia, J M, Prieto F (2009) The Pareto-positive Stable Distribution: A New Descriptive Model for City Size Data. *Physica A* 388: 4170-4191.

Stanley M, Amaral L, Buldyrev S, Havlin, S, Heiko L, Maass P, Salinger M, Stanley H (1996) Scaling behaviour in the growth of companies. *Nature* 379 (6568):804–806.

Spagat M, Johnson N, Weezel S (2018) Fundamental patterns and predictions of event size distributions in modern wars and terrorist campaigns. *PLoS ONE* 13(10):e0204639 (https://doi.org/10.1371/journal.pone.0204639).

United Health Foundation Report (2019) America's Health Rankings (https://www.americashealthrankings.org).

Viscusi, W K (2009) Valuing Risks of Death from Terrorism and Natural Disasters Journal of Risk and Uncertainty 38(3):191-213.

Viscusi, W K (2020) Pricing the Global Health Risks of the COVID-19 Pandemic Journal of Risk and Uncertainty 61(1):101-128.

Weitzman M (2009) On Modeling and Interpreting the Economics of Catastrophic Climate Change. *The Review of Economics and Statistics* 91(1):1-19.

Weitzman M (2011) Fat-Tailed Uncertainty in the Economics of Catastrophic Climate Change. *Review of Environmental Economics and Policy* 5(2):275-292.

West, G, Brown J, Enquist B (1997) A general model for the origin of allometric scaling laws in biology *Science* 276(5309):122–126.

Wong F, Collins J (2020) Evidence that coronavirus superspreading is fat-tailed. *Proceedings of the National Academy of Sciences* 117(47): 29416-29418.

World Health Organization Report (2000) Annex Table 1: Health system attainment and performance in all Member States, ranked by eight measures, estimates for 1997 (https://www.who.int/whr/2000/en/whr00_annex_en.pdf?ua=1)

Appendix: Pareto Decay for $k$ order statistics for each year of the data