PNAS RESEARCH ARTICLE | APPLIED MATHEMATICS
POPULATION BIOLOGY

COVID-19 cases and deaths in the United States follow Taylor's law for heavy-tailed distributions with infinite variance

Joel E. Cohen[a,b,c,d,1], Richard A. Davis[c], and Gennady Samorodnitsky[e]

Contributed by Joel E. Cohen; received May 28, 2022; accepted August 9, 2022; reviewed by John Nolan and Zhengjun Zhang

The spatial and temporal patterns of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) cases and COVID-19 deaths in the United States are poorly understood. We show that variations in the cumulative reported cases and deaths by county, state, and date exemplify Taylor's law of fluctuation scaling. Specifically, on day 1 of each month from April 2020 through June 2021, each state's variance (across its counties) of cases is nearly proportional to its squared mean of cases. COVID-19 deaths behave similarly. The lower 99% of counts of cases and deaths across all counties are approximately lognormally distributed. Unexpectedly, the largest 1% of counts are approximately Pareto distributed, with a tail index that implies a finite mean and an infinite variance. We explain why the counts across the entire distribution conform to Taylor's law with exponent two using models and mathematics. The finding of infinite variance has practical consequences. Local jurisdictions (counties, states, and countries) that are planning for prevention and care of largely unvaccinated populations should anticipate the rare but extremely high counts of cases and deaths that occur in distributions with infinite variance. Jurisdictions should prepare collaborative responses across boundaries, because extremely high local counts of cases and deaths may vary beyond the resources of any local jurisdiction.

COVID-19 | lognormal–Pareto distribution | Taylor's law | fluctuation scaling | variance function

Significance

Variations in the cumulative reported SARS-CoV-2 cases and COVID-19 deaths by US county, state, and date exemplify Taylor's law of fluctuation scaling, a widespread ecological and epidemiological pattern. Specifically, on day 1 of each month from April 2020 through June 2021, each state's variance (across its counties) of cases is nearly proportional to its squared mean of cases. COVID-19 deaths behave similarly. The largest 1% of counts are approximately Pareto distributed, with a finite mean and an infinite variance. Finding infinite variance has practical consequences. Local jurisdictions (counties, states, and countries) that plan for prevention and care of largely unvaccinated people should anticipate rare but extremely high counts of cases and deaths, by preparing collaborative responses across boundaries.

By 9 May 2022, 226 countries and territories reported more than 517 million confirmed cases of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and more than 6.27 million deaths. The United States reported more confirmed cases (83.6 million) and more deaths (1,025,000) than any other country (1).

We report that cumulative counts of cases and deaths by county in states of the United States are consistent with Taylor's power law of fluctuation scaling (henceforth TL), which will be explained below. The exponent of TL is not distinguishable from two, which means that the coefficient of variation of counts by county is not distinguishable from a constant across states on any given date. The counties with the lowest 99% of counts have a distribution of counts well approximated by the lognormal distribution, while the counties with the highest 1% of counts have a distribution of counts consistent with a Pareto distribution with an infinite variance. Simple statistical models based on the lognormal–Pareto distribution reproduce TL scaling with exponent equal to two. Mathematical results explain the observed and simulated TL scaling in multiple samples from heavy-tailed distributions. Planning for prevention and care in a largely unvaccinated population should anticipate the rare but extremely high counts of cases and deaths that occur in distributions with infinite variance.

We now describe the data and TL. We test TL, the lognormal model for the lowest 99% of counts, and the Pareto model for the extreme upper tail, and give evidence for an infinite variance of counts. We then review our empirical and theoretical findings. *SI Appendix* reports simulations and mathematical analyses of models that interpret the empirical findings. We state precisely and prove mathematically that multiple samples from heavy-tailed distributions obey TL with exponent two under certain conditions. *SI Appendix* also gives additional discussion.

1. Data

In the United States, *The New York Times* has tabulated cumulative cases and cumulative deaths at the end of each day since the first reported confirmed case on 20 January 2021 (2). Cumulative cases and cumulative deaths are reported according to their location in a primary subdivision of the United States (the 50 states, plus possessions, territories, and Washington, D.C., all referred to as "states" henceforth) and, within each state, by secondary subdivision (county, parish, borough, or other equivalents of county, all referred to as "counties" henceforth). From now on, "cases" refer to cumulative cases and "deaths"

Author contributions: J.E.C. designed research; J.E.C., R.A.D., and G.S. performed research; J.E.C., R.A.D., and G.S. contributed new reagents/analytic tools; J.E.C. analyzed data; and J.E.C., R.A.D., and G.S. wrote the paper.

Reviewers: J.N., American University; and Z.Z., University of Wisconsin–Madison.

Competing interest statement: J.N. was one of 11 co-PIs on a Multidisciplinary University Research Initiative grant from the US Army Research Office in 2012–2018 along with R.A.D. and G.S., coauthors of this paper. That grant ended at the end of 2018, approximately 42 mo ago, overlapping a few months with the 48-mo exclusion period.

See online for related content such as Commentaries.

[1]To whom correspondence may be addressed. Email: cohen@rockefeller.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2209234119/-/DCSupplemental.

Published September 12, 2022.

refer to cumulative deaths on a given date. We shall use "count" or "counts" to refer to either cases or deaths.

We downloaded the file us-counties.csv with 1,436,628 lines of data on 19 June 2021. All calculations use Matlab (version 2021a). We select the first day of each of the 15 mo from April 2020 through June 2021 (keeping 47,004 lines of data). We exclude county–days with reported counts of zero (keeping 46,956 lines of data for positive cases and 37,649 lines for positive deaths), and we sort the counties by state. There were 843 month-state combinations, an average of $56.2 = 843/15$ states each month for 15 mo. We then exclude month–state combinations with six or fewer counties, leaving 718 month–state combinations with an average of $47.9 = 718/15$ states each month for 15 mo. We then compute the means and the variances of counts over the counties within each state on each of the 15 dates. Because each remaining state has seven or more counties, no remaining month–state combination has zero mean or zero variance over the counties within a state.

We attempt no adjustments for possible underreporting of cases or deaths due to COVID-19. If, as seems likely, different jurisdictions had different propensities to test, systematic data that would make it possible to adjust for such differences are not available. High counts of cases could either overwhelm local testing capacity, resulting in undercounts of tested cases, or could elicit supplemental test resources, resulting in unusually high reported cases. Such possible limitations of the data seem unlikely to have generated the systematic patterns we shall describe.

2. Taylor's Law (TL)

On each first day of 15 mo, the (mean, variance) pairs, one point for each retained state on that date, closely approximate TL, for both cases and deaths (Fig. 1). Specifically, on each date, each state's sample variance of the count (over its counties) is approximately proportional to some power of that state's sample mean of the count (over its counties). Equivalently, the logarithm of a state's sample variance is approximately a linear function of the logarithm of that state's sample mean. On log–log coordinates, there is no visual indication of systematic curvature. The ranges of the axes of all panels are the same within each figure, to make it easy see that, as time passes and the counts increase, the cloud of observed (mean, variance) pairs shifts from the lower left corner to the upper right corner.

In statistical language, TL says that the counts have a power variance function (3–5). Explicitly, for real constants $k > 0$ and b, both independent of the state i, but depending on the date and whether the counts are cases or deaths, TL proposes, and we find empirically, that

$$\text{sample variance of state } i \approx k \times (\text{sample mean of state } i)^b, \quad i = 1, 2, \ldots. \qquad [1]$$

Let $\log = \log_{10}$ here and throughout the data analysis, and let $a = \log_{10} k$. (In *SI Appendix*, the simulations use $\log_{10}$, and the mathematical theorems and proofs use the natural logarithm.) Then the power-law form of TL in Eq. **1** is equivalent to the linear log–log form of TL displayed in Fig. 1,

$$\log(\text{sample variance of state } i) \approx a + b \times \log(\text{sample mean of state } i), \; i = 1, 2, \ldots \qquad [2]$$

or the ratio form of TL,

$$\text{sample variance of state } i/(\text{sample mean of state } i)^b \approx k, \quad i = 1, 2, \ldots. \qquad [3]$$

The exponent b in Eqs. **1** and **3** is the same as the slope b in Eq. **2**, so b may be called either the slope or the exponent of TL. These specifications of TL intentionally leave vague the error model behind the approximation $\approx$.

TL is an empirical regularity widely observed in many sciences, including ecology, infectious disease epidemiology, human demography, financial statistics, earth sciences, and other physical sciences (6, 7).

The Poisson distribution, a common model of purely random variation in counts, has a variance equal to its mean. As the mean of a Poisson distribution increases to larger values, the Poisson distribution increasingly approximates a normal distribution with variance equal to the mean. Both the Poisson distribution and its normal approximation with variance equal to the mean follow TL with $k = b = 1, a = 0$. If states had different average counts per county and a Poisson distribution of counts over the counties within each state, then the states' means and variances of the counts would approximate TL with $k = b = 1, a = 0$. Graphically, on (log mean, log variance) coordinates, a family of Poisson distributions with varying mean will lie on or near a line of slope one through the origin. Fig. 1 shows that the Poisson distribution does not describe even approximately how the sample variance of counts by county within states relates to the sample mean. We infer that other sources of variation besides purely random fluctuation influence the counts, such as heterogeneity or contagion.

Only once in Fig. 1 does the 95% CI of the slope exclude $b = 2$. For deaths in April 2020, the estimated slope is 3.003 with 95% CI (2.685, 3.321). In all the remaining 29 instances of TL ($29 =$ 15 mo $\times$ 2 counts (cases or deaths) -1), the estimated slope b of TL in Eq. **2** is statistically indistinguishable from two. When $b = 2$, the ratio (sample variance)/(sample mean)^2 is independent of any positive rescaling of the original measurements.

In every month, for both cases and deaths, the lower bound of the 95% CI of the slope b exceeds one, excluding Poisson variation among counties in cases and deaths.

The parameters a, b of the log–log form of TL Eq. **2** (Fig. 2) show no substantial trends over time. For both counts, early in the period of observation, the intercept a rises slightly while the slope b falls slightly, as if the fitted straight line that represents TL were rotating slightly clockwise. In the second half of the period of observation, there is no suggestion of change in the TL parameters.

3. Lognormal Model of TL with Slope Two

To understand why TL holds and why the slope b in Eq. **2** approximates two (except for deaths in April 2020), we observe that the spread of infection could be modeled by a variety of stochastic multiplicative processes. For example, the supercritical discrete generation Galton–Watson branching process with finite mean > 1 and finite variance > 0 of the offspring distribution satisfies TL with slope two asymptotically in time (ref. 8, p. 33). The supercritical continuous-time birth and death process with a birth rate per individual that strictly exceeds the death rate per individual also satisfies TL with slope two asymptotically in time (ref. 8, pp. 33–34). These and other similar examples provide prototypes of explanations of why TL holds with slope two in the COVID-19 counts. But the details of transmission of infection and death from COVID-19 are not adequately described by these simple models.

A more robust explanation is required that does not depend on the details of transmission. The lognormal and Weibull distributions are limiting distributions of large families of mechanisms,

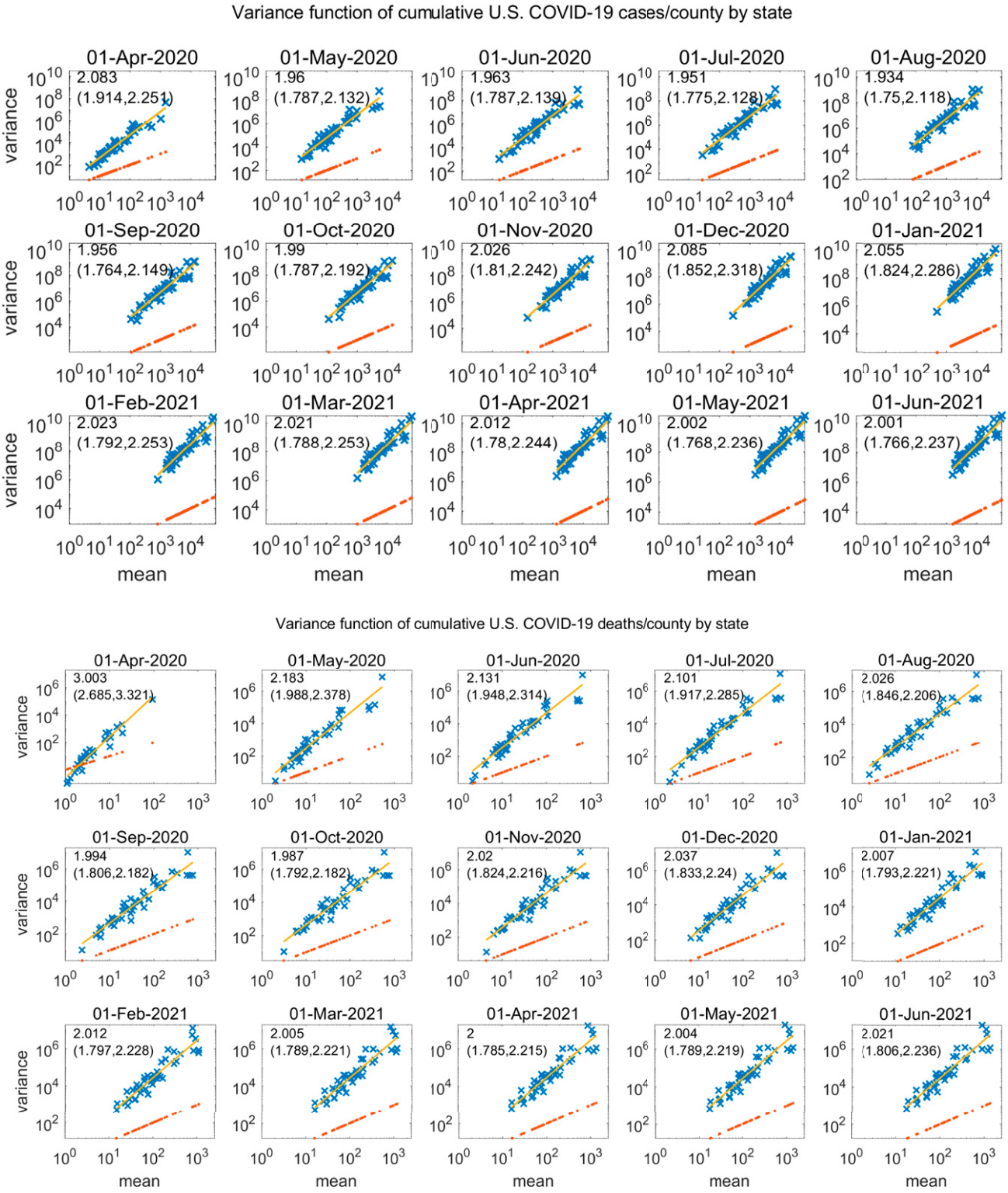

Fig. 1. Cumulative US COVID-19 cases (*Upper*) and deaths (*Lower*) by state are well described by TL. Each blue × marker shows, on log–log coordinates, the (mean, variance) of the number of cases across counties within one state. The yellow straight line fitted by ordinary least squares to the blue × markers (log mean, log variance) is the estimated TL. In all 15 mo, R^2 is between 0.90 and 0.92 for cases and between 0.87 and 0.91 for deaths. The estimated slope b of TL is the top left figure in each panel. An approximate 95% CI of the slope is given in parentheses below the estimated slope b. As the months pass, the estimates of b become increasingly and remarkably close to two. The red dots in a straight line near the bottom of each panel show a hypothetical variance equal to the mean, as in a family of Poisson distributions with parameters equal to the mean count of each state. The similarity in the bottom five panels reflects the temporary slowdown in new COVID-19 cases in the first half of 2021.

thanks to the central limit theorem (9). For example, Yule-type multiplicative growth and division processes converge asymptotically in time to the lognormal distribution and the Weibull distribution under different conditions (10).

So far, we have analyzed the distribution of counts over counties within each state separately. Because any single state has few counties for the purpose of discriminating between lognormal and Weibull distributions or of examining the upper tail of the

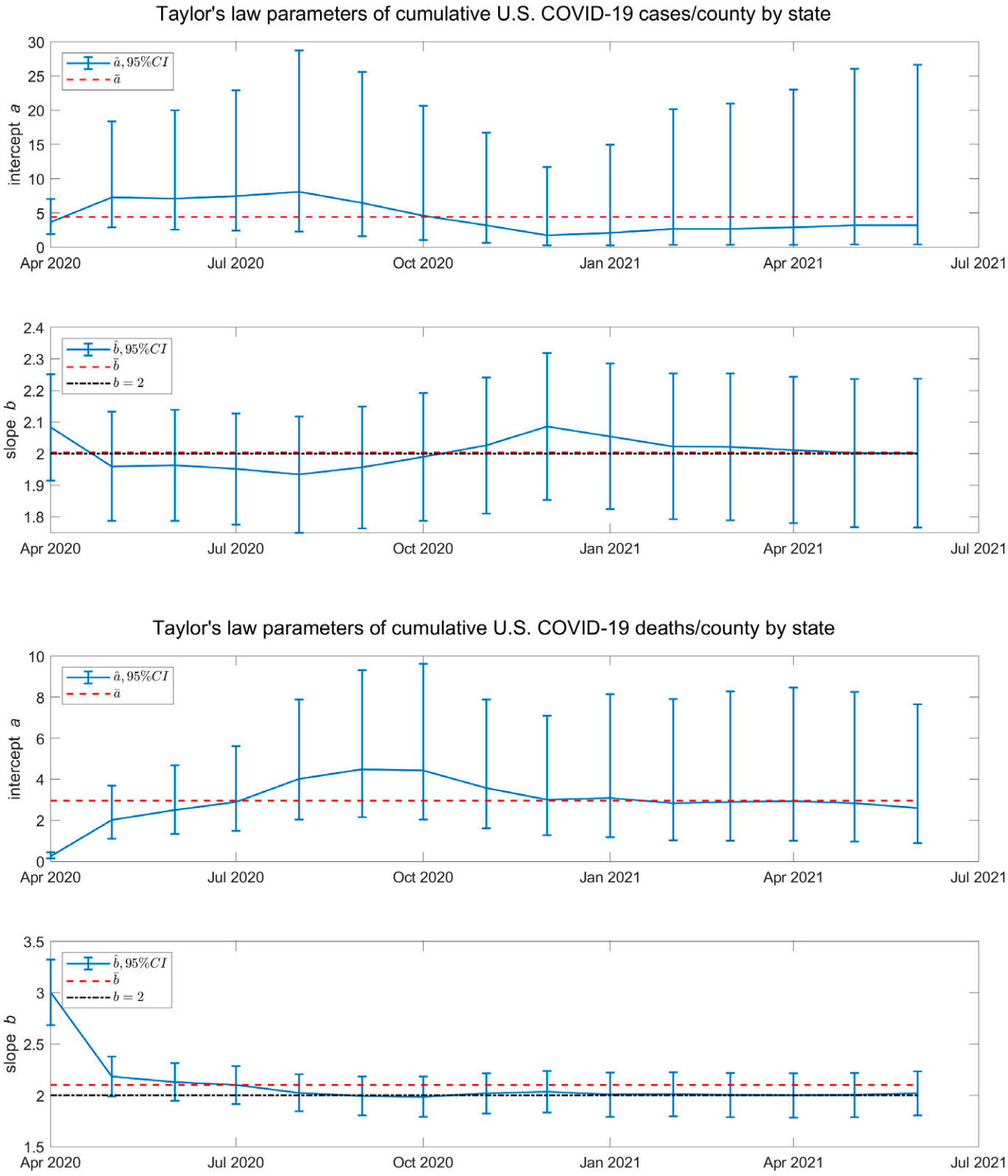

Fig. 2. Intercept *a* (panels 1 and 3) and slope *b* (panels 2 and 4) of TL Eq. **2** fitted to cumulative US COVID-19 cases (panels 1 and 2) and deaths (panels 3 and 4). The solid blue line joins the point estimate of the parameter on each date. The upper and lower error bars show the 95% CI. The horizontal dashed red line shows the average value of the point estimates of the parameter over the period of observation. In panels 2 and 4 for the slope, the horizontal dash-dotted black line shows $b = 2$. This line falls within the CI in every month (except for deaths in April 2020) and nearly coincides with the dashed red line for $\bar{b}$.

survival curve, we now (temporarily) analyze the distribution of counts over all counties within the United States. In this analysis, all counties with a positive count are considered, regardless of the state in which they occur.

Fig. 3 plots, on log–log coordinates, the empirical survival curve $\Pr(X > x)$ of the count X as a function of the positive number x for all US counties with positive count on day 1 of each of 15 mo, along with the survival curves of lognormal and Weibull

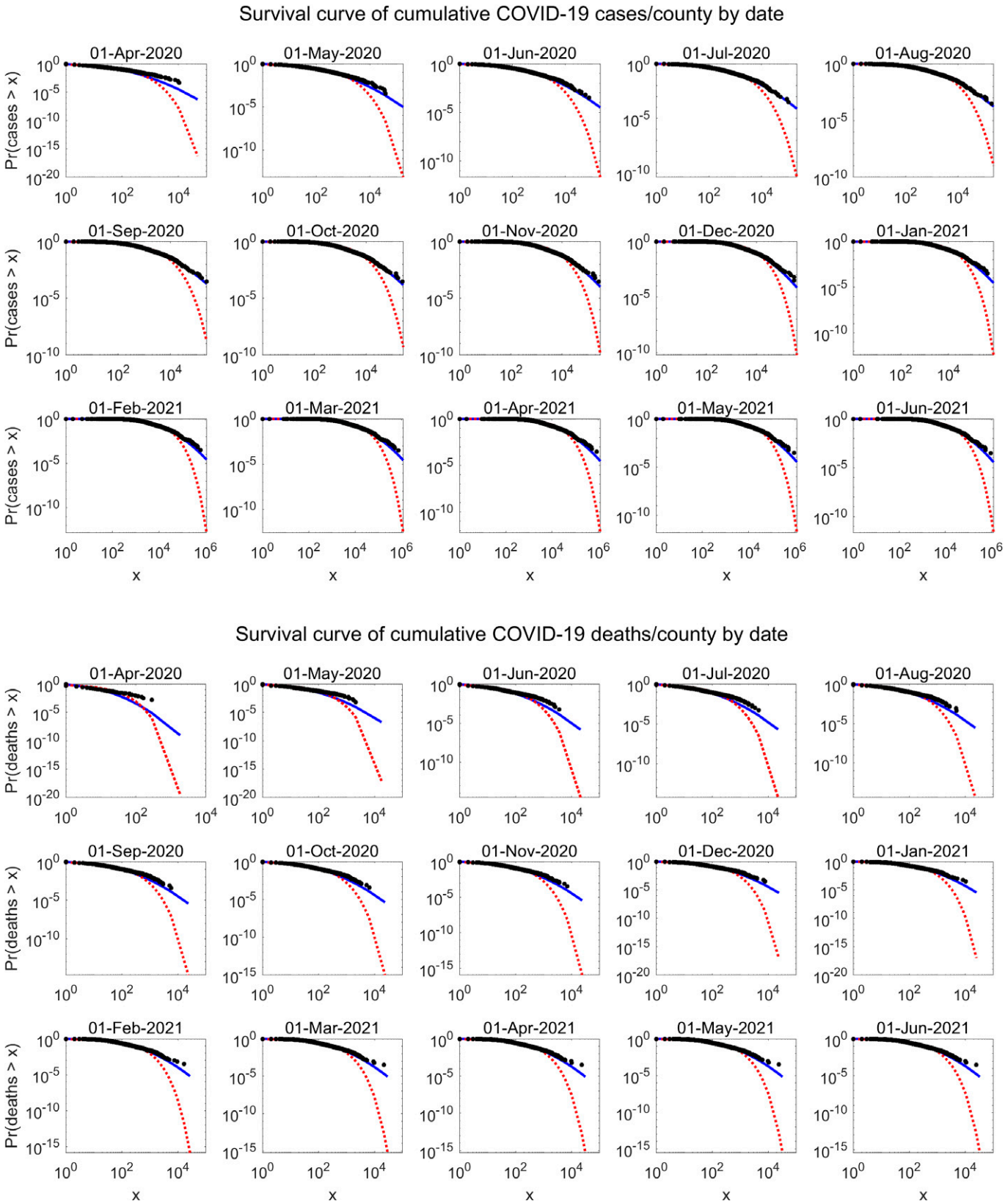

Fig. 3. Empirical survival curve $\Pr(X > x)$ as a function of $x =$ cumulative COVID-19 cases (*Upper*) and deaths (*Lower*), for all US counties included in the previous analyses, on log–log coordinates. The solid blue curve is the lognormal distribution fitted by maximum likelihood to all counties, including those in the upper tail. The dotted red curve is the Weibull distribution fitted by maximum likelihood to all counties, including those in the upper tail. Each black dot $(x, P(X > x))$ represents the count x in one U.S. county and the fraction $P(X > x)$ of all counties with a count larger than x.

distributions fitted by maximum likelihood to the counts of all the counties included in each panel. If cases were Pareto distributed, the empirical survival curve would be a downward sloping straight line on log–log coordinates. The plots are clearly concave on log–log coordinates, not linear. The curvature appears to increase as the months pass. Both the lognormal and the Weibull survival

curves describe roughly 99% or more of the empirical survival curve (corresponding to the interval from 10^0 at the top of the vertical axis down to 10^{-2}). However, the Weibull survival curves fall off much faster than both the empirical survival curves and the lognormal survival curves. Henceforth, we disregard the Weibull distribution as a model for these data.

The lognormal distribution, while far better than the Weibull, is imperfect: For the largest values of the threshold x, the lognormal survival curve always falls off faster than the empirical survival curve. This discrepancy means that extremely large counts have higher probability in the data than predicted by the best-fitting lognormal distribution. Extremely high counts are more likely than the lognormal model would predict, despite the good agreement with the lognormal model for 99% of counts. In the top percentile upper tail where the fitted lognormal drops below the data, the empirical survival curve appears to fall roughly linearly on log–log coordinates, like a Pareto distribution. We examine the top percentile upper tail of the empirical survival curve in Section 4.

A positive-valued random variable $Y(\mu, \sigma^2)$ with real parameters μ and $\sigma^2 \geq 0$ is lognormal if $\log(Y(\mu, \sigma^2))$ is normal with mean μ and variance σ^2. The mean and the variance of $Y(\mu, \sigma^2)$ are

$$E[Y(\mu, \sigma^2)] = \exp(\mu + \sigma^2/2), \quad [4]$$

$$var[Y(\mu, \sigma^2)] = [\exp(\sigma^2) - 1]\exp(2\mu + \sigma^2) = [\exp(\sigma^2) - 1]\{E[Y(\mu, \sigma^2)]\}^2. \quad [5]$$

This variance function Eq. **5** of the lognormal distribution can exhibit various behaviors, depending on the relation between σ and μ (Appendix in ref. 11). If each state's distribution of counts is well approximated by a lognormal distribution, and if the parameter μ of $Y(\mu, \sigma^2)$ changes while σ^2 remains constant (or varies little, compared to variation in μ) from state to state on a given date, then the coefficient $[\exp(\sigma^2) - 1]$ on the right side of Eq. **5** remains constant (or varies little) while the mean $E[Y(\mu, \sigma^2)]$ changes. Then the variance changes in proportion to the square of the mean. Thus, if σ^2 is constant (or varies little) and if μ varies from state to state on a given date by amounts that are large compared to the variation in σ^2, this variance function Eq. **5** obeys TL with $b = 2$.

To see whether the counts, state by state, support this proposed explanation of our empirical finding in Fig. 1 that TL holds with $b \approx 2$ in 29 of 30 cases, we return to data analysis at the state level. We fit lognormal distributions by maximum likelihood to all the positive counts per county of each state on each date separately. Fig. 4 plots, for each date, the state-specific estimates of the parameter μ on the abscissa and, on the ordinate, σ (the square root of the parameter σ^2, to get both axes on the same scale).

For cases (Fig. 4, *Upper*), the slope of a straight line fitted by ordinary least squares to the points (μ, σ), one point per state, does not differ significantly from zero in every month after April 2020. The cloud of points moves to the right from $\mu \in (0, 6)$ in April 2020 to $\mu \in (6, 11)$ in June 2021, while $\sigma \in (0, 3)$ over the entire period. At each date, and over all months, μ ranges much more widely than σ.

For deaths (Fig. 4, *Lower*), the slope of a straight line fitted by ordinary least squares to the points (μ, σ), one point per state, declines steadily from April 2020 through November 2020. From October 2020 through June 2021, the 95% CI of the slope includes zero. The cloud of points moves to the right from $\mu \in (0, 3)$ in April 2020 to $\mu \in (1, 8)$ in June 2021, while $\sigma \in (0, 3)$ over the entire period. Again, at each date, and over all months, μ ranges more widely than σ. The slope of σ as a linear function of μ exceeds zero only early in the pandemic.

To a first approximation, the model of a lognormal distribution with fixed parameter σ^2 and changing parameter μ describes the empirical lognormal parameter estimates well for cases in all months except April 2020, and well for deaths from October 2020 onward, thereby explaining why, in most cases, TL holds with $b \approx 2$.

This explanation is incomplete in at least three respects. First, the model leaves open the question of why the lognormal parameters behave as they do. Second, the model does not interpret why, in the early months of the epidemic, σ is larger in states with larger μ. For cases, the increase of σ with μ is limited to April 2020. For deaths, the increase of σ with μ is more dramatic and extends over several early months. Third, the model does not interpret why, for both counts in every month, the extreme upper tail of the empirical survival curve of all counties in the United States falls like a Pareto distribution, more slowly than the lognormal distribution.

The observation that, in the early months of the epidemic, σ is larger in states with larger μ explains why TL slopes are larger than two in the early months. When σ is larger in states with larger μ, it is obvious from the variance function Eq. **5** that the variance will increase faster than the square of the expectation, and a fitted TL will have an estimated slope larger than two. For example (Appendix in ref. 11), if $\sigma^2 = \mu$, then TL holds approximately, and its slope is $2 + 2/3$, while, if μ is constant and only σ^2 varies, then TL holds approximately with slope four, as for tornadoes (*SI Appendix*). These examples provide a qualitative insight into why, in the early months, TL slopes larger than two are associated with σ being larger in states with larger μ (for cases, April 2020; for deaths, April 2020 through roughly October 2020).

4. Lognormal–Pareto Model for the Survival Curves of Counts

The survival curves of counts in all counties of the United States (Fig. 3) are consistently well described by a lognormal distribution for the lower 99% or more of values of the count. For the highest ∼1% of counts, the empirical survival curves all fall more slowly than the fitted lognormal survival curves. On log–log coordinates, the empirical survival curves fall approximately linearly. This combination of lognormal and Pareto behavior was known for decades from graphical examples (ref. 12, pp. 110–111 and figures 1 and 2; ref. 13, p. 219 and figure 1; ref. 14, pp. 3, 5, and 6 and figures 1–3) before it was formalized and named the lognormal–Pareto distribution (15–17). The lognormal–Pareto distribution (in its version with a continuous and differentiable probability density function [pdf]) is specified by a threshold $\theta > 0$, a tail index $\alpha > 0$, and a scatter $\sigma > 0$. For $x > 0$, its pdf $f(x) := rf_1(x; \mu, \sigma^2, \theta) + (1 - r)f_2(x; \theta, \alpha)$, $r \in (0, 1)$ is a weighted sum of a lognormal distribution right-truncated at θ with pdf

$$f_1(x; \mu, \sigma^2, \theta) := \left[\Phi((\log(\theta) - \mu)/\sigma)x\sigma\sqrt{2\pi}\right]^{-1} \exp\{-((\log(\theta) - \mu)/\sigma)^2/2\}\mathbf{1}_{\{0<x\leq\theta\}}$$

and a Pareto distribution with left threshold θ and pdf $f_2(x; \theta, \alpha) := \alpha\theta^\alpha x^{-(\alpha+1)}\mathbf{1}_{\{\theta<x\}}$. Because $f(x)$ is required to be continuous and differentiable in x, parameters r and μ are functions of the other parameters (16, 17).

To examine in greater detail the upper tail of the survival curves in Fig. 3, we plot $\log \Pr(X > x)$ as a function of $\log x$ for only the counties with the highest 1% of cases (Fig. 5, *Upper*) or deaths (Fig. 5, *Lower*). The number of such counties ranges from a low of

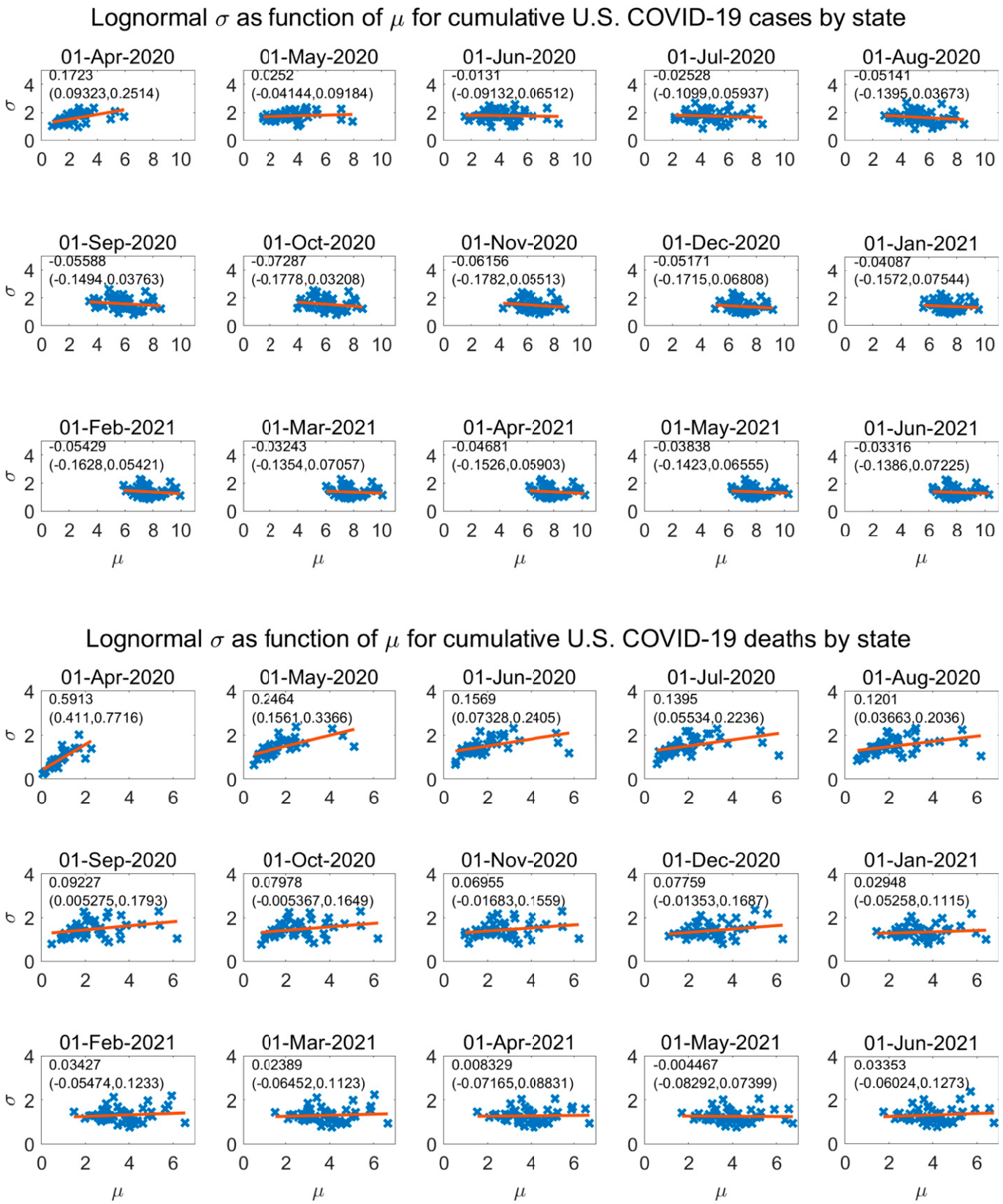

Fig. 4. Lognormal distributions are fitted by maximum likelihood to the cases (*Upper*) or deaths (*Lower*) per county within each state separately, that is, fitted to $x_{1j}, \ldots, x_{r_j j}$ separately for each state $j = 1, \ldots, c$, where r_j is the number of counties in state j. Each panel plots the lognormal parameter σ_j of each state j as a function of the lognormal μ_j (blue $\times$ markers). To a first approximation, μ_j varies over a considerably wider range than does σ_j. A solid red straight line is fitted using ordinary least squares. The slope of the fitted line is given in the upper left corner of each panel. The 95% CI, given below the slope, includes zero for all months except April 2020.

six (once only, for deaths in April 2020, before COVID-19 deaths had spread widely) to 32 (Fig. 6). On each date, we consider the counts x (of cases or deaths) of all counties regardless of state.

In each panel of Fig. 5, the horizontal axis is labeled, on the left, with the lowest count included in the largest 1% of counts and, on the right, with the highest count. On visual inspection,

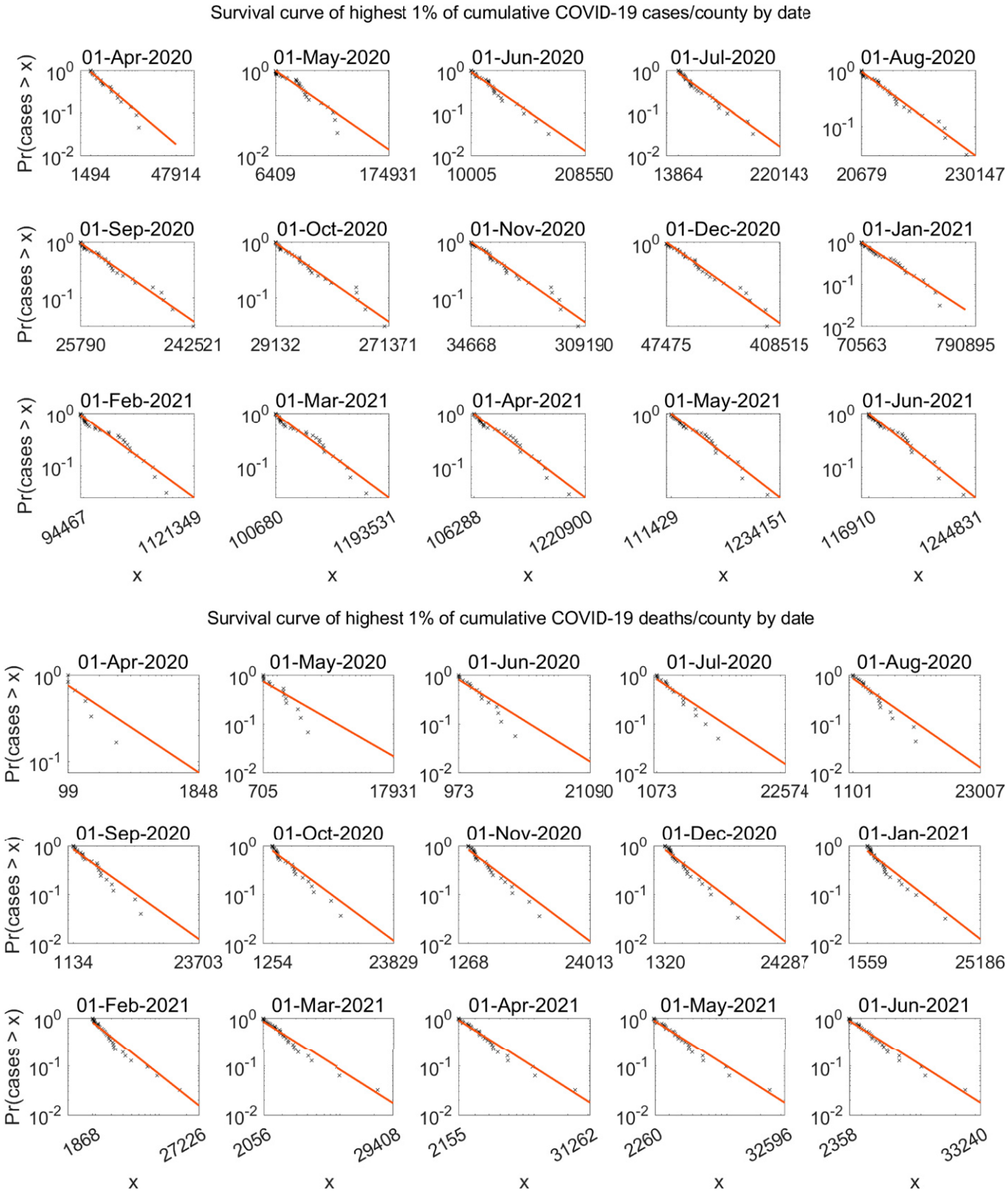

Fig. 5. Top percentile of the empirical survival curve $\Pr(X > x)$ as a function of the cumulative count x of COVID-19 cases (*Upper*) or COVID-19 deaths (*Lower*) of one county on that date, on log–log coordinates. Each county is represented by one blue $\times$. The orange straight line is fitted by the method of ref. 21.

a linear relationship of $\log \Pr(X > x)$ to $\log x$ seems reasonable, as specified in a Pareto distribution. If this linear relationship is written $\log \Pr(X > x) = \beta - \alpha \log x$, α is called the tail index (or simply the index) of the Pareto distribution. The value of the tail index determines whether the Pareto distribution (and hence also the lognormal–Pareto distribution) has finite or infinite mean (in case $\alpha > 1$ or $\alpha \in (0, 1]$, respectively) and, if the mean is finite, whether the variance is finite or infinite (in case $\alpha > 2$ or $\alpha \in (1, 2)$, respectively).

For the counties with the largest 1% of counts, we estimate α using the Hill estimator (ref. 18 and ref. 19, p. 69, section 3.2.2). Let $X_{(1)} \geq \ldots \geq X_{(m+1)}$ be the $m + 1$ largest order statistics of

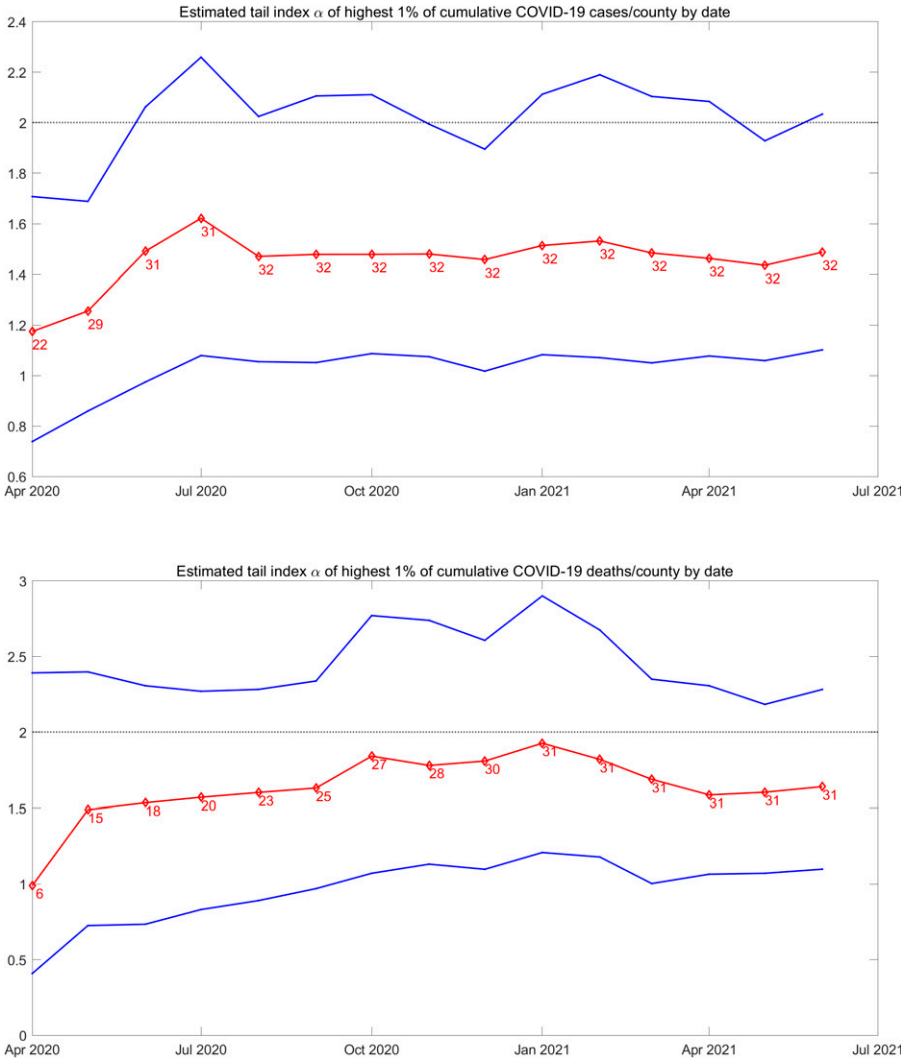

Fig. 6. For each month (horizontal axis), Hill estimates (ref. 18 and ref. 19, p. 69, section 3.2.2) (vertical axis) of the tail index α (red diamond markers, red solid line) of the empirical survival curve $\Pr(X > x)$ for the counties with the largest 1% of cases (*Upper*) and deaths (*Lower*), and 95% CIs (solid blue lines) of the point estimates of α based on 1,000 bootstrap samples with replacement of the counties with the largest 1% of counts. The number with each data point is the number of counties in the top 1% of counties on that date; it corresponds to $m + 1$ in Eq. **6**.

all counts $X_k,\ k = 1, \ldots, n$ of the n counties on a given date. Since the Pareto model describes only the largest counts, inference about α will be based on only the largest 1% of counts. The maximum likelihood estimator of α based on the largest 1% of counts on a given date is (using the natural logarithm here)

$$\hat{\alpha} = \left[\frac{1}{m} \sum_{k=1}^{m} \log(X_{(k)}) - \log(X_{(m+1)}) \right]^{-1}. \qquad [6]$$

Empirically, all estimates $\hat{\alpha}$ fall in (1, 2), with one exception (Fig. 6). For cases, the lowest $\hat{\alpha} = 1.17$ was based on the 22 counties with the highest cases among the 2,214 US counties with reported cases on 1 April 2020. For cases, all remaining $\hat{\alpha}$ are based on 29 to 32 counties and fall between 1.26 (in May 2020) and 1.62 (in July 2020). For deaths, the lowest estimate of $\hat{\alpha} = 0.99$ on 1 April 2020 was based on only six counties, the top 1% of the 573 counties in the United States with reported deaths. All remaining estimates of $\hat{\alpha}$ for deaths are based on 15 to 31 counties and fall between 1.49 (in May 2020) and 1.93 (in January 2021). When the Pareto component of a lognormal–Pareto distribution has $\alpha \in (1, 2)$, the whole distribution has a finite mean and an infinite variance.

These results pose the challenge of reconciling a Pareto distribution of the largest counts, having tail index α in $(1, 2)$, with TL, having slope or exponent $b \approx 2$. The simulations and mathematical analyses in *SI Appendix* (Theorem 3) meet this challenge.

5. Summary of Simulations and Mathematical Analyses in *SI Appendix*

SI Appendix describes simulations that give evidence that TL with slope two holds for samples with the largest means drawn from heavy-tailed distributions. One simulation assumes independent observations, which is mathematically convenient but empirically implausible. A second simulation considers moderately and highly dependent observations. *SI Appendix* then states and proves three theorems inspired by the simulations and gives an example. These results are the mathematical heart of the paper. We then review our main theoretical findings and some of their limitations.

6. Discussion

Here we review and discuss our main empirical findings and some of their limitations. We also propose a practical consequence of our findings for planning the management of COVID-19

cases and deaths, especially in largely unvaccinated populations. *SI Appendix* gives some examples of other empirical data that could be, but have not yet been, analyzed using our approach in this paper.

We have demonstrated striking regularities in the variability among counties within states of reported counts of cumulative SARS-CoV-2 cases and cumulative COVID-19 deaths. In the United States, on the first day of each of 15 mo from April 2020 through June 2021, omitting the first few months of the pandemic, the variance of cases and, separately, deaths is nearly proportional to the square of the mean count of cases or deaths (across the counties within each state or other primary administrative unit of the United States). The approximately power-law relationship of the variance to the mean illustrates TL, a widespread pattern in ecology and epidemiology. To our knowledge, this pattern has not been recognized previously for SARS-CoV-2 cases and COVID-19 deaths.

The estimated slope b of TL closely approximates two after the first few months, for both cases and deaths. The slope $b = 2$ is the only positive value of b such that the coefficient of variation (SD divided by mean) is the same for all samples. Only for slope $b = 2$ is TL Eq. **2** scale invariant, with the same parameters a, b regardless of the scale on which observations are measured (formally, if $\sigma_X^2 = a\mu_X^2$ and $c > 0$, then $\sigma_{cX}^2 = a\mu_{cX}^2$). For example, if $b = 2$, then the parameters a, b are the same whether counts are measured in terms of individuals or millions of individuals.

We estimate that the largest 1% of counts of cases and deaths by county on each date are approximately Pareto distributed and that the upper tail of the empirical survival curve of counts by county has a tail index α between one and two. If that is correct, the underlying distributions of cases and deaths, except for, possibly, deaths in April 2020, have finite mean and infinite variance. This finding has implications for planning prevention and care: Facility and resource planning should prepare for rare but extremely high counts. No single county, state, region, or country can prepare in isolation for unboundedly high counts. Cooperative exchanges of support should be planned cooperatively.

Beare and Toda (20) analyzed *The New York Times* cumulative counts of cases by county on 31 March 2020, the day before the first observation we use here. The left portion of their survival curve, $\log \Pr\{X > x\}$ as a function of $\log x$, is concave, and hence not Pareto: Figure 1 of ref. 20 qualitatively resembles each panel in our Fig. 5. They found that the upper tail of the distribution of the number of cases for the top 6.2% of counties by number of cases reasonably approximated a straight line (on log–log coordinates) with estimated tail index 0.930 (SE 0.081). Hence their estimated tail index did not differ from one but was clearly less than two. Our estimate of the tail index of the top 1% of counties by number of cases on 1 April 2020 (Fig. 5, first panel), lies between 1.1 and 1.2, not far from the results of Beare and Toda (20). Our calculations were completed before we learned of ref. 20. In all months (except for deaths in April 2020), our estimated tail index of cases is greater than one but less than two.

Heavy-tailed distributions have also been used in modeling COVID-19 superspreader events, in which a primary infected individual infects an exceptionally large number of secondary individuals. That different use is not pursued here.

Our data analysis suffers from multiple limitations, some due to the data used and some due to our analyses. For example, we made no effort to estimate or correct underreporting of cases or deaths. We did not examine case fatality rates. We did not relate counts to the population at risk, whether treated as a population total or adjusted for age structure.

6.1. Data, Materials, and Software Availability. Previously published data were used for this work (https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties.csv).

ACKNOWLEDGMENTS. J.E.C.'s research was partially supported by Columbia University's Earth Institute. R.A.D.'s research was partially supported by NSF Grant DMS 2015379 to Columbia University. G.S.'s research was partially supported by NSF Grant DMS-2015242 to Cornell University.

Author affiliations: [a]Laboratory of Populations, The Rockefeller University & Columbia University, New York, NY 10065; [b]Earth Institute, Columbia University, New York, NY 10027; [c]Department of Statistics, Columbia University, New York, NY 10027; [d]Department of Statistics, University of Chicago, Chicago, IL 60637; and [e]School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853

1. Worldometer, *Countries where COVID-19 has spread*. https://www.worldometers.info/coronavirus/countries-where-coronavirus-has-spread/. Accessed 9 May 2022.
2. New York Times, *nytimes covid-19-data public*. https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties.csv. Accessed 19 June 2021.
3. S. K. Bar-Lev, P. Enis, Reproducibility and natural exponential families with power variance functions. *Ann. Stat.* **14**, 1507–1522 (1986).
4. S. K. Bar-Lev, O. Stramer, Characterizations of natural exponential families with power variance functions by zero regression properties. *Probab. Theory Relat. Fields* **76**, 509–522 (1987).
5. M. Davidian, R. J. Carroll, Variance function estimation. *J. Am. Stat. Assoc.* **82**, 1079–1091 (1987).
6. Z. Eisler, I. Bartos, J. Kertész, Fluctuation scaling in complex systems: Taylor's law and beyond. *Adv. Phys.* **57**, 89–142 (2008).
7. R. A. Taylor, *Taylor's Power Law: Order and Pattern in Nature* (Academic, 2019).
8. J. E. Cohen, Stochastic population dynamics in a Markovian environment implies Taylor's power law of fluctuation scaling. *Theor. Popul. Biol.* **93**, 30–37 (2014).
9. J. Aitchison, J. Brown, *The Lognormal Distribution with Special Reference to Its Uses in Economics* (Cambridge University Press, Cambridge, 1957).
10. S. Goh, H. W. Kwon, M. Y. Choi, Discriminating between Weibull distributions and log-normal distributions emerging in branching processes. *J. Phys. A Math. Theor.* **47**, 225101 (2014).
11. M. K. Tippett, J. E. Cohen, Tornado outbreak variability follows Taylor's power law of fluctuation scaling and increases dramatically with severity. *Nat. Commun.* **7**, 10668 (2016).
12. W. W. Badger, "An entropy-utility model for the size distribution of income" in *Mathematical Models as a Tool for the Social Sciences*, B. J. West, Ed. (Gordon and Breach, New York, 1980), pp. 87–120.
13. E. W. Montroll, M. F. Shlesinger, Maximum entropy formalism, fractals, scaling phenomena, and 1 $/f$ noise: A tale of tails. *J. Stat. Phys.* **32**, 209–230 (1983).
14. W. Souma, "Physics of personal income" in *Empirical Science of Financial Fluctuations: The Advent of Econophyscs*, H. Takayasu, Ed. (Springer Japan KK, Tokyo, 2002), pp. 342–352.
15. K. Cooray, M. M. Ananda, Modeling actuarial data with a composite lognormal-Pareto model. *Scand. Actuar. J.* **2005**, 321–334 (2005).
16. D. P. M. Scollnik, On composite lognormal-Pareto models. *Scand. Actuar. J.* **2007**, 321–334 (2007).
17. M. Bee, Estimation of the lognormal-Pareto distribution using probability weighted moments and maximum likelihood. *Commun. Stat. Comput* **44**, 2040–2060 (2015).
18. B. M. Hill, A simple general approach to inference about the tail of a distribution. *Ann. Stat.* **3**, 1163–1174 (1975).
19. L. de Haan, A. Ferreira, *Extreme Value Theory: An Introduction* (Springer, New York, 2006).
20. B. K. Beare, A. A. Toda, On the emergence of a power law in the distribution of COVID-19 cases. *Physica D* **412**, 132649 (2020).
21. X. Gabaix, R. Ibragimov, Rank minus 1/2: A simple way to improve the OLS estimation of tail exponents. *J. Bus. Econ. Stat.* **29**, 24–39 (2011).

PNAS
www.pnas.org

Supplementary Information for

COVID-19 cases and deaths in the United States follow Taylor's law for heavy-tailed distributions with infinite variance

Joel E. Cohen, Richard A. Davis and Gennady Samorodnitsky

Corresponding Author Joel E. Cohen.
E-mail: cohen@rockefeller.edu

This PDF file includes:

Supplementary text
Figs. S1 to S2
Legend for Dataset S1
SI References

Other supplementary materials for this manuscript include the following:

Dataset S1

Supporting Information Text

In this Appendix, Section 1 describes simulations of heavy-tailed counts that give evidence that TL with slope 2 holds for the largest means and variances. One simulation assumes independent observations, which is mathematically convenient but empirically implausible. A second simulation considers moderately and highly dependent observations. Section 2 states three theorems inspired by the simulations and gives an example as a special case of one of the theorems. Section 3 proves the theorems. Section 4 gives some examples of other empirical data that could be, but have not yet been, analyzed using our approach in this paper.

1. Simulating multiple samples from heavy-tailed distributions

Here we report simulations that strongly suggest that TL (Main text Eq. [2]) with slope $b \approx 2$ holds in the extreme upper tail of a heavy-tailed (including but not limited to Pareto) distribution of counts with tail index α between 0 and 2. Our idealized statistical model assumes $c > 1$ states, each with $r = c$ counties. This idealization ignores that, in reality, different states have different numbers of counties and that the number of states only approximates the average number of counties per state.

We consider a general setting of a data matrix for our simulations and theoretical results. Specifically, let $\mathbf{X} := (X_{ij})_{r\times c}$ be an $r \times c$ matrix. (Mnemonic: r for rows, here the number of counties per state; c for columns, here the number of states.) We interpret the jth column $X_{:j}$ of $\mathbf{X}$ as the counts in the counties of state j, for $j = 1, \ldots, c$. Collectively, the columns of $\mathbf{X}$ give a collection of c samples (one sample per state), each sample containing r observations, here COVID-19 counts (one observation per county). To study the asymptotic theory behind TL, we let the sample size in each column grow without limit, $r \to \infty$, and c converges to infinity with r. We emphasize this relation by writing $c = c(r)$. We have one such matrix for cases, another for deaths. We assume that the element (or count) X_{ij} in row i and column j is a positive regularly varying (RV) rv with (upper tail) index α between 0 and 2.

By definition, X is regularly varying with index α (and we write $X \in RV(\alpha)$) if and only if

$$\forall t > 0, \lim_{x\to\infty} \frac{\Pr(X > tx)}{\Pr(X > x)} = \frac{1}{t^\alpha}. \qquad [1]$$

The RV rvs include the lognormal-Pareto, the Pareto, the stable, and many other distributions.

The sample mean and the sample variance of the jth column $X_{:j}$ of $\mathbf{X}$ are, by definition,

$$\bar{X}_{:j} := r^{-1}\sum_{i=1}^{r} X_{ij}, \quad \widehat{\mathrm{var}}(X_{:j}) := (r-1)^{-1}\sum_{i=1}^{r}(X_{ij} - \bar{X}_{:j})^2, \qquad j = 1, \ldots, c. \qquad [2]$$

If this collection of c sample means and c sample variances approximates Main text Eq. [2], then we say that TL holds.

We simulate $RV(\alpha)$ rvs in four ways to confirm that the extreme upper tail index is independent of the regularly varying distribution being simulated. Let N be normally distributed with mean 0 and variance 1 and let U, U_1, U_2, U_3 be independent rvs uniformly distributed on $(0, 1)$. For every element X_{ij} of the $r \times c$ matrix $\mathbf{X}$ independently, we compute

$$X_{ij} \stackrel{d}{=} |N|^{-1/\alpha} \in RV(\alpha) \qquad \text{(power of } |N|) \qquad [3]$$

$$X_{ij} \stackrel{d}{=} U^{-1/\alpha} \in RV(\alpha) \qquad \text{(Pareto with domain } [1, \infty)) \qquad [4]$$

$$X_{ij} \stackrel{d}{=} (U_1U_2)^{-1/\alpha} \in RV(\alpha) \qquad \text{(product of two iid Paretos)} \qquad [5]$$

$$X_{ij} \stackrel{d}{=} (U_1U_2U_3)^{-1/\alpha} \in RV(\alpha) \qquad \text{(product of three iid Paretos)} \qquad [6]$$

We first verify that all four methods are in $RV(\alpha)$. Method Eq. (3) is in $RV(\alpha)$ by (1, p. 10, Corollary 4.4). The product Eq. (5) of two iid Pareto rvs X, Y with left threshold 1 and tail index α is in $RV(\alpha)$ because $P(XY > x) = (\alpha \log x + 1)x^{-\alpha}$. The product Eq. (6) of three (or any finite number of) iid Pareto rvs with tail index α is in $RV(\alpha)$ by (2, p. 4499, Lemma 3).

For the simulations in Figure S1, we set $\alpha = 1/2, 1, 3/2$ and $r = c = 100$ and 1000. In general, the estimated slope on log-log coordinates of the five largest (mean, variance) pairs is close to 2, and is closer to 2 for the larger sample size and the smaller α.

This model assumes statistical independence both among observations within columns (counts of counties within states) and among observations in different columns (counts of counties in different states). This mathematically convenient assumption seems empirically implausible.

In the next simulation, instead of rc independent modeled counts X_{ij}, we simulate dependent modeled counts X_{ij}. As for the simulations in Figure S1, we set $\alpha = 1/2,\ 1,\ 3/2$ and $r = c = 100,\ 1000$. In Figure S2, each observation is the product of a Pareto rv with tail index α times the exponential of a $Normal(0, 1)$ rv that is correlated with every other $Normal(0, 1)$ rv in the same simulation with correlation $\rho = 0,\ 0.5,\ 0.9$. This model is analyzed in Theorem 2 in Section 2 below.

The simulation results in Figure S2 suggest, as before, that regardless of the correlation $\rho \in [0, 1)$, regardless of $\alpha \in (0, 2)$, in some limit as $r = c \to \infty$, the extreme upper tail of the variance function is increasingly well approximated by TL (Main text Eq. [2]) with intercept $a = \log_{10} r$ and slope $b = 2$. The following mathematical results make this claim more precise and prove it (apart from the special case $\alpha = 1$).

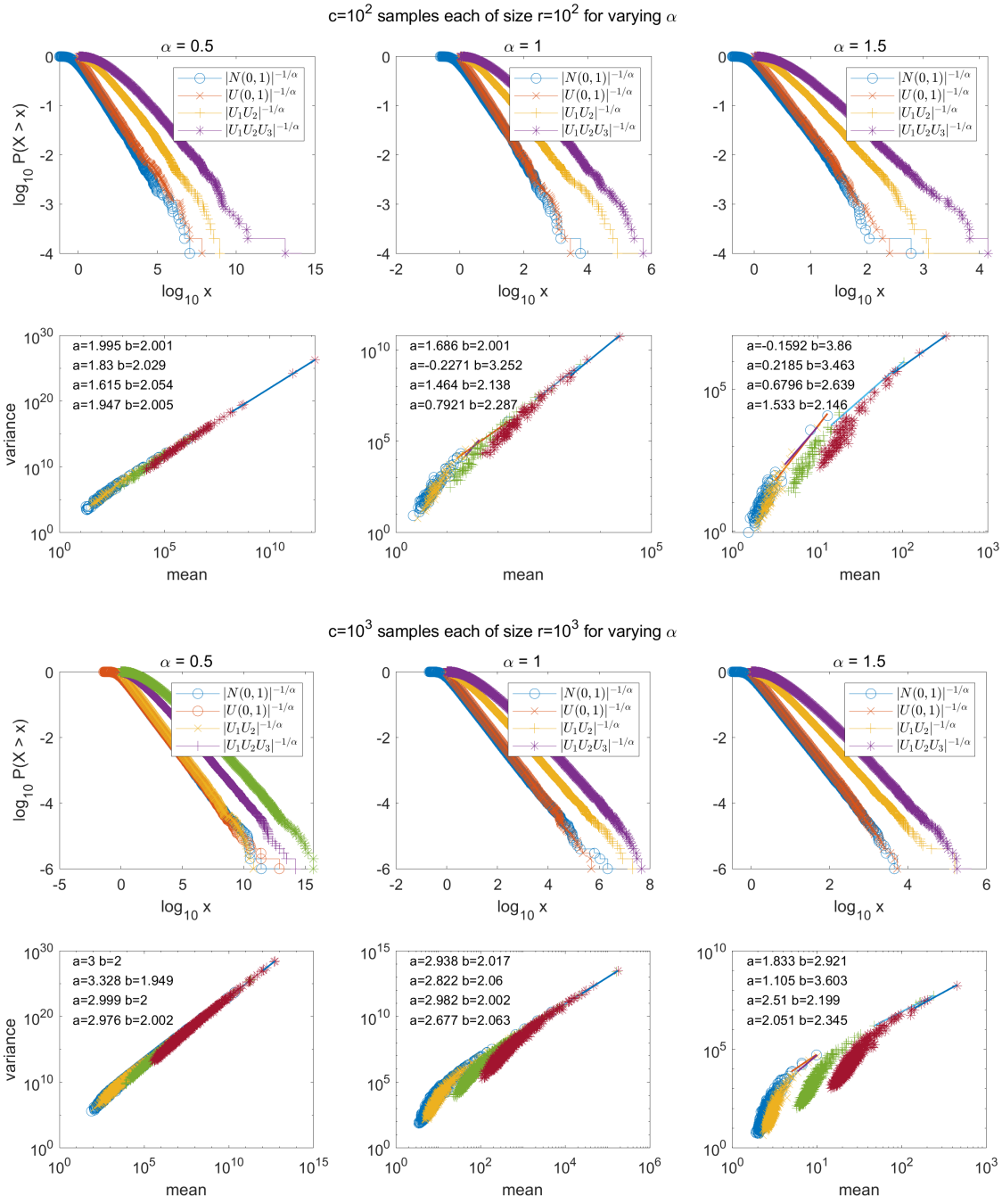

Fig. S1. For *independent* observations (counts), the empirical survival curves (rows 1, 3) of one simulation of $rc = 10^4$ (row 1) or $rc = 10^6$ (row 3) observations by each of four methods of simulation Eq. (3)–Eq. (6); and sample variance functions (rows 2, 4) of one simulation of $\mathbf{X}$ of size $r \times c = 10^2 \times 10^2$ (row 2) or $r \times c = 10^3 \times 10^3$ (row 4) by each of the same four methods of simulation. The vector of rc simulated values used for each empirical survival curve is rearranged into an $r \times c$ matrix $\mathbf{X}$ for each variance function, but each method, sample size, and value of α is simulated independently. Straight lines are fitted to the five observations with the largest means. The intercept a and slope b of each line are in the top left corner of each variance-function panel in the same order as the four methods of simulation are listed in the top right corner of each empirical survival curve panel.

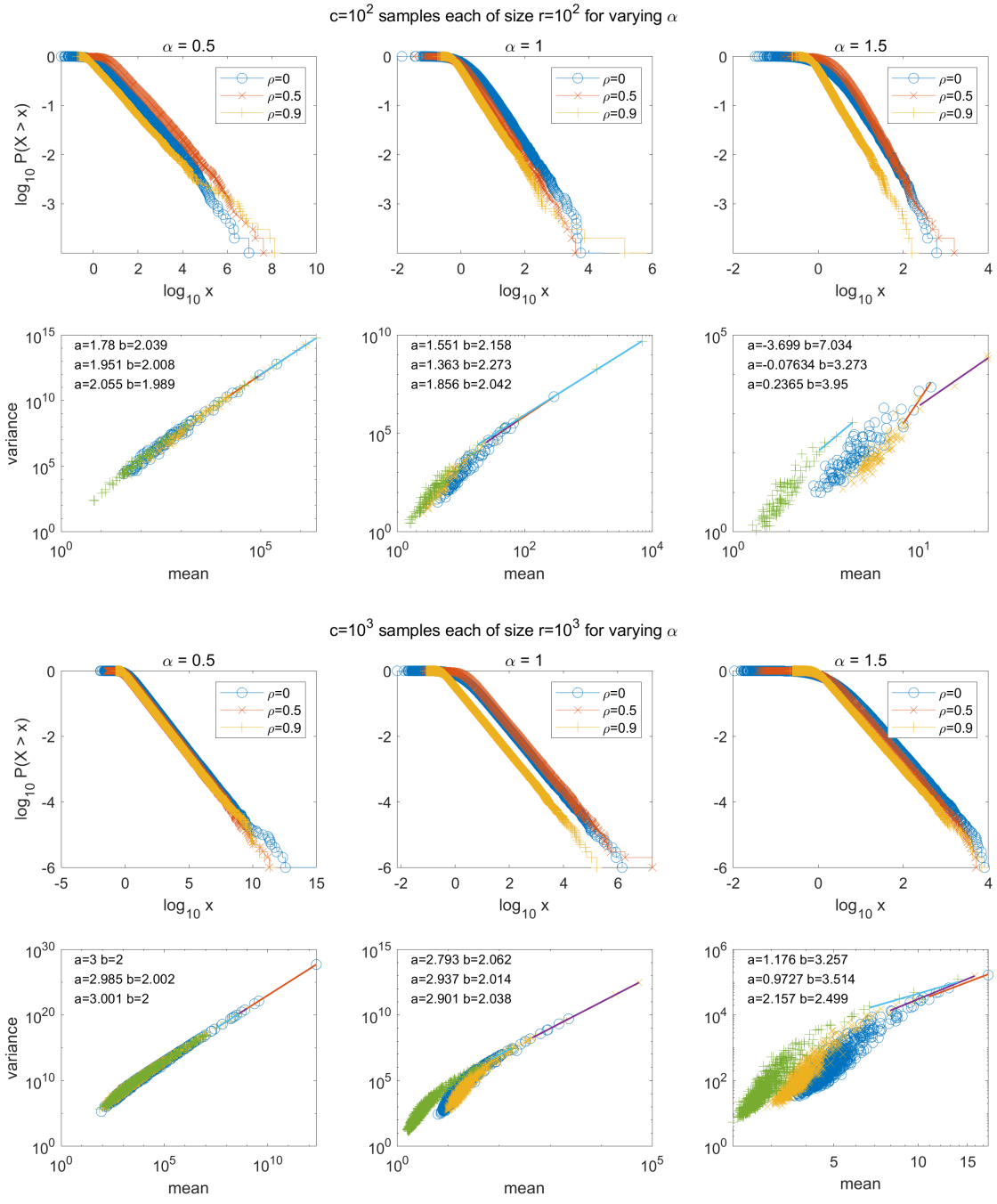

Fig. S2. For *dependent* observations (counts), the empirical survival curves (rows 1, 3) of one simulation of $rc = 10^4$ (row 1) or $rc = 10^6$ (row 3) observations and sample variance functions (rows 2, 4) of one simulation of $\mathbf{X}$ of size $r \times c = 10^2 \times 10^2$ (row 2) or $r \times c = 10^3 \times 10^3$ (row 4). In each panel, each observation is the product of a Pareto rv with tail index α times the exponential of a $Normal(0, 1)$ rv that is correlated with every other $Normal(0, 1)$ rv in the same $\mathbf{X}$ with correlation $\rho = 0,\ 0.5,\ 0.9$. This model is analyzed in Theorem 2. The vector of rc simulated values used for each empirical survival curve is rearranged into an $r \times c$ matrix $\mathbf{X}$ for each variance function, but each correlation, sample size, and value of α is simulated independently. Straight lines are fitted to the five observations with the largest means. The intercept a and slope b of each line are in the top left corner of each variance-function panel in the same order as the correlations listed in the top right corner of each empirical survival curve panel.

2. Mathematical theorems and example

This mathematical section supports, clarifies, and extends the conjectures based on the simulations just reported. The theorems demonstrate that TL (Main text Eq. [2]) with slope $b = 2$ describes the variance function of (log mean, log variance) pairs with the largest sample means in a sufficiently large matrix of observations of nonnegative $RV(\alpha)$ rvs, $\alpha \in (0,1) \cup (1,2)$, asymptotically as the sample size and the magnitude of the sum of the observations get large.

The case $\alpha = 1$ remains open mathematically, although the simulation results suggest that $\alpha = 1$ is not qualitatively different from $\alpha \in (0,1) \cup (1,2)$. The theorems also show that TL describes the variance function of the upper tails of means and variances in the presence of certain forms of dependence and heterogeneity of marginal distributions between and within samples. Proofs are in Section 3.

In this Section 2, log refers to the natural logarithm. The base of the logarithm has no effect on the slope b in TL (Main text Eq. [2]), though the intercept does depend on the base. We assume the background on regularly varying and stable laws in (1). The sample mean and the sample variance of the jth column $X_{:j}$ of $\mathbf{X}$ are defined in Eq. (2) in Section 1.

Theorems 1 and 2 assume $0 < \alpha < 1$. Theorem 1 assumes that the elements of the matrix $\mathbf{X}$ are identically distributed, that different columns are identically distributed, and that elements within each column are, conditionally on a σ-field $\mathcal{G}$, independent. Theorem 1, Eq. (9), guarantees that, with probability converging to one, TL (Main Text Eq. [2]) holds with $b = 2$ as $r \to \infty$ for those columns $X_{:j}$ of $\mathbf{X}$ with sample sums $r\bar{X}_{:j}$ that exceed some large threshold $x(r)$. The theshold $x(r)$ will converge to infinity at some rate as specified in the theorems.

Theorem 3 assumes $1 < \alpha < 2$, an assumption consistent with our analysis of COVID-19 counts, so the expectation of any one observation is finite. Then Eq. (12) guarantees the same version of TL when the sum of the sample elements is sufficiently larger than the expectation of any one observation.

Theorem 1. *Let every element $X_{ij}, i = 1, \ldots, r; j = 1, \ldots, c$ of the array $\mathbf{X}$ be nonnegative and have the same distribution in $RV(\alpha)$, $0 < \alpha < 1$, and assume that all columns are equally distributed. Assume also that the entries within each column are, conditionally on a σ-field $\mathcal{G}$, independent. Let $x(r)$ be a sequence of thresholds satisfying*

$$\lim_{r\to\infty} x(r) = \infty \qquad \textit{and} \qquad \lim_{r\to\infty} r \Pr(X_{11} > x(r)) = 0. \tag{7}$$

Let the number c of columns depend on the number r of rows in such a way that the function $c = c(r) \to \infty$ and satisfies

$$\lim_{r\to\infty} c(r) r^2 E\left[\max_{i=1,\ldots,r}\left(\left[\frac{1}{x(r)}\int_0^{x(r)} \Pr\big(X_{i,1} > x \;\big|\; \mathcal{G}\big)\, dx\right]^2\right)\right] = 0. \tag{8}$$

Then for any $\varepsilon > 0$,

$$\lim_{r\to\infty} \Pr\left(\left|\log \frac{\widehat{\mathrm{var}}(X_{:j})}{r(\bar{X}_{:j})^2}\right| > \varepsilon \textit{ for some } \; j = 1, \ldots, c(r) \textit{ such that } r\bar{X}_{:j} > x(r)\right) = 0. \tag{9}$$

Remark: As noted above, Eq. (9) implies that with probability approaching one, $\log \widehat{\mathrm{var}}(X_{:j}) \approx \log(r) + 2\log(\bar{X}_{:j})$ for those columns j in which the sample mean is large, i.e., $\bar{X}_{:j} > x(r)/r$.

Example: a special case of Theorem 1. For $0 < \alpha < 1$, let X be Pareto with survival function $\Pr(X > x) = x^{-\alpha}$ for $x \geq 1$. Let every element $X_{ij}, i = 1, \ldots, r; j = 1, \ldots, c$ of the array $\mathbf{X}$ be iid as X. Then an example of $x(r)$ is $x(r) = r^p$ for any $p > 1/\alpha > 1$, because if $p > 1/\alpha$, then $0 > 1 - \alpha p$ and

$$r \Pr(X > x(r)) = r \Pr(X > r^p) = r(r^p)^{-\alpha} = r^{1-\alpha p} \to 0 \text{ as } r \to \infty.$$

This satisfies Eq. (7). An example of $c(r)$ is $c(r) = r^q$ for any $0 < q < 2\alpha p - 2$ (and here $2\alpha p - 2 > 0$ because $p > 1/\alpha$), since under these assumptions, in Eq. (8) we have, with $\mathcal{G}$ being the trivial σ-field, $\Pr\big(X_{i,1} > x \;\big|\; \mathcal{G}\big) = \Pr(X_{i,1} > x)$ and hence

$$\begin{aligned} &E\left(\max_{i=1,\ldots,r}\left[\frac{1}{x(r)}\int_0^{x(r)} \Pr\big(X_{i,1} > x \;\big|\; \mathcal{G}\big)\, dx\right]^2\right) \\ &= \left[\frac{1}{x(r)}\int_0^{x(r)} \Pr\big(X > x\big)\, dx\right]^2 = \left[\frac{1}{r^p}\int_1^{r^p} x^{-\alpha}\, dx\right]^2 \\ &= \left[\frac{1 - r^{p(\alpha-1)}}{(1-\alpha)r^{p\alpha}}\right]^2 \end{aligned}$$

and

$$c(r) r^2 \left[\frac{1 - r^{p(\alpha-1)}}{(1-\alpha)r^{p\alpha}}\right]^2 \sim \frac{(r^q)(r^2)(r^{-2\alpha p})}{(1-\alpha)^2} = \frac{r^{q+2(1-\alpha p))}}{(1-\alpha)^2} \to 0 \text{ as } r \to \infty.$$

This satisfies Eq. (8).

The next theorem is another illustration of Theorem 1.

Theorem 2. *Let* $\{Z_{ij} \mid i = 1, \dots, r;\ j = 1, \dots, c\}$ *be iid Pareto-distributed with tail index* $\alpha \in (0, 1)$. *Let* $\{G_{ij}\}$ *be Gaussian with mean* 0 *and covariance function* $\gamma(\cdot, \cdot)$, *i.e.,* $Cov(G_{ij}, G_{i'j'}) = \gamma(i - i', j - j')$, *independent of the Pareto random variables. Given* $\mathcal{G} = \sigma(G_{ij},\ i = 1, \dots, r;\ j = 1, \dots, c)$, *the process* $X_{ij} := Z_{ij}e^{G_{ij}}$ *is conditionally independent. Then* X_{ij} *are identically distributed and regularly varying with index* α. *Moreover, if* $x(r)$ *and* $c(r)$ *are chosen as in the Example above, then* Eq. (9) *holds.*

The process $X_{ij} := Z_{ij}e^{G_{ij}}$ is sometimes referred to as a stochastic volatility (SV) random field. In this case, the volatility process $\sigma_{ij} = e^{G_{ij}}$ is assumed to be a stationary log-Gaussian random field, while the noise terms Z_{ij} are iid and heavy-tailed. The random field version generalizes the notion of a SV process that is commonly used for modeling financial time series such as log-returns, e.g., (3). Basic properties of heavy-tailed SV processes can be found in (4) and extensions of these models to the spatial (random field) setting in (5) and (6); the latter includes environmental applications. The idea is to scale the noise Z_{ij} by a random process $\sigma_{ij} = e^{G_{ij}}$ that allows dependence among the rows and columns of the data matrix X. While the SV random field is strictly stationary, the sample paths can have very non-stationary looking realizations. This feature makes them flexible models for many phenomena.

Theorem 3. *Suppose the elements* X_{ij} *of the* $r \times c$ *matrix* $\mathbf{X}$ *are nonnegative, iid, and in* $RV(\alpha)$, $1 < \alpha < 2$. *Let* $x(r)$ *satisfy*

$$\lim_{r\to\infty} x(r) = \infty \qquad \textit{and} \qquad \lim_{r\to\infty} r \Pr(X_{11} > x(r)) = 0. \tag{10}$$

Let the number c *of columns depend on the number* r *of rows in such a way that the function* $c = c(r) \to \infty$ *and satisfies*

$$\lim_{r\to\infty} c(r) r^2 \big(\Pr(X_{11} > x(r))\big)^2 = 0. \tag{11}$$

Then for any $\varepsilon > 0$,

$$\lim_{r\to\infty} \Pr\left(\left|\log \frac{\widehat{\mathrm{var}}(X_{:j})}{r(\bar{X}_{:j})^2}\right| > \varepsilon \textit{ for some } \ j = 1, \dots, c(r) \textit{ such that } r\big(\bar{X}_{:j} - EX_{11}\big) > x(r)\right) = 0. \tag{12}$$

3. Proofs of mathematical theorems

A. Proof of Theorem 1.

Proof. The probability in Eq. (9) does not exceed

$$\begin{aligned} & c(r) \Pr\left(\left|\log \frac{\widehat{\mathrm{var}}(X_{:1})}{r(\bar{X}_{:1})^2}\right| > \varepsilon,\ r\bar{X}_{:1} > x(r)\right) \\ =& c(r) \Pr\left(\left|\log\left[\frac{\sum_{i=1}^r X_{i,1}^2}{\left(\sum_{i=1}^r X_{i,1}\right)^2} - \frac{1}{r}\right]\right| > \varepsilon,\ \sum_{i=1}^r X_{i,1} > x(r)\right) \\ \le& c(r) \Pr\left(\log \frac{\left(\sum_{i=1}^r X_{i,1}\right)^2}{\sum_{i=1}^r X_{i,1}^2} > \varepsilon/2,\ \sum_{i=1}^r X_{i,1} > x(r)\right), \end{aligned} \tag{13}$$

with the last inequality valid for large r.

Next we prove that

$$\lim_{r\to\infty} c(r) \Pr\left(\sum_{i=1}^r X_{i,1} > x(r),\ X_{i,1} \le \frac{x(r)}{4},\ i = 1, \dots, r\right) = 0. \tag{14}$$

Indeed, let

$$\begin{aligned} T_1 &= \inf\left\{n = 1, 2, \dots, r : \sum_{i=1}^n X_{i,1} > \frac{x(r)}{4}\right\}, \\ T_2 &= \inf\left\{n = T_1 + 1, \dots, r : \sum_{i=T_1+1}^n X_{i,1} > \frac{x(r)}{4}\right\}, \end{aligned}$$

and $T_2 = \infty$ if either infimum is over an empty set. Then the probability in Eq. (14) is bounded above by

$$\begin{aligned} & \Pr\left(T_2 < \infty,\ X_{i,1} \le \frac{x(r)}{4},\ i = 1, \dots, r\right) \\ \le& E\left[\Pr\left(\sum_{i=1}^r X_{i,1} > \frac{x(r)}{4},\ X_{i,1} \le \frac{x(r)}{4},\ i = 1, \dots, r \,\middle|\, \mathcal{G}\right)\right]^2. \end{aligned} \tag{15}$$

Using Markov's inequality,

$$\begin{aligned}\Pr\left(\sum_{i=1}^{r} X_{i,1} > \frac{x(r)}{4},\ X_{i,1} \le \frac{x(r)}{4},\ i=1,\dots,r \,\middle|\, \mathcal{G}\right) &\le \Pr\left(\sum_{i=1}^{r} X_{i,1}\mathbf{1}\left(X_{i,1} \le \frac{x(r)}{4}\right) > \frac{x(r)}{4} \,\middle|\, \mathcal{G}\right)\\ &\le \frac{E\left[\sum_{i=1}^{r} X_{i,1}\mathbf{1}\left(X_{i,1} \le x(r)/4\right) \,\middle|\, \mathcal{G}\right]}{(x(r)/4)}\\ &\le \frac{4r}{x(r)} \max_{i=1,\dots,r} E\left[X_{i,1}\mathbf{1}\left(X_{i,1} \le \frac{x(r)}{4}\right) \,\middle|\, \mathcal{G}\right]\\ &\le \frac{4r}{x(r)} \max_{i=1,\dots,r} \int_0^{x(r)/4} \Pr\big(X_{i,1} > x \mid \mathcal{G}\big)dx.\end{aligned}$$

Now Eq. (14) follows from Eq. (15) and Eq. (8).

Next, for any $\delta > 0$,

$$\begin{aligned}&c(r)\Pr\big(X_{i,1} > \delta x(r) \text{ for 2 or more of } i=1,\dots,r\big) \qquad [16]\\ \le& c(r)\frac{r(r-1)}{2} \max_{i=1,\dots,r} E\big[\Pr\big(X_{i,1} > \delta x(r) \mid \mathcal{G}\big)\big]^2\\ \le& c(r)r^2 \max_{i=1,\dots,r} E\left[\frac{1}{\delta x(r)}\int_0^{\delta x(r)} \Pr\big(X_{i,1} > x \mid \mathcal{G}\big)\,dx\right]^2 \to 0\end{aligned}$$

as $r \to \infty$ by assumption Eq. (8).

It follows from Eq. (13), Eq. (14) and Eq. (16) that for any $0 < \tau < 1$

$$\begin{aligned}&\Pr\left(\left|\log\frac{\widehat{\mathrm{var}}(X_{:j})}{r(\bar{X}_{:j})^2}\right| > \varepsilon \text{ for some } \ j=1,\dots,c(r) \text{ such that } r\bar{X}_{:j} > x(r)\right) \qquad [17]\\ \le& c(r)\Pr\left(\log\frac{\left(\sum_{i=1}^{r} X_{i,1}\right)^2}{\sum_{i=1}^{r} X_{i,1}^2} > \varepsilon/2, \text{ and for some } j=1,\dots,r,\right.\\ &\qquad \left. X_{j,1} > (1-\tau)x(r) \text{ and } \sum_{i=1}^{r}\mathbf{1}(i\ne j)X_{i,1} \le \tau x(r)\right) + o(1).\end{aligned}$$

On the event described in the probability in the right side of Eq. (17), let J be the unique j satisfying the condition in that event. Then

$$\begin{aligned}1 &\le \frac{\left(\sum_{i=1}^{r} X_{i,1}\right)^2}{\sum_{i=1}^{r} X_{i,1}^2} = \frac{X_{J,1}^2 + 2X_{J,1}\sum_{i=1}^{r}\mathbf{1}(i\ne J)X_{i,1} + \left(\sum_{i=1}^{r}\mathbf{1}(i\ne J)X_{i,1}\right)^2}{\sum_{i=1}^{r} X_{i,1}^2}\\ &\le 1 + \frac{2\sum_{i=1}^{r}\mathbf{1}(i\ne J)X_{i,1}}{X_{J,1}} + \frac{\left(\sum_{i=1}^{r}\mathbf{1}(i\ne J)X_{i,1}\right)^2}{X_{J,1}^2}\\ &\le 1 + \frac{2\tau}{1-\tau} + \frac{\tau^2}{(1-\tau)^2} = \frac{1}{(1-\tau)^2},\end{aligned}$$

and if $\tau > 0$ is small enough, the logarithm of $1/(1-\tau)^2$ does not exceed $\epsilon/2$. This proves the claim of the theorem. □

B. Proof of Theorem 2.

Proof. To show X_{i1} has Pareto-like tails, note that

$$\Pr(X_{i1} > x) = \Pr(Z_{i1}e^{G_{i1}} > x) = x^{-\alpha}E\left[e^{\alpha G_{i1}}1_{\{e^{G_{i1}}<x\}}\right] + \Pr(e^{G_{i1}} > x)$$

and since $e^{G_{i1}}$ has all moments, we have, as $x \to \infty$,

$$x^{\alpha}\Pr(X_{i1} > x) \to Ee^{\alpha G_{i1}}.$$

Also,

$$\begin{aligned}\frac{1}{x(r)}\int_0^{x(r)} P(X_{i1} > x \mid \mathcal{G})dx &= \frac{1}{x(r)}\int_0^{x(r)} x^{-\alpha}e^{\alpha G_{i1}}1_{\{e^{G_{i1}}<x\}}dx + \frac{1}{x(r)}\min\{e^{G_{i1}}, x(r)\}\\ &\le Ke^{\alpha G_{i1}}x(r)^{-\alpha},\end{aligned}$$

where K is a generic constant whose value may change from line to line. It follows that

$$c(r)r^2E\left[\max_{i=1,\dots,r}\left(\left[\frac{1}{x(r)}\int_0^{x(r)}\Pr\big(X_{i,1}>x\mid\mathcal{G}\big)\,dx\right]^2\right)\right]\le Kc(r)r^2x(r)^{-2\alpha}E\left[\max_{i=1,\dots,r}\{e^{2\alpha G_{i1}}\}\right]. \quad [18]$$

We next show that there exists a constant $d_1>2\sqrt{2}\alpha\sigma$, where $\sigma^2=\gamma(0,0)$, such that

$$E\left[\max_{i=1,\dots,r}\{e^{2\alpha G_{i1}}\}\right]\le\exp\{d_1(\log r)^{1/2}\} \quad [19]$$

for all large r. Condition Eq. (9) follows easily by applying this bound together with Eq. (18). We have

$$\begin{aligned}
E\left[\max_{i=1,\dots,r}\{e^{2\alpha G_{i1}}\}\right] &= \int_0^\infty \Pr(\max_{i=1,\dots,r}\{e^{2\alpha G_{i1}}\}>x))\,dx\\
&= \int_0^\infty \Pr(\max_{i=1,\dots,r}\{G_{i1}\}>\log x/(2\alpha))\,dx\\
&\le \exp\{d_1(\log r)^{1/2}\}+r\int_{\exp\{c(\log r)^{1/2}\}}^\infty \Pr(G_{11}>\log x/(2\alpha))\,dx\\
&\le \exp\{d_1(\log r)^{1/2}\}+rK\int_{\exp\{d_1(\log r)^{1/2}\}}^\infty \exp\{-(\log x)^2/(8\alpha^2\sigma^2)\}\,dx
\end{aligned}$$

where we have used the relation $1-\Phi(x)\sim x^{-1}\phi(x)$ as $x\to\infty$, $\Phi(x)$ is the standard normal distribution, and $\phi(x)=\Phi'(x)$. By a change of variables, the second term is equal to

$$\begin{aligned}
rK\int_{d_1(\log r)^{1/2}/(2\alpha\sigma)}^\infty \exp\{-y^2/2+2\alpha\sigma y\}\,dy &\le rK\int_{d_1(\log r)^{1/2}/(2\alpha\sigma)}^\infty \exp\{-d_2y^2/2\}\,dy\\
&\le rK\exp\{-d_2d_1^2(\log r)/(8\alpha^2\sigma^2)\},
\end{aligned}$$

where the first inequality follows for any $0<d_2<1$ and r large. By choosing d_2 sufficiently close to 1 such that $d_2d_1^2/(8\alpha^2\sigma^2)>1$, the bound of the last term is

$$rKo(r^{-1})=o(1).$$

This establishes the bound in Eq. (19) and completes the proof. □

C. Proof of Theorem 3.

Proof. The probability in Eq. (12) does not exceed

$$\begin{aligned}
&c(r)\Pr\left(\left|\log\frac{\widehat{\mathrm{var}}(X_{:1})}{r(\bar X_{:1})^2}\right|>\varepsilon,\ r\bar X_{:1}>rEX_{11}+x(r)\right) \quad [20]\\
=&c(r)\Pr\left(\left|\log\left[\frac{\sum_{i=1}^r X_{i,1}^2}{\left(\sum_{i=1}^r X_{i,1}\right)^2}-\frac{1}{r}\right]\right|>\varepsilon,\ \sum_{i=1}^r X_{i,1}>rEX_{11}+x(r)\right)\\
\le&c(r)\Pr\left(\log\frac{\left(\sum_{i=1}^r X_{i,1}\right)^2}{\sum_{i=1}^r X_{i,1}^2}>\varepsilon/2,\ \sum_{i=1}^r X_{i,1}>rEX_{11}+x(r)\right),
\end{aligned}$$

with the last inequality valid for large r.

Next we prove that

$$\lim_{r\to\infty}c(r)\Pr\left(\sum_{i=1}^r X_{i,1}>rEX_{11}+x(r),\ X_{i,1}\le\frac{x(r)}{4},\ i=1,\dots,r\right)=0. \quad [21]$$

Indeed, let

$$\begin{aligned}
T_1&=\inf\left\{n=1,2,\dots,r:\sum_{i=1}^n(X_{i,1}-EX_{11})>\frac{x(r)}{4}\right\},\\
T_2&=\inf\left\{n=T_1+1,\dots,r:\sum_{i=T_1+1}^n(X_{i,1}-EX_{11})>\frac{x(r)}{4}\right\},
\end{aligned}$$

and $T_2 = \infty$ if either infimum is over an empty set. Then the probability in Eq. (21) is bounded above by

$$\begin{aligned} &\Pr\left(T_2 < \infty,\ X_{i,1} \le \frac{x(r)}{4},\ i = 1, \ldots, r\right) \\ &\le \left[\Pr\left(\sum_{i=1}^{r}(X_{i,1} - EX_{11}) > \frac{x(r)}{4},\ X_{i,1} \le \frac{x(r)}{4},\ i = 1, \ldots, r\right)\right]^2. \end{aligned} \quad [22]$$

By the assumptions, the law of X_{11} is in the domain of attraction of an α-stable law with $1 < \alpha < 2$. Therefore, there is a $q > 0$ such that, for all r sufficiently large,

$$Pr\left(\sum_{i=1}^{r}(X_{i,1} - EX_{11}) < 0\right) \ge q.$$

Therefore, if $(Y_{i,1})$ is an independent copy of $(X_{i,1})$, then

$$\begin{aligned} &qPr\left(\sum_{i=1}^{r}(X_{i,1} - EX_{11}) > \frac{x(r)}{4}, X_{i,1} \le \frac{x(r)}{4},\ i = 1, \ldots, r,\right) \\ \le &Pr\left(\sum_{i=1}^{r}(X_{i,1} - EX_{11}) > \frac{x(r)}{4}, X_{i,1} \le \frac{x(r)}{4},\ i = 1, \ldots, r,\right) Pr\left(\sum_{i=1}^{r}(Y_{i,1} - EY_{11}) < 0\right) \\ = &Pr\left(\sum_{i=1}^{r}(X_{i,1} - EX_{11}) > \frac{x(r)}{4},\ X_{i,1} \le \frac{x(r)}{4},\ i = 1, \ldots, r,\ \sum_{i=1}^{r}(Y_{i,1} - EY_{11}) < 0\right) \\ \le &\Pr\left(\sum_{i=1}^{r}(X_{i,1} - Y_{i,1}) > \frac{x(r)}{4},\ X_{i,1} \le \frac{x(r)}{4},\ i = 1, \ldots, r\right). \end{aligned}$$

Now using Markov's inequality,

$$\begin{aligned} &\Pr\left(\sum_{i=1}^{r}(X_{i,1} - EX_{11}) > \frac{x(r)}{4},\ X_{i,1} \le \frac{x(r)}{4},\ i = 1, \ldots, r\right) \\ \le &q^{-1}\Pr\left(\sum_{i=1}^{r}(X_{i,1} - Y_{i,1}) > \frac{x(r)}{4},\ X_{i,1} \le \frac{x(r)}{4},\ i = 1, \ldots, r\right) \\ \le &q^{-1}\Pr\left(\sum_{i=1}^{r}\big(X_{i,1} \wedge x(r)/4 - Y_{i,1} \wedge x(r)/4\big) > \frac{x(r)}{4}\right) \\ \le &q^{-1}\frac{E\left[\sum_{i=1}^{r}\big(X_{i,1} \wedge x(r)/4 - Y_{i,1} \wedge x(r)/4\big)\right]^2}{(x(r)^2/16)} \\ = &\frac{16r}{qx(r)^2}E\big(X_{1,1} \wedge x(r)/4 - Y_{1,1} \wedge x(r)/4\big)^2 \\ \le &\frac{32r}{qx(r)^2}E\big(X_{1,1} \wedge x(r)/4\big)^2. \end{aligned}$$

By Karamata's theorem, $E\big[(X_{11})_+ \wedge x(r)/4\big]^2 \sim \frac{\alpha}{2-\alpha}x^2(r)\Pr(X_{11} > x(r)/4)/16$, from which Eq. (21) follows from Eq. (11) and Eq. (22).

The rest of the argument is identical to that in Theorem 1. □

Our derivations of TL with $b = 2$ for the extreme upper tail of positive regularly varying distributions with tail index $\alpha \in (0, 1) \cup (1, 2)$ extend related previous findings that assumed finite means and finite variances (7–10).

Our theoretical analyses leave room for further development. For example, we have no results on $\alpha = 1$, on the speed of convergence to 2 of the slope b of TL for the extreme upper tail, or on variable sample sizes $r_j,\ j = 1, \ldots, c$. More realistic models are needed to describe the dependence spatially (among counties within and between states) and temporally (from day to day, month to month, year to year) in the counts of cases and deaths.

4. Supplementary discussion: Other examples of the same data structure

The data structure analyzed here is a large set of c samples, and each sample contains a large number of observations $r_j,\ j = 1, \ldots, c$. In the simplest case, which we simulate and analyze mathematically here, each sample has $r_j = r$ observations. To show that our simulations and mathematical analysis have scientific relevance beyond COVID-19, we give some published examples of the same or very similar data structures in ecology (vole population density), census counts (U.S. county human populations), meteorology (tornados in outbreaks), and age-specific human death rates. None of these empirical studies of TL checked whether the upper extremes of their data could usefully be described as heavy-tailed.

Cohen and Saitoh (12) reported annual trapping measurements of the population abundance of a vole (then an economically important rodent pest of cultivated forests) in Hokkaido, Japan, at 85 forest stations from 1962 to 1992, giving a 31×85 matrix of abundances, one row per year, one column per trapping station. They tested and confirmed a temporal TL and a spatial TL. They demonstrated that the time-series at different stations (the columns of the data matrix) were correlated, but not with the same correlation for all pairs of stations. Using an autoregressive time series model, they showed that, at each station, each year's abundance was influenced by the abundance of at least the two prior years. Thus their rectangular array of counts demonstrates dependence between columns (correlations between different trapping stations) and within columns (serial autocorrelations).

Xu and Cohen (11) analyzed the censused counts of numbers of people in every county (or equivalent) that has ever existed in every territory or state of the United States that has ever been censused in any of the 23 U.S. decennial censuses from 1790 through 2010. As not every county was censused 23 times, these data do not fill a rectangular matrix but may be thought of as occupying an array with 23 rows, one row per census, and 4,128 columns, one column per county. An element in the array is blank if a county was not included in a census. Xu and Cohen (11) tested temporal TLs, spatial TLs, and spatial hierarchical TLs and quadratic generalizations of these log-linear variance functions. They demonstrated temporal and spatial dependence of the county population time series.

Tippett and Cohen (13) showed that, for the years 1954-2014, in the continental United States, in outbreaks consisting of multiple tornadoes (of intensity F1+ on the Fujita or Enhanced Fujita scales) closely spaced in time, the annual mean number of tornadoes per outbreak rose on average from year to year, and the annual variance of the number of tornadoes per outbreak increased on average from year to year more than four times faster. The variance function was well described by TL Main text Eq. [2] when outbreaks were grouped by year (analogous to state in our COVID-19 study), and the mean and variance of the numbers of tornadoes were computed for the outbreaks (analogous to our counties) in each year. There were no statistically significant trends in the number of F1+ tornadoes per year or in the number of outbreaks, but the tornadoes were increasingly clustered into outbreaks. In terms of our matrix model, $c = 61$ is the number of years of observation. The number r_j of outbreaks in each year $j = 1, \ldots, c$ averaged between 22 and 25 (13, Figure 2a).

Tippett and Cohen (13) did not study the possible temporal and spatial dependence among the number of tornadoes in different outbreaks. They noted that the Fujita-kilometers values (another measure of the intensity of a tornado outbreak) "for the 1974 Super Outbreak and the 25-28 April 2011 tornado outbreak are more than 26 standard deviations above the mean of the data on an arithmetic scale and more than six standard deviations above the mean of the log-transformed data, when means and standard deviations are calculated after withholding the two extreme values." The upper percentiles of the annual distribution of tornadoes per outbreak increased much faster than the lower percentiles (14).

In the age-specific death rates of 12 countries with high-quality human mortality data, separately by sex and country, from 1960 to 2009, the variance over years (corresponding to counties in our COVID-19 study) of death rates at a given age (corresponding to states in our COVID-19 study) was well described by a power function of the mean over years of death rates at that age, with one (mean over time, variance over time) pair for each age group from 0 to 100+ (Bohk, Rau and Cohen (15)). This is TL. The estimated slope of Main text Eq. [2] satisfied $1.5 < \hat{b} < 2$ but differed from country to country. The Gompertz model of human mortality, which asserts that the age-specific death rate grows exponentially with age, predicts TL with slope $b = 2$ when the modal age at death in the Gompertz model increases linearly with time and the growth rate of mortality with age is constant in time (16). That the latter assumption is empirically false helps explain why the estimated slope of TL was generally less than the value 2 theoretically predicted from the Gompertz model. Bohk, Rau and Cohen (15, 16) did not model the dependence of age-specific death rates over time and across ages.

SI Dataset S1 (us-counties.csv)

NewYorkTimes. nytimes covid-19-data public. Accessed 19 June 2021. https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties.csv.

References

1. JE Cohen, RA Davis, G Samorodnitsky, Heavy-tailed distributions, correlations, kurtosis, and Taylor's law of fluctuation scaling. *Proc. Royal Soc. (UK) A* **476**, 20200610 (2020).
2. S Nadarajah, Exact distribution of the product of m gamma and n Pareto random variables. *J. Comput. Appl. Math.* **235**, 4496–4512 (2011) doi:10.1016/j.cam.2011.04.018.
3. SJ Taylor, Modeling stochastic volatility: A review and comparative study. *Math. finance* **4**, 183–204 (1994).

4. RA Davis, T Mikosch, Probabilistic properties of stochastic volatility models in *Handbook of financial time series.* (Springer), pp. 255–267 (2009).
5. MB Palacios, MFJ Steel, Non-gaussian bayesian geostatistical modeling. *J. Am. Stat. Assoc.* **101**, 604–618 (2006).
6. W Huang, K Wang, F Jay Breidt, RA Davis, A class of stochastic volatility models for environmental applications. *J. Time Ser. Analysis* **32**, 364–377 (2011).
7. JE Cohen, M Xu, WSF Schuster, Stochastic multiplicative population growth predicts and interprets Taylor's power law of fluctuation scaling. *Proc. Royal Soc. UK B* **280**, 20122955 (2013) doi:10.1098/rspb.2012.2955.
8. JE Cohen, Stochastic population dynamics in a Markovian environment implies Taylor's power law of fluctuation scaling. *Theor. Popul. Biol.* **93**, 30–37 (2014) doi:10.1016/j.tpb.2014.01.001.
9. JE Cohen, M Xu, Random sampling of skewed distributions implies Taylor's power law of fluctuation scaling. *Proc. Natl. Acad. Sci. USA* **112** (2015).
10. DC Reuman, L Zhao, L Sheppard, P Reid, JE Cohen, Synchrony affects Taylor's law in theory and data. *Proc. Natl. Acad. Sci. USA* **114**, 6788–6793 (2017).
11. M Xu, JE Cohen, Spatial and temporal autocorrelations affect Taylor's law for us county populations: descriptive and predictive models. *PLoS ONE* **16**, e0245062 (2021).
12. JE Cohen, T Saitoh, Population dynamics, synchrony, and environmental quality of Hokkaido voles lead to temporal and spatial Taylor's laws. *Ecology* **97**, 3402–3413 (2016).
13. MK Tippett, JE Cohen, Tornado outbreak variability follows Taylor's power law of fluctuation scaling and increases dramatically with severity. *Nat. Commun.* **7**, 10668 (2016).
14. MK Tippett, C Lepore, JE Cohen, More tornadoes in the most extreme U.S. tornado outbreaks. *Science* **354**, 1419–1423 (2016) doi:10.1126/science.aah7393.
15. C Bohk, R Rau, JE Cohen, Taylor's power law in human mortality. *Demogr. Res.* **33**, 589–610 (2015).
16. JE Cohen, C Bohk-Ewald, R Rau, Gompertz, Makeham, and Siler models explain Taylor's law in human mortality data. *Demogr. Res.* **38**, 773–842 (2018).