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 $\alpha = 1$).

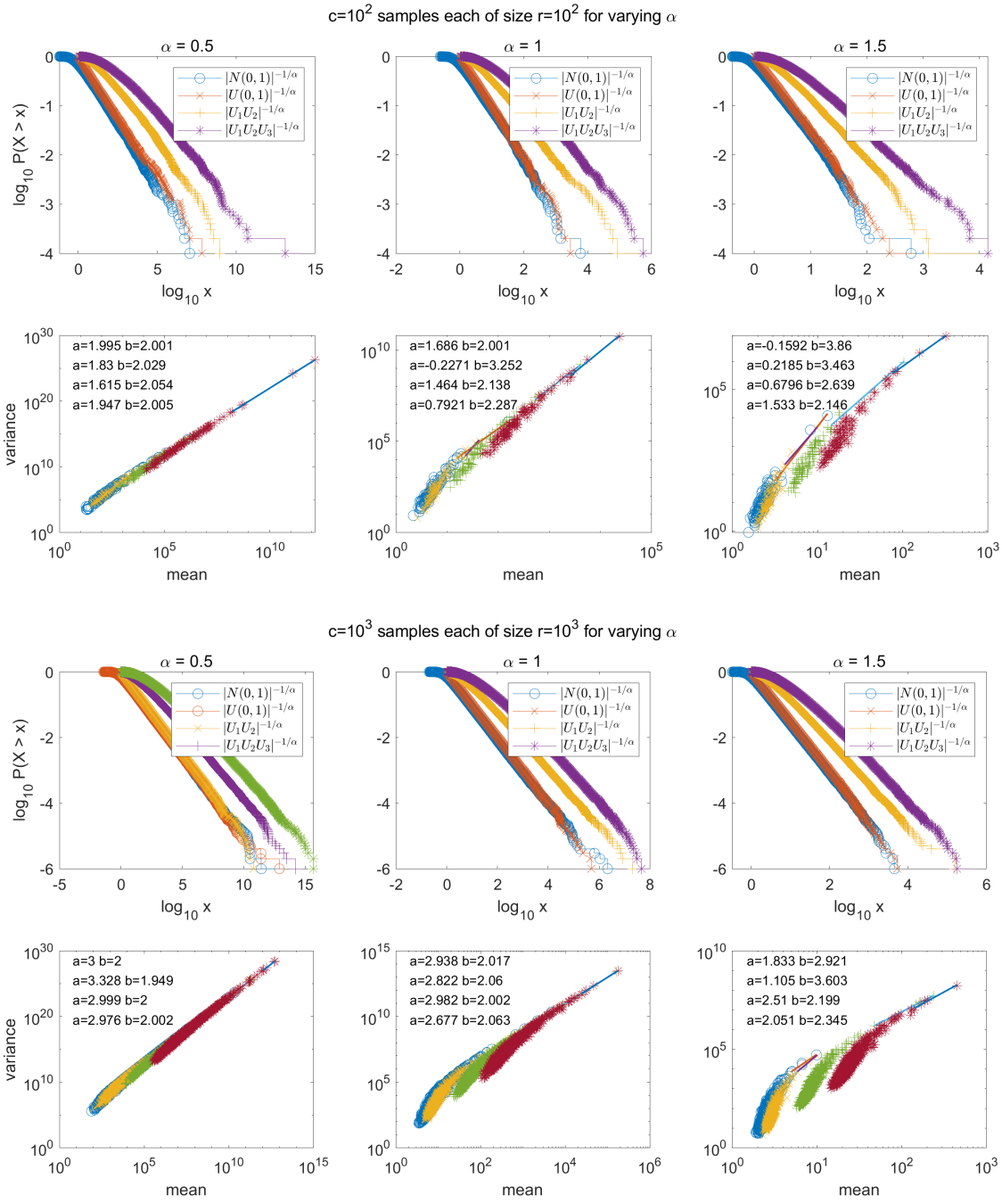

Fig. S1. For *independent* observations (counts), the empirical survival curves (rows 1, 3) of one simulation of $rc = 10^4$ (row 1) or $rc = 10^6$ (row 3) observations by each of four methods of simulation Eq. (3)–Eq. (6); and sample variance functions (rows 2, 4) of one simulation of $\mathbf{X}$ of 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.

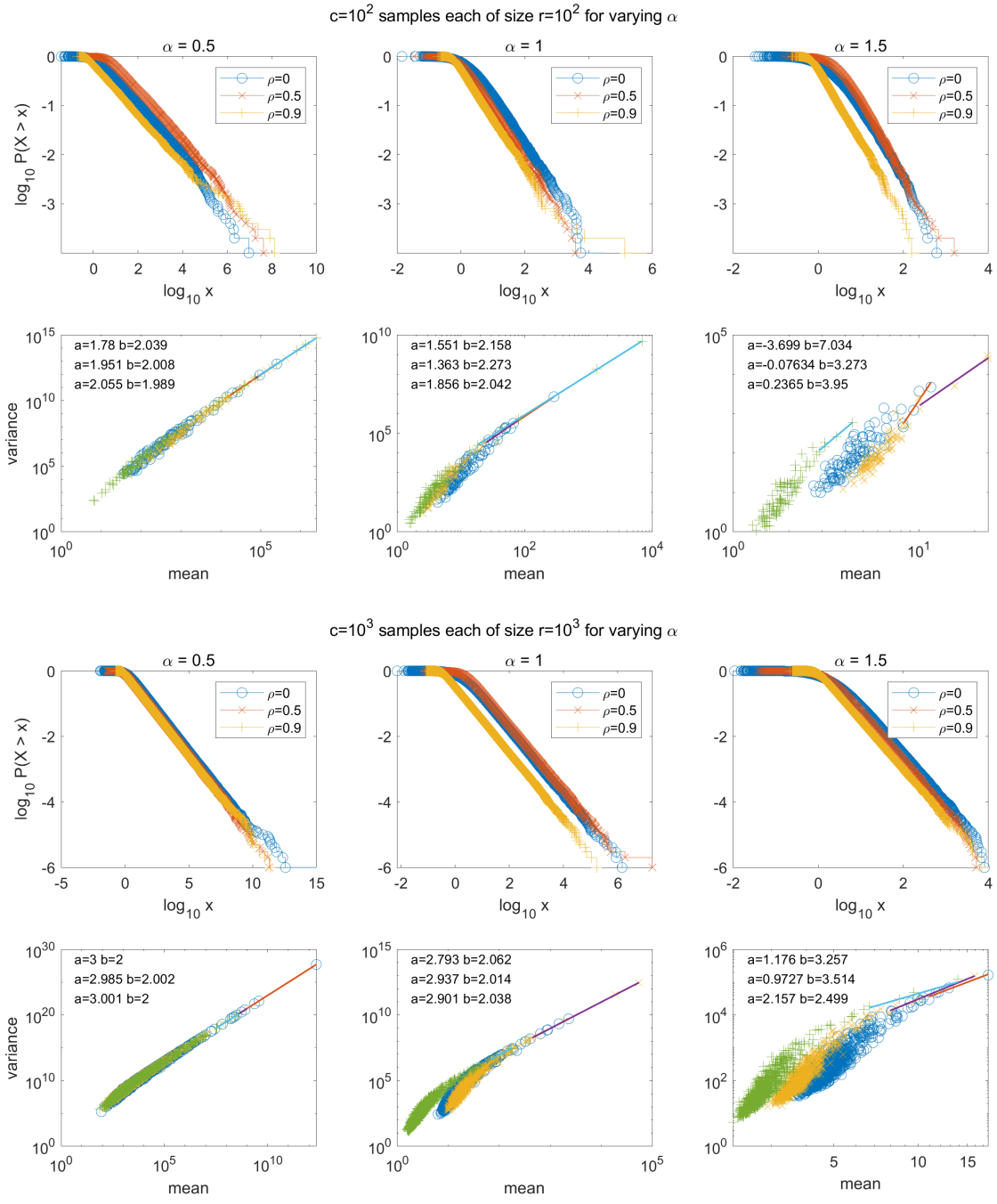

Fig. S2. For *dependent* observations (counts), the empirical survival curves (rows 1, 3) of one simulation of $rc = 10^4$ (row 1) or $rc = 10^6$ (row 3) observations and sample variance functions (rows 2, 4) of one simulation of $\mathbf{X}$ of size $r \times c = 10^2 \times 10^2$ (row 2) or $r \times c = 10^3 \times 10^3$ (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).