## Research

Check for updates

**Author for correspondence:**
Joel E. Cohen
e-mail: joel.cohen@rockefeller.edu

# Heavy-tailed distributions, correlations, kurtosis and Taylor's Law of fluctuation scaling

Joel E. Cohen[1,2,3], Richard A. Davis[4] and Gennady Samorodnitsky[5]

[1]Laboratory of Populations, The Rockefeller University and Columbia University, New York, NY, USA
[2]Earth Institute, Department of Statistics, Columbia University, New York, NY, USA
[3]Department of Statistics, University of Chicago, Chicago, IL, USA
[4]Department of Statistics, Columbia University, New York, NY, USA
[5]School of Operations Research and Information Engineering, Cornell University, Ithaca, NY, USA

JEC, 0000-0002-9746-6725

Pillai & Meng (Pillai & Meng 2016 *Ann. Stat.* **44**, 2089–2097; p. 2091) speculated that 'the dependence among [random variables, rvs] can be overwhelmed by the heaviness of their marginal tails ⋯'. We give examples of statistical models that support this speculation. While under natural conditions the sample correlation of regularly varying (RV) rvs converges to a generally random limit, this limit is zero when the rvs are the reciprocals of powers greater than one of arbitrarily (but imperfectly) positively or negatively correlated normals. Surprisingly, the sample correlation of these RV rvs multiplied by the sample size has a limiting distribution on the negative half-line. We show that the asymptotic scaling of Taylor's Law (a power-law variance function) for RV rvs is, up to a constant, the same for independent and identically distributed observations as for reciprocals of powers greater than one of arbitrarily (but imperfectly) positively correlated normals, whether those powers are the same or different. The correlations and heterogeneity do not affect the asymptotic scaling. We

## THE ROYAL SOCIETY
PUBLISHING

analyse the sample kurtosis of heavy-tailed data similarly. We show that the least-squares estimator of the slope in a linear model with heavy-tailed predictor and noise unexpectedly converges much faster than when they have finite variances.

**2**

royalsocietypublishing.org/journal/rspa   *Proc. R. Soc. A* **476**: 20200610

# 1. Introduction

## (a) Motivations

This work has a practical motivation and a theoretical motivation. Both are driven by the need to cope with exceptionally large events such as outbreaks of human, animal and plant diseases; earth events like cyclones, tsunamis, earthquakes, volcanoes, fires, droughts and floods; and extreme fluctuations in finance, insurance, trade, production and employment, among others. Both motivations are enabled by the on-going development since the work of Paul Lévy in 1924 of a systematic theory of random events with infinite mean [1–3]

For example, one practical motivation is the need to understand the effect of heavy-tailed data on an empirical regularity called Taylor's Law (TL) or Taylor's power law in the biological sciences, and fluctuation scaling in the physical sciences. TL asserts that the sample variance is approximately proportional to some power of the sample mean in a set of samples. Explicitly, for some constants $a > 0$ and $b$, both of which are independent of the sample $i$,

$$\text{sample variance of sample } i \approx a \times (\text{sample mean of sample } i)^b, i = 1, 2, \ldots$$

or equivalently

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

or equivalently

$$\frac{\text{sample variance of sample } i}{(\text{sample mean of sample } i)^b} \approx a, i = 1, 2, \ldots.$$

These specifications of TL intentionally leave vague the error model hinted at by the approximation $\approx$. TL is very widely confirmed in many sciences, including ecology, infectious disease epidemiology, human demography, financial statistics, earth sciences and other physical sciences, as reviewed by Eisler *et al.* [4] and Taylor [5].

In statistics, a variance function is a function that specifies how the variance is related to the mean in a set of samples. To statisticians, TL is simply a power-law variance function. An early, perhaps the first, occurrence of a power-law variance function in ecology appeared in Bliss [6]. From 1941 to 2017, users of TL assumed, usually without saying so, that observations were drawn from distributions with finite mean and finite variance. At the same time, heavy-tailed data were increasingly recognized in insurance, income distributions, earthquake magnitudes, financial fluctuations, meteorology and other sciences. Heavy-tailed data distributions have some or all infinite moments, such as infinite variance or infinite mean. Heavy-tailed distributions have much greater probabilities of extremely large observations than light-tailed distributions, such as the normal or Gaussian distribution. But samples from a heavy-tailed distribution always have a finite sample mean and finite sample variance. It is important to know under what conditions samples from a heavy-tailed distribution obey TL and, when TL holds, how the exponent $b$ of TL relates to the parameter(s) of the underlying heavy-tailed distribution. Brown *et al.* [7] showed that certain random samples (sets of independently and identically distributed [iid] observations) from heavy-tailed distributions with infinite mean obey TL. These results are greatly extended here. In particular, we ask whether and when TL holds if observations are not iid, either because the observations are dependent or because the observations come from different distributions.

In addition to its empirical motivation, this work has a theoretical motivation. Drton & Xiao [8] conjectured, and proved when $m = 2$, a surprising and beautiful proposition subsequently proved

for any integer $m > 1$ by Pillai & Meng [9]. For integer $m > 1$, let $X = (X_1, \ldots, X_m)$ and $Y = (Y_1, \ldots, Y_m)$ be independent copies of a multivariate normal vector $N(0, \Sigma)$, where $\Sigma = \{\sigma_{ij}\} \geq 0$ is any $m \times m$ covariance matrix with $\sigma_{jj} > 0, 1 \leq j \leq m$. Let $(w_1, \ldots, w_m)$ be weights, $w_j \geq 0, j = 1, \ldots, m$, and $\Sigma_{j=1}^m w_j = 1$. Then $\Sigma_{j=1}^m w_j X_j / Y_j$ has the standard Cauchy distribution on the real line with probability density function (pdf) $1/[\pi(1 + z^2)]$. The surprise is that the covariance matrix $\Sigma$ has no effect on the distribution of $\Sigma_{j=1}^m w_j X_j / Y_j$.

Pillai & Meng [9, p. 2091] remarked: 'A theoretical speculation from this unexpected result is that for a set of random variables $\cdots$, the dependence among them can be overwhelmed by the heaviness of their marginal tails $\cdots$. We invite the reader to ponder with us whether this is a pathological phenomenon or something profound'. The present work supports the speculation of Pillai & Meng [9] by giving multiple examples in which heavy tails outweigh correlations and heterogeneity of distributions. Here is a preview of coming attractions.

## (b) Organization of this paper

Section 2 reviews the definitions and some properties of regularly varying (RV), including heavy-tailed, random variables (rvs), those in which the survival function or upper tail is asymptotically a power function with negative exponent $-\alpha$, where $\alpha > 0$. An important classical limit theorem stated here is that sums of appropriately centred and normalized RV rvs converge to stable rvs with index $\alpha$ if $\alpha < 2$. We illustrate the uses of this theorem by deriving the limiting stable laws that describe the asymptotic behaviour of the sums of squared rvs and the sums of squared deviations from the sample mean.

Section 3 analyses the asymptotic behaviour of Taylor's Law when the observations are iid heavy-tailed RV rvs; or when the observations are RV rvs with distinct tail indices (heterogeneity); or when the observations are the reciprocals of powers greater than one of arbitrarily (but imperfectly) positively or negatively dependent normals. These results provide far-reaching generalizations of Taylor's Law. The ratio of the sample variance to a power of the sample mean, which is the focus of Taylor's Law, is formally similar to the sample kurtosis, which is the ratio of the fourth central moment to a power (the square) of the sample variance. Thus, we use the methods developed to analyse Taylor's Law to derive the asymptotic behaviour of the sample kurtosis when the data are iid RV rvs. In principle, these results could be extended to heterogeneous and dependent data.

Section 4 shows that under natural conditions the sample correlation of a pair of RV rvs converges to a generally random limit, which we express in terms of stable laws. This limit is zero in case the rvs are the reciprocals of powers greater than one of two arbitrarily (but imperfectly) positively or negatively dependent normal rvs. Surprisingly, in this case the sample correlation multiplied by the sample size has a limiting distribution that is concentrated on the negative half-line, regardless of whether the two normals are positively or negatively dependent. These findings support the speculation of Pillai & Meng [9].

Section 5 compares the behaviour of the least-squares estimator of $\eta$ in a linear model $Y = \eta X + Z$ under two different assumptions about $X$ and $Z$: first, that $X$ and $Z$ are heavy-tailed RV rvs with asymptotically equivalent tails; and second, that $X$ is heavy-tailed but the noise variable $Z$ is light-tailed, possibly Gaussian. In both cases the estimator converges much faster than in the case when $X$ and $Z$ have finite variances. When $Z$ is light-tailed, the sample correlation between $X$ and $Y$ converges in probability to 1.

Section 6 investigates by numerical simulation how rapidly some of the correlated, heterogeneous, RV rvs described in the preceding sections converge in distribution to their limit rvs expressed as functions of stable laws. The examples are drawn from Taylor's Law, bivariate correlation, and the linear model.

For the sake of readability, §§2–5 give only a few simple proofs. Appendix A gives more detailed mathematical background on heavy-tailed distributions and point process methods. It then proves the major claims made regarding the limit theory for Taylor's Law and the limit

theory for the sample correlations of heavy-tailed data including the special example of the reciprocals of powers greater than one of correlated normals.

## 2. Background on heavy tails and regular variation

For heavy-tailed data when the population variance is infinite, the limit behaviour of the sample mean, sample variance, extreme order statistics and related statistics, is widely studied [1,2,10–14]. When the variance is finite, the classical central limit theorem implies that the sample mean, centred by its mean and normalized by its standard deviation, converges to a normal distribution. When the variance is infinite, the suitably centred and normalized partial sums also converge in distribution when the data are regularly varying (RV), with Pareto upper tails. So regular variation is often the starting point for modelling various statistics of univariate heavy-tailed data.

To be specific, assume the random variable (rv) $X$ is positive and has distribution function $F(x) := P(X \leq x), x > 0$ and survival function $\bar{F}(x) := 1 - F(x) = P(X > x)$. We say that $X$ (or $F$) is RV with index $\alpha > 0$, and we shall henceforth write $F \in RV(\alpha)$, if

$$\lim_{t \to \infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-\alpha} \quad \text{for all } x > 0. \tag{2.1}$$

If $F$ satisfies (2.1), then $F$ has a Pareto upper tail:

$$\bar{F}(x) = x^{-\alpha} L(x), \quad \lim_{t \to \infty} \frac{L(tx)}{L(t)} = 1, \quad \text{for all } x > 0. \tag{2.2}$$

We say $L(x)$ is a slowly varying function (at infinity). Examples of slowly varying functions are $L(x) = \log x$, $\log \log x$ and their reciprocals. Assumption (2.1) implies that $X$ has finite moments only up to order $\alpha$: $EX^p < \infty$ for $p < \alpha$ and $EX^p = \infty$ for $p > \alpha$. From (2.2), $EX^\alpha = \int_0^\infty P(X^\alpha > x) \, dx = \int_0^\infty x^{-1} L(x^{1/\alpha}) \, dx$, so that the finiteness of $EX^\alpha$ depends on the integrability of $x^{-1} L(x^{1/\alpha})$ for $x$ large. For example, $EX^\alpha = \infty$ if $L(x)$ is constant, whereas $EX^\alpha < \infty$ if $L(x) \sim (1/\log x)^{1+\epsilon}$ for some $\epsilon > 0$. Since we are interested here only in heavy-tailed distributions with infinite variance, we shall assume that $\alpha < 2$ throughout.

Let $X_1, \ldots, X_n$ be iid with distribution $F$. The sum $\sum_{t=1}^n X_t$, appropriately centred and normalized, converges in distribution as the sample size $n \to \infty$. To this end, define the sequence of normalizing constants $\{a_n : n = 1, 2, \ldots\}$ as the sequence of $1 - 1/n$ quantiles of $F$, i.e.

$$a_n = \inf \left\{ x : F(x) \geq 1 - \frac{1}{n} \right\}. \tag{2.3}$$

If $F$ is continuous, then $\bar{F}(a_n) = 1/n$. By replacing $t$ with $a_n$ in (2.1) it can be shown that

$$nP(X > a_n x) = n\bar{F}(a_n x) \to x^{-\alpha} \quad \text{for all } x > 0 \text{ as } n \to \infty \tag{2.4}$$

and $a_n = n^{1/\alpha} \tilde{L}(n)$ for some slowly varying function $\tilde{L}$. If $F$ is Pareto, i.e. if $\bar{F}(x) = x^{-\alpha}$ for $x \geq 1$, then $a_n = n^{1/\alpha}$.

A classical limit theorem [15] describes sums of heavy-tailed rvs. Let $S_\alpha$ be a stable rv with index $\alpha \in (0, 2)$ and shape parameter $\beta = 1$. When $\alpha \in (0, 1)$, $S_\alpha$ takes only non-negative real values. When $\alpha \in [1, 2)$, $S_\alpha$ takes negative real values, but the left tail is lighter than that of a normal distribution. When $\alpha = 2$, $S_\alpha$ is normally distributed.

Define

$$b_n = 0 \text{ for } \alpha < 1 \quad \text{and} \quad b_n = E[X \mathbf{1}_{\{X \leq a_n\}}] \text{ for } \alpha \in [1, 2). \tag{2.5}$$

Then

$$a_n^{-1} \sum_{t=1}^n (X_t - b_n) \xrightarrow{d} S_\alpha, \tag{2.6}$$

where $\xrightarrow{d}$ denotes 'converges in distribution to'. For $\alpha \in (1, 2)$, $b_n$ can be replaced by the (finite) mean $EX$.

**Example 2.1 (sums of squares).** Because $X^2 \in RV(\alpha/2)$, applying (2.6) to the sums of squares gives, for any $\alpha \in (0, 2)$,

$$a_n^{-2} \sum_{t=1}^n X_t^2 \xrightarrow{d} S_{\alpha/2} \quad \text{as } n \to \infty. \tag{2.7}$$

The sum of squared deviations from the sample mean $\bar{X}_n := n^{-1} \sum_{t=1}^n X_t$ has the same limit:

$$a_n^{-2} \sum_{t=1}^n (X_t - \bar{X}_n)^2 \xrightarrow{d} S_{\alpha/2} \quad \text{as } n \to \infty. \tag{2.8}$$

The convergences in (2.6) and (2.8) are joint, i.e.

$$\left( a_n^{-1} \sum_{t=1}^n (X_t - b_n), a_n^{-2} \sum_{t=1}^n (X_t - \bar{X}_n)^2 \right) \xrightarrow{d} (S_\alpha, S_{\alpha/2}) \quad \text{as } n \to \infty. \tag{2.9}$$

The remarkable feature of this example is that the limit rvs on the right sides of (2.7) and (2.8) are identical. To give insight into why subtracting the sample mean in (2.8) makes no difference to the limit rv, we offer a brief proof. For the remainder of this article, for the sake of readability, we shall defer many detailed proofs to appendix A. For example, the nature of the dependence between $S_\alpha$ and $S_{\alpha/2}$ and the proof of (2.9) are given in theorem A.3.

*Proof of* (2.8). The sum of squared deviations from the sample mean on the left of (2.8) decomposes into

$$a_n^{-2} \sum_{t=1}^n (X_t - \bar{X}_n)^2 = a_n^{-2} \sum_{t=1}^n (X_t - b_n)^2 - n^{-1} \left( a_n^{-1} \sum_{t=1}^n (X_t - b_n) \right)^2. \tag{2.10}$$

By virtue of (2.6), the second term on the right converges to zero in probability. Since $a_n^2 = n^{+2/\alpha} \tilde{L}^2(n)$, it follows that $a_n^2 > n^{1+\epsilon}$ for some small $\epsilon > 0$ and all $n$ large. Moreover, since for $\alpha > 1$, $b_n \to EX$ and for $\alpha = 1$, $b_n$ is slowly varying [15], $nb_n/a_n^2 \to 0$ and $b_n/a_n \to 0$. Using these facts and expanding the first term on the right of (2.10) gives

$$a_n^{-2} \sum_{t=1}^n X_t^2 - 2a_n^{-1} b_n a_n^{-1} \sum_{t=1}^n (X_t - b_n) - na_n^{-2} b_n^2 = a_n^{-2} \sum_{t=1}^n X_t^2 + a_n^{-1} b_n O_p(1) - na_n^{-2} b_n^2$$

$$= a_n^{-2} \sum_{t=1}^n X_t^2 + o_p(1),$$

from which (2.8) is immediate. ∎

# 3. Heavy tails, Taylor's Law and kurtosis

## (a) Taylor's Law with iid heavy-tailed data

If the RV heavy-tailed iid rvs $X_1, \ldots, X_n$ have tail index $\alpha < 1$, then the centring constants are $b_n = 0$ and the population mean and population variance are infinite. In one of its many forms, TL considers the ratio of the sample variance divided by a power $b$, $0 < b < \infty$, of the sample mean,

$$W_n = \frac{n^{-1} \sum_{t=1}^n (X_t - \bar{X}_n)^2}{\bar{X}_n^b}. \tag{3.1}$$

The order of the numerator is $n^{-1} a_n^2$ while the order of the denominator is $n^{-b} a_n^b$. For $W_n$ to converge in distribution, the order of the numerator and denominator must match, which requires

that

$$b = \frac{(2 - \alpha)}{(1 - \alpha)}. \tag{3.2}$$

For this $b$,

$$\frac{1}{\tilde{L}^{2-b}(n)} W_n \xrightarrow{d} \frac{S_{\alpha/2}}{S_\alpha^b}, \tag{3.3}$$

where $\tilde{L}(n)$ is as defined for $a_n$ following (2.4). The numerator $S_{\alpha/2}$ and the denominator $S_\alpha^b$ are dependent, and the dependence is described explicitly in theorem A.3. Brown *et al.* [7] obtained expression (3.2) for $b$ and some of the moments of the limiting distribution.

## (b) Taylor's Law with heterogeneous heavy-tailed data

In the previous subsection, the heavy-tailed observations $\{X_t\}$ are assumed to be iid. Here we analyse independent data (or rvs) that are heterogeneous, i.e. not identically distributed.

Suppose that the observations are independent but come from two rvs $U$ and $V$, where $U \in RV(\alpha)$ with $\alpha \in (0, 1)$ and $V$ is dominated by $U$ in the sense that

$$\frac{P(V > x)}{P(U > x)} \to 0 \quad \text{as } x \to \infty. \tag{3.4}$$

Further assume that the data $X_1, \ldots, X_n$ consist of a random number $v_n < n$ of observations distributed as $U$ and $n - v_n$ of observations distributed as $V$, where $v_n/n \xrightarrow{P} p \in (0, 1]$ as $n \to \infty$. Here $\xrightarrow{P}$ means 'converges in probability to'. This formulation can be interpreted quite generally: $V$ could represent a union of several pooled rvs, each with tails lighter than the tail of $U$. Letting $a_n$ be the $1 - 1/n$ quantile of $F_U$, it is easy to show that

$$a_n^{-1} \sum_{t=1}^{n} X_t = a_n^{-1} \sum_{t=1}^{v_n} U_t + a_n^{-1} \sum_{t=1}^{n-v_n} V_t$$

$$= a_n^{-1} \sum_{t=1}^{[np]} U_t + a_n^{-1} \sum_{t=1}^{[n(1-p)]} V_t + o_p(1).$$

Since $P(V > x) = o(P(U > x))$ for $x$ large, the argument in the proof of theorem 4.1 in appendix A(c) can be used to show that

$$a_n^{-1} \sum_{t=1}^{[n(1-p)]} V_t \xrightarrow{P} 0.$$

If $f(x), g(x)$ are real-valued functions of real $x$ and $g(x) > 0$ for all sufficiently large $x$, then $f(x) \sim g(x)$ means $\lim_{x \to \infty} f(x)/g(x) = 1$. Then, on the other hand,

$$a_n^{-1} \sum_{t=1}^{[np]} U_t = \frac{a_{np}}{a_n} a_{np}^{-1} \sum_{t=1}^{[np]} U_t$$

$$\sim p^{1/\alpha} a_{np}^{-1} \sum_{t=1}^{[np]} U_t$$

$$\xrightarrow{d} p^{1/\alpha} S_\alpha.$$

A similar result holds for the sample variance so that the convergence in distribution for TL in (3.3) also holds for such non-identically distributed data. The dependence between numerator $S_{\alpha/2}$ and the denominator $S_\alpha^b$ is described explicitly in theorem A.3.

## (c) Taylor's Law with correlated heavy-tailed data

Here we suppose the observations (or rvs) are pairwise correlated and identically distributed. We shall show that the scaling exponent $b$ in TL remains the same as (3.2), but the limit changes. Specifically, let $N_0, N_1, \ldots$ be iid standard normal rvs and let $0 \le \rho < 1$. Let $Z_i = \sqrt{\rho} N_0 + \sqrt{1 - \rho} N_i$, $i = 1, 2, \ldots$. Then for $i = 1, \ldots$, each $Z_i$ is normal, $EZ_i = 0$, $\text{Var}(Z_i) = 1$, and for $i \ne j \ge 1$, $\text{Cov}(Z_i, Z_j) = \rho$. Let $0 < \alpha < 1$ and define $X_i := 1/(Z_i^2)^{1/(2\alpha)}$, $i = 1, 2, \ldots$. Define $W_n$ in terms of $X_i$ as in (3.1) and $b$ as in (3.2). Then

$$W_n \xrightarrow{d} \frac{S_{\alpha/2}}{S_\alpha^b} \left[ (1 - \rho) \exp \left\{ \frac{\rho}{1 - \rho} N^2 \right\} \right]^{1/(2(1-\alpha))}, \tag{3.5}$$

where $S_{\alpha/2}, S_\alpha$ are non-negative stable rvs given in (A 10) and (A 11) with tail indices $\alpha/2, \alpha \in (0, 1)$, respectively, and shape parameter $\beta = 1$, and $N$ is a standard normal rv independent of $S_{\alpha/2}$ and $S_\alpha$. As previously, the numerator $S_{\alpha/2}$ and the denominator $S_\alpha^b$ are dependent because $\Gamma_j$ in (A 10) is the same as $\Gamma_j$ in (A 11), for each $j = 1, \ldots$.

To prove (3.5), it is enough to prove that for every fixed real $z$, if $Z_i := \sqrt{\rho} z + \sqrt{1 - \rho} N_i$, $i = 1, 2, \ldots$, then

$$W_n \xrightarrow{d} \frac{S_{\alpha/2}}{S_\alpha^b} \left[ (1 - \rho) \exp \left\{ \frac{\rho}{1 - \rho} z^2 \right\} \right]^{1/(2(1-\alpha))}. \tag{3.6}$$

To prove (3.6), we observe that $X_1, X_2, \ldots$ are iid (because they are conditional on fixed $z$) and

$$P(X_1 > t) \sim \sqrt{\frac{2}{\pi}} (1 - \rho)^{-1/2} \exp \left\{ -\frac{\rho}{2(1 - \rho)} z^2 \right\} t^{-\alpha} \text{ as } t \to \infty.$$

In particular, the tail of $X$ is multiplied by the factor of

$$h(\rho) = (1 - \rho)^{-1/2} \exp \left\{ -\frac{\rho}{2(1 - \rho)} z^2 \right\}$$

in comparison with the case $\rho = 0$ in §3(a). That means that, in the limit, every term $\Gamma_j^{-1/\alpha}$ in both $S_{\alpha/2}$ and $S_\alpha$ ends up being multiplied by $h(\rho)^{1/\alpha}$. Therefore, the numerator in the limit in (3.3) gets multiplied by $h(\rho)^{2/\alpha}$, while the denominator gets multiplied by $h(\rho)^{(2-\alpha)/(\alpha(1-\alpha))}$. Doing the arithmetic proves (3.6).

The heavy-tailed version of TL also holds widely for stationary mixing sequences. Here we assume that $\{X_t\}$ is a strictly stationary time series with $X_1 \in RV(\alpha)$, $\alpha \in (0, 1)$. If the time series satisfies the mixing and dependence conditions $D$ and $D'$ in [16], then the limit theorem that holds for TL for the iid case also holds exactly for the stationary case. This follows directly from Davis [16, theorem 4 and its corollary]. The mixing condition $D$ governs how fast certain events become independent as their time separation increases. Condition $D'$ prohibits clustering of pairs of nearby observations when both are large. While condition $D$ is rather mild, condition $D'$ is more restrictive.

Alternatively, in mixing stationary time series that are RV, i.e. where all the finite-dimensional distributions are multivariate RV, the point process convergence in theorem A.1 can be extended, but now the limit point process includes clusters of points that are linked to the Poisson points $\Gamma_i^{-1/\alpha}$ [10,17]. From these point process convergence results, the limiting form of Taylor's Law (3.3) is a ratio of the same two stable rvs with tail indices $\alpha/2$ and $\alpha$ in the numerator and denominator as in (3.3), but the numerator and denominator may need to be multiplied by different constant scale factors.

The results in §§3(b) and (c) may be combined. Explicitly, as in §3(b), suppose that $U \in RV(\alpha)$ with $\alpha \in (0, 1)$ and $U$ dominates $V$ in the sense of (3.4). Assume that the data $X_1, \ldots, X_n$ consist of a random number $\nu_n < n$ of observations distributed as $U$ and $n - \nu_n$ observations distributed as $V$, where $\nu_n/n \xrightarrow{P} p \in (0, 1]$ as $n \to \infty$. In addition, as in §3(c), suppose that the $\nu_n < n$ observations $X_i$ distributed as $U$ satisfy $X_i := 1/(Z_i^2)^{1/(2\alpha)} \in RV(\alpha)$ and, when $X_i, X_j$ are both distributed as $U$, then $\{Z_i\}$ are correlated standard normals with $\text{Cov}(Z_i, Z_j) = \rho$ if $i \ne j \ge 1$. The $n - \nu_n$ of observations

distributed as $V$ may be arbitrarily dependent or independent. Then TL holds as described in (3.5). We simulate this model in §6(a).

## (d) Sample kurtosis

The kurtosis is often used to measure the heaviness of the tail of a distribution function. For a random variable $X$ with finite fourth moment and finite variance $\sigma^2$, the kurtosis is defined by $\kappa = E(X - EX)^4/\sigma^4$. A normal random variable has kurtosis 3. A rv with $\kappa > 3$ is said to be leptokurtic, indicating tails fatter than those of a normal rv. Since we are primarily concerned with heavy-tailed data in which the population variance is infinite, we focus our attention on the kurtosis $\kappa$ as opposed to the excess kurtosis $\kappa - 3$.

If a rv has infinite second moment, then the population kurtosis does not exist. Nevertheless, given a sample $X_1, \ldots, X_n$, one can still compute the sample mean $\bar{X}_n$ and the sample kurtosis

$$\hat{\kappa}_n := \frac{n^{-1} \sum_{t=1}^n (X_t - \bar{X}_n)^4}{\left(n^{-1} \sum_{t=1}^n (X_t - \bar{X}_n)^2\right)^2} = n \frac{\sum_{t=1}^n (X_t - \bar{X}_n)^4}{\left(\sum_{t=1}^n (X_t - \bar{X}_n)^2\right)^2}. \tag{3.7}$$

Assume that $X_1, \ldots, X_n$ are iid as $X \in RV(\alpha)$ with $\alpha \in (0, 2)$. From the argument of example 2.1 (see also theorem A.3), it follows that

$$\left(a_n^{-2} \sum_{t=1}^n (X_t - \bar{X}_n)^2, a_n^{-4} \sum_{t=1}^n (X_t - \bar{X}_n)^4\right) \xrightarrow{d} \left(S_{\alpha/2}, S_{\alpha/4}\right) \tag{3.8}$$

where $S_\gamma = \sum_{i=1}^\infty \Gamma_i^{-1/\gamma}$. By the continuous mapping theorem, we obtain

$$n^{-1} \hat{\kappa}_n \xrightarrow{d} \frac{S_{\alpha/4}}{S_{\alpha/2}^2}. \tag{3.9}$$

By Cauchy's inequality, the limit random variable on the right of (3.9) is bounded above by 1. As the tails get heavier, i.e. as $\alpha \downarrow 0$, the limit random variable becomes more concentrated at 1 because, for $\gamma < 1$,

$$S_\gamma = \Gamma_1^{-1/\gamma} + \Gamma_1^{-1/\gamma} \sum_{i=2}^\infty \left(\frac{\Gamma_i}{\Gamma_1}\right)^{-1/\gamma}.$$

Since each term in the sum is bounded by 1 and $\Gamma_i \sim i$ a.s. as $i \to \infty$, it follows that $S_\gamma \sim \Gamma_1^{-1/\gamma}$ as $\gamma \downarrow 0$. Thus, the limit random variable on the right of (3.9) converges to 1 as $\alpha \downarrow 0$.

If $\alpha \in (2, 4)$, then $X$ has finite variance $\sigma^2$ but infinite fourth moment. In this case, the same argument shows that

$$n a_n^{-4} \hat{\kappa}_n \xrightarrow{d} \frac{S_{\alpha/4}}{[\sigma^2]^2}. \tag{3.10}$$

The rate of convergence, however, is roughly $n^{1-4/\alpha} L(n)$, where $L$ is a slowly varying function, which is rather slow especially as $\alpha \uparrow 4$.

## 4. Heavy tails and sample correlations

The correlation of a pair of rvs is defined as their covariance divided by the product of their standard deviations. It cannot be defined for pairs of rvs with infinite variance, such as all those in $RV(\alpha)$ with $\alpha \in (0, 2)$. Nevertheless, the *sample* correlation (4.4) is a well defined rv. For heavy-tailed time series, the sample autocorrelation can have unexpected and desirable properties [11,12]. Here we give conditions under which the sample correlation of dependent heavy-tailed rvs converges to zero (in probability or almost surely, depending on detailed assumptions) as the sample size gets large.

## (a) Bivariate regular variation and asymptotic independence

We define regular variation for a pair $(X, Y)$ of heavy-tailed rvs. Assume that $X$ and $Y$ are identically distributed with RV distribution $F \in RV(\alpha)$ as in (2.1). Our results require only that their distributions have asymptotically equivalent tails, i.e. $\lim_{s \to \infty} P(X > s)/P(Y > s) = C \in (0, \infty)$. We say that $X$ and $Y$ are asymptotically independent if

$$P(X > t | Y > t) := \frac{P(X > t, Y > t)}{P(Y > t)} \to 0 \quad \text{as } t \to \infty. \tag{4.1}$$

Of course, if $X$ and $Y$ are independent then they are also asymptotically independent. When $t$ is replaced by $a_n$ defined in (2.3), asymptotic independence is equivalent to

$$nP(X > a_n, Y > a_n) \to 0 \quad \text{as } n \to \infty. \tag{4.2}$$

If $X$ and $Y$ are not asymptotically independent, we quantify their dependence in the tails of the distribution by assuming that there exists a measure $\nu(\cdot)$ on $([0, \infty)^2, \mathcal{B}([0, \infty)^2))$ such that, for $A_{x,y} = [0, x] \times [0, y]$ and $A_{x,y}^c$ the set complement of $A_{x,y}$,

$$nP(X > a_n x \text{ or } Y > a_n y) = nP(a_n^{-1}(X, Y) \in A_{x,y}^c) \to \nu(A_{x,y}^c) \text{ as } n \to \infty \tag{4.3}$$

for all $x, y \geq 0$ with $\max\{x, y\} > 0$.

If $x = 0, y > 0$, then $\nu(A_{0,y}^c) = y^{-\alpha}$. Similarly, $\nu(A_{x,0}^c) = x^{-\alpha}$. It follows that $X$ and $Y$ are asymptotically independent if and only if $\nu([x, \infty) \times [y, \infty)) = 0$ for all $x > 0, y > 0$.

## (b) Sample correlation of heavy-tailed data

Define the sample correlation between the pairs $(X_i, Y_i)$, $t = 1, \dots, n$, in the conventional way by

$$\rho_n := \frac{\sum_{t=1}^n (X_t Y_t - \bar{X}_n \bar{Y}_n)}{\sqrt{\sum_{t=1}^n (X_t - \bar{X}_n)^2} \sqrt{\sum_{t=1}^n (Y_t - \bar{Y}_n)^2}}. \tag{4.4}$$

**Theorem 4.1.** *Assume the distribution of $(X, Y)$ satisfies (4.3) and $\alpha \in (0, 2)$. Then*

$$\rho_n \xrightarrow{d} U := \frac{S_{\alpha/2,0}}{\sqrt{S_{\alpha/2,1} S_{\alpha/2,2}}} \quad \text{as } n \to \infty, \tag{4.5}$$

*where $S_{\alpha/2,0}, S_{\alpha/2,1},$ and $S_{\alpha/2,2}$ are stable rvs with joint distribution specified in appendix A(c), (A 18)–(A 20). Moreover,*

$$\rho_n \xrightarrow{P} 0 \quad \text{as } n \to \infty \tag{4.6}$$

*if and only if $X$ and $Y$ are asymptotically independent.*

Appendix A(c) proves this result. Although the stable rvs in the numerator and denominator of (4.5) have heavy tails, the ratio on the right of (4.5) does not have heavy tails because of self-normalization.

When $\rho_n \xrightarrow{P} 0$, i.e. when $X$ and $Y$ are asymptotically independent, it is often possible to find normalizing constants $c_n \to \infty$ such that $c_n \rho_n$ converges in distribution to a non-degenerate rv. For example, when $X$ and $Y$ are independent, $c_n = n^{1/\alpha} L(n)$ for some slowly varying function $L$ [11,12]. Theorem 4.5 gives another example in which $X$ and $Y$ are the reciprocals of powers of correlated bivariate normal rvs with mean 0 and variance 1. Such dependent $X$ and $Y$ are asymptotically independent.

We now extend theorem 4.1 to the case when $X$ and $Y$ are RV rvs with different tail indices. Appendix A(c) gives the proof of theorem 4.2.

**Theorem 4.2.** *If $X \in RV(\alpha)$ and $Y \in RV(\beta)$ with $0 < \alpha < \beta < 2$, then*

$$\rho_n \xrightarrow{d} U := \frac{S_{\alpha\beta/(\alpha+\beta),0}}{\sqrt{S_{\alpha/2,1}S_{\beta/2,2}}} \quad \text{as } n \to \infty, \tag{4.7}$$

*where $S_{\alpha\beta/(\alpha+\beta),0}$, $S_{\alpha/2,1}$ and $S_{\beta/2,2}$ are defined in (A 21), (A 19) and (A 22), respectively. If $X$ and $Y^{\beta/\alpha}$ are asymptotically independent, then $\rho_n \xrightarrow{P} 0$ as $n \to \infty$ since then $U = 0$ a.s.*

We give a quick and easy way to construct dependent, but asymptotically independent, rvs.

**Theorem 4.3.** *Let $U$ and $V$ be two positive rvs with positive and continuous marginal densities in a neighbourhood of zero. Suppose that in a neighbourhood of the origin, $U$ and $V$ have a joint density function $f_{U,V}(u,v)$ satisfying $f_{U,V}(u,v) \le a(u^2 + v^2)^{-\theta/2}$ for some $a > 0$ and $0 \le \theta < 1$. Then, for any $c > 1$ (or $\alpha = 1/c \in (0,1)$), $X := 1/U^c \in RV(1/c)$, $Y := 1/V^c \in RV(1/c)$, $X$ and $Y$ are asymptotically independent, and, for a random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ iid as $(X, Y)$, we have $\rho_n \xrightarrow{P} 0$ as in (4.6).*

*Proof.* We have

$$P(X > x) = P(1/U^c > x) = P(U < x^{-1/c}) = \int_0^{x^{-1/c}} f_U(u)\,\mathrm{d}u \sim x^{-1/c}f_U(0) \quad \text{as } x \to \infty,$$

where $f_U(u)$ is the marginal pdf of $U$. It follows that $X \in RV(1/c)$. A similar argument holds for $Y$. Next,

$$P(X > x, Y > x) = P\left(\frac{1}{U^c} > x, \frac{1}{V^c} > x\right) = \int_0^{x^{-1/c}} \int_0^{x^{-1/c}} f_{U,V}(u,v)\,\mathrm{d}u\mathrm{d}v$$

$$\le a \int_0^{x^{-1/c}} \int_0^{x^{-1/c}} (u^2 + v^2)^{-\theta/2}\,\mathrm{d}u\mathrm{d}v$$

$$\le 2\pi a x^{-(2-\theta)/c}.$$

Dividing the right side by $x^{-1/c}f_V(0)$ and letting $x \to \infty$, the limit becomes zero, proving (4.1). Then theorem 4.1 entails (4.6). ∎

**Corollary 4.4.** *Suppose $(U, V)$ has a bivariate normal distribution with means 0, variances 1 and correlation $\rho \in (-1, 1)$. Set $X := 1/|U|^c, Y := 1/|V|^c$ for $c > 1$. Then $X \in RV(1/c)$, $Y \in RV(1/c)$, and $X$ and $Y$ are asymptotically independent with $\rho_n \xrightarrow{P} 0$ as $n \to \infty$. Further, $\rho_n \xrightarrow{a.s.} 0$ as $n \to \infty$.*

**Theorem 4.5.** *Under the assumptions of corollary 4.4,*

$$P(XY > x) \sim \frac{2}{c\pi(1 - \rho^2)^{1/2}} x^{-1/c} \log x \quad \text{as } x \to \infty.$$

*Moreover,*

$$n\rho_n \xrightarrow{d} -R_1R_2 \quad \text{as } n \to \infty \tag{4.8}$$

*where $R_1$ and $R_2$ are iid rvs with $R_1 = S_\alpha/\sqrt{S_{\alpha/2}}$ and $S_\alpha$ and $S_{\alpha/2}$ are given in (A 10) and (A 11) with $\alpha = 1/c < 1$.*

It is curious that, regardless of $\rho \in (-1, 1)$, $n\rho_n$ converges in distribution to a rv that is negative almost surely. In this situation ($\alpha < 1$), the mean is infinite and the tail of the product $XY$ is similar to the tails of $X$ and $Y$ apart from a slowly varying function. The limit behaviour of the numerator of (4.4) is governed by the product of the partial sums of the $X_i$s and $Y_i$s, which in view of theorem 4.5 is of smaller order. In other words, $\rho_n$ behaves essentially like the product of two $t$-statistics, one for the $X_i$s and one for the $Y_i$s.

Since the mean here is infinite, it is reasonable also to define a version $\tilde{\rho}_n$ of the sample correlation without the correction for the mean:

$$\tilde{\rho}_n := \frac{\sum_{i=1}^n X_i Y_i}{\sqrt{\sum_{i=1}^n X_i^2}\sqrt{\sum_{i=1}^n Y_i^2}}.$$

**Corollary 4.6.** *The assumptions of corollary 4.4 imply*

$$\left(\frac{n}{\log n}\right)^c \tilde{\rho}_n \xrightarrow{d} (1 - \rho^2)^{-c/2} \frac{S_1}{\sqrt{S_2 S_3}} \quad \text{as } n \to \infty, \tag{4.9}$$

*where $S_1$, $S_2$, $S_3$ are independent stable rvs with representations given in (A 10) and indices $\alpha$ equal to $1/c$, $1/(2c)$ and $1/(2c)$, respectively.*

Here the stable rvs in the numerator and denominator of (4.9) have heavy tails and are independent, so the ratio on the right of (4.9) does have heavy tails. We comment after (5.5) below on the confidence intervals implied by the similar situation there (a ratio of independent stable rvs).

Whereas the limit rv in (4.8) is independent of the population correlation $\rho$, the limit rv in (4.9) is scaled by a function of the population correlation $\rho$. The rate of convergence $(n/\log n)^c$ is rather fast, essentially of order $n^2$ for $c = 2$. This is much faster than the standard $n^{1/2}$ rate when the variance is finite. Corollary 4.4, theorem 4.5 and corollary 4.6 are proved in appendix A(d).

## 5. Heavy tails in a linear model

Suppose $X$ and $Z$ are two heavy-tailed rvs and

$$Y = \eta X + Z, \quad \eta > 0. \tag{5.1}$$

Given a sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ of $n$ independent observations, what is a useful way to estimate $\eta$? Surprisingly, even though the data have infinite variance, the least-squares estimator for $\eta$ performs remarkably well. Specifically, the least-squares estimator is weakly consistent, as we now show.

Assume that $X$ and $Z$ are independent positive rvs with distribution functions $F_X$ and $F_Z$, respectively, with asymptotically equivalent tails, i.e.

$$\lim_{s \to \infty} \frac{\bar{F}_Z(s)}{\bar{F}_X(s)} = C \in (0, \infty). \tag{5.2}$$

If $X \in RV(\alpha)$, then it is elementary that

$$\frac{P(Y > x)}{P(X > x)} \to \eta^\alpha + C \quad \text{as } x \to \infty.$$

Thus $Y$ also has a heavy-tailed distribution with the same tail index $\alpha$. However, if $C = \infty$, then $Y$ inherits its heavy tail from the noise $Z$. Here we will assume $C < \infty$.

Since the vector $(X, Z)$ is bivariate regularly varying and since $(X, Y)$ is a linear transformation of $(X, Z)$, it follows directly from Basrak *et al.* [18, p. 113, Proposition A.1] that $(X, Y)$ is bivariate RV.

The angular measure of $(X, Y)$ in (A 5) has mass at $\arctan \eta$ and $\pi/2$, namely,

$$G(\cdot) = \frac{(\eta^2 + 1)^{\alpha/2}}{(\eta^2 + 1)^{\alpha/2} + C} \varepsilon_{\arctan \eta}(\cdot) + \frac{C}{(\eta^2 + 1)^{\alpha/2} + C} \varepsilon_{\pi/2}(\cdot). \tag{5.3}$$

It follows directly from theorem 4.1 that if $(X_1, Y_1), \ldots, (X_n, Y_n)$ are iid observations from the linear model (5.1) with $0 < C < \infty$, then $\rho_n \xrightarrow{d} U$ where $U$ is given in (4.5). On the other hand, if $C = 0$, which would allow the noise variable $Z$ to be light-tailed, including normal, then the form of the limit variable $U$ in (4.5) (see also (A 14)–(A 20)) implies that $\rho_n \xrightarrow{P} 1$ as $n \to \infty$. To see this, the distribution $G$ in (5.3) for the angular part of the limit measure corresponds to point mass at $\arctan \eta$. It follows now that the rvs $\Theta_i$ appearing in (A 18)–(A 20) are all constant and equal to $\arctan \eta$ with probability one. As long as $\eta > 0$, then $\arctan \eta \in (0, \pi/2)$ and hence $S_{\alpha/2,0}^2/(S_{\alpha/2,1} S_{\alpha/2,2}) = 1$.

11

royalsocietypublishing.org/journal/rspa *Proc. R. Soc. A* **476**: 20200610

The least-squares estimator of $\eta$ minimizes $\sum_{t=1}^{n}(Y_t - \eta X_t)^2$. It is given by

$$\hat{\eta} = \frac{\sum_{t=1}^{n} X_t Y_t}{\sum_{t=1}^{n} X_t^2}.$$

Plugging $Y_t = \eta X_t + Z_t$ into the summands in the numerator gives

$$\hat{\eta} - \eta = \frac{\sum_{t=1}^{n} X_t Z_t}{\sum_{t=1}^{n} X_t^2}. \tag{5.4}$$

Applying theorem 4.5 as in the proof of (4.5), it is easy to see that, if $a_n$ is the $1 - 1/n$ quantile of $F_X$ and if $\alpha \in (0, 2)$, then $a_n^{-2} \sum_{t=1}^{n} X_t Z_t \xrightarrow{P} 0$, $a_n^{-2} \sum_{t=1}^{n} X_t^2 \xrightarrow{d} S_{\alpha/2,1}$ as $n \to \infty$, and hence $\hat{\eta}$ is weakly consistent for $\eta$ (i.e., as $n \to \infty$, $\hat{\eta}$ converges in probability to $\eta$ for every $\eta > 0$).

If $\alpha \in (0, 1)$ and $F_X$ is asymptotic to a Pareto distribution $\bar{F}_x(x) \sim K x^{-\alpha}, K \in (0, \infty)$ as $x \to \infty$, then as in corollary 4.6,

$$\left( \frac{n}{\log n} \right)^{1/\alpha} (\hat{\eta} - \eta) \xrightarrow{d} C^{1/\alpha} \frac{S_1}{S_2} \quad \text{as } n \to \infty, \tag{5.5}$$

where $S_1$ and $S_2$ are independent stable rvs given by (A 10) with indices $\alpha$ and $\alpha/2$, respectively. For example, if $\alpha = 1/2$, then the rate of convergence of the least-squares estimate is extremely fast, approximately of order $n^2$, which is substantially faster than the convergence rate of order $\sqrt{n}$ when the variances of $X$ and $Z$ are finite. In particular, an approximate 95% confidence interval for $\eta$, assuming $C = 1$ and $\alpha = 1/2$, is $(\hat{\eta} - c_{0.975}(\log n/n)^2, \hat{\eta} - c_{0.025}(\log n/n)^2)$, where $c_\gamma$ is the $\gamma$-quantile of the distribution for $S_1/S_2$. The two minus signs in the lower and upper bounds of the confidence interval are the surprising consequence of the fact that $\hat{\eta}$ is always bigger than $\eta$, as can be seen from (5.4), even for finite $n$ and as $n \to \infty$. This inequality is an artifact of $X$ and $Z$ being positive. The width of this confidence interval is $(c_{0.975} - c_{0.025})(\log n/n)^2$. On the other hand, if the variances of $X$ and $Z$ are finite, the width of the 95% confidence interval for $\eta$ based on the least squares estimate is $3.92(\sigma_Z/\sigma_X)n^{-1/2}$. While the difference in quantiles $c_{0.975} - c_{0.025}$ coming from the ratio of two independent stable rvs can be large, this potentially large width is more than overcome by the fast rate of the scaling $(\log n/n)^2$ for large $n$.

# 6. Simulations of Taylor's Law, correlations and a linear model

The purpose of this section is to investigate by numerical simulation how rapidly some of the statistics based on dependent heterogeneous, RV rvs described in §§3–5 converge in distribution to their limit rvs expressed as functions of stable laws.

## (a) Taylor's Law with heterogeneous, dependent, heavy tails

We simulate the heterogeneous, dependent, RV rvs described in §§3(b),(c) with three different values of the pairwise correlations $\rho = 0, 0.5, 0.9999$ of the standard normals $Z_t$ and with observations

$$X_t := \frac{1}{(Z_t^2)^{1/(2\alpha_t)}} \in RV(\alpha_t), \ t = 1, 2, \ldots, n.$$

In this numerical example (figure 1), for every tenth value of $t$, $\alpha_t = 0.1$, and the remaining nine of every ten values have $\alpha_t = 0.9$. Despite having nine of every ten exponents equal to 0.9, the numerals $p = 1, \ldots, 7$ plotted at the location of (sample mean, sample variance) = $(\bar{X}_n, \hat{\sigma}_n^2)$ for the three samples of each size $n = 10^p$ fall along the solid line with slope $b = 2.1111 = (2 - 0.1)/(1 - 0.1)$ on log-log coordinates, which is the slope in (3.2) when $\alpha = 0.1$. The solid lines in figure 1 are not fitted to the markers but are calculated *a priori* as the $b$th power of the abscissa. The values of $\alpha_t > 0.1$ have no effect on the slope because the mean and variance are determined by the largest observations, which arise from the RV rvs with the smallest tail exponent, here 0.1.
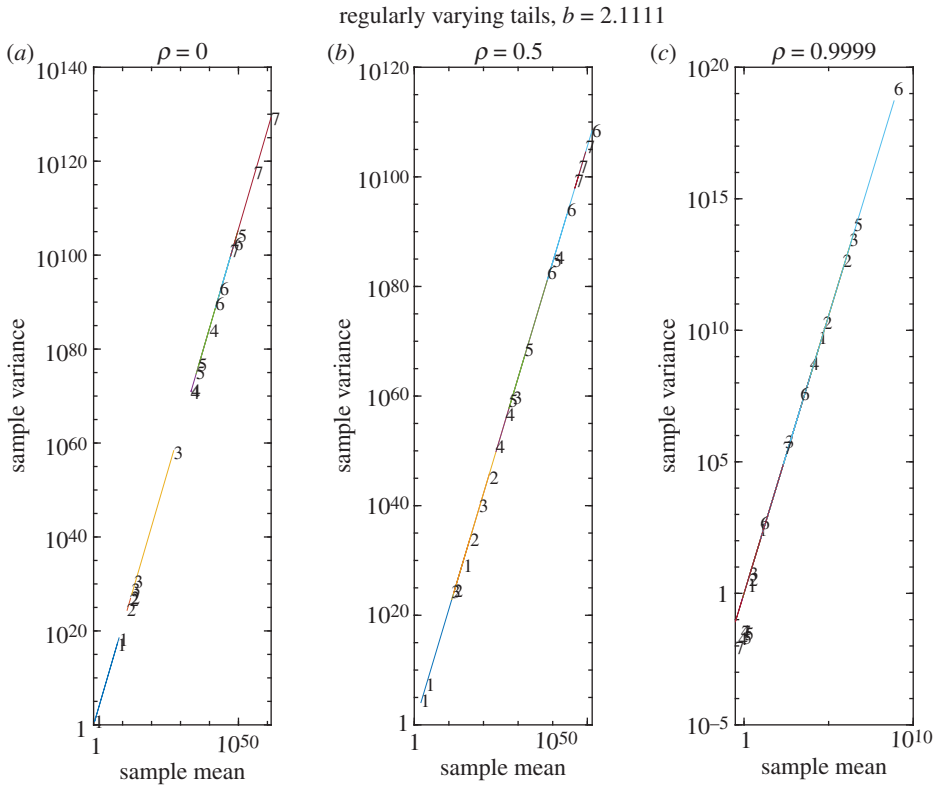
**Figure 1.** Scatterplots on log–log axes of (sample mean $\bar{X}_n$, sample variance $\hat{\sigma}_n^2$) in three samples of $X_t := 1/(2Z_t^2)^{1/[2\alpha_t]} \in RV(\alpha_t), t = 1, 2, \ldots, n$ of each sample size $n = 10^p, p = 1, 2, \ldots, 7$ with pairwise correlation $\rho = 0, 0.5, 0.9999$ (panels $(a)$, $(b)$ and $(c)$). For example, in the left panel, the three appearances of the numeral '7' represent $(\bar{X}_n, \hat{\sigma}_n^2)$ in three samples of size $n = 10^7$. The solid lines $(\bar{X}_s)^b$ have the same slope $b = 2.1111 = (2 - 0.1)/(1 - 0.1)$ on log–log axes, regardless of the value of the pairwise correlation $\rho = 0, 0.5, 0.9999$ (panels $(a)$, $(b)$ and $(c)$). (Online version in colour.)

## (b) Bivariate correlation

In §6(b), we simulate $(X, Y)$ described in corollary 4.4 and theorem 4.5. In this case, although $X$ and $Y$ are generated by arbitrarily correlated (with $|\rho| < 1$) standard normal variates, $X$ and $Y$ are asymptotically independent stable laws. The purpose of these simulations is to shed light on this question: Does the sample correlation $\rho_n$ (4.4) of $n$ iid copies of $(X, Y)$ approach its almost sure limit, 0, or its rescaled limiting distribution (4.8) rapidly enough as $n \to \infty$ that, for sample sizes plausibly encountered in empirical applications, the limiting value or rescaled limiting distribution are useful approximations? We ask a similar question below in §6(c), where we simulate $(X, Y)$ for the linear model in (5.1).

Here we generate $n$ iid pairs $(U, V)$ normally distributed with mean 0, variance 1, and correlation $\rho \in (-1, +1)$, for sample sizes $n = 10^p, p = 1, 2, 3, 4$, and $\rho = +0.9$. From each $(U, V)$ we compute $(X, Y) := (1/|U|^2, 1/|V|^2)$. The $n \times 2$ matrix in which each row is one $(X, Y)$ pair of $1/2$-stable rvs constitutes one simulation. For each sample size $n$, we generate $s = 10^4$ simulations. For each simulation, we record the sample correlation $\rho_n$ between the $n$ realizations of $X$ and the $n$ realizations of $Y$ and calculate $n\rho_n$ (the left side of (4.8)).

Figure 2 shows that the frequency histogram of the sample correlation $\rho_n$ changes from a bimodal distribution with one mode at negative values and another near $\rho_n = 1$ when $n = 10$ (top left) to a unimodal distribution to the left of 0 for large sample sizes $n = 10^4$ (bottom left).

The right side of (4.8) is $-R_1 R_2$ where $R_1$ and $R_2$ are iid rvs with $R_1 = S_\alpha/\sqrt{S_{\alpha/2}}$ and $S_\alpha$ and $S_{\alpha/2}$ are given in (A 10) and (A 11) with $\alpha = 1/c = 1/2$. To simulate each value of $S_\alpha$ and

10 000 simulations of samples of size $n$ of $(X,Y)$, $\rho = 0.9$, $\alpha = 0.5$
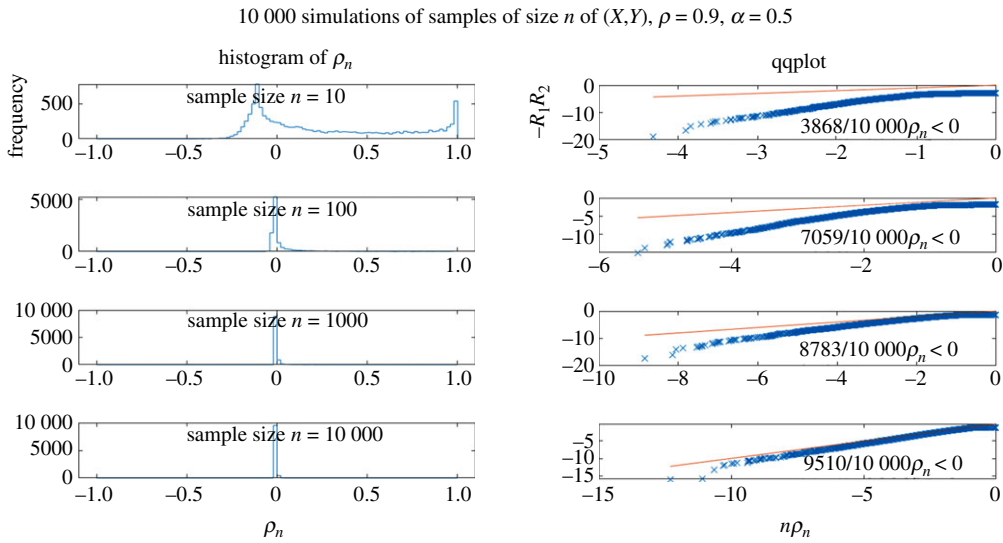


**Figure 2.** For sample sizes $n = 10^p$, $p = 1, 2, 3, 4$, we generate $n$ iid pairs $(U, V)$ normally distributed with mean 0, variance 1 and correlation $\rho = +0.9$ and then $(X, Y) := (1/|U|^2, 1/|V|^2)$. The $n \times 2$ matrix in which each row is one $(X, Y)$ pair constitutes one simulation. For each sample size $n$, we generate $s = 10^4$ simulations. For each simulation, we record the sample correlation $\rho_n$ between the $n$ realizations of $X$ and the $n$ realizations of $Y$. The left column of this figure shows the frequency histogram of the $s = 10^4$ simulations of $\rho_n$, for sample sizes from $n = 10$ (top row) to $n = 10^4$ (bottom row). The histogram becomes increasingly concentrated on $(-1, 0)$. The right column of this figure shows, on the ordinate with cross symbols, the order statistics of $-R_1 R_2$, the limit rv on the right side of (4.8), and, on the abscissa and in the diagonal solid line of slope 1, the corresponding order statistics of $n\rho_n$ (see text for details of computation). For samples of size $n = 10$, the sample correlation $\rho_n$ was negative in 3868 of 10 000 samples. For samples of size $n = 10^4$, the sample correlation $\rho_n$ was negative in 9510 of 10 000 samples. The plots on the right include only and all the negative values of $\rho_n$ and the corresponding order statistics of $-R_1 R_2$. (4.8) predicts, as $n \to \infty$, that the fraction of simulations in which $\rho_n < 0$ should approach 1, and that the cross symbols markers should approach the diagonal line. The right panels support both predictions. (Online version in colour.)

its corresponding $S_{\alpha/2}$, we generate $10^5$ iid unit exponential variates $E_i$, compute the $10^5$ partial sums $\Gamma_t := \sum_{i=1}^{t} E_i, t = 1, \ldots, 10^5$, raise each $\Gamma_t$ to the power $-1/\alpha$ (for $S_\alpha$) or $-2/\alpha$ (for $S_{\alpha/2}$), and sum the $10^5$ values of $\Gamma_t^{-1/\alpha}$ or $\Gamma_t^{-2/\alpha}$. From each such pair $(S_\alpha, S_{\alpha/2})$ using the same values of $\Gamma_t$, we compute one value of $R_1 = S_\alpha / \sqrt{S_{\alpha/2}}$. An independent repetition of this procedure generates one iid value of $R_2$ and thence one value of $-R_1 R_2$.

For each sample size $n$, we sort the $s$ simulated values of $n\rho_n$ (the left side of (4.8)) in increasing order, and we sort the $s$ simulated values of $-R_1 R_2$ (the right side of (4.8)) in increasing order. If the distribution $n\rho_n$ approaches the distribution of $-R_1 R_2$ as $n \to \infty$ as theorem 4.5 asserts, then the empirical quantile–quantile plot of $-R_1 R_2$ on the vertical axis and $n\rho_n$ on the horizontal axis should fall along a diagonal straight line of slope 1 through the origin.

However, the absolute error $n\rho_n - (-R_1 R_2)$ of the approximation in (4.8) is unbounded as $n \to \infty$ because there is a non-zero (though asymptotically vanishing) probability that $\rho_n$ will be near 1, which case $n\rho_n$ will be near $n$ while $-R_1 R_2$ is negative with probability 1, so the error could be close to $n$. Consequently, including positive values of $\rho_n$ in the quantile–quantile plots in the right column of figure 2 would obscure the comparison of $n\rho_n$ with $-R_1 R_2$ on $(-\infty, 0]$ where asymptotically an increasing proportion of the values of $\rho_n$ will fall. For each sample size $n$, we report the number of simulations with $\rho_n < 0$ in the lower right corner of each panel in the right column of figure 2 and we restrict the quantile–quantile plot to these values and to the corresponding order statistics of $-R_1 R_2$. As (4.8) predicts, as $n \to \infty$, the fraction of simulations

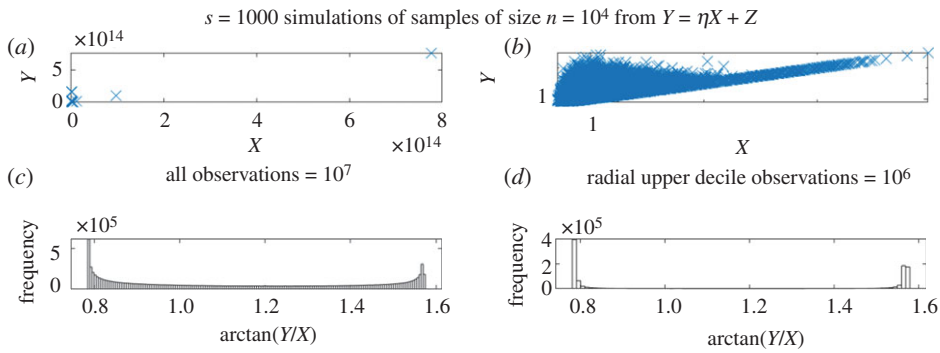**Figure 3.** Left to right, top to bottom: (*a*) Scatterplot on linear scales of $(X, Y)$, $Y = X + Z$ with $X, Z$ iid 1/2-stable, in a single sample of size $n = 10^4$. (*b*) Scatterplot on logarithmic scales of the same sample shown on the left. (*c*) Frequency histogram of $n \times s = 10^4 \times 10^3 = 10^7$ values of arctan($Y/X$) from the pooled $s$ simulations of samples of size $n$. The modal values approximate arctan(1) on the left and $\pi/2$ on the right. (*d*) Frequency histogram of arctan($Y/X$) from the $10^6$ values of $(X, Y)$ for which the radius in polar coordinates $R := (X^2 + Y^2)^{1/2}$ falls in the upper decile of all $10^7$ values of the radius. The values of arctan($Y/X$) are more concentrated at arctan(1) and $\pi/2$. (Online version in colour.)

in which $\rho_n < 0$ approaches 1 and the order statistics of $n\rho_n$ approach the corresponding order statistics of $-R_1 R_2$.

## (c) Linear model

We simulate the linear model (5.1) with slope $\eta = 1$ under two different assumptions: (i) that the independent variable $X$ and the noise (or random perturbation) variable $Z$ are iid 1/2-stable, so that in (5.2) we have $C = 1$; and (ii) that $X$ is 1/2-stable and $Z$ is standard normal, which is light-tailed, and $X$ and $Z$ are independent, so that in (5.2) $C = 0$.

Under (i), assuming $X$ and $Z$ are iid $\alpha$-stable with $\alpha = 1/2$, $\eta = C = 1$, the angular measure (5.3) simplifies to

$$G(\cdot) = \frac{2^{1/4}}{2^{1/4} + 1}\varepsilon_{\arctan 1}(\cdot) + \frac{1}{2^{1/4} + 1}\varepsilon_{\pi/2}(\cdot)$$

$$\approx 0.5432\varepsilon_{0.7854}(\cdot) + 0.4568\varepsilon_{1.5708}(\cdot).$$

The limiting distribution of $\rho_n$, given by the right side of (4.5), depends on the three stable rvs $S_{1/4,0}, S_{1/4,1}$ and $S_{1/4,2}$ with joint distribution specified in appendix A(c), (A 18)–(A 20). To simulate their joint distribution, for each realization $\omega$, each $\Gamma_t$ is a cumulative sum of $t$ iid exponential rv with parameter 1, as described in greater detail in §6(b). From (5.3), in the limit of large $n$, each $\Theta_t = \arctan(1) \approx 0.7854$ with probability 0.5432 and $\Theta_t = \pi/2 \approx 1.5708$ with probability 0.4568, as calculated numerically just above. We sum the summands on the right sides of (A 18)–(A 20) until the sums quasi-converge.

Under (ii), with $\alpha = 1/2, \eta = 1$, $C = 0$, the angular measure (5.3) simplifies to

$$G(\cdot) = \varepsilon_{\arctan 1}(\cdot) \approx \varepsilon_{0.7854}(\cdot)$$

and the limiting distribution of $\rho_n$ is a point mass at 1.

As above, $n = 10^p, p = 1, 2, 3, 4$, denotes the sample size or number of realizations of $(X, Y)$. For each sample size $n$, we simulate $s = 1000$ samples each with $n$ copies of $(X, Y)$ and plot the frequency histogram of the $s$ values of the sample correlation coefficient $\rho_n$ for these $s$ samples. We compare this frequency histogram with the limiting distribution calculated in §5.
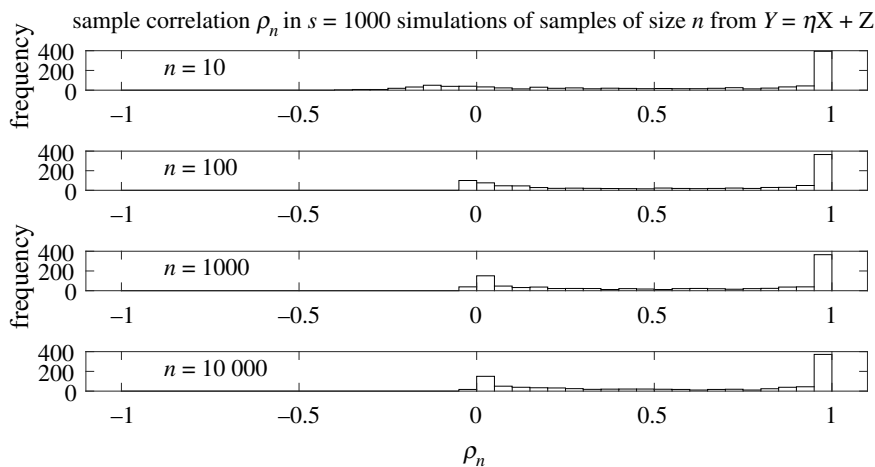
**Figure 4.** Frequency histograms of the sample correlation coefficient $\rho_n$ of samples of $(X, Y)$, $Y = \eta X + Z$ when $Z$ is 1/2-stable (heavy-tailed) as in figure 3. We assume $\eta = 1$. Top to bottom: Sample sizes are (a) $n = 10$, (b) $n = 10^2$, (c) $n = 10^3$, and (d) $n = 10^4$. Modal sample correlation coefficients are near 0 and near 1.
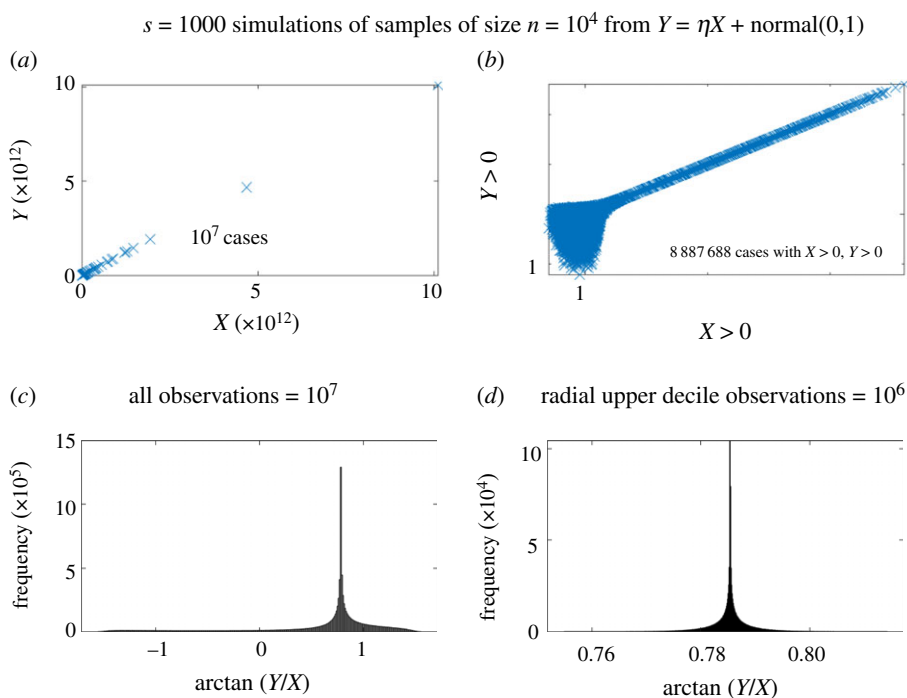


**Figure 5.** Left to right, top to bottom: (a) Scatterplot on linear scales of $(X, Y)$ where $Y = X + U$ with $X$ iid 1/2-stable, $U$ iid standard normal. We have $s = 10^3$ samples each of size $n = 10^4$, giving a total of $n \times s = 10^4 \times 10^3 = 10^7$ pairs $(X, Y)$. (b) Scatterplot on logarithmic scales of only the 8 887 329 pairs $(X, Y)$ where $X > 0, Y > 0$. The same simulations appear in (A) and (B) but only the positive pairs are selected for (B). (c) Frequency histogram of $10^7$ values of $\arctan(Y/X)$ from $s$ iid simulations of samples of size $n$. The modal values centre on $\arctan(1) = \pi/4 \approx 0.7854$. (d) Frequency histogram of $\arctan(Y/X)$ from the $10^6$ values of $(X, Y)$ for which the radius in polar coordinates $R := (X^2 + Y^2)^{1/2}$ falls in the upper decile of all $10^7$ values. The values are more concentrated at $\arctan(1)$. (Online version in colour.)

| $n$ | medianheavyZ | rangeheavyZ | medianlightZ | rangelightZ |
|-----|--------------|-------------|--------------|-------------|
| 10 | 0.8254 | 1.3622 | 0.99981 | 0.36106 |
| 100 | 0.68704 | 1.0494 | 1 | 0.0011082 |
| 1000 | 0.59068 | 1.0035 | 1 | $6.1644 \times 10^{-7}$ |
| 10 000 | 0.6971 | 1.0003 | 1 | $3.1278 \times 10^{-10}$ |

**Figure 6.** Column 1: sample size $n$. Column 2: median of the $s = 1000$ values of the sample correlation coefficient $\rho_n$ of $(X, Y)$, $Y = \eta X + Z$ with $X$, $Z$ iid 1/2-stable, $\eta = 1$. The noise term is heavy–tailed. Column 3: range (highest minus lowest) of the $s = 1000$ values of the sample correlation coefficient in column 2. Column 4: median of the $s = 1000$ values of the sample correlation coefficient of $(X, Y)$, $Y = \eta X + U$ with $X$ iid 1/2-stable, $U$ iid standard normal, $\eta = 1$. The noise term is light–tailed. Column 5: range (highest minus lowest) of the $s = 1000$ values of the sample correlation coefficient of the sample correlation coefficient in column 4. With heavy-tailed noise, the sample correlation coefficient is increasingly covers the full interval $[0, 1]$ as sample size $n$ increases. With light-tailed noise, the sample correlation coefficient is increasingly tightly concentrated near 1 as sample size $n$ increases.

To simulate $s$ samples of size $n$ of the 1/2-stable rv $X$, we let each element of the $n \times s$ matrix **X** be iid distributed as the reciprocal of the square of an iid standard normal. Each column of **X** represents one sample of size $n$ of $X$.

Under assumption (i) that $X$ and $Z$ are iid 1/2-stable, we also simulate $s$ samples of size $n$ of the 1/2-stable rv $Z$ by exactly the same procedure, independently of **X**. Each column of the $n \times s$ matrix **Z** represents one sample of size $n$ of $Z$. Then $Y = \eta X + Z$ gives an $n \times s$ matrix containing $s$ samples of size $n$ of $Y$ according to (5.1). We assume $\eta = 1$ here, so $Y = X + Z$. The top row of figure 3 shows one sample of size $n = 10^4$ under assumption (i). The modal values of arctan$(Y/X)$ from $s$ iid simulations of samples of size $n = 10^4$ (second row of figure 3) approximate arctan(1) on the left and $\pi/2$ on the right (lower left panel), with increasing concentration at these modal values when the radius in polar coordinates $R := (X^2 + Y^2)^{1/2}$ falls in the upper decile of all $10^7$ values of $R$ (lower right panel). For samples of increasing size, the sample correlation coefficient $\rho_n$ is increasingly concentrated on the interval $[0, 1]$, with modal frequencies at the extremes (figures 4 and 6).

Under assumption (ii) that $X$ is 1/2-stable, $Z$ is standard normal, and $X$ and $Z$ are independent, we generate another independent $n \times s$ matrix **X** as above and an $n \times s$ matrix **N** with iid standard normal elements $N$. Then $Y = \eta X + N = X + N$. The top row of figure 5 shows one simulated sample of size $n = 10^4$ under assumption (ii). The top right panel shows on log–log coordinates only the pairs $(X, Y)$ of this sample where $X > 0, Y > 0$. Unlike the previous case of heavy-tailed noise, here the modal values of arctan$(Y/X)$ from the pooled $s$ simulations of samples of size $n = 10^4$ (lower row of figure 5) centre on and are near arctan(1) $= \pi/4 \approx 0.7854$, with increasing concentration at this modal value when the radius in polar coordinates $R := (X^2 + Y^2)^{1/2}$ falls in the upper decile of all $10^7$ values of $R$ (lower right panel). Under assumption (ii), for samples of increasing size, the sample correlation coefficient $\rho_n$ is concentrated around 1, with a range that decreases very rapidly toward 0 (figure 6): these results are consistent with the theory and figure 5c,d where the histogram of arctan$(Y/X)$ is heavily concentrated at $\pi/4$.

# Appendix A

## (a) Heavy-tailed distributions and point process methods: background

Point process methods provide indispensable insight into limit theory for statistics arising from heavy-tailed data. The methods have been championed by Resnick [1,2,19], and others. We will discuss the relevant theory in the setting of §4 and leave the general formulation to the references in Resnick [2].

Here the Euclidean space $E$ is defined as $E = (0, \infty]$ or $(0, \infty]^2 \setminus (0, 0)$.[1] A point measure $\xi$ on $E$ is a non-negative integer-valued measure on the Borel $\sigma$-field, $\mathcal{B}(E)$, having the form

$$\xi(\cdot) = \sum_{t=1}^{\infty} \varepsilon_{x_t}(\cdot),$$

where $\varepsilon_x(\cdot)$ is the delta measure with unit mass at $x$, and $x_i \in E$. That is, for $A \in \mathcal{B}(E)$,

$$\varepsilon_x(A) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

We say that a point measure is Radon on $E$ if it is finite on all compact sets.

We define the vague topology on the class $M_p(E)$ of Radon point measures in terms of the space $C_K(E)$ of continuous functions on $E$ with compact support. If $\xi_n, \xi \in M_p(E)$, then $\xi_n \xrightarrow{v} \xi$ if and only if for every $f \in C_K(E)$,

$$\xi_n(f) := \int_E f \, d\xi_n \to \xi(f) := \int_E f \, d\xi.$$

A point process is a random element of $M_p(E)$ defined on some probability space $(\Omega, \mathcal{F}, P)$. A Poisson point process $\xi$ is defined in terms of a Radon intensity measure $\mu$ and is often called a Poisson random measure, denoted by $\mathrm{PRM}(\mu)$ [1]. Such a point process satisfies the following assumptions:

(i) For relatively compact $A \in \mathcal{B}(E)$, $\xi(A)$ has a Poisson distribution with mean $\mu(A)$.
(ii) For relatively compact and disjoint $A, B \in \mathcal{B}(E)$, $\xi(A)$ and $\xi(B)$ are independent.

The Laplace functional $L_\xi(f)$ of a point process $\xi(\cdot) = \sum_{t=1}^{\infty} \varepsilon_{X_t}(\cdot)$ is

$$L_\xi(f) := E(\exp\{-\xi(f)\}) = E\left(\exp\left\{-\sum_{t=1}^{\infty} f(X_t)\right\}\right) \tag{A 1}$$

for non-negative $f \in C_K(E)$. When $\xi$ is a $\mathrm{PRM}(\mu)$, the Laplace functional becomes [1, Sect. 5.3.2]

$$L_\xi(f) = \exp\left\{-\int_E \left(1 - e^{-f}\right) d\mu\right\}. \tag{A 2}$$

A sequence of point processes $\xi_n$ converges in distribution to the point process $\xi$ if and only if their respective Laplace functionals also converge, i.e.

$$\xi_n \xrightarrow{d} \xi \quad \text{if and only if } L_{\xi_n}(f) \to L_\xi(f) \text{ for all non-negative } f \in C_K(E). \tag{A 3}$$

A powerful result about the convergence of point processes of heavy-tailed one-dimensional data is:

**Theorem A.1.** *Suppose $\{X_t\}$ is an iid sequence of observations with distribution $F \in RV(\alpha)$ as in (2.1), and define $a_n$ by (2.3). Let $\{E_t\}$ be an iid sequence of unit exponential rvs, and define $\Gamma_t := \sum_{i=1}^{t} E_i, t =$*

---

[1] Here $E$ corresponds to a one-point compactification of $[0, \infty)$ or $[0, \infty)^2$ with the origin removed. Relatively compact sets are sets bounded away from the origin, i.e. $A \in \mathcal{B}(E)$ is relatively compact if, for some small $\epsilon > 0$, $A \subset \{x : \epsilon \le x\}$ or in the two-dimensional case $A \subset \{(x, y) : \epsilon \le \|(x, y)\|\}$.

$1, 2, \ldots$. *Then the sequence of point processes* $N_n(\cdot)$ *defined on the left in* (A 4) *converges in distribution to the point process* $N(\cdot)$ *defined on the right in* (A 4) *as* $n \to \infty$:

$$N_n(\cdot) := \sum_{t=1}^{n} \varepsilon_{a_n^{-1} X_t}(\cdot) \xrightarrow{d} N(\cdot) := \sum_{t=1}^{\infty} \varepsilon_{\Gamma_t^{-1/\alpha}}(\cdot). \tag{A 4}$$

*The limit point process N is a* $\mathrm{PRM}(\alpha x^{-\alpha-1}\,\mathrm{d}x)$.

This result follows directly from Resnick [1, theorem 5.3].

We recall from elementary stochastic processes that the $\Gamma_t$'s are points of a homogeneous Poisson process on $(0, \infty)$ with rate one. By a change of variables, the points $\Gamma_t^{-1/\alpha}$ are then points of a Poisson process with intensity measure $\alpha x^{-\alpha-1}\,\mathrm{d}x$.

To consider sample correlations between pairs of rvs, we extend regular variation to bivariate random vectors. The nonnegative bivariate random vector $(X, Y)$ is said to be RV if the radial part $R = (X^2 + Y^2)^{1/2}$ is $RV(\alpha)$ and the angular part $\Theta = \arctan(Y/X)$ becomes independent of $R$ as $R \to \infty$. More precisely, if $\tilde{a}_n$ is the $1 - 1/n$ quantile of $R$ and $G$ is a distribution function on $[0, \pi/2]$, then for all $r > 0$,

$$nP(R > \tilde{a}_n r) \to r^{-\alpha} \quad \text{and} \quad P(\Theta \in \cdot \mid R > \tilde{a}_n) \xrightarrow{d} G(\cdot) \quad \text{as } n \to \infty. \tag{A 5}$$

The limiting relations (A 5) are equivalent to

$$nP((\tilde{a}_n^{-1} R, \Theta) \in (\cdot, \cdot)) \xrightarrow{v} \tilde{v}(\mathrm{d}r, \mathrm{d}\theta) := \alpha r^{-\alpha-1}\,\mathrm{d}r \times G(\mathrm{d}\theta) \quad \text{as } n \to \infty \tag{A 6}$$

in polar coordinates or, in the original Euclidean coordinates,

$$nP(\tilde{a}_n^{-1}(X, Y) \in (\cdot, \cdot)) \xrightarrow{v} v(\mathrm{d}x, \mathrm{d}y) \quad \text{as } n \to \infty, \tag{A 7}$$

where $\xrightarrow{v}$ represents vague convergence of measures on $(0, \infty]^2$ and $v(\mathrm{d}x, \mathrm{d}y)$ is the measure in rectangular coordinates corresponding to the polar coordinate version $\tilde{v}(\mathrm{d}r, \mathrm{d}\theta)$.

The point process convergence that drives much of the limit theory for bivariate heavy-tailed data is:

**Theorem A.2.** *Suppose* $\{(X_t, Y_t), t = 1, 2, \ldots\}$ *is an iid sequence of observations with distribution that satisfies* (A 6) *or* (A 7). *Then the sequence of point processes* $N_n(\cdot)$ *converges in distribution as* $n \to \infty$ *to a* $\mathrm{PRM}(v)$ $N(\cdot)$, *where*

$$N_n(\cdot) := \sum_{t=1}^{n} \varepsilon_{\tilde{a}_n^{-1}(X_t, Y_t)}(\cdot) \xrightarrow{d} N(\cdot) := \sum_{t=1}^{\infty} \varepsilon_{\Gamma_t^{-1/\alpha}(\cos \Theta_t, \sin \Theta_t)}(\cdot) \quad \text{as } n \to \infty. \tag{A 8}$$

*Here* $\{\Theta_t\}$ *is an iid sequence with distribution* $G(\mathrm{d}\theta)$ *in* (A 6) *that is also independent of the sequence* $\{\Gamma_t\}$.

The proof of this result is in Resnick [1, theorem 5.3].

The limit point process $N(\cdot)$ in theorem A.2 has an instructive representation. The $\Gamma_j^{-1/\alpha}$ and $\Theta_j$ correspond to the limit of the radial and angular parts, respectively, of the pairs $\tilde{a}_n^{-1}(X_j, Y_j)$. To verify that $N$ is a $\mathrm{PRM}(v)$, we use Laplace functionals. Specifically, by conditioning on the Poisson

points $\Gamma_j$ and by using the independence of the $\Theta_i$, we have

$$
\begin{aligned}
L_N(f) &= E\left(\exp\left\{-\sum_{j=1}^{\infty} f\left(\Gamma_j^{-1/\alpha}(\cos\Theta_j, \sin\Theta_j)\right)\right\}\right) \\
&= E\left(\prod_{j=1}^{\infty}\int_{[0,\pi/2]}\left(\exp\left\{-f\left(\Gamma_j^{-1/\alpha}(\cos\theta, \sin\theta)\right)\right\}G(d\theta)\right)\right) \\
&= E\left(\exp\left\{\sum_{j=1}^{\infty} h(\Gamma_j^{-1/\alpha})\right\}\right), \quad \left[h(r) := -\log\int_{[0,\pi/2]}\left(\exp\{-f\left(r(\cos\theta, \sin\theta)\right)\}G(d\theta)\right)\right] \\
&= \exp\left\{-\int_{(0,\infty)}(1-\exp\{-h(r)\})\alpha r^{-\alpha-1}\,dr\right\} \\
&= \exp\left\{-\int_{(0,\infty)}(1-\exp\{-f(x,y)\})\nu(dx,dy)\right\}
\end{aligned}
$$

as desired.

## (b) Limit theory for Taylor's Law

The following theorem describes the joint convergence of the partial sums and partial sums of squares in example 2.1 and the limiting behaviour of Taylor's Law.

**Theorem A.3.** *Suppose $\{X_t\}$ is an iid sequence of observations with $F \in RV(\alpha)$ as in (2.1). Define $a_n$ by (2.3) and $b_n$ by (2.5). Then*

$$
\left(a_n^{-1}\sum_{t=1}^{n}(X_t - b_n), a_n^{-2}\sum_{t=1}^{n}(X_t - \bar{X}_n)^2\right) \xrightarrow{d} \left(S_\alpha, S_{\alpha/2}\right) \quad as\ n \to \infty. \tag{A 9}
$$

*The limit rvs $S_\alpha$ and $S_{\alpha/2}$ are stable and defined through the points of the point process $N$ in theorem A.2. Namely,*

$$
S_\alpha = \begin{cases}
\sum_{j=1}^{\infty}\left(\Gamma_j^{-1/\alpha} - E[\Gamma_j^{-1/\alpha}\mathbf{1}_{\{\Gamma_j^{-1/\alpha} \le 1\}}]\right), & if\ \alpha \in [1,2), \\
\sum_{j=1}^{\infty}\Gamma_j^{-1/\alpha}, & if\ \alpha \in (0,1),
\end{cases} \tag{A 10}
$$

*and*

$$
S_{\alpha/2} = \sum_{j=1}^{\infty}\Gamma_j^{-2/\alpha}. \tag{A 11}
$$

*For $\alpha \in (0,1)$, Taylor's Law (3.3) follows: for $b = (2-\alpha)/(1-\alpha)$,*

$$
\frac{na_n^{-2}}{n^b a_n^{-b}}W_n = \frac{1}{\tilde{L}^{2-b}(n)}W_n \xrightarrow{d} \frac{\sum_{i=1}^{\infty}\Gamma_i^{-2/\alpha}}{\left(\sum_{i=1}^{\infty}\Gamma_i^{-1/\alpha}\right)^b} \quad as\ n \to \infty. \tag{A 12}
$$

*Proof.* This is a special case of a standard result in, for example, LePage *et al.* [20, theorems 1 and 1'], where the derivation of (2.10) from (2.8) is required. ∎

The representations in (A 10) and (A 11) are often referred to as the LWZ representation of a stable distribution. Davis [16] gives the companion result when $\{X_t\}$ is a stationary time series satisfying certain dependence conditions.

## (c) Limit theory for the sample correlation

*Proof of theorem* 4.1. The condition (4.3) implies the vague convergence in (A 7). Indeed, (4.3) implies that (A 7) holds for sets of the form $A_{x,y}^c = ([0,x] \times [0,y])^c$, which can then be extended to

all two-dimensional Borel sets that are bounded away from $(0, 0)$. Moreover, the constants $a_n$ and $\tilde{a}_n$ are related via

$$nP(R > \tilde{a}_n) \to 1 \quad \text{and} \quad nP(R > a_n) \to \nu(\{(x, y) : x^2 + y^2 > 1\}) \quad \text{as } n \to \infty,$$

which implies

$$\tilde{a}_n \sim \nu^{1/\alpha}(\{(x, y) : x^2 + y^2 > 1\}) a_n \quad \text{as } n \to \infty. \tag{A 13}$$

For $0 < \epsilon < 1$ fixed, consider the function $f_\epsilon(x, y) = xy \mathbf{1}_{\{\epsilon < xy \le 1/\epsilon\}}$, which has compact support on $E$ and is continuous except on a set of $\nu$ measure 0. By the weak convergence of $N_n$ to $N$ in theorem A.2, it follows that

$$N_n(f_\epsilon) = \tilde{a}_n^{-2} \sum_{t=1}^{n} X_t Y_t \mathbf{1}_{\{\tilde{a}_n^2 \epsilon < X_t Y_t \le \tilde{a}_n^2/\epsilon\}}$$

$$\xrightarrow{d} N(f_\epsilon) = \sum_{t=1}^{\infty} \Gamma_t^{-2/\alpha} \cos \Theta_t \sin \Theta_t \mathbf{1}_{\{\epsilon < \Gamma_t^{-2/\alpha} \cos \Theta_t \sin \Theta_t \le 1/\epsilon\}} \quad \text{as } n \to \infty. \tag{A 14}$$

Since $\Gamma_t \sim t$ a.s. by the strong law of large numbers, the series $\sum_{t=1}^{\infty} \Gamma_t^{-2/\alpha}$ is summable a.s. Hence

$$N(f_\epsilon) \xrightarrow{a.s.} N(f_0) = \sum_{t=1}^{\infty} \Gamma_t^{-2/\alpha} \cos \Theta_t \sin \Theta_t \quad \text{as } \epsilon \downarrow 0. \tag{A 15}$$

Moreover, for any $\delta > 0$,

$$P(|N_n(f_\epsilon) - N_n(f_0)| > \delta) \le P\left(\tilde{a}_n^{-2} \sum_{t=1}^{n} X_t Y_t \mathbf{1}_{\{X_t Y_t > \tilde{a}_n^2/\epsilon\}} > \frac{\delta}{2}\right)$$

$$+ P\left(\tilde{a}_n^{-2} \sum_{t=1}^{n} X_t Y_t \mathbf{1}_{\{X_t Y_t \le \tilde{a}_n^2 \epsilon\}} > \frac{\delta}{2}\right)$$

$$= I + II.$$

Term $I$ is bounded by

$$P\left(\bigcup_{t=1}^{n} \left\{\tilde{a}_n^{-2} X_t Y_t > \frac{1}{\epsilon}\right\}\right) \le nP\left(\tilde{a}_n^{-2} X_1 Y_1 > \frac{1}{\epsilon}\right)$$

$$\xrightarrow{n \to \infty} \nu\left\{(x, y) : xy \ge \frac{1}{\epsilon}\right\} \to 0 \quad \text{as } \epsilon \to 0. \tag{A 16}$$

Hence $\lim_{\epsilon \to 0} \limsup_{n \to \infty} I = 0$. Markov's inequality bounds term $II$ by

$$II \le 2\delta^{-1} n \tilde{a}_n^{-2} E\left[X_1 Y_1 \mathbf{1}_{\{X_1 Y_1 \le \tilde{a}_n^2 \epsilon\}}\right].$$

Next

$$\frac{P(XY > a_n^2)}{P(X^2 > a_n^2)} = \frac{nP(XY > a_n^2)}{nP(X^2 > a_n^2)} \to \nu(\{(x, y) : xy > 1\}) \quad \text{as } n \to \infty,$$

where the limit is positive if $X$ and $Y$ are asymptotically dependent, or zero if $X$ and $Y$ are asymptotically independent. Either way, it follows that there exists a finite positive constant $C$ such that, for all $z > 0$,

$$P(XY > z) \le CP(X^2 > z).$$

Consequently

$$E\left[XY \mathbf{1}_{\{XY \le \tilde{a}_n^2 \epsilon\}}\right] = \int_0^{\tilde{a}_n^2 \epsilon} P(XY > z)\, dz \le \int_0^{\tilde{a}_n^2 \epsilon} CP(X^2 > z)\, dz = CEX^2 \mathbf{1}_{\{X^2 \le \tilde{a}_n^2 \epsilon\}}.$$

Now from Karamata's theorem [1, theorem 2.1], we have $E(X^2 \mathbf{1}_{\{X^2 \leq z\}}) \sim 2/(2 - \alpha) z P(X^2 > z)$ as $z \to \infty$ and hence for some $K \in (0, \infty)$,

$$\lim_{\epsilon \to 0} \limsup_{n \to \infty} II \leq \lim_{\epsilon \to 0} \limsup_{n \to \infty} 2 C \delta^{-1} n \tilde{a}_n^{-2} E\left[ X^2 \mathbf{1}_{\{X^2 \leq \tilde{a}_n^2 \epsilon\}} \right] = \lim_{\epsilon \to 0} K \epsilon^{1 - \alpha/2} = 0. \tag{A 17}$$

Combining (A 14), (A 16) and (A 17), we conclude that

$$\tilde{a}_n^{-2} \sum_{t=1}^{n} X_t Y_t \xrightarrow{d} S_{\alpha/2,0} := \sum_{t=1}^{\infty} \Gamma_t^{-2/\alpha} \cos \Theta_t \sin \Theta_t \quad \text{as } n \to \infty. \tag{A 18}$$

Essentially, the same point process argument, although somewhat easier, can be used to show

$$\tilde{a}_n^{-2} \sum_{t=1}^{n} X_t^2 \xrightarrow{d} S_{\alpha/2,1} := \sum_{t=1}^{\infty} \Gamma_t^{-2/\alpha} \cos^2 \Theta_t \quad \text{as } n \to \infty \tag{A 19}$$

and

$$\tilde{a}_n^{-2} \sum_{t=1}^{n} Y_t^2 \xrightarrow{d} S_{\alpha/2,2} := \sum_{t=1}^{\infty} \Gamma_t^{-2/\alpha} \sin^2 \Theta_t \quad \text{as } n \to \infty, \tag{A 20}$$

where the convergences in (A 18)–(A 20) are joint. Applying the continuous mapping theorem gives (4.5).

Finally, by the form of the limit, one sees that the numerator $S_{\alpha/2,0} = 0$ a.s. if and only if the angular measure concentrates at 0 and $\pi/2$, which is equivalent to asymptotic independence. ∎

*Proof of theorem 4.2.* If $X \in RV(\alpha)$ and $Y \in RV(\beta)$ with $0 < \alpha < \beta < 2$, then $Y^{\beta/\alpha} \in RV(\alpha)$. So replacing the condition on $(X, Y)$ in theorem 4.1 with the assumption that $(X, Y^{\beta/\alpha})$ is bivariate RV, theorem A.2 implies that the same point process convergence in (A 8) holds with $Y_t$ replaced by $Y_t^{\beta/\alpha}$. Using the same proof as for theorem 4.1 with the function $f_\epsilon(x, y) = xy^{\alpha/\beta} \mathbf{1}_{\{\epsilon < xy^{\alpha/\beta} \leq 1/\epsilon\}}$ implies the analogue of (A 18), namely,

$$\tilde{a}_n^{-1-\alpha/\beta} \sum_{t=1}^{n} X_t Y_t \xrightarrow{d} S_{\alpha\beta/(\alpha+\beta),0} := \sum_{t=1}^{\infty} \Gamma_t^{-1/\alpha - 1/\beta} \cos \Theta_t \sin^{\alpha/\beta} \Theta_t \quad \text{as } n \to \infty. \tag{A 21}$$

From the same point process result, we obtain

$$\tilde{a}_n^{-2\alpha/\beta} \sum_{t=1}^{n} (Y_t - \bar{Y}_n)^2 \xrightarrow{d} S_{\beta/2,2} := \sum_{t=1}^{\infty} \Gamma_t^{-2/\beta} \sin^{2\alpha/\beta} \Theta_t \quad \text{as } n \to \infty. \tag{A 22}$$

The remainder of the proof is identical to that for theorem 4.1. ∎

## (d) Limit theory for correlation of heavy-tailed data from correlated normal random variables

In this section, we prove theorem 4.5, corollaries 4.6 and 4.4.

*Proof of theorem 4.5 and corollary 4.6.* Recall that $X := 1/|U|^c, Y := 1/|V|^c$ for some $c > 1$ where $(U, V)$ is bivariate normal with means 0, variances 1, and correlation $\rho \in (-1, 1)$. Let $\phi(x) := (2\pi)^{-1/2} \exp(-x^2/2), x \in (-\infty, +\infty)$, be the standard normal pdf. It follows directly from proposition 4.3 that

$$P(X > x) \sim x^{-1/c} 2\phi(0) = x^{-1/c} \sqrt{\frac{2}{\pi}}.$$

Now setting $a_n = (2/\pi)^{c/2} n^c$, it follows that $nP(X > a_n) \to 1$ as $n \to \infty$.

Establishing

$$\Pr(XY > x) \sim \frac{2}{c\pi(1-\rho^2)^{1/2}}x^{-1/c}\log x \quad \text{as } x \to \infty \tag{A 23}$$

is equivalent to proving

$$\Pr(|UV| < \epsilon) \sim \frac{2}{\pi(1-\rho^2)^{1/2}}\epsilon \log\left(\frac{1}{\epsilon}\right) \quad \text{as } \epsilon \downarrow 0. \tag{A 24}$$

The probability on the left side of (A 24) can be written

$$\Pr(|UV| < \epsilon) = \left(\frac{1}{2\pi}\right)^{1/2}\int_{-\infty}^{\infty} e^{-x^2/2}\left[\Phi\left(\frac{\epsilon/|x| - \rho x}{(1-\rho^2)^{1/2}}\right) - \Phi\left(\frac{-\epsilon/|x| - \rho x}{(1-\rho^2)^{1/2}}\right)\right]dx, \tag{A 25}$$

where $\Phi$ is the standard normal cdf. For an upper bound, notice that the integral in the right side of (A 25) does not exceed

$$2\epsilon + \int_{|x|>\epsilon} e^{-x^2/2}\left[\Phi\left(\frac{\epsilon/|x| - \rho x}{(1-\rho^2)^{1/2}}\right) - \Phi\left(\frac{-\epsilon/|x| - \rho x}{(1-\rho^2)^{1/2}}\right)\right]dx$$

$$\leq 2\epsilon + \left(\frac{1}{2\pi}\right)^{1/2}\int_{|x|>\epsilon} e^{-x^2/2}\frac{2\epsilon}{|x|(1-\rho^2)^{1/2}}\,dx$$

since

$$\Phi(b) - \Phi(a) \leq \left(\frac{1}{2\pi}\right)^{1/2}(b-a)$$

for any $a < b$. We conclude that

$$P(|UV| < \epsilon) \leq \left(\frac{2}{\pi}\right)^{1/2}\epsilon + \frac{2}{\pi(1-\rho^2)^{1/2}}\epsilon\int_{\epsilon}^{\infty}\frac{1}{x}e^{-x^2/2}\,dx.$$

Since by L'Hôpital's rule

$$\int_{\epsilon}^{\infty}\frac{1}{x}e^{-x^2/2}\,dx \sim \log\left(\frac{1}{\epsilon}\right) \quad \text{as } \epsilon \downarrow 0,$$

we conclude that

$$\limsup_{\epsilon \downarrow 0}\frac{P(|UV| < \epsilon)}{\epsilon\log(1/\epsilon)} \leq \frac{2}{\pi(1-\rho^2)^{1/2}}. \tag{A 26}$$

It remains to obtain a matching lower bound of the right side of (A 25). Now we use the bound, for any $a < b$,

$$\Phi(b) - \Phi(a) \geq \left(\frac{1}{2\pi}\right)^{1/2}(b-a)\exp\left\{\frac{-2\max(a^2, b^2)}{2}\right\}.$$

Choose $\theta = \theta(\epsilon) \downarrow 0$ as $\epsilon \downarrow 0$ such that

$$\log\left(\frac{1}{\theta}\right) = o\left(\log\left(\frac{1}{\epsilon}\right)\right) \quad \text{as } \epsilon \downarrow 0. \tag{A 27}$$

Let $M$ be a large number. We have by (A 25),

$$P(|UV| < \epsilon) \geq \frac{1}{2\pi}\int_{\epsilon M \leq |x| \leq \theta} e^{-x^2/2}\frac{2\epsilon}{|x|(1-\rho^2)^{1/2}}\exp\left\{-\frac{1}{2}\left(\frac{\epsilon}{|x|} + |\rho x|\right)^2\right\}dx$$

$$= \frac{2}{\pi(1-\rho^2)^{1/2}}\epsilon\int_{\epsilon M}^{\theta}\frac{1}{x}e^{-x^2/2}\exp\left\{-\frac{\epsilon^2}{2x^2} - \epsilon|\rho| - \frac{\rho^2 x^2}{2}\right\}dx$$

$$\geq \frac{2}{\pi(1-\rho^2)^{1/2}}\epsilon\exp\left\{-\frac{1}{2M^2} - \epsilon - \theta^2\right\}\int_{\epsilon M}^{\theta}\frac{1}{x}\,dx.$$

By (A 27) we conclude that

$$\limsup_{\epsilon \downarrow 0}\frac{P(|UV| < \epsilon)}{\epsilon\log(1/\epsilon)} \geq \frac{2}{\pi(1-\rho^2)^{1/2}}\exp\{-1/2M^2\}.$$

Letting $M \to \infty$ provides a lower bound matching (A 26) and, hence, proves (A 23).

The normalizing constant associated with the distribution of $XY$ is given by $\tilde{a}_n = K(n \log n)^c$, where $K = (2/\pi)^c(1-\rho^2)^{-c/2}$ so that $nP(XY > \tilde{a}_n) \to 1$ as $n \to \infty$. Note that $\tilde{a}_n/a_n \to \infty$. Also, for $M > 0$ fixed, large $n$, and all $x, y > 0$,

$$nP(XY > \tilde{a}_n x, X > a_n y) \leq nP(X > a_n M) + nP(XY > \tilde{a}_n x, X > a_n y, X \leq a_n M)$$

$$\leq nP(X > a_n M) + nP\left(Y > \frac{\tilde{a}_n}{a_n}\frac{x}{M}, X > a_n y\right). \tag{A 28}$$

As $n \to \infty$, the first term converges to $M^{-1/c}$, while the second term is

$$nP\left(|U| \leq (y^{-1/c}a_n^{-1/c}), |V| \leq \text{const}\left(\frac{\tilde{a}_n}{a_n}\right)^{-1/c}\right) \sim Cn(a_n^{-1/c})\left(\frac{\tilde{a}_n}{a_n}\right)^{-1/c}\varphi_\rho(0,0),$$

where $C$ is a generic constant and $\varphi_\rho$ is the joint density of $(U, V)$. Since $na_n^{-1/c} \to (2/\pi)^{-1/2}$ and $\tilde{a}_n/a_n \to \infty$ as $n \to \infty$, it follows that this term goes to 0 as $n \to \infty$. Consequently, the $\limsup_{n\to\infty}$ followed by the limit as $M \to \infty$ of the left side of (A 28) is 0. In other words, $X$ and $XY$ are asymptotically independent. From this property [12], we have the point process convergence

$$N_n := \sum_{t=1}^{n} \varepsilon_{(a_n^{-1}X_t, a_n^{-1}Y_t, \tilde{a}_n^{-1}X_tY_t)} \xrightarrow{d} N := \sum_{i=1}^{\infty}\left[\varepsilon_{(\Gamma_{i,1}^{-c},0,0)} + \varepsilon_{(0,\Gamma_{i,2}^{-c},0)} + \varepsilon_{(0,0,\Gamma_{i,3}^{-c})}\right] \quad \text{as } n \to \infty, \tag{A 29}$$

where $\{\Gamma_{i,1}\}, \{\Gamma_{i,2}\}$, and $\{\Gamma_{i,3}\}$ are iid copies of the points $\{\Gamma_i\}$ defined in theorem A.1.

From this point process, by the same argument as used in the proof of theorem 4.1 (see appendix A(c)), we obtain

$$\left(a_n^{-2}\sum_{t=1}^{n} X_t^2, a_n^{-2}\sum_{t=1}^{n} Y_t^2, \tilde{a}_n^{-1}\sum_{t=1}^{n} X_tY_t\right) \xrightarrow{d} (S_{1/(2c),1}, S_{1/(2c),2}, S_{1/c}) \quad \text{as } n \to \infty, \tag{A 30}$$

where the limit rvs are independent and given by

$$(S_{1/(2c),1}, S_{1/(2c),2}, S_{1/c}) := \left(\sum_{i=1}^{\infty}\Gamma_{i,1}^{-1/(2c)}, \sum_{i=1}^{\infty}\Gamma_{i,2}^{-1/(2c)}, \sum_{i=1}^{\infty}\Gamma_{i,3}^{-1/c}\right). \tag{A 31}$$

Since $a_n^2/\tilde{a}_n = (1-\rho^2)^{c/2}(n/\log n)^c$, we have by the continuous mapping theorem

$$\left(\frac{n}{\log n}\right)^c \tilde{\rho}_n \xrightarrow{d} (1-\rho^2)^{-c/2}\frac{S_{1/c}}{\sqrt{S_{1/(2c),1}S_{1/(2c),2}}} \quad \text{as } n \to \infty, \tag{A 32}$$

as desired.

To prove corollary 4.6 and (4.9), we note that by theorem 4.3, $X$ and $Y$ are asymptotically independent. By the earlier argument for (2.10), it follows that, as $n \to \infty$,

$$\left(a_n^{-1}\sum_{t=1}^{n} X_t, a_n^{-2}\sum_{t=1}^{n}(X_t - \bar{X}_n)^2, a_n^{-1}\sum_{t=1}^{n} Y_t, a_n^{-2}\sum_{t=1}^{n}(Y_t - \bar{Y}_n)^2\right) \xrightarrow{d} (S_{1/c,1}, S_{1/(2c),1}, S_{1/c,2}, S_{1/(2c),2}),$$

where $S_{1/c,i}, S_{1/(2c),i}$ are given by (A 10) and (A 11) with $\alpha = 1/c$ and are independent for $i = 1, 2$. Then by the continuous mapping theorem

$$\frac{-\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{\sqrt{\sum_{t=1}^{n}(X_t - \bar{X}_n)^2}\sqrt{\sum_{t=1}^{n}(Y_t - \bar{Y}_n)^2}} \xrightarrow{d} -R_1R_2, \tag{A 33}$$

where $R_i = S_{1/c,i}/\sqrt{S_{1/(2c),i}}$, $i = 1, 2$. Also, since $na_n^{-2}\tilde{a}_n \to 0$ as $n \to \infty$, it follows that $na_n^{-2}\sum_{i=1}^{n} X_iY_i \xrightarrow{P} 0$ as $n \to \infty$. ∎

*Proof of corollary* 4.4. Let $0 < \alpha < \gamma < 1$, and let $S \in RV(\gamma)$. Then for some $c > 0, P(S > x) \sim cx^{-\gamma}$ as $x \to \infty$ so the assumption $\bar{F}_X \in RV(\alpha)$ implies that $P(X > x)/P(S > x) \to \infty$ as $x \to \infty$. Therefore, a constant $C$ exists such that $P(X > x) > P(S > x)$ for all real $x \geq C$. Therefore, $P(X + C > x) > P(S > x)$ for all $x \geq C > 0$, while for all $x < C$, by definition of $C$ we have $P(X + C > x) = 1 \geq P(S > x)$.

24

royalsocietypublishing.org/journal/rspa   Proc. R. Soc. A **476**: 20200610

Consequently, a constant $C$ exists such that $P(X + C > x) \geq P(S > x)$ for all $x \in \mathbb{R}$. That is, by definition, $X + C$ is stochastically larger than $S$, or $X + C \geq_{st} S$. When $0 < \gamma < \alpha < 1$, the argument works with $S$ and $X$ exchanged, hence there exists a constant $C > 0$ such that $S + C \geq_{st} X$.

For the next results, we require a well-known inequality (A 34): if $\gamma \in (0, 1)$, then for some $c > 0$, a $\gamma$-stable rv $S$ satisfies

$$P(S \leq \epsilon) \leq \exp(-c\epsilon^{\gamma/(\gamma-1)}) \quad \text{as } \epsilon \downarrow 0. \tag{A 34}$$

We prove (A 34) here for ease of reference. For any $\theta > 0$, if $S$ is $\gamma$-stable, then its Laplace transform is $E\exp(-\theta S) = \exp(-\theta^{\gamma})$ [7]. For any $\epsilon > 0$, $\theta > 0$, using Markov's inequality for the following inequality, we have

$$P(S \leq \epsilon) = P(e^{-\theta S} \geq e^{-\theta\epsilon}) \leq e^{+\theta\epsilon} E\exp(-\theta S) = \exp(-\theta^{\gamma} + \theta\epsilon). \tag{A 35}$$

Now pick any $a \in (0, 1)$ and set $\theta = a\epsilon^{1/(\gamma-1)}$ to get $-\theta^{\gamma} = -a^{\gamma}\epsilon^{\gamma/(\gamma-1)}$ and $\theta\epsilon = a\epsilon^{1/(\gamma-1)+1} = a\epsilon^{\gamma/(\gamma-1)}$. Therefore $-\theta^{\gamma} + \theta\epsilon = -a^{\gamma}\epsilon^{\gamma/(\gamma-1)} + a\epsilon^{\gamma/(\gamma-1)} = a\epsilon^{\gamma/(\gamma-1)}(1 - a^{\gamma-1})$. Since $\gamma \in (0, 1)$ and $a \in (0, 1)$, we have $a^{1-\gamma} < 1$ or $1 - a^{\gamma-1} < 0$ or $c := -a(1 - a^{\gamma-1}) > 0$ in (A 34), which is now proved.

Let $X_i, i = 1, 2, \ldots$ be iid copies of $X \in RV(\alpha)$. We consider two cases: first that $0 < \tau < 1/\alpha$, then that $1/\alpha < \tau < \infty$.

Suppose first that $0 < \tau < 1/\alpha$. We claim that

$$n^{-\tau} \sum_{i=1}^{n} X_i \to \infty \quad \text{a.s. as } n \to \infty. \tag{A 36}$$

We may and will assume that $\tau \geq 1$. Choose $\alpha < \gamma < 1/\tau$, and choose $C > 0$ such that $X + C \geq_{st} S$, where $S$ is $\gamma$-stable. Let $S_1, S_2, \ldots$ be iid copies of $S$. Then for any $t > 0$

$$
\begin{aligned}
P\left(n^{-\tau} \sum_{i=1}^{n} X_i \leq t\right) &\leq P\left(\sum_{i=1}^{n} S_i - Cn \leq tn^{\tau}\right) \\
&\leq P\left(\sum_{i=1}^{n} S_i \leq (t + C)n^{\tau}\right) \\
&= P\left(n^{1/\gamma} S_1 \leq (t + C)n^{\tau}\right) \\
&= P\left(S_1 \leq (t + C)n^{-(1/\gamma - \tau)}\right) \\
&\leq \exp\left\{-cn^{(1-\gamma\tau)/(1-\gamma)}\right\}
\end{aligned}
$$

for some $c = c(t) > 0$. Therefore, for any $t > 0$,

$$\sum_{n=1}^{\infty} P\left(n^{-\tau} \sum_{i=1}^{n} X_i \leq t\right) < \infty,$$

so by the first Borel–Cantelli lemma [21, p. 201], (A 36) holds.

Next, suppose that $\tau > 1/\alpha$. We claim that

$$n^{-\tau} \sum_{i=1}^{n} X_i \to 0 \quad \text{a.s. as } n \to \infty. \tag{A 37}$$

Choose $1/\tau < \gamma < \alpha$, and choose $C > 0$ such that $S + C \geq_{st} X$, where $S$ is a $\gamma$-stable rv. Let $S_1, S_2, \ldots$ be iid copies of $S$. Then for any $t > 0$, for all $n$ large enough,

$$P\left(n^{-\tau} \sum_{i=1}^n X_i > t\right) \leq P\left(\sum_{i=1}^n S_i + Cn > tn^\tau\right)$$

$$\leq P\left(\sum_{i=1}^n S_i > \frac{tn^\tau}{2}\right)$$

$$= P\left(n^{1/\gamma} S_1 > \frac{tn^\tau}{2}\right)$$

$$= P\left(S_1 > \frac{tn^{\tau-1/\gamma}}{2}\right)$$

$$\leq cn^{-(\gamma\tau-1)}$$

for some $c > 0$. Therefore, for any $t > 0$,

$$\sum_{m=1}^\infty P\left((2^m)^{-\tau} \sum_{i=1}^{2^m} X_i > t\right) < \infty,$$

so by the first Borel–Cantelli lemma [21, p. 201],

$$(2^m)^{-\tau} \sum_{i=1}^{2^m} X_i \to 0 \quad \text{a.s. as } m \to \infty.$$

Therefore, with

$$m_n = \lfloor \log_2 n \rfloor \to \infty \quad \text{as } n \to \infty$$

we have

$$n^{-\tau} \sum_{i=1}^n X_i \leq (2^{m_n})^{-\tau} \sum_{i=1}^{2^{m_n+1}} X_i$$

$$= 2^\tau (2^{m_n+1})^{-\tau} \sum_{i=1}^{2^{m_n+1}} X_i \to 0 \quad \text{a.s. as } n \to \infty,$$

proving (A 37).

To prove that $\rho_n \to 0$ a.s. as $n \to \infty$, we pick any $\delta \in (c, 2c)$ and multiply numerator $N_n$ and denominator $D_n$ of $\rho_n$ by $n^{-\delta}$ to get

$$\rho_n = \frac{n^{-\delta} N_n}{n^{-\delta} D_n}.$$

It follows from (A 37) that $n^{-\delta} N_n \to 0$ a.s. as $n \to \infty$, and from (A 36) that $n^{-\delta} D_n \to \infty$ a.s. as $n \to \infty$. Therefore, $\rho_n \to 0$ a.s. as $n \to \infty$. ∎

Shapiro [22] proved a special case of our results, namely, powers of reciprocals of rvs with a continuous density that is positive at 0 have Pareto-like tails.

# References

1. Resnick SI. 2007 *Heavy-tail phenomena: probabilistic and statistical modeling*. Berlin, Germany: Springer Science & Business Media.
2. Resnick SI. 2008 *Extreme values, regular variation and point processes*. Berlin, Germany: Springer.
3. Zolotarev VM. 1986 *One-dimensional stable distributions*. Translations of Mathematical Monographs, vol. 65. Providence, RI: American Mathematical Society.
4. Eisler Z, Bartos I, Kertész J. 2008 Fluctuation scaling in complex systems: Taylor's law and beyond. *Adv. Phys.* **57**, 89–142. (doi:10.1080/00018730801893043)

26

royalsocietypublishing.org/journal/rspa    *Proc. R. Soc. A* **476**: 20200610

5. Taylor RA. 2019 *Taylor's power law: order and pattern in nature*. New York, NY: Academic Press.
6. Bliss CI. 1941 Statistical problems in estimating populations of Japanese beetle larvae. *J. Econ. Entomol.* **34**, 221–232. (doi:10.1093/jee/34.2.221)
7. Brown M, Cohen JE, de la Peña VH. 2017 Taylor's law, via ratios, for some distributions with infinite mean. *J. Appl. Probab.* **54**, 657–669. (doi:10.1017/jpr.2017.25)
8. Drton M, Xiao H. 2016 Wald tests of singular hypotheses. *Bernoulli* **22**, 38–59. (doi:10.3150/14-BEJ620)
9. Pillai NS, Meng X-L. 2016 An unexpected encounter with Cauchy and Lévy. *Ann. Stat.* **44**, 2089–2097. (doi:10.1214/15-AOS1407)
10. Davis RA, Hsing T. 1995 Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Ann. Probab.* **23**, 879–917. (doi:10.1214/aop/1176988294)
11. Davis RA, Resnick S. 1985 More limit theory for the sample correlation function of moving averages. *Stoch. Process. Appl.* **20**, 257–279. (doi:10.1016/0304-4149(85)90214-5)
12. Davis RA, Resnick S. 1986 Limit theory for the sample correlation function of moving averages. *Ann. Stat.* **14**, 533–558. (doi:10.1214/aos/1176349937)
13. Samorodnitsky G. 2016 *Stochastic processes and long range dependence*. Springer Series in Operations Research and Financial Engineering, vol. 26. Berlin, Germany: Springer.
14. Samorodnitsky G, Taqqu M. 1994 *Stable non-Gaussian stable processes: stochastic models with infinite variance*. London, UK: Chapman and Hall.
15. Feller W. 1971 *An introduction to probability theory and its applications*, vol. 2. New York, NY: John Wiley and Sons.
16. Davis RA. 1983 Stable limits for partial sums of dependent random variables. *Ann. Probab.* **11**, 262–269. (doi:10.1214/aop/1176993595)
17. Davis RA, Mikosch T. 1998 Limit theory for the sample ACF of stationary process with heavy tails with applications to ARCH. *Ann. Stat.* **26**, 2049–2080. (doi:10.1214/aos/1024691368)
18. Basrak B, Davis RA, Mikosch T. 2002 Regular variation of GARCH processes. *Stoch. Process. Appl.* **99**, 95–115. (doi:10.1016/S0304-4149(01)00156-9)
19. Resnick SI. 1986 Point processes, regular variation and weak convergence. *Adv. Appl. Probab.* **18**, 66–138. (doi:10.1017/S0001867800015597)
20. LePage R, Woodroofe M, Zinn J. 1981 Convergence to a stable distribution via order statistics. *Ann. Probab.* **9**, 624–632. (doi:10.1214/aop/1176994367)
21. Feller W. 1968 *An introduction to probability theory and its applications*, vol. 1. New York, NY: John Wiley and Sons.
22. Shapiro JM. 1977 On domains of normal attraction to stable distributions. *Houston J. Math.* **3**, 539–542.