PNAS RESEARCH ARTICLE | APPLIED MATHEMATICS
POPULATION BIOLOGY

Generalized Taylor's law for infinite-mean heavy-tailed data under dependence and heterogeneity

Pok Him Cheng[a,1], Joel E. Cohen[a,b,c,d,1,2], Hok Kan Ling[e,1], and Sheung Chi Phillip Yam[f,1]

Affiliations are included on p. 11.

Contributed by Joel E. Cohen; received April 4, 2026; accepted August 31, 2026; reviewed by Vali Asimit and Qihe Tang

Taylor's law, also known as fluctuation scaling in physics and the power-law variance function in statistics, is an empirical pattern widely observed across fields including ecology, physics, finance, and epidemiology. It states that the variance of a sample scales as some power of the mean of the sample. We study generalizations of Taylor's law in the context of heavy-tailed distributions with infinite mean and variance. We establish the probabilistic limit and analyze the associated convergence rates. Our results extend the existing literature by relaxing the assumption that the observed random variables are independently and identically distributed to accommodate dependence and heterogeneity among them. This generalization enables application of Taylor's law to dependent time series and network-structured data. We support the theoretical developments by extensive simulations, and the practical relevance through applications to real network data.

heavy-tailed distribution | dependence | infinite mean | network data

Significance

Many phenomena exhibit rare but extremely large fluctuations. Examples include financial losses, epidemic outbreaks, and measurements on networked systems. In such settings, classical assumptions of finite mean, finite variance, and independence often fail. We show that, even when the underlying distributions have infinite mean and variance and the data are dependent or heterogeneous, under some conditions the sample variance may be a power of the sample mean, a relationship widely known as Taylor's law. These results extend fluctuation scaling theory to some time series and network data and provide a rigorous foundation for interpreting empirical scaling patterns in complex systems.

Introduction

The study of heavy-tailed distributions has gained significant attention in various fields of research, including finance, economics, operational risk management, network analysis, disaster analysis, and extreme event modeling. For example, ref. 1 provided a comprehensive introduction to heavy-tailed distributions in finance; ref. 2 explored methods for computing portfolio Value-at-Risk when underlying risk factors have heavy-tailed distributions; ref. 3 modeled heavy-tailed loss processes in operational risk; ref. 4 discussed the consequences of the heavy-tailed behavior of flood peak distributions; ref. 5 studied estimation methods for the tail degree in large-scale social networks. ref. 6 considered operational risk management under such heavy-tailed models, and ref. 7 used heavy-tailed distributions to model disasters. These distributions exhibit rare, extreme events that occur more frequently than in normal distributions and have practical applications.

Taylor's law (8), also known as fluctuation scaling in physics, and as a power-law variance function in statistics, is an empirical pattern observed in various fields such as ecology, physics, finance, and epidemiology. Taylor's law states that the variance of a sample is proportional to a power of the sample mean. For a random variable X with mean $\mu_X > 0$ and variance $\mathrm{Var}_X > 0$, Taylor's law postulates that, as X ranges over a set of two or more random variables considered to be somehow comparable,

$$\mathrm{Var}_X = a\mu_X^b, \qquad [1]$$

where $a > 0$ and b are real constants, the same for all random variables in the set. Most applications have $b > 0$. In its logarithmic form, Taylor's law becomes

$$\log(\mathrm{Var}_X) = \log(a) + b\log(\mu_X), \qquad [2]$$

where the slope b in Eq. **2** is the exponent b in the power-law variance–mean relationship Eq. **1**. Equivalently,

$$\frac{\log(\mathrm{Var}_X)}{\log(\mu_X)} = \frac{\log(a)}{\log(\mu_X)} + b.$$

Replacing the population variance Var_X by the sample variance and replacing the population mean μ_X by the sample mean yields a sample version of Taylor's law even when the population quantities Var_X, μ_X may be infinite.

Ref. 9 reviewed many applications of Taylor's law. Ref. 10 discussed the general applicability of the scaling relationship of Taylor's law, reviewed literature, and presented

Author contributions: P.H.C., J.E.C., H.K.L., and S.C.P.Y. designed research; performed research; analyzed data; and wrote the paper.

Reviewers: V.A., Bayes Business School; and Q.T., University of New South Wales.

Competing interest statement: Q.T. co-authored a paper with H.K.L. and S.C.P.Y.: Cheung, K.C., et al., On additivity of tail comonotonic risks. Scandinavian Actuarial Journal, 2019(10), pp. 837-866 (The authors of the present paper do not believe this publication 7 years ago poses any conflict of interest with his reviewing the present paper objectively).

Copyright © 2026 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]P.H.C., J.E.C., H.K.L., and S.C.P.Y. contributed equally to this work.

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

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2611888123/-/DCSupplemental.

Published October 1, 2026.

Downloaded from https://www.pnas.org by Joel Cohen on October 4, 2026 from IP address 174.231.48.236.

new empirical data and model calculations; (11) showed that monthly crime reports in the United Kingdom followed Taylor's law relationships approximately; (12) extended Taylor's law to higher moments and gave conditions on the ratio of sample size per random variable and the number of random variables under which estimates of the slope reflect underlying distributions or are artifactual; (13) showed that human demographic data from Norway follow Taylor's law and proposed a simple model of some features of the results. These references focused on light-tailed data where the population mean and variance exist. Ref. 14 introduced a dynamic Taylor's law for dependent samples using self-normalized expressions involving Bernstein blocks. They discussed variations of Taylor's law under dependence, considering ergodic behaviors and stationary time series of light-tailed random variables under weak dependence and strong mixing assumptions of a strictly stationary time series.

Taylor's law is also related to the Tweedie–Bar-Lev–Enis family of exponential dispersion models (15–18), for which the variance functions have the power form $\mathrm{Var}_X = \phi\mu_X^p$, where $\phi > 0$ is a dispersion parameter. This connection has been studied in the context of Taylor's ecological power law and fluctuation scaling (19–21). These works provide one particular distributional framework for Taylor's law when the population mean and variance are finite. The present article studies a complementary heavy-tailed setting in which the population mean and variance are infinite; in this setting, Taylor's law is formulated in terms of sample moments and established through convergence in probability.

We now review Taylor's law for heavy-tailed distributions with infinite mean and therefore also infinite variance. To the best of our knowledge, ref. 22 were the first to extend Taylor's law to distributions with infinite mean and variance. They showed that a modified version of Taylor's law remains valid for α-stable distributions with tail index $\alpha \in (0, 1)$. Ref. 23 showed that heavy-tailed distributions can overwhelm the dependence among random variables; that under certain conditions, the sample correlation of regularly varying random variables converges to a random limit; and that the sample kurtosis converges in distribution. They discussed Taylor's law for heterogeneous and correlated heavy-tailed data, instead of focusing solely on independently and identically distributed (henceforth i.i.d.) data. Ref. 24 extended the work of ref. 22 by analyzing the asymptotic behavior of the sample (local) upper and lower semivariances, higher sample moments, skewness, and kurtosis. They applied their results to various financial ratios, including the Sharpe ratio, the Sortino ratio, the Omega index, the upside potential ratio, and the Farinelli–Tibiletti ratio, when returns have a nonnegative heavy-tailed distribution.

Ref. 25 considered Taylor's law when the tail index satisfies $\alpha \in (1, 2)$, so that the random variable has finite mean and infinite variance. They argued that this setting for heavy-tailed random variables is more relevant to some real data. They were also the first to consider an array of heavy-tailed random variables and investigated the patterns of COVID-19 cases and deaths in the United States using Taylor's law.

We concentrate on establishing limits in probability $\xrightarrow{\mathbb{P}}$, rather than weak distributional limits, for several generalizations of Taylor's law. Inspired by ref. 26, we use Karamata's theorem (27, page 26, Proposition 1.5.8) to establish the moments of truncated heavy-tailed random variables and to approximate heavy-tailed random variables with infinite moments. This approach accommodates dependence through Condition A(p), a covariance-type weak-dependence condition on shifted truncated powers of the observations.

The rest of this article is organized as follows. Section 1 presents Taylor's law for i.i.d. infinite-mean heavy-tailed data and introduces the truncation scheme used throughout the paper. Section 2 extends the results to dependent heavy-tailed data through Condition A(p), verifies this condition under standard weak-dependence assumptions, and treats central moments and correlated heavy-tailed data. Section 3 considers finite-cluster heterogeneous data and network-indexed data. We explain why the same logarithmic exponent appears in all three settings: The leading logarithmic exponent is governed by the dominant tail index; under our assumptions, dependence and finite-cluster heterogeneity affect terms that are negligible on the logarithmic scale.

Since our primary condition for establishing these laws is based on the covariances of truncated random variables, our results apply to a collection of random variables such as those on a network graph. Section 4 illustrates Taylor's law for heavy-tailed network data using three real-world datasets on network graphs. Our results might be applicable in many real-life situations such as social network data.

Section 5 summarizes the article and discusses possible future work. While our main text considers higher moments and higher central moments, *SI Appendix* gives extensions of these results to semivariances, rate refinements under additional conditions on tail regularity, proofs of the mathematical claims made in the main text, supplementary empirical analyses, and simulations illustrating the convergence discussed in the main text.

1. Taylor's Law for i.i.d. Infinite-Mean Heavy-Tailed Data

1.1. Setting and Notation. Unless otherwise specified, in Sections 1 and 2 we assume that the random variables $X_1, \ldots, X_n$ have a common cumulative distribution function F_X with finite lower endpoint $x_F := \inf\{x : F_X(x) > 0\} > -\infty$. Let

$$Y_i := X_i - x_F.$$

Standing marginal assumption: We assume that the survival function of Y_i satisfies

$$\overline{F}(x) := \mathbb{P}(Y_i > x) = x^{-\alpha} l(x), \qquad x > 0, \qquad \textbf{[3]}$$

where $\alpha > 0$ and l is slowly varying at infinity. Examples of slowly varying functions include $\log(x)$, $(\log(x))^p$ for $p \in \mathbb{R}$, $\log(\log(x))$, and the constant function. Refs. 27–29 give more details about slowly varying functions. We refer to Eq. **3**, together with the common distribution and finite lower endpoint assumption above, as the standing marginal assumption.

It is well known that $\overline{F}(x) = x^{-\alpha} l(x)$ for some slowly varying function l if and only if F belongs to the maximum domain of attraction of the Fréchet distribution, which has cumulative distribution function $\Phi_\alpha(x) = \exp\{-x^{-\alpha}\}$ for $x > 0$ and $\alpha > 0$ (30, Theorem 3.3.7). This class of distributions contains, for example, the Pareto and the α-stable distributions.

Let $\{a_n\}$ and $\{b_n\}$ be deterministic sequences with $b_n \neq 0$ for all sufficiently large n. We write $a_n \sim b_n$ if $a_n/b_n \to 1$ as $n \to \infty$. For two positive sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \asymp b_n$ if there exist constants $0 < c < C < \infty$ such that $cb_n \le a_n \le Cb_n$ for all sufficiently large n. Let $\{r_n\}$ be a positive deterministic sequence. We write $a_n = O(r_n)$ if there exist constants $C < \infty$ and N such that $|a_n| \le Cr_n$ for all $n \ge N$, and $a_n = o(r_n)$

Downloaded from https://www.pnas.org by Joel Cohen on October 4, 2026 from IP address 174.231.48.236.

if $a_n/r_n \to 0$. For random variables $\{Z_n\}$, we write $Z_n = O_p(r_n)$ if, for every $\varepsilon > 0$, there exist constants $M < \infty$ and N such that $\sup_{n \geq N} \mathbb{P}(|Z_n| > Mr_n) \leq \varepsilon$, and we write $Z_n = o_p(r_n)$ if, for every $\varepsilon > 0$, $\mathbb{P}(|Z_n| > \varepsilon r_n) \to 0$. For $\{Z_n\}$, we say its rate of convergence to 0 is r_n if $r_n Z_n = O_p(1)$ and $(r_n Z_n)^{-1} = O_p(1)$. We let $\sum_{i \neq j}$ denote $\sum_{i,j=1,\ldots,n, i \neq j}$.

For $p > 0$, denote the shifted pth sample moment by $M_{n,p}^Y := \frac{1}{n}\sum_{i=1}^n Y_i^p$. When no confusion can arise, we write $M_{n,p}$ for $M_{n,p}^Y$. For $p \in \mathbb{N}$, define the raw integer sample moment $M_{n,p}^X := \frac{1}{n}\sum_{i=1}^n X_i^p$. For $h, k > \alpha$, where α is the tail index of X and Y, define

$$\iota_\alpha(h,k) := \frac{h-\alpha}{k-\alpha},$$

which is the limit of the ratio of the logarithm of the hth to the kth moments in various Taylor's laws considered here, e.g., *Theorem* 1.3.

The notation introduced in this subsection will also be used in later sections, where independence is relaxed.

1.2. Truncation Scheme and Truncated Moments. When $\alpha \in (0,1)$, the mean and variance of F are infinite (31, 32). More generally, for any $p > \alpha$, the population moment $\mathbb{E}(Y^p)$ is infinite. Thus, although $M_{n,p}$ is finite for any fixed sample size, it does not estimate a finite population moment. The usual finite-moment laws of large numbers and central limit theorems do not describe the relevant asymptotic behavior.

We first explain the idea used to address this difficulty. Consider the truncated pth moment $d_{n,p} := \mathbb{E}(Y^p \mathbb{1}(Y \leq t_n))$, which is finite for each n because $t_n < \infty$. Here t_n is a deterministic sequence diverging to infinity. The goal is to choose $\{t_n\}$ so that $d_{n,p}$ provides the appropriate deterministic scale for $M_{n,p}$. Let $c_n \to \infty$ with $\log c_n = o(\log n)$. When the truncation level t_n is chosen so that

$$n\mathbb{P}(Y > t_n) \sim \frac{1}{c_n} \to 0,$$

we can show that $M_{n,p}/d_{n,p} = O_p(1)$. To establish a lower bound for $M_{n,p}/d_{n,p}$, we introduce another truncation level v_n, chosen so that

$$n\mathbb{P}(Y > v_n) \sim c_n \to \infty.$$

Together, these two truncation levels allow us to show that $\log(M_{n,p}/d_{n,p}) = O_p(\log c_n)$. This bound, together with the growth order of $d_{n,p}$, is the main ingredient in proving the probability limits for the various generalizations of Taylor's law considered below.

The following lemma gives the growth rate of truncated moments for an arbitrary sequence $a_n \to \infty$.

Lemma 1.1 (Truncated moment asymptotics). *Suppose that $Y \geq 0$ has a survival function $\overline{F}(x) = x^{-\alpha} l(x)$ for $\alpha > 0$. Let $\tilde{Y}_n := Y\mathbb{1}(Y \leq a_n)$, where $a_n \to \infty$ as $n \to \infty$. Then, for $p > \alpha$, as $n \to \infty$, we have*

(a)

$$\mathbb{E}[\tilde{Y}_n^p] \sim \frac{\alpha}{p-\alpha} d_n^{p-\alpha} l(a_n);$$

(b)

$$\frac{\log \mathbb{E}[\tilde{Y}_n^p]}{\log a_n} \to p - \alpha.$$

The next lemma shows that the two truncation levels t_n, v_n described above can be chosen for any slowly varying function l.

Lemma 1.2 (Existence of truncation levels). *Fix $\alpha > 0$. Let $\{c_n\}$ be a positive sequence such that $c_n \to \infty$ and $\log c_n = o(\log n)$. For any positive slowly varying function l, there exist sequences $t_n \to \infty$ and $v_n \to \infty$ such that*

$$\frac{n l(t_n)}{t_n^\alpha} \sim \frac{1}{c_n} \quad [4]$$

and

$$\frac{n l(v_n)}{v_n^\alpha} \sim c_n. \quad [5]$$

Consequently, any such choice of t_n and v_n satisfies

$$\frac{\log t_n}{\log n} \to \frac{1}{\alpha} \quad \text{and} \quad \frac{\log v_n}{\log n} \to \frac{1}{\alpha}. \quad [6]$$

Hereafter we take $\{c_n\}$ to be a positive sequence such that $c_n \to \infty$ and $\log c_n = o(\log n)$. Such a positive sequence c_n exists: For example, $c_n := 1 + \log(n)$ for every positive integer n. We take $\{t_n\}$ and $\{v_n\}$ to satisfy Eqs. **4** and **5**, respectively.

For notational simplicity, define

$$\Delta_n := \frac{\log c_n + |\log l(t_n)|}{\log n}. \quad [7]$$

By Eq. **6**, the slow variation of l, and the assumption $\log c_n = o(\log n)$, we have $\Delta_n \to 0$. Throughout the sequel, deterministic error terms of order $O(\Delta_n)$ are absorbed into $O_p(\Delta_n)$.

1.3. The i.i.d. Taylor's Law. Taylor's law holds empirically when the sample variance scales approximately in direct proportion to a power of the sample mean. When the scaling of higher-order moments or higher centered moments is of interest, we may replace the variance or the mean or both in Eq. **2** with higher (centered) moments (or semimoments). Ref. 12 applied the generalized Taylor's law for higher sample moments to tree abundance across plots in the Black Rock Forest and abundances of carabid beetles in different sites in the Netherlands. Refs. 12 and 24 gave more examples and discussion.

The variance and higher-moment Taylor's laws in Refs. 22 and 24 that are most directly comparable to ours are established for i.i.d. nonnegative observations with stable or approximately stable distributions. In these cases, the survival function satisfies $\overline{F}_X(x) \sim C_\alpha x^{-\alpha}$ for some constant $C_\alpha > 0$ (33, p. 448). Our results allow a more general regularly varying tail: After shifting by the finite lower endpoint x_F, the variables $Y_i = X_i - x_F$ satisfy $\overline{F}_Y(x) = x^{-\alpha} l(x)$, where l is a general slowly varying function and need not converge to a positive constant.

Theorem 1.3 (i.i.d. Taylor's law). *Let $X_1, X_2, \ldots$ be i.i.d. random variables satisfying the standing marginal assumption with $\alpha > 0$ [not necessarily in $(0,1)$]. Let $h_1, h_2 > \alpha$. Then*

$$\frac{\log M_{n,h_1}}{\log M_{n,h_2}} \xrightarrow{\mathbb{P}} \iota_\alpha(h_1, h_2).$$

If $\alpha \in (0,1)$, then the sample variance $V_n := \frac{1}{n}\sum_{i=1}^n (X_i - M_{n,1}^X)^2$ obeys

$$\frac{\log V_n}{\log |M_{n,1}^X|} \xrightarrow{\mathbb{P}} \iota_\alpha(2,1).$$

Downloaded from https://www.pnas.org by Joel Cohen on October 4, 2026 from IP address 174.231.48.236.

2. Taylor's Law for Dependent Heavy-Tailed Data

2.1. Condition A(p): Weak Dependence of Truncated Powers. To establish the probability limits of Taylor's law for dependent data, we impose Condition A(p), a covariance-type condition on the shifted truncated random variables $\check{Y}_i := Y_i \mathbb{1}(Y_i \le v_n)$. We view Condition A($p$) as a weak-dependence assumption, since it is satisfied under the sufficient conditions developed in Section 2.3 for several standard weak-dependence settings, including m-dependent sequences, mixing sequences, and autoregressive processes.

Condition A(p): Here and below where noted, assume that the shifted truncated random variables $\check{Y}_i$ satisfy

$$\sum_{i \neq j} \mathrm{Cov}(\check{Y}_i^p, \check{Y}_j^p) = o(v_n^{2p} c_n^2). \qquad [8]$$

2.2. Higher-Moment Taylor's Law Under Condition A(p).

Lemma 2.1 (Log-moment expansion under Condition A(p)). *Assume the standing marginal assumption with $\alpha > 0$ [not necessarily in $(0,1)$). If Condition A[p] holds for $p > \alpha$, then*

$$\frac{\log M_{n,p}}{\log n} \xrightarrow{\mathbb{P}} \frac{p-\alpha}{\alpha}. \qquad [9]$$

More precisely,

$$\frac{\log M_{n,p}}{\log n} = \frac{p-\alpha}{\alpha} + O_p(\Delta_n).$$

Theorem 2.2 (Higher-moment Taylor's law under Condition A(p)). *Assume the standing marginal assumption with $\alpha > 0$ [not necessarily in $(0,1)$]. Let $h_1, h_2 > \alpha$ and $h_1 \neq h_2$. Suppose that Condition A(p) holds for $p = h_1$ and $p = h_2$. Then*

(a)

$$\frac{\log M_{n,h_1}}{\log M_{n,h_2}} \xrightarrow{\mathbb{P}} \iota_\alpha(h_1, h_2).$$

(b) More precisely,

$$R^{hm}_{n,h_1,h_2} := \frac{\log M_{n,h_1}}{\log M_{n,h_2}} - \iota_\alpha(h_1,h_2) = O_p(\Delta_n). \qquad [10]$$

Remark 2.3. Part (a) is the main Taylor's law for higher moments of the shifted random variables $Y_i = X_i - x_F$. In particular, for $\alpha \in (0,1)$ and $k \in \mathbb{N}$,

$$\frac{\log M_{n,k}}{\log M_{n,1}} \xrightarrow{\mathbb{P}} \frac{k-\alpha}{1-\alpha}.$$

Part (b) gives a general upper bound on the rate of convergence in part (a), and may be viewed as a quantitative refinement of the convergence in probability in (a).

Additional rate refinements for this and related Taylor's law limits, under further regularity assumptions on the slowly varying factor l, are given in *SI Appendix*, section D.

Theorem 2.2 is formulated for the shifted moments of $Y_i = X_i - x_F$. For integer powers, the corresponding logarithmic conclusions also hold for the raw moments of the original variables X_i, because the finite shift contributes only lower-order terms on the logarithmic scale. The absolute values below account for possible finite-sample sign changes of raw odd moments when $x_F < 0$.

Corollary 2.4 (Raw integer moments). *Under the assumptions of Theorem 2.2, let h_1, h_2 be positive integers such that $h_1, h_2 > \alpha$. Then*

$$\frac{\log |M^X_{n,h_1}|}{\log |M^X_{n,h_2}|} \xrightarrow{\mathbb{P}} \iota_\alpha(h_1,h_2).$$

More precisely,

$$\frac{\log |M^X_{n,h_1}|}{\log |M^X_{n,h_2}|} - \iota_\alpha(h_1,h_2) = O_p(\Delta_n).$$

The same finite-shift argument also applies to the corresponding higher-moment results in the conditionally weak-dependent and heterogeneous settings, with the appropriate notation for the sample moments. To avoid repetition, we state the corollary only for *Theorem* 2.2; analogous logarithmic limits and upper bounds for raw integer moments are omitted.

2.3. Verifying Condition A(p). Condition A(p) is a verifiable weak-dependence condition in several standard settings. We now give sufficient conditions for Condition A(p).

An m-dependent random sequence is a random sequence such that there exists a nonnegative integer m and for every positive integer k, $\{X_i\}_{i \le k}$ is independent of $\{X_i\}_{i \ge k+m+1}$. A 0-dependent random sequence is independent.

Theorem 2.5 (m-dependent sequences). *Assume the standing marginal assumption with $\alpha > 0$ [not necessarily in $(0,1)$]. If $\{X_i, i \ge 1\}$ is an m-dependent sequence, then Condition A(p) holds for any $p > \alpha$.*

Now we show that Condition A(p) holds under some commonly used mixing conditions. Let $Z := (Z_k, k \in \mathbb{Z})$ be a sequence of random variables on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{F}_j^r := \sigma(Z_k, j \le k \le r)$ be the σ-field generated by $Z_j, \ldots, Z_r$. For any σ-field $\mathcal{A} \subset \mathcal{F}$, let $\mathcal{L}^2_{\mathrm{Real}}(\mathcal{A})$ denote the space of square-integrable, $\mathcal{A}$-measurable (real-valued) random variables. For any two σ-fields $\mathcal{A}$ and $\mathcal{B} \subset \mathcal{F}$, define the following measures of dependence:

$$\alpha(\mathcal{A},\mathcal{B}) := \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|,$$

$$\phi(\mathcal{A},\mathcal{B}) := \sup_{A \in \mathcal{A}, B \in \mathcal{B}, \mathbb{P}(A)>0} \left|P(B|A) - P(B)\right|,$$

$$\psi(\mathcal{A},\mathcal{B}) := \sup_{A \in \mathcal{A}, B \in \mathcal{B}, \mathbb{P}(A)>0, \mathbb{P}(B)>0} \left|\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)\mathbb{P}(B)} - 1\right|,$$

$$\rho(\mathcal{A},\mathcal{B}) := \sup_{f \in \mathcal{L}^2_{\mathrm{Real}}(\mathcal{A}), g \in \mathcal{L}^2_{\mathrm{Real}}(\mathcal{B})} \left|\mathrm{Corr}(f,g)\right|,$$

$$\beta(\mathcal{A},\mathcal{B}) := \sup \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} \left|\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)\right|,$$

where the last supremum is taken over all pairs of (finite) partitions $\{A_1, \ldots, A_I\}$ and $\{B_1, \ldots, B_J\}$ of Ω such that $A_i \in \mathcal{A}$ for each i and $B_j \in \mathcal{B}$ for each j. For any positive integer n, the strong mixing coefficient is defined as

$$\alpha(n) = \sup_{j \in \mathbb{Z}} \alpha(\mathcal{F}^j_{-\infty}, \mathcal{F}^\infty_{j+n}).$$

Other types of mixing coefficients can be defined similarly by replacing α with their respective Greek letters.

Downloaded from https://www.pnas.org by Joel Cohen on October 4, 2026 from IP address 174.231.48.236.

The random sequence Z is then said to be strongly mixing or α-mixing if $\alpha(n) \to 0$ as $n \to \infty$, and similarly, ϕ-mixing if $\phi(n) \to 0$ as $n \to \infty$. The same applies to all the other types of mixing above. These mixing conditions are commonly considered (34–36). Ref. 14, who considered Taylor's law for a time series of light-tailed random variables with finite moments, also introduced strong mixing conditions and the corresponding mixing rates.

Theorem 2.6 (Strong-mixing criterion for Condition $A(p)$). *Assume the standing marginal assumption with $\alpha > 0$. Let $X_1, X_2, \ldots$ be a strictly stationary sequence of random variables with common finite lower endpoint x_F. Suppose that $\{X_i : i \geq 1\}$ is strongly mixing with strong mixing coefficient $\alpha(k)$. Assume that there exists a sequence of positive integers $r_n \to \infty$ such that $r_n = o(c_n)$ and*

$$\frac{n}{c_n^2} \sum_{k=r_n+1}^{n-1} \alpha(k) \to 0. \qquad [11]$$

Then Condition A(p) holds for every $p > \alpha$.

The conditions in *Theorem* 2.6 are easy to verify when the strong mixing coefficients decay geometrically; the following corollary considers this important special case.

Corollary 2.7 (Geometric strong mixing). *Assume the conditions of Theorem 2.6, and suppose that the strong mixing coefficients satisfy $\alpha(k) \leq C\rho^k$, $k \geq 1$, for some constants $C > 0$ and $\rho \in (0, 1)$. If, in addition, $c_n / \log n \to \infty$, then Condition A(p) holds for every $p > \alpha$. In particular, one may take $c_n = (\log n)^\gamma$ for any $\gamma > 1$.*

The strong mixing condition is among the weakest of the commonly used mixing conditions. For example, ϕ-mixing implies α-mixing, and analogous coefficient inequalities relate ψ-, ρ-, and β-mixing to strong mixing. These relationships are reviewed in ref. 37. The following corollary gives convenient sufficient conditions for Condition A(p) under these stronger mixing notions. The rate assumptions stated in the corollary ensure, through the standard inequalities between mixing coefficients, that Eq. **11** in *Theorem* 2.6 holds.

Corollary 2.8 (Other mixing criteria for Condition A(p)). *Assume the standing marginal assumption with $\alpha > 0$. Let $X_1, X_2, \ldots$ be a strictly stationary sequence of random variables with common finite lower endpoint x_F. If there exists a sequence of positive integers $r_n \to \infty$ such that $r_n = o(c_n)$ and $\{X_i; i \geq 1\}$ is any one of the following:*

1. *ψ-mixing with $\lim_{n\to\infty} \frac{n}{c_n^2} \sum_{k=r_n+1}^{n-1} \psi(k) = 0$;*
2. *ϕ-mixing with $\lim_{n\to\infty} \frac{n}{c_n^2} \sum_{k=r_n+1}^{n-1} \phi(k) = 0$;*
3. *ρ-mixing with $\lim_{n\to\infty} \frac{n}{c_n^2} \sum_{k=r_n+1}^{n-1} \rho(k) = 0$;*
4. *β-mixing with $\lim_{n\to\infty} \frac{n}{c_n^2} \sum_{k=r_n+1}^{n-1} \beta(k) = 0$,*

then Condition A(p) holds for any $p > \alpha$.

We next demonstrate that a first-order autoregressive model known as AR(1) satisfies Condition A(p) for any $p > \alpha$ under certain regularity conditions.

Theorem 2.9 (AR(1) criterion for Condition A(p)). *Fix $x_F \in \mathbb{R}$ and $\beta_1 \in (0, 1)$. Let $\{\varepsilon_t : t \in \mathbb{Z}\}$ be i.i.d. nonnegative random variables with survival function $\mathbb{P}(\varepsilon_1 > y) = y^{-\alpha}\ell_\varepsilon(y)$, $y > 0$, where $\alpha > 0$ and ℓ_ε is slowly varying. Assume that ε_1 has a bounded density f_ε, and that there exists a finite partition of $\mathbb{R}$ into intervals such that, on each interval, f_ε is monotone. Let $\{Y_t : t \in \mathbb{Z}\}$ be the strictly stationary solution of*

$$Y_t = \beta_1 Y_{t-1} + \varepsilon_t, \qquad t \in \mathbb{Z},$$

and put $X_t := x_F + Y_t$. If $c_n / \log n \to \infty$, then $\{X_t : t \geq 1\}$ satisfies Condition A(p) for every $p > \alpha$.

The density condition in *Theorem* 2.9 is used only as a sufficient regularity condition for verifying the weak-dependence requirement in Condition A(p); see ref. 38.

2.4. Central Moments and the Variance–Mean Taylor's Law.

For each integer $k \geq 2$, define the kth sample central moment as $M_{n,k}^{c,X} := \frac{1}{n}\sum_{i=1}^n (X_i - M_{n,1}^X)^k$ and $M_{n,k}^{c,Y} := \frac{1}{n}\sum_{i=1}^n (Y_i - M_{n,1}^Y)^k$. Since $Y_i = X_i - x_F$, we have $M_{n,k}^{c,X} = M_{n,k}^{c,Y}$.

Theorem 2.10 (Central-moment Taylor's law). *Assume the standing marginal assumption with $\alpha \in (0, 1)$.*

(a) Let $k \geq 2$ be an integer. If Condition A(p) holds for $p = 1$ and $p = k$, then

$$\frac{\log |M_{n,k}^{c,X}|}{\log |M_{n,1}^X|} \xrightarrow{\mathbb{P}} \iota_\alpha(k, 1).$$

More precisely,

$$R_{n,k,1}^{hcm1} := \frac{\log |M_{n,k}^{c,X}|}{\log |M_{n,1}^X|} - \iota_\alpha(k, 1) = O_p(\Delta_n).$$

(b) Let $h_1, h_2 \geq 2$ be distinct integers. If Condition A(p) holds for $p = h_1$ and $p = h_2$, then

$$\frac{\log |M_{n,h_1}^{c,X}|}{\log |M_{n,h_2}^{c,X}|} \xrightarrow{\mathbb{P}} \iota_\alpha(h_1, h_2).$$

More precisely,

$$R_{n,h_1,h_2}^{hcm2} := \frac{\log |M_{n,h_1}^{c,X}|}{\log |M_{n,h_2}^{c,X}|} - \iota_\alpha(h_1, h_2) = O_p(\Delta_n).$$

Remark 2.11. Taking $k = 2$ in *Theorem* 2.10(a) gives the variance–mean form of Taylor's law:

$$\frac{\log V_n}{\log |M_{n,1}^X|} \xrightarrow{\mathbb{P}} \frac{2-\alpha}{1-\alpha}, \qquad [12]$$

where $V_n := M_{n,2}^{c,X}$ is the sample variance. This special case corresponds to the sample variance–sample mean form studied in refs. 22, Section 6 and 24, Eq. (10).

2.5. Conditionally Weak Dependence and Correlated Data.

In Section 2.2, we assumed weak dependence, Condition A(p), among the shifted truncated random variables. Here, in the spirit of ref. 25, Theorems 1 and 2, we allow all the random variables to be highly correlated, provided that after conditioning on a σ-field $\mathcal{G}$, the shifted truncated variables have weak enough dependence and their conditional truncated mean is not asymptotically negligible. Of course, a special case is that the random variables are conditionally independent given $\mathcal{G}$.

The next theorem gives the higher-moment Taylor's law for correlated heavy-tailed data. Corresponding statements for higher central moments, analogous to *Theorem* 2.10, are similar.

Downloaded from https://www.pnas.org by Joel Cohen on October 4, 2026 from IP address 174.231.48.236.

Theorem 2.12 (Taylor's law with correlated heavy-tailed data). *Assume the standing marginal assumption with $\alpha > 0$. Suppose that $\mathcal{G}$ is a σ-field. For $p = h_1, h_2 > \alpha$, assume that*

$$\mathbb{E}\left(\frac{\left|\sum_{i\neq j}\mathrm{Cov}(\breve{Y}_i^p, \breve{Y}_j^p \mid \mathcal{G})\right|}{v_n^{2p}c_n^2}\right) = o(1), \qquad [13]$$

and that

$$\lim_{\eta\downarrow 0}\limsup_{n\to\infty}\mathbb{P}\left(\sum_{i=1}^{n}\mathbb{E}(\breve{Y}_i^p \mid \mathcal{G}) < \eta n\mathbb{E}(\breve{Y}_i^p)\right) = 0. \qquad [14]$$

Then

$$\frac{\log M_{n,h_1}^Y}{\log M_{n,h_2}^Y} \xrightarrow{\mathbb{P}} \iota_\alpha(h_1,h_2).$$

More precisely,

$$R_{n,h_1,h_2}^{corr} := \frac{\log M_{n,h_1}^Y}{\log M_{n,h_2}^Y} - \iota_\alpha(h_1,h_2) = O_p(\Delta_n).$$

Examples satisfying the assumptions of *Theorem* 2.12, together with their verifications, are given in *SI Appendix*, section B.5. These examples include two conditionally independent constructions adapted from refs. 23 and 25, and a conditionally 1-dependent construction showing that conditional independence is not necessary.

Our results go beyond the results of refs. 23 and 25 in two ways: We allow the random variables to remain weakly dependent after conditioning on a σ-field $\mathcal{G}$, and, under Eqs. **13** and **14**, we establish Taylor's law not only for the variance but also for higher moments.

3. Dependent, Heterogeneous, and Network-Indexed Heavy-Tailed Data

The preceding section considered dependent observations with a common marginal tail. We now consider two extensions motivated by complex systems, where observations are often neither identically distributed nor naturally arranged in a linear or temporal order, but may instead form an arbitrary finite collection with interactions encoded by a graph or network. First, we consider finite-cluster heterogeneity, where one cluster has an asymptotically dominant right tail. Second, we consider observations indexed by an arbitrary finite set, with network-indexed data as an important example. In both settings, dependence is controlled through covariance conditions on truncated powers, and the same logarithmic limit holds under the assumptions below.

3.1. Finite-Cluster Heterogeneity with One Dominant Tail. In this section, we consider a heterogeneous setting in which the observations come from a fixed finite number of clusters. The clusters may have different distributions, but one cluster has an asymptotically dominant right tail. Ref. 23 considered a related two-cluster setting for Taylor's law, focusing on the scaling relationship between the sample mean and the sample variance. We extend their results to multiple clusters and higher moments, and relax their assumption of independence among random variables.

Let $J \geq 2$ be fixed. For each n, let $\mathcal{I}_{n,1}, \ldots, \mathcal{I}_{n,J}$ be a partition of $\{1,\ldots,n\}$, and write $n_{r,n} := |\mathcal{I}_{n,r}|$ and $\pi_{r,n} := n_{r,n}/n$ for $r = 1,\ldots,J$. Assume that

$$\pi_{r,n} \to \pi_r, \qquad r = 1,\ldots,J, \qquad [15]$$

where $\pi_r \geq 0$ and $\sum_{r=1}^J \pi_r = 1$. Without loss of generality, we label the dominant cluster as cluster 1 and assume

$$\pi_1 > 0. \qquad [16]$$

For $i \in \mathcal{I}_{n,r}$, the random variable $X_{n,i}$ has the same distribution as the other observations in cluster r. Let $x_{F,r}$ be the finite lower endpoint of the distribution in cluster r. Since J is fixed, $x_F := \min_{1\leq r\leq J} x_{F,r} > -\infty$. Define the shifted nonnegative observations $Y_{n,i} := X_{n,i} - x_F$, $i = 1,\ldots,n$. For $i \in \mathcal{I}_{n,r}$, let $\overline{F}_r(x)$ denote the cluster-r survival function after the common shift. We assume that cluster 1 has tail

$$\overline{F}_1(x) = x^{-\alpha}l(x), \qquad x \geq x_0, \qquad [17]$$

for some $x_0 > 0$, where $\alpha \in (0,1)$ and l is slowly varying. The other clusters are asymptotically lighter-tailed than cluster 1:

$$\lim_{x\to\infty}\frac{\overline{F}_r(x)}{\overline{F}_1(x)} = 0, \qquad r = 2,\ldots,J. \qquad [18]$$

As in Section 1, let $\{c_n\}$ be a positive sequence such that $c_n \to \infty$ and $\log c_n = o(\log n)$. Choose $\{t_n\}$ and $\{v_n\}$ according to the dominant tail:

$$\frac{nl(t_n)}{t_n^\alpha} \sim \frac{1}{c_n}, \qquad \frac{nl(v_n)}{v_n^\alpha} \sim c_n. \qquad [19]$$

For $p > 0$, define the shifted pth sample moment

$$M_{n,p}^Y := \frac{1}{n}\sum_{r=1}^{J}\sum_{i\in\mathcal{I}_{n,r}} Y_{n,i}^p.$$

Define the truncated shifted variables $\breve{Y}_{n,i} := Y_{n,i}\mathbb{1}(Y_{n,i} \leq v_n)$, $i = 1,\ldots,n$.

Remark 3.1. Condition Eq. **18** includes the case in which

$$\overline{F}_r(x) \sim x^{-\alpha_r}l_r(x), \qquad r = 1,\ldots,J,$$

where l_r is slowly varying and $\alpha_1 < \alpha_r$ for every $r \geq 2$. It also allows clusters with the same power index as the dominant cluster, provided their slowly varying factors are asymptotically smaller, i.e. $l_r(x)/l_1(x) \to 0$.

Theorem 3.2 (Finite cluster heterogeneity). *Assume Eqs. **15–19**. Let $h_1, h_2 > \alpha$. Suppose that, for $p = h_1$ and $p = h_2$,*

$$\sum_{i\neq j}\mathrm{Cov}\left(\breve{Y}_{n,i}^p, \breve{Y}_{n,j}^p\right) = o\left(v_n^{2p}c_n^2\right). \qquad [20]$$

Then

$$\frac{\log M_{n,h_1}^Y}{\log M_{n,h_2}^Y} \xrightarrow{\mathbb{P}} \iota_\alpha(h_1,h_2). \qquad [21]$$

More precisely,

$$R_{n,h_1,h_2}^{clust} := \frac{\log M_{n,h_1}^Y}{\log M_{n,h_2}^Y} - \iota_\alpha(h_1,h_2) = O_p(\Delta_n). \qquad [22]$$

Downloaded from https://www.pnas.org by Joel Cohen on October 4, 2026 from IP address 174.231.48.236.

3.2. Network-Indexed Data and Local Dependence. Although we have stated the generalizations of Taylor's law for a denumerable set of random variables $\{X_1, X_2, \ldots, X_n, \ldots\}$ with a common finite lower endpoint x_F, the results clearly also hold for a set of random variables $\{X_i : i \in \mathcal{I}_n\}$ with an arbitrary nonempty index set $\mathcal{I}_n$ of cardinality n, after putting $Y_i = X_i - x_F$, and modifying the notation in the corresponding conditions accordingly. For example, for Taylor's law for higher moments in *Theorem* 2.2, Condition A(p) becomes

$$\sum_{i,j\in\mathcal{I}_n, i\neq j} \mathrm{Cov}(\breve{Y}_i^p, \breve{Y}_j^p) = o(v_n^{2p} c_n^2). \qquad [23]$$

Here is an example where Eq. **23** is satisfied in a network graph.

Example 3.3. *For a natural number n, let $G_n = (V_n, E_n)$ be an undirected or directed graph, where $V_n = \{v_1, \ldots, v_n\}$ is a set of n vertices and E_n is a nonempty set of edges. Define $[n] := \{1, \ldots, n\}$ and let $d(v_i, v_j)$ be the shortest unweighted distance between v_i and v_j according to E_n, with $d(v_i, v_j) = \infty$ if no such path exists. For a directed graph, $d(v_i, v_j)$ is computed after ignoring edge directions. Let $U_{n,h}$ be the subset of $\mathcal{S}_n := \{(i,j) : i,j \in [n]\}$ such that the unweighted distance between two nodes is h:*

$$U_{n,h} := \{(i,j) \in \mathcal{S}_n : d(v_i, v_j) = h\}.$$

$U_{n,h}$ describes a specific network structure.

Suppose that each node v_i is associated with a random variable X_i, that $X_1, \ldots, X_n$ have a common finite lower endpoint x_F, and define $Y_i = X_i - x_F$. Assume that $Y_1, \ldots, Y_n$ have a common survival function $\overline{F}(x) = x^{-\alpha} l(x)$, where l is slowly varying and $\alpha > 0$. Assume further that, for the $p > \alpha$ under consideration,

(i) $\mathrm{Cov}(\breve{Y}_i^p, \breve{Y}_j^p) = 0$ if $d(v_i, v_j) > M$, where M is a fixed nonnegative integer;

(ii) for all $i \in [n]$, $|\{j : (i,j) \in U_{n,1}\}| \leq K$, where K is a fixed positive integer. (In graph-theoretic language, the degree of every vertex does not exceed K.)

Condition (i) is satisfied, for example, if X_i and X_j are independent whenever $d(v_i, v_j) > M$, since independence is preserved under deterministic transformations; hence $\breve{Y}_i^p$ and $\breve{Y}_j^p$ are independent whenever $d(v_i, v_j) > M$.

*Then the network version of Condition A(p), namely Eq. **23**, holds. SI Appendix, section 4 gives the proof.*

3.3. Why the Same Logarithmic Limit Appears. The main results in the i.i.d., dependent, and heterogeneous settings yield the same asymptotic relationship:

$$R_{n,h_1,h_2} := \frac{\log M_{n,h_1}}{\log M_{n,h_2}} \xrightarrow{\mathbb{P}} \frac{h_1 - \alpha}{h_2 - \alpha}. \qquad [24]$$

Under the respective assumptions, Taylor's law is invariant with respect to the dependence and heterogeneity considered. Such insensitivity may not be too surprising, given that the result is on a logarithmic scale. Indeed, in the homogeneous cases, we have $\log M_{n,h} = \log d_{n,h} + O_p(\log c_n)$ and by *Lemma* 1.1, $\log d_{n,h} = \frac{h-\alpha}{\alpha} \log n + O(\log c_n) + O(|\log l(t_n)|)$. Since $\log c_n = o(\log n)$ and $\log l(t_n) = o(\log n)$, these error terms are negligible when we consider ratios of logarithmic sample moments, leading to the same limit depending only on α. The same type of asymptotic expansion holds in the heterogeneous case when one cluster has a dominant tail.

3.4. Estimation and Application. When Eq. **24** holds, solving for α and applying the continuous mapping theorem gives consistent estimators of α:

$$\hat{\alpha}_{n,h_1,h_2} := \frac{h_1 - h_2 R_{n,h_1,h_2}}{1 - R_{n,h_1,h_2}}.$$

Ref. 24, section 6A gave the analogous variance-based estimator for $h_1 = 2, h_2 = 1$ corresponding to Eq. **12**,

$$\hat{\alpha}^c_{n,2,1} = \left(2 - \frac{\log V_n}{\log |M^X_{n,1}|}\right) \Big/ \left(1 - \frac{\log V_n}{\log |M^X_{n,1}|}\right). \qquad [25]$$

In the sample version of Taylor's law with infinite-mean data, the quantity $\hat{b}_n := (\log V_n)/(\log M^X_{n,1})$ can be interpreted as an estimate of the Taylor's law exponent b, where b is defined in Eq. **2**. In the setting considered here, Eq. **12** gives $\hat{b}_n \xrightarrow{\mathbb{P}} b = (2-\alpha)/(1-\alpha)$. If $\alpha \in (0,1)$, then

$$b - 2 = \frac{\alpha}{1-\alpha} > 0.$$

Thus, for sufficiently large sample sizes, the estimated Taylor's law exponent $\hat{b}_n$ exceeds 2 with high probability.

4. Network Data Illustration

In this section, we illustrate Taylor's law for network data, as introduced in Section 3.2, using three real-world datasets. To the best of our knowledge, these examples are the first to explore the application of Taylor's law in network data.

For each dataset, we estimate the heavy-tail index α using the stabilizing region of the alternative Hill plot (39). For each dataset, the Hill estimate, based on the stabilizing region of the alternative Hill plot, gives a heavy-tail index α less than 1. We also compare this estimate with the Taylor's law estimator Eq. **25**. We find that the two estimates generally agree closely.

To illustrate Taylor's law for each dataset, we construct subsamples of varying sizes from the full dataset and, for each subsample, we plot log-variance on the vertical axis against log-mean on the horizontal axis. All the plots reveal a clear linear relationship, consistent with Taylor's law for network data. Except when we explicitly state otherwise, we use the natural logarithm, base e.

The empirical analyses below should be interpreted as finite-sample tail diagnostics for node-associated values in finite networks. They are not formal tests that the underlying distribution of any network-generating mechanism has infinite mean. The Hill, Pareto, Taylor's law, and generalized Pareto distribution estimates summarize different aspects of the observed upper tail and are used here descriptively, because the node-associated observations are heterogeneous and network-derived rather than independent observations from a common population distribution.

4.1. Wikipedia Talk Dataset. Wikipedia is a free encyclopedia collaboratively written by volunteers worldwide. Each registered user has a talk page that serves as a space for communication and coordination. Both the user and others can edit this page to exchange messages or discuss edits to articles. The Wikipedia Talk dataset captures all user interactions via talk page edits from the inception of Wikipedia through January 2008 (40). This network represents the communication structure within the

Downloaded from https://www.pnas.org by Joel Cohen on October 4, 2026 from IP address 174.231.48.236.

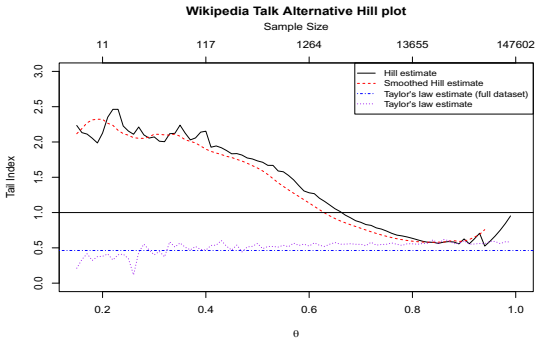


Fig. 1. Alternative Hill plot and tail index estimates implied by Taylor's law for Wikipedia talk dataset. The Hill and smoothed Hill estimates are plotted against the threshold parameter θ, shown on the bottom x-axis. The Taylor's law estimate based on subsamples of the full dataset has sample sizes indicated on the top x-axis. The Hill estimate at $\theta = 0.85$ is 0.563 (99% CI: 0.554 to 0.572).

Wikipedia community. Each node corresponds to a user, and a directed edge from user A to user B indicates that user A edited the talk page of user B. For each user, the total number of edits they made to other users' talk pages is treated as a node attribute. We use these values to illustrate Taylor's law. Talk-page edit counts may exhibit a heavy tail because participation and coordination in a collaborative online community are highly uneven: A few highly active users may edit many users' talk pages, while most active users make relatively few such edits. The dataset contains 2,394,385 users in total. We removed users who never interacted with others, as they are not active participants. The remaining dataset consists of 147,602 users. The maximum number of edits by a single user is 100,022, while the average is 34 and the sample variance is 160,858.

In Fig. 1, the alternative Hill plot displays estimates of the tail index as a function of the threshold parameter $\theta \in [0, 1]$, where the number of upper order statistics used in the Hill estimate is $\lceil n^{\theta} \rceil$. This rescaling facilitates interpretation and improves accuracy (39). In the figure, the black solid line represents the Hill estimate and the red dashed line represents its smoothed version as proposed in ref. 39, computed as the simple average of the subsequent $\lceil n^{\theta} \rceil$ Hill estimates. The figure shows that the Hill estimate stabilizes for $\theta \in [0.8, 0.9]$. Following ref. 41, we construct a 99% asymptotic CI for $1/\alpha$ using the normal approximation. We then obtain a CI for α by taking the reciprocals of the two endpoints in reverse order. The tail index inferred from this stabilization region closely aligns with that implied by Taylor's law using the full dataset (blue dash-dot horizontal line), and is between 0 and 1.

Fig. 1 also plots the estimates Eq. **25** from Taylor's law using increasingly large random subsamples of the full dataset (purple dotted line, slightly fluctuating), with the corresponding sample sizes indicated on the top x-axis. While the Hill plot is constructed using varying numbers of upper order statistics from the full dataset, the Taylor's law estimates are computed from fixed-size subsamples. Tail index estimates from Eq. **25** converge toward and align with the estimate obtained from the Hill plot.

To demonstrate Taylor's law, we generate a sequence of samples of increasing sizes, evenly spaced on a logarithmic scale, drawn from the full dataset. The sequence is created by taking uniform increments from $\log_{10}(500)$ to $\log_{10}(N)$, where N is the total number of samples. Each value is then transformed using the base-10 exponential function and rounded to the nearest integer. This approach results in denser sampling at smaller sizes and sparser sampling at larger sizes, minimizing redundancy where estimates stabilize. For each subsample, we compute the corresponding log-mean and log-variance. Fig. 2 plots the log-variances versus the log-means of these subsamples. A linear fit approximated the points, in rough agreement with Taylor's law with slope 4.0, greater than 2 as expected. We constructed two-sided nominal 95% and 99% Student t-based CIs for the intercept and the slope. Because the log-mean and log-variance pairs are computed from overlapping subsamples of the same dataset, the reported intervals should be interpreted as approximate model-based measures of uncertainty rather than intervals with exact finite-sample coverage.

4.2. Epinions Dataset. The Epinions dataset is a bipartite rating network from Epinions, an online platform for product reviews (42). Nodes represent users and products. Edges connect users to the products they have rated; see refs. 43 and 44 for more details. We use the number of edges connected to each product node (i.e., the number of reviews received by each product) to illustrate a network Taylor's law. Product-review counts may exhibit tail heaviness because attention is highly uneven across products. For example, visibility, age, category, and demand can cause some products to accumulate many more reviews than others, and existing reviews may further attract future reviews. The dataset includes 635,268 user nodes and 120,492 product nodes. The maximum number of edges connected to a product node is 162,202, with an average of 113.4 edges per product, and a sample variance of 1,129,884. We apply the same methods as used for the Wikipedia talk dataset.

In Fig. 3, the Hill estimate appears to stabilize around $\theta \in [0.75, 0.85]$. Although there is a slight discrepancy between the tail index inferred from Taylor's law and that estimated from the Hill plot, both indicate that the tail index is less than 1. Fig. 4 shows a clear linear relationship between the log-variances and log-means of the subsamples, consistent with Taylor's law with slope 3.1, greater than 2 as expected.

4.3. DBpedia Dataset. The DBpedia dataset includes a bipartite network representing affiliations between notable individuals and

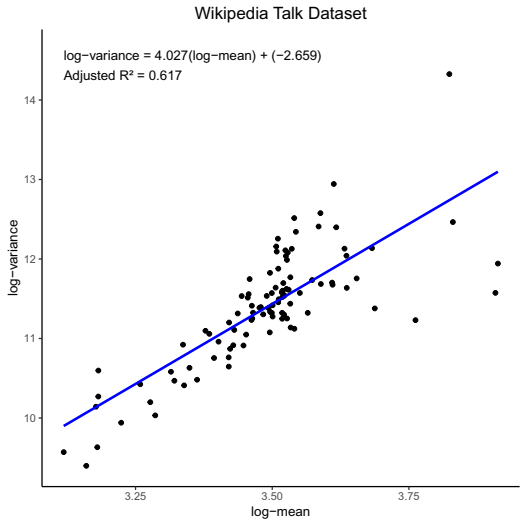


Fig. 2. Log-variance versus log-mean across subsamples for Wikipedia talk dataset. The analysis includes 100 pairs (log-mean, log-variance). The regression line was fitted using ordinary least squares. The 95% and 99% CIs for the intercept are $(-4.864, -0.455)$ and $(-5.577, 0.258)$ respectively. For the slope, the 95% CI is $(3.396, 4.658)$, while the 99% CI is $(3.191, 4.863)$. The adjusted R^2 value of the regression model is 0.617.

Downloaded from https://www.pnas.org by Joel Cohen on October 4, 2026 from IP address 174.231.48.236.

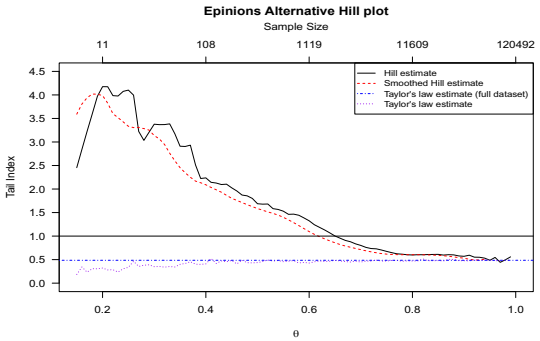


Fig. 3. Alternative Hill plot and tail index estimates implied by Taylor's law for Epinions dataset. The Hill and smoothed Hill estimates are plotted against the threshold parameter θ, shown on the bottom x-axis. The Taylor's law estimate based on subsamples of the full dataset has sample sizes indicated on the top x-axis. The Hill estimate at $\theta = 0.8$ is 0.539 (99% CI: 0.535 to 0.543).

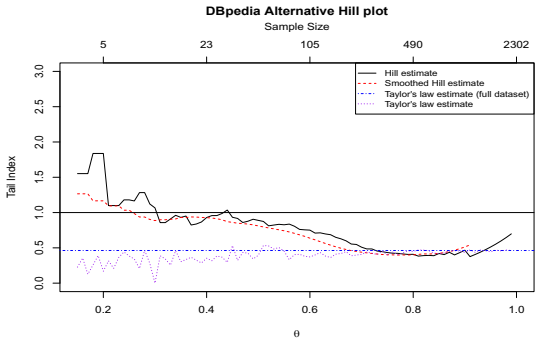


Fig. 5. Alternative Hill plot and tail index estimates implied by Taylor's law for DBpedia dataset. The Hill and smoothed Hill estimates are plotted against the threshold parameter θ, shown on the bottom x-axis. The Taylor's law estimate based on subsamples of the full dataset has sample sizes indicated on the top x-axis. The Hill estimate at $\theta = 0.8$ is 0.409 (99% CI: 0.366 to 0.463).

countries worldwide, extracted from Wikipedia through the DBpedia project (45). The set of "countries" in this dataset includes not only current nations but also former states, empires, kingdoms, and certain country-like entities; (46) gives more details.

We use the number of edges connected to each country node (i.e., the number of entities affiliated with each country) to illustrate a network Taylor's law. The number of entities affiliated with each country may exhibit a heavy upper tail because country nodes are highly heterogeneous aggregation units. Countries differ greatly in population size, historical duration, geopolitical prominence, and coverage in Wikipedia and DBpedia. Since the dataset also includes former countries, empires, kingdoms, and some country-like entities, a few country nodes can accumulate affiliations across many periods and domains, while most have relatively few affiliated entities. The dataset contains 590,112 entity nodes and 2,302 country nodes. On average, each country node is connected to 276.8 edges, with a maximum of 111,132 edges and a sample variance of 9,700,633. We again apply the same analysis as before. The only difference is that, because this graph has fewer nodes, we generate the log-variance versus log-mean plot by progressively sampling the data starting from 10 nodes instead of 500 nodes in the logarithmic sequence.

In Fig. 5, the Hill estimate appears to stabilize around $\theta \in [0.75, 0.85]$. The corresponding tail index aligns closely with the estimate inferred from Taylor's law; both suggest that the tail index is less than 1. Fig. 6 reveals a clear approximately linear relationship between the log-variances and log-means of the subsamples, approximately consistent with Taylor's law with slope 2.8, greater than 2 as expected.

4.4. Further Empirical Analysis and Remarks.

4.4.1. Hill and Taylor's law diagnostics. The fact that the Hill plots for the three datasets stabilize in similar regions around $\theta \approx 0.8$ is noteworthy, but it should not be interpreted as a universal numerical rule. In the alternative Hill plot, θ determines the number $k = \lceil n^{\theta} \rceil$ of upper order statistics used in the Hill estimate; it is not itself a tail-index estimate or a fixed fraction of the data. For i.i.d. regularly varying data, the usual consistency theory for the Hill estimator requires only $k \to \infty$ and $k/n \to 0$, and the practical choice of k is well known to be delicate (39).

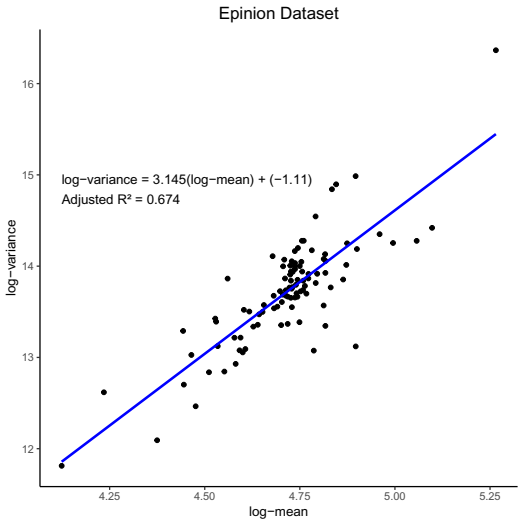


Fig. 4. Log-variance versus log-mean across subsamples for Epinions dataset. The analysis includes 100 pairs (log-mean, log-variance). The regression line was fitted using ordinary least squares. The 95% and 99% CIs for the intercept are $(-3.166, 0.945)$ and $(-3.831, 1.610)$ respectively. For the slope, the 95% CI is $(2.709, 3.580)$, while the 99% CI is $(2.568, 3.721)$. The adjusted R^2 value of the regression model is 0.674.

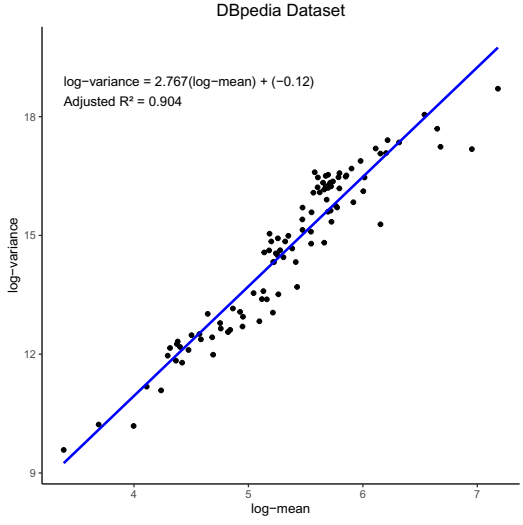


Fig. 6. Log-variance versus log-mean across subsamples for DBpedia dataset. A total of 100 pairs of (log-mean, log-variance) were included in the analysis. The regression line was fitted using ordinary least squares. The 95% and 99% CIs for the intercept are $(-1.091, 0.850)$ and $(-1.405, 1.164)$ respectively. For the slope, the 95% CI is $(2.587, 2.946)$, while the 99% CI is $(2.528, 3.005)$. The adjusted R^2 value of the regression model is 0.904.

Downloaded from https://www.pnas.org by Joel Cohen on October 4, 2026 from IP address 174.231.48.236.

Formal optimal choices of k, and hence of θ, depend on the assumed tail model and loss criterion; they can be derived under additional second-order tail assumptions, and adaptive procedures for estimating such choices have been proposed (47, 48). However, these consistency and optimality results need not apply directly to dependent and heterogeneous network-derived observations.

In all three examples (Figs. 2, 4, and 6), the lower bounds of the 99% CIs for the estimated slopes in the log sample variance versus log sample mean plots exceed 2, as expected from Section 3.4. These estimated slopes are consistent with considerable heterogeneity in activity or affiliation: Most nodes have few connections or interactions, while a small number of nodes have very many.

4.4.2. Negative binomial and Pareto fits. To see directly whether the data are better approximated by an overdispersed (variance > mean) light-tailed distribution or a heavy-tailed distribution, we fitted both the negative binomial distribution and the Pareto distribution to the node-associated values for the three data examples (Fig. 7). We fitted the negative binomial distribution using `fitdist()` from the `fitdistrplus` package in `R`, which estimates parameters by maximum likelihood. We fitted the Pareto distribution by least squares using `epareto()` from the `VGAM` package in `R`; see ref. 49 for additional discussion of least-squares fitting for the Pareto distribution.

To compare the data with the fitted models visually, we plotted the log of the estimated survival function based on the empirical distribution function and overlaid the corresponding functions from the fitted Pareto and negative binomial distributions. We log-transformed the horizontal axis so that the Pareto distribution's survival function would appear as a straight line.

In all three datasets, the Pareto distribution fit the empirical data substantially better than the negative binomial distribution for moderate values of x, but the data exhibit lighter tails than those implied by the fitted Pareto distribution. The discrepancies between the empirical upper tails and the fitted Pareto upper tails, though of uncertain origin, are more likely to be due to heterogeneity and dependence in the network data than to the absence of a slowly varying factor from the Pareto distribution. In a regularly varying function, as we assume here, a slowly varying factor would be more influential for low to moderate values of the random variable and asymptotically negligible in affecting the slope for large values of the random variable. Despite this uncertainty about the origin of the deviations from a Pareto distribution, it is clear that the Pareto distribution approximates the data better than the negative binomial distribution.

In addition, the tail indices from the Hill estimator and those from the Pareto distribution fitted by least squares do not differ substantially. The estimates are reported in the captions of Figs. 1, 3, 5, and 7.

4.4.3. Generalized Pareto fit and threshold choice. To describe the upper tail of the data better, we also applied the peaks-over-threshold approach and fitted a generalized Pareto distribution via maximum likelihood estimation using the `fitgpd()` function from the `POT` package in `R`. Threshold selection remains a difficult problem in peaks-over-threshold analysis (50). The thresholds here were selected diagnostically rather than by applying a universally optimal rule. We selected thresholds using graphical diagnostics by jointly considering threshold-stability plots for the fitted GPD shape and modified-scale parameters, visual agreement between the fitted GPD survival curve and the empirical upper-tail survival curve (51); the corresponding threshold-stability plots and DBpedia threshold-comparison plots are provided in *SI Appendix*, section H.

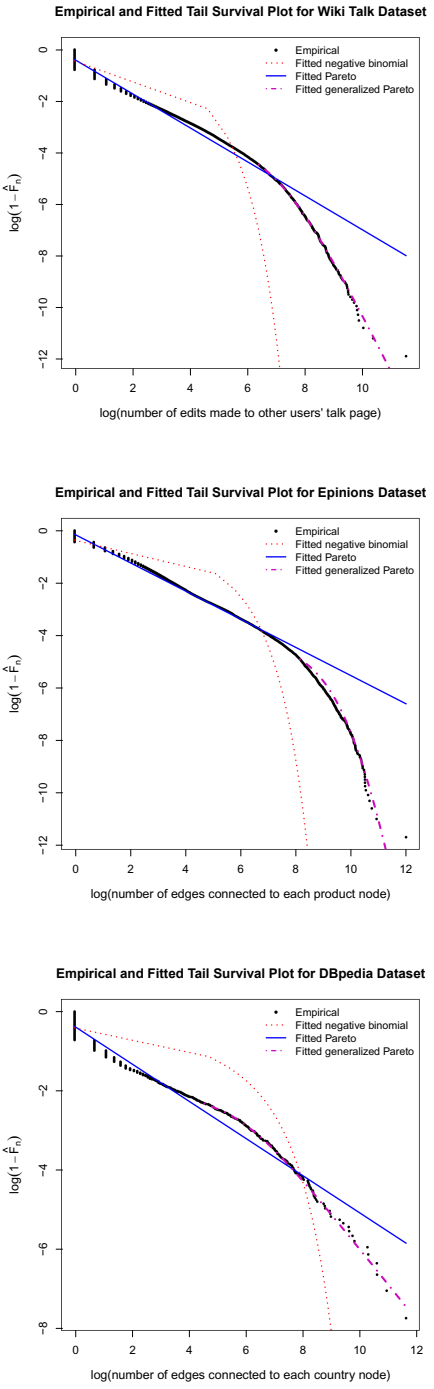


Fig. 7. Plots of $\log(1 - \hat{F}_n)$, where $\hat{F}_n$ denotes the empirical distribution function or one of the fitted distributions (negative binomial, Pareto, or generalized Pareto). For the Wiki Talk, Epinions, and DBpedia datasets, the estimated tail indices with 99% nonparametric bootstrap CIs based on the least-squares Pareto fits are 0.661 (0.655 to 0.667), 0.539 (0.535 to 0.543), and 0.471 (0.443 to 0.512), respectively. The corresponding Pareto-equivalent tail-index estimates $\alpha_{\mathrm{GPD}} = 1/\xi$ based on maximum-likelihood GPD fits are 2.231 (1.748 to 2.675), 5.869 (3.571 to 16.472), and 0.899 (0.637 to 1.238), respectively; the reported GPD intervals are 99% likelihood-based CIs computed by the `POT` package and transformed from the corresponding intervals for the GPD shape parameter ξ. The corresponding thresholds used in fitting the generalized Pareto distribution are 6.39, 8.16, and 4.55 on the log scale.

For the Wiki Talk and Epinions datasets, these diagnostics gave broadly consistent choices: The selected thresholds were near the lower end of regions where the fitted quantities were comparatively stable and the fitted GPD survival curves agreed

Downloaded from https://www.pnas.org by Joel Cohen on October 4, 2026 from IP address 174.231.48.236.

reasonably well with the empirical upper tails. For the DBpedia dataset, the threshold-stability diagnostics were less decisive because the sample is smaller and high thresholds leave relatively few exceedances. We therefore placed more weight on the fitted upper-tail survival curve and chose a threshold that provided better agreement with the empirical tail while retaining enough exceedances.

The corresponding estimated survival functions for values above the selected thresholds are plotted in Fig. 7. The generalized Pareto distribution generally provides a better description of the far upper tail than a single Pareto distribution fitted over a broader range. For the GPD fits, we report the Pareto-equivalent tail index $\alpha_{GPD} = 1/\xi$, where ξ is the usual GPD shape parameter. Under the fitted GPD model, $\alpha_{GPD} > 1$ corresponds to a finite mean and $\alpha_{GPD} > 2$ corresponds to a finite variance.

The GPD estimates should not be expected to agree exactly with the Hill, least-squares Pareto, or Taylor-law estimates, because the different estimates summarize different aspects of the data. The Hill estimates near $\theta \approx 0.8$ use the $k = \lceil n^{\theta} \rceil$ largest observations and therefore summarize a relatively broad portion of the upper tail. The least-squares Pareto estimates provide global approximations to the empirical survival curves, while the Taylor-law estimates are inferred from variance–mean scaling across broad subsamples of the full datasets. Neither is a high-threshold tail-only fit. By contrast, the peaks-over-threshold GPD fits use only exceedances above selected high thresholds. For Wiki Talk and Epinions, in particular, the selected thresholds leave only about 1.1% and 0.7% of the observations, respectively, as exceedances, so the corresponding GPD estimates are based on the most extreme observed tail portions. One possible finite-sample explanation for the discrepancy between the GPD estimates and the other estimates is that the farthest observed tail may appear lighter, with the empirical log-survival curve steepening at the largest observations relative to the broader upper-tail behavior captured by the Hill, Pareto, and Taylor-law estimates. Such apparent steepening could arise from finite-sample effects, finite observation windows, heterogeneity, dependence, or clustering of the largest observations. This behavior can lead to a smaller fitted GPD shape parameter ξ and hence to a larger Pareto-equivalent tail-index estimate $\alpha_{GPD} = 1/\xi$. *SI Appendix*, section H.2 includes a simulation study illustrating how such a discrepancy can arise.

4.4.4. Interpretation and implications for network data. The observed scaling may provide a way to detect heterogeneity across subgraphs: If the log-variance versus log-mean relationship in a subset deviates notably from the overall trend, it may indicate a distinct structural pattern. Additionally, the scaling exponent may serve as a diagnostic tool for evaluating generative models of network data, as not all models reproduce the level of dispersion observed in these examples.

5. Discussion

We present a framework for establishing the limits in probability of the ratios of moments that arise in various forms of Taylor's law when the data can be heavy-tailed with infinite mean and tail index $\alpha \in (0, 1)$, dependent, and heterogeneous. We explore Taylor's law for higher (centered) moments and semivariances, along with its most commonly known form Eq. **2**. We introduce a covariance condition [Condition A(p)] that leads to Taylor's law for dependent data, and we propose mixing conditions that satisfy this covariance condition. We extend our results to heterogeneous and correlated heavy-tailed-dependent data and to an arbitrary index set, illustrated by network-dependent data. To our knowledge, networks are a previously unexplored setting for Taylor's law. The theoretical results are supported with three empirical examples and extensive simulations.

A question to consider in the future is whether the assumption that data are i.i.d. when $\alpha \in (1, 2)$, which is considered in ref. 25, can be relaxed to allow weak dependence. In addition, more sophisticated network structures from real-world social network data can be used to test the descriptive ability of Taylor's law in new domains.

Data, Materials, and Software Availability. The datasets analyzed in this study are publicly available from the Stanford Network Analysis Project [Wikipedia Talk network; (40)], Network Repository [Epinions product ratings; (42)], and Netzschleuder [DBpedia person-country affiliations; (45)].

ACKNOWLEDGMENTS. We thank reviewers V.A. and Q.T. and a member of the PNAS Editorial Board for very careful, constructive suggestions, and improvements. J.E.C. thanks Roseanne Benjamin for help during this work. H.K.L. acknowledges The Natural Sciences and Engineering Research Council of Canada Grant RGPIN/03124-2021. S.C.P.Y. acknowledges grants HKGRF-14301321 ("General Theory . . . Classical Problems"), HKGRF-14300123 ("Well-posedness . . . Applications"), HKGRF-14300025 ("A Generic Theory for Stochastic Control against Fractional Brownian Motions"), and G-CUHK411/23. He thanks The University of Texas at Dallas for the kind invitation to be a Visiting Professor in Naveen Jindal School of Management.

Author affiliations: [a]Department of Statistics, Columbia University, New York, NY 10027; [b]Laboratory of Populations, The Rockefeller University, New York, NY 10065-6399; [c]Climate School, Columbia University, New York, NY 10027; [d]Department of Statistics, University of Chicago, Chicago, IL 60637; [e]Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada; and [f]Department of Statistics and Data Science, The Chinese University of Hong Kong, China

1. S. T. Rachev, *Handbook of Heavy Tailed Distributions in Finance: Handbooks in Finance, Book 1* (Elsevier, 2003).
2. P. Glasserman, P. Heidelberger, P. Shahabuddin, Portfolio value-at-risk with heavy-tailed risk factors. *Math. Finance* **12**, 239–269 (2002).
3. G. W. Peters, P. V. Shevchenko, *Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk* (John Wiley & Sons, 2015).
4. B. Merz *et al.*, Understanding heavy tails of flood peak distributions. *Water Resour. Res.* **58**, e2021WR030506 (2022).
5. Y. H. Eom, H. H. Jo, Tail-scope: Using friends to estimate heavy tails of degree distributions in large-scale complex networks. *Sci. Rep.* **5**, 09752 (2015).
6. J. Nešlehová, P. Embrechts, V. Chavez-Demoulin, Infinite mean models and the LDA for operational risk. *J. Oper. Risk* **1**, 3–25 (2006).
7. V. F. Pisarenko, M. V. Rodkin, *Heavy-Tailed Distributions in Disaster Analysis* (Springer Science & Business Media, 2010), vol. 30.
8. L. R. Taylor, Aggregation, variance and the mean. *Nature* **189**, 732–735 (1961).
9. R. A. Taylor, *Taylor's Power Law: Order and Pattern in Nature* (Academic Press, 2019).
10. Z. Eisler, I. Bartos, J. Kertész, Fluctuation scaling in complex systems: Taylor's law and beyond. *Adv. Phys.* **57**, 89–142 (2008).
11. Q. S. Hanley, S. Khatun, A. Yosef, R. M. Dyer, Fluctuation scaling, Taylor's law, and crime. *PLoS One* **9**, e109004 (2014).
12. A. Giometto, M. Formentin, A. Rinaldo, J. E. Cohen, A. Maritan, Sample and population exponents of generalized Taylor's law. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7755–7760 (2015).
13. J. E. Cohen, M. Xu, H. Brunborg, Taylor's law applies to spatial variation in a human population. *Genus* **69**, 25–60 (2013).
14. V. H. de la Peña, P. Doukhan, Y. Salhi, A dynamic Taylor's law. *J. Appl. Probab.* **59**, 584–607 (2022).
15. M. C. K. Tweedie, "An index which distinguishes between some important exponential families" in *Statistics: Applications and New Directions: Proc. Indian Statistical Institute Golden Jubilee International Conference* (1984), pp. 579–604.
16. S. K. Bar-Lev, P. Enis, Reproducibility and natural exponential families with power variance functions. *Ann. Stat.* **14**, 1507–1522 (1986).
17. S. K. Bar-Lev, O. Stramer, Characterizations of natural exponential families with power variance functions by zero regression properties. *Probab. Theory Relat. Fields* **76**, 509–522 (1987).
18. S. Bar-Lev, Independent, tough identical results: The class of Tweedie on power variance functions and the class of Bar-Lev and Enis on reproducible natural exponential families. *Int. J. Stat. Probab.* **9**, 30–35 (2020).

Downloaded from https://www.pnas.org by Joel Cohen on October 4, 2026 from IP address 174.231.48.236.

19. W. S. Kendal, Taylor's ecological power law as a consequence of scale invariant exponential dispersion models. *Ecol. Complex.* **1**, 193–209 (2004).
20. W. S. Kendal, B. Jørgensen, Taylor's power law and fluctuation scaling explained by a central-limit-like convergence. *Phys. Rev. E* **83**, 066115 (2011).
21. W. S. Kendal, B. Jørgensen, Tweedie convergence: A mathematical basis for Taylor's power law, 1/*f* noise, and multifractality. *Phys. Rev. E* **84**, 066120 (2011).
22. M. Brown, J. E. Cohen, V. H. de la Peña, Taylor's law, via ratios, for some distributions with infinite mean. *J. Appl. Probab.* **54**, 657–669 (2017).
23. J. E. Cohen, R. A. Davis, G. Samorodnitsky, Heavy-tailed distributions, correlations, kurtosis and Taylor's law of fluctuation scaling. *Proc. R. Soc. A* **476**, 20200610 (2020).
24. M. Brown, J. E. Cohen, C. F. Tang, S. C. P. Yam, Taylor's law of fluctuation scaling for semivariances and higher moments of heavy-tailed data. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2108031118 (2021).
25. J. E. Cohen, R. A. Davis, G. Samorodnitsky, Covid-19 cases and deaths in the United States follow Taylor's law for heavy-tailed distributions with infinite variance. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2209234119 (2022).
26. A. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Springer, 1933).
27. N. H. Bingham, C. M. Goldie, J. L. Teugels, *Regular Variation* (Cambridge University Press, 1987), vol. 27.
28. E. Seneta, *Regularly Varying Functions* (Springer, 1976), vol. 508.
29. S. Foss, D. Korshunov, S. Zachary, *An Introduction to Heavy-Tailed and Subexponential Distributions* (Springer Series in Operations Research and Financial Engineering, Springer, ed. 2, 2013).
30. P. Embrechts, C. Klüppelberg, T. Mikosch, *Modelling Extremal Events: For Insurance and Finance* (Springer Science & Business Media, 2013), vol. 33.
31. J. Nair, A. Wierman, B. Zwart, *The Fundamentals of Heavy Tails* (Cambridge University Press, 2022), vol. 53.
32. S. Borak, W. Härdle, R. Weron, *Statistical Tools for Finance and Insurance* (Springer, Berlin, 2005), pp. 21–44.
33. W. Feller, *An Introduction to Probability Theory and Its Applications* (John Wiley & Sons, New York, ed. 2, 1971), vol. 2.
34. Z. Lin, C. Lu, *Limit Theory for Mixing Dependent Random Variables* (Springer Science & Business Media, 1997), vol. 378.
35. F. Merlevede, M. Peligrad, The functional central limit theorem under the strong mixing condition. *Ann. Probab.* **28**, 1336–1352 (2000).
36. A. Quintela-Del-Río, Hazard function given a functional variable: Non-parametric estimation under strong mixing conditions. *J. Nonparametric Stat.* **20**, 413–430 (2008).
37. R. C. Bradley, Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2**, 107–144 (2005).
38. D. W. Andrews, "First order autoregressive processes and strong mixing" (Cowles Foundation Discuss. Pap. 664, 1983).
39. S. Resnick, C. Stărică, Smoothing the Hill estimator. *Adv. Appl. Probab.* **29**, 271–293 (1997).
40. J. Leskovec, D. Huttenlocher, J. Kleinberg, Data from "Wikipedia talk network." Stanford Network Analysis Project. https://snap.stanford.edu/data/wiki-Talk.html. Accessed 27 January 2025.
41. S. Cheng, L. Peng, Confidence intervals for the tail index. *Bernoulli* **7**, 751–760 (2001).
42. R. A. Rossi, N. K. Ahmed, Data from "rec-epinions-user-ratings." Network Repository. https://networkrepository.com/rec_epinions_user_ratings.php. Accessed 27 January 2025.
43. M. Richardson, R. Agrawal, P. Domingos, "Trust management for the semantic web" in *The Semantic Web – ISWC 2003*, D. Fensel, K. Sycara, J. Mylopoulos, Eds. (Springer, 2003), pp. 351–368.
44. P. Massa, P. Avesani, "Controversial users demand local trust metrics: An experimental study on epinions.com community" in *Proceedings of the Twentieth National Conference on Artificial Intelligence*, M. Veloso, S. Kambhampati, Eds. (AAAI Press, Menlo Park, CA, 2005), vol. 1, pp. 121–126.
45. S. Auer *et al.*, Data from "Person-country affiliations (DBpedia, 2016)." Netzschleuder. https://networks.skewed.de/net/dbpedia_country. Accessed 27 January 2025.
46. S. Auer *et al.*, "DBpedia: A nucleus for a web of open data" in *The Semantic Web*, K. Aberer et al., Eds. (Springer, 2007), pp. 722–735.
47. P. Hall, A. H. Welsh, Adaptive estimates of parameters of regular variation. *Ann. Stat.* **13**, 331–341 (1985).
48. H. Drees, E. Kaufmann, Selecting the optimal sample fraction in univariate extreme value estimation. *Stoch. Process. Appl.* **75**, 149–172 (1998).
49. N. L. Johnson, S. Kotz, N. Balakrishnan, *Continuous Univariate Distributions* (John Wiley & Sons, 1995), vol. 2.
50. C. Scarrott, A. MacDonald, A review of extreme value threshold estimation and uncertainty quantification. *Revstat. Stat. J.* **10**, 33–60 (2012).
51. S. Coles, *An Introduction to Statistical Modeling of Extreme Values, Springer Series in Statistics* (Springer, London, 2001).

Downloaded from https://www.pnas.org by Joel Cohen on October 4, 2026 from IP address 174.231.48.236.



Supporting Information for

Generalized Taylor's Law for Infinite-Mean Heavy-Tailed Data under Dependence and Heterogeneity

Pok Him Cheng, Joel E. Cohen, Hok Kan Ling, and Sheung Chi Phillip Yam

Joel E. Cohen
cohen@rockefeller.edu

This PDF file includes:

Supporting text
Figs. S1 to S31
SI References

Supporting Information Text

A. Proofs for Section 2

We first state two results that are well-known in probability theory.

Lemma A.1. *Let $Y \geq 0$ be a nonnegative random variable and $p > 0$. Then $\mathbb{E}(Y^p) = \int_0^\infty p y^{p-1} \mathbb{P}(Y > y) dy$.*

For a proof of Lemma A.1, see (1) or (2).

Proposition A.2 (Karamata's Theorem, e.g., Proposition 1.5.8 in (3))**.** *If the function l is slowly varying and l is locally bounded on $[c_0, \infty)$ and if $\alpha > -1$, then*

$$\int_{c_0}^{x} t^\alpha l(t) dt \sim \frac{x^{\alpha+1}}{\alpha+1} l(x).$$

Proof of Lemma 2.1. For notational simplicity, throughout this proof we write $\tilde{Y} = \tilde{Y}_n$.

(a) The survival function of $\tilde{Y}$ for any $x > 0$ is

$$\mathbb{P}(\tilde{Y} > x) = \mathbb{P}(x < Y \leq a_n) \mathbb{1}(x \leq a_n) = \left[\overline{F}(x) - \overline{F}(a_n)\right] \mathbb{1}(x \leq a_n).$$

Thus, by Lemma A.1,

$$\begin{aligned}\mathbb{E}[\tilde{Y}^p] &= p \int_0^\infty y^{p-1} \mathbb{P}(\tilde{Y} > y) dy = p \int_0^{a_n} (y^{p-1} \overline{F}(y) - y^{p-1} \overline{F}(a_n)) dy \\ &= p \int_0^{a_n} y^{p-1} \overline{F}(y) dy - a_n^p \overline{F}(a_n).\end{aligned}$$

Since $\overline{F}(y) = y^{-\alpha} l(y)$,

$$\mathbb{E}[\tilde{Y}^p] = p \int_0^{a_n} y^{p-1-\alpha} l(y) dy - a_n^{p-\alpha} l(a_n).$$

Fix $c_0 > 0$. By Proposition A.2, for $p - 1 - \alpha > -1$, or equivalently, $p > \alpha$,

$$\int_{c_0}^{a_n} y^{p-1-\alpha} l(y) dy \sim \frac{a_n^{p-\alpha} l(a_n)}{p-\alpha}, \quad \text{as } n \to \infty.$$

Since

$$\int_{c_0}^{a_n} y^{p-1-\alpha} l(y) dy \leq \int_0^{a_n} y^{p-1-\alpha} l(y) dy = \int_0^{c_0} y^{p-1-\alpha} l(y) dy + \int_{c_0}^{a_n} y^{p-1-\alpha} l(y) dy,$$

and

$$\int_0^{c_0} y^{p-1-\alpha} l(y) dy = \int_0^{c_0} y^{p-1} \overline{F}(y) dy \leq \int_0^{c_0} y^{p-1} dy = \frac{c_0^p}{p}.$$

Since $a_n^{p-\alpha} l(a_n) \to \infty$ because $a_n^{p-\alpha} l(a_n)$ is regularly varying with positive index $p - \alpha$, we have

$$\int_0^{a_n} y^{p-1-\alpha} l(y) dy \sim \frac{a_n^{p-\alpha} l(a_n)}{p-\alpha}, \quad \text{as } n \to \infty.$$

Thus, as $n \to \infty$,

$$\frac{\mathbb{E}(\tilde{Y}^p)}{a_n^{p-\alpha} l(a_n)} \to \frac{p}{p-\alpha} - 1 = \frac{\alpha}{p-\alpha}.$$

(b) Part (a) implies that

$$\log \mathbb{E}[\tilde{Y}^p] - \log \left[\frac{\alpha}{p-\alpha} a_n^{p-\alpha} l(a_n)\right] \to 0, \quad \text{as } n \to \infty.$$

Thus, as $n \to \infty$,

$$\frac{\log \mathbb{E}[\tilde{Y}^p]}{\log a_n} \to p - \alpha,$$

as $\frac{\log l(a_n)}{\log a_n} \to 0$; also see Proposition 1.3.6 (i) in (3). □

Proof of Lemma 2.2. We first consider the case when l is continuous. Let

$$g(x) := \frac{x^\alpha}{l(x)}.$$

Since l is slowly varying and positive, $g(x) \to \infty$ as $x \to \infty$. Fix $a > 0$. Since $c_n \to \infty$, for each sufficiently large n, we have

$$g(a) < nc_n.$$

Since $g(x) \to \infty$, there exists $b > a$ such that

$$g(b) > nc_n.$$

By continuity, there exists $t_n \in (a, b)$ such that $g(t_n) = nc_n$. Equivalently,

$$\frac{nl(t_n)}{t_n^\alpha} = \frac{1}{c_n}.$$

Moreover, $t_n \to \infty$. Indeed, if a subsequence of t_n were bounded, then, since $t_n \geq a$, the corresponding values $g(t_n)$ would be bounded, contradicting $g(t_n) = nc_n \to \infty$.

The proof for v_n is identical, using $g(v_n) = n/c_n$. Since $\log c_n = o(\log n)$, we have $c_n = n^{o(1)}$, and hence $n/c_n \to \infty$. Thus, for all sufficiently large n, there exists $v_n \to \infty$ such that

$$g(v_n) = \frac{n}{c_n}.$$

Equivalently,

$$\frac{nl(v_n)}{v_n^\alpha} = c_n.$$

Now consider a general positive slowly varying function l. By Proposition 1.3.4 in (3), there exists a continuous positive slowly varying function l_0 such that

$$l(x) \sim l_0(x), \qquad x \to \infty.$$

Applying the continuous case to l_0, there exist sequences $t_n \to \infty$ and $v_n \to \infty$ such that

$$\frac{nl_0(t_n)}{t_n^\alpha} = \frac{1}{c_n}$$

and

$$\frac{nl_0(v_n)}{v_n^\alpha} = c_n.$$

Since $t_n \to \infty$ and $v_n \to \infty$, we have

$$\frac{l(t_n)}{l_0(t_n)} \to 1, \qquad \frac{l(v_n)}{l_0(v_n)} \to 1.$$

Therefore,

$$\frac{nl(t_n)}{t_n^\alpha} = \frac{nl_0(t_n)}{t_n^\alpha}\frac{l(t_n)}{l_0(t_n)} \sim \frac{1}{c_n},$$

and similarly,

$$\frac{nl(v_n)}{v_n^\alpha} = \frac{nl_0(v_n)}{v_n^\alpha}\frac{l(v_n)}{l_0(v_n)} \sim c_n.$$

It remains to prove the logarithmic convergence statements in Eq. (6). From Eq. (4), we have

$$\alpha \log t_n = \log n + \log c_n + \log l(t_n) + o(1).$$

Since $t_n \to \infty$ and l is slowly varying, Proposition 1.3.6 (i) in (3) gives

$$a_n := \frac{\log l(t_n)}{\log t_n} \to 0.$$

Then

$$\alpha \log t_n = \log n + \log c_n + a_n \log t_n + o(1).$$

Dividing by $\log n$, we obtain

$$(\alpha - a_n)\frac{\log t_n}{\log n} = 1 + \frac{\log c_n}{\log n} + o(1).$$

Since $a_n \to 0$ and $\log c_n = o(\log n)$, it follows that

$$\frac{\log t_n}{\log n} \to \frac{1}{\alpha}.$$

The proof for the statement for v_n is similar and is therefore omitted. □

Proof of Theorem 2.3. When the random variables are i.i.d., Condition A(p) is satisfied for any $p > \alpha$. Indeed, $\sum_{i \neq j} \mathrm{Cov}(\breve{Y}_i^p, \breve{Y}_j^p) = 0$. Therefore, the claims follow from Theorem 3.2 and Theorem 3.10. □

B. Proofs and examples for Section 3

B.1. Proofs for Subsection 3.2. We introduce Lemmas B.1–B.3 before proving Lemma 3.1 and Theorem 3.2. Recall that t_n satisfies Eq. (4) and that c_n satisfies $c_n \to \infty$ and $\log c_n = o(\log n)$.

For $i = 1, \ldots, n$, let $\tilde{Y}_i := Y_i \mathbb{1}(Y_i \le t_n)$. For $p > \alpha$, denote the pth truncated sample moment $\tilde{M}_{n,p} := \frac{1}{n}\sum_{i=1}^n \tilde{Y}_i^p$ and its corresponding population moment $d_{n,p} := \mathbb{E}(\tilde{Y}_1^p)$. The following lemma provides an upper bound for the ratio $M_{n,p}/d_{n,p}$ by truncating Y_i at t_n.

Lemma B.1 (Upper bound for $M_{n,p}/d_{n,p}$)**.** *Let $Y_1, \ldots, Y_n, \ldots$ be a sequence of nonnegative random variables with a common survival function $\overline{F}(x) = x^{-\alpha} l(x)$, where l is slowly varying and $\alpha > 0$ (not necessarily in $(0,1)$). For $p > \alpha$, we have*

$$\frac{M_{n,p}}{d_{n,p}} = O_p(1).$$

Proof of Lemma B.1. Let $\varepsilon > 0$. Choose a constant $C > 0$ such that $1/C < \varepsilon/2$. By Eq. (4), there exists N such that for all $n \ge N$, $nl(t_n)/t_n^\alpha < \varepsilon/2$. Then

$$\begin{aligned}\mathbb{P}\left(\frac{M_{n,p}}{d_{n,p}} > C\right) &= \mathbb{P}\left(\frac{M_{n,p}}{d_{n,p}} > C, \tilde{M}_{n,p} = M_{n,p}\right) + \mathbb{P}\left(\frac{M_{n,p}}{d_{n,p}} > C, \tilde{M}_{n,p} \ne M_{n,p}\right) \\ &\le \mathbb{P}\left(\frac{\tilde{M}_{n,p}}{d_{n,p}} > C\right) + \mathbb{P}(\tilde{M}_{n,p} \ne M_{n,p}). \end{aligned} \tag{S.1}$$

For the first term in Eq. (S.1), since $\tilde{M}_{n,p}/d_{n,p}$ is non-negative, by Markov's inequality,

$$\mathbb{P}\left(\frac{\tilde{M}_{n,p}}{d_{n,p}} > C\right) \le \frac{\mathbb{E}(\tilde{M}_{n,p}/d_{n,p})}{C} = \frac{1}{C} < \frac{\varepsilon}{2}.$$

For the second term in Eq. (S.1),

$$\mathbb{P}(\tilde{M}_{n,p} \ne M_{n,p}) \le \mathbb{P}\left(\cup_{i=1}^n \{Y_i > t_n\}\right) \le n\mathbb{P}(Y > t_n) = \frac{nl(t_n)}{t_n^\alpha} < \frac{\varepsilon}{2}. \qquad \square$$

Next, Lemma B.2 provides a lower bound for $M_{n,p}/d_{n,p}$. Consider the truncated term $\breve{Y}_i := Y_i \mathbb{1}(Y_i \le v_n)$. Recall that v_n satisfies Eq. (5). Define $\breve{M}_{n,p} := \frac{1}{n}\sum_{i=1}^n \breve{Y}_i^p$ and $\breve{d}_{n,p} := \mathbb{E}(\breve{Y}_1^p)$.

Lemma B.2 (Lower Bound for $M_{n,p}/d_{n,p}$)**.** *Let $Y_1, \ldots, Y_n, \ldots$ be a sequence of nonnegative random variables with a common survival function $\overline{F}(x) = x^{-\alpha} l(x)$, where l is slowly varying and $\alpha > 0$ (not necessarily in $(0,1)$). If Condition A(p) holds for $p > \alpha$, then*

$$\frac{M_{n,p}}{d_{n,p}} \ge \frac{c_{p,\alpha,\delta'}(1 + o_p(1))}{c_n^{\delta'}},$$

for any $\delta' > 2(p/\alpha - 1) > 0$ and $c_{p,\alpha,\delta'} > 0$ is some constant depending only on p, α, δ'.

Proof of Lemma B.2. Write

$$nM_{n,p} \ge (n\breve{M}_{n,p} - n\breve{d}_{n,p}) + n\breve{d}_{n,p}. \tag{S.2}$$

For any $p > \alpha$, by Lemma 2.1 and Eq. (5),

$$\frac{n\breve{d}_{n,p}}{v_n^p} \sim \frac{\frac{\alpha}{p-\alpha} n v_n^{p-\alpha} l(v_n)}{v_n^p} \sim c_n \frac{\alpha}{p-\alpha}. \tag{S.3}$$

For any $x > 0$, by Markov's inequality,

$$\begin{aligned}\mathbb{P}\left(\frac{|n\breve{M}_{n,p} - n\breve{d}_{n,p}|}{v_n^p c_n} > x\right) &\le \frac{\mathrm{Var}(n\breve{M}_{n,p} - n\breve{d}_{n,p})}{v_n^{2p} c_n^2 x^2} \\ &\le \frac{n^2 \,\mathrm{Var}(\breve{M}_{n,p})}{v_n^{2p} c_n^2 x^2} \\ &\le \frac{n\breve{d}_{n,2p} + \sum_{i \ne j} \mathrm{Cov}(\breve{Y}_i^p, \breve{Y}_j^p)}{v_n^{2p} c_n^2 x^2} \to 0, \end{aligned} \tag{S.4}$$

where the last convergence follows from Eq. (S.3), with p replaced by $2p$, so that $n\breve{d}_{n,2p}/v_n^{2p} c_n^2 x^2 \to 0$, the fact that $c_n \to \infty$ and Eq. (8). In view of Eq. (S.2)–Eq. (S.4),

$$\frac{nM_{n,p}}{v_n^p c_n} \ge o_p(1) + \frac{\alpha}{p-\alpha}.$$

Then

$$\frac{M_{n,p}}{d_{n,p}} = \frac{nM_{n,p}}{v_n^p c_n} \cdot \frac{v_n^p c_n}{nd_{n,p}} \geq \left(o_p(1) + \frac{\alpha}{p-\alpha} \right) \frac{v_n^p c_n}{nd_{n,p}}.$$

By Lemma 2.1,

$$\frac{v_n^p c_n}{nd_{n,p}} \sim \frac{v_n^p c_n}{nt_n^{p-\alpha} l(t_n) \frac{\alpha}{p-\alpha}} = \frac{p-\alpha}{\alpha} \frac{t_n^\alpha}{nl(t_n)} c_n \left(\frac{v_n}{t_n} \right)^p \sim \frac{p-\alpha}{\alpha} c_n^2 \left(\frac{v_n}{t_n} \right)^p.$$

Therefore,

$$\frac{M_{n,p}}{d_{n,p}} \geq (1 + o_p(1)) c_n^2 \left(\frac{v_n}{t_n} \right)^p.$$

The preceding lower bound together with $M_{n,p}/d_{n,p} = O_p(1)$ in Lemma B.1 imply that

$$c_n^2 \left(\frac{v_n}{t_n} \right)^p = O(1).$$

Indeed, otherwise a subsequence along which $c_n^2 (v_n/t_n)^p \to \infty$ would force $M_{n,p}/d_{n,p} \to \infty$ in probability, contradicting Lemma B.1. Hence

$$\frac{v_n}{t_n} = O(c_n^{-2/p}) = o(1).$$

In particular, $v_n \leq t_n$ for all sufficiently large n.

By the definitions of v_n and t_n,

$$\left(\frac{v_n}{t_n} \right)^p \sim \left(\frac{nl(v_n)}{c_n} \cdot \frac{1}{nc_n l(t_n)} \right)^{p/\alpha} = \left(\frac{1}{c_n^2} \frac{l(v_n)}{l(t_n)} \right)^{p/\alpha}. \tag{S.5}$$

By Potter's Theorem (see Theorem 1.5.6 in (3)), for any $0 < \delta < \alpha$, there exists N such that for all $n \geq N$,

$$\frac{l(t_n)}{l(v_n)} \leq 2 \max \left\{ \left(\frac{t_n}{v_n} \right)^\delta, \left(\frac{t_n}{v_n} \right)^{-\delta} \right\}.$$

Since $v_n \leq t_n$ for all sufficiently large n, we have

$$\frac{l(t_n)}{l(v_n)} \leq 2 \left(\frac{t_n}{v_n} \right)^\delta. \tag{S.6}$$

Then Eq. (S.5) and Eq. (S.6) give

$$\left(\frac{v_n}{t_n} \right)^p \geq (1 + o(1)) \left(\frac{1}{c_n^2} \right)^{p/\alpha} \cdot \frac{1}{2^{p/\alpha}} \left(\frac{v_n}{t_n} \right)^{\delta p/\alpha}.$$

Therefore

$$\left(\frac{v_n}{t_n} \right)^{p(1-\delta/\alpha)} \geq (1 + o(1)) \left(\frac{1}{2c_n^2} \right)^{p/\alpha}.$$

Thus

$$c_n^2 \left(\frac{v_n}{t_n} \right)^p \geq (1 + o(1)) c_n^2 \left(\frac{1}{2c_n^2} \right)^{\frac{p}{\alpha-\delta}} = \frac{1 + o(1)}{2^{\frac{p}{\alpha-\delta}}} \left(\frac{1}{c_n^2} \right)^{\frac{p}{\alpha-\delta} - 1}.$$

The claim follows as given any $\delta' > 2(p/\alpha - 1)$, we can choose $0 < \delta < \alpha$ such that $2(p/(\alpha - \delta) - 1) = \delta'$, and the constant $c_{p,\alpha,\delta'}$ can be chosen as $\frac{1}{2^{p/(\alpha-\delta)}}$. □

Lemma B.3. *Under the conditions in Lemma B.2,*

$$\log \frac{M_{n,p}}{d_{n,p}} = O_p(\log c_n).$$

Proof of Lemma B.3. Fix $\varepsilon \in (0, 1)$. By Lemmas B.1 and B.2, with probability greater than $1 - \varepsilon$, for all large n, there exists some $C > 0$ such that

$$\frac{c_{p,\alpha,\delta'}}{2c_n^{\delta'}} \leq \frac{M_{n,p}}{d_{n,p}} \leq C.$$

Therefore, for all sufficiently large n,

$$\left| \log \frac{M_{n,p}}{d_{n,p}} \right| \leq |\log C| + |\log(c_{p,\alpha,\delta'}/2)| + \delta' \log c_n.$$ □

Proof of Lemma 3.1. We write:

$$\log \frac{M_{n,p}}{n^{\frac{p-\alpha}{\alpha}}} = \log \frac{M_{n,p}}{d_{n,p}} + \log \frac{d_{n,p}}{n^{\frac{p-\alpha}{\alpha}}}.$$

By Lemma 2.1 and Eq. (4),

$$\frac{d_{n,p}}{n^{\frac{p-\alpha}{\alpha}}} \sim \frac{\frac{\alpha}{p-\alpha} t_n^{p-\alpha} l(t_n)}{n^{\frac{p-\alpha}{\alpha}}} \sim \frac{\frac{\alpha}{p-\alpha} (nl(t_n)c_n)^{\frac{p-\alpha}{\alpha}} l(t_n)}{n^{\frac{p-\alpha}{\alpha}}} = \frac{\alpha}{p-\alpha} c_n^{\frac{p-\alpha}{\alpha}} l(t_n)^{\frac{p}{\alpha}}.$$

Thus, by Lemma B.3,

$$\frac{\log M_{n,p}}{\log n} - \frac{p-\alpha}{\alpha} = \frac{\log \frac{M_{n,p}}{d_{n,p}}}{\log n} + \frac{\log \frac{d_{n,p}}{n^{\frac{p-\alpha}{\alpha}}}}{\log n} = O_p\left(\frac{\log c_n}{\log n}\right) + O\left(\frac{|\log l(t_n)|}{\log n}\right) = O_p(\Delta_n).$$

Finally, Eq. (9) follows as

$$\frac{\log l(t_n)}{\log n} = \frac{\log l(t_n)}{\log t_n} \cdot \frac{\log t_n}{\log n} \to 0 \cdot \frac{1}{\alpha} = 0,$$

where the convergence follows from the fact that l is slowly varying, $t_n \to \infty$, and Eq. (6). □

Now we are ready to prove Theorem 3.2.

Proof of Theorem 3.2. It suffices to prove part (b) as Part (a) follows immediately because $\log c_n = o(\log n)$ and $\log l(t_n)/\log n \to 0$.

We write

$$\log \frac{M_{n,h_1}}{M_{n,h_2}^{\imath_\alpha(h_1,h_2)}} = \log \frac{M_{n,h_1}}{d_{n,h_1}} + \log \frac{d_{n,h_1}}{d_{n,h_2}^{\imath_\alpha(h_1,h_2)}} + \imath_\alpha(h_1,h_2) \log \frac{d_{n,h_2}}{M_{n,h_2}}. \tag{S.7}$$

By Lemma B.3,

$$\log \frac{M_{n,h_1}}{d_{n,h_1}} = O_p(\log c_n) \text{ and } \log \frac{d_{n,h_2}}{M_{n,h_2}} = O_p(\log c_n).$$

Also

$$\frac{d_{n,h_1}}{d_{n,h_2}^{\imath_\alpha(h_1,h_2)}} \sim \frac{\frac{\alpha}{h_1-\alpha} t_n^{h_1-\alpha} l(t_n)}{[\frac{\alpha}{h_2-\alpha} t_n^{h_2-\alpha} l(t_n)]^{\imath_\alpha(h_1,h_2)}} = \left(\frac{\alpha}{h_1-\alpha}\right)\left(\frac{\alpha}{h_2-\alpha}\right)^{-\imath_\alpha(h_1,h_2)} l(t_n)^{(h_2-h_1)/(h_2-\alpha)}.$$

Thus

$$\log \frac{d_{n,h_1}}{d_{n,h_2}^{\imath_\alpha(h_1,h_2)}} = o(1) + C_1 + C_2 \log l(t_n), \tag{S.8}$$

where C_1, C_2 are some constants depending only on α, h_1, h_2. Combining Eq. (S.7)–Eq. (S.8), we obtain

$$\log \frac{M_{n,h_1}}{M_{n,h_2}^{\imath_\alpha(h_1,h_2)}} = \log M_{n,h_1} - \imath_\alpha(h_1,h_2) \log M_{n,h_2} = O_p(\log c_n) + O(|\log l(t_n)|). \tag{S.9}$$

Dividing both sides of Eq. (S.9) by $\log M_{n,h_2}$ gives

$$\frac{\log M_{n,h_1}}{\log M_{n,h_2}} - \imath_\alpha(h_1,h_2) = O_p\left(\frac{\log c_n}{\log M_{n,h_2}}\right) + O\left(\frac{|\log l(t_n)|}{\log M_{n,h_2}}\right).$$

The claim in Eq. (10) then follows since $\log M_{n,h_2}/\log n \xrightarrow{\mathbb{P}} \frac{h_2-\alpha}{\alpha}$ as shown in Lemma 3.1. □

Proof of Corollary 3.4. Fix $p \in \mathbb{N}$. Since $X_i = Y_i + x_F$, the binomial expansion gives

$$M_{n,p}^X = M_{n,p} + \sum_{r=0}^{p-1} \binom{p}{r} x_F^{p-r} M_{n,r},$$

where $M_{n,0} = 1$. We first show that $M_{n,r}/M_{n,p} \xrightarrow{\mathbb{P}} 0$ for every $0 \le r < p$. Indeed, for any $a > 0$ and $y \ge 0$, $y^r \le a^r + a^{r-p} y^p$. Applying this inequality with $y = Y_i$, summing over $i = 1, \ldots, n$, and rearranging the terms, we obtain

$$\frac{M_{n,r}}{M_{n,p}} \le \frac{a^r}{M_{n,p}} + a^{r-p}.$$

By Lemma 3.1, $M_{n,p} \to \infty$ in probability, since $p > \alpha$. Thus the first term on the right converges to 0 in probability for each fixed a, and then letting $a \to \infty$ gives $M_{n,r}/M_{n,p} \xrightarrow{\mathbb{P}} 0$.

It follows that

$$\frac{M_{n,p}^X}{M_{n,p}} = 1 + o_p(1), \qquad \log|M_{n,p}^X| = \log M_{n,p} + o_p(1).$$

Applying this relation with $p = h_1$ and $p = h_2$, write

$$\log|M_{n,h_j}^X| = \log M_{n,h_j} + \Delta_{n,j}, \qquad \Delta_{n,j} = o_p(1), \quad j = 1, 2.$$

Hence

$$\frac{\log|M_{n,h_1}^X|}{\log|M_{n,h_2}^X|} - \frac{\log M_{n,h_1}}{\log M_{n,h_2}} = \frac{\log M_{n,h_1} + \Delta_{n,1}}{\log M_{n,h_2} + \Delta_{n,2}} - \frac{\log M_{n,h_1}}{\log M_{n,h_2}} = \frac{\Delta_{n,1}\log M_{n,h_2} - \Delta_{n,2}\log M_{n,h_1}}{\log M_{n,h_2}\big(\log M_{n,h_2} + \Delta_{n,2}\big)} = o_p\left(\frac{1}{\log n}\right),$$

where the last step follows from Lemma 3.1. Thus

$$\frac{\log|M_{n,h_1}^X|}{\log|M_{n,h_2}^X|} - \imath_\alpha(h_1, h_2) = R_{n,h_1,h_2}^{\mathrm{hm}} + o_p\left(\frac{1}{\log n}\right).$$

The claimed upper bound of the rate follows from Theorem 3.2 and $c_n \to \infty$, since $o_p(1/\log n)$ is absorbed into $O_p(\log c_n/\log n)$. The probability limit follows immediately. □

B.2. Proofs for Subsection 3.3.

Proof of Theorem 3.5. Fix $p > \alpha$. Since $Y_i = X_i - x_F$ is a deterministic transformation of X_i, the sequence $\{Y_i\}$ is also m-dependent. Therefore, $\mathrm{Cov}(\breve{Y}_i^p, \breve{Y}_j^p) = 0$ whenever $|i-j| > m$. For any i, j,

$$\left|\mathrm{Cov}(\breve{Y}_i^p, \breve{Y}_j^p)\right| \le \sqrt{\mathrm{Var}(\breve{Y}_i^p)\,\mathrm{Var}(\breve{Y}_j^p)} = \mathrm{Var}(\breve{Y}_1^p) \le \mathbb{E}(\breve{Y}_1^{2p}).$$

Hence,

$$\frac{\left|\sum_{i\neq j}\mathrm{Cov}(\breve{Y}_i^p, \breve{Y}_j^p)\right|}{v_n^{2p}c_n^2} \le \frac{2nm\mathbb{E}(\breve{Y}_1^{2p})}{v_n^{2p}c_n^2} \sim \frac{2nm\frac{\alpha}{2p-\alpha}v_n^{2p-\alpha}l(v_n)}{v_n^{2p}c_n^2} = \frac{2m\alpha}{2p-\alpha}\frac{nl(v_n)}{v_n^\alpha c_n^2} \to 0,$$

where the asymptotic equivalence follows from Lemma 2.1 and the last convergence follows from Eq. (5) and $c_n \to \infty$. Therefore, Condition A(p) is satisfied. □

Proof of Theorem 3.6. Fix $p > \alpha$. Since $Y_i = X_i - x_F$ is a deterministic transformation of X_i, the sequence $\{Y_i : i \ge 1\}$ is also strongly mixing with the same strong mixing coefficient. We split the covariance sum into nearby and far-apart pairs:

$$\sum_{i\neq j}\mathrm{Cov}(\breve{Y}_i^p, \breve{Y}_j^p) = \sum_{0<|i-j|\le r_n}\mathrm{Cov}(\breve{Y}_i^p, \breve{Y}_j^p) + \sum_{|i-j|>r_n}\mathrm{Cov}(\breve{Y}_i^p, \breve{Y}_j^p).$$

We first control the nearby pairs. As in the proof of Theorem 3.5, we have

$$\left|\mathrm{Cov}(\breve{Y}_i^p, \breve{Y}_j^p)\right| \le \mathbb{E}(\breve{Y}_1^{2p}).$$

Since the number of ordered pairs satisfying $0 < |i-j| \le r_n$ is at most $2nr_n$, we have

$$\left|\sum_{0<|i-j|\le r_n}\mathrm{Cov}(\breve{Y}_i^p, \breve{Y}_j^p)\right| \le 2nr_n\,\mathbb{E}(\breve{Y}_1^{2p}).$$

By Lemma 2.1 and the choice of v_n, we obtain

$$\frac{2nr_n\,\mathbb{E}(\breve{Y}_1^{2p})}{v_n^{2p}c_n^2} \sim \frac{2\alpha}{2p-\alpha}\frac{r_n nl(v_n)}{v_n^\alpha c_n^2} \sim \frac{2\alpha}{2p-\alpha}\frac{r_n}{c_n} = o(1),$$

where the last equality follows from $r_n = o(c_n)$.

We next control the far-apart pairs. Since $\breve{Y}_i^p \le v_n^p$, Lemma 1.2 in (4) gives, for $i \neq j$,

$$\left|\mathrm{Cov}(\breve{Y}_i^p, \breve{Y}_j^p)\right| \le 4v_n^{2p}\alpha(|i-j|).$$

Hence

$$\left|\sum_{|i-j|>r_n}\mathrm{Cov}(\breve{Y}_i^p, \breve{Y}_j^p)\right| \le 4v_n^{2p}\sum_{|i-j|>r_n}\alpha(|i-j|) = 8v_n^{2p}\sum_{k=r_n+1}^{n-1}(n-k)\alpha(k) \le 8nv_n^{2p}\sum_{k=r_n+1}^{n-1}\alpha(k).$$

Therefore, by Eq. (11),

$$\frac{\left|\sum_{|i-j|>r_n} \mathrm{Cov}(\breve{Y}_i^p, \breve{Y}_j^p)\right|}{v_n^{2p} c_n^2} \le \frac{8n}{c_n^2} \sum_{k=r_n+1}^{n-1} \alpha(k) \to 0.$$

Combining the nearby and far-apart bounds gives

$$\sum_{i \neq j} \mathrm{Cov}(\breve{Y}_i^p, \breve{Y}_j^p) = o(v_n^{2p} c_n^2).$$

Thus Condition A(p) holds. □

Proof of Corollary 3.7. Choose a constant $K > 1/|\log \rho|$ and set $r_n := \lceil K \log n \rceil$. Then $r_n \to \infty$, and since $c_n / \log n \to \infty$, we have $r_n / c_n \to 0$. It remains to verify Eq. (11). Since $0 < \rho < 1$,

$$\sum_{k=r_n+1}^{n-1} \alpha(k) \le C \sum_{k=r_n+1}^{\infty} \rho^k = \frac{C\rho^{r_n+1}}{1-\rho} \le C' \rho^{r_n} \le C' n^{-K|\log \rho|},$$

where $C' > 0$ is a constant. Therefore,

$$\frac{n}{c_n^2} \sum_{k=r_n+1}^{n-1} \alpha(k) \le C' \frac{n^{1-K|\log \rho|}}{c_n^2} \to 0,$$

because $K|\log \rho| > 1$ and $c_n \to \infty$. The conclusion follows from Theorem 3.6. □

Proof of Corollary 3.8. By (5), for any two σ-fields $\mathcal{A}, \mathcal{B} \subset \mathcal{F}$,

$$2\alpha(\mathcal{A}, \mathcal{B}) \le \beta(\mathcal{A}, \mathcal{B}) \le \phi(\mathcal{A}, \mathcal{B}) \le \frac{1}{2}\psi(\mathcal{A}, \mathcal{B}),$$
$$4\alpha(\mathcal{A}, \mathcal{B}) \le \rho(\mathcal{A}, \mathcal{B}) \le \psi(\mathcal{A}, \mathcal{B}).$$

Hence the same inequalities hold for the corresponding sequence coefficients. Thus each of the four assumptions implies strong mixing and, for some constant $C < \infty$ depending only on the type of mixing, $\alpha(k) \le C\eta(k)$, where η denotes the corresponding ψ-, ϕ-, ρ-, or β-mixing coefficient. Therefore

$$\frac{n}{c_n^2} \sum_{k=r_n+1}^{n-1} \alpha(k) \le C \frac{n}{c_n^2} \sum_{k=r_n+1}^{n-1} \eta(k) \to 0.$$

Theorem 3.6 now implies Condition A(p). □

Proof of Theorem 3.9. Choose $q \in (0, \alpha)$. Since ε_1 is regularly varying with tail index α, we have $\mathbb{E}|\varepsilon_1|^q < \infty$. Since $\beta_1 \in (0, 1)$, the strictly stationary solution of the AR(1) recursion admits the causal representation

$$Y_t = \sum_{j=0}^{\infty} \beta_1^j \varepsilon_{t-j}, \qquad t \in \mathbb{Z}.$$

By the tail asymptotics for infinite weighted sums of i.i.d. regularly varying random variables (6, Lemma 3.4), applied with $\psi_j = \beta_1^j$ and $Z_j = \varepsilon_{t-j}$, we know that Y_t is regularly varying with tail index α.

Under the conditions for ε_t, by Theorems 1 and 2 and the remark following Theorem 2 in (7), $\{Y_t\}$ is a strong mixing process, and there exist constants $C_0 > 0$ and an integer $s_0 \ge 1$ such that the strong mixing coefficient of $\{Y_t\}$ satisfies

$$\alpha(k) \le C_0 \mathbb{E}(Y_1^q)(\beta_1^q)^k, \qquad k \ge s_0,$$

where $\mathbb{E}(Y_1^q) < \infty$. Since $\alpha(k) \le 1$ for $1 \le k < s_0$, after increasing the constant if necessary, there exist constants $C > 0$ and $\rho \in (0, 1)$ such that

$$\alpha(k) \le C\rho^k, \qquad k \ge 1.$$

Since $c_n / \log n \to \infty$, Corollary 3.7 implies that Condition $A(p)$ holds for every $p > \alpha$. □

B.3. Proofs for Subsection 3.4. We will use Lemma B.4(b) in the proof of Lemma B.5 and Lemma B.4(a) in the proof of Lemma G.1.

Lemma B.4. *Let $\alpha \in (0,1)$ and suppose that $Y_i = X_i - x_F$ have common survival function $\overline{F}(y) = y^{-\alpha} l(y)$, where l is slowly varying. Then the following statements hold.*

(a) For every $h > 1$,

$$\frac{(M_{n,1}^Y)^h}{d_{n,h}} = O_p\left(\frac{1}{(nc_n)^{h-1}}\right).$$

(b) For every integer $k \geq 2$,

$$R_{n,k} := \sum_{j=0}^{k-1} \binom{k}{j} M_{n,j}^Y (M_{n,1}^Y)^{k-j},$$

where $M_{n,0}^Y := 1$, satisfies

$$\frac{R_{n,k}}{d_{n,k}} = O_p\left(\frac{1}{nc_n}\right).$$

Proof of Lemma B.4. (a) By Lemma 2.1, for every $p > \alpha$,

$$d_{n,p} \sim \frac{\alpha}{p-\alpha} t_n^{p-\alpha} l(t_n).$$

Since $t_n^\alpha \sim nc_n l(t_n)$, for every $h > \alpha$,

$$\frac{d_{n,1}^h}{d_{n,h}} \sim C_{h,\alpha} \frac{t_n^{h(1-\alpha)} l(t_n)^h}{t_n^{h-\alpha} l(t_n)} = C_{h,\alpha} \left(\frac{l(t_n)}{t_n^\alpha}\right)^{h-1} \sim C_{h,\alpha} \frac{1}{(nc_n)^{h-1}}, \tag{S.10}$$

where $C_{h,\alpha}$ is a positive constant depending only on h and α. If $h > 1$, by Lemma B.1, $M_{n,1}^Y / d_{n,1} = O_p(1)$. Hence

$$\frac{(M_{n,1}^Y)^h}{d_{n,h}} = \left(\frac{M_{n,1}^Y}{d_{n,1}}\right)^h \frac{d_{n,1}^h}{d_{n,h}} = O_p\left(\frac{1}{(nc_n)^{h-1}}\right).$$

(b) The term in $R_{n,k}$ corresponding to $j = 0$ is handled by part (a) with $h = k$:

$$\frac{(M_{n,1}^Y)^k}{d_{n,k}} = O_p\left(\frac{1}{(nc_n)^{k-1}}\right) = O_p\left(\frac{1}{nc_n}\right).$$

For $1 \leq j \leq k-1$, since $\alpha \in (0,1)$, both $j > \alpha$ and $1 > \alpha$. Therefore Lemma B.1 gives

$$\frac{M_{n,j}^Y}{d_{n,j}} = O_p(1), \qquad \frac{M_{n,1}^Y}{d_{n,1}} = O_p(1).$$

Moreover, by Lemma 2.1 and $t_n^\alpha \sim nc_n l(t_n)$,

$$\begin{aligned}\frac{d_{n,j} d_{n,1}^{k-j}}{d_{n,k}} &\sim C_{j,k,\alpha} \frac{t_n^{j-\alpha} l(t_n) t_n^{(1-\alpha)(k-j)} l(t_n)^{k-j}}{t_n^{k-\alpha} l(t_n)} \\ &= C_{j,k,\alpha} \left(\frac{l(t_n)}{t_n^\alpha}\right)^{k-j} = O\left(\frac{1}{(nc_n)^{k-j}}\right),\end{aligned}$$

where $C_{j,k,\alpha}$ is a positive constant depending only on j, k and α. Hence, for $1 \leq j \leq k-1$,

$$\frac{M_{n,j}^Y (M_{n,1}^Y)^{k-j}}{d_{n,k}} = \frac{M_{n,j}^Y}{d_{n,j}} \left(\frac{M_{n,1}^Y}{d_{n,1}}\right)^{k-j} \frac{d_{n,j} d_{n,1}^{k-j}}{d_{n,k}} = O_p\left(\frac{1}{(nc_n)^{k-j}}\right).$$

Since $j \leq k-1$, we have $k - j \geq 1$, and therefore

$$O_p\left(\frac{1}{(nc_n)^{k-j}}\right) = O_p\left(\frac{1}{nc_n}\right).$$

Summing over the finitely many indices $j = 0, \ldots, k-1$ gives

$$\frac{R_{n,k}}{d_{n,k}} = O_p\left(\frac{1}{nc_n}\right).$$

□

To prove Theorem 3.10, we introduce Lemma B.5, which is analogous to Lemma B.3.

Lemma B.5. *Let $\alpha \in (0,1)$ and let $k \geq 2$ be an integer. Suppose that $Y_i = X_i - x_F$ have common survival function $\overline{F}(y) = y^{-\alpha} l(y)$, where l is slowly varying. If Condition A(k) holds, then*

$$\log \frac{|M_{n,k}^{c,X}|}{d_{n,k}} = O_p(\log c_n).$$

Proof of Lemma B.5. Fix $k \geq 2$. Since central moments are invariant under shifts,

$$M_{n,k}^{c,X} = M_{n,k}^{c,Y} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - M_{n,1}^{Y})^k.$$

Expanding the right-hand side gives

$$M_{n,k}^{c,X} = M_{n,k}^{Y} + \sum_{j=0}^{k-1}\binom{k}{j}(-1)^{k-j} M_{n,j}^{Y}(M_{n,1}^{Y})^{k-j},$$

where $M_{n,0}^{Y} := 1$. Hence

$$|M_{n,k}^{c,X}| \leq M_{n,k}^{Y} + R_{n,k}, \tag{S.11}$$

where

$$R_{n,k} := \sum_{j=0}^{k-1}\binom{k}{j} M_{n,j}^{Y}(M_{n,1}^{Y})^{k-j}.$$

In view of Eq. (S.11), Lemma B.4, and Lemma B.1, we obtain

$$\frac{|M_{n,k}^{c,X}|}{d_{n,k}} = O_p(1). \tag{S.12}$$

It remains to obtain a lower bound. From the same expansion, we have

$$|M_{n,k}^{c,X}| \geq M_{n,k}^{Y} - R_{n,k}.$$

Hence, by Lemma B.2 and Lemma B.4, for any fixed $\delta' > 2(k/\alpha - 1)$,

$$\frac{|M_{n,k}^{c,X}|}{d_{n,k}} \geq \frac{c_{k,\alpha,\delta'}(1 + o_p(1))}{c_n^{\delta'}} - O_p\left(\frac{1}{nc_n}\right).$$

Since $\log c_n = o(\log n)$, we have $1/(nc_n) = o(c_n^{-\delta'})$. Thus, with probability tending to one,

$$\frac{|M_{n,k}^{c,X}|}{d_{n,k}} \geq \frac{c_{k,\alpha,\delta'}}{2c_n^{\delta'}}. \tag{S.13}$$

The upper bound Eq. (S.12) and the lower bound Eq. (S.13) imply

$$\log \frac{|M_{n,k}^{c,X}|}{d_{n,k}} = O_p(\log c_n). \qquad \square$$

Proof of Theorem 3.10. (a) Since Condition A(1) holds and $\alpha \in (0,1)$, Lemma B.2 with $p = 1$ implies $M_{n,1}^{Y} \to \infty$ in probability. Since $M_{n,1}^{X} = M_{n,1}^{Y} + x_F$, we have $|M_{n,1}^{X}|/M_{n,1}^{Y} \xrightarrow{\mathbb{P}} 1$. Thus, by Lemma B.3

$$\log \frac{|M_{n,1}^{X}|}{d_{n,1}} = \log \frac{|M_{n,1}^{X}|}{M_{n,1}^{Y}} + \log \frac{M_{n,1}^{Y}}{d_{n,1}} = O_p(\log c_n). \tag{S.14}$$

For $k \geq 2$, write

$$\log \frac{|M_{n,k}^{c,X}|}{|M_{n,1}^{X}|^{\imath_\alpha(k,1)}} = \log \frac{|M_{n,k}^{c,X}|}{d_{n,k}} + \log \frac{d_{n,k}}{d_{n,1}^{\imath_\alpha(k,1)}} + \imath_\alpha(k,1) \log \frac{d_{n,1}}{|M_{n,1}^{X}|}$$

With Lemma B.5 and Eq. (S.14), the rest of the proof follows the same argument as in the proof of Theorem 3.2 and we therefore omit the details.

(b) For $h_1, h_2 \geq 2$ with $h_1 \neq h_2$, write

$$\log \frac{|M_{n,h_1}^{c,X}|}{|M_{n,h_2}^{c,X}|^{\imath_\alpha(h_1,h_2)}} = \log \frac{|M_{n,h_1}^{c,X}|}{d_{n,h_1}} + \log \frac{d_{n,h_1}}{d_{n,h_2}^{\imath_\alpha(h_1,h_2)}} + \imath_\alpha(h_1,h_2) \log \frac{d_{n,h_2}}{|M_{n,h_2}^{c,X}|}.$$

With Lemma B.5, the rest of the proof follows the same argument as in the proof of Theorem 3.2 and we therefore omit the details. $\square$

B.4. Proof for Subsection 3.5.

Proof of Theorem 3.12. It suffices to prove that, for each $p \in \{h_1, h_2\}$,

$$\log\left(\frac{M_{n,p}^{Y}}{d_{n,p}}\right) = O_p(\log c_n). \tag{S.15}$$

Indeed, once Eq. (S.15) is established, the rest of the proof is similar to the proof of Theorem 3.2.

Fix $p \in \{h_1, h_2\}$. The upper bound

$$\frac{M_{n,p}^{Y}}{d_{n,p}} = O_p(1) \tag{S.16}$$

follows from the same argument as in Lemma B.1, applied to the shifted variables Y_i.

It remains to prove the corresponding lower bound. Let

$$S_{n,p} := \sum_{i=1}^{n} \breve{Y}_i^p,$$

and define

$$\mu_{n,p}(\mathcal{G}) := \sum_{i=1}^{n} \mathbb{E}(\breve{Y}_i^p \mid \mathcal{G}).$$

We first show that

$$\frac{S_{n,p} - \mu_{n,p}(\mathcal{G})}{v_n^p c_n} = o_p(1). \tag{S.17}$$

For any $x > 0$, by Markov's inequality and the tower property,

$$\mathbb{P}\left(\frac{|S_{n,p} - \mu_{n,p}(\mathcal{G})|}{v_n^p c_n} > x\right) \le \frac{\mathbb{E}\{\mathrm{Var}(S_{n,p} \mid \mathcal{G})\}}{x^2 v_n^{2p} c_n^2}.$$

Moreover,

$$\mathrm{Var}(S_{n,p} \mid \mathcal{G}) \le \sum_{i=1}^{n} \mathbb{E}(\breve{Y}_i^{2p} \mid \mathcal{G}) + \left|\sum_{i \ne j} \mathrm{Cov}(\breve{Y}_i^p, \breve{Y}_j^p \mid \mathcal{G})\right|.$$

Taking expectations and using Lemma 2.1, we obtain

$$\begin{aligned}
\frac{\mathbb{E}\{\mathrm{Var}(S_{n,p} \mid \mathcal{G})\}}{v_n^{2p} c_n^2} &\le \frac{n\mathbb{E}(\breve{Y}_1^{2p})}{v_n^{2p} c_n^2} + \mathbb{E}\left(\frac{\left|\sum_{i \ne j} \mathrm{Cov}(\breve{Y}_i^p, \breve{Y}_j^p \mid \mathcal{G})\right|}{v_n^{2p} c_n^2}\right) \\
&= (1 + o(1))\frac{\alpha}{2p - \alpha}\frac{nl(v_n)}{v_n^{\alpha} c_n^2} + o(1) \\
&= O(c_n^{-1}) + o(1) = o(1).
\end{aligned}$$

where the last line uses Eq. (5) and Eq. (13). This proves Eq. (S.17).

Next, by Lemma 2.1 and Eq. (5),

$$n\breve{d}_{n,p} := n\mathbb{E}(\breve{Y}_1^p) \sim \frac{\alpha}{p - \alpha} v_n^p c_n. \tag{S.18}$$

By Eq. (14), for every $\varepsilon > 0$ there exists $\eta > 0$ such that, for all sufficiently large n,

$$\mathbb{P}\left(\mu_{n,p}(\mathcal{G}) \ge \eta n\breve{d}_{n,p}\right) \ge 1 - \varepsilon.$$

Combining this with Eq. (S.17) and Eq. (S.18), we have, for some constant $C_{\varepsilon,p} > 0$,

$$S_{n,p} \ge C_{\varepsilon,p} v_n^p c_n \tag{S.19}$$

with probability at least $1 - 2\varepsilon$ for all sufficiently large n.

Since $nM_{n,p}^{Y} \ge S_{n,p}$, Eq. (S.19) gives

$$\frac{M_{n,p}^{Y}}{d_{n,p}} \ge C_{\varepsilon,p}\frac{v_n^p c_n}{nd_{n,p}}$$

with probability at least $1 - 2\varepsilon$ for all sufficiently large n. By the same deterministic Potter-bound comparison used in the proof of Lemma B.2, for any $\delta' > 2(p/\alpha - 1)$ there is a constant $C_{p,\delta'} > 0$ such that

$$\frac{v_n^p c_n}{nd_{n,p}} \ge C_{p,\delta'} c_n^{-\delta'}$$

for all sufficiently large n. Therefore,

$$\frac{M_{n,p}^{Y}}{d_{n,p}} \ge C'_{\varepsilon,p,\delta'} c_n^{-\delta'} \tag{S.20}$$

with probability at least $1 - 2\varepsilon$ for all sufficiently large n.

Combining Eq. (S.16) and Eq. (S.20) gives Eq. (S.15). □

B.5. Examples for the conditionally weak-dependent extension. Three examples illustrate this framework. In the first two examples, the variables become conditionally independent after conditioning on a suitable σ-field; hence Eq. (13) follows immediately, while Eq. (14) follows from a direct conditional-tail calculation and Karamata's theorem. The third example is included to show that conditional independence is not required. We give the detailed verification for this third example; the verification for the first two examples is similar and simpler.

Example B.6. *This example is considered in (8, Section 3(c)). Let $N_0, N_1, \ldots$ be i.i.d. standard normal random variables and let $0 \le \rho < 1$. Let $Z_i := \sqrt{\rho}N_0 + \sqrt{1-\rho}N_i$ for $i = 1, \ldots, n$. Then each Z_i is a standard normal with $\mathrm{Cov}(Z_i, Z_j) = \rho$ when $i \ne j$. For $\alpha \in (0,1)$, define $Y_i := 1/(Z_i^2)^{1/(2\alpha)}$. Then Y_i is regularly varying. Conditional on $\mathcal{G} := \sigma(N_0)$, Z_i's are independent. Thus Y_i's are also independent conditional on $\sigma(N_0)$. Fix $x_F \in \mathbb{R}$, let $X_i = Y_i + x_F$. In this example, Conditions* Eq. (13) *and* Eq. (14) *hold, so our Taylor's law holds.*

Example B.7. *This example is considered in (9, Theorem 2). Let $\{G_i\}$ be a stationary Gaussian sequence with mean 0 and covariance function $\gamma(\cdot,\cdot)$, i.e., $\mathrm{Cov}(G_i, G_j) = \gamma(i,j)$. Let Z_i's be i.i.d. Pareto-distributed with tail index $\alpha \in (0,1)$. Let $Y_i := Z_i e^{G_i}$. Then Y_i are identically distributed and regularly varying with index α. Conditioning on $\mathcal{G} := \sigma(G_i, i \in \mathbb{N})$, Y_i's are independent. Fix $x_F \in \mathbb{R}$, let $X_i = Y_i + x_F$. In this example, Conditions* Eq. (13) *and* Eq. (14) *hold, so our Taylor's law holds.*

Example B.8. *Let $\alpha > 0$. Let $\{U_i : i \in \mathbb{Z}\}$ be i.i.d. nonnegative random variables with survival function*

$$\overline{H}(u) := \mathbb{P}(U_i > u) = (1+u)^{-\alpha}, \qquad u \ge 0.$$

Let Λ be independent of $\{U_i : i \in \mathbb{Z}\}$, with $\mathbb{P}(\Lambda = 1) = \mathbb{P}(\Lambda = 2) = 1/2$, and let $\mathcal{G} := \sigma(\Lambda)$. Fix $x_F \in \mathbb{R}$. Define

$$Z_i := \max(U_{i-1}, U_i), \qquad Y_i := \Lambda Z_i, \qquad X_i := x_F + Y_i.$$

Then X_i has finite lower endpoint x_F, and $Y_i = X_i - x_F$ is regularly varying with tail index α.

Conditional on $\mathcal{G}$, the sequence $\{Y_i\}$ is 1-dependent but not independent, since Y_i and Y_{i+1} both depend on the common innovation U_i. In the following, we verify that Eq. (13) *and* Eq. (14) *hold for every $p > \alpha$. Therefore Theorem 3.12 applies.*

Verification of Example B.8. First, we verify the tail behavior. Since

$$\mathbb{P}(U_i > u) = (1+u)^{-\alpha}, \qquad u \ge 0,$$

we have, for $z \ge 0$,

$$\begin{aligned}
\mathbb{P}(Z_i > z) &= 1 - \mathbb{P}(U_{i-1} \le z, U_i \le z) \\
&= 1 - \left(1 - (1+z)^{-\alpha}\right)^2 \\
&= 2(1+z)^{-\alpha} - (1+z)^{-2\alpha}.
\end{aligned}$$

Hence

$$\mathbb{P}(Z_i > z) \sim 2z^{-\alpha}, \qquad z \to \infty.$$

Since Λ is independent of $\{U_i : i \in \mathbb{Z}\}$ and takes only the values 1 and 2,

$$\begin{aligned}
\mathbb{P}(Y_i > x) &= \mathbb{E}\left[\mathbb{P}\left(Z_i > \frac{x}{\Lambda} \,\middle|\, \Lambda\right)\right] \\
&= \mathbb{E}\left[2\left(1 + \frac{x}{\Lambda}\right)^{-\alpha} - \left(1 + \frac{x}{\Lambda}\right)^{-2\alpha}\right] \\
&\sim 2\mathbb{E}(\Lambda^\alpha) x^{-\alpha}.
\end{aligned}$$

Thus Y_i is regularly varying with tail index α. In particular, if

$$l(x) := x^\alpha \mathbb{P}(Y_i > x),$$

then l is slowly varying and

$$l(x) \to 2\mathbb{E}(\Lambda^\alpha).$$

We now verify Eq. (13). Conditional on $\mathcal{G}$, the sequence $\{Y_i\}$ is 1-dependent because Y_i is a function of Λ, U_{i-1}, U_i. Therefore,

$$\mathrm{Cov}(\breve{Y}_i^p, \breve{Y}_j^p \mid \mathcal{G}) = 0 \qquad \text{whenever } |i-j| > 1.$$

For $|i-j| = 1$,

$$\begin{aligned}
\left|\mathrm{Cov}(\breve{Y}_i^p, \breve{Y}_j^p \mid \mathcal{G})\right| &\le \left\{\mathrm{Var}(\breve{Y}_i^p \mid \mathcal{G})\,\mathrm{Var}(\breve{Y}_j^p \mid \mathcal{G})\right\}^{1/2} \\
&\le \mathbb{E}(\breve{Y}_1^{2p} \mid \mathcal{G}),
\end{aligned}$$

where the last inequality uses the fact that, conditional on $\mathcal{G}$, the variables $\breve{Y}_i$ are identically distributed. There are at most $2n$ ordered pairs with $|i-j|=1$. Thus, by Lemma 2.1, the definition of v_n and the fact that $c_n \to 0$, we have

$$\mathbb{E}\left(\frac{\left|\sum_{i\neq j}\operatorname{Cov}(\breve{Y}_i^p, \breve{Y}_j^p \mid \mathcal{G})\right|}{v_n^{2p}c_n^2}\right) \leq \frac{2n\mathbb{E}(\breve{Y}_1^{2p})}{v_n^{2p}c_n^2} \sim \frac{2\alpha}{2p-\alpha}\frac{nl(v_n)}{v_n^{\alpha}c_n^2} \sim \frac{2\alpha}{2p-\alpha}\frac{1}{c_n} \to 0.$$

Hence Eq. (13) holds.

It remains to verify Eq. (14). Since the variables are conditionally identically distributed given $\mathcal{G}$,

$$\mu_{n,p}(\mathcal{G}) = n\mathbb{E}(\breve{Y}_1^p \mid \mathcal{G}).$$

Moreover,

$$\mathbb{E}(\breve{Y}_1^p \mid \mathcal{G}) = \Lambda^p \mathbb{E}\left[Z_1^p \mathbb{1}\left(Z_1 \leq \frac{v_n}{\Lambda}\right)\middle|\Lambda\right].$$

Since $\mathbb{P}(Z_1 > z) \sim 2z^{-\alpha}$, Karamata's theorem gives, for $p > \alpha$,

$$\mathbb{E}\left[Z_1^p \mathbb{1}\left(Z_1 \leq \frac{v_n}{\Lambda}\right)\middle|\Lambda\right] \sim \frac{2\alpha}{p-\alpha}\left(\frac{v_n}{\Lambda}\right)^{p-\alpha}.$$

Because Λ takes only the two values 1 and 2, the preceding asymptotic is uniform over the possible values of Λ. Hence

$$\mathbb{E}(\breve{Y}_1^p \mid \mathcal{G}) \sim \frac{2\alpha}{p-\alpha}\Lambda^{\alpha}v_n^{p-\alpha}.$$

On the other hand, applying Lemma 2.1 to Y_1 gives

$$\breve{d}_{n,p} = \mathbb{E}(\breve{Y}_1^p) \sim \frac{\alpha}{p-\alpha}v_n^{p-\alpha}l(v_n) \sim \frac{2\alpha}{p-\alpha}\mathbb{E}(\Lambda^{\alpha})v_n^{p-\alpha}.$$

Therefore,

$$\frac{\mu_{n,p}(\mathcal{G})}{n\breve{d}_{n,p}} = \frac{\mathbb{E}(\breve{Y}_1^p \mid \mathcal{G})}{\breve{d}_{n,p}} \xrightarrow{\mathbb{P}} \frac{\Lambda^{\alpha}}{\mathbb{E}(\Lambda^{\alpha})}.$$

Since $\Lambda \in \{1,2\}$,

$$\frac{\Lambda^{\alpha}}{\mathbb{E}(\Lambda^{\alpha})} \geq \frac{1}{\mathbb{E}(\Lambda^{\alpha})} > 0.$$

Thus, for every sufficiently small $\eta > 0$,

$$\mathbb{P}\left(\frac{\mu_{n,p}(\mathcal{G})}{n\breve{d}_{n,p}} < \eta\right) \to 0,$$

which implies

$$\lim_{\eta\downarrow 0}\limsup_{n\to\infty}\mathbb{P}\left(\mu_{n,p}(\mathcal{G}) < \eta n\breve{d}_{n,p}\right) = 0.$$

Therefore, Eq. (14) also holds. Since both conditions hold for every $p > \alpha$, Theorem 3.12 applies. □

C. Proofs for Section 4

C.1. Proofs for Subsection 4.1. To prove Theorem 4.2, we establish Lemmas C.1–C.3.

Lemma C.1. *Assume* Eq. (15), Eq. (16), Eq. (17), *and* Eq. (18). *For $p > \alpha$ and $a_n \to \infty$, define*

$$d_{r,p}(a_n) := \mathbb{E}\left[Y_r^p \mathbb{1}(Y_r \leq a_n)\right], \qquad r = 1, \ldots, J,$$

where Y_r denotes a generic shifted random variable from cluster r, and define

$$D_{n,p}(a_n) := \frac{1}{n}\sum_{r=1}^{J} n_{r,n}d_{r,p}(a_n).$$

Then

$$D_{n,p}(a_n) \sim \pi_1\frac{\alpha}{p-\alpha}a_n^{p-\alpha}l(a_n). \tag{S.21}$$

Consequently,

$$D_{n,p}(t_n) \sim \pi_1\frac{\alpha}{p-\alpha}t_n^{p-\alpha}l(t_n) \tag{S.22}$$

and

$$\frac{nD_{n,p}(v_n)}{v_n^p c_n} \sim \pi_1\frac{\alpha}{p-\alpha}. \tag{S.23}$$

Proof of Lemma C.1. For the dominant cluster, the same argument as in the proof of Lemma 2.1 gives

$$d_{1,p}(a_n) \sim \frac{\alpha}{p-\alpha} a_n^{p-\alpha} l(a_n).$$

For $r \geq 2$, we have

$$d_{r,p}(a_n) = o\left(a_n^{p-\alpha} l(a_n)\right). \tag{S.24}$$

Indeed, by Eq. (18), for every $\varepsilon > 0$, there exists $A < \infty$ such that

$$\overline{F}_r(y) \leq \varepsilon \overline{F}_1(y), \qquad y \geq A.$$

Thus

$$\begin{aligned} d_{r,p}(a_n) &\leq p \int_0^{a_n} y^{p-1} \overline{F}_r(y)\, dy \\ &\leq p \int_0^{A} y^{p-1} \overline{F}_r(y)\, dy + \varepsilon p \int_A^{a_n} y^{p-1} \overline{F}_1(y)\, dy. \end{aligned}$$

The first term is $O(1)$, and Karamata's theorem gives

$$\int_A^{a_n} y^{p-1} \overline{F}_1(y)\, dy \sim \frac{a_n^{p-\alpha} l(a_n)}{p-\alpha}.$$

Since $\varepsilon > 0$ is arbitrary, Eq. (S.24) follows. Combining Eq. (S.24) with Eq. (15) and Eq. (16) proves Eq. (S.21). Taking $a_n = t_n$ gives Eq. (S.22). Taking $a_n = v_n$ and using Eq. (19) gives Eq. (S.23). □

Lemma C.2. *Assume the conditions of Lemma C.1. Fix $p > \alpha$ and suppose that*

$$\sum_{i \neq j} \mathrm{Cov}\left(\breve{Y}_{n,i}^p, \breve{Y}_{n,j}^p\right) = o\left(v_n^{2p} c_n^2\right). \tag{S.25}$$

Let $D_{n,p} := D_{n,p}(t_n)$. Then

$$\log \frac{M_{n,p}^Y}{D_{n,p}} = O_p(\log c_n). \tag{S.26}$$

Proof of Lemma C.2. The upper bound follows by the same truncation argument used in Lemma B.1. Let

$$\widetilde{Y}_{n,i} := Y_{n,i} \mathbb{1}(Y_{n,i} \leq t_n), \qquad \widetilde{M}_{n,p}^Y := \frac{1}{n} \sum_{i=1}^n \widetilde{Y}_{n,i}^p.$$

For any $C > 0$,

$$\begin{aligned} \mathbb{P}\left(\frac{M_{n,p}^Y}{D_{n,p}} > C\right) &\leq \mathbb{P}\left(\frac{\widetilde{M}_{n,p}^Y}{D_{n,p}} > C\right) + \mathbb{P}\left(M_{n,p}^Y \neq \widetilde{M}_{n,p}^Y\right) \\ &\leq \frac{1}{C} + \sum_{r=1}^J n_{r,n} \overline{F}_r(t_n). \end{aligned}$$

By Eq. (17)-Eq. (19),

$$\sum_{r=1}^J n_{r,n} \overline{F}_r(t_n) = o(1).$$

Hence

$$\frac{M_{n,p}^Y}{D_{n,p}} = O_p(1). \tag{S.27}$$

The lower bound follows by the same argument used in Lemma B.2. Let

$$\breve{M}_{n,p}^Y := \frac{1}{n} \sum_{i=1}^n \breve{Y}_{n,i}^p.$$

By Markov's inequality,

$$\mathbb{P}\left(\frac{|n\breve{M}_{n,p}^Y - nD_{n,p}(v_n)|}{v_n^p c_n} > x\right) \leq \frac{\sum_{i=1}^n \mathbb{E}(\breve{Y}_{n,i}^{2p}) + \sum_{i \neq j} \mathrm{Cov}(\breve{Y}_{n,i}^p, \breve{Y}_{n,j}^p)}{v_n^{2p} c_n^2 x^2}. \tag{S.28}$$

By Lemma C.1 with $2p$ in place of p,

$$\frac{\sum_{i=1}^{n}\mathbb{E}(\breve{Y}_{n,i}^{2p})}{v_n^{2p}c_n^2} = O\left(\frac{1}{c_n}\right) = o(1),$$

and the covariance term is $o(1)$ after normalization by Eq. (S.25). Therefore, by Eq. (S.23) and Eq. (S.28),

$$\frac{n\breve{M}_{n,p}^{Y}}{v_n^p c_n} = \pi_1\frac{\alpha}{p-\alpha} + o_p(1). \tag{S.29}$$

Since $M_{n,p}^{Y} \geq \breve{M}_{n,p}^{Y}$, Eq. (S.29), Eq. (S.22), and Eq. (19) imply

$$\frac{M_{n,p}^{Y}}{D_{n,p}} \geq (1+o_p(1))c_n^2\left(\frac{v_n}{t_n}\right)^p.$$

The Potter-bound calculation used in Lemma B.2 gives, for every $\delta' > 2(p/\alpha - 1)$,

$$c_n^2\left(\frac{v_n}{t_n}\right)^p \geq C_{\alpha,p,\delta'}c_n^{-\delta'}$$

for some constant $C_{\alpha,p,\delta'} > 0$. Hence

$$\frac{M_{n,p}^{Y}}{D_{n,p}} \geq C_{\alpha,p,\delta'}c_n^{-\delta'}(1+o_p(1)). \tag{S.30}$$

Combining Eq. (S.27) and Eq. (S.30) proves Eq. (S.26). □

Lemma C.3. *Assume the conditions of Lemma C.1. Fix $p > \alpha$ and suppose that* Eq. (S.25) *holds. Then*

$$\frac{\log M_{n,p}^{Y}}{\log n} = \frac{p-\alpha}{\alpha} + O_p\left(\frac{\log c_n}{\log n}\right) + O\left(\frac{|\log l(t_n)|}{\log n}\right). \tag{S.31}$$

Proof of Lemma C.3. By Lemma C.2,

$$\log M_{n,p}^{Y} = \log D_{n,p} + O_p(\log c_n). \tag{S.32}$$

By Eq. (S.22),

$$\log D_{n,p} = (p-\alpha)\log t_n + \log l(t_n) + O(1).$$

Since $t_n^\alpha \sim nc_n l(t_n)$,

$$\frac{\log D_{n,p}}{\log n} = \frac{p-\alpha}{\alpha} + O\left(\frac{\log c_n}{\log n}\right) + O\left(\frac{|\log l(t_n)|}{\log n}\right). \tag{S.33}$$

Combining Eq. (S.32) and Eq. (S.33) proves Eq. (S.31). □

Proof of Theorem 4.2. The theorem assumes Eq. (20) for $p = h_1$ and $p = h_2$. Hence the fixed-p condition Eq. (S.25) holds for each $p = h_1, h_2$. Applying Lemma C.3 with $p = h_1$ and $p = h_2$ gives, for $j = 1, 2$,

$$\log M_{n,h_j}^{Y} = \frac{h_j-\alpha}{\alpha}\log n + O_p(\log c_n) + O(|\log l(t_n)|).$$

Therefore,

$$\log M_{n,h_1}^{Y} - \frac{h_1-\alpha}{h_2-\alpha}\log M_{n,h_2}^{Y} = O_p(\log c_n) + O(|\log l(t_n)|). \tag{S.34}$$

Moreover,

$$\frac{\log M_{n,h_2}^{Y}}{\log n} \xrightarrow{\mathbb{P}} \frac{h_2-\alpha}{\alpha}. \tag{S.35}$$

Dividing Eq. (S.34) by $\log M_{n,h_2}^{Y}$ and using Eq. (S.35) gives Eq. (22), and hence Eq. (21). □

C.2. Proofs for Subsection 4.2. In this section, we provide the details of Example 4.3. Fix $p > \alpha$ and $k \in [n]$. By Conditions (i) and (ii),

$$\begin{aligned}\left|\sum_{j:j\neq k}\operatorname{Cov}(\breve{Y}_k^p,\breve{Y}_j^p)\right| &\le \sum_{m=1}^{M}\sum_{j:(k,j)\in U_{n,m}}\left|\operatorname{Cov}(\breve{Y}_k^p,\breve{Y}_j^p)\right| \\ &\le \sum_{m=1}^{M} K^m \operatorname{Var}(\breve{Y}_k^p) \\ &\le C_1\mathbb{E}(\breve{Y}_k^{2p}),\end{aligned}$$

where $C_1 := \sum_{m=1}^{M} K^m$. The first inequality follows from Condition (i), which removes all terms with network distance larger than M. The second inequality follows from Condition (ii), since the number of vertices at distance m from a fixed vertex is at most K^m. Therefore,

$$\left|\sum_{i\neq j}\operatorname{Cov}(\breve{Y}_i^p,\breve{Y}_j^p)\right| \le nC_1\mathbb{E}(\breve{Y}_1^{2p}).$$

As a result,

$$\frac{\left|\sum_{i\neq j}\operatorname{Cov}(\breve{Y}_i^p,\breve{Y}_j^p)\right|}{v_n^{2p}c_n^2} \le \frac{C_1 n\mathbb{E}(\breve{Y}_1^{2p})}{v_n^{2p}c_n^2} \sim \frac{C_1 n\frac{\alpha}{2p-\alpha}v_n^{2p-\alpha}l(v_n)}{v_n^{2p}c_n^2} = \frac{C_1\alpha}{2p-\alpha}\frac{nl(v_n)}{v_n^{\alpha}c_n^2} \to 0,$$

where the asymptotic equivalence follows from Lemma 2.1 applied to Y, and the last convergence follows from Eq. (5) and $c_n \to \infty$. Therefore, Eq. (23) is satisfied.

D. Rate refinements under additional tail regularity

In addition to establishing the limit in probability, we may also obtain the rate of convergence of $\frac{\log M_{n,h_1}}{\log M_{n,h_2}} - \iota_\alpha(h_1,h_2)$ under some conditions on the slowly varying function l in Theorem 3.2. Let $\alpha > 0$. For some $\delta \in (0,\alpha)$, we define

$$\begin{aligned} l_{1n} &:= \left(2\max\{l(n),l(n)^{-1},1\}\sup_{n^{\frac{1}{\alpha+\delta}}\le x\le n^{\frac{1+\delta}{\alpha-\delta}}} l(x)\right)^{1/\alpha}, \\ l_{2n} &:= \left(\frac{1}{2}\inf_{n^{\frac{1}{\alpha+\delta}}\le x\le n^{\frac{1+\delta}{\alpha-\delta}}} l(x)\right)^{1/\alpha}, \\ I_{n,l} &:= [n^{1/\alpha}l_{2n}, n^{1/\alpha}l_{1n}], \\ s_{1n,l} &:= \frac{|\log(\sup_{x\in I_{n,l}} l(x))|}{\log n}, \\ s_{2n,l} &:= \frac{|\log(\inf_{x\in I_{n,l}} l(x))|}{\log n}. \end{aligned}$$

In view of Eq. (10), a two-sided rate statement requires an additional calibration of the sequence c_n to the slowly varying factor l.

Theorem D.1 (Rate refinements under tail regularity)**.** *Assume the standing marginal assumption with $\alpha > 0$ (not necessarily in $(0,1)$). Let $h_1, h_2 > \alpha$ and $h_1 \neq h_2$. Throughout this theorem, Condition A(p) is understood with the sequence c_n specified in the corresponding case below and its associated truncation level v_n.*

(a) Suppose that $\lim_{x\to\infty} l(x) = \infty$. Let $c_n \to \infty$ be such that $1 \le c_n \le l(n)$, $\log c_n = o(\log n)$, and

$$\frac{\log(\inf_{x\in I_{n,l}} l(x))}{\log c_n} \to \infty.$$

Suppose that Condition A(p) holds for $p = h_1$ and $p = h_2$. Then $R^{hm}_{n,h_1,h_2} = O_p(s_{1n,l})$ and $(R^{hm}_{n,h_1,h_2})^{-1} = O_p(s^{-1}_{2n,l})$.

(b) Suppose that $\lim_{x\to\infty} l(x) = 0$. Let $c_n \to \infty$ be such that $1 \le c_n \le l(n)^{-1}$, $\log c_n = o(\log n)$, and

$$\frac{|\log(\sup_{x\in I_{n,l}} l(x))|}{\log c_n} \to \infty.$$

Suppose that Condition A(p) holds for $p = h_1$ and $p = h_2$. Then $R^{hm}_{n,h_1,h_2} = O_p(s_{2n,l})$ and $(R^{hm}_{n,h_1,h_2})^{-1} = O_p(s^{-1}_{1n,l})$.

Analogous rate refinements to those in Theorem D.1 also hold for Theorems 3.10, 3.12, 4.2, F.3, and F.5.

In view of the results in Theorem D.1, the following remark gives sufficient conditions on the slowly varying function such that the rate of convergence of R^{hm}_{n,h_1,h_2} can be established.

Remark D.2. *Suppose that either* $\lim_{x\to\infty} l(x) = \infty$ *or* $\lim_{x\to\infty} l(x) = 0$. *If*

$$\left|\log\left(\sup_{x\in I_{n,l}} l(x)\right)\right| \asymp \left|\log\left(\inf_{x\in I_{n,l}} l(x)\right)\right|, \tag{S.36}$$

then the rate of convergence of $R^{\mathrm{hm}}_{n,h_1,h_2}$ *may be taken to be* $r_n := \frac{\log n}{\left|\log\left(\sup_{x\in I_{n,l}} l(x)\right)\right|}$. *Equivalently, by* Eq. (S.36), $r_n \asymp \frac{\log n}{\left|\log\left(\inf_{x\in I_{n,l}} l(x)\right)\right|}$.

Several examples of slowly varying functions illustrate the results on the rate of convergence in Theorem D.1 and the discussion in Remark D.2.

Example D.3. *(a) Suppose that* $l(x) = \exp((\log x)^\beta)$ *for some* $\beta \in (0,1)$, *which is known to be slowly varying (3, p. 16). Then,* $l(x) \to \infty$ *and*

$$\left|\log\left(\sup_{x\in I_{n,l}} l(x)\right)\right| \asymp \left|\log\left(\inf_{x\in I_{n,l}} l(x)\right)\right| \asymp (\log n)^\beta.$$

Hence the rate of convergence may be taken to satisfy $r_n \asymp (\log n)^{1-\beta}$.

(b) Suppose that $l(x) = \exp(-(\log x)^\beta)$ *for some* $\beta \in (0,1)$. *It can be shown that* l *is slowly varying, similar to the function* $x \mapsto \exp((\log x)^\beta)$. *Then,* $l(x) \to 0$ *and*

$$\left|\log\left(\sup_{x\in I_{n,l}} l(x)\right)\right| \asymp \left|\log\left(\inf_{x\in I_{n,l}} l(x)\right)\right| \asymp (\log n)^\beta.$$

Hence the rate of convergence may be taken to satisfy $r_n \asymp (\log n)^{1-\beta}$.

(c) Suppose that $l(x) = \log(x) f(x)$ *where* $0 < a \le f(x) \le b < \infty$, *and suppose that* l *is slowly varying. Then,* $l(x) \to \infty$ *and*

$$\left|\log\left(\sup_{x\in I_{n,l}} l(x)\right)\right| \asymp \left|\log\left(\inf_{x\in I_{n,l}} l(x)\right)\right| \asymp \log\log n.$$

Hence the rate of convergence may be taken to satisfy $r_n \asymp \frac{\log n}{\log\log n}$.

Verification of Eq. (S.36) **in Example D.3**

We verify Eq. (S.36) for the three slowly varying functions in Example D.3. Let

$$A_n := \left[n^{1/(\alpha+\delta)}, n^{(1+\delta)/(\alpha-\delta)}\right].$$

(a) Suppose that $l(x) = \exp((\log x)^\beta)$ for some $\beta \in (0,1)$. For $x \in A_n$, we have $\log x \asymp \log n$, and hence $\log l(x) = O((\log n)^\beta) = o(\log n)$. Since $l(n) \to \infty$, the definitions of l_{1n} and l_{2n} imply

$$\log l_{1n} = O((\log n)^\beta) = o(\log n), \qquad \log l_{2n} = O((\log n)^\beta) = o(\log n).$$

Therefore, for $i = 1, 2$,

$$\log(n^{1/\alpha} l_{in}) = \frac{1}{\alpha}\log n + o(\log n).$$

Since l is increasing, the infimum and supremum of l over $I_{n,l}$ are attained at the left and right endpoints of $I_{n,l}$, respectively. Thus

$$\left|\log\left(\inf_{x\in I_{n,l}} l(x)\right)\right| = \left(\log(n^{1/\alpha} l_{2n})\right)^\beta \sim \alpha^{-\beta}(\log n)^\beta,$$
$$\left|\log\left(\sup_{x\in I_{n,l}} l(x)\right)\right| = \left(\log(n^{1/\alpha} l_{1n})\right)^\beta \sim \alpha^{-\beta}(\log n)^\beta.$$

Hence Eq. (S.36) holds, and both sides are of order $(\log n)^\beta$.

(b) Suppose that $l(x) = \exp(-(\log x)^\beta)$ for some $\beta \in (0,1)$. For $x \in A_n$, we have $|\log l(x)| = O((\log n)^\beta) = o(\log n)$. Since $l(n) \to 0$, the definitions of l_{1n} and l_{2n} imply

$$\log l_{1n} = O((\log n)^\beta) = o(\log n), \qquad \log l_{2n} = O((\log n)^\beta) = o(\log n).$$

Therefore, for $i = 1, 2$,

$$\log(n^{1/\alpha} l_{in}) = \frac{1}{\alpha} \log n + o(\log n).$$

Since l is decreasing, the supremum and infimum of l over $I_{n,l}$ are attained at the left and right endpoints of $I_{n,l}$, respectively. Thus

$$\left| \log \left(\sup_{x \in I_{n,l}} l(x) \right) \right| = \left(\log(n^{1/\alpha} l_{2n}) \right)^\beta \sim \alpha^{-\beta} (\log n)^\beta,$$
$$\left| \log \left(\inf_{x \in I_{n,l}} l(x) \right) \right| = \left(\log(n^{1/\alpha} l_{1n}) \right)^\beta \sim \alpha^{-\beta} (\log n)^\beta.$$

Hence Eq. (S.36) holds, and both sides are of order $(\log n)^\beta$.

(c) Suppose that $l(x) = \log(x) f(x)$, where $0 < a \le f(x) \le b < \infty$, and suppose that l is slowly varying. For $x \in A_n$, we have $\log x \asymp \log n$, and hence $l(x) \asymp \log n$ uniformly over A_n. Since also $l(n) \asymp \log n$, the definitions of l_{1n} and l_{2n} imply

$$\log l_{1n} = O(\log \log n), \qquad \log l_{2n} = O(\log \log n).$$

Consequently, uniformly for $x \in I_{n,l}$,

$$\log x = \frac{1}{\alpha} \log n + O(\log \log n).$$

Thus $l(x) = \log(x) f(x) \asymp \log n$ uniformly over $I_{n,l}$. Therefore,

$$\log \left(\inf_{x \in I_{n,l}} l(x) \right) = \log \log n + O(1), \qquad \log \left(\sup_{x \in I_{n,l}} l(x) \right) = \log \log n + O(1).$$

Hence Eq. (S.36) holds, and both sides are of order $\log \log n$.

The next example shows that when $l(x)$ does not diverge to infinity or go to 0, the upper bound obtained in Theorem 3.2(b) can be arbitrarily close to the true rate.

Example D.4. *Let $X_1, \ldots, X_n$ be i.i.d. Pareto random variables with survival function $\mathbb{P}(X_i > x) = x^{-\alpha}$ for $i = 1, \ldots, n$ and $0 < \alpha < 1$ for $x \ge 1$. In this case, the slowly varying function is $l(x) = 1$. From Theorem 3.2(b), for any $h_1, h_2 > \alpha$, we have* Eq. (10). *Since $l(x) = 1$, $\log l(t_n) = \log 1 = 0$ and thus the second summand on the right side of* Eq. (10) *is 0. As a result, the upper bound of the convergence rate of $\frac{\log M_{n,h_1}}{\log M_{n,h_2}} - \imath_\alpha(h_1, h_2)$ to 0 is $\frac{\log c_n}{\log n}$; indeed, $\log c_n$ can be chosen to diverge arbitrarily slowly, and the upper bound of the convergence rate can be arbitrarily close to $\frac{1}{\log n}$.*

This result coincides with that of (8). Define $M_{n,2}^{c,X} := n^{-1} \sum_{i=1}^n (X_i - M_{n,1})^2$ and $b = \frac{2-\alpha}{1-\alpha}$ and $W_n := \frac{M_{n,2}^{c,X}}{(M_{n,1})^b}$. From Section 3(a) of (8), we have that $W_n \xrightarrow{d} S_{\alpha/2}/S_\alpha^b$, where S_α is a stable random variable with index $\alpha \in (0,2)$ and shape parameter $\beta = 1$, which implies that

$$\log M_{n,2}^c - b \log M_{n,1} \xrightarrow{d} \log \left(S_{\alpha/2}/S_\alpha^b \right). \tag{S.37}$$

From Eq. (S.37), we obtain

$$\log n \left(\frac{\log M_{n,1}}{\log n} \left(\frac{\log M_{n,2}^c}{\log M_{n,1}} - b \right) \right) \xrightarrow{d} \log \left(S_{\alpha/2}/S_\alpha^b \right).$$

By Lemma 3.1, $\frac{\log M_{n,1}}{\log n} \xrightarrow{\mathbb{P}} \frac{1-\alpha}{\alpha}$. Thus, the convergence rate of $\frac{\log(M_{n,2}^c)}{\log(M_{n,1})} - \frac{2-\alpha}{1-\alpha}$ is $\frac{1}{\log n}$. Our results in this article imply that the convergence rate is arbitrarily close to $\frac{1}{\log n}$. Thus the two results coincide, despite the fact that we are interested in finding the probability limit with its convergence rate while (8) specified a distributional limit without giving a rate of convergence.

E. Proofs for Subsection D

Lemma E.1. *Assume the assumptions used in Theorem D.1.*

(a) If $l(x) \to \infty$, then $t_n \in I_{n,l}$ for all sufficently large n and

$$\log c_n + |\log l(t_n)| \le C \log \left(\sup_{x \in I_{n,l}} l(x) \right).$$

(b) If $l(x) \to 0$, then $t_n \in I_{n,l}$ for all sufficently large n and

$$\log c_n + |\log l(t_n)| \le C \log \left(\inf_{x \in I_{n,l}} l(x) \right).$$

Proof of Lemma E.1. (a) As $l(x) \to \infty$ and $t_n \to \infty$, we have $l(t_n) > 1$ and $l(x) > 1$ for all sufficiently large n and x. Recall that $c_n \to \infty$ is such that $1 \le c_n \le l(n)$ and

$$\frac{\log(\inf_{x \in I_{n,l}} l(x))}{\log c_n} \to \infty. \tag{S.38}$$

From $nl(t_n)/t_n^\alpha \sim 1/c_n$, we have $t_n^\alpha \le 2nc_n l(t_n)$ for all sufficiently large n. As $l(x)/x^{\delta/2} \to 0$ and $c_n \le l(n)$, we have for all sufficiently large n,

$$t_n^{\alpha-\delta} \le 2nc_n \frac{l(t_n)}{t_n^\delta} \le 2nl(n) \frac{l(t_n)}{t_n^\delta} \le 2n^{1+\delta/2} \frac{l(t_n)}{t_n^\delta} \le n^{1+\delta}.$$

Thus, $t_n \le n^{(1+\delta)/(\alpha-\delta)}$ for all sufficiently large n. On the other hand, as $c_n \ge 1$,

$$\frac{nl(t_n)}{t_n^\alpha} \sim \frac{1}{c_n} \le 1,$$

and so for all sufficiently large n,

$$t_n^{\alpha+\delta} \ge \frac{1}{2} n t_n^\delta l(t_n) \ge n t_n^{\delta/2} l(t_n).$$

As $t_n^{\delta/2} l(t_n) \to \infty$ by Proposition 1.3.6 (v) in (3), we have $t_n^{\alpha+\delta} \ge n$ and so $t_n \ge n^{1/(\alpha+\delta)}$ for all large n. From $n^{1/(\alpha+\delta)} \le t_n \le n^{(1+\delta)/(\alpha-\delta)}$ and $t_n^\alpha \sim c_n nl(t_n)$, we have

$$\frac{1}{2} n \inf_{n^{1/(\alpha+\delta)} \le x \le n^{(1+\delta)/(\alpha-\delta)}} l(x) \le t_n^\alpha \le 2l(n) n \sup_{n^{1/(\alpha+\delta)} \le x \le n^{(1+\delta)/(\alpha-\delta)}} l(x).$$

Thus $t_n \in I_{n,l}$. Then, in view of Eq. (S.38), for all large n,

$$\log c_n + |\log l(t_n)| \le 2 \log \left(\sup_{x \in I_{n,l}} l(x) \right).$$

(b) As $l(x) \to 0$ and $t_n \to \infty$, we have $l(t_n) < 1$ and $l(x) < 1$ for all large n and x. Recall that $c_n \to \infty$ is such that $1 \le c_n \le l(n)^{-1}$ and

$$\frac{|\log(\sup_{x \in I_{n,l}} l(x))|}{\log c_n} \to \infty. \tag{S.39}$$

Using the same argument as in the proof of (b)(i), we have $t_n \in I_{n,l}$ for all sufficiently large n. Thus, by Eq. (S.39), for all sufficiently large n,

$$\log c_n + |\log l(t_n)| \le 2 \left| \log \left(\inf_{x \in I_{n,l}} l(x) \right) \right|.$$

□

Proof of Theorem D.1.

(a) In view of Theorem 3.2 (b) and Lemma E.1(a), we have $R^{\mathrm{hm}}_{n,h_1,h_2} = O_p(s_{1n,l})$.

For the reciprocal bound, let

$$L_{n,-} := \log \left(\inf_{x \in I_{n,l}} l(x) \right).$$

Since $l(x) \to \infty$ and $t_n \in I_{n,l}$ for all sufficiently large n, we have $L_{n,-} \to \infty$ and $\log l(t_n) \geq L_{n,-}$. Also, by Eq. (S.38), $\log c_n = o(L_{n,-})$. From Eq. (S.8),

$$\log \frac{d_{n,h_1}}{d_{n,h_2}^{\imath_\alpha(h_1,h_2)}} = o(1) + C_1 + C_2 \log l(t_n),$$

where $C_2 = (h_2 - h_1)/(h_2 - \alpha) \neq 0$. Hence, for all sufficiently large n,

$$\left| \log \frac{d_{n,h_1}}{d_{n,h_2}^{\imath_\alpha(h_1,h_2)}} \right| \geq \frac{|C_2|}{2} \log l(t_n) \geq \frac{|C_2|}{2} L_{n,-}.$$

On the other hand, by Lemma B.3, for every $\varepsilon > 0$, there exists $C > 0$ such that, with probability at least $1 - \varepsilon$, for all sufficiently large n,

$$\left| \log \frac{M_{n,h_1}}{d_{n,h_1}} \right| + \left| \imath_\alpha(h_1,h_2) \log \frac{d_{n,h_2}}{M_{n,h_2}} \right| \leq C \log c_n \leq \frac{|C_2|}{4} L_{n,-}.$$

Using Eq. (S.7) and the reverse triangle inequality, we obtain, with probability at least $1 - \varepsilon$, for all sufficiently large n,

$$\left| \log \frac{M_{n,h_1}}{M_{n,h_2}^{\imath_\alpha(h_1,h_2)}} \right| \geq \left| \log \frac{d_{n,h_1}}{d_{n,h_2}^{\imath_\alpha(h_1,h_2)}} \right| - \left| \log \frac{M_{n,h_1}}{d_{n,h_1}} \right| - \left| \imath_\alpha(h_1,h_2) \log \frac{d_{n,h_2}}{M_{n,h_2}} \right| \geq \frac{|C_2|}{4} L_{n,-}.$$

Finally, dividing both sides of the above inequality by $|\log M_{n,h_2}|$, with probability at least $1 - \varepsilon$, we have

$$\left| R^{\mathrm{hm}}_{n,h_1,h_2} \right| = \left| \frac{\log M_{n,h_1}}{\log M_{n,h_2}} - \imath_\alpha(h_1,h_2) \right| \geq \frac{|C_2|}{4} \frac{L_{n,-}}{\log M_{n,h_2}}.$$

Since $s_{2n,l} = L_{n,-}/\log n$, $(R^{\mathrm{hm}}_{n,h_1,h_2})^{-1} = O_p(s_{2n,l}^{-1})$ in view of Lemma 3.1.

(b) In view of Theorem 3.2 (b) and Lemma E.1(b), we have $R^{\mathrm{hm}}_{n,h_1,h_2} = O_p(s_{2n,l})$.

For the reciprocal bound, let

$$L_{n,+} := \left| \log \left(\sup_{x \in I_{n,l}} l(x) \right) \right|.$$

Since $l(x) \to 0$ and $t_n \in I_{n,l}$ for all sufficiently large n, we have $L_{n,+} \to \infty$ and $|\log l(t_n)| \geq L_{n,+}$. Also, by Eq. (S.39), $\log c_n = o(L_{n,+})$. From Eq. (S.8),

$$\log \frac{d_{n,h_1}}{d_{n,h_2}^{\imath_\alpha(h_1,h_2)}} = o(1) + C_1 + C_2 \log l(t_n),$$

where $C_2 = (h_2 - h_1)/(h_2 - \alpha) \neq 0$. Hence, for all sufficiently large n,

$$\left| \log \frac{d_{n,h_1}}{d_{n,h_2}^{\imath_\alpha(h_1,h_2)}} \right| \geq \frac{|C_2|}{2} |\log l(t_n)| \geq \frac{|C_2|}{2} L_{n,+}.$$

The rest of the argument is the same as in (a). □

F. Taylor's Law for Semivariances, Central Lower and Upper Moments of Heavy-tailed Data

In this section, we extend and discuss Taylor's law for semivariances and central lower and upper moments of *dependent* data; see (10) for more discussion. The semivariance measures only the risk of unfavorable deviations from the mean, while the variance considers equally both upside and downside deviations from the mean. Central lower and upper moments generalize the semivariance to powers other than 2 of deviations from the mean.

Throughout this section, we use the notation from the main text: $Y_i := X_i - x_F$, $M_{n,1} := n^{-1} \sum_{i=1}^n Y_i$, and $M^X_{n,1} := n^{-1} \sum_{i=1}^n X_i = M_{n,1} + x_F$. Since $M_{n,1} \to \infty$ in probability under the assumptions of the theorems below,

$$\mathbb{P}(M^X_{n,1} > 0) \to 1, \quad \frac{|M^X_{n,1}|}{M^X_{n,1}} \xrightarrow{\mathbb{P}} 1, \quad \frac{|M^X_{n,1}|}{M_{n,1}} \xrightarrow{\mathbb{P}} 1.$$

Consequently,

$$\frac{\log |M^X_{n,1}|}{\log M_{n,1}} \xrightarrow{\mathbb{P}} 1.$$

All logarithms involving $|M^X_{n,1}|$ are understood on the event $|M^X_{n,1}| > 0$, whose probability tends to one. Moreover,

$$M^X_{n,1} - X_i = M_{n,1} - Y_i \quad \text{and} \quad X_i - M^X_{n,1} = Y_i - M_{n,1}.$$

Also, $X_i \le M_{n,1}^X$ if and only if $Y_i \le M_{n,1}$. All occurrences of Condition A(p) in this section refer to the shifted truncated variables $\breve{Y}_i := Y_i \mathbb{1}(Y_i \le v_n)$, and all truncation levels below are on the shifted Y-scale.

Denote $a_+ := \max\{a, 0\}$. Let $N_n^- := \sum_{i=1}^n \mathbb{1}(X_i \le M_{n,1}^X)$ and $N_n^+ := \sum_{i=1}^n \mathbb{1}(X_i > M_{n,1}^X)$. For any $h > 0$, the sample hth central lower moments and sample hth central local lower moments are defined as

$$M_{n,h}^- := \frac{1}{n}\sum_{i=1}^n [(M_{n,1}^X - X_i)_+]^h \quad \text{and} \quad M_{n,h}^{-*} := \frac{1}{N_n^-}\sum_{i=1}^n [(M_{n,1}^X - X_i)_+]^h.$$

Similarly, the sample hth central upper moments and sample hth central local upper moments are defined as

$$M_{n,h}^+ := \frac{1}{n}\sum_{i=1}^n [(X_i - M_{n,1}^X)_+]^h \quad \text{and} \quad M_{n,h}^{+*} := \frac{1}{N_n^+}\sum_{i=1}^n [(X_i - M_{n,1}^X)_+]^h.$$

When $h = 2$, $M_{n,2}^-, M_{n,2}^{-*}, M_{n,2}^+$, and $M_{n,2}^{+*}$ are also called the sample lower semivariance, sample local lower semivariance, sample upper semivariance and sample local upper semivariance, respectively.

F.1. Central Lower and Upper Moments. We recall that $\{c_n\}$ is a positive sequence such that $c_n \to \infty$ and $\log c_n = o(\log n)$ and that $\tilde{Y} := Y\mathbb{1}(Y \le t_n)$. Define $d_{n,1} := \mathbb{E}(\tilde{Y})$. For $\delta' > 0$, define $b_n := d_{n,1}/c_n^{2\delta'}$ and $\tilde{b}_n := d_{n,1}c_n$.

Theorem F.1 (hth Central Lower Moments)**.** *Let $X_1, \ldots, X_n, \ldots$ be a sequence of random variables with common finite lower endpoint x_F. Put $Y_i := X_i - x_F$, and suppose that the Y_i's have common survival function $\overline{F}(x) = x^{-\alpha}l(x)$, where l is slowly varying and $\alpha \in (0, 1)$. Suppose that Condition A(p) holds for $p = 1$.*

(i) Suppose that for some $a > 0$,

$$\sum_{i \ne j} \mathrm{Cov}(\mathbb{1}(Y_i \le a), \mathbb{1}(Y_j \le a)) = o(n^2). \tag{S.40}$$

Then, for any $h > 0$,

$$\frac{\log M_{n,h}^-}{\log |M_{n,1}^X|} \xrightarrow{\mathbb{P}} h.$$

More precisely,

$$\frac{\log M_{n,h}^-}{\log |M_{n,1}^X|} = h + O_p\left(\frac{1}{\log n}\right).$$

(ii) For some $\delta' > 2(1/\alpha - 1) > 0$, suppose that

$$\sum_{i \ne j} \mathrm{Cov}(\mathbb{1}(Y_i \le b_n), \mathbb{1}(Y_j \le b_n)) = o(n^2). \tag{S.41}$$

Then, for any $h > 0$,

$$\frac{M_{n,h}^-}{|M_{n,1}^X|^h} \xrightarrow{\mathbb{P}} 1.$$

Remark F.2. *In* Eq. (S.40),

$$\mathrm{Cov}(\mathbb{1}(Y_i \le a), \mathbb{1}(Y_j \le a)) = \mathbb{P}(Y_i \le a, Y_j \le a) - \mathbb{P}(Y_i \le a)\mathbb{P}(Y_j \le a).$$

Equivalently, this condition can be written with $\mathbb{1}(X_i \le x_F + a)$ and $\mathbb{1}(X_j \le x_F + a)$. This condition is "weaker" than strong mixing as we only require the condition to hold for one value of a on the shifted scale, equivalently for one threshold $x_F + a$ above the lower endpoint of the support of X. The condition guarantees that $\frac{1}{n}\sum_{i=1}^n \mathbb{1}(Y_i \le a)$ is bounded away from 0 in probability.

Theorem F.3 (hth Central Upper Moments)**.** *Let $X_1, \ldots, X_n, \ldots$ be a sequence of random variables with common finite lower endpoint x_F. Put $Y_i := X_i - x_F$, and suppose that the Y_i's have common survival function $\overline{F}(x) = x^{-\alpha}l(x)$, where l is slowly varying and $\alpha \in (0, 1)$.*

(a) If Condition A(p) holds for $p = 1$ and $p = h$ with $\alpha < h < 1$, and if Eq. (S.41) *also holds for some $\delta' > 2(1/\alpha - 1) > 0$, then*

$$\frac{M_{n,h}^+}{|M_{n,1}^X|^h} = o_p(1).$$

(b) Suppose that Condition A(p) holds for $p = 1$ and $p = h$ with $h > 1$. Then

$$\frac{\log M_{n,h}^+}{\log |M_{n,1}^X|} \xrightarrow{\mathbb{P}} \imath_\alpha(h, 1).$$

More precisely,

$$R_{n,h,1}^{cum} := \frac{\log M_{n,h}^+}{\log |M_{n,1}^X|} - \imath_\alpha(h, 1) = O_p\left(\frac{\log c_n}{\log n}\right) + O_p\left(\frac{|\log l(t_n)|}{\log n}\right).$$

F.2. Local Central Lower and Upper Moments.

Theorem F.4 (hth Local Central Lower Moments)**.** *Let $X_1, \ldots, X_n, \ldots$ be a sequence of random variables with common finite lower endpoint x_F. Put $Y_i := X_i - x_F$, and suppose that the Y_i's have common survival function $\overline{F}(x) = x^{-\alpha} l(x)$, where l is slowly varying and $\alpha \in (0,1)$. If Condition A(p) for $p = 1$ and* Eq. (S.40) *hold, then for any $h > 0$,*

$$\frac{\log M_{n,h}^{-*}}{\log |M_{n,1}^X|} \xrightarrow{\mathbb{P}} h.$$

More precisely,

$$\frac{\log M_{n,h}^{-*}}{\log |M_{n,1}^X|} = h + O_p\left(\frac{1}{\log n}\right).$$

To establish Taylor's law for the hth local central upper moment, we impose an additional condition for the truncated terms:

$$\sum_{i \neq j} \mathrm{Cov}(\mathbb{1}(Y_i > \tilde{b}_n), \mathbb{1}(Y_j > \tilde{b}_n)) = o\left(n^{2\alpha}\left(\frac{l(\tilde{b}_n)}{l(t_n)c_n}\right)^2\right). \tag{S.42}$$

Theorem F.5 (hth Local Upper Central Moments)**.** *Let $X_1, \ldots, X_n, \ldots$ be a sequence of random variables with common finite lower endpoint x_F. Put $Y_i := X_i - x_F$, and suppose that the Y_i's have common survival function $\overline{F}(x) = x^{-\alpha} l(x)$, where l is slowly varying and $\alpha \in (0,1)$. Suppose that Condition A(p) holds for $p = 1$, that* Eq. (S.42) *holds, and that the δ' in the definition of b_n satisfies $\delta' > 2(1/\alpha - 1)$. For $h > 1$, define $\imath_+(h) := (h - \alpha^2)/(1 - \alpha)$. Let*

$$L_n := \log c_n + |\log l(t_n)| + |\log l(b_n)| + |\log l(\tilde{b}_n)|.$$

(a) If Condition A(p) also holds for $p = h$, then

$$\frac{\log M_{n,h}^{+*}}{\log |M_{n,1}^X|} \xrightarrow{\mathbb{P}} \imath_+(h).$$

More precisely,

$$R_{n,h,1}^{lucm} := \frac{\log M_{n,h}^{+*}}{\log |M_{n,1}^X|} - \imath_+(h) = O_p\left(\frac{L_n}{\log n}\right).$$

(b) Furthermore, under the assumptions in part (a), suppose that $l(b_n)/l(t_n) \to c_1$ and $l(\tilde{b}_n)/l(t_n) \to c_2$ for some constants $c_1, c_2 > 0$. Then

$$R_{n,h,1}^{lucm} = O_p\left(\frac{\log c_n + |\log l(t_n)|}{\log n}\right).$$

Remark F.6. *For some choices of l, we have $l(b_n)/l(t_n) \to c_1 > 0$ and $l(\tilde{b}_n)/l(t_n) \to c_2 > 0$. For example, consider $l(t) = \log t$. As $n \to \infty$,*

$$d_{n,1} c_n \sim C(t_n^\alpha)^{1/\alpha - 1} l(t_n) c_n = C n^{1/\alpha - 1} c_n^{1/\alpha} l(t_n)^{1/\alpha}.$$

Therefore, as $n \to \infty$,

$$\frac{l(\tilde{b}_n)}{l(t_n)} = \frac{\log \tilde{b}_n}{\frac{1}{\alpha}\log t_n^\alpha} = \frac{o(1) + \log C + (\frac{1}{\alpha} - 1)\log n + \frac{1}{\alpha}\log c_n + \frac{1}{\alpha}\log l(t_n)}{\frac{1}{\alpha}(\log n + \log l(t_n) + \log c_n)} \to \frac{\frac{1}{\alpha} - 1}{\frac{1}{\alpha}} = 1 - \alpha.$$

Similarly, as $n \to \infty$,

$$\frac{l(b_n)}{l(t_n)} \to 1 - \alpha.$$

G. Proofs for Section F

G.1. Proofs for Section F.1.

Proof of Theorem F.1. (i) We have an upper bound:

$$\frac{M_{n,h}^-}{M_{n,1}^h} = \frac{\frac{1}{n}\sum_{i:Y_i \le M_{n,1}} (M_{n,1} - Y_i)^h}{M_{n,1}^h} \le \frac{\frac{1}{n}\sum_{i:Y_i \le M_{n,1}} M_{n,1}^h}{M_{n,1}^h} \le 1. \tag{S.43}$$

Fix $a > 0$. Since, on $M_{n,1} > a$, each $Y_i \le a$ is included in the sum and satisfies $(1 - Y_i/M_{n,1})^h \ge (1 - a/M_{n,1})^h$, we have

$$\frac{M_{n,h}^-}{M_{n,1}^h} = \frac{1}{n} \sum_{i:Y_i \le M_{n,1}} \left(1 - \frac{Y_i}{M_{n,1}}\right)^h \ge \left[\frac{1}{n}\sum_{i=1}^n \mathbb{1}(Y_i \le a)\right] \left(1 - \frac{a}{M_{n,1}}\right)^h \mathbb{1}(M_{n,1} > a). \tag{S.44}$$

Since Condition A(p) holds for $p = 1$, from the lower bound for $M_{n,1}/d_{n,1}$ in Lemma B.2, we know that $M_{n,1}$ diverges to ∞ in probability. Thus, as $a > 0$ is fixed,

$$\left(1 - \frac{a}{M_{n,1}}\right)^h \mathbb{1}(M_{n,1} > a) \xrightarrow{\mathbb{P}} 1. \tag{S.45}$$

For any $\varepsilon > 0$, by Markov's inequality,

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n \mathbb{1}(Y_i \le a) - F(a)\right| > \varepsilon\right) &\le \frac{\operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{1}(Y_i \le a)\right)}{\varepsilon^2} \\ &= \frac{F(a)(1-F(a))}{n\varepsilon^2} + \frac{\sum_{i\ne j}\operatorname{Cov}(\mathbb{1}(Y_i \le a), \mathbb{1}(Y_j \le a))}{n^2\varepsilon^2}.\end{aligned}$$

The last expression goes to 0 by Eq. (S.40). This together with Eq. (S.44) and Eq. (S.45) gives

$$\frac{M_{n,h}^-}{M_{n,1}^h} \ge (F(a) + o_p(1))(1 + o_p(1)) = F(a) + o_p(1). \tag{S.46}$$

In view of Eq. (S.43) and Eq. (S.46), we have

$$\log \frac{M_{n,h}^-}{M_{n,1}^h} = O_p(1),$$

which implies that

$$\log M_{n,h}^- - h \log M_{n,1} = O_p(1).$$

Since $|M_{n,1}^X|/M_{n,1} \xrightarrow{\mathbb{P}} 1$,

$$\log |M_{n,1}^X| - \log M_{n,1} = \log \frac{|M_{n,1}^X|}{M_{n,1}} = o_p(1).$$

Therefore,

$$\log M_{n,h}^- - h \log |M_{n,1}^X| = O_p(1).$$

Dividing both sides of the last equation by $\log |M_{n,1}^X|$, we obtain

$$\frac{\log M_{n,h}^-}{\log |M_{n,1}^X|} = h + O_p\left(\frac{1}{\log |M_{n,1}^X|}\right).$$

The claim in the theorem then follows since $\log |M_{n,1}^X|/\log n \xrightarrow{\mathbb{P}} \frac{1-\alpha}{\alpha}$, which follows from $\log |M_{n,1}^X|/\log M_{n,1} \xrightarrow{\mathbb{P}} 1$ and Lemma 3.1.

(ii) As in (i), we have

$$\left[\frac{1}{n}\sum_{i=1}^n \mathbb{1}(Y_i \le b_n)\right] \left(1 - \frac{b_n}{M_{n,1}}\right)^h \mathbb{1}(M_{n,1} > b_n) \le \frac{M_{n,h}^-}{M_{n,1}^h} \le 1. \tag{S.47}$$

By Lemma B.2,

$$\frac{c_{\alpha,1,\delta'}(1 + o_p(1))}{c_n^{\delta'}} \le \frac{M_{n,1}}{d_{n,1}}.$$

Thus, by the definition of b_n,

$$\frac{b_n}{M_{n,1}} = O_p(c_n^{-\delta'}) = o_p(1).$$

As a result,

$$\mathbb{1}(M_{n,1} > b_n) \xrightarrow{\mathbb{P}} 1 \text{ and } \left(1 - \frac{b_n}{M_{n,1}}\right)^h \xrightarrow{\mathbb{P}} 1.$$

By the same argument as in (i),

$$\frac{1}{n}\sum_{i=1}^n \mathbb{1}(Y_i \le b_n) = o_p(1) + F(b_n) \to 1, \tag{S.48}$$

as $b_n \to \infty$. In view of Eq. (S.47)–Eq. (S.48),

$$\frac{M_{n,h}^{-}}{M_{n,1}^{h}} \xrightarrow{\mathbb{P}} 1.$$

Since $|M_{n,1}^{X}|/M_{n,1} \xrightarrow{\mathbb{P}} 1$, we have

$$\frac{M_{n,h}^{-}}{|M_{n,1}^{X}|^{h}} = \frac{M_{n,h}^{-}}{M_{n,1}^{h}} \left(\frac{M_{n,1}}{|M_{n,1}^{X}|} \right)^{h} \xrightarrow{\mathbb{P}} 1. \qquad \square$$

To prove Theorem F.3, we establish a lemma analogous to Lemma B.3.

Lemma G.1. *Let $X_1, \ldots, X_n, \ldots$ be a sequence of random variables with common finite lower endpoint x_F. Put $Y_i := X_i - x_F$, and suppose that the Y_i's have common survival function $\overline{F}(x) = x^{-\alpha} l(x)$, where l is slowly varying and $\alpha \in (0,1)$. Suppose that Condition A(p) holds for $p = h > 1$. Then*

$$\log \frac{M_{n,h}^{+}}{d_{n,h}} = O_p\left(\log c_n\right).$$

Proof of Lemma G.1. To establish the claimed result, we shall obtain upper and lower bounds of $M_{n,h}^{+}/d_{n,h}$.

For an upper bound for $M_{n,h}^{+}/d_{n,h}$, we first note that:

$$\begin{aligned} M_{n,h}^{+} &= \frac{1}{n}\sum_{i=1}^{n} [(Y_i - M_{n,1})_+]^h \le \frac{1}{n}\sum_{i=1}^{n} |Y_i - M_{n,1}|^h \\ &\le 2^{h-1}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i^h + M_{n,1}^h \right) = 2^{h-1}(M_{n,h} + M_{n,1}^h). \end{aligned} \tag{S.49}$$

By Lemma B.1,

$$\frac{M_{n,h}}{d_{n,h}} = O_p(1). \tag{S.50}$$

Thus, by Eq. (S.49), Eq. (S.50), and Lemma B.4, we have

$$\frac{M_{n,h}^{+}}{d_{n,h}} \le 2^{h-1}\left(\frac{M_{n,h}}{d_{n,h}} + \frac{M_{n,1}^h}{d_{n,h}} \right) = 2^{h-1}\left(O_p(1) + O_p\left(\frac{1}{(nc_n)^{h-1}} \right) \right) = O_p(1), \tag{S.51}$$

where the last equality follows as $h > 1$ and $c_n \to \infty$.

We now establish a lower bound for $M_{n,h}^{+}/d_{n,h}$. As $|a-b| = (a-b)_+ + (b-a)_+$ for any $a, b \in \mathbb{R}$, we have

$$M_{n,h}^{+} = \frac{1}{n}\sum_{i=1}^{n} |Y_i - M_{n,1}|^h - \frac{1}{n}\sum_{i=1}^{n} (M_{n,1} - Y_i)_+^h. \tag{S.52}$$

Using the inequality $(a+b)^p \le 2^{p-1}(a^p + b^p)$ for $a, b \ge 0$ and $p \ge 1$, we have

$$\frac{1}{n}\sum_{i=1}^{n} Y_i^h = \frac{1}{n}\sum_{i=1}^{n} |Y_i - M_{n,1} + M_{n,1}|^h \le 2^{h-1}\left(\frac{1}{n}\sum_{i=1}^{n} |Y_i - M_{n,1}|^h + M_{n,1}^h \right).$$

Thus,

$$\frac{1}{n}\sum_{i=1}^{n} |Y_i - M_{n,1}|^h \ge \frac{1}{n2^{h-1}}\sum_{i=1}^{n} Y_i^h - M_{n,1}^h = \frac{M_{n,h}}{2^{h-1}} - M_{n,1}^h.$$

As $(M_{n,1} - Y_i)_+ \le M_{n,1}$, we have

$$\frac{1}{n}\sum_{i=1}^{n} (M_{n,1} - Y_i)_+^h \le M_{n,1}^h. \tag{S.53}$$

In view of Eq. (S.52)–Eq. (S.53),

$$M_{n,h}^{+} \ge \frac{M_{n,h}}{2^{h-1}} - 2M_{n,1}^h.$$

Then, by Lemma B.2 and Lemma B.4,

$$\frac{M_{n,h}^{+}}{d_{n,h}} \ge \frac{c_{h,\alpha,\delta'}(1 + o_p(1))}{2^{h-1} c_n^{\delta'}} - O_p\left(\frac{1}{(nc_n)^{h-1}} \right), \tag{S.54}$$

for some constant $\delta' > 2(h/\alpha - 1) > 0$ and $c_{h,\alpha,\delta'} > 0$. Since $\log c_n = o(\log n)$,

$$\frac{c_n^{\delta'}}{(nc_n)^{h-1}} \to 0.$$

Thus the second term on the right-hand side of Eq. (S.54) is $o_p(c_n^{-\delta'})$.

Finally, in view of Eq. (S.51) and Eq. (S.54), as $h > 1$, for any $\varepsilon \in (0,1)$, with probability greater than $1-\varepsilon$, for all sufficiently large n, there exists some $C > 0$ such that

$$\frac{1}{2} \cdot \frac{c_{h,\alpha,\delta'}}{2^{h-1} c_n^{\delta'}} \le \frac{M_{n,h}^+}{d_{n,h}} \le C.$$

Therefore, for all sufficiently large n,

$$\left| \log \frac{M_{n,h}^+}{d_{n,h}} \right| \le |\log C| + |\log(c_{h,\alpha,\delta'}/2)| + h \log 2 + \delta' \log c_n. \qquad \square$$

Proof of Theorem F.3. (a) We first show that

$$\frac{\frac{1}{n}\sum_{i=1}^n |Y_i - M_{n,1}|^h}{M_{n,1}^h} - 1 = o_p(1).$$

Following the proof of Theorem B.1 of (10), we have

$$\left| \frac{\frac{1}{n}\sum_{i=1}^n |Y_i - M_{n,1}|^h}{M_{n,1}^h} - 1 \right| \le \frac{M_{n,h}}{M_{n,1}^h}.$$

By Lemmas B.1 and B.2, and Eq. (S.10),

$$\frac{M_{n,h}}{M_{n,1}^h} = \frac{M_{n,h}}{d_{n,h}} \cdot \frac{d_{n,h}}{d_{n,1}^h} \cdot \left(\frac{d_{n,1}}{M_{n,1}} \right)^h = O_p(1) O((nc_n)^{h-1}) O_p(c_n^{h\delta'}) = o_p(1),$$

as $h < 1$ and $\log c_n = o(\log n)$. Since $M_{n,1}/|M_{n,1}^X| \xrightarrow{\mathbb{P}} 1$, we have

$$\frac{\frac{1}{n}\sum_{i=1}^n |Y_i - M_{n,1}|^h}{|M_{n,1}^X|^h} = 1 + o_p(1).$$

By Theorem F.1 (ii),

$$\frac{M_{n,h}^+}{|M_{n,1}^X|^h} = \frac{\frac{1}{n}\sum_{i=1}^n |Y_i - M_{n,1}|^h}{|M_{n,1}^X|^h} - \frac{M_{n,h}^-}{|M_{n,1}^X|^h} = (1 + o_p(1)) - (1 + o_p(1)) = o_p(1).$$

(b) With Lemma G.1, the proof of the result is similar to that of Theorem 3.2. Since $\log |M_{n,1}^X| - \log M_{n,1} = o_p(1)$ and $\log |M_{n,1}^X| / \log n \xrightarrow{\mathbb{P}} (1-\alpha)/\alpha$, replacing $\log M_{n,1}$ by $\log |M_{n,1}^X|$ does not affect the stated rate. $\square$

G.2. Proofs for Section F.2. Let $M_{n,1}^* := \frac{1}{n}\sum_{i=1}^n Y_i^*$, where Y_i^*'s have the same joint distribution as Y_i's and are independent of Y_i's when used jointly. Recall that $d_{n,1} = \mathbb{E}(Y_1 \mathbb{1}(Y_1 \le t_n))$, $b_n = d_{n,1}/c_n^{2\delta'}$, and $\tilde{b}_n = d_{n,1} c_n$ for $\delta' > 0$. To prove Theorems F.4 and F.5, we establish Lemmas G.2–G.4.

Lemma G.2. *Let $X_1, \ldots, X_n, \ldots$ be a sequence of random variables with common finite lower endpoint x_F. Put $Y_i := X_i - x_F$, and suppose that the Y_i's have common survival function $\overline{F}(x) = x^{-\alpha} l(x)$, where l is slowly varying and $\alpha \in (0,1)$.*

(i) If Condition A(p) holds for $p = 1$, then

$$\frac{\sum_{i=1}^n \mathbb{1}(Y_i > M_{n,1}^*)}{n^\alpha} = O_p\left(\frac{l(b_n) c_n^{2\delta'\alpha + \alpha - 1}}{l(t_n)} \right),$$

where $\delta' > 2(1/\alpha - 1)$.

(ii) If Condition A(p) holds for $p = 1$ and Eq. (S.42) *holds, then*

$$\left(\frac{\sum_{i=1}^n \mathbb{1}(Y_i > M_{n,1}^*)}{n^\alpha} \right)^{-1} = O_p\left(\frac{l(t_n) c_n}{l(\tilde{b}_n)} \right).$$

Proof of Lemma G.2. (i) Let $C > 0$ and

$$g(n) := \frac{n\mathbb{P}(Y_1 > b_n)}{n^\alpha}.$$

We have

$$\mathbb{P}\left(\frac{\sum_{i=1}^n \mathbb{1}(Y_i > M_{n,1}^*)}{n^\alpha} > Cg(n)\right) = \int_0^{b_n} \mathbb{P}\left(\frac{\sum_{i=1}^n \mathbb{1}(Y_i > t)}{n^\alpha} > Cg(n)\right) dF_{M_{n,1}^*}(t) + \int_{b_n}^{\infty} \mathbb{P}\left(\frac{\sum_{i=1}^n \mathbb{1}(Y_i > t)}{n^\alpha} > Cg(n)\right) dF_{M_{n,1}^*}(t), \quad [S.55]$$

where $F_{M_{n,1}^*}(\cdot)$ is the cumulative distribution function of $M_{n,1}^*$. For the first term on the right side of Eq. (S.55), since any probability is bounded by 1, we have

$$\int_0^{b_n} \mathbb{P}\left(\frac{\sum_{i=1}^n \mathbb{1}(Y_i > t)}{n^\alpha} > Cg(n)\right) dF_{M_{n,1}^*}(t) \le \mathbb{P}\left(M_{n,1}^* \le b_n\right) = 1 - \mathbb{P}\left(M_{n,1} > b_n\right),$$

as $M_{n,1}^* \overset{d}{=} M_{n,1}$. By the definition of b_n and Lemma B.2, for some constant C' independent of n,

$$\frac{M_{n,1}}{b_n} = c_n^{2\delta'} \frac{M_{n,1}}{d_{n,1}} \ge C' c_n^{\delta'}(1 + o_p(1)).$$

As $c_n \to \infty$, we have

$$\mathbb{P}(M_{n,1} > b_n) \to 1.$$

Since for any $t > b_n$,

$$\mathbb{P}\left(\frac{\sum_{i=1}^n \mathbb{1}(Y_i > t)}{n^\alpha} > Cg(n)\right) \le \mathbb{P}\left(\frac{\sum_{i=1}^n \mathbb{1}(Y_i > b_n)}{n^\alpha} > Cg(n)\right).$$

For the second term on the right side of Eq. (S.55), we have

$$\int_{b_n}^{\infty} \mathbb{P}\left(\frac{\sum_{i=1}^n \mathbb{1}(Y_i > t)}{n^\alpha} > Cg(n)\right) dF_{M_{n,1}^*}(t) \le \mathbb{P}\left(\frac{\sum_{i=1}^n \mathbb{1}(Y_i > b_n)}{n^\alpha} > Cg(n)\right).$$

By Markov's inequality and the definition of $g(n)$,

$$\mathbb{P}\left(\frac{\sum_{i=1}^n \mathbb{1}(Y_i > b_n)}{n^\alpha} > Cg(n)\right) \le \frac{n\mathbb{P}(Y_1 > b_n)}{Cn^\alpha g(n)} = \frac{1}{C}. \quad [S.56]$$

In view of Eq. (S.55)–Eq. (S.56), we have shown that

$$\frac{\sum_{i=1}^n \mathbb{1}(Y_i > M_{n,1}^*)}{n^\alpha} = O_p(g(n)).$$

Finally, by Lemma 2.1, for some constant C',

$$g(n) = \frac{n^{1-\alpha} l(b_n) c_n^{2\delta'\alpha}}{d_{n,1}^\alpha} \sim C' \frac{n^{1-\alpha} l(b_n) c_n^{2\delta'\alpha}}{(t_n^{1-\alpha} l(t_n))^\alpha} = C' \frac{n^{1-\alpha} l(b_n) c_n^{2\delta'\alpha}}{(nl(t_n)c_n)^{1-\alpha} l(t_n)^\alpha} = C' \frac{l(b_n) c_n^{2\delta'\alpha+\alpha-1}}{l(t_n)}.$$

Thus, the result in part (i) of the lemma follows.

(ii) Let $C > 0$ and

$$g_2(n) := \frac{l(\tilde{b}_n)}{l(t_n)c_n}.$$

We have

$$\mathbb{P}\left(\left(\frac{\sum_{i=1}^n \mathbb{1}(Y_i > M_{n,1}^*)}{n^\alpha}\right)^{-1} > \frac{C}{g_2(n)}\right) = \int_0^{\tilde{b}_n} \mathbb{P}\left(\frac{\sum_{i=1}^n \mathbb{1}(Y_i > t)}{n^\alpha} < \frac{g_2(n)}{C}\right) dF_{M_{n,1}^*}(t) + \int_{\tilde{b}_n}^{\infty} \mathbb{P}\left(\frac{\sum_{i=1}^n \mathbb{1}(Y_i > t)}{n^\alpha} < \frac{g_2(n)}{C}\right) dF_{M_{n,1}^*}(t). \quad [S.57]$$

For the second term on the right hand side of Eq. (S.57),

$$\int_{\tilde{b}_n}^{\infty} \mathbb{P}\left(\frac{\sum_{i=1}^n \mathbb{1}(Y_i > t)}{n^\alpha} < \frac{g_2(n)}{C}\right) dF_{M^*_{n,1}}(t) \le 1 - F_{M^*_{n,1}}(\tilde{b}_n).$$

We have

$$1 - F_{M^*_{n,1}}(\tilde{b}_n) = \mathbb{P}(M^*_{n,1} > \tilde{b}_n) = \mathbb{P}\left(\frac{M_{n,1}}{d_{n,1}c_n} > 1\right) \to 0,$$

where the last convergence follows as $\frac{M_{n,1}}{d_{n,1}} = O_p(1)$ by Lemma B.1 so that $\frac{M_{n,1}}{d_{n,1}c_n} = o_p(1)$. For the first term on the right hand side of Eq. (S.57), when $0 \le t \le \tilde{b}_n$,

$$\mathbb{P}\left(\frac{\sum_{i=1}^n \mathbb{1}(Y_i > t)}{n^\alpha} < \frac{g_2(n)}{C}\right) \le \mathbb{P}\left(\frac{\sum_{i=1}^n \mathbb{1}(Y_i > \tilde{b}_n)}{n^\alpha} < \frac{g_2(n)}{C}\right)$$

and so

$$\int_0^{\tilde{b}_n} \mathbb{P}\left(\frac{\sum_{i=1}^n \mathbb{1}(Y_i > t)}{n^\alpha} < \frac{g_2(n)}{C}\right) dF_{M^*_{n,1}}(t) \le \mathbb{P}\left(\frac{\sum_{i=1}^n \mathbb{1}(Y_i > \tilde{b}_n)}{n^\alpha} < \frac{g_2(n)}{C}\right),$$

as $F_{M^*_{n,1}}(\tilde{b}_n) \le 1$. Write

$$\frac{\sum_{i=1}^n \mathbb{1}(Y_i > \tilde{b}_n)}{n^\alpha} = \frac{\sum_{i=1}^n (\mathbb{1}(Y_i > \tilde{b}_n) - \overline{F}(\tilde{b}_n))}{n^\alpha} + \frac{n\overline{F}(\tilde{b}_n)}{n^\alpha}.$$

For some constant $C' > 0$ independent of n, by Lemma 2.1,

$$\frac{n\overline{F}(\tilde{b}_n)}{n^\alpha} = \frac{n^{1-\alpha} l(\tilde{b}_n)}{d_{n,1}^\alpha c_n^\alpha} \sim C' \frac{n^{1-\alpha} l(\tilde{b}_n)}{t_n^{(1-\alpha)\alpha} l(t_n)^\alpha c_n^\alpha} = C' \frac{l(\tilde{b}_n)}{l(t_n)c_n} = C' g_2(n).$$

For any $\varepsilon > 0$, by Markov's inequality,

$$\begin{aligned} &\mathbb{P}\left(\left|\frac{\sum_{i=1}^n (\mathbb{1}(Y_i > \tilde{b}_n) - \overline{F}(\tilde{b}_n))}{n^\alpha}\right| > \varepsilon g_2(n)\right) \\ &\le \frac{n\overline{F}(\tilde{b}_n)(1 - \overline{F}(\tilde{b}_n))}{n^{2\alpha}\varepsilon^2 g_2^2(n)} + \frac{\sum_{i \ne j} \mathrm{Cov}(\mathbb{1}(Y_i > \tilde{b}_n), \mathbb{1}(Y_j > \tilde{b}_n))}{n^{2\alpha}\varepsilon^2 g_2^2(n)}. \end{aligned} \tag{S.58}$$

The first term on the right side of Eq. (S.58) obeys

$$\frac{n\overline{F}(\tilde{b}_n)(1 - \overline{F}(\tilde{b}_n))}{n^{2\alpha}\varepsilon^2 g_2^2(n)} \sim \frac{1}{n^\alpha \varepsilon^2 g_2(n)} \to 0.$$

The second term on the right side of Eq. (S.58) goes to 0 by Eq. (S.42). Thus we have shown that

$$\frac{\sum_{i=1}^n \mathbb{1}(Y_i > \tilde{b}_n)}{n^\alpha} = (C' + o_p(1)) g_2(n).$$

Finally,

$$\mathbb{P}\left(\frac{\sum_{i=1}^n \mathbb{1}(Y_i > \tilde{b}_n)}{n^\alpha} < \frac{g_2(n)}{C}\right) = \mathbb{P}\left(C' + o_p(1) < \frac{1}{C}\right).$$

The last probability can be made arbitrarily small by picking C large enough. □

Lemma G.3. *Let $X_1, \ldots, X_n, \ldots$ be a sequence of random variables with common finite lower endpoint x_F. Put $Y_i := X_i - x_F$, and suppose that the Y_i's have common survival function $\overline{F}(x) = x^{-\alpha} l(x)$, where l is slowly varying and $\alpha \in (0, 1)$.*

(i) If Condition A(p) holds for $p = 1$, then

$$\frac{N_n^+}{n^\alpha} = O_p\left(\frac{l(b_n) c_n^{2\delta'\alpha + \alpha - 1}}{l(t_n)}\right),$$

where $\delta' > 2(1/\alpha - 1)$.

(ii) If Condition A(p) holds for $p = 1$ and Eq. (S.42) *holds, then*

$$\left(\frac{N_n^+}{n^\alpha}\right)^{-1} = O_p\left(\frac{l(t_n)c_n}{l(\tilde{b}_n)}\right).$$

Proof of Lemma G.3. (i) With Lemma G.2 (i), the proof follows using the argument in the proof of Lemma A.3 in (10), with Y_i replacing X_i, and is therefore omitted.

(ii) Consider $k > 1$ independent samples, each of size n, having the same joint distribution as $Y_1, \ldots, Y_n$:

$$\{Y_1^{(1)}, \ldots, Y_n^{(1)}\}, \ldots, \{Y_1^{(k)}, \ldots, Y_n^{(k)}\}.$$

Denote $M_{n,1}^{(j)} := \frac{1}{n}\sum_{i=1}^n Y_i^{(j)}$. Since $N_n^+ = \sum_{i=1}^n \mathbb{1}(Y_i > M_{n,1})$, we have

$$N_n^+ \geq \sum_{i=1}^n \left[\mathbb{1}\left(Y_i > \max_{1\leq j\leq k} M_{n,1}^{(j)} \right) - \mathbb{1}\left(\max_{1\leq j\leq k} M_{n,1}^{(j)} < Y_i \leq M_{n,1} \right) \right]. \tag{S.59}$$

We claim that

$$\sum_{i=1}^n \mathbb{1}\left(Y_i > \max_{1\leq j\leq k} M_{n,1}^{(j)} \right) = \min_{1\leq j\leq k} \sum_{i=1}^n \mathbb{1}\left(Y_i > M_{n,1}^{(j)} \right). \tag{S.60}$$

Indeed, for any $j = 1, \ldots, k$,

$$\sum_{i=1}^n \mathbb{1}\left(Y_i > \max_{1\leq j\leq k} M_{n,1}^{(j)} \right) \leq \sum_{i=1}^n \mathbb{1}\left(Y_i > M_{n,1}^{(j)} \right).$$

Thus

$$\sum_{i=1}^n \mathbb{1}\left(Y_i > \max_{1\leq j\leq k} M_{n,1}^{(j)} \right) \leq \min_{1\leq j\leq k} \sum_{i=1}^n \mathbb{1}\left(Y_i > M_{n,1}^{(j)} \right). \tag{S.61}$$

On the other hand, since $\max_{1\leq j\leq k} M_{n,1}^{(j)}$ is equal to one of $M_{n,1}^{(j)}$, we have

$$\min_{1\leq j\leq k} \sum_{i=1}^n \mathbb{1}\left(Y_i > M_{n,1}^{(j)} \right) \leq \sum_{i=1}^n \mathbb{1}\left(Y_i > \max_{1\leq j\leq k} M_{n,1}^{(j)} \right). \tag{S.62}$$

In view of Eq. (S.61) and Eq. (S.62), we have Eq. (S.60). From Eq. (S.59) and Eq. (S.60), we obtain

$$\begin{aligned} N_n^+ &\geq \min_{1\leq j\leq k} \sum_{i=1}^n \mathbb{1}\left(Y_i > M_{n,1}^{(j)} \right) - \sum_{i=1}^n \mathbb{1}\left(\max_{1\leq j\leq k} M_{n,1}^{(j)} < Y_i \leq M_{n,1} \right) \\ &\geq \min_{1\leq j\leq k} \sum_{i=1}^n \mathbb{1}\left(Y_i > M_{n,1}^{(j)} \right) - n\mathbb{1}\left(\max_{1\leq j\leq k} M_{n,1}^{(j)} < M_{n,1} \right), \end{aligned} \tag{S.63}$$

where the second inequality follows as

$$\mathbb{1}\left(\max_{1\leq j\leq k} M_{n,1}^{(j)} < Y_i \leq M_{n,1} \right) \leq \mathbb{1}\left(\max_{1\leq j\leq k} M_{n,1}^{(j)} < M_{n,1} \right).$$

Denote

$$g_2(n) := \frac{l(\tilde{b}_n)}{l(t_n)c_n}.$$

Let $\varepsilon > 0$. Choose k large enough so that $(k+1)^{-1} < \varepsilon/2$. For any $C > 0$, by Eq. (S.63),

$$\begin{aligned} &\mathbb{P}\left(\left(\frac{N_n^+}{n^\alpha} \right)^{-1} > \frac{C}{g_2(n)} \right) = \mathbb{P}\left(\frac{N_n^+}{n^\alpha} < \frac{g_2(n)}{C} \right) \\ &\leq \mathbb{P}\left(\frac{1}{n^\alpha} \min_{1\leq j\leq k} \sum_{i=1}^n \mathbb{1}\left(Y_i > M_{n,1}^{(j)} \right) < \frac{g_2(n)}{C} \right) + \mathbb{P}\left(\max_{1\leq j\leq k} M_{n,1}^{(j)} < M_{n,1} \right). \end{aligned} \tag{S.64}$$

Since $M_{n,1}^{(1)}, \ldots, M_{n,1}^{(k)}, M_{n,1}$ are independent and have a common distribution, by symmetry,

$$\mathbb{P}\left(\max_{1\leq j\leq k} M_{n,1}^{(j)} < M_{n,1} \right) \leq \frac{1}{k+1} < \frac{\varepsilon}{2}. \tag{S.65}$$

For any random variables $Z_1, \ldots, Z_k$, $\mathbb{P}(\min_{j=1,\ldots,k} Z_j < c) \leq \mathbb{P}(\cup_j \{Z_j < c\}) \leq \sum_{j=1}^k \mathbb{P}(Z_j < c)$. Thus

$$\begin{aligned} \mathbb{P}\left(\frac{1}{n^\alpha} \min_{1\leq j\leq k} \sum_{i=1}^n \mathbb{1}(Y_i > M_{n,1}^{(j)}) < \frac{g_2(n)}{C} \right) &\leq k\mathbb{P}\left(\frac{1}{n^\alpha} \sum_{i=1}^n \mathbb{1}(Y_i > M_{n,1}^{(1)}) < \frac{g_2(n)}{C} \right) \\ &= k\mathbb{P}\left(\left(\frac{1}{n^\alpha} \sum_{i=1}^n \mathbb{1}(Y_i > M_{n,1}^{(1)}) \right)^{-1} > \frac{C}{g_2(n)} \right). \end{aligned}$$

By Lemma G.2 (ii), there exists a $C > 0$ such that for all large n, the last probability is smaller than $\frac{\varepsilon}{2k}$. This together with Eq. (S.64) and Eq. (S.65) shows that there exists $C > 0$ such that for all large n,

$$\mathbb{P}\left(\left(\frac{N_n^+}{n^\alpha}\right)^{-1} > \frac{C}{g_2(n)}\right) < \varepsilon. \qquad \square$$

Lemma G.4. *Let $X_1, \ldots, X_n, \ldots$ be a sequence of random variables with common finite lower endpoint x_F. Put $Y_i := X_i - x_F$, and suppose that the Y_i's have common survival function $\overline{F}(x) = x^{-\alpha} l(x)$, where l is slowly varying and $\alpha \in (0,1)$. If Condition A(p) holds for $p = 1$, then*

$$\frac{N_n^-}{n} \xrightarrow{\mathbb{P}} 1.$$

Proof of Lemma G.4. By Lemma G.3 (i), we have

$$\frac{N_n^+}{n^\alpha} = O_p\left(\frac{l(b_n) c_n^{2\delta'\alpha+\alpha-1}}{l(t_n)}\right).$$

Thus, $\frac{N_n^+}{n} = o_p(1)$ as $\frac{l(b_n) c_n^{2\delta'\alpha+\alpha-1}}{l(t_n) n^{1-\alpha}} \to 0$ by Potter's bound and $\log c_n = o(\log n)$. The proof is then completed by noting that

$$1 - \frac{N_n^+}{n} = \frac{N_n^-}{n} \le 1. \qquad \square$$

Proof of Theorem F.4. By Lemma G.4, $\frac{N_n^-}{n} \xrightarrow{\mathbb{P}} 1$ and so $\log(\frac{N_n^-}{n}) \xrightarrow{\mathbb{P}} 0$. Since Eq. (S.40) holds, by the proof of Theorem F.1,

$$\log \frac{M_{n,h}^-}{M_{n,1}^h} = O_p(1).$$

Thus

$$\log \frac{M_{n,h}^{-*}}{M_{n,1}^h} = \log \frac{M_{n,h}^-}{M_{n,1}^h} - \log \frac{N_n^-}{n} = O_p(1) + o_p(1) = O_p(1).$$

Hence

$$\log M_{n,h}^{-*} - h \log M_{n,1} = O_p(1).$$

Since $\log |M_{n,1}^X| - \log M_{n,1} = o_p(1)$, we have

$$\log M_{n,h}^{-*} - h \log |M_{n,1}^X| = O_p(1).$$

Dividing both sides by $\log |M_{n,1}^X|$ gives

$$\frac{\log M_{n,h}^{-*}}{\log |M_{n,1}^X|} = h + O_p\left(\frac{1}{\log n}\right),$$

because $\log |M_{n,1}^X| / \log n \xrightarrow{\mathbb{P}} (1-\alpha)/\alpha$. $\square$

Proof of Theorem F.5. (a) Write

$$\log \frac{M_{n,h}^{+*}}{M_{n,1}^{\imath_+(h)}} = \log \frac{M_{n,h}^+}{d_{n,h}} + \log \frac{d_{n,h}}{d_{n,1}^{\imath_+(h)}} + \imath_+(h) \log \frac{d_{n,1}}{M_{n,1}} + \log \frac{n}{N_n^+}. \qquad \text{[S.66]}$$

By Lemmas G.1 and B.3,

$$\log \frac{M_{n,h}^+}{d_{n,h}} + \imath_+(h) \log \frac{d_{n,1}}{M_{n,1}} = O_p(\log c_n).$$

By Lemma 2.1, for some constant $C' > 0$,

$$\frac{d_{n,h}}{d_{n,1}^{\imath_+(h)}} \sim C' \frac{t_n^{h-\alpha} l(t_n)}{t_n^{\imath_+(h)(1-\alpha)} l(t_n)^{\imath_+(h)}} = C' t_n^{\alpha^2-\alpha} l(t_n)^{1-\imath_+(h)}$$

and thus

$$\begin{aligned} \log \frac{d_{n,h}}{d_{n,1}^{\imath_+(h)}} &= O(1) + (\alpha-1) \log t_n^\alpha + (1 - \imath_+(h)) \log l(t_n) \\ &= (\alpha-1) \log n + O(|\log l(t_n)|) + O(\log c_n). \end{aligned}$$

By Lemma G.3,

$$\log \frac{n^\alpha}{N_n^+} = O_p(\log c_n) + O_p(|\log l(t_n)|) + O_p(|\log l(b_n)|) + O_p(|\log l(\tilde{b}_n)|).$$

Therefore

$$\begin{aligned}\log \frac{d_{n,h}}{d_{n,1}^{\imath_+(h)}} + \log \frac{n}{N_n^+} &= \log \frac{n^\alpha}{N_n^+} + O(|\log l(t_n)|) + O(\log c_n) \\ &= O_p(\log c_n) + O_p(|\log l(t_n)|) + O_p(|\log l(b_n)|) + O_p(|\log l(\tilde{b}_n)|).\end{aligned}$$

Combining the preceding displays gives

$$\log M_{n,h}^{+*} - \imath_+(h) \log M_{n,1} = O_p(L_n).$$

Since $\log |M_{n,1}^X| - \log M_{n,1} = o_p(1)$, it follows that

$$\log M_{n,h}^{+*} - \imath_+(h) \log |M_{n,1}^X| = O_p(L_n).$$

Therefore

$$\frac{\log M_{n,h}^{+*}}{\log |M_{n,1}^X|} - \imath_+(h) = O_p\left(\frac{L_n}{\log |M_{n,1}^X|}\right) = O_p\left(\frac{L_n}{\log n}\right).$$

(b) Since $l(b_n)/l(t_n) \to c_1$ and $l(\tilde{b}_n)/l(t_n) \to c_2$ for some constants $c_1, c_2 > 0$, we have

$$|\log l(b_n)| + |\log l(\tilde{b}_n)| = O(|\log l(t_n)|) + O(1).$$

Hence

$$L_n = O(\log c_n + |\log l(t_n)|).$$

The result follows from part (a). □

H. Threshold Diagnostics for Peaks-over-threshold GPD Fits

This section provides threshold diagnostics for the peaks-over-threshold generalized Pareto distribution (GPD) fits used in the empirical network data analysis. For a high threshold u, the GPD approximation models the conditional excess distribution of $X - u$ given $X > u$. The choice of u involves the usual bias–variance tradeoff: a lower threshold retains more exceedances but may introduce bias because the GPD approximation may not yet be adequate, whereas a higher threshold improves the asymptotic justification but increases estimation variability.

H.1. Threshold diagnostics for the empirical network datasets. Thresholds were selected diagnostically using graphical criteria (11). We considered threshold-stability plots for the fitted GPD shape and modified scale parameters, together with visual agreement between the fitted GPD survival curve and the empirical upper-tail survival curve. For the Wiki Talk and Epinions datasets, these diagnostics gave broadly consistent choices: the selected thresholds were near the lower end of regions where the fitted quantities were comparatively stable, and the fitted GPD survival curves agreed reasonably well with the empirical upper tails. For the DBpedia dataset, the threshold-stability diagnostics were less decisive because the sample is smaller and high thresholds leave relatively few exceedances. We therefore selected the DBpedia threshold by placing more weight on the fitted upper-tail survival curve. The selected thresholds should be interpreted as diagnostic choices rather than universally optimal thresholds. Since the observations are node-derived quantities from finite networks, the GPD fits and the associated confidence intervals should be interpreted as descriptive summaries of the empirical upper tail rather than as formal *i.i.d.* inference.

H.2. Simulation with clustered extremes. As discussed in the main text, the GPD estimates should not be expected to agree exactly with the Hill, least-squares Pareto, or Taylor-law estimates, because these estimates summarize different aspects of the data. We now give a simple simulation showing that such disagreement can occur even when the marginal distribution of the observations has tail index $\alpha < 1$, provided that extreme observations appear in clusters.

Let m be the number of latent clusters and let L be the cluster size, so that $n = mL$. For $j = 1, \ldots, m$, let

$$\mathbb{P}(Z_j > z) = z^{-\alpha}, \qquad z \geq 1,$$

where $0 < \alpha < 1$, and let $Z_1, \ldots, Z_m$ be independent. For $r = 1, \ldots, L$, let $B_{j,r}$ be independent copies of a bounded positive random variable, independent of $Z_1, \ldots, Z_m$. In the simulation below, we take

$$B_{j,r} \sim \text{Uniform}(1 - \delta, 1 + \delta),$$

with a small value of δ. We then define

$$X_{j,r} = Z_j B_{j,r}, \qquad j = 1, \ldots, m, \quad r = 1, \ldots, L.$$

Thus observations in the same cluster share the same heavy-tailed factor Z_j and differ only by a small multiplicative perturbation.

The marginal distribution of $X_{j,r}$ is still regularly varying with tail index α. Indeed, for $x \geq 1 + \delta$,

$$\mathbb{P}(X_{j,r} > x) = \mathbb{E}\left[\mathbb{P}(Z_j > x/B_{j,r} \mid B_{j,r})\right] = x^{-\alpha}\mathbb{E}(B_{j,r}^{\alpha}).$$

Hence the true marginal tail index is α, and the marginal mean is infinite when $\alpha < 1$. However, because large observations occur in clusters, a high-threshold exceedance set may contain many observations but only a small number of distinct latent cluster levels. The effective number of extreme tail observations available to a high-threshold GPD fit may therefore be much smaller than the raw number of exceedances.

We used $n = 120{,}000$, $L = 200$, $\alpha = 0.6$, and $\delta = 0.01$. We then applied the same graphical diagnostics and estimation methods as in the real-data examples in the main text. Figure S5 shows that the estimates in the alternative Hill plot stabilize around $\theta \approx 0.7$, with corresponding Hill estimate 0.609. The tail-index estimate implied by Taylor's law is close to this value, and the least-squares Pareto fit gives 0.601. By contrast, Figure S6 suggests a GPD threshold near $u = 350$. At and above this threshold, the fitted GPD shape parameter is positive and below 1, corresponding to a Pareto-equivalent tail-index estimate $\alpha_{\mathrm{GPD}} = 1/\xi$ greater than 1. The empirical and fitted upper-tail survival plots in Figure S7 also show that the fitted GPD describes the selected upper tail better than the broader Pareto fit, although it does not capture the most extreme observations perfectly.

This simulation illustrates that, for clustered heavy-tailed data, different finite-sample tail diagnostics need not give closely agreeing tail-index estimates. In particular, a high-threshold GPD fit may suggest $\alpha_{\mathrm{GPD}} > 1$ even when the true marginal tail index is $\alpha < 1$, because the extreme exceedances may correspond to only a small number of effective latent tail observations. This behavior is not specific to the particular simulated sample shown here: repeating the simulation with different random seeds produced qualitatively similar discrepancies in additional runs. Because the threshold choice in peaks-over-threshold analysis is partly graphical and diagnostic, we do not attach a numerical frequency to this occurrence. The purpose of the simulation is to show that such disagreement can arise naturally from clustered extremes, rather than to estimate the probability of the phenomenon under this model. Heterogeneity, finite observation windows, and other forms of dependence may further complicate such finite-sample diagnostics in real data.

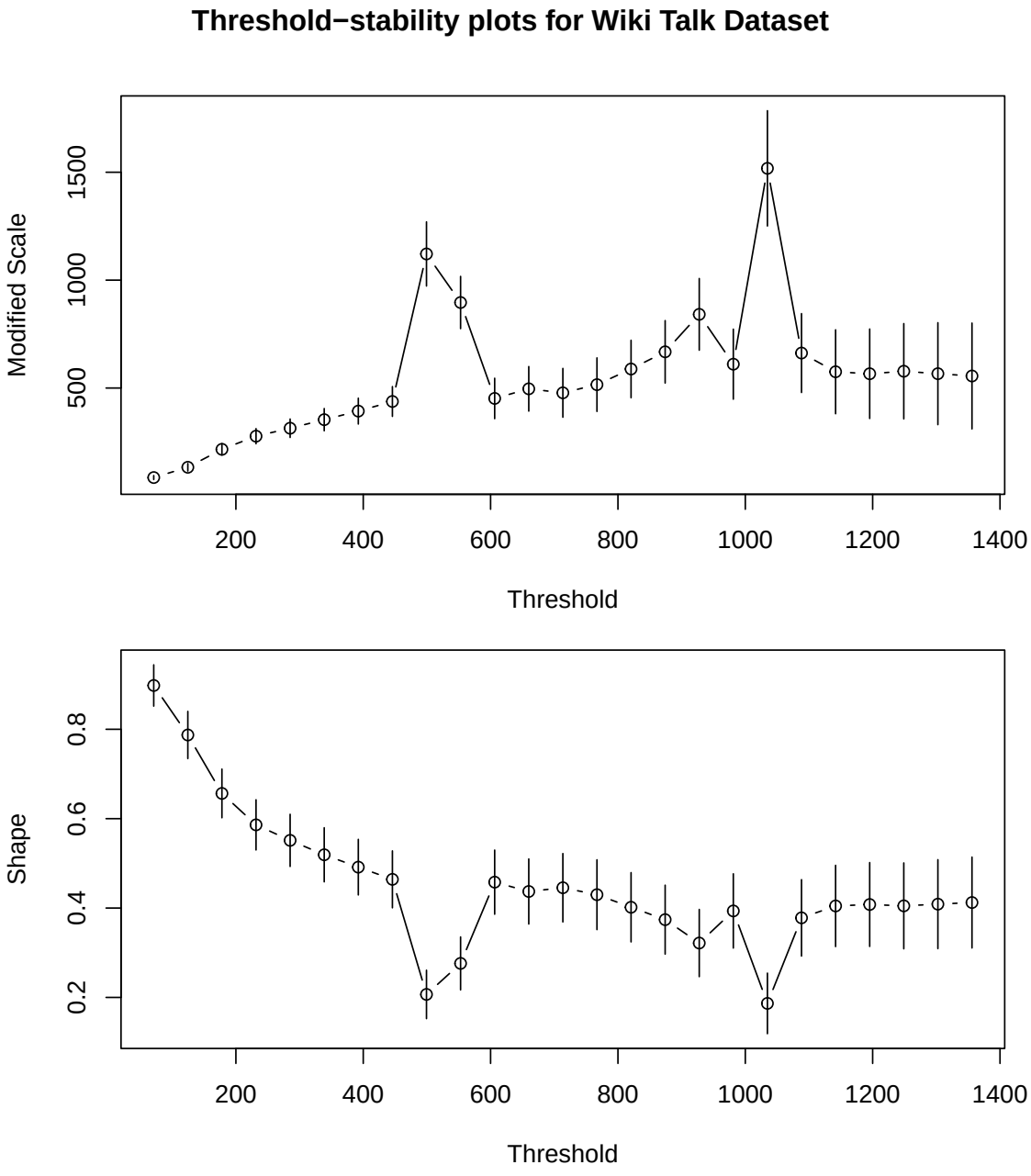


Fig. S1. Threshold-stability diagnostics for the peaks-over-threshold GPD fit to the Wiki Talk dataset. The upper panel shows the fitted modified scale parameter, and the lower panel shows the fitted shape parameter, plotted against the threshold. The selected threshold $u = 600$ lies near the lower end of a region where both fitted quantities are comparatively stable as the threshold increases.

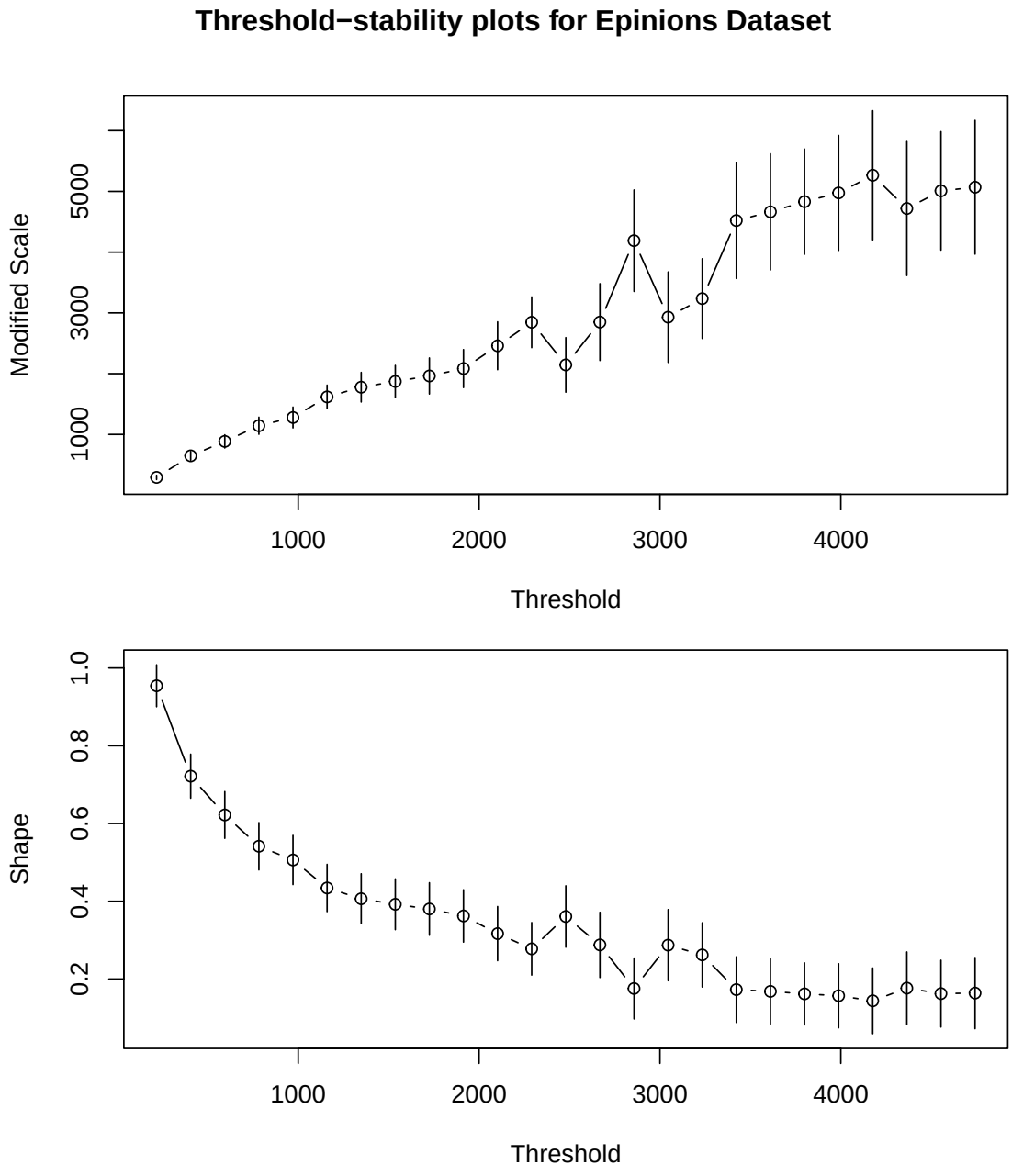


Fig. S2. Threshold-stability diagnostics for the peaks-over-threshold GPD fit to the Epinions dataset. The upper panel shows the fitted modified scale parameter, and the lower panel shows the fitted shape parameter, plotted against the threshold. The selected threshold $u = 3500$ lies near the lower end of a region where both fitted quantities are comparatively stable as the threshold increases.

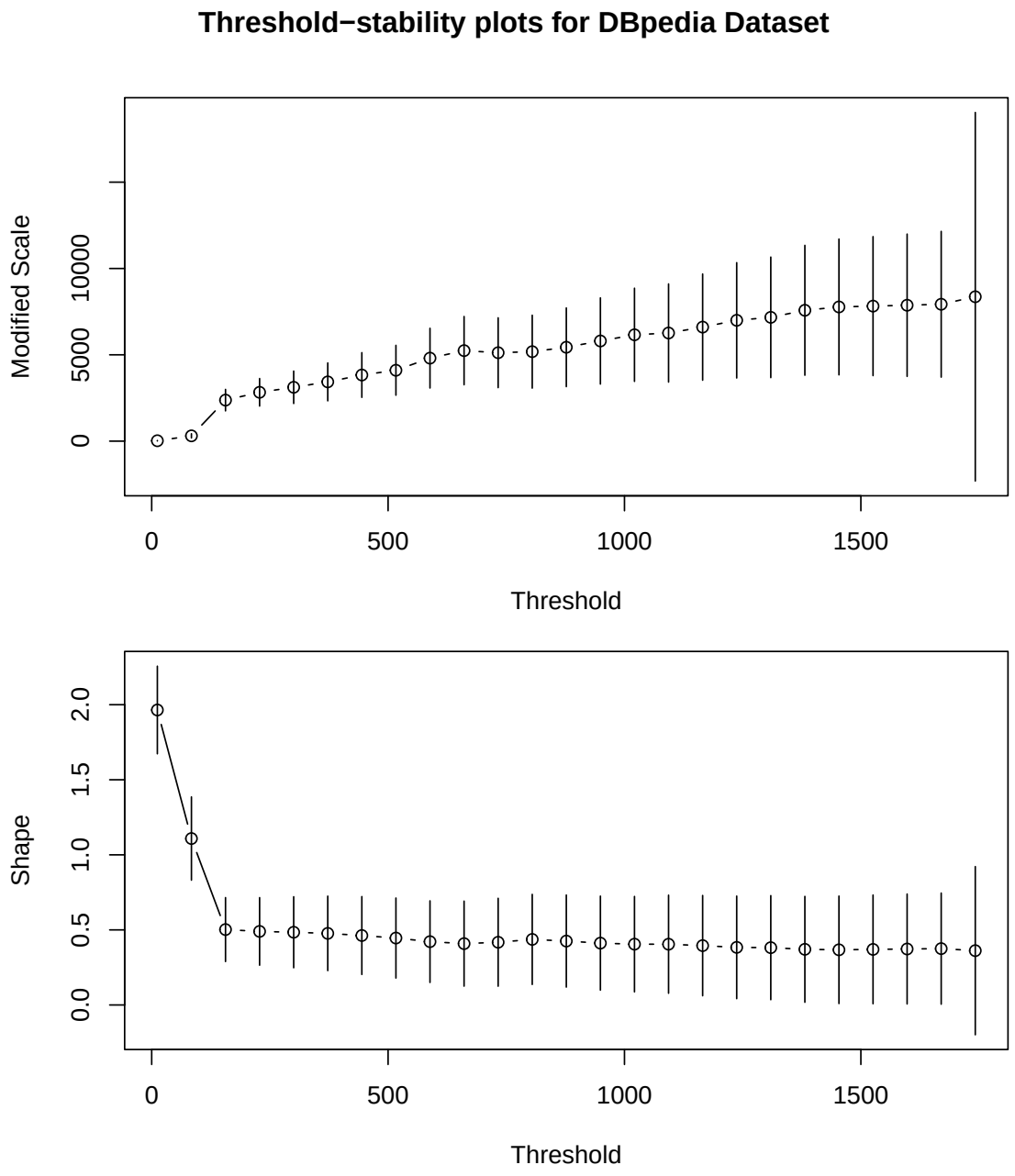
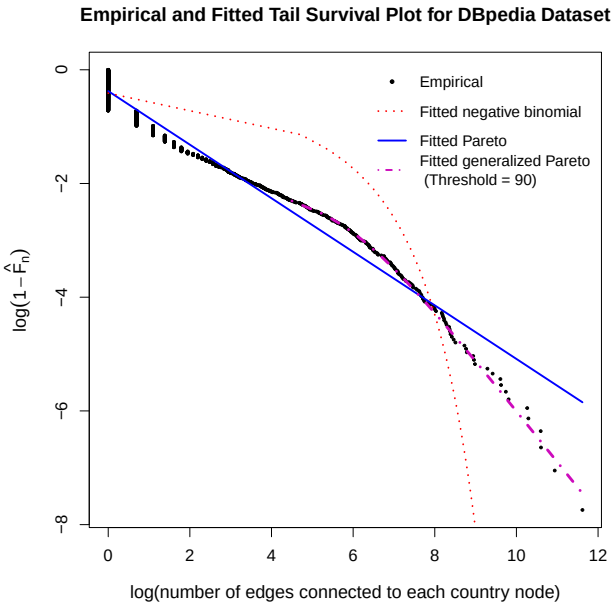
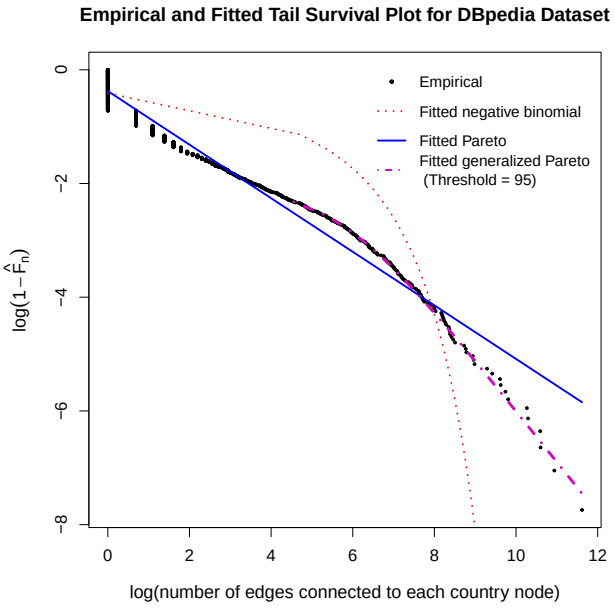


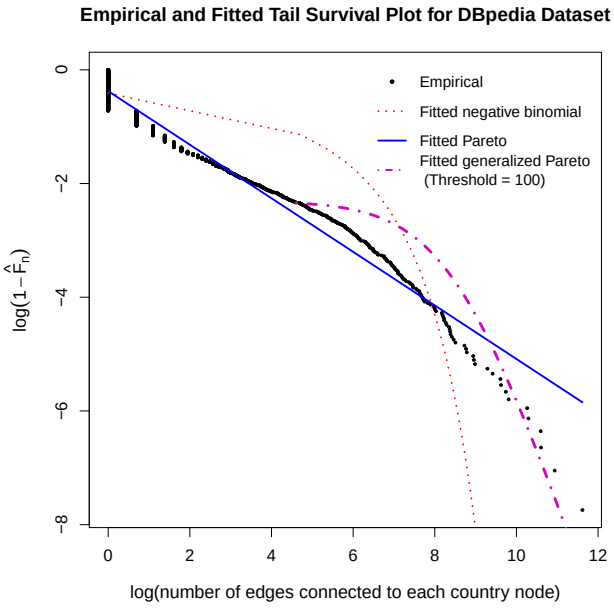
Fig. S3. Threshold-stability diagnostics for the peaks-over-threshold GPD fit to the DBpedia dataset. The upper panel shows the fitted modified scale parameter, and the lower panel shows the fitted shape parameter, plotted against the threshold. The fitted shape parameter becomes comparatively stable for thresholds above roughly 100, whereas the modified scale parameter is less stable. For this smaller dataset, the threshold-stability diagnostics were therefore used together with the fitted upper-tail survival plots in Figure S4.



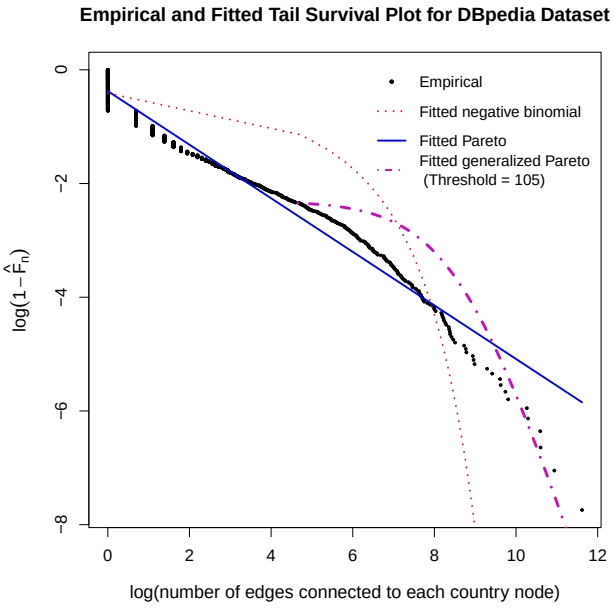
(a) Threshold $u = 90$.



(b) Threshold $u = 95$.



(c) Threshold $u = 100$.



(d) Threshold $u = 105$.

Fig. S4. Comparison of empirical and fitted upper-tail survival curves for the DBpedia dataset under several candidate GPD thresholds. The fitted GPD curve is sensitive to the threshold choice. In this dataset, the lower candidate thresholds shown here give better visual agreement with the empirical upper-tail survival curve while retaining more exceedances for estimation.

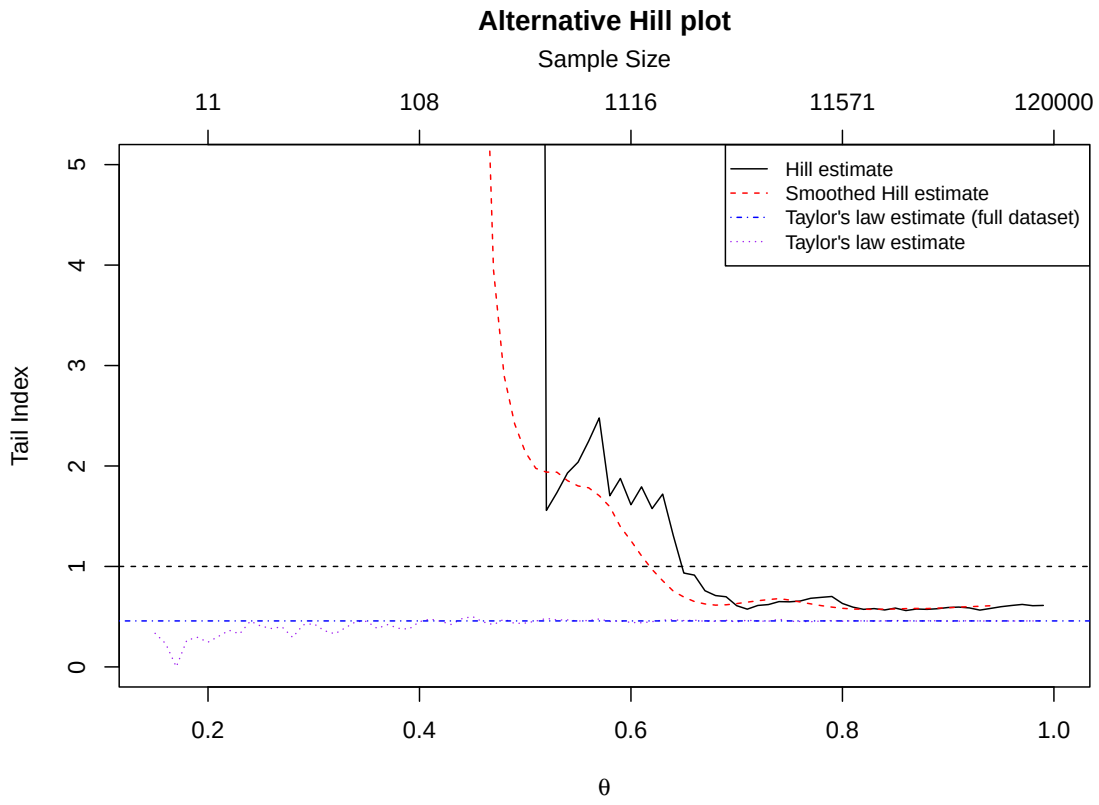


Fig. S5. Alternative Hill plot for the simulated clustered heavy-tailed data. For small values of θ, the Hill estimates are much larger and fall outside the displayed vertical-axis range; the plot therefore focuses on the range where the estimates are more stable. The smoothed Hill estimate appears to stabilize around $\theta = 0.7$.

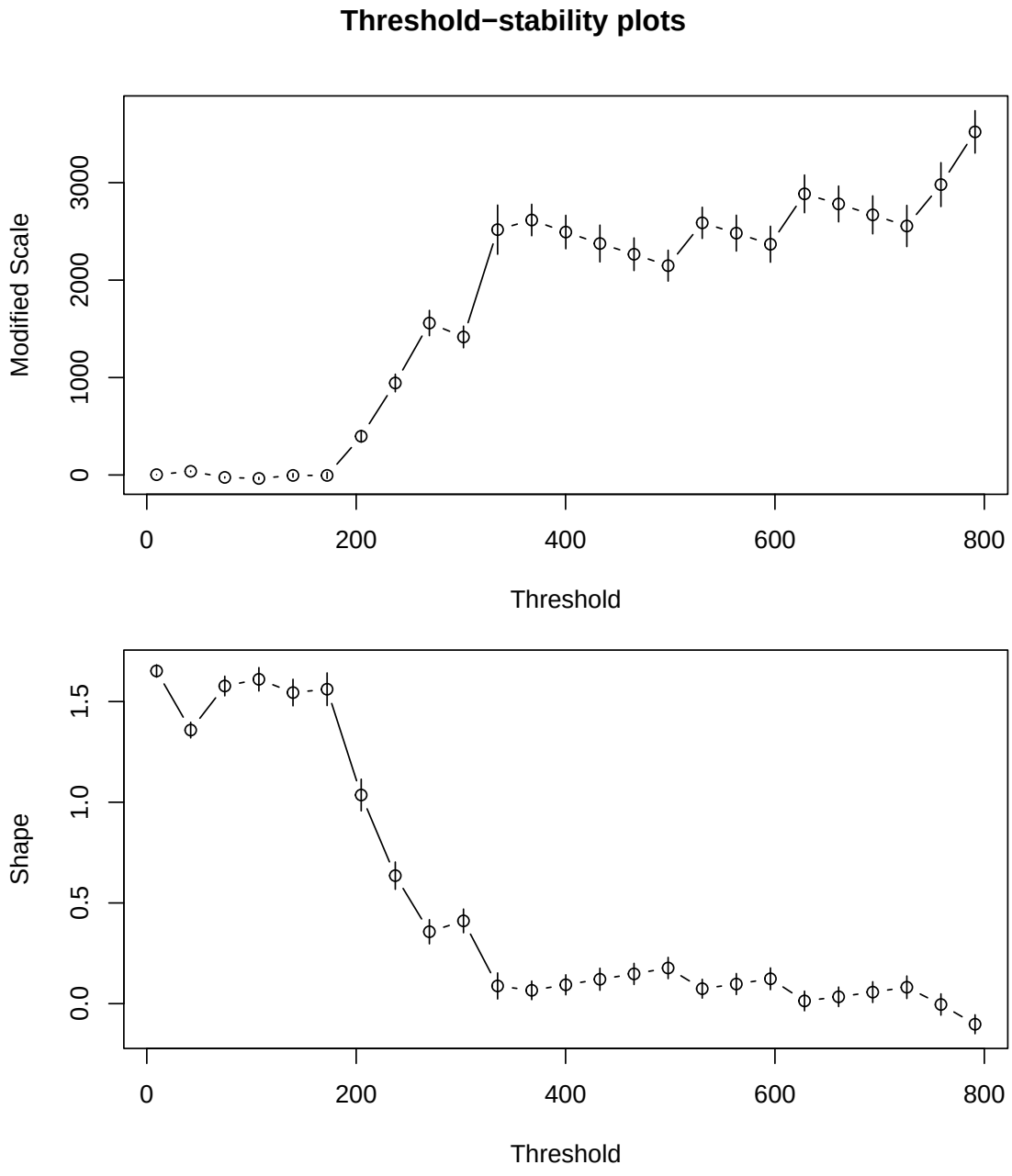


Fig. S6. Threshold-stability diagnostics for the peaks-over-threshold GPD fit to the simulated clustered heavy-tailed data. The upper panel shows the fitted modified scale parameter, and the lower panel shows the fitted shape parameter, plotted against the threshold. The selected threshold $u = 350$ lies near the lower end of a region where both fitted quantities are comparatively stable as the threshold increases.

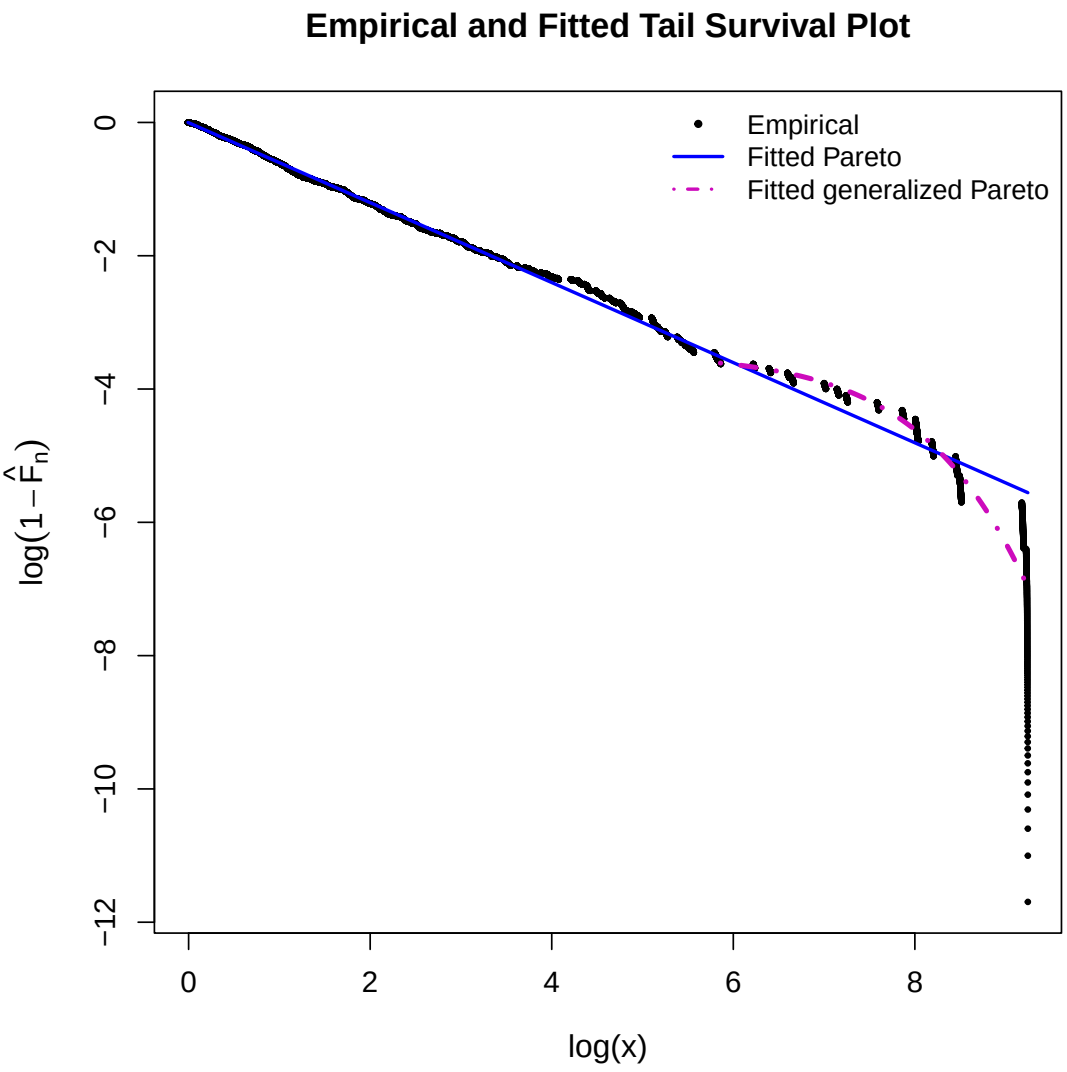


Fig. S7. Empirical and fitted log-survival functions for the simulated clustered heavy-tailed data. The plot compares the empirical upper-tail survival curve with the fitted Pareto survival curve and the generalized Pareto survival curve fitted above the selected threshold $u = 350$ ($\log u = 5.86$).

I. Simulation Studies

The following simulations illustrate the results discussed in the main text. We use the `R` packages `EnvStats`, `stabledist` to simulate heavy-tailed distributions and `ggplot2` to visualize the results.

In each of the following figures, we simulate samples of size $n = 10^p$ for $p = 1, \ldots, 7$. In panels (a)–(d) of each figure, we plot the locations of (a) $(\log(M_{n,1}), \log(V_n))$ for variance, (b) $(\log(M_{n,1}), \log(M^+_{n,2}))$ for upper semivariance, (c) $(\log(M_{n,1}), \log(M^c_{n,3}))$ for third central moment, and (d) $(\log(M_{n,1}), \log(M^{+*}_{n,2}))$ for local upper semivariance. In each plot, three occurrences of $1, 2, \ldots, 7$ represent the location of the ratios with sample size $10^1, 10^2, \ldots, 10^7$ in three different samples, respectively.

The dashed straight line in each plot has a slope equal to the theoretical limit and an intercept estimated from the samples when the slope is fixed to be the theoretical limit. From the logarithmic form of Taylor's law in Eq. (2), it is evident that the slopes represent the asymptotic ratio of the moments being plotted. For example, in each panel (a), where $\frac{\log(V_n))}{\log(M_{n,1})} \xrightarrow{\mathbb{P}} \frac{2-\alpha}{1-\alpha}$, the slope is set to be $b = \frac{2-\alpha}{1-\alpha}$. Theoretically, the points in each graph should lie close to the dashed straight line.

I.1. Taylor's Law for i.i.d. Data. Simulations of i.i.d. heavy-tailed random variables, generated by three different distributions with three different choices of α each, illustrate the asymptotic results of Theorem 3.10 with $h_1 = 2$ and $h_2 = 1$, Theorem F.3, Theorem 3.10 with $h_1 = 3$ and $h_2 = 1$, and Theorem F.5, in the following nine figures.

We observe that the points lie closer to the dashed line when $\alpha = 0.2$ than when $\alpha = 0.5$ and 0.7. Moreover, the points with "7" are also generally closer to the line than the other points, as they are based on the largest sample size.

1. Suppose that $X_1, \ldots, X_n$ are i.i.d. Pareto random variables with location parameter $x_m = 0.5$ and scale parameter α, denoted by Pareto$(0.5, \alpha)$, with survival function $\mathbb{P}(X_i > x) = x^{-\alpha}$ for $i = 1, \ldots, n$, $0 < \alpha < 1$, and $x \geq x_m$. Figures S8, S9, and S10 show the results when $\alpha = 0.2, 0.5$, and 0.8, respectively. We employ the function `rpareto` in the **R** package **EnvStats** to simulate Pareto random variables.

2. As a special case of the survival function $\overline{F}(x) = x^{-\alpha} l(x)$, where l is slowly varying and $\alpha > 0$, we define the cumulative distribution function

$$F_{1,\alpha}(x) = 1 - x^{-\alpha} \log(x) e\alpha, \quad x \geq \exp(\alpha^{-1}). \tag{S.67}$$

   Let $x_{m,\alpha} := \exp(\alpha^{-1})$. We used the acceptance-rejection method (12) to simulate random variables with the distribution function $F_{1,\alpha}$. Specifically, a Pareto distribution with the same location parameter $x_{m,\alpha}$ and a shape parameter of $\alpha - 0.05$ (i.e., Pareto$(x_{m,\alpha}, \alpha - 0.05)$) was used as the proposal distribution. This ensures that the two distributions have the same support, while the proposal distribution has a heavier tail than $F_{1,\alpha}$.

   Figures S11, S12, and S13 show the results when $\alpha = 0.2, 0.5$, and 0.8, respectively.

3. Consider the stable distribution $F(c, \alpha)$ on the nonnegative half line with Laplace transform

$$\mathcal{L}(s) = \mathbb{E}(e^{-(sX)}) = e^{-(cs)^\alpha}, \quad s \geq 0, \quad c > 0, \quad 0 < \alpha < 1.$$

   According to the tail approximation of the stable distribution (13), the survival function of $X \sim F(c, \alpha)$ satisfies $\mathbb{P}(X > x) \sim Cx^{-\alpha}$ for some constant $C > 0$. Thus, it has a tail index of α. Figures S14, S15, and S16 present the results of simulations for $\alpha = 0.2$, 0.5, and 0.8. We used the `rstable` function from the **R** package **stabledist** to generate random variables following the stable distribution.

I.2. Taylor's Law for an AR(1) Model. We simulated two cases of the first-order autoregressive model discussed in Section 3.3, Theorem 3.9:

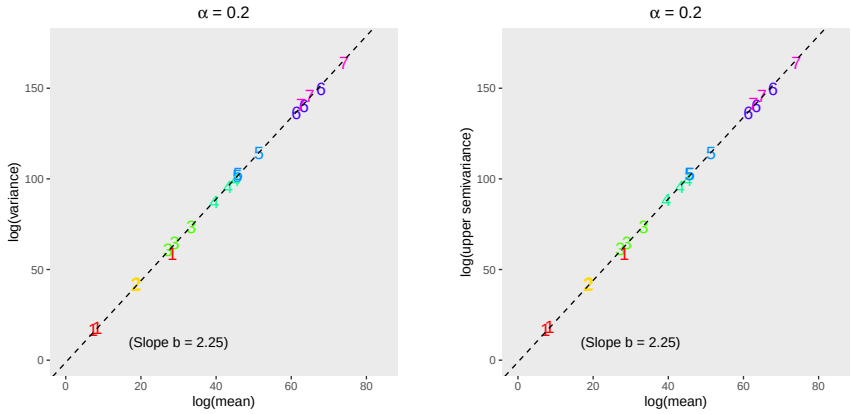
$$X_t = \beta_1 X_{t-1} + \epsilon_t$$

for $t = 1, \ldots, n$, $\beta_1 \in (0, 1)$. In the first case, $\epsilon_t \sim Pareto(1, \alpha)$. In the second case, $\epsilon_t \sim F(1, \alpha)$ with $\alpha = 0.5$. Both distributions satisfy the conditions given in Theorem 3.9. Indeed, the Pareto distribution has a decreasing density and any stable distribution is unimodal (13), and both densities are also bounded.

Figures S17 and S18 show the results of simulations of AR(1) models when the noise follows Pareto and stable distributions, respectively, with $\alpha = 0.5$, and $\beta_1 = 0.8$.

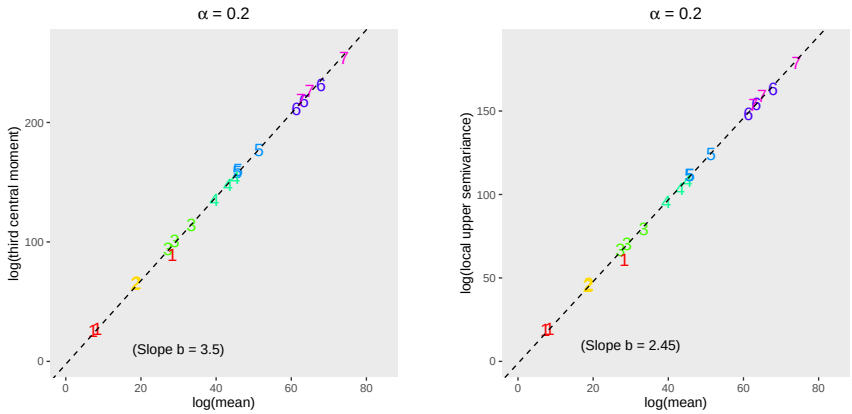
The sample size was $n = 10^p$ for $p = 1, 2, \ldots, 7$.

When there is autoregressive dependence in the simulated time series, the points in the figures lie further from the straight dashed line than when the random variables are i.i.d., suggesting that, even though our results hold for weak dependence asymptotically, the asymptotic results may not be very accurate for finite samples.



(a) Taylor's law

(b) Taylor's law for upper semivariance



(c) Taylor's law for third central moment

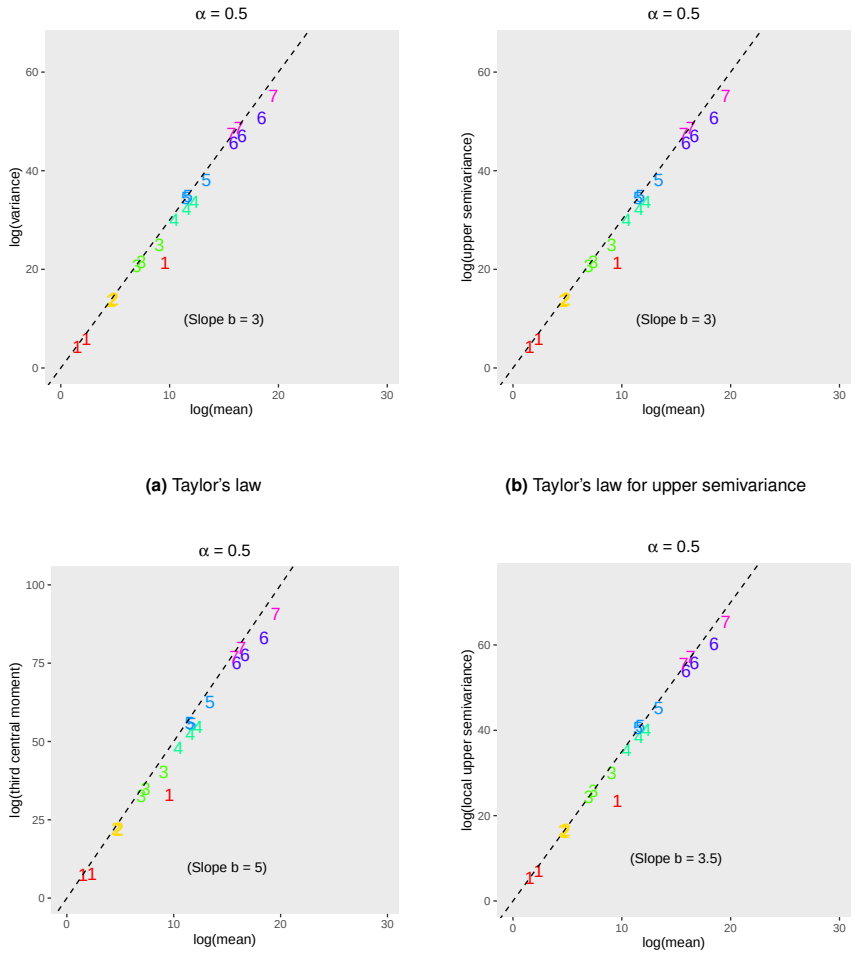
(d) Taylor's law for local upper semivariance

Fig. S8. Scatterplots of (a) $(\log(M_{n,1}), \log(V_n))$, (b) $(\log(M_{n,1}), \log(M^{+}_{n,2}))$, (c) $(\log(M_{n,1}), \log(M^{c}_{n,3}))$, and (d) $(\log(M_{n,1}), \log(M^{+*}_{n,2}))$, corresponding to Theorem 3.10 with $h_1 = 2$ and $h_2 = 1$, Theorem F.3, Theorem 3.10 with $h_1 = 3$ and $h_2 = 1$, and Theorem F.5, respectively, when $X_i \overset{\text{i.i.d.}}{\sim} \text{Pareto}(0.5, 0.2)$ with sample size 10^p for $p = 1, 2, \ldots, 7$.

I.3. Taylor's Law with Heterogeneous Data. Here we simulate the setting described in Section 4.1. Given a value of $p^* \in (0, 1)$, we first generate u_n from a binomial distribution with n trials and a probability of success p^* on each trial. Next, we generate u_n random variables $X_{i,U}$ for $i = 1, \ldots, u_n$, each with a survival function $\overline{F}_U(x) = x^{-\alpha_1} l_1(x)$, where l_1 is slowly varying and $\alpha_1 \in (0, 1)$. Similarly, we generate $n - u_n$ random variables $X_{i,V}$ for $i = 1, \ldots, n - u_n$, each with a survival function $\overline{F}_V(x) = x^{-\alpha_2} l_2(x)$, where l_2 is slowly varying and $\alpha_2 > \alpha_1$. Consequently,

$$\lim_{x \to \infty} \frac{\mathbb{P}(X_{1,V} > x)}{\mathbb{P}(X_{1,U} > x)} = 0.$$

For each simulation setting, we specify the value of p^*, the distributions, and the corresponding values of α_1 and α_2 in the captions of Figures S19–S22. The straight dashed line in each panel represents the corresponding limit of the smaller value among α_1 and α_2, in agreement with Theorem 4.2. In general, the points lie close to the line. Figures S19 and S20 show the simulation results from heterogeneous heavy-tailed data generated by Pareto and stable distributions, respectively, with $\alpha_1 = 0.5$ and $\alpha_2 = 0.8$. Figures S21 and S22 show the simulation results from heterogeneous heavy-tailed data generated from a stable distribution and the distribution function $F_{1,\alpha}$ defined in Eq. (S.67), with $\alpha_1 = 0.2$ and $\alpha_2 = 0.7$. The asymptotic results depend only on α_1, the smaller α, in the heterogeneous case. Hence there is only one straight dashed line in each panel of these figures. In all these figures, the points lie close to or on the line in almost all settings, suggesting that the asymptotic theory usefully approximates the behavior of our finite samples.



(a) Taylor's law

(b) Taylor's law for upper semivariance

(c) Taylor's law for third central moment

(d) Taylor's law for local upper semivariance

Fig. S9. Scatterplots of (a) $(\log(M_{n,1}), \log(V_n))$, (b) $(\log(M_{n,1}), \log(M^{+}_{n,2}))$, (c) $(\log(M_{n,1}), \log(M^{c}_{n,3}))$, and (d) $(\log(M_{n,1}), \log(M^{+*}_{n,2}))$, corresponding to Theorem 3.10 with $h_1 = 2$ and $h_2 = 1$, Theorem F.3, Theorem 3.10 with $h_1 = 3$ and $h_2 = 1$, and Theorem F.5, respectively, when $X_i \overset{\text{i.i.d.}}{\sim}$ Pareto$(0.5, 0.5)$ with sample size 10^p for $p = 1, 2, \ldots, 7$.

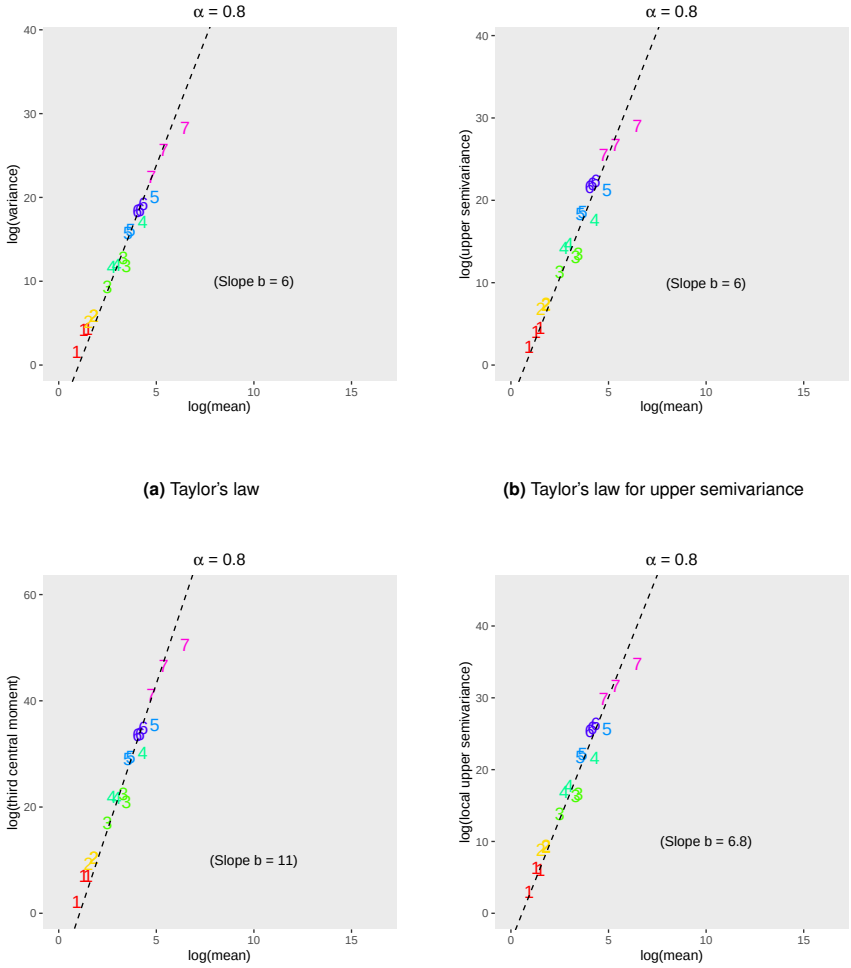
I.4. Taylor's Law with Correlated Heavy-tailed Data. Simulations of correlated heavy-tailed data specified in Examples B.6 and B.7 yielded Figures S23 to S24. The asymptotic ratio of moments describes the finite-sample simulations well.

I.5. Taylor's Law for Heavy-tailed Network Data. Here we illustrate Taylor's law using network data. We generate random graphs based on the Erdős–Rényi–Gilbert model (14). In this model, each edge has a fixed probability of being present or absent, independently of the other edges. These random graphs are generated with the `erdos.renyi.game` function from the R package **igraph**. The edges for each node are generated through independent Bernoulli trials with a success probability p, that is, each pair of nodes has a probability p of being joined by an edge. We illustrate these random graphs with two examples in Figure S25, each containing 20 nodes with $p = 0.2$.

Define $\Xi^n_{j,1}$ to be the set of nodes in the graph that are at distance 1 from node j, i.e., $\Xi^n_{j,1} = \{k : d(j,k) = 1, 1 \le k \le n\}$, where the distance between two nodes is defined as the length of the shortest path connecting them. We associate each node with a random variable defined via:

$$X_j := Z_j + \frac{\sum_{k \in \Xi^n_{j,1}} Z_k}{|\Xi^n_{j,1}|}, \qquad [S.68]$$

where Z_j are the i.i.d. heavy-tailed random variables. We consider $n = 100, 500, 1000, 5000, 10000$ with $p = \frac{10}{n-1}$ for $n = 100$ and 500, and $p = \frac{100}{n-1}$ for $n = 1000, 5000$, and 10000. We also limit the maximum number of edges to 20 when $n = 100$ or 500 and to 200 when $n = 1000, 5000$, or 10000. These bounds are larger than twice the expected number of edges. The actual



(a) Taylor's law

(b) Taylor's law for upper semivariance

(c) Taylor's law for third central moment

(d) Taylor's law for local upper semivariance

Fig. S10. Scatterplots of (a) $(\log(M_{n,1}), \log(V_n))$, (b) $(\log(M_{n,1}), \log(M_{n,2}^{+}))$, (c) $(\log(M_{n,1}), \log(M_{n,3}^{c}))$, and (d) $(\log(M_{n,1}), \log(M_{n,2}^{+*}))$, corresponding to Theorem 3.10 with $h_1 = 2$ and $h_2 = 1$, Theorem F.3, Theorem 3.10 with $h_1 = 3$ and $h_2 = 1$, and Theorem F.5, respectively, when $X_i \overset{\text{i.i.d.}}{\sim} \text{Pareto}(0.5, 0.8)$ with sample size 10^p for $p = 1, 2, \ldots, 7$.

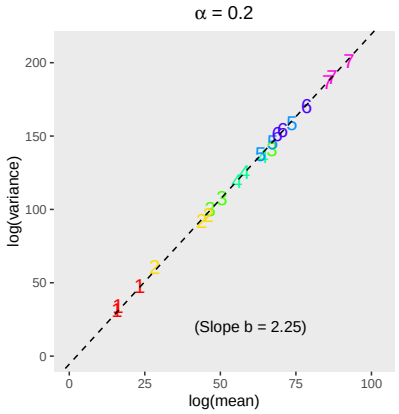
number of edges is unlikely to exceed these bounds for large n. If the number of simulated edges exceeds these bounds, we regenerate another random graph until the constraint is satisfied.

Under this setting, if $d(X_i, X_j) \geq 3$, then $\text{Cov}(X_i, X_j) = 0$, since $X_i = Z_i + \frac{\sum_{k \in \Xi_{i,1}^n} Z_k}{|\Xi_{i,1}^n|}$ and $X_j = Z_j + \frac{\sum_{k \in \Xi_{j,1}^n} Z_k}{|\Xi_{j,1}^n|}$ have no overlapping Z_k for $k \in \mathbb{N}$. On the other hand, we restrict $|\Xi_{i,1}^n| \leq K$ for a positive constant K across different cases. Thus, this setting satisfies the two conditions in Example 4.3, and hence the covariance condition Eq. (23) holds.

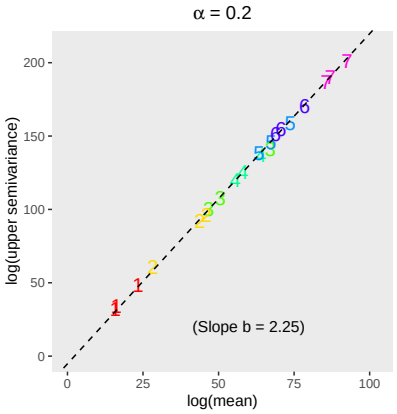
If each edge (j, k) were associated with a positive weight, $w(j, k) > 0$, then a natural generalization would be to define

$$X_j := Z_j + \frac{\sum_{k \in \Xi_{j,1}^n} w(j,k) Z_k}{\sum_{k \in \Xi_{j,1}^n} w(j,k)}.$$

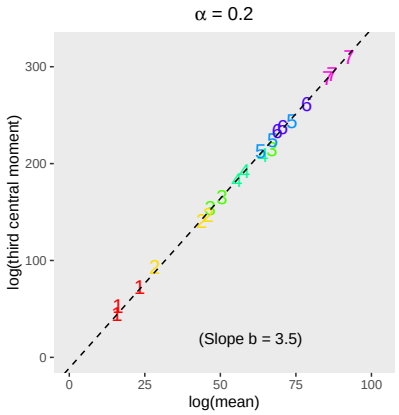
Figures S26–S31 illustrate Taylor's laws when the heavy-tailed data are connected by a network graph, and Z_i's are generated from Pareto or stable distributions, with $\alpha = 0.2, 0.5$, and 0.8. In general, under these network settings, the theoretical asymptotic results of Taylor's laws fit the finite-sample heavy-tailed network data quite well.



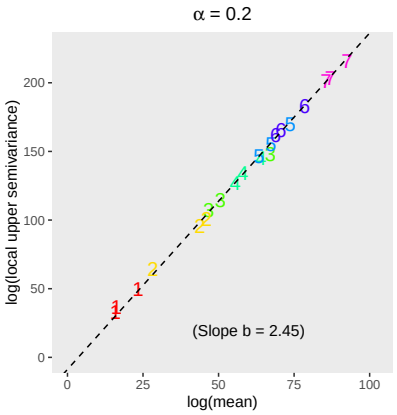
(a) Taylor's law



(b) Taylor's law for upper semivariance

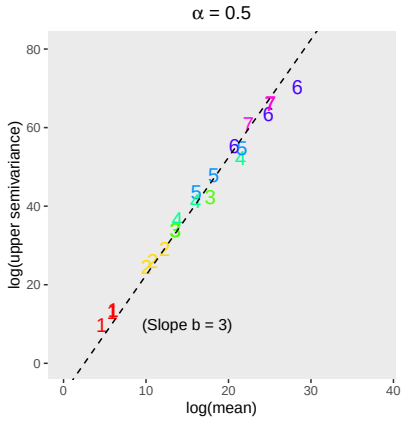


(c) Taylor's law for third central moment

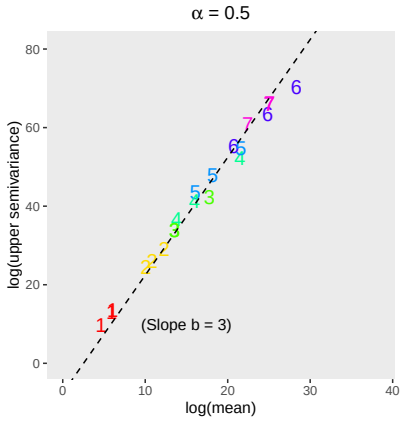


(d) Taylor's law for local upper semivariance

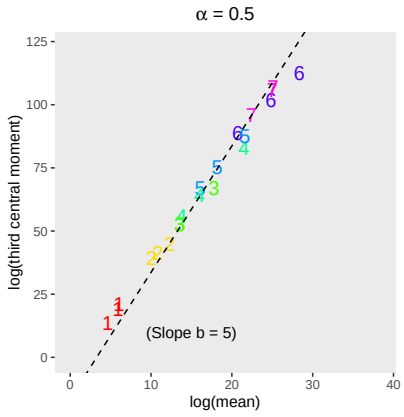
Fig. S11. Scatterplots of (a) $(\log(M_{n,1}), \log(V_n))$, (b) $(\log(M_{n,1}), \log(M_{n,2}^{+}))$, (c) $(\log(M_{n,1}), \log(M_{n,3}^{c}))$, and (d) $(\log(M_{n,1}), \log(M_{n,2}^{+*}))$, corresponding to Theorem 3.10 with $h_1 = 2$ and $h_2 = 1$, Theorem F.3, Theorem 3.10 with $h_1 = 3$ and $h_2 = 1$, and Theorem F.5, respectively, when $F_{1,0.2}$ in Eq. (S.67) is the distribution function of i.i.d. X_i with sample size 10^p for $p = 1, 2, \ldots, 7$.



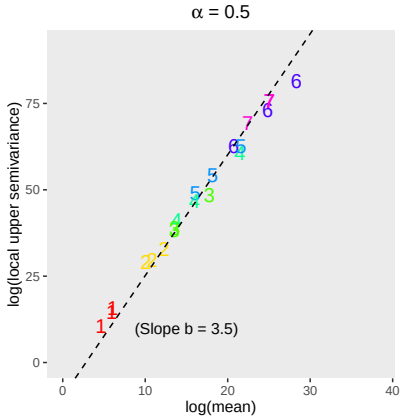
(a) Taylor's law



(b) Taylor's law for upper semivariance

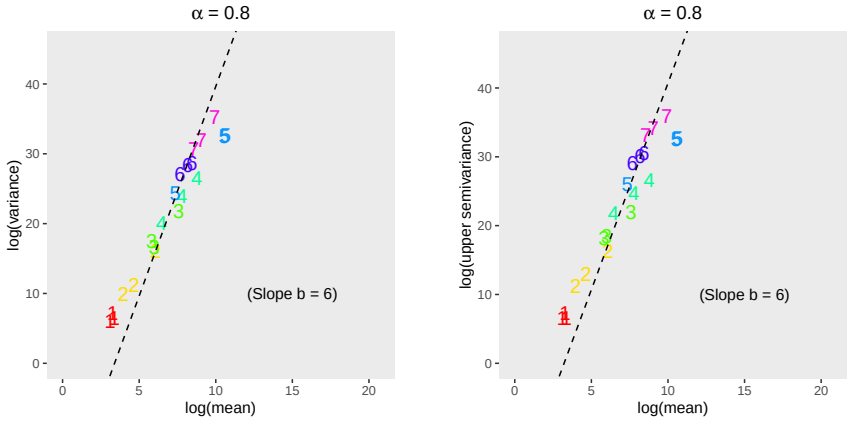


(c) Taylor's law for third central moment



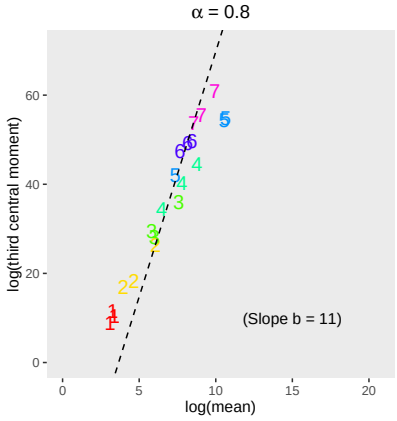
(d) Taylor's law for local upper semivariance

Fig. S12. Scatterplots of (a) $(\log(M_{n,1}), \log(V_n))$, (b) $(\log(M_{n,1}), \log(M^{+}_{n,2}))$, (c) $(\log(M_{n,1}), \log(M^{c}_{n,3}))$, and (d) $(\log(M_{n,1}), \log(M^{+*}_{n,2}))$, corresponding to Theorem 3.10 with $h_1 = 2$ and $h_2 = 1$, Theorem F.3, Theorem 3.10 with $h_1 = 3$ and $h_2 = 1$, and Theorem F.5, respectively, when $F_{1,0.5}$ in Eq. (S.67) is the distribution function of i.i.d. X_i with sample size 10^p for $p = 1, 2, \ldots, 7$.

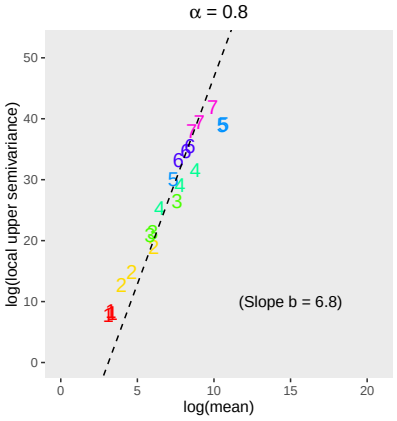


(a) Taylor's law

(b) Taylor's law for upper semivariance

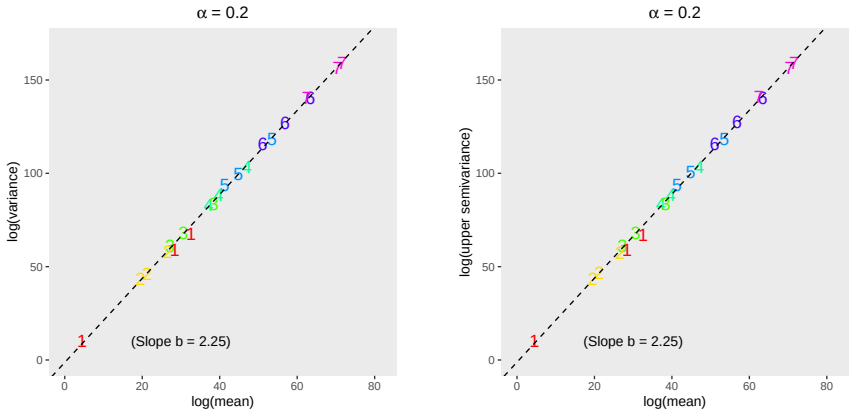


(c) Taylor's law for third central moment



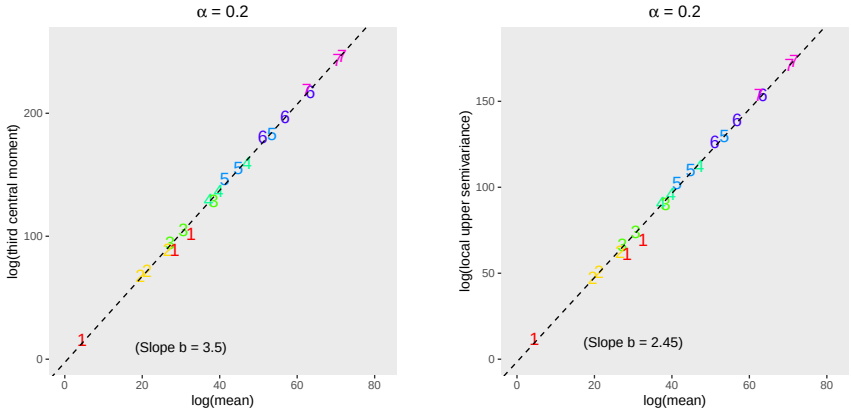
(d) Taylor's law for local upper semivariance

Fig. S13. Scatterplots of (a) $(\log(M_{n,1}), \log(V_n))$, (b) $(\log(M_{n,1}), \log(M_{n,2}^{+}))$, (c) $(\log(M_{n,1}), \log(M_{n,3}^{c}))$, and (d) $(\log(M_{n,1}), \log(M_{n,2}^{+*}))$, corresponding to Theorem 3.10 with $h_1 = 2$ and $h_2 = 1$, Theorem F.3, Theorem 3.10 with $h_1 = 3$ and $h_2 = 1$, and Theorem F.5, respectively, when $F_{1,0.8}$ in Eq. (S.67) is the distribution function of i.i.d. X_i with sample size 10^p for $p = 1, 2, \ldots, 7$.



(a) Taylor's law

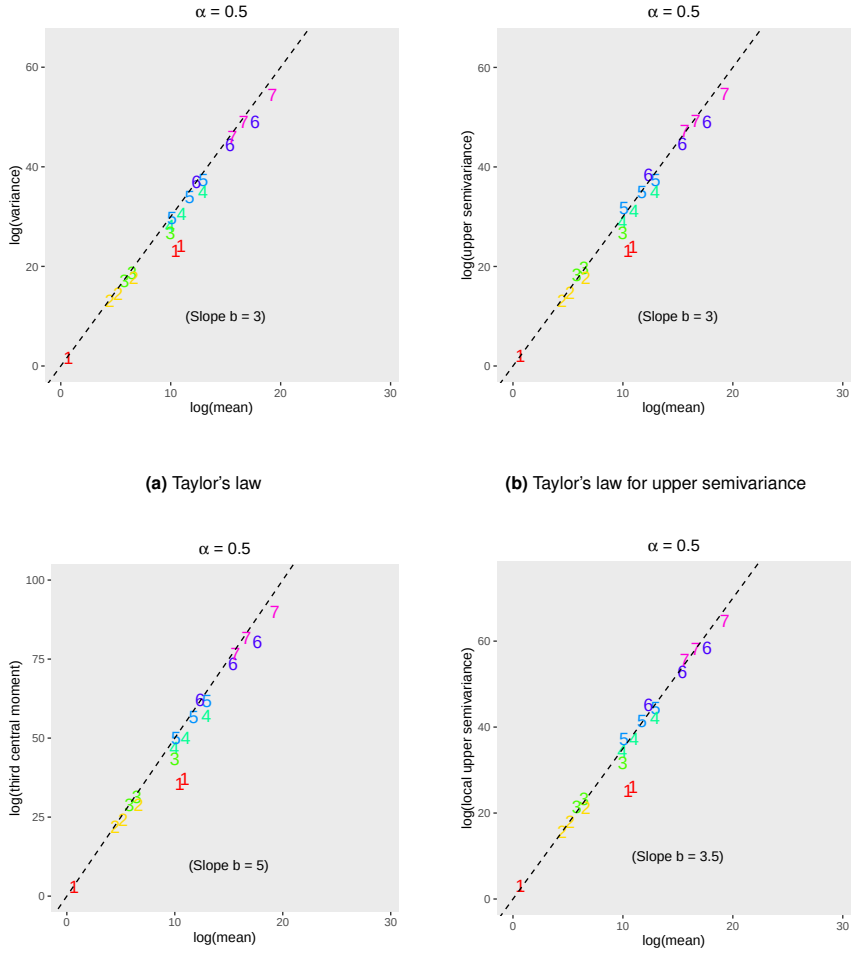
(b) Taylor's law for upper semivariance



(c) Taylor's law for third central moment

(d) Taylor's law for local upper semivariance

Fig. S14. Scatterplots of (a) $(\log(M_{n,1}), \log(V_n))$, (b) $(\log(M_{n,1}), \log(M^{+}_{n,2}))$, (c) $(\log(M_{n,1}), \log(M^{c}_{n,3}))$, and (d) $(\log(M_{n,1}), \log(M^{+*}_{n,2}))$, corresponding to Theorem 3.10 with $h_1 = 2$ and $h_2 = 1$, Theorem F.3, Theorem 3.10 with $h_1 = 3$ and $h_2 = 1$, and Theorem F.5, respectively, when $X_i \overset{\text{i.i.d.}}{\sim} F(1, 0.2)$ with sample size 10^p for $p = 1, 2, \ldots, 7$.



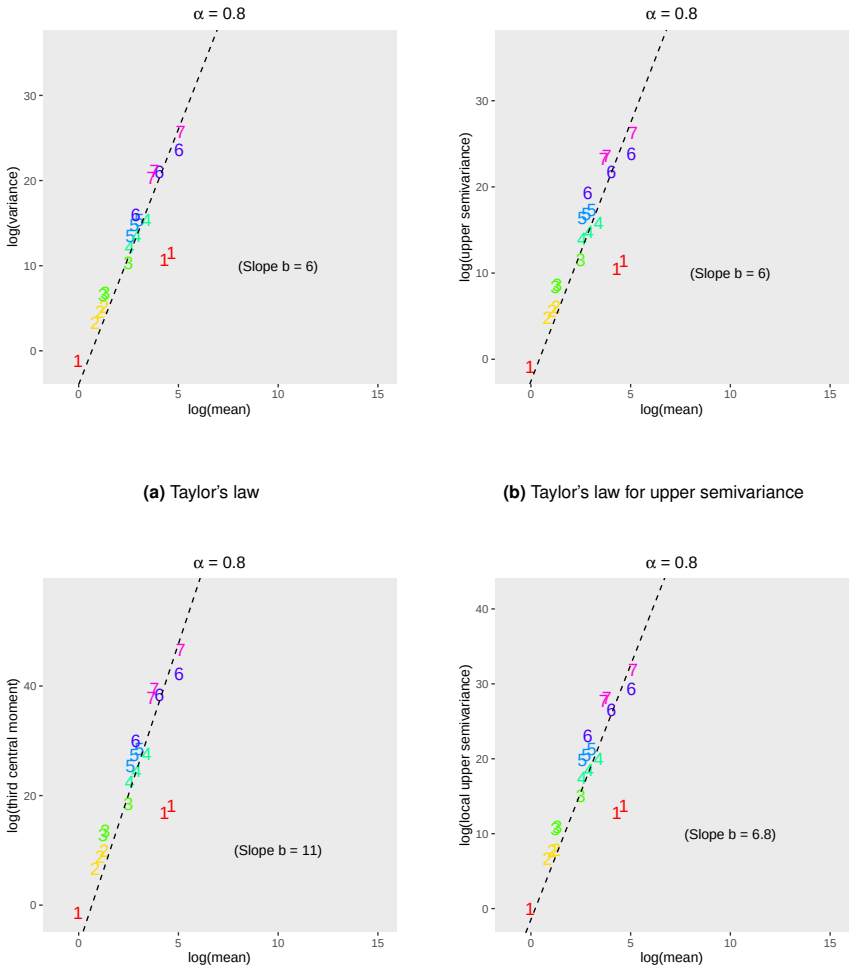
(a) Taylor's law

(b) Taylor's law for upper semivariance

(c) Taylor's law for third central moment

(d) Taylor's law for local upper semivariance

Fig. S15. Scatterplots of (a) $(\log(M_{n,1}), \log(V_n))$, (b) $(\log(M_{n,1}), \log(M_{n,2}^{+}))$, (c) $(\log(M_{n,1}), \log(M_{n,3}^{c}))$, and (d) $(\log(M_{n,1}), \log(M_{n,2}^{+*}))$, corresponding to Theorem 3.10 with $h_1 = 2$ and $h_2 = 1$, Theorem F.3, Theorem 3.10 with $h_1 = 3$ and $h_2 = 1$, and Theorem F.5, respectively, when $X_i \overset{\text{i.i.d.}}{\sim} F(1, 0.5)$ with sample size 10^p for $p = 1, 2, \ldots, 7$.



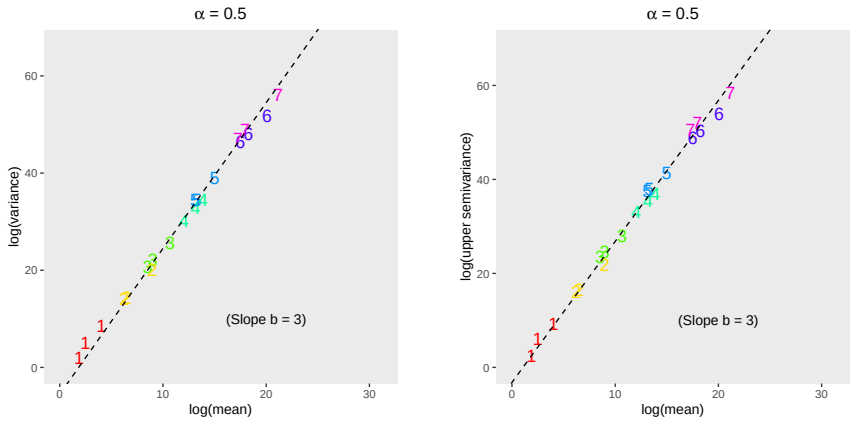
(a) Taylor's law

(b) Taylor's law for upper semivariance

(c) Taylor's law for third central moment

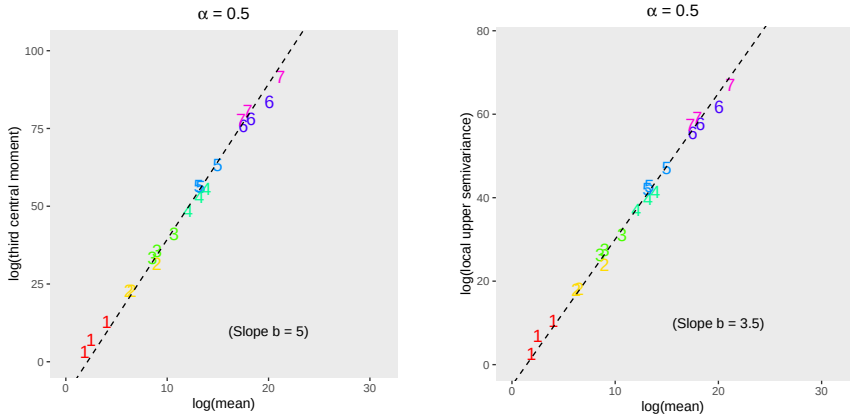
(d) Taylor's law for local upper semivariance

Fig. S16. Scatterplots of (a) $(\log(M_{n,1}), \log(V_n))$, (b) $(\log(M_{n,1}), \log(M_{n,2}^{+}))$, (c) $(\log(M_{n,1}), \log(M_{n,3}^{c}))$, and (d) $(\log(M_{n,1}), \log(M_{n,2}^{+*}))$, corresponding to Theorem 3.10 with $h_1 = 2$ and $h_2 = 1$, Theorem F.3, Theorem 3.10 with $h_1 = 3$ and $h_2 = 1$, and Theorem F.5, respectively, when $X_i \overset{\text{i.i.d.}}{\sim} F(1, 0.8)$ with sample size 10^p for $p = 1, 2, \ldots, 7$.



(a) Taylor's law

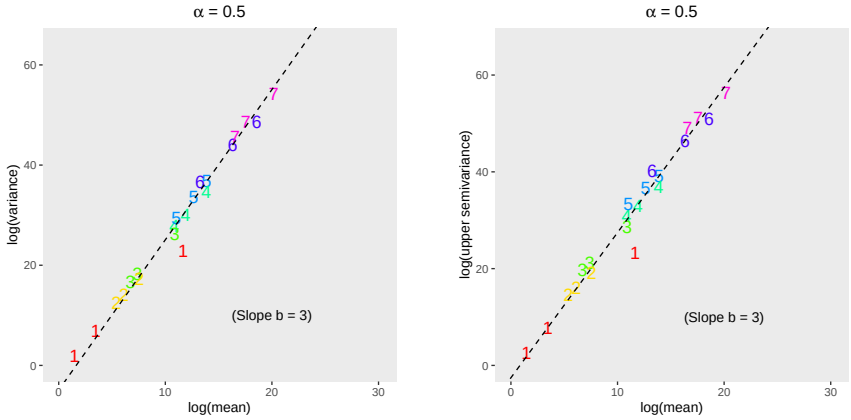
(b) Taylor's law for upper semivariance



(c) Taylor's law for third central moment

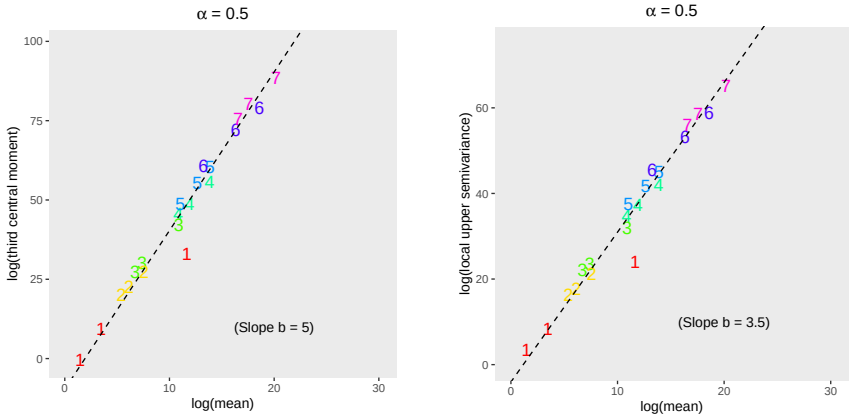
(d) Taylor's law for local upper semivariance

Fig. S17. Scatterplots of (a) $(\log(M_{n,1}), \log(V_n))$, (b) $(\log(M_{n,1}), \log(M_{n,2}^{+}))$, (c) $(\log(M_{n,1}), \log(M_{n,3}^{c}))$, and (d) $(\log(M_{n,1}), \log(M_{n,2}^{+*}))$, corresponding to Theorem 3.10 with $h_1 = 2$ and $h_2 = 1$, Theorem F.3, Theorem 3.10 with $h_1 = 3$ and $h_2 = 1$, and Theorem F.5, respectively, when X_i follows an AR(1) model with $\beta_1 = 0.8$, and $\epsilon_i \overset{\text{i.i.d.}}{\sim} \text{Pareto}(1, 0.5)$ with sample size 10^p for $p = 1, 2, \ldots, 7$.



(a) Taylor's law

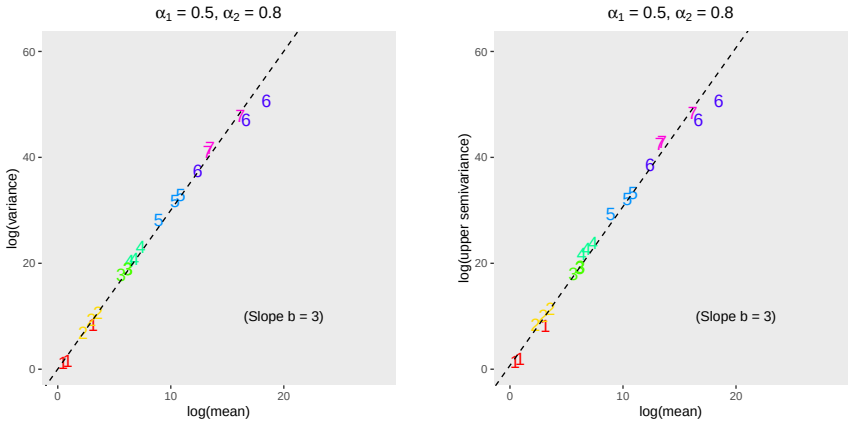
(b) Taylor's law for upper semivariance



(c) Taylor's law for third central moment

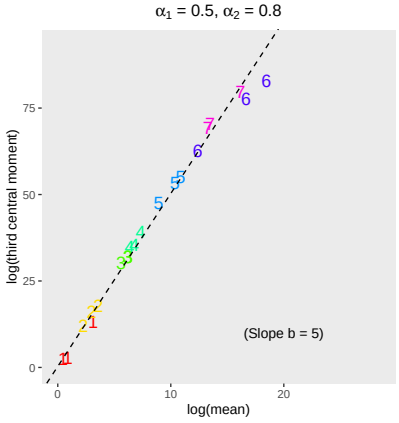
(d) Taylor's law for local upper semivariance

Fig. S18. Scatterplots of (a) $(\log(M_{n,1}), \log(V_n))$, (b) $(\log(M_{n,1}), \log(M^{+}_{n,2}))$, (c) $(\log(M_{n,1}), \log(M^{c}_{n,3}))$, and (d) $(\log(M_{n,1}), \log(M^{+*}_{n,2}))$, corresponding to Theorem 3.10 with $h_1 = 2$ and $h_2 = 1$, Theorem F.3, Theorem 3.10 with $h_1 = 3$ and $h_2 = 1$, and Theorem F.5, respectively, when X_i follows an AR(1) model with $\beta_1 = 0.8$, and $\epsilon_i \overset{\text{i.i.d.}}{\sim} F(1, 0.5)$ with sample size 10^p for $p = 1, 2, \ldots, 7$.

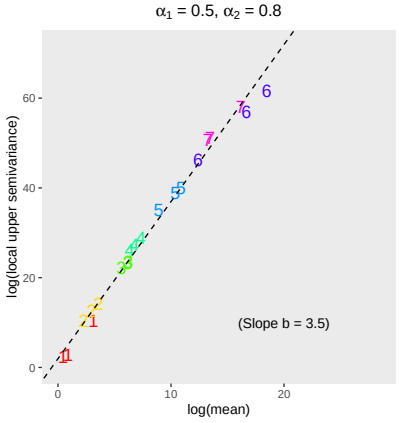


(a) Taylor's law

(b) Taylor's law for upper semivariance

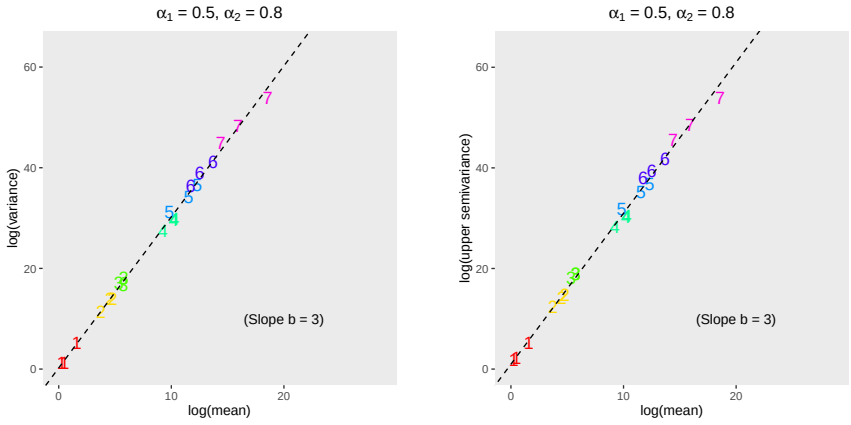


(c) Taylor's law for third central moment



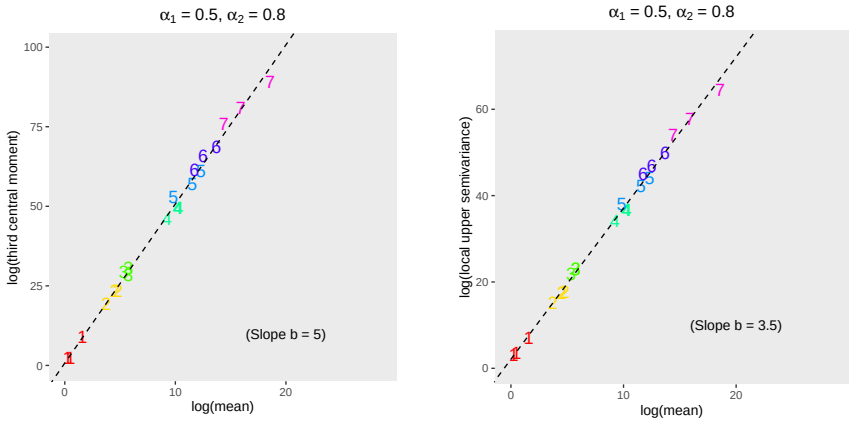
(d) Taylor's law for local upper semivariance

Fig. S19. Scatterplots of (a) $(\log(M_{n,1}), \log(V_n))$, (b) $(\log(M_{n,1}), \log(M_{n,2}^{+}))$, (c) $(\log(M_{n,1}), \log(M_{n,3}^{c}))$, and (d) $(\log(M_{n,1}), \log(M_{n,2}^{+*}))$, corresponding to Theorem 3.10 with $h_1 = 2$ and $h_2 = 1$, Theorem F.3, Theorem 3.10 with $h_1 = 3$ and $h_2 = 1$, and Theorem F.5, respectively, when $X_{i,U} \overset{\text{i.i.d.}}{\sim} \text{Pareto}(0.5, 0.5)$, $X_{i,V} \overset{\text{i.i.d.}}{\sim} \text{Pareto}(0.8, 0.8)$, and $p^* = 0.6$, with sample size 10^p for $p = 1, 2, \ldots, 7$.



(a) Taylor's law

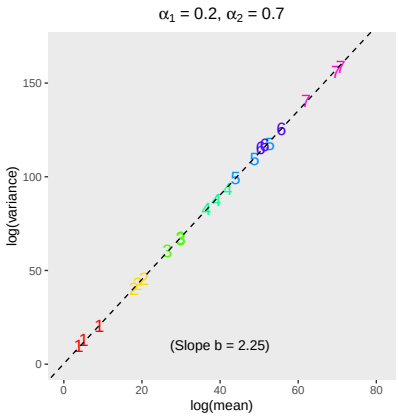
(b) Taylor's law for upper semivariance



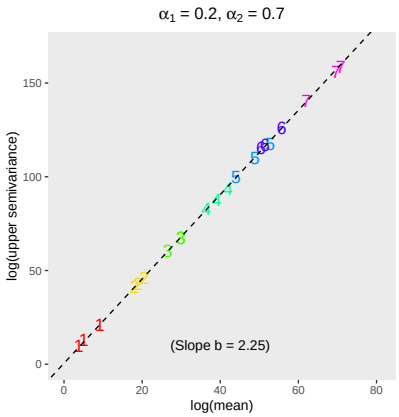
(c) Taylor's law for third central moment

(d) Taylor's law for local upper semivariance

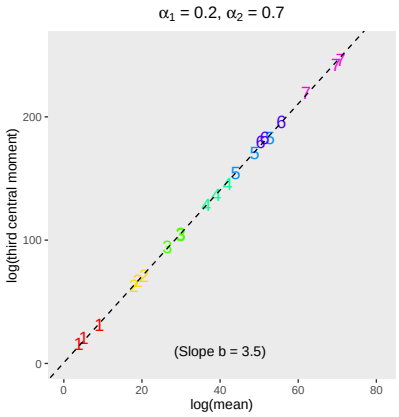
Fig. S20. Scatterplots of (a) $(\log(M_{n,1}), \log(V_n))$, (b) $(\log(M_{n,1}), \log(M_{n,2}^{+}))$, (c) $(\log(M_{n,1}), \log(M_{n,3}^{c}))$, and (d) $(\log(M_{n,1}), \log(M_{n,2}^{+*}))$, corresponding to Theorem 3.10 with $h_1 = 2$ and $h_2 = 1$, Theorem F.3, Theorem 3.10 with $h_1 = 3$ and $h_2 = 1$, and Theorem F.5, respectively, when $X_{i,U} \overset{\text{i.i.d.}}{\sim} F(1, 0.5)$, $X_{i,V} \overset{\text{i.i.d.}}{\sim} F(1, 0.8)$, and $p^* = 0.3$, with sample size 10^p for $p = 1, 2, \ldots, 7$.



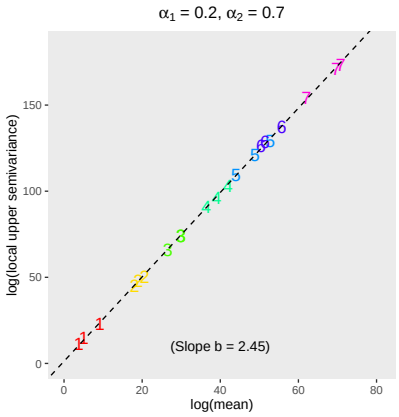
(a) Taylor's law



(b) Taylor's law for upper semivariance

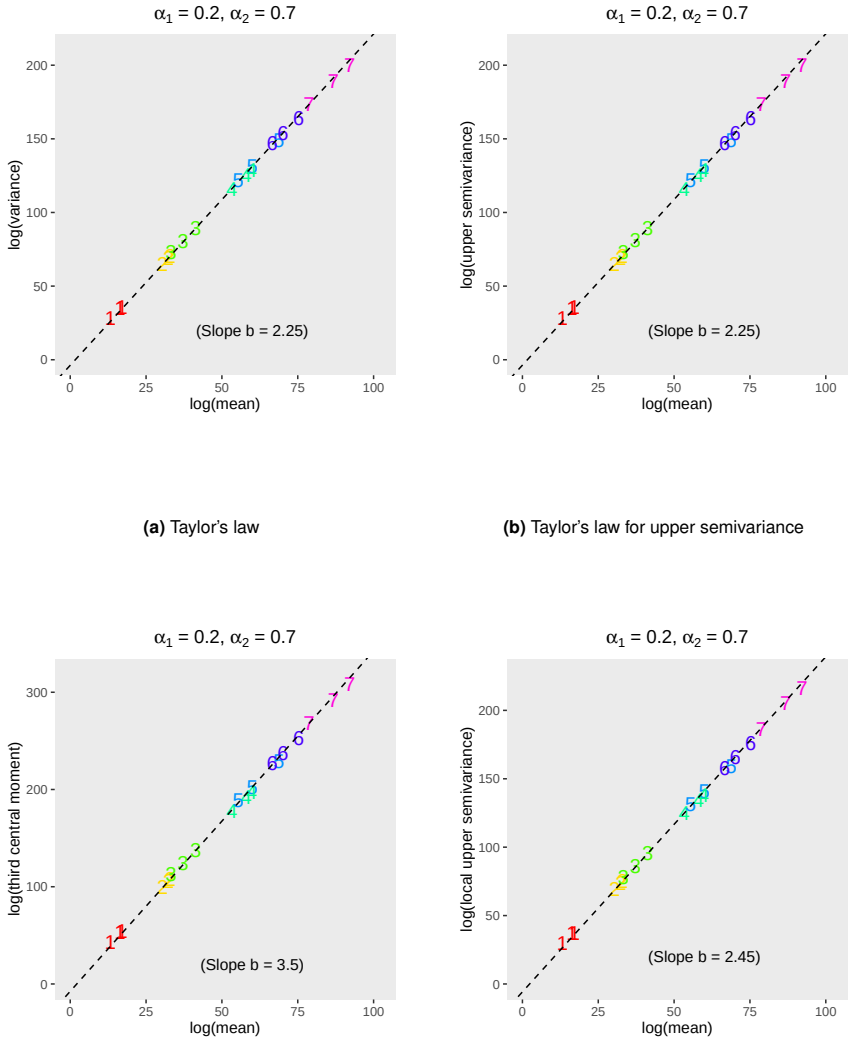


(c) Taylor's law for third central moment



(d) Taylor's law for local upper semivariance

Fig. S21. Scatterplots of (a) $(\log(M_{n,1}), \log(V_n))$, (b) $(\log(M_{n,1}), \log(M_{n,2}^{+}))$, (c) $(\log(M_{n,1}), \log(M_{n,3}^{c}))$, and (d) $(\log(M_{n,1}), \log(M_{n,2}^{+*}))$, corresponding to Theorem 3.10 with $h_1 = 2$ and $h_2 = 1$, Theorem F.3, Theorem 3.10 with $h_1 = 3$ and $h_2 = 1$, and Theorem F.5, respectively, when $X_{i,U} \overset{\text{i.i.d.}}{\sim} F(1, 0.2)$, $X_{i,V} \overset{\text{i.i.d.}}{\sim} F(1, 0.7)$, and $p^* = 0.3$, with sample size 10^p for $p = 1, 2, \ldots, 7$.



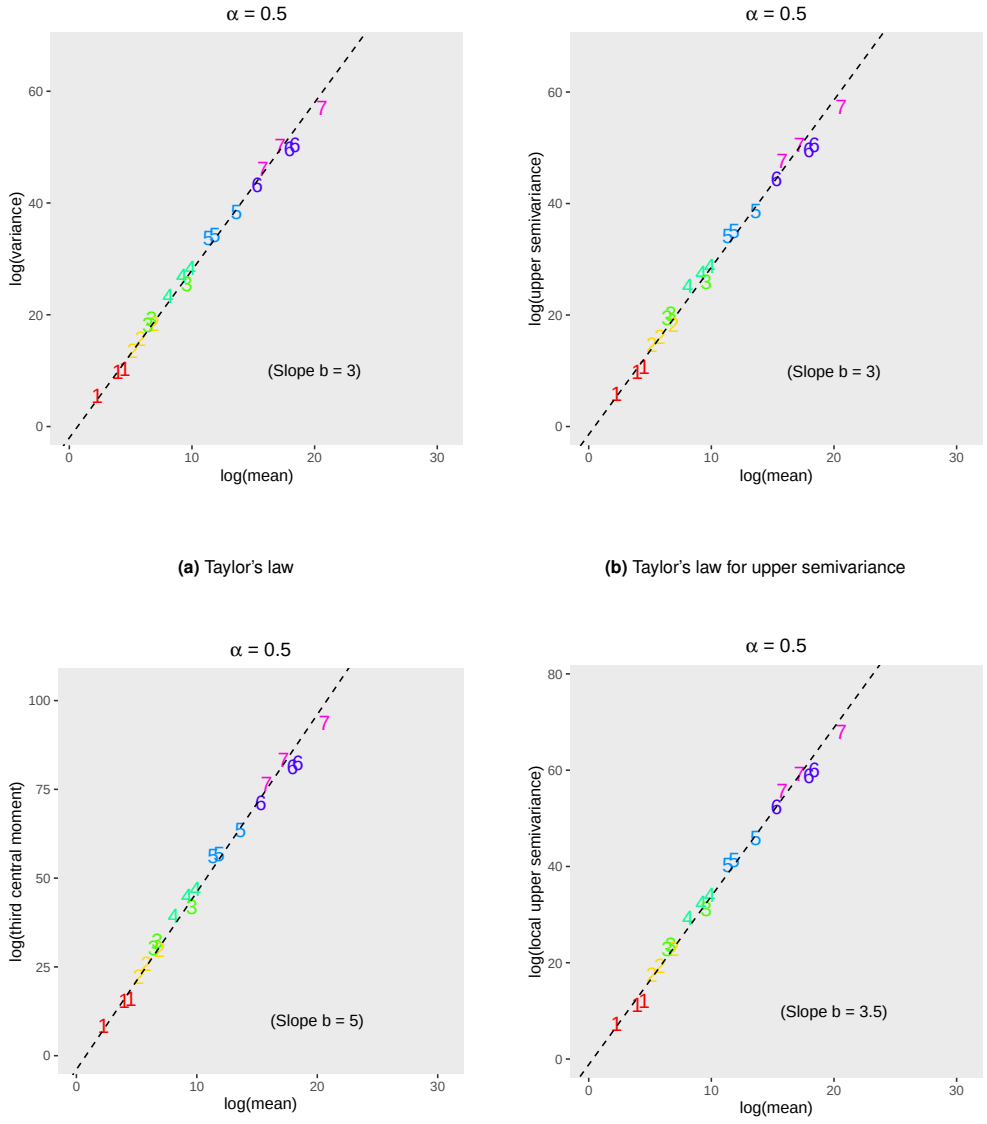
(a) Taylor's law

(b) Taylor's law for upper semivariance

(c) Taylor's law for third central moment

(d) Taylor's law for local upper semivariance

Fig. S22. Scatterplots of (a) $(\log(M_{n,1}), \log(V_n))$, (b) $(\log(M_{n,1}), \log(M_{n,2}^{+}))$, (c) $(\log(M_{n,1}), \log(M_{n,3}^{c}))$, and (d) $(\log(M_{n,1}), \log(M_{n,2}^{+*}))$, corresponding to Theorem 3.10 with $h_1 = 2$ and $h_2 = 1$, Theorem F.3, Theorem 3.10 with $h_1 = 3$ and $h_2 = 1$, and Theorem F.5, respectively, when $X_{i,U} \overset{\text{i.i.d.}}{\sim} F_{1,0.2}$, $X_{i,V} \overset{\text{i.i.d.}}{\sim} F_{1,0.7}$, and $p^* = 0.3$, with sample size 10^p for $p = 1, 2, \ldots, 7$.



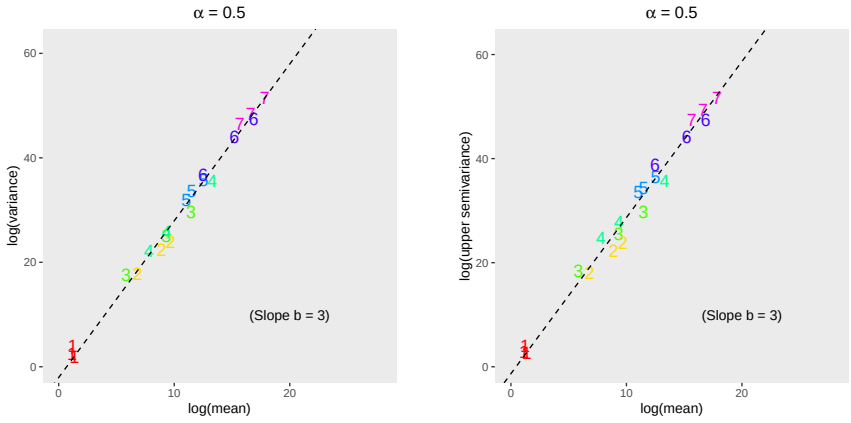
(a) Taylor's law

(b) Taylor's law for upper semivariance

(c) Taylor's law for third central moment

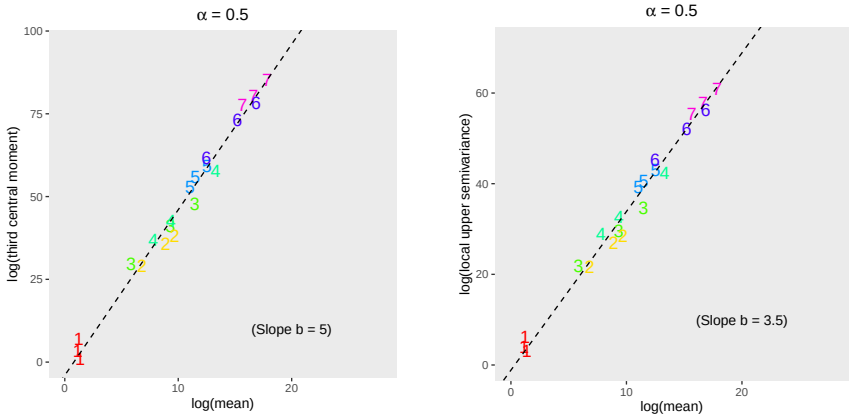
(d) Taylor's law for local upper semivariance

Fig. S23. Scatterplots of (a) $(\log(M_{n,1}), \log(V_n))$, (b) $(\log(M_{n,1}), \log(M_{n,2}^{+}))$, (c) $(\log(M_{n,1}), \log(M_{n,3}^{c}))$, and (d) $(\log(M_{n,1}), \log(M_{n,2}^{+*}))$, corresponding to Theorem 3.10 with $h_1 = 2$ and $h_2 = 1$, Theorem F.3, Theorem 3.10 with $h_1 = 3$ and $h_2 = 1$, and Theorem F.5, respectively. The simulation follows Example B.6 with $\alpha = 0.5$ and $\rho = 0.1$ and sample size 10^p for $p = 1, 2, \ldots, 7$.



(a) Taylor's law

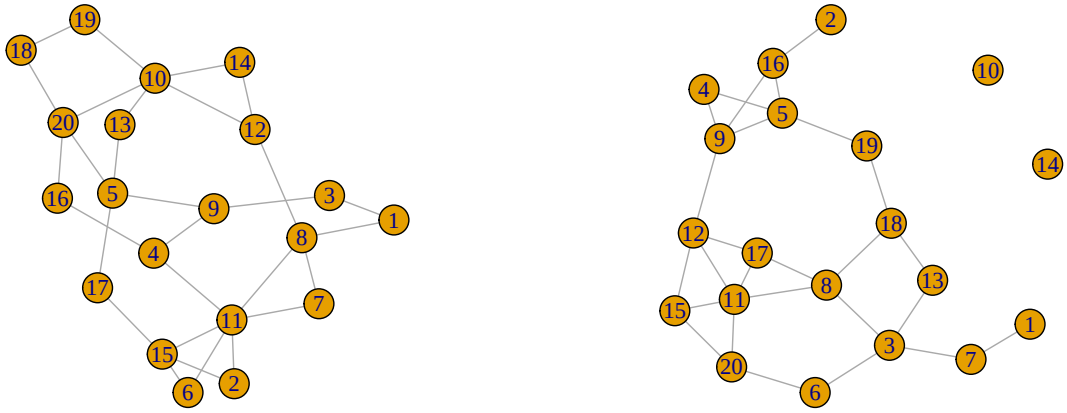
(b) Taylor's law for upper semivariance



(c) Taylor's law for third central moment

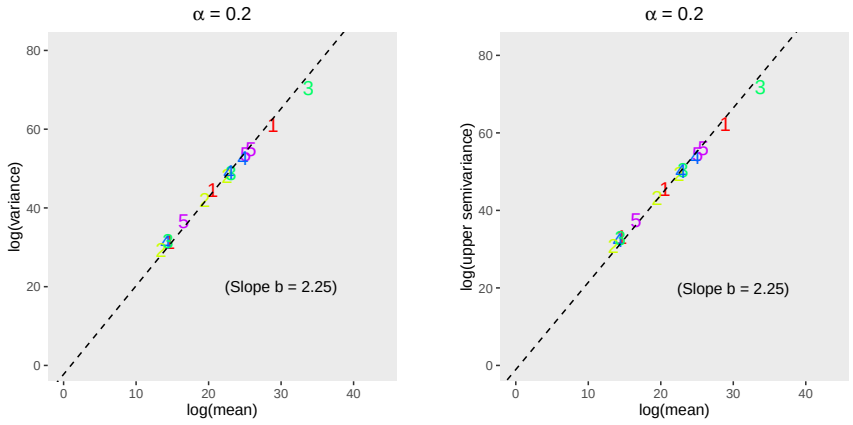
(d) Taylor's law for local upper semivariance

Fig. S24. Scatterplots of (a) $(\log(M_{n,1}), \log(V_n))$, (b) $(\log(M_{n,1}), \log(M_{n,2}^{+}))$, (c) $(\log(M_{n,1}), \log(M_{n,3}^{c}))$, and (d) $(\log(M_{n,1}), \log(M_{n,2}^{+*}))$, corresponding to Theorem 3.10 with $h_1 = 2$ and $h_2 = 1$, Theorem F.3, Theorem 3.10 with $h_1 = 3$ and $h_2 = 1$, and Theorem F.5, respectively. The simulation follows Example B.7 with $\alpha = 0.5$ and $\gamma(X_i, X_j) = \exp(-|i-j|/100)$ and sample size 10^p for $p = 1, 2, \ldots, 7$.



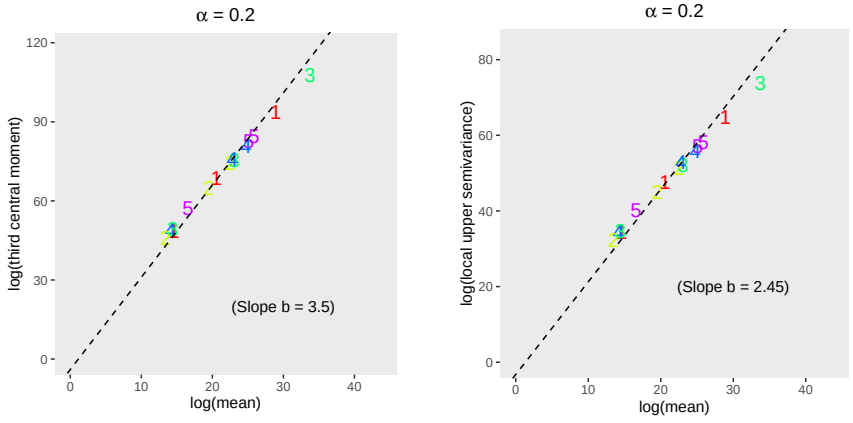
(a) Example 1 **(b)** Example 2

Fig. S25. Two subgraphs each containing 20 nodes from the network generated by the random graph.



(a) Taylor's law

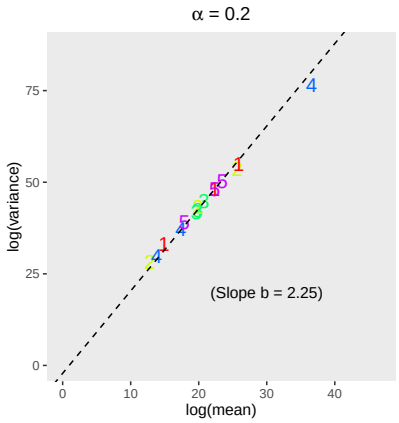
(b) Taylor's law for upper semivariance



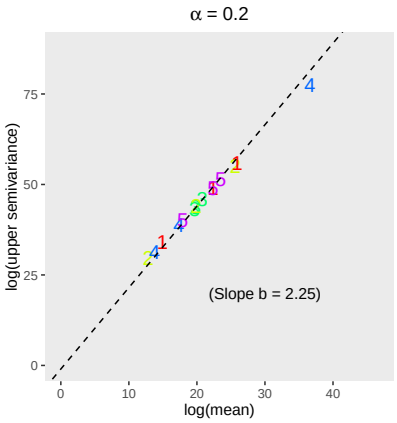
(c) Taylor's law for third central moment

(d) Taylor's law for local upper semivariance

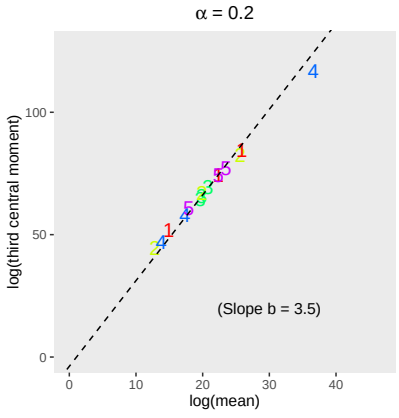
Fig. S26. Scatterplots of (a) $(\log(M_{n,1}), \log(V_n))$, (b) $(\log(M_{n,1}), \log(M_{n,2}^{+}))$, (c) $(\log(M_{n,1}), \log(M_{n,3}^{c}))$, and (d) $(\log(M_{n,1}), \log(M_{n,2}^{+*}))$, corresponding to Theorem 3.10 with $h_1 = 2$ and $h_2 = 1$, Theorem F.3, Theorem 3.10 with $h_1 = 3$ and $h_2 = 1$, and Theorem F.5, respectively, when $Z_i \overset{\text{i.i.d.}}{\sim}$ Pareto$(0.5, 0.2)$ with a sample size of $100, 500, 1000, 5000, 10000$, which are labeled as "1", . . ., "5" in the figures, respectively.



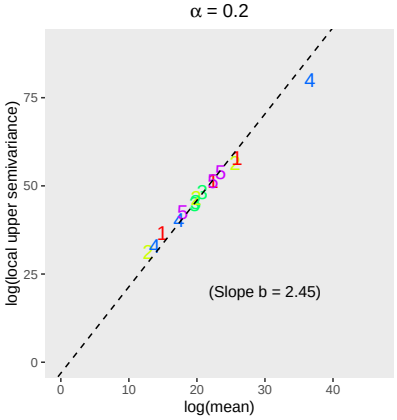
(a) Taylor's law



(b) Taylor's law for upper semivariance

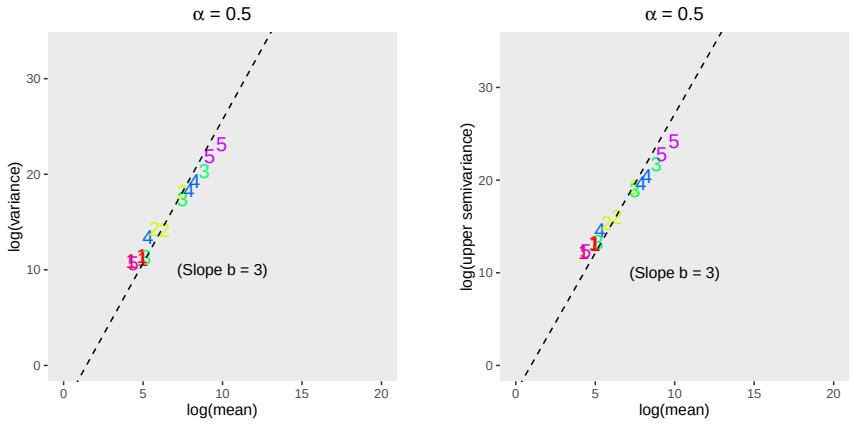


(c) Taylor's law for third central moment



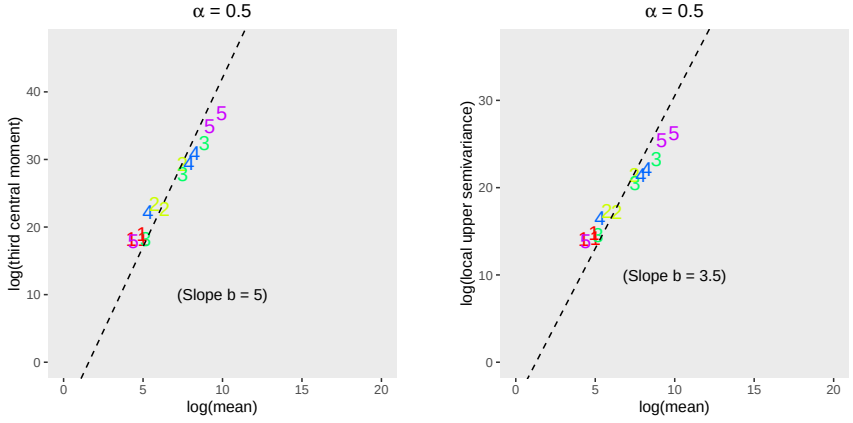
(d) Taylor's law for local upper semivariance

Fig. S27. Scatterplots of (a) $(\log(M_{n,1}), \log(V_n))$, (b) $(\log(M_{n,1}), \log(M_{n,2}^{+}))$, (c) $(\log(M_{n,1}), \log(M_{n,3}^{c}))$, and (d) $(\log(M_{n,1}), \log(M_{n,2}^{+*}))$, corresponding to Theorem 3.10 with $h_1 = 2$ and $h_2 = 1$, Theorem F.3, Theorem 3.10 with $h_1 = 3$ and $h_2 = 1$, and Theorem F.5, respectively, when $Z_i \overset{\text{i.i.d.}}{\sim} F(1, 0.2)$ with a sample size of $100, 500, 1000, 5000, 10000$, which are labeled as "1", . . ., "5" in the figures, respectively.



(a) Taylor's law

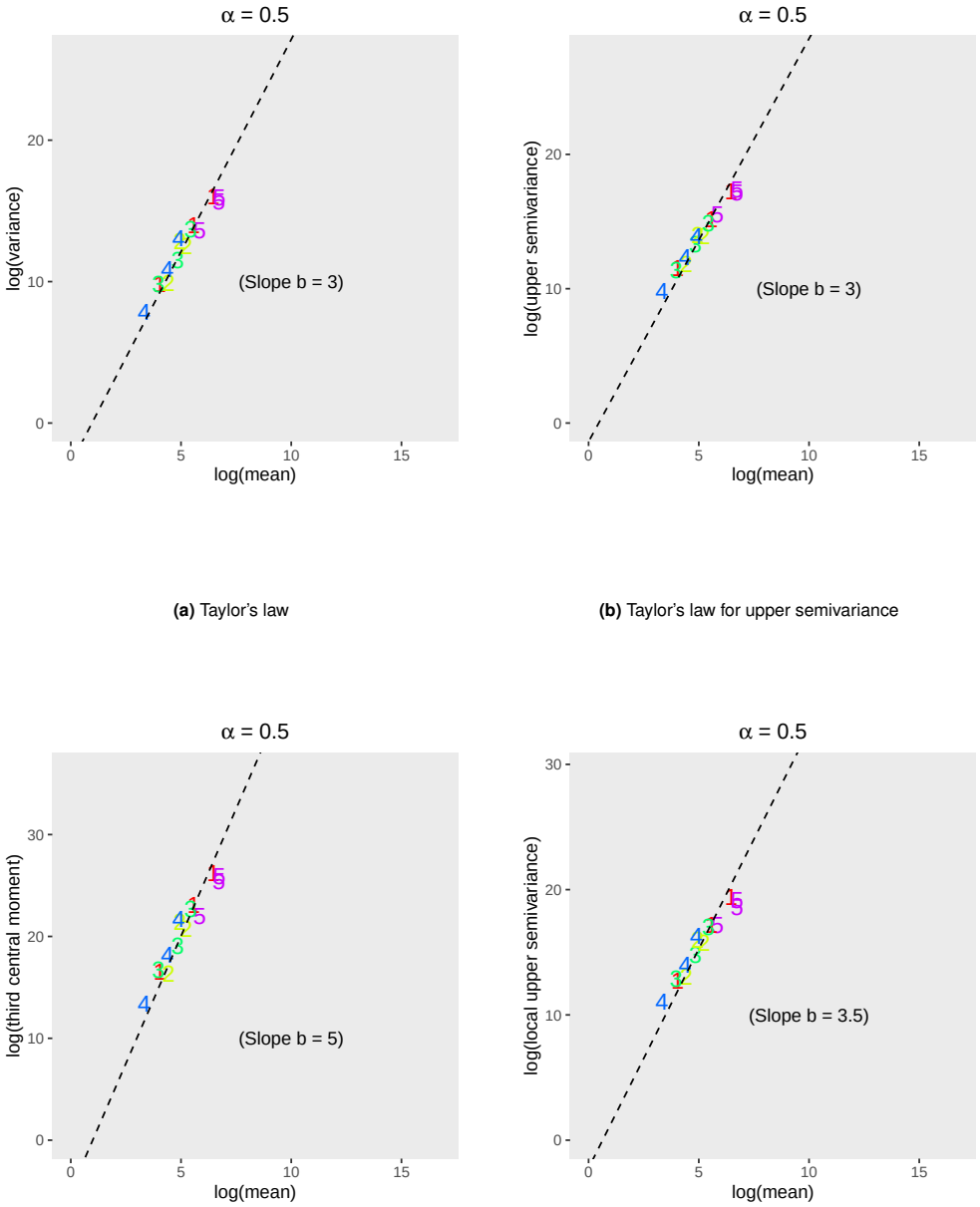
(b) Taylor's law for upper semivariance



(c) Taylor's law for third central moment

(d) Taylor's law for local upper semivariance

Fig. S28. Scatterplots of (a) $(\log(M_{n,1}), \log(V_n))$, (b) $(\log(M_{n,1}), \log(M_{n,2}^{+}))$, (c) $(\log(M_{n,1}), \log(M_{n,3}^{c}))$, and (d) $(\log(M_{n,1}), \log(M_{n,2}^{+*}))$, corresponding to Theorem 3.10 with $h_1 = 2$ and $h_2 = 1$, Theorem F.3, Theorem 3.10 with $h_1 = 3$ and $h_2 = 1$, and Theorem F.5, respectively, when $Z_i \overset{\text{i.i.d.}}{\sim} \text{Pareto}(0.5, 0.5)$ with a sample size of $100, 500, 1000, 5000, 10000$, which are labeled as "1", . . ., "5" in the figures, respectively.



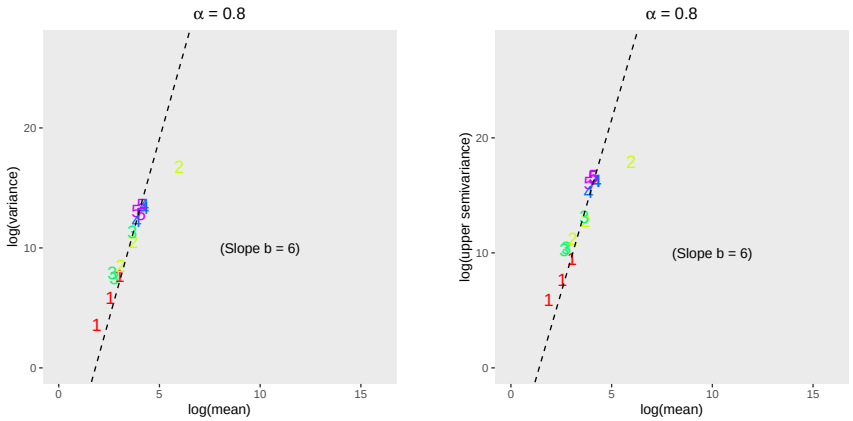
(a) Taylor's law

(b) Taylor's law for upper semivariance

(c) Taylor's law for third central moment

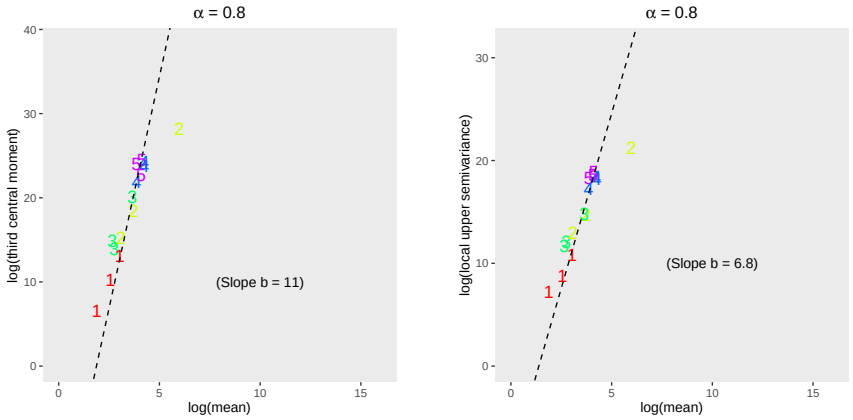
(d) Taylor's law for local upper semivariance

Fig. S29. Scatterplots of (a) $(\log(M_{n,1}), \log(V_n))$, (b) $(\log(M_{n,1}), \log(M_{n,2}^{+}))$, (c) $(\log(M_{n,1}), \log(M_{n,3}^{c}))$, and (d) $(\log(M_{n,1}), \log(M_{n,2}^{+*}))$, corresponding to Theorem 3.10 with $h_1 = 2$ and $h_2 = 1$, Theorem F.3, Theorem 3.10 with $h_1 = 3$ and $h_2 = 1$, and Theorem F.5, respectively, when $Z_i \overset{\text{i.i.d.}}{\sim} F(1, 0.5)$ with a sample size of $100, 500, 1000, 5000, 10000$, which are labeled as "1", . . ., "5" in the figures, respectively.



(a) Taylor's law

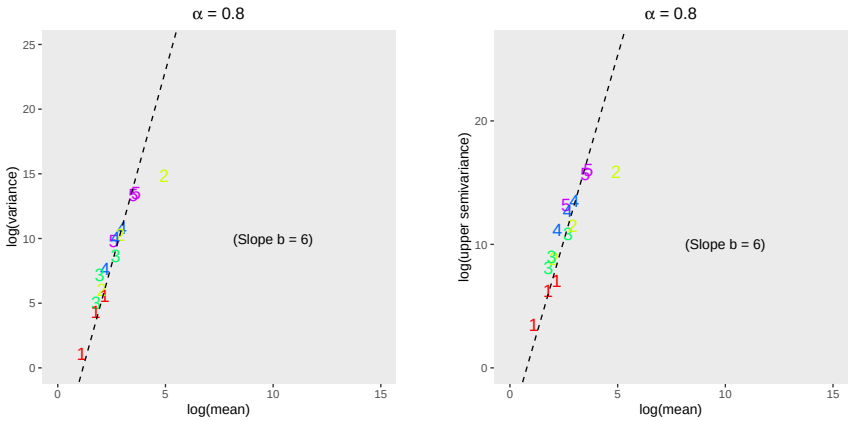
(b) Taylor's law for upper semivariance



(c) Taylor's law for third central moment

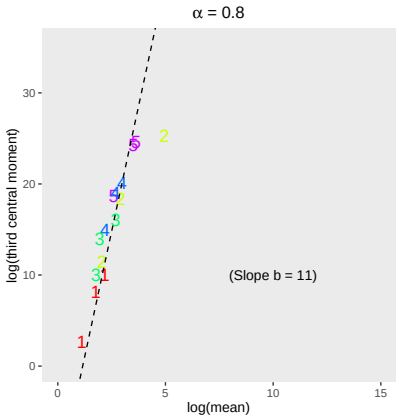
(d) Taylor's law for local upper semivariance

Fig. S30. Scatterplots of (a) $(\log(M_{n,1}), \log(V_n))$, (b) $(\log(M_{n,1}), \log(M^{+}_{n,2}))$, (c) $(\log(M_{n,1}), \log(M^{c}_{n,3}))$, and (d) $(\log(M_{n,1}), \log(M^{+*}_{n,2}))$, corresponding to Theorem 3.10 with $h_1 = 2$ and $h_2 = 1$, Theorem F.3, Theorem 3.10 with $h_1 = 3$ and $h_2 = 1$, and Theorem F.5, respectively, when $Z_i \overset{\text{i.i.d.}}{\sim} \text{Pareto}(0.5, 0.8)$ with a sample size of $100, 500, 1000, 5000, 10000$ which are labeled as "1", . . ., "5" in the figures, respectively.

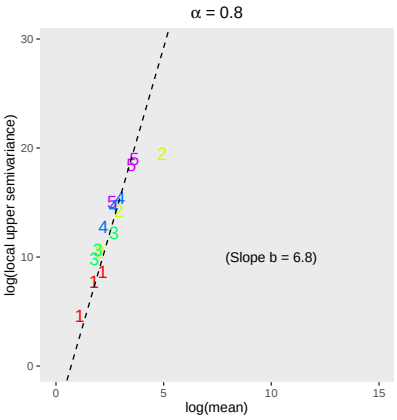


(a) Taylor's law

(b) Taylor's law for upper semivariance



(c) Taylor's law for third central moment



(d) Taylor's law for local upper semivariance

Fig. S31. Scatterplots of (a) $(\log(M_{n,1}), \log(V_n))$, (b) $(\log(M_{n,1}), \log(M_{n,2}^{+}))$, (c) $(\log(M_{n,1}), \log(M_{n,3}^{c}))$, and (d) $(\log(M_{n,1}), \log(M_{n,2}^{+*}))$, corresponding to Theorem 3.10 with $h_1 = 2$ and $h_2 = 1$, Theorem F.3, Theorem 3.10 with $h_1 = 3$ and $h_2 = 1$, and Theorem F.5, respectively, when $Z_i \overset{\text{i.i.d.}}{\sim} F(1, 0.8)$ with a sample size of $100, 500, 1000, 5000, 10000$, which are labeled as "1", . . ., "5" in the figures, respectively.

References

1. S Chakraborti, F Jardim, E Epprecht, Higher-order moments using the survival function: The alternative expectation formula. *The Am. Stat.* **73**, 191–194 (2019).
2. Y Liu, A general treatment of alternative expectation formulae. *Stat. & Probab. Lett.* **166**, 108863 (2020).
3. NH Bingham, CM Goldie, JL Teugels, *Regular variation.* (Cambridge University Press) Vol. 27, (1987).
4. IA Ibragimov, Some limit theorems for stationary processes. *Theory Probab. & Its Appl.* **7**, 349–382 (1962).
5. RC Bradley, Basic properties of strong mixing conditions. a survey and some open questions. *Probab. Surv.* **2**, 107–144 (2005).
6. HA Jessen, T Mikosch, Regularly varying functions. *Publ. de l'Institut Mathématique* **80**, 171–192 (2006).
7. DW Andrews, First order autoregressive processes and strong mixing. *Cowles Foundation Discuss. Pap. 664* (1983).
8. JE Cohen, RA Davis, G Samorodnitsky, Heavy-tailed distributions, correlations, kurtosis and Taylor's law of fluctuation scaling. *Proc. Royal Soc. A* **476**, 20200610 (2020).
9. JE Cohen, RA Davis, G Samorodnitsky, Covid-19 cases and deaths in the United States follow Taylor's law for heavy-tailed distributions with infinite variance. *Proc. Natl. Acad. Sci.* **119**, e2209234119 (2022).
10. M Brown, JE Cohen, CF Tang, SCP Yam, Taylor's law of fluctuation scaling for semivariances and higher moments of heavy-tailed data. *Proc. Natl. Acad. Sci.* **118**, e2108031118 (2021).
11. S Coles, *An Introduction to Statistical Modeling of Extreme Values*, Springer Series in Statistics. (Springer, London), (2001).
12. SM Ross, *Simulation.* (Academic Press), (2022).
13. JP Nolan, Univariate stable distributions. *Springer Ser. Oper. Res. Financial Eng.* **10**, 978–3 (2020).
14. EN Gilbert, Random graphs. *The Annals Math. Stat.* **30**, 1141–1144 (1959).