# Heavy-Tailed Phenomena

# and Their Modeling [1]


Thomas Mikosch

University of Copenhagen

www.math.ku.dk/∼mikosch/Vilnius


[1]Vilnius, 17-22 September, 2007





Contents

# 1. Evidence of heavy tails in real-life data

## 1.1. Heavy tails in finance.

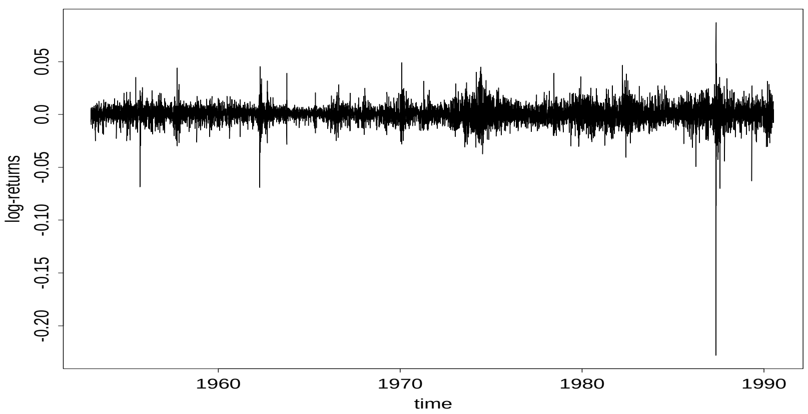


Figure 1. Plot of **9558** *S&P500* daily log-returns from January *2, 1953*, to December *31, 1990*. The year marks indicate the beginning of the calendar year.



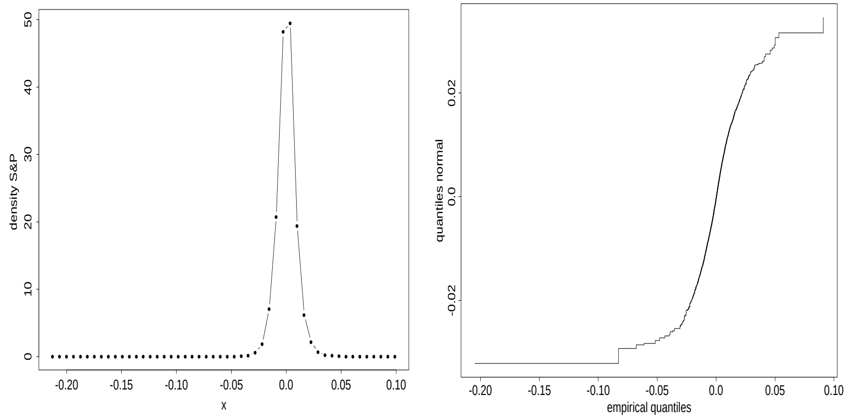


Figure 2. Left: Density plot of the *S&P500* data. The limits on the $x$-axis indicate the range of the data. QQ-plot of the *S&P500* data against the normal distribution.



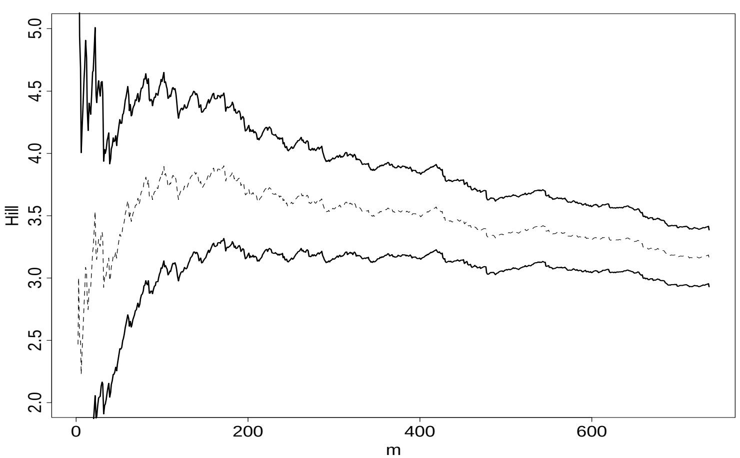


Figure 3. Hill plot (dotted line) for the *S&P500 data* with **95%** asymptotic confidence bounds. The Hill estimator approximates the tail index $\alpha$ in the model $P(X_1 > x) \sim c\,x^{-\alpha}$ as a function of the $m$ upper order statistics in the return sample.



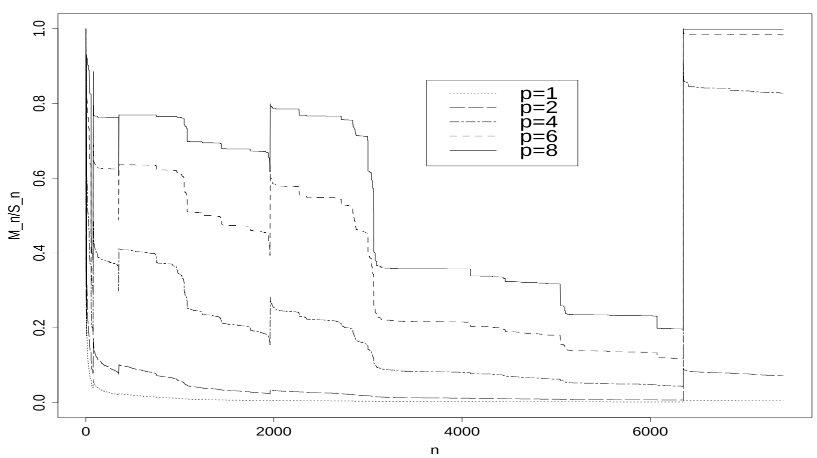


Figure 4. Plot of the ratio $T_n(p) = \max_{i=1,\dots,n} |X_i|^p/(|X_1|^p + \cdots + |X_n|^p)$ for the *S&P500* data for various values of $p$. If $E|X_1|^p < \infty$ and the data came from a stationary ergodic model, the ratio should converge to zero a.s., by virtue of the strong law of large numbers.

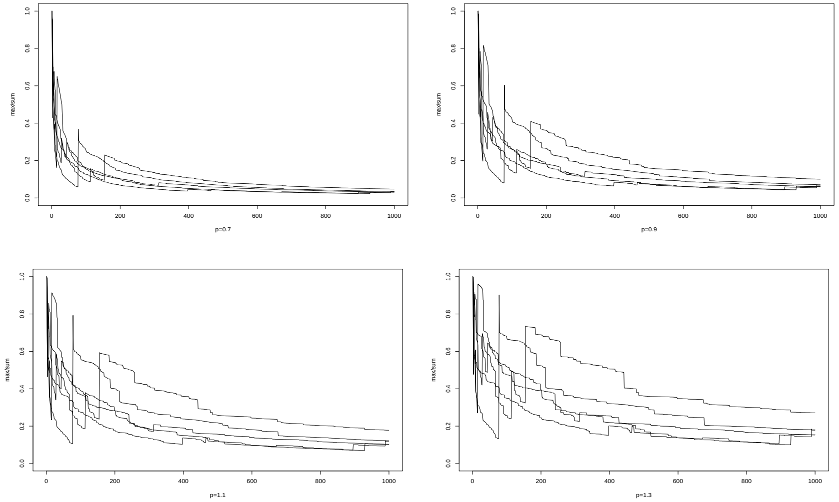


FIGURE 5. Plot of the ratio $T_n(p) = \max_{i=1,\dots,n} |X_i|^p/(|X_1|^p + \dots + |X_n|^p)$ for iid simulated Cauchy variables with tail $P(|X| > x) \sim c\,x^{-1}$ and $p = 0.7, 0.9, 1.1, 1.3$



## 1.2. Heavy tails in insurance.

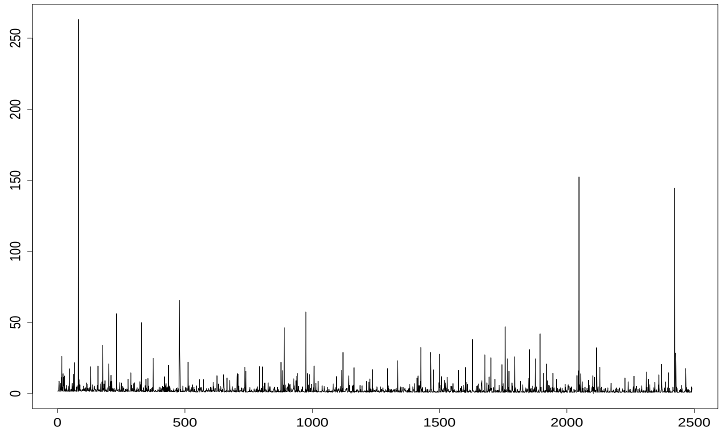


FIGURE 6. Danish fire insurance data.



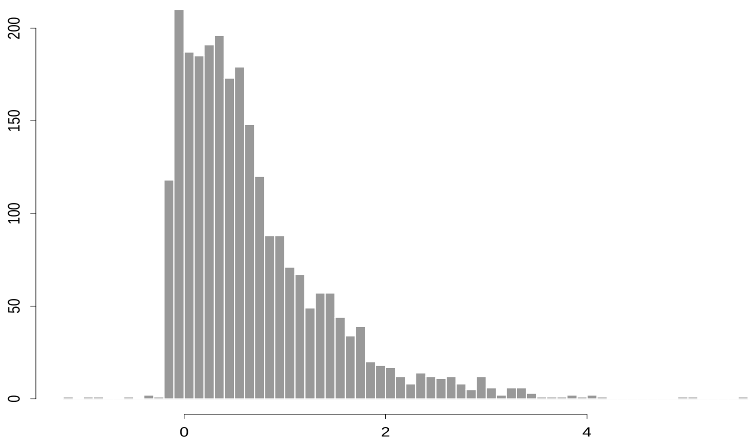


FIGURE 7. Histogram of the logarithmic Danish fire insurance data.



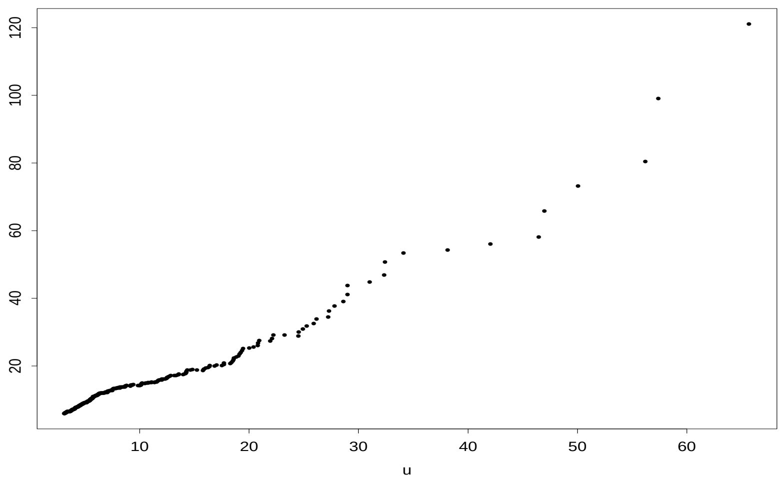


FIGURE 8. Empirical mean excess function of the Danish fire insurance data.

## 1.3. Heavy tails in teletraffic.

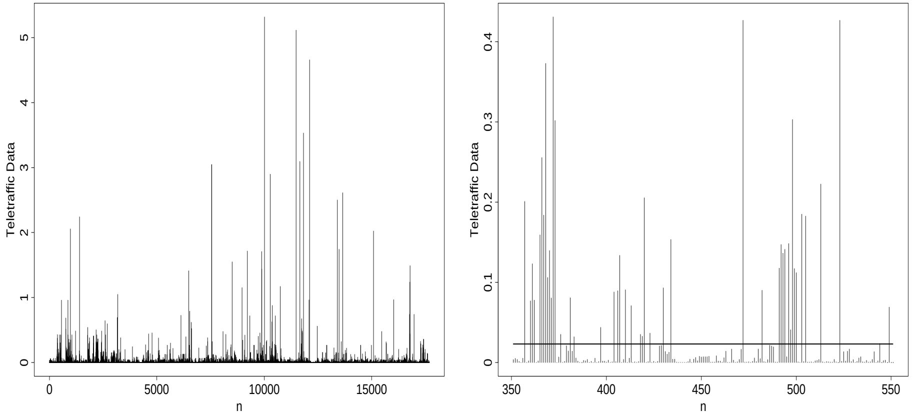


Figure 9. Time series of transmission durations (BU data).



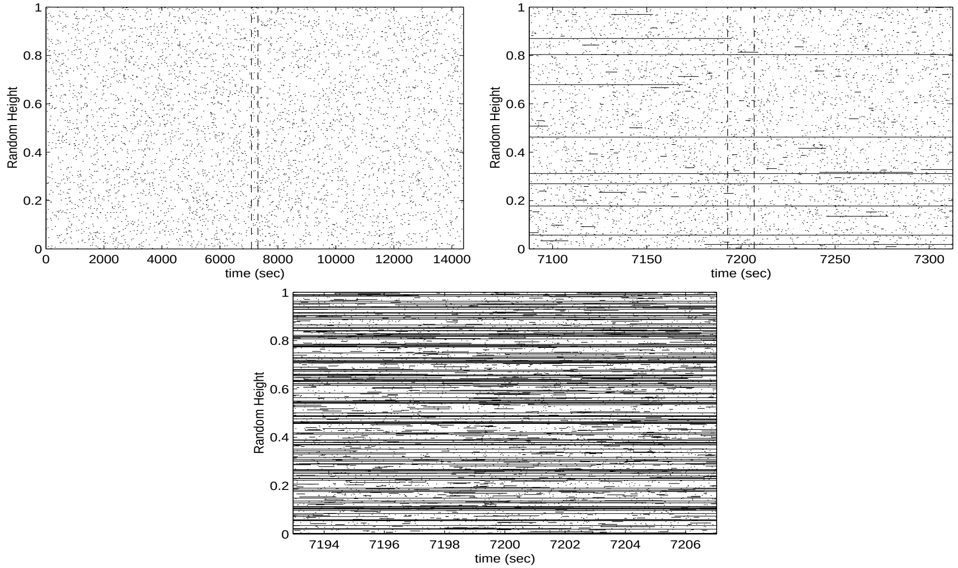


Figure 10. Mice and elephants plots (S. Marron).



## 2. Objectives in modeling heavy-tailed phenomena

- **Extreme value theory and statistics for spatio-temporal structures with heavy-tailed components: time series and random fields.**
- **Study the influence of very large values in these structures, e.g. large file sizes in telecommunications, large claims in insurance, large losses/gains in finance, impact of large numbers of citations in science, storms in hydrology,. . . .**
- **Study extremal dependence in multivariate structures. Find measures of dependence beyond covariances. The interplay between tails and extremal dependence.**



- **Infinite-dimensional structures: extreme value theory for stochastic processes and random fields, e.g., the maximal height of sea waves along a coast line, maximum temperature or maximum rainfall in a given area,. . . .**
- **Use results from extreme value theory to build reasonable models for description of heavy-tailed phenomena, e.g. in teletraffic (ON-OFF process, infinite source Poisson model,. . . ) financial time series analysis (stochastic volatility models, GARCH models,. . . )**

## 3. Concepts of heavy-tailed distributions, see Embrechts et al. (1997)

### 3.1. Long-tailed distributions.

A positive random variable $X$ and its distribution $F$ with tail $\overline{F} = 1 - F$ are long-tailed if

$$\frac{\overline{F}(x-y)}{\overline{F}(x)} \to 1\,, \quad x \to \infty\,, \quad y \in \mathbb{R}\,. \tag{3.1}$$

- Notice that

$$P(X > x + y \mid X > x) \to 1\,, \quad x \to \infty\,, \quad y > 0\,.$$

- If $X > x$ exceeded $X$ it is very likely to exceed any higher level $x + y$.
- This notion is inconvenient since the class of long-tailed distributions is "too large".



- (3.1) is equivalent to slow variation of $L(x) = \overline{F}(\log x)$, i.e.

$$\frac{L(cx)}{L(x)} \to 1\,, \quad x \to \infty\,, \quad c > 0\,.$$

This means that $\overline{F}(x) = L(\mathrm{e}^x)$.

- A slowly varying function has Karamata representation, see Bingham et al. (1987)

$$L(x) = c(x)\, \exp\left\{ - \int_z^x \frac{\varepsilon(t)}{t}\, dt \right\}\,, \quad x \ge z\,,$$

for some functions $\varepsilon(t) \to 0$ and $c(t) \to c > 0$ as $t \to \infty$.

- It satisfies for any $\delta > 0$ and sufficiently large $x$

$$x^{-\delta} \le L(x) \le x^{\delta}\,.$$

- In particular, for long-tailed $F$ and any $\delta > 0$,

$$\mathrm{e}^{\delta x} \overline{F}(x) = \mathrm{e}^{\delta\, x} L(\mathrm{e}^x) \to \infty\,, \quad x \to \infty\,,$$



hence

$$E\mathrm{e}^{\delta X} = \int_0^\infty P(\mathrm{e}^{\delta X} > y)\, dy = \infty\,.$$

- Exponential moments do not exist.
- It is reasonable to define "heavy-tailedness" in relation to some probabilistic structure, e.g. sums, maxima,..., where one observes a phase transition of the behavior when changing from "light" to "heavy" tails.
- A simple example: CLT behavior for an iid sequence $(X_i)$ with $EX^2 < \infty$ or $EX^2 = \infty$. In the first case, normal limits and $\sqrt{n}$ normalizations appear, in the latter case infinite variance stable limits and normalizations $n^{1/\alpha}\ell(n)$, $\alpha \in (0,2)$, may occur.



### 3.2. Subexponential distributions.

- A positive random variable $X$ and its distribution $F$ are subexponential if for iid copies $X_i$ of $X$ and any (some) $n \ge 2$, with $S_n = X_1 + \cdots + X_n$, $M_n = \max(X_1, \ldots, X_n)$,

$$\frac{P(S_n > x)}{P(M_n > x)} \sim \frac{P(S_n > x)}{n\,\overline{F}(x)} \to 1\,, \quad x \to \infty\,.$$

- Subexponential distributions are long-tailed, see EKM, p. 41.
- Subexponential distributions do not have finite moment generating function. See EKM, p. 42, and p. 19 above.
- Examples of subexponential distributions: regularly varying distributions $\overline{F}(x) = x^{-\alpha} L(x)$, $\alpha > 0$, $L$ slowly varying (power law tails), log-normal, heavy-tailed Weibull $\overline{F}(x) = \mathrm{e}^{-x^\tau}$, $\tau \in (0,1)$, $\overline{F}(x) = \mathrm{e}^{-x/\log^\gamma(x)}$, $\gamma > 0$.

- Examples of non-subexponential distributions: exponential, gamma, (truncated) normal, any distribution with finite upper endpoint.
- Subexponential distributions are considered as natural heavy-tailed distributions in the context of insurance mathematics, queuing, storage, dam, renewal theory.
- In insurance mathematics one is interested in the tail behavior of a random walk $(S_{N(t)})$ (total claim amount) with iid step sizes $X_i$ (claim sizes), independent of the counting process $N$ (claim arrivals). For a fixed $t > 0$, with $p_n = P(N(t) = n)$,

$$P(S_{N(t)} > x) = \sum_{n=1}^{\infty} p_n \, P(S_n > x)\,.$$



- If $X_i$ is subexponential, then $P(S_n > x)/\overline{F}(x) \le K(\varepsilon)\,(1+\varepsilon)^n$.

See EKM p. 41.

- Hence if $Ee^{hN(t)} < \infty$ for some $h > 0$,

$$\frac{P(S_{N(t)} > x)}{\overline{F}(x)} = \sum_{n=1}^{\infty} p_n \frac{P(S_n > x)}{\overline{F}(x)} \sim \sum_{n=1}^{\infty} p_n \, n = EN(t)\,.$$

- The total claim amount $S_{N(t)}$ of an insurance portfolio at high threshold:

$$P(S_{N(t)} > x) \sim EN(t)\,\overline{F}(x)\,.$$

- The right-hand side decays to zero much slower than exponential or normal tails.



- If $N$ is homogeneous Poisson with intensity $\lambda > 0$, the ruin probability in the portfolio is given by ($u$ initial capital, $c$ premium rate)

$$\begin{aligned}\psi(u) &= P\left(\inf_{t\ge 0}\,(u + c\,t - S_{N(t)}) < 0\right)\\ &= \rho(1+\rho)^{-1} \sum_{n=1}^{\infty} (1+\rho)^{-n}\, P(S_n^* > u)\,,\end{aligned}$$

where $\rho = c/(\lambda EX) - 1$ is assumed positive (net profit condition) and $(S_n^*)$ is a random walk with iid positive step sizes with distribution $F_*(x) = (EX)^{-1}\int_0^x \overline{F}(t)\,dt$.



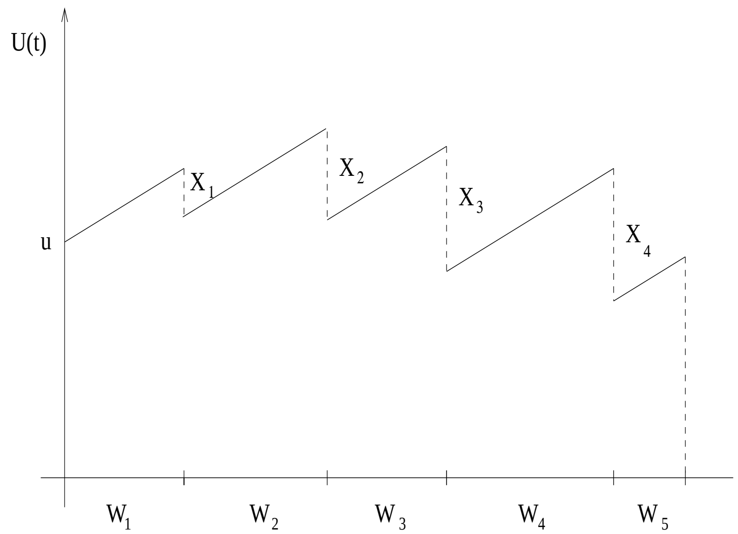


FIGURE 11. The risk process $U(t) = u + ct - S_{N(t)}$ and ruin.

- **TFAE** EKM, p. 581

  **(1)** $F_*$ is subexponential,

  **(2)** $1-\psi$ is subexponential,

  **(3)** the following relation holds as $u \to \infty$

$$\frac{\psi(u)}{\overline{F}_*(u)} \sim \rho(1+\rho)^{-1} \sum_{n=1}^{\infty} (1+\rho)^{-n}\, n = \rho^{-1}\,.$$

- This is in stark contrast to the light-tailed Cramér-Lundberg case. See EKM, p. 29
- Under general conditions, if the claim size distribution $F$ has finite moment generating function in some neighborhood of the origin: for some $C, r > 0$,

$$\psi(u) \sim C\, \mathrm{e}^{-r\,u}\,, \quad u \to \infty\,.$$



- **Message:**

  Light-tailed claims are not dangerous. Ruin is very unlikely (exponentially small probability) for large initial capital $u$.

  Heavy-tailed claims are dangerous. Ruin is not unlikely even for large initial capital $u$. One large claim may cause ruin of the insurance company.





| Losses | Date | Event | Country |
|---|---|---|---|
| 20 511 | 08/24/92 | Hurricane “Andrew” | US, Bahamas |
| 19 301 | 09/11/01 | Terrorist attack on WTC, Pentagon and other buildings | US |
| 16 989 | 01/17/94 | Northridge earthquake in California | US |
| 7 456 | 09/27/91 | Tornado “Mireille” | Japan |
| 6 321 | 01/25/90 | Winter storm “Daria” | Europe |
| 6 263 | 12/25/99 | Winter storm “Lothar” | Europe |
| 6 087 | 09/15/89 | Hurricane “Hugo” | P. Rico, US |
| 4 749 | 10/15/87 | Storm and floods | Europe |
| 4 393 | 02/26/90 | Winter storm “Vivian” | Europe |
| 4 362 | 09/22/99 | Typhoon “Bart” hits the south of the country | Japan |
| 3 895 | 09/20/98 | Hurricane “Georges” | US, Caribbean |
| 3 200 | 06/05/01 | Tropical storm “Allison”; flooding | US |
| 3 042 | 07/06/88 | Explosion on “Piper Alpha” offshore oil rig | UK |
| 2 918 | 01/17/95 | Great “Hanshin” earthquake in Kobe | Japan |
| 2 592 | 12/27/99 | Winter storm “Martin” | France, Spain, CH |
| 2 548 | 09/10/99 | Hurricane “Floyd”, heavy down-pours, flooding | US, Bahamas |
| 2 500 | 08/06/02 | Rains, flooding | Europe |
| 2 479 | 10/01/95 | Hurricane “Opal” | US, Mexico |
| 2 179 | 03/10/93 | Blizzard, tornadoes | US, Mexico, Canada |
| 2 051 | 09/11/92 | Hurricane “Iniki” | US, North Pacific |
| 1 930 | 04/06/01 | Hail, floods and tornadoes | US |
| 1 923 | 10/23/89 | Explosion at Philips Petroleum | US |
| 1 864 | 09/03/79 | Hurricane “Frederic” | US |
| 1 835 | 09/05/96 | Hurricane “Fran” | US |
| 1 824 | 09/18/74 | Tropical cyclone “Fifi” | Honduras |
| 1 771 | 09/03/95 | Hurricane “Luis” | Caribbean |
| 1 675 | 04/27/02 | Spring storm with several tornadoes | US |
| 1 662 | 09/12/88 | Hurricane “Gilbert” | Jamaica |
| 1 620 | 12/03/99 | Winter storm “Anatol” | Europe |
| 1 604 | 05/03/99 | Series of 70 tornadoes in the Midwest | US |
| 1 589 | 12/17/83 | Blizzard, cold wave | US, Mexico, Canada |
| 1 585 | 10/20/91 | Forest fire which spread to urban area | US |
| 1 570 | 04/02/74 | Tornados in 14 states | US |
| 1 499 | 04/25/73 | Flooding on the Mississippi | US |
| 1 484 | 05/15/98 | Wind, hail and tornadoes (MN, IA) | US |
| 1 451 | 10/17/89 | “Loma Prieta” earthquake | US |
| 1 436 | 08/04/70 | Hurricane “Celia” | US |
| 1 409 | 09/19/98 | Typhoon “Vicki” | Japan, Philippines |
| 1 358 | 01/05/98 | Cold spell with ice and snow | Canada, US |
| 1 340 | 05/05/95 | Wind, hail and flooding | US |

**Table 3.2.18** *The 40 most costly insurance losses* 1970 − 2002*. Losses are in million \$US indexed to* 2002 *prices. The table is taken from* Sigma [73] *with friendly permission of Swiss Re Zurich.*



## 3.3. Regularly varying distributions.

- A positive measurable function $f$ on $(0,\infty)$ is regularly varying with index $\rho \in \mathbb{R}$ if $f(x) = L(x)\, x^{\rho}$ for some slowly varying function $L$. Bingham et al. (1987)
- A positive measurable function $f$ on $(0,\infty)$ is regularly varying if and only if as $x \to \infty$, $\dfrac{f(cx)}{f(x)} \to c^{\rho}\,, c > 0$.
- A positive random variable $X$ and its distribution $F$ are regularly varying with index $\alpha > 0$ if for some slowly varying function $L$

$$\overline{F}(x) = P(X > x) = \frac{L(x)}{x^{\alpha}}\,, \quad x > 0\,.$$

- Regularly varying distributions are subexponential see e.g. Feller (1971), EKM p. 37, hence long-tailed.
- Examples:

  Pareto,

  log-gamma,

  Fréchet,

  absolute values/positive parts of

  infinite variance stable,

  Cauchy,

  student



- Regular variation is a natural condition in the context of extreme value theory and limit theory for partial sums of iid random variables: For an iid non-negative sequence $(X_i)$ with distribution $\overline{F}(x) = P(X > x) = \dfrac{L(x)}{x^\alpha}$, $x > 0$,

Fréchet limit: $n^{-1/\alpha}\,\ell(n)\,M_n \stackrel{d}{\rightarrow} Y_M \sim \Phi_\alpha\,,\ \alpha > 0\,,$

Stable limit: $n^{-1/\alpha}\,\ell(n)\,(S_n - b_n) \stackrel{d}{\rightarrow} Y_S \sim P_\alpha\,,\quad \alpha \in (0,2]\,.$

See Feller (1971), Ibragimov, Linnik (1971), Petrov (1975,1995) for sums; Galambos (1978), Leadbetter et al. (1983), Resnick (1987,2006), EKM for maxima



- There is also joint convergence for $\alpha \in (0,2)$ see Resnick (1986):

$$n^{-1/\alpha}\ell(n)(M_n, S_n - b_n) \stackrel{d}{\rightarrow} (Y_M, Y_S)\,,$$

  and $Y_S, Y_M$ are dependent.
- By the continuous mapping theorem,

$$(S_n - b_n)/M_n \stackrel{d}{\rightarrow} Y_S/Y_M\,.$$

- In particular, for $\alpha \in (0,1)$,

$$S_n/M_n \stackrel{d}{\rightarrow} Y_S/Y_M$$

  and for $\alpha \in (1,2)$

$$(S_n - n\,EX)/M_n \stackrel{d}{\rightarrow} Y_S/Y_M\,.$$



- If $\alpha \geq 2$, there exist $a_n, b_n \to \infty$ such that

$$(a_n^{-1}(S_n - n\,EX)\,, b_n^{-1}\,M_n) \stackrel{d}{\rightarrow} (Y_S, Y_M)\,,$$

  $Y_S, Y_M$ are independent, $Y_S \sim N(0,1)$, $Y_M \sim \Phi_\alpha$ and $b_n/a_n \to 0$.
- In particular,

$$M_n/(S_n - n\,EX) \stackrel{P}{\rightarrow} 0\,.$$

- For general iid non-negative $X_i$, $M_n/S_n \stackrel{P}{\rightarrow} 0$ if and only if $EX < \infty$ or $P(X > x) = L(x)x^{-1}$ for some slowly varying $L$ O'Brien (1980), and $M_n/S_n \stackrel{P}{\rightarrow} 1$ if and only if $\overline{F}(x) = L(x)$ Arov and Bobrov (1960), Maller and Resnick (1984).

## 3.4. Alternative definitions of regular variation.

- $X > 0$ is regularly varying with index $\alpha > 0$ if and only if

$$\frac{P(X > tx)}{P(X > x)} \to t^{-\alpha}, \quad x \to \infty, \quad t > 0.$$

- Replacing $x$ by $a_n$ with $P(X > a_n) \sim n^{-1}$ or $a_n = F^{\leftarrow}(1 - 1/n)$, one can show equivalence with

$$n\,P(a_n^{-1}X > t) \to t^{-\alpha}, \quad n \to \infty, \quad t > 0. \tag{3.2}$$

- For iid copies $X_i$ of $X$, (3.2) has interpretation

$$E\Big(\sum_{i=1}^{n} I_{(t,\infty)}(X_i/a_n)\Big) \to t^{-\alpha}, \quad t > 0.$$



Recalling Poisson's limit theorem, (3.2) is equivalent to

$$\sum_{i=1}^{n} I_{(t,\infty)}(X_i/a_n) \xrightarrow{d} \text{Poisson}(t^{-\alpha}), \quad t > 0,$$

- (3.2) is equivalent to the point process convergence

$$N_n = \sum_{i=1}^{n} \varepsilon_{X_i/a_n} \xrightarrow{d} N \sim \text{PRM}(\mu),$$

where $\mu(t, \infty] = t^{-\alpha}$ defines the mean measure of the limiting Poisson random measure $N$ with state space $(0, \infty]$.

- (3.2) can be shown to be equivalent to vague convergence of the measures

$$n\,P(a_n^{-1}X \in \cdot) \xrightarrow{v} \mu(\cdot) \quad \text{on } (0, \infty].$$



# 4. Some point process theory

The theory of point processes plays a central role in extreme value theory. Applications include:

- Derivation of joint limiting distribution of order statistics, i.e., $k$th largest order statistic, limiting distribution of maximum and minimum, etc.
- Calculation of limit distribution of exceedances of a high level.
- Extensions to stationary processes.
- Provides a useful tool in heavy-tailed case for deriving limiting behavior of various statistics, e.g., sample mean, sample autocovariances, etc., which are often determined by the behavior of the extreme order statistics.



## 4.1. Basic results on convergence of extremes.

- Suppose $(X_i)$ is an iid sequence with common distribution $F$.
- Assume that there exist sequences of constants $a_n > 0$ and $b_n$ such that

$$P(a_n^{-1}(M_n - b_n) \le x) = F^n(a_n x + b_n) \to H(x) \tag{4.1}$$

for all $x$, where $M_n = \max(X_1, \ldots, X_n)$ and $H$ is a nondegenerate distribution function.

- By the Fisher-Tippett theorem (1928), (see Leadbetter et al. (1983), Resnick (1987), EKM p. 121) $H$ has to be an extreme value distribution of which there are only three types.

Example: $H(x) = \Phi_\alpha(x) = e^{-x^{-\alpha}}$, $x > 0$, for some $\alpha > 0$.

Fréchet distribution.

(4.1) holds iff $\overline{F}$ is regularly varying with index $\alpha$ and then one can choose $b_n = 0$ and $a_n = F^{\leftarrow}(1 - 1/n)$ or such that $\overline{F}(a_n) \sim n^{-1}$.

- Taking logarithms and using a Taylor series expansion, (4.1) holds if and only if for any $x \in \mathbb{R}$,

$$n\,P(a_n^{-1}(X - b_n) > x) \to -\log H(x)\,. \tag{4.2}$$

(If $H(x) = 0$ we interpret $-\log H(x)$ as $\infty$.)

Example: In the Fréchet case, (4.2) becomes

$$n\,P(X > a_n x) \to -\log \Phi_\alpha(x) = x^{-\alpha}\,, \quad x > 0\,.$$

- Now (4.2) can be strengthened to the statement,

$$n\,P(a_n^{-1}(X - b_n) \in B) \to \nu(B) \tag{4.3}$$



for all suitably chosen Borel sets $B$, where the measure $\nu$ is defined by its value on intervals of the form $(a, b]$ as

$$\nu(a, b] = \log H(b) - \log H(a)\,.$$

- The convergence in (4.3) can be connected with the convergence in distribution of a sequence of point processes.
- For a bounded Borel set $B$ in the product space $(0, \infty) \times \mathbb{R}$, define the sequence of point processes $(N_n)$ by

$$\begin{aligned} N_n(B) &= \#\{(t/n, a_n^{-1}(\,X_t - b_n)) \in B\,, t = 1, 2, \ldots\} \\ &= \sum_{t=1}^{\infty} \varepsilon_{(t/n, a_n^{-1}(X_t - b_n))}(B)\,, \end{aligned}$$

where $\varepsilon_y$ is the Dirac measure at the point $y$.



Properties:

- If $B$ is the rectangle $(a, b] \times (c, d]$ with $0 \le a < b < \infty$ and $-\infty < c < d < \infty$, then since the $X_j$ are iid, $N_n(B) \sim \text{Bin}([nb] - [na], p_n)$ ($[s]$ = integer part of $s$), and

$$p_n = P(a_n^{-1}(X_1 - b_n) \in (c, d])\,.$$

- Provided $\nu(c, d] < \infty$, it follows from (4.3) that $N_n(B)$ converges in distribution to a Poisson random variable $N(B)$ with mean $\mu(B) = (b - a)\,\nu(c, d]$.
- In fact, we have the stronger point process convergence,

$$N_n \stackrel{d}{\to} N\,,$$



where $N$ is a Poisson process on $(0, \infty) \times \mathbb{R}$ with mean measure $\mu(dt, dx) = dt \times \nu(dx)$ and $\stackrel{d}{\to}$ denotes convergence in distribution of point processes.

- Recall that a point process $N$ is a Poisson process or Poisson random measure with mean measure $\mu$ (PRM($\mu$)) and state space $E \subset \overline{\mathbb{R}}^d$ if $N(A)$ is Poisson($\mu$(A)) distributed and $N(A_1), \ldots, N(A_m)$ are independent if the $A_i$'s are disjoint.

4.2. Convergence for point processes. For our purposes, $N_n \stackrel{d}{\to} N$ for point processes $N_n, N$ means that for any collection of *bounded* Borel sets $B_1, \ldots, B_k$ for which $P(N(\partial B_j) > 0) = 0$, $j = 1, \ldots, k$, we have

$$(N_n(B_1), \ldots, N_n(B_k)) \stackrel{d}{\to} (N(B_1), \ldots, N(B_k))$$

on $\mathbb{R}^k$. See Leadbetter et al. (1983), Resnick (1987), EKM, Chapter 4.

**Technical remarks:**

- In the heavy-tailed case, the state space of the point process is often defined to be $(0,\infty)\times([-\infty,\infty]\setminus\{0\})$.
- For the space, $[-\infty,\infty]\setminus\{0\}$, the roles of zero and infinity have been interchanged so that bounded sets are now those sets which are bounded away from 0.
- A bounded set on the product space is contained in the rectangle $[0,c]\times([-\infty,-d]\cup[d,\infty])$ for some positive and finite constants $c$ and $d$. Under this topology, the mean measure of the limit Poisson process is ensured to be finite on all bounded Borel sets.



**Application:**

Define $X_{(n-1)}$ to be the second largest among $X_1,\ldots,X_n$. The event $\{a_n^{-1}(X_{(n-1)}-b_n)\le y\}$ is the same as $\{N_n((0,1]\times(y,\infty))\le 1\}$, we conclude from $N_n\xrightarrow{d}N$ that

$$\begin{aligned}P(a_n^{-1}(X_{(n-1)}-b_n)\le y) &= P(N_n((0,1]\times(y,\infty))\le 1)\\ &\to P(N((0,1]\times(y,\infty))\le 1)\\ &= H(y)\,(1-\log H(y))\,.\end{aligned}$$



Similarly, the joint limiting distribution of $(X_{(n-1)},M_n)$ can be calculated by noting that for $y\le x$,

$$\begin{aligned}&\{a_n^{-1}(M_n-b_n)\le x, a_n^{-1}(X_{(n-1)}-b_n)\le y\}\\ &=\{N_n((0,1]\times(x,\infty))=0, N_n((0,1]\times(y,x])\le 1\}.\end{aligned}$$

Hence,

$$\begin{aligned}&P(a_n^{-1}(M_n-b_n)\le x, a_n^{-1}(X_{(n-1)}-b_n)\le y)\\ &= P(N_n((0,1]\times(x,\infty))=0, N_n((0,1]\times(y,x])\le 1)\\ &\to P(N((0,1]\times(x,\infty))=0, N((0,1]\times(y,x])\le 1))\\ &= H(y)(1+\log H(x)-\log H(y))\,.\end{aligned}$$



## 5. Dependence and tails for financial data, see Mikosch (2003)

### 5.1. “Stylized facts”. Log-returns

$$\begin{aligned}X_t &= \log(P_t/P_{t-1}) = \log P_t - \log P_{t-1}\\ &= \log\left(1+\frac{P_t-P_{t-1}}{P_{t-1}}\right)\\ &\approx \frac{P_t-P_{t-1}}{P_{t-1}}, \quad t=0,1,2\ldots.\end{aligned}$$

$P_t$ is a speculative price (share price, stock index, FX rate, ...)

Scale: days, weeks, hours, ..., high frequency data.

Why log-returns? $(X_t)$ is unit free. Common belief: prices $P_t$ “increase” exponentially on average, $(X_t)$ is stationary.

### 5.1.1. Marginal distribution.

- sample mean close to zero
- sample variance of order $10^{-5}, 10^{-6}, \ldots$
- distribution is roughly symmetric in the center
- density is sharply peaked at zero (leptokurtic)
- data are non-Gaussian
- heavy tails on both sides,

$$P(X_t > x) \approx x^{-\alpha} \quad \text{as } x \to \infty,\ \alpha \in (2, 5)$$

- The tails of log-returns do not seem to be balanced (asymmetric). One often estimates different tail parameters $\alpha$ for the left and right tails. The left tail is often heavier than the right tail.



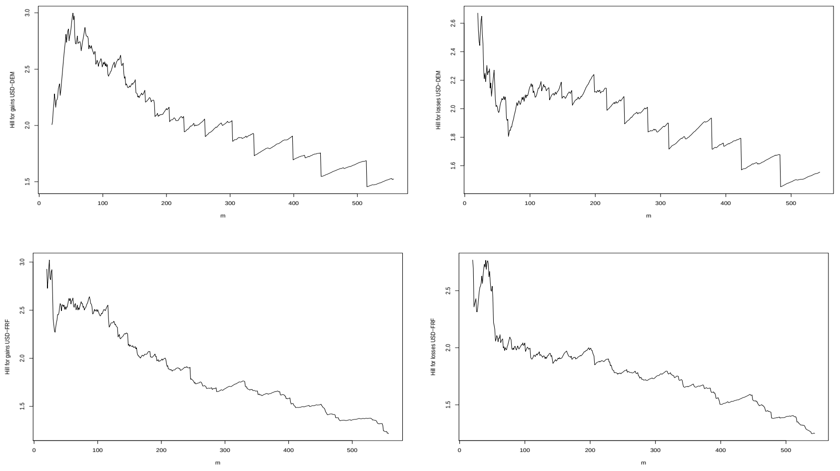

Figure 12. Hill estimation (based on up to 50% of the order statistics) for 5 minute foreign exchange rate log-returns, USD-DEM (top) and USD-FRF (bottom). Left: gains. Right: losses.



- Since the 1960s, Benoit Mandelbrot has propagated scaling behavior for various phenomena in nature and society, space and time (Zipf's law, fractals, Hurst phenomenon, ...).
- One of his claims is that *log-returns come from an infinite variance stable distribution.* This would imply

$$P(X_t > x) \sim c\, x^{-\alpha} \quad \text{for some } \alpha \in (0, 2).$$

- There is no statistical evidence for this claim.



- The Black-Scholes model for the price $P_t$ of a risky asset prescribes that $\log P_t = ct + \sigma B_t$, $t \geq 0$, where $\sigma, c > 0$ and $B$ is Brownian motion. This means that log-returns would be iid Gaussian: $X_t = c + \sigma\,(B_t - B_{t-1})$, $t = 1, 2, \ldots$. This assumption is convenient from a mathematical point of view but it is not realistic.
- Various attempts have been made to get more realistic models in continuous time, e.g. with Lévy process driven SDE: Eberlein; Barndorff-Nielsen and Shephard; Madan and Seneta,....

## 5.1.2. Dependence, autocorrelations, clustering of extremes.

- Classical time series analysis: main goal is second order structure of (Gaussian) stationary time series $(X_t)$
- This structure is determined by

  autocovariance function (ACVF)

$$\gamma_X(h) = \operatorname{cov}(X_0, X_h)\,, \quad h \in \mathbb{Z}\,.$$

autocorrelation function (ACF)

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}\,, \quad h \in \mathbb{Z}\,.$$

- ACF determines dependence structure of stationary Gaussian $(X_t)$.



- ACF used for parameter estimation, model testing, prediction of Gaussian/non-Gaussian time series (ARMA, FARIMA,...)

  Brockwell and Davis (1991,1996)

- Since one does not know the ACF/ACVF of real-life data one needs to estimate them: sample ACVF and sample ACF

$$\gamma_{n,X}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(X_t - \overline{X}_n)(X_{t+h} - \overline{X}_n)$$

$$\rho_{n,X}(h) = \frac{\gamma_{n,X}(h)}{\gamma_{n,X}(0)}\,, \quad h \in \mathbb{Z}\,.$$

- If $(X_t)$ is *stationary ergodic*, $\operatorname{var}(X_t) < \infty$,

$$\gamma_{n,X}(h) \stackrel{\text{a.s.}}{\to} \gamma_X(h)\,, \quad \rho_{n,X}(h) \stackrel{\text{a.s.}}{\to} \rho_X(h)\,.$$



## The ACF stylized fact

- sample ACF $\rho_{n,X}$ of returns are negligible (possible exception: 1st lag)
- sample ACFs $\rho_{n,|X|}$, $\rho_{n,X^2}$ are positive and decay very slowly (typical for "long" time series)
- This is often interpreted as long memory or long-range dependence (LRD), see Samorodnitsky and Taqqu (1994), Doukhan et al. (2003)



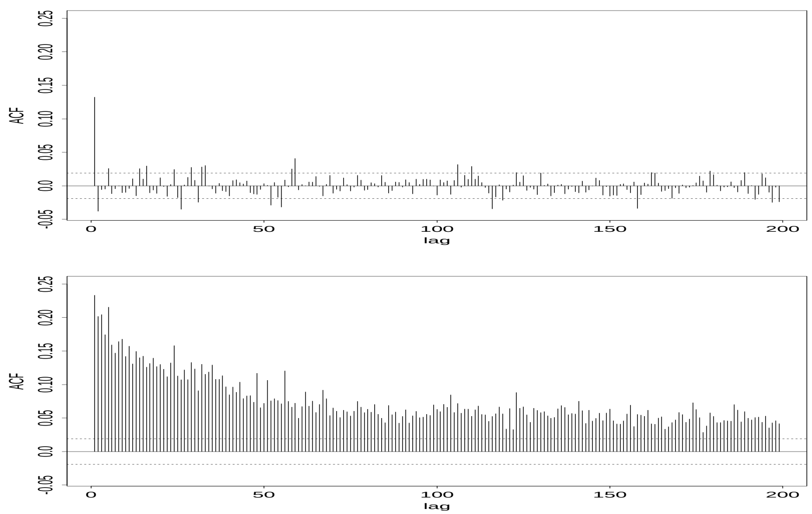


Figure 13. Sample ACFs for the log-returns *(top)* and absolute log-returns *(bottom)* of the *S&P500*. Here and in what follows, the horizontal lines in graphs displaying sample ACFs are set as the **95%** confidence bands $(\pm 1.96/\sqrt{n})$ corresponding to the ACF of iid Gaussian noise.

## The extremal stylized fact

**High/low level exceedances of returns tend to appear in clusters.**



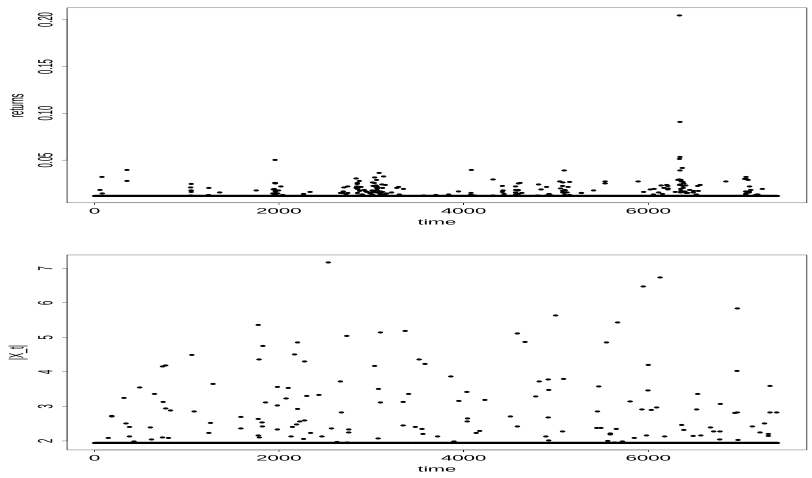


Figure 14. *Top:* Absolute returns $|X_t|$ of the *S&P500* series for which both $|X_t|$ and $|X_{t+1}|$ exceed the $87\%$ quantile of the data. The latter is indicated by the bottom line. *Bottom:* The same kind of plot for an iid sequence from a student distribution with $4$ degrees of freedom. In the former case pairwise exceedances occur in clusters, in the latter case exceedances appear uniformly scattered through time.



## The extremal index Leadbetter et al. (1983), EKM, Section 8.1

- Assume $(X_t)$ iid and $M_n = \max(X_1, \ldots, X_n)$. Then

$$P(M_n \le x) = [P(X_1 \le x)]^n \,.$$

- This is incorrect for a dependent sequence $(X_t)$. For strictly stationary $(X_t)$ (often) for some $x_n \uparrow$

$$P(M_n \le x_n) = [P(X_1 \le x_n)]^{\theta n} + o(1)\,,$$

for some $\theta \in [0,1]$, the extremal index.

- A possible interpretation of $\theta > 0$: By definition,

$$P(M_n \le x_n) \approx [P(\widetilde{X}_1 \le x_n)]^{n\theta} \approx P(\widetilde{M}_{[n\theta]} \le x_n)\,,$$

where $\widetilde{M}_n = \max(\widetilde{X}_1, \ldots, \widetilde{X}_n)$ and $(\widetilde{X}_t)$ are iid with the same marginal distribution as $X_1$.



- This means that

$$P(M_{[n/\theta]} \le x_n) \approx P(\widetilde{M}_n \le x_n)\,,$$

i.e., the maximum of $n/\theta$ dependent $X_i$'s has roughly the same distribution as the maximum of $n$ iid $\widetilde{X}_i$'s for a sufficiently high threshold $x_n$.

- This is due to the fact that large values of $X_i$ have roughly the same value in a small time interval: they generate a cluster of exceedances of $x_n$.
- $1/\theta$ is the expected (asymptotic) cluster size.

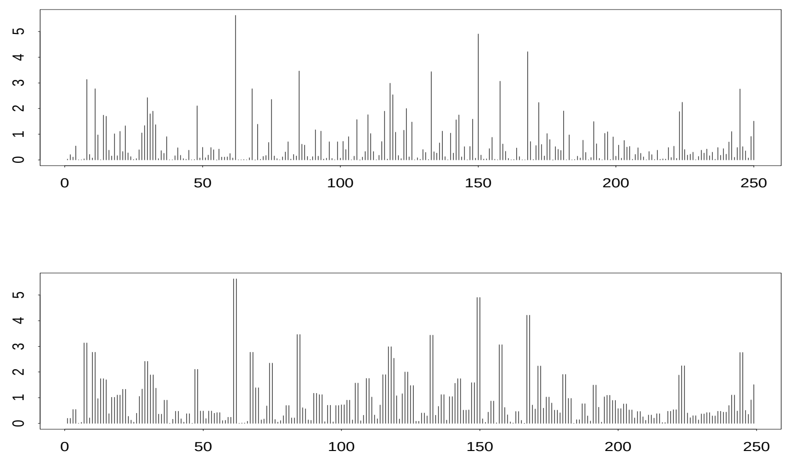


Figure 15. A sequence of iid random variables $Y_i$ (Top) with distribution function $\sqrt{F}$, where $F$ is standard exponential. Bottom: the sequence of pairwise maxima $\max(Y_i, Y_{i+1})$ with distribution $F$. By construction, extremes appear in clusters of size 2. The extremal index is $\theta = 1/2$.



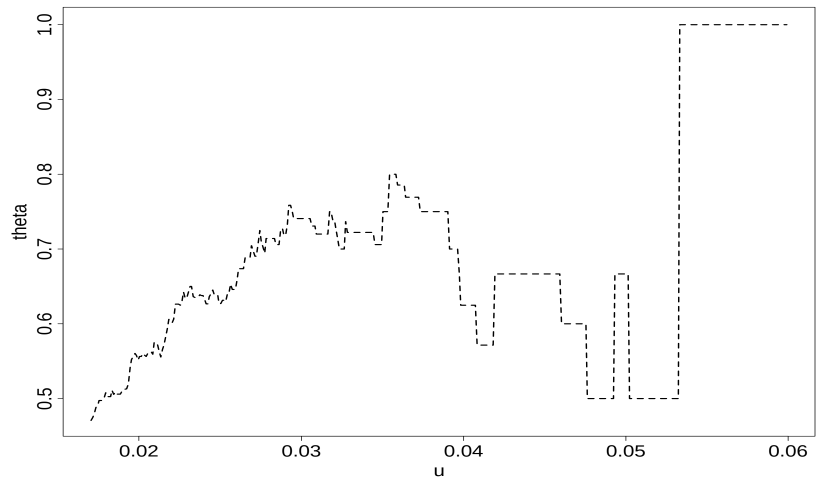


Figure 16. Point estimation of the extremal index for the *S&P500* data. The estimators are based on the upper order statistics exceeding the threshold $u$. The smallest $u$ is the $97\%$ quantile of the data. The estimator is reliable where it is stable (between $0.03$ and $0.04$) resulting in an extremal index of about $0.7$.



## 5.2. Can classical time series analysis model returns?

- Classical time series analysis is about linear processes

$$X_t = \sum_{j=0}^{\infty} \psi_j \, Z_{t-j}\,, \quad (Z_t) \text{ iid}$$

in particular ARMA$(p,q)$ processes:

$$\begin{aligned} \varphi(B) X_t &= X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p}\,, \\ = \theta(B) Z_t &= Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}\,, \end{aligned}$$

where

$$\varphi(z) = 1 - \sum_{i=1}^{p} \varphi_i \, z^i\,, \quad \theta(z) = 1 + \sum_{j=1}^{q} \theta_j \, z^j\,,$$

and $B^k A_t = A_{t-k}$ is the backshift operator.



**Examples.** AR(2) process

$$(1 - \varphi_1 B - \varphi_2 B^2) X_t = X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2} = Z_t\,.$$

MA(2) process

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} = (1 + \theta_1 B + \theta_2 B^2) Z_t\,.$$

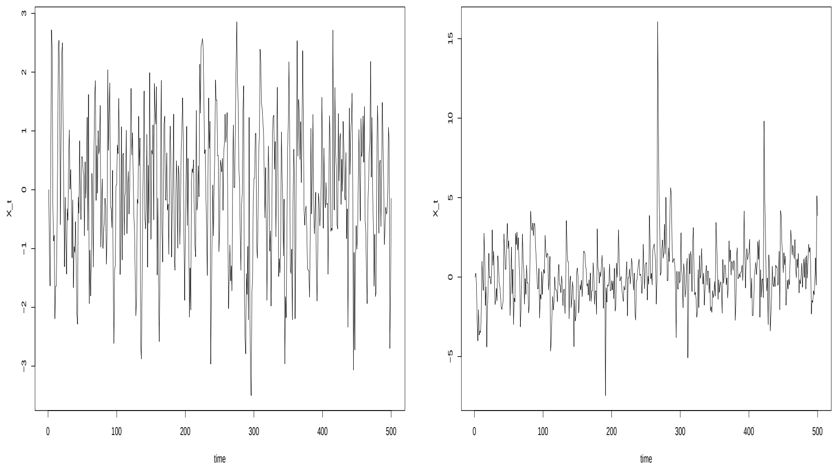


FIGURE 17. Simulation of AR(1) process $X_t = 0.5X_{t-1} + Z_t$ with iid standard normal noise (left) and student noise with 3 degrees of freedom (right).



## Which stylized facts can be explained by linear processes?

- **Heavy tails** of $X_t$ are only possible if the noise $Z_t$ has heavy tails, see p. 121.
- **ACF dependence.** $(X_t)$ must be iid or MA(1).

**Examples.** The ACFs of the iid sequences $(X_t)$ and $(|X_t|)$ are negligible at all lags.

The ACFs of an MA(1) process $(X_t)$ and of $(|X_t|)$ are zero at lags $\geq 2$ since $X_t$ and $X_{t+2}$ are independent.

A linear process cannot explain the complicated dependence structure of the sequences $(|X_t|)$ and $(X_t^2)$.

**Conclusion: We need a "non-linear" model.**



## 5.3. Multiplicative models.

$$X_t = \mu + \sigma_t \, Z_t\,, \quad t \in \mathbb{Z}\,.$$

We assume

- $(Z_t)$ iid mean zero (or symmetric) *noise*, $EZ_t^2 = 1$
- $\sigma_t$ and $Z_t$ independent
- $(\sigma_t)$ *volatility sequence* (unobservable) stationary
- $(X_t)$ stationary
- $\mu = 0$



### 5.3.1. Why this model?

- **Conditional forecast** of $X_t$ given $\sigma_t = f(\text{past})$. Then $\mathcal{L}(X_t \mid \text{past})$ is known, e.g. $Z_t \sim N(0,1)$ and $X_t \mid \text{past} \sim N(0, \sigma_t^2)$. (Conditional VaR)
- $$\rho_X(h) = \text{corr}(X_0, X_h) = 0\,, \quad h \neq 0\,,$$ in agreement with stylized facts.
- Although the whole time series $(X_t)$ is stationary one can model changing *conditional variance* over time quite flexibly. ("volatility clusters")

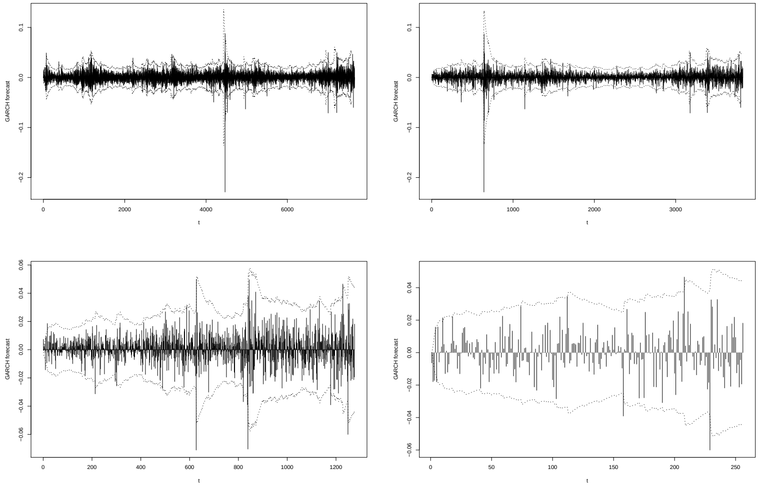

FIGURE 18. One day **95%** distributional forecasts of log-returns of the S&P500 composite stock index (from top left, top right, bottom left to bottom right: **30**, **15**, **5**, **1** years of data) based on a GARCH$(\mathbf{1},\mathbf{1})$ model with iid standard normal noise and parameters $\alpha_0 = 10^{-6}$, $\alpha_1 = 0.07$, $\beta_1 = 0.96$. The extreme values of the log-returns are not correctly captured by the model.



## 5.3.2. The ARCH family. Engle (1982)

$$X_t = \sigma_t Z_t, \quad (Z_t) \text{ iid}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2, \quad t \in \mathbb{Z}.$$

**for $\alpha_0 > 0$, certain $\alpha_i \geq 0$, $\alpha_p > 0$.**

**ARCH$(p)$ autoregressive conditionally heteroscedastic process of order $p$**



## WHY AR?

$$\nu_t = X_t^2 - \sigma_t^2 = \sigma_t^2 (Z_t^2 - 1),$$

**is white noise (zero correlations, constant variance) if $(\sigma_t^2)$ strictly stationary and $EX_t^4 < \infty$.**

$$\varphi(B) X_t^2 = \alpha_0 + \nu_t, \quad t \in \mathbb{Z},$$

**where**

$$\varphi(z) = 1 - \sum_{i=1}^{p} \alpha_i z^i,$$

**Notice: $(\nu_t)$ is not iid. Since it involves the sequence $(X_t)$ one knows a priori nothing about its existence and properties.**

**Problem: ARCH$(p)$ does not fit returns well unless $p$ is large.**



## 5.3.3. The GARCH model. Bollerslev (1986), Taylor (1986)

$$\varphi(B) X_t^2 = \alpha_0 + \beta(B) \nu_t, \quad t \in \mathbb{Z},$$

**where**

$$\varphi(z) = 1 - \sum_{i=1}^{p} \alpha_i z^i - \sum_{j=1}^{q} \beta_j z^j, \; \beta(z) = 1 + \sum_{j=1}^{q} \beta_j z^j,$$

**for certain $\alpha_i, \beta_j \geq 0$, $\alpha_p \beta_q > 0$.**

**Generalized ARCH$(p,q)$ (GARCH$(p,q)$)**

$$X_t = \sigma_t Z_t,$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \quad t \in \mathbb{Z}.$$

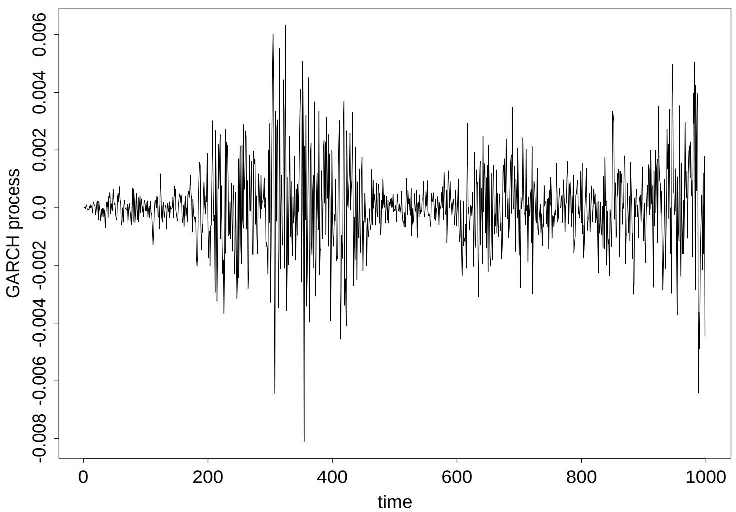


FIGURE 19. Simulation of a GARCH(1,1) process with Gaussian noise.



## WHY GARCH?

- Relation to ARMA processes (in general not very helpful for studying theoretical properties because the noise $\nu_t = X_t^2 - \sigma_t^2$ and $X_t^2$ are closely related)

- Reasonable fit to data (for not too long series) even for GARCH$(1,1)$ with *only 3 parameters*: residuals are iid-like.

- "Uncomplicated" statistical estimation of GARCH parameters.

  See Berkes, Horváth and Kokoszka (2003), Straumann, M. (2006) Straumann's Springer Lecture Notes in Statistics (2005).



### 5.3.4. An easier model: The stochastic volatility model see Davis and M. (2006).

$$X_t = \sigma_t \, Z_t \,, \quad (Z_t) \text{ iid centered or symmetric}$$

- $(\sigma_t)$ strictly stationary
- $(Z_t)$ and $(\sigma_t)$ independent
- Often $\log \sigma_t$ is assumed to be a Gaussian linear process:
$$\log \sigma_t = \sum_{j=0}^{\infty} c_j \, \eta_{t-j} \,, \quad \eta_j \sim N(0,1) \text{ iid}$$
- No feedback between $(\sigma_t)$ and $(Z_t)$
- Dependence modeled via $(\sigma_t)$, tails via $(Z_t)$
- Estimation more difficult since the likelihood function is not explicit. See Breidt and Carriquiry (1996); Shephard (1996) for a survey.



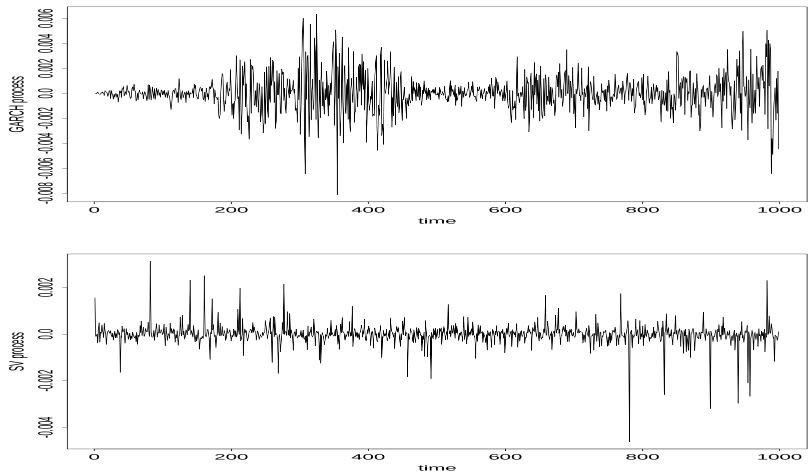


FIGURE 20. *Top:* 1000 simulated values from the GARCH$(1,1)$ model $X_t = (0.0001 + 0.1X_{t-1}^2 + 0.9\sigma_{t-1}^2)^{0.5} Z_t$ for iid standard normal $(Z_t)$. *Bottom:* 1000 simulated values from the stochastic volatility model $X_t = e^{Y_t} Z_t$ for iid student noise $(Z_t)$ with $4$ degrees of freedom, $Y_t = 0.5Y_{t-1} + 0.3\eta_{t-1} + \eta_t$ is an ARMA(1,1) process with iid standard normal noise $(\eta_t)$.

## 5.4. The stationarity problem.

### 5.4.1. The SV model.

$$X_t = \sigma_t\, Z_t\,, \quad (Z_t) \text{ iid}$$

$(\sigma_t)$ and $(Z_t)$ independent.

The log-volatility sequence

$$\log \sigma_t = \sum_{i=1}^{\infty} c_i\, \eta_{t-i}\,, \quad \text{iid } \eta_i \sim N(0,1)$$

is *strictly stationary* if and only if $\sum_i c_i^2 < \infty$.

$(X_t)$ is stationary if and only if $(\sigma_t)$ is stationary.



### 5.4.2. The GARCH model.

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i\, X_{t-i}^2 + \sum_{j=1}^{q} \beta_j\, \sigma_{t-j}^2\,.$$

The sequence $X_t = \sigma_t Z_t$ is stationary if $(\sigma_t)$ is stationary.

Example: the GARCH$(1,1)$ case

Write

$$A_t = \alpha_1\, Z_{t-1}^2 + \beta_1\,, \quad B_t = \alpha_0 \quad \text{and} \quad Y_t = \sigma_t^2\,.$$

Then

$$\sigma_t^2 = \alpha_0 + \alpha_1\, X_{t-1}^2 + \beta_1\, \sigma_{t-1}^2 = \alpha_0 + (\alpha_1 Z_{t-1}^2 + \beta_1)\, \sigma_{t-1}^2$$

or

$$Y_t = A_t\, Y_{t-1} + B_t\,. \tag{5.1}$$



- $A_t$ and $Y_{t-1}$ are independent.
- $((A_t, B_t))$ is iid.
- (5.1) is a stochastic recurrence equation (SRE) or random coefficient autoregressive model.
- For constant coefficient AR model ($A_t = \varphi$), $|\varphi| < 1$ is NASC for existence of a unique stationary *causal* solution.

Iterate back and notice that $B_t = B_1$,

$$Y_t = A_t \cdots A_{t-r} Y_{t-r-1} + \sum_{i=t-r}^{t} A_t \cdots A_{i+1}\, B_1\,.$$



Notice

$$\sum_{i=-\infty}^{t} A_t \cdots A_{i+1} = 1 + \sum_{i=-\infty}^{t-1} \exp\{(t-i)\, [\, \frac{1}{t-i} \sum_{j=i+1}^{t} \log A_j\, ]\, \}\,. \tag{5.2}$$

For fixed $t$, the SLLN gives as $i \to -\infty$,

$$\frac{1}{t-i} \sum_{j=i+1}^{t-1} \log A_j \overset{\text{a.s.}}{\to} E \log A_1\,,$$

Hence, if $-\infty \le E \log A_1 < 0$, (5.2) converges a.s. for every fixed $t$ and

$$\widetilde{Y}_t = \sum_{i=-\infty}^{t} A_t \cdots A_{i+1}\, B_1\,, \quad t \in \mathbb{Z}\,,$$

is a strictly stationary solution to SRE.

For any other stationary solution $(\widehat{Y}_t)$ to SRE:

$$|\widetilde{Y}_t - \widehat{Y}_t| = A_t \cdots A_{t-r}\, |\widetilde{Y}_{t-r-1} - \widehat{Y}_{t-r-1}|\,,$$

and since $A_t \cdots A_{t-r}$ and $|\widetilde{Y}_{t-r-1} - \widehat{Y}_{t-r-1}|$ are independent, the LLN and $E\log A_1 < 0$ imply $\widetilde{Y}_t = \widehat{Y}_t$ a.s. for every $t$.

**Theorem 5.1.** (Nelson (1990), Bougerol and Picard (1992a,b)) There exists an a.s. unique non-vanishing strictly stationary causal (i.e., depending only on past and present values of the $Z$'s) solution of SRE with $B_t \equiv \alpha_0$ if and only if

$$\alpha_0 > 0 \text{ and } E\log(\alpha_1 Z_1^2 + \beta_1) < 0.$$



- If $\alpha_0 = 0$, $X_t \equiv 0$ is a solution.
- Since $\log(\beta_1) \le E\log(\alpha_1 Z_1^2 + \beta_1) < 0$, $0 \le \beta_1 < 1$ is necessary for existence of a non-trivial solution.
- $E\log(A_1) = E\log(\alpha_1 Z_1^2 + \beta_1) < 0$ means that $A_t < 1$ "on log-average".



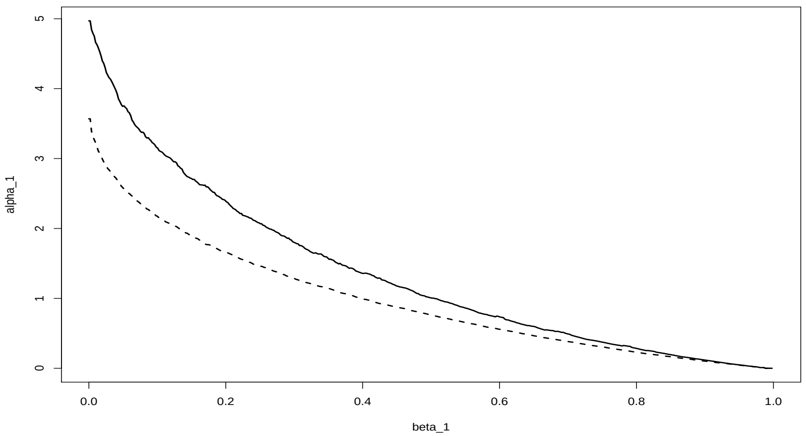


FIGURE 21. The $(\alpha_1, \beta_1)$-areas below the two curves guarantee the existence of a stationary GARCH$(1,1)$ process. *Solid line:* IID student noise with $4$ degrees of freedom with variance $1$. *Dotted line:* IID standard normal noise.



## The general GARCH case

$$\mathbf{Y}_t = \left(\sigma_{t+1}^2, \ldots, \sigma_{t-q+2}^2, X_t^2, \ldots, X_{t-p+2}^2\right)',$$

$$\mathbf{A}_t = \begin{pmatrix} \alpha_1 Z_t^2 + \beta_1 & \beta_2 & \cdots & \beta_{q-1} & \beta_q & \alpha_2 & \alpha_3 & \cdots & \alpha_p \\ 1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 \\ Z_t^2 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix},$$

$$\mathbf{B}_t = (\alpha_0, 0, \ldots, 0)',$$

Then

$$\mathbf{Y}_t = \mathbf{A}_t \mathbf{Y}_{t-1} + \mathbf{B}_t, \quad t \in \mathbb{Z}.$$

- This is a stochastic recurrence equation.
- Notice: $((A_t, B_t))$ iid, $Y_{t-1}$ and $(A_t, B_t)$ independent.
- Existence of stationary solution depends on top Lyapunov exponent

$$\gamma = \inf \{n^{-1} \, E \log \|A_n \cdots A_1\|\} < 0 \,,$$

$\|\cdot\|$ operator norm corresponding to norm $|\cdot|$.

- In general, $\gamma$ cannot be calculated explicitly.
- According to the subadditive ergodic theorem,

$$n^{-1} \log \|A_n \cdots A_1\| \stackrel{\text{a.s.}}{\to} \gamma \,.$$

- For GARCH(1,1), $A_n \cdots A_1 = \prod_{t=1}^n (\alpha_1 Z_t^2 + \beta_1)$, and so $\gamma = E \log(\alpha_1 Z_1^2 + \beta_1)$.



Theorem 5.2. (Bougerol and Picard (1992a,b))

The GARCH$(p, q)$ SRE has the a.s. unique strictly stationary non-vanishing causal (i.e., depending only on past and present $Z$'s) solution

$$Y_t = \sum_{i=-\infty}^{t} A_t \cdots A_{i+1} B_i = B_t + \sum_{i=-\infty}^{t-1} A_t \cdots A_{i+1} B_i \,,$$

if and only if $\alpha_0 > 0$ and $\gamma < 0$. Moreover, this solution is also ergodic.

- $\alpha_0 > 0$ ensures that $X_t \equiv 0$ a.s. is excluded.
- $\sum_{j=1}^q \beta_j < 1$ is a necessary condition for $\gamma < 0$.
- $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$ is sufficient for $\gamma < 0$ if $EZ_1^2 = 1$ and $EZ_1 = 0$, and ensures that $EX_t^2 < \infty$.



Corollary 5.3. The GARCH$(p, q)$ process $X_t = \sigma_t Z_t$, $t \in \mathbb{Z}$, with an iid mean zero and unit variance noise $(Z_t)$ has a non-vanishing strictly stationary causal ergodic version if and only if $\alpha_0 > 0$ and $\gamma < 0$.

The integrated GARCH process (IGARCH) Engle and Bollerslev (1986)

$$\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j = 1,$$

IGARCH process has infinite variance; see p. 116. This is not desirable (e.g. ACF would not make sense) and is in contrast to statistical evidence.




Figure 22. The estimated values of $\alpha_1 + \beta_1$ for an increasing sample of the *S&P500* log-returns.

- **How can we get heavy tails in these models in order to achieve agreement with statistical evidence from returns?**
- **What is a "multivariate tail"?**
- **How does one model/measure extremal dependence beyond covariances?**



## 6. An excursion to multivariate power law distributions: regular variation

- **There is statistical evidence that the tails of various real-life data, e.g. returns in finance, claim sizes in insurance, ON-periods in teletraffic, are well modeled by power laws.**
- **A probabilistic/analytical concept to describe *power law behavior* is regular variation.**
- **A regularly varying function is a power function which is (possibly) perturbed by a function of "lower order".**
- **This is in agreement with real-life data. In this case we can never expect that they come from a distribution with a pure power law tail.**



- **Regular variation is a natural condition which arises in various contexts in extreme value theory and the weak limit theory for sums of iid and weakly dependent stationary sequences.**



### 6.1. The univariate case, see Bingham et al. (1987).

- **A positive function on $(0,\infty)$ $L$ is slowly varying if it satisfies**

$$L(cx)/L(x)\to 1 \quad \text{as } x\to\infty, \text{ for each } c>0.$$

- **The random variable $X$ and its distribution $F$ are regularly varying with index $\alpha>0$ if there exist $p,q\ge 0$, $p+q=1$, and a slowly varying function $L$ such that as $x\to\infty$,**

$$\overline{F}(x) = P(X>x)\sim p\,\frac{L(x)}{x^\alpha}, \qquad (6.1)$$

$$F(-x) = P(X\le -x)\sim q\,\frac{L(x)}{x^\alpha}.$$

- **This is also called a tail balance condition.**

- It is a semi-parametric assumption about the tails; $L$ is not specified. In particular, its behavior in any neighborhood of the origin is not of interest.

Examples

Pareto

$$\overline{F}(x) = \frac{\kappa^\alpha}{(x+\kappa)^\alpha}\,, \quad x \ge 0\,, \quad \kappa > 0\,, \alpha > 0\,,$$

student

$$f(x) = \frac{\Gamma((n+1)/2)}{\Gamma(n/2)\sqrt{\pi n}}\,(1+x^2/n)^{-(n+1)/2}\,, \quad x \in \mathbb{R}\,,$$

log-gamma

$$f(x) = \frac{\alpha^\beta}{\Gamma(\beta)}\,(\log x)^{\beta-1}\,x^{-\alpha-1}\,, \quad x \ge 1\,, \quad \alpha, \beta > 0\,.$$



Further: (infinite variance) stable distributions, Cauchy, Burr, Fréchet

Notice: By virtue of Karamata's theorem, e.g. Bingham et al. (1987), for a regularly varying density $f$ with index $-\alpha$, $\alpha > 1$,

$$\overline{F}(x) = \int_x^\infty f(t)\,dt \sim (\alpha-1)^{-1} x\,f(x)\,.$$

- The tail balancing condition (6.1) is equivalent to: for any $t > 0$,

$$\lim_{x\to\infty} \frac{P(X > tx)}{P(|X| > x)} = p\,t^{-\alpha}\,, \quad \lim_{x\to\infty} \frac{P(X \le -tx)}{P(|X| > x)} = q\,t^{-\alpha}\,.$$



- The limiting expressions determine a measure $\mu$ on $\overline{\mathbb{R}}\backslash\{0\}$ given by

$$\mu(dt) = p\,\alpha\,t^{-\alpha-1}\,I_{(0,\infty]}(t)\,dt + q\,\alpha\,|t|^{-\alpha-1}\,I_{[-\infty,0)}(t)\,dt\,.$$

- In particular,

$$\lim_{x\to\infty} \frac{P(x^{-1}X \in (t,\infty])}{P(|X| > x)} = p\,t^{-\alpha} = \mu(t,\infty]\,,$$

$$\lim_{x\to\infty} \frac{P(x^{-1}X \in [-\infty,-t])}{P(|X| > x)} = q\,t^{-\alpha} = \mu[-\infty,-t]\,.$$

- Regular variation is equivalent to:

$$\lim_{x\to\infty} \frac{P(X \in xA)}{P(|X| > x)} = \mu(A) \text{ (vague convergence)}$$

for every Borel set $A$ bounded away from zero with $\mu(\partial A) = 0$.



- For such an $A$,

$$\mu(t\,A) = \lim_{x\to\infty} \frac{P(X \in x\,t\,A)}{P(|X| > t\,x)}\,\frac{P(|X| > t\,x)}{P(|X| > x)} = t^{-\alpha}\,\mu(A)\,.$$

- Hence $\mu$ satisfies the homogeneity property

$$\mu(tA) = t^{-\alpha}\mu(A) \quad \text{for every } t > 0.$$

- In spherical coordinates:

$$P(\theta = 1) = p \text{ and } P(\theta = -1) = q.$$

For any $t > 0$ and set $S \subset \{-1, 1\}$,

$$\lim_{x\to\infty} \frac{P\left(|X| > tx\,, X/|X| \in S\right)}{P(|X| > x)} = t^{-\alpha}\,P(\theta \in S)\,.$$

## Applications of univariate regular variation

$X_i \sim F$ iid

Partial maxima $M_n = \max(X_1, \ldots, X_n)$.

- If

$$\lim_{n\to\infty} P(c_n^{-1}(M_n - d_n) \le x) = H(x)\,, \quad x \in \mathbb{R}\,,$$

  $H$ is an extreme value distribution and $F$ is in the maximum domain of attraction of $H$ ($F \in \mathrm{MDA}(H)$).

- The Fréchet distribution is an extreme value distribution:

$$H(x) = \Phi_\alpha(x) = \exp\{-x^{-\alpha}\}\,, \quad x > 0\,.$$



- $F \in \mathrm{MDA}(\Phi_\alpha)$ if and only if $\overline{F} = 1 - F$ is regularly varying with index $-\alpha$.

- $\overline{F}(c_n) \sim n^{-1}$, i.e., $c_n = n^{1/\alpha}\ell(n)$ for some slowly varying $\ell$, and

$$c_n^{-1} M_n \xrightarrow{d} Y \sim \Phi_\alpha\,.$$

Partial sums $S_n = X_1 + \cdots + X_n$.

- If

$$\lim_{n\to\infty} P(c_n^{-1}(S_n - d_n) \le x) = H(x)\,, \quad x \in \mathbb{R}\,,$$

  $H$ is stable and $F$ is in the domain of attraction of the stable distribution $H$ ($F \in \mathrm{DA}(H)$).



- If $H$ has infinite variance (non-Gaussian), it is regularly varying with almost exact power law tails for some $\alpha < 2$,

$$H_\alpha(-x) \sim qcx^{-\alpha}\,, \quad \overline{H}_\alpha(x) \sim pcx^{-\alpha}\,, \quad x > 0\,.$$

- $F \in \mathrm{DA}(H_\alpha)$ if and only if $F$ is regularly varying with index $\alpha$.
- For $P(|X_1| > c_n) \sim n^{-1}$, i.e., $c_n = n^{1/\alpha}\ell(n)$ some slowly varying $\ell$,

$$c_n^{-1}(S_n - d_n) \xrightarrow{d} Y \sim H_\alpha\,.$$

- Only in a very few cases the density of an $\alpha$-stable distribution has a pleasant form, including the Cauchy distribution and the inverse Gaussian see Feller (1971), Samorodnitsky and Taqqu (1994).



## 6.2. Multivariate regular variation, see Resnick (1986,1987,2006).

- $\mathrm{X} \in \mathbb{R}^d$ and its distribution are regularly varying with index $\alpha > 0$:

  there exists $\Theta \in \mathbb{S}^{d-1}$ such that for any $t > 0$, $S \subset \mathbb{S}^{d-1}$ with $P(\Theta \in \partial S) = 0$,

$$\lim_{x\to\infty} \frac{P\left(|\mathrm{X}| > tx\,, \widetilde{\mathrm{X}} \in S\right)}{P(|\mathrm{X}| > x)} = t^{-\alpha} P(\Theta \in S)\,, \tag{6.2}$$

  where $\widetilde{\mathrm{x}} = \mathrm{x}/|\mathrm{x}|$. (weak convergence)

- $P_\Theta$ is the spectral measure of X.

- Another way of writing (6.2) is

$$\lim_{x\to\infty}\frac{P\left(|\mathrm{X}|>tx\,,\widetilde{\mathrm{X}}\in S\right)}{P(|\mathrm{X}|>x)}$$

$$=\lim_{x\to\infty}\frac{P(|\mathrm{X}|>tx)}{P(|\mathrm{X}|>x)}\lim_{x\to\infty}P\left(\widetilde{\mathrm{X}}\in S\;\Big|\;|\mathrm{X}|>x\right)$$

$$=t^{-\alpha}P(\Theta\in S)\,.$$

- There is "asymptotic independence" between the radial and spherical parts of X if |X| is large enough.
- The right-hand side of (6.2) can be interpreted as the $\mu$-measure of

$$A(t,S)=\{\mathrm{x}:\ |\mathrm{x}|>t\,,\quad \widetilde{\mathrm{x}}\in S\}\ ,$$



- and (6.2) can be written as

$$\lim_{x\to\infty}\frac{P\left(x^{-1}\mathrm{X}\in\ A(t,S)\right)}{P(|\mathrm{X}|>x)}=\mu(A(t,S))=t^{-\alpha}\,\mu(A(1,S))\,.$$

The values $\mu(A(t,S))$ determine a Radon measure $\mu$ (i.e., finite on compact sets) on $\overline{\mathbb{R}}^d\backslash\{0\}$.

- (6.2) is equivalent to vague convergence:

$$\lim_{x\to\infty}\frac{P(x^{-1}\mathrm{X}\in A)}{P(|\mathrm{X}|>x)}=\mu(A)\,,$$

for every $A$ bounded away from 0 with $\mu(\partial A)=0$, and $\mu$ is a measure on $\overline{\mathbb{R}}^d\backslash\{0\}$ with homogeneity property

$$\mu(tA)=t^{-\alpha}\mu(A)\,,\quad t>0\,.$$

- Vague convergence

$$\frac{P(\mathrm{X}\in x\,\cdot)}{P(|\mathrm{X}|>x)}\stackrel{v}{\to}\ \mu(\cdot)\,.$$



## Sequential definition of regular variation.

- Choose $P(|\mathrm{X}|>c_n)\sim n^{-1}$.
- Then regular variation of X with index $\alpha$ and limiting measure $\mu$ is equivalent to Hult and Lindskog (2005,2006a)

$$n\,P(c_n^{-1}\mathrm{X}\in\cdot)\ \stackrel{v}{\to}\ \mu(\cdot)\,,\tag{6.3}$$

$$n\,P\left(|\mathrm{X}|>tc_n\,,\widetilde{\mathrm{X}}\in\cdot\right)\ \stackrel{w}{\to}\ t^{-\alpha}P(\Theta\in\cdot)\,,\quad t>0\,.$$

- It is often more convenient to work with this equivalent notion of regular variation.



## Examples of multivariate regularly varying distributions

multivariate student

multivariate $F$

multivariate Cauchy

multivariate stable

multivariate Fréchet extreme value distributions

## 6.3. Special cases of dependence.

### 6.3.1. Example: Total independence.

- $\mathbf{X} = (X_1, X_2)$ iid $F$, $X_i > 0$, regularly varying with index $\alpha$, $|\mathbf{x}| = \max(x_1, x_2)$.
- Notice that

$$\begin{aligned} n\,P(|\mathbf{X}| > c_n) &= n\,(1 - P(\max(X_1, X_2) \le c_n)) \\ &= [n\,\overline{F}(c_n)]\,(1 + F(c_n)) \\ &\sim 2\,[n\,\overline{F}(c_n)] \sim 1 \end{aligned}$$

and therefore

$$\overline{F}(c_n) \sim 0.5\,n^{-1}\,.$$

- Hence for positive $x, y$,

$$\frac{P(X_1 > c_n\,x\,, X_2 > c_n\,y)}{P(|\mathbf{X}| > c_n)} \sim \frac{\overline{F}(x\,c_n)\,\overline{F}(y\,c_n)}{2\,\overline{F}(c_n)} \to 0\,.$$

- The spectral measure is concentrated at the intersection of the axes and the unit circle.



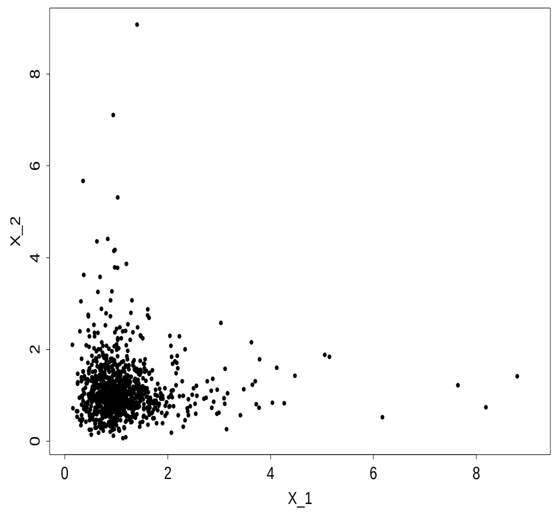


Figure 23. Plot of **1000** realizations from an iid sample $(\mathbf{X}_i)$ with iid components satisfying $P(X_1 > x) \sim x^{-3}$. The $\mathbf{X}_i$'s with large distance from the origin are concentrated along the axes.



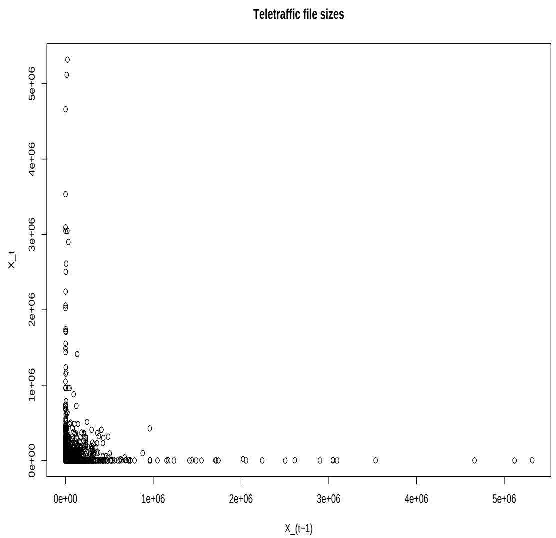


Figure 24. Scatterplot of file sizes of teletraffic data.



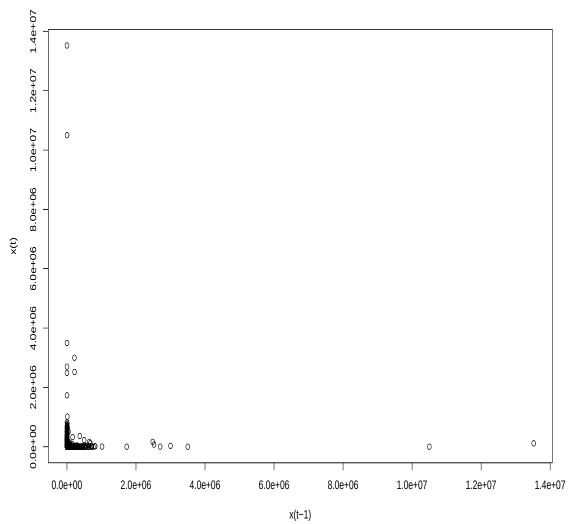


Figure 25. Scatterplot of US fire insurance losses.

### 6.3.2. Example: Total dependence.

- $\mathbf{X} = (X, X)$ for some regularly varying $X > 0$ with index $\alpha > 0$, $|\mathbf{x}| = \max(x_1, x_2)$.

- Then, since $P(|\mathbf{X}| > c_n) = P(X > c_n) \sim 1/n$,

$$n\,P(|\mathbf{X}| > c_n\,, \widetilde{\mathbf{X}} \in S) \;=\; n\,P(X > c_n\,, (1,1) \in S)$$
$$\sim\; I_{(1,1)}(S)\,.$$

- The spectral measure is degenerate and concentrated at the intersection of the unit sphere with the line $x = y$.



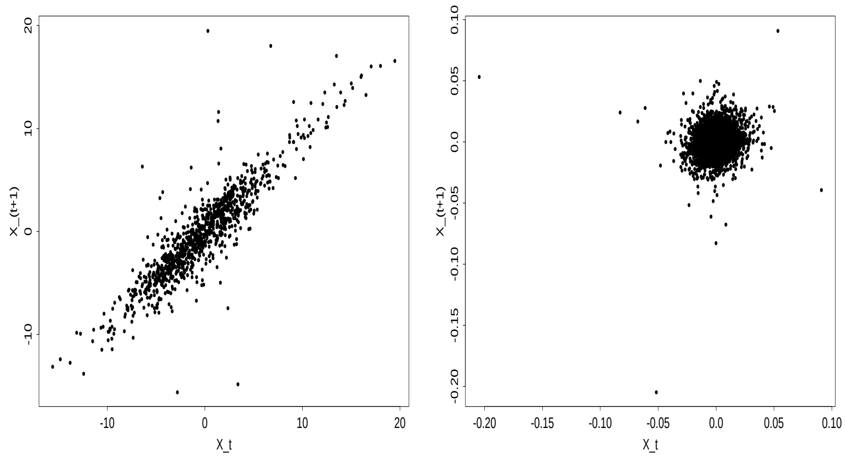


FIGURE 26. Left: Plot of $\mathbf{1000}$ lagged vectors $\mathbf{X}_t = (X_t, X_{t+1})$ for the AR(1) process $X_{t+1} = \mathbf{0.9}X_t + Z_t$ for iid symmetric regularly varying noise $(Z_t)$ with tail index $\mathbf{1.8}$. The vectors $\mathbf{X}_t$ with a large norm $|\mathbf{X}_t|$ are typically concentrated along the line $y = \mathbf{0.9}x$. Right: Scatterplot of the pairs $(X_t, X_{t+1})$ of the daily log-returns $X_t$ of the *S&P500* series. The extremes in the series do not tend to cluster around the axes.



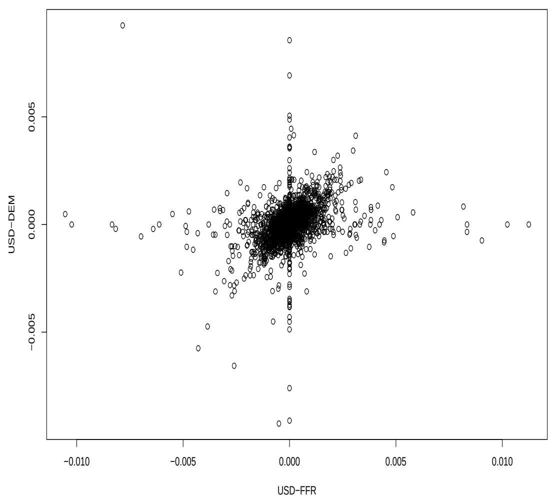


FIGURE 27. Scatterplot of 5 minute foreign exchange rate log-returns, USD-DEM against USD-FRF.



### 6.3.3. A toy model: The "2-dimensional Pareto distribution".

- Assume

$$\mathbf{X} = R\,(\cos\Phi\,, \sin\Phi)\,,$$

where

$$P(R > r) = r^{-\alpha}\,, \quad r \ge 1\,, \quad \text{for some } \alpha > 0,$$

and $R$ is independent of $\Phi$ with distribution on $(-\pi, \pi]$.

- Take Euclidean norm. Then $\mathbf{X}$ is regularly varying with index $\alpha$ and $\Theta = (\cos\Phi, \sin\Phi)$:

$$\frac{P(|\mathbf{X}| > t\,x\,, \widetilde{\mathbf{X}} \in S)}{P(|\mathbf{X}| > x)} = \frac{P(R > t\,x\,, \Theta \in S)}{P(R > x)}$$
$$= \frac{P(R > t\,x)}{P(R > x)}\,P(\Theta \in S)$$
$$= t^{-\alpha}\,P(\Theta \in S)\,,$$

provided $\min(t\,x, x) \ge 1$.

- The knowledge of the distribution of $\Phi$ allows for some straightforward interpretation of the two-dimensional dependence in the tails.

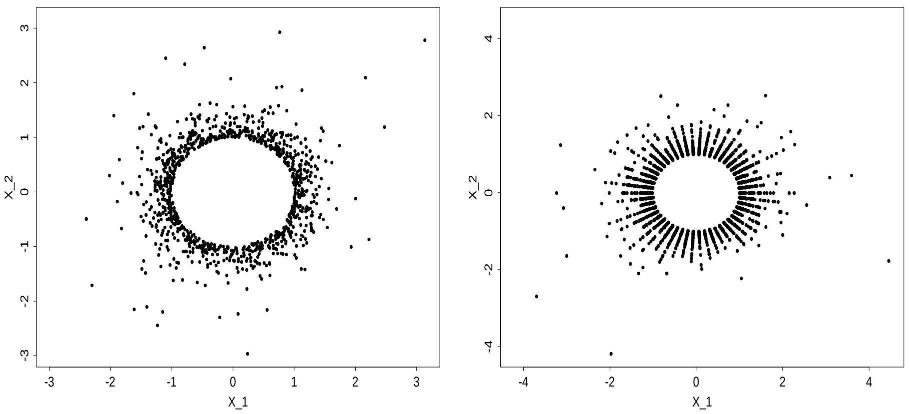


FIGURE 28. IID vectors $\mathbf{X}_i$ from the toy model with tail index $\alpha = 5$. Left: $\Phi$ is uniform on $(-\pi, \pi]$. Right: $\Phi$ has a discrete uniform distribution on the points $2\pi i/50$.



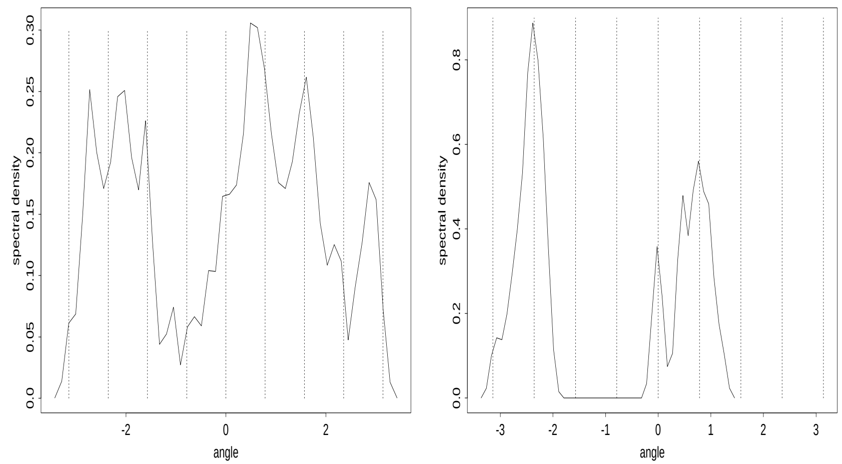


FIGURE 29. *Left:* Spectral density estimation for the pairs $(X_t, X_{t+1})$ of the *S&P500* returns $(X_t)$. The vertical lines indicate multiples of $\pi/4$. *Right:* Spectral density estimation for the pairs $(X_t, Y_t)$, where $(X_t)$ is the log-return series of the DAX, $(Y_t)$ the corresponding series for the CAC40 for the period from September $21$, $1988$, till August $24$, $1995$.



## 6.4. Where does multivariate regular variation appear?

6.4.1. **In extreme value theory.** $(\mathrm{X}_i)$ iid with positive components. Regular variation of $\mathrm{X}_i$ with index $\alpha$ is NASC for

$$c_n^{-1}\left(\max_{i\le n} X_i^{(1)}, \ldots, \max_{i\le n} X_i^{(d)}\right) \stackrel{d}{\to} \mathrm{Y} \sim \Phi_\alpha\,,$$

where $\Phi_\alpha$ has Fréchet marginals. de Haan and Resnick (1977), see Resnick (1987)

6.4.2. **In summation theory.** $(\mathrm{X}_i)$ iid regularly varying with index $\alpha \in (0,2)$, NASC for

$$a_n^{-1}\left(\mathrm{X}_1 + \cdots + \mathrm{X}_n - \mathrm{b}_n\right) \stackrel{d}{\to} \mathrm{Y} \sim \mathrm{S}_\alpha\,,$$

where $\mathrm{S}_\alpha$ is $\alpha$-stable, $\alpha \in (0,2)$. Rvačeva (1962)



6.4.3. **In stochastic recurrence equations.**

- $((\mathrm{A}_i, \mathrm{B}_i))$ iid pairs of non-negative matrices and vectors, $(\mathrm{X}_t)$ stationary solution to the stochastic recurrence equation

$$\mathrm{X}_t = \mathrm{A}_t\,\mathrm{X}_{t-1} + \mathrm{B}_t\,, \quad t \in \mathbb{Z}\,. \tag{6.4}$$

- Under general conditions,

$$P\left((\mathrm{x}, \mathrm{X}_1) > x\right) \sim c(\mathrm{x})\, x^{-\alpha}\,, \quad x \to \infty\,, \tag{6.5}$$

  for some $\alpha > 0$, and $c(\mathrm{x}) \neq 0$ for $\mathrm{x} \in \mathbb{R}_+^d \backslash \{0\}$. Kesten (1973)
- In general, there is no Cramér-Wold device for regular variation (Kesten (1973), Hult, Lindskog (2006b)).

- However, for $\mathrm{A}_t, \mathrm{B}_t$ with non-negative components (6.5), for all $\mathrm{x} \neq 0$, is equivalent to regular variation of X. See Basrak, Davis, M. (2002a) for non-integer $\alpha$, Boman and Lindskog (2007) in the general case.

### 6.4.4. GARCH processes are regularly varying.

- Recall from p. 80 that the squared GARCH$(p,q)$ vector

$$\left(\sigma_{t+1}^2, \ldots, \sigma_{t-q+2}^2, X_t^2, \ldots, X_{t-p+2}^2\right)'$$

satisfies a stochastic recurrence equation of type (6.4).

- This fact and Kesten's result ensure that the finite-dimensional distributions of the GARCH sequences $(X_t)$, $(|X_t|)$, $(\sigma_t)$ are regularly varying with some index $\alpha > 0$ provided the noise



$(Z_t)$ has a density with infinite support. The spectral measures are not concentrated on the axes. Basrak, Davis, M. (2002b).

- In the general stochastic recurrence equation (6.4), $\alpha$ is given as the solution to the equation

$$0 = \lim_{n\to\infty} \frac{1}{n} \log E\|\mathrm{A}_n \cdots \mathrm{A}_1\|^{\alpha/2}. \qquad (6.6)$$

- Recall from p. 74 that for GARCH$(1,1)$, $A_t = \alpha_1 Z_{t-1}^2 + \beta_1$. Then (6.6) degenerates to

$$E(\alpha_1 Z^2 + \beta_1)^{\alpha/2} = 1\,.$$

- This equation can be solved numerically for nice densities of $Z$.



Table 1. Results for $\alpha$ when $\alpha_1 = 0.1$. Top: Standard normal noise. Bottom: Unit variance student noise with 4 degrees of freedom.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\beta_1$ | 0.9 | 0.8 | 0.7 | 0.6 | 0.5 | 0.4 | 0.3 | 0.2 | 0.1 | 0.0 |
| $\alpha$ | 2.0 | 12.5 | 16.2 | 18.5 | 20.2 | 21.7 | 23.0 | 24.2 | 25.4 | 26.5 |
| $\beta_1$ | 0.81 | 0.82 | 0.83 | 0.84 | 0.85 | 0.86 | 0.87 | 0.88 | 0.89 | |
| $\alpha$ | 11.9 | 11.3 | 10.7 | 9.9 | 9.1 | 8.1 | 7.0 | 5.6 | 4.0 | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\beta_1$ | 0.9 | 0.8 | 0.7 | 0.6 | 0.5 | 0.4 | 0.3 | 0.2 | 0.1 | 0.0 |
| $\alpha$ | 2.0 | 3.68 | 3.83 | 3.88 | 3.91 | 3.92 | 3.93 | 3.93 | 3.94 | 3.94 |
| $\beta_1$ | 0.81 | 0.82 | 0.83 | 0.84 | 0.85 | 0.86 | 0.87 | 0.88 | 0.89 | |
| $\alpha$ | 3.65 | 3.61 | 3.56 | 3.49 | 3.41 | 3.29 | 3.13 | 2.90 | 2.54 | |



### 6.4.5. The IGARCH(1,1) case $\alpha_1 + \beta_1 = 1$.

- Since $EZ_1^2 = 1$,

$$EA_1^{\alpha/2} = E(\alpha_1 Z_1^2 + \beta_1)^{\alpha/2} = 1$$

has the unique solution $\alpha = 2$.

- Kesten yields regular variation with index 2 for the finite-dimensional distributions of IGARCH:

$$x^2 P\left(|\mathrm{X}_d| > tx\,, \widetilde{\mathrm{X}}_d \in \cdot\right) \stackrel{v}{\to} c\,t^{-2}\,P(\Theta \in \cdot) \quad \text{as } x \to \infty.$$

- In particular, $X_t$ has infinite variance.

- This remains valid in the general case:

$$\begin{aligned} E\sigma_1^2 &= \alpha_0 + EX_1^2 \sum_{i=1}^{p} \alpha_i + E\sigma_1^2 \sum_{j=1}^{q} \beta_j \\ &= \alpha_0 + E\sigma_1^2 \left[ \sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j \right] \\ &= \alpha_0 + E\sigma_1^2 . \end{aligned}$$

  Since $\alpha_0 > 0$ is necessary for stationarity in the non-degenerate case, the only possible conclusion is that $E\sigma_1^2 = EX_1^2 = \infty$.
- Similar arguments apply to the case $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j > 1$.
- It is an open question as to whether the finite-dimensional distributions of the general IGARCH$(p, q)$ process are regularly varying with index $\alpha = 2$.



## 6.5. Operations on regularly varying random vectors, see Jessen, M. (2006).

### 6.5.1. Linear combinations.

- Linear combinations of the components of a regularly varying random vector X are regularly varying.
- Write

$$A_{\mathrm{x}} = \{\mathrm{y} : (\mathrm{x}, \mathrm{y}) > 1\} \quad \text{for } \mathrm{x} \neq 0.$$

- Then

$$\lim_{x\to\infty} \frac{P((\mathrm{x}, \mathrm{X}) > x)}{P(|\mathrm{X}| > x)} = \lim_{x\to\infty} \frac{P(x^{-1}\mathrm{X} \in A_{\mathrm{x}})}{P(|\mathrm{X}| > x)} = \mu(A_{\mathrm{x}}),$$

  hence $(\mathrm{x}, \mathrm{X})$ is univariate regularly varying with the same index as X provided that $\mu(A_{\mathrm{x}}) > 0$.



- Now suppose that the following condition holds:

$$\text{(6.7)} \qquad \text{For all } \mathrm{x} \neq 0, \quad \lim_{u\to\infty} \frac{P((\mathrm{x}, \mathrm{X}) > u)}{u^{-\alpha} L(u)} = w(\mathrm{x}) \quad \text{exists},$$

  for some slowly varying $L$, $\alpha > 0$, where $w$ is a finite-valued function, $w(\mathrm{x}) \neq 0$ for at least one $\mathrm{x} \neq 0$.
- Then

$$w(t\mathrm{x}) = t^{-\alpha} w(\mathrm{x}), \qquad \text{for all } t > 0,\ \mathrm{x} \neq 0.$$

**Theorem 6.1.** (Basrak, Davis, M. (2002a), Boman and Lindskog (2007)) X is regularly varying with index $\alpha > 0$ with a unique spectral measure

(1) if X satisfies (6.7) for some non-integer $\alpha > 0$,

(2) if X assumes values in $[0, \infty)^d$ and satisfies (6.7) for some $\alpha$ integer.



There exist examples for $\alpha$ integer when (6.7) does not imply regular variation of X, Hult and Lindskog (2006b), i.e., there is no Cramér-Wold device for regular variation.

### 6.5.2. Aggregation.

- If $\mathrm{X} = (X_1, \ldots, X_n)$ is regularly varying with index $\alpha$, so is $X_1 + \cdots + X_n$, and then, for example,

$$\frac{P(X_1 + \cdots + X_n > x)}{P(|\mathrm{X}| > x)} \to \mu(\{\mathrm{x} : x_1 + \cdots + x_n > 1\}).$$

- If a vector of returns $(X_1, \ldots, X_n)$ is regularly varying with index $\alpha$ (e.g. GARCH), the aggregated vector is regularly varying with the same index $\alpha$.

6.5.3. **Infinite moving averages.**

- $(Z_t)$ iid regularly varying with index $\alpha > 0$ and tail balance condition

$$P(Z_1 > x) \sim p\, P(|Z_1| > x)$$

$$P(Z_1 \le -x) \sim q\, P(|Z_1| > x)\,,$$

where $p + q = 1$.

- Then $(Z_1, \ldots, Z_n)$ is regularly varying with spectral measure at the intersection of the unit sphere $\mathbb{S}^{n-1}$ and the axes.
- Hence linear combinations of finitely many $Z_i$'s are regularly varying.



- This remains valid for infinite moving averages (e.g. FARIMA)

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}\,.$$

- Then (see M., Samorodnitsky (2000) for most general conditions)

$$\lim_{x\to\infty} \frac{P(X_1 > x)}{P(|Z_1| > x)} = \sum_{j=-\infty}^{\infty} [p\, (\psi_j)_+^{\alpha} + q\, (\psi_j)_-^{\alpha}]\,.$$

- More generally, the finite-dimensional distributions of $(X_t)$ are regularly varying.
- Only heavy-tailed input causes heavy-tailed output.
- For non-linear processes, light-tailed input (e.g. normal) can cause heavy-tailed output (e.g. GARCH, see p. 114).



6.5.4. **Products.**

- $X$ and $Y$ independent, non-negative. $X$ is regularly varying with index $\alpha > 0$, $EY^{\alpha+\epsilon} < \infty$ for some $\epsilon > 0$.
- Then $XY$ is regularly varying with index $\alpha$, and Breiman (1965)

$$P(XY > x) \sim EY^{\alpha}\, P(X > x) \quad \text{as } x \to \infty.$$

Example. $P(X > x) = x^{-\alpha}$, $x \ge 1$. Then

$$P(XY > x) = E[P(XY > x \mid Y)]$$

$$= \int_0^{\infty} P(X > x/y)\, dP(Y \le y)$$

$$= \int_0^{x} (y/x)^{\alpha}\, dP(Y \le y) + P(Y > x) \sim x^{-\alpha} EY^{\alpha}\,.$$



- $\mathrm{X} \in \mathbb{R}^d$ regularly varying, A is a $q \times d$ matrix, possibly random, independent of X with

$$0 < E\|\mathrm{A}\|^{\alpha+\epsilon} < \infty \quad \text{for some } \epsilon > 0.$$

- Then

$$\frac{P\left(x^{-1}\mathrm{AX} \in \cdot\right)}{P(|\mathrm{X}| > x)} \xrightarrow{v} \widetilde{\mu}(\cdot) := E[\mu \circ \mathrm{A}^{-1}(\cdot)]\,,$$

where $\xrightarrow{v}$ denotes vague convergence in $\overline{\mathbb{R}}^d \backslash \{0\}$.

Multivariate Breiman, see Basrak, Davis, M. (2002b).

6.5.5. **Application: The tails of stochastic volatility models,** Davis and M. (2001a,b, 2007b).

$$X_t = \sigma_t Z_t\,, \quad \log \sigma_t = \sum_{i=0}^{\infty} c_i \eta_{t-i}\,.$$

$\eta_i \sim N(0,1)$ iid, independent of iid $Z_t$ regularly varying with index $\alpha > 0$ such that $P(Z_1 > x)/P(|Z_1| > x) \to p > 0$.

- Then $\sigma_t$ is log-normal, hence $E\sigma_t^p < \infty$ for all $p > 0$.
- One-dimensional Breiman implies

$$P(X_1 > x) = P(\sigma_1 Z_1 > x) \sim E\sigma_1^\alpha\, P(Z_1 > x)$$

$$P(X_1 \le -x) \sim E\sigma_1^\alpha\, P(Z_1 \le -x)\,. \tag{6.8}$$



The one-dimensional marginals of an stochastic volatility model are regularly varying with index $\alpha$. See also Breidt and Davis (1998) for the case of a *light-tailed stochastic volatility model* with $Z$ normal.

- Multivariate Breiman implies

$$(X_1, \dots, X_n)' = \operatorname{diag}(\sigma_1, \dots, \sigma_n)\,(Z_1\, \dots, Z_n)'$$

is regularly varying since $\mathrm{Z} = (Z_1\, \dots, Z_n)'$ is regularly varying, independent of $\operatorname{diag}(\sigma_1, \dots, \sigma_n)$.



- We have for $i \neq j$, by Breiman,

$$\frac{P(X_i > x\,, X_j > x)}{P(|X_1| > x)} = \frac{E[P(\sigma_i Z_i > x\,, \sigma_j Z_j > x \mid (\sigma_t))]}{P(|X_1| > x)} = \frac{E[P(\sigma_i Z_1 > x \mid \sigma_i)\, P(\sigma_j Z_1 > x \mid \sigma_j)]}{P(\sigma_1 |Z_1| > x)} \sim \frac{E(\sigma_i\sigma_j)^\alpha [P(Z_1 > x)]^2}{E\sigma_1^\alpha P(|Z_1| > x)} \to 0\,.$$

- Similar calculations for left/right tails.
- Hence the spectral measures of the finite-dimensional distributions of $(X_t)$ are concentrated at the intersection of the unit spheres and the axes, as in the iid case, see p. 101
- Upper tail dependence coefficients for $i \neq j$ are zero:

$$\lambda_U = \lim_{x\to\infty} P(X_i > x \mid X_j > x) = 0\,.$$

as in the iid case.



- The extremal index (see p. 55) in this stochastic volatility model is $\theta = 1$ as in the iid case.
- On the other hand, the ACFs of $(|X_t|)$ and $(X_t^2)$ (if exist) can decay to zero arbitrarily slowly, i.e., extremal dependence is not related to the ACF.
- In order to get heavy tails (regular variation) in the considered stochastic volatility model with lognormal volatility one needs that the multiplicative noise $(Z_t)$ is regularly varying.
- This is in contrast to GARCH. In this case, light-tailed multiplicative noise (e.g. normal) causes regular variation of the volatility sequence $(\sigma_t)$ with index $\alpha > 0$.

- In particular, by Breiman

$$P(X_t > x) = P(\sigma_t Z_t > x) \sim EZ_+^\alpha P(\sigma > x) \sim EZ_+^\alpha c\, x^{-\alpha}$$

$$P(X_t \le -x) = P(\sigma_t Z_t \le -x) \sim EZ_-^\alpha P(\sigma > x) \sim EZ_-^\alpha c\, x^{-\alpha}.$$

- Compare with the corresponding result (6.8) for a SV model.



## 6.6. Convergence of componentwise maxima and regular variation.

- $\mathrm{X}_i = (X_i, Y_i)$ iid regularly varying with non-negative components.
- The componentwise maxima

$$M_n(X) = \max_{i \le n} X_i \quad \text{and} \quad M_n(Y) = \max_{i \le n} Y_i.$$

- Write $C_{\mathrm{a}} = [0, \mathrm{a}]^c$, where $\mathrm{a} \in [0, \infty]^2 \backslash \{0\}$.
- Then by a Taylor expansion and regular variation

$$P(c_n^{-1} M_n(X) \le a_1, c_n^{-1} M_n(Y) \le a_2)$$
$$= (1 - P(c_n^{-1} \mathrm{X}_1 \in C_{\mathrm{a}}))^n$$
$$\sim \exp\{-n\, P(c_n^{-1} \mathrm{X}_1 \in C_{\mathrm{a}})\} \to \exp\{-\mu(C_{\mathrm{a}})\}.$$



- The totality of the relations

$$n\, P(c_n^{-1} \mathrm{X}_1 \in C_{\mathrm{a}}) \to \mu(C_{\mathrm{a}})$$

for a non-null measure $\mu$ is equivalent to regular variation.

- Therefore: convergence in distribution of the partial maxima of iid regularly varying vectors $\mathrm{X}_i$ with non-negative components is equivalent to the point process convergence above. See Resnick (1987)
- An alternative derivation Let $N_n = \sum_{t=1}^n \varepsilon_{(X_t, Y_t)/c_n}$. Then regular variation of $(X, Y)$ is equivalent to $N_n \xrightarrow{d} N$ for a PRM($\mu$)), $N$, with $\mu(tA) = t^{-\alpha} \mu(A)$.



- Then

$$P(c_n^{-1} M_n(X) \le a_1, c_n^{-1} M_n(Y) \le a_2) = P(N_n(C_{\mathrm{a}}) = 0)$$
$$\to P(N(C_{\mathrm{a}}) = 0) = \mathrm{e}^{-\mu(C_{\mathrm{a}})}.$$

## 6.7. Maximum domains of attraction and copulas Galambos (1978), Resnick (1987).

- Let $\mathrm{X}_i = (X_i, Y_i)$ be iid vectors (the restriction to $d = 2$ is inessential) with continuous distribution $F$.
- $F \in \mathrm{MDA}(H)$ for some extreme value distribution $H$ if the componentwise maxima converge in distribution:

$$[c_{n,X}^{-1}(M_n(X) - d_{n,X}), c_{n,Y}^{-1}(M_n(Y)) - d_{n,Y})] \xrightarrow{d} \mathrm{Y} \sim H.$$

- It is (theoretically) possible to transform both marginals of $\mathbf{X}_i$ to some standard distribution.
- In extreme value theory it is common to transform both marginals to unit Fréchet marginals with distribution function $\Phi_1(x) = \mathrm{e}^{-x^{-1}}$, $x > 0$.
- After the transformation $F_* \in \mathrm{MDA}(H_*)$ for an extreme value distribution $H_*$ with unit Fréchet marginals.
- In particular,

$$n^{-1}(M_n(X_*), M_n(Y_*)) \xrightarrow{d} \mathbf{Z}_* \sim H_* \,.$$

- This means that $F \in \mathrm{MDA}(H)$ if and only if $F_*$ is regularly varying with index 1 and unit Fréchet marginals.



- Alternatively, one can transform the marginals to uniform $U(0,1)$ (by the quantile transform $F^{\leftarrow}$).
- The resulting multivariate distribution on $[0,1]^2$ is called the copula of $F$.
- If $F$ is an extreme value distribution, the copula is called extreme value copula.
- It is known Galambos (1978) that $F \in \mathrm{MDA}(H)$ for some extreme value distribution $H$ if and only if the copula $C$ of $F$, i.e.,

$$F(x_1, x_2) = C(F_X(x_1), F_Y(x_2)) \,, \quad x_1, x_2 \in \mathbb{R} \,,$$

satisfies

$$C^t(x^{1/t}, y^{1/t}) \to D(x,y) \,, \quad t \to \infty \,, \qquad (6.9)$$

for some copula $D$ with $D(x^t, y^t) = D^t(x,y)$.



- McNeil, Frey, Embrechts (2005) say that $C$ is in the copula domain of attraction of $D$ if (6.9) holds.
- This notion does not add anything to the theory.
- On the contrary, this notion is confusing since it suggests that copula domain of attraction and MDA of an extreme value distribution are different objects.



## 6.8. Upper and lower tail dependence coefficients.

- Let $\mathbf{X} = (X_1, X_2) \sim F$ have identical marginals with support on $(0, \infty)^2$.
- The limits (if exist)

$$\lambda_U = \lim_{x \to \infty} P(X_1 > x \mid X_2 > x) \,,$$
$$\lambda_L = \lim_{x \downarrow 0} P(X_1 \le x \mid X_2 \le x)$$

are the upper/lower tail dependence coefficients of $F$.
- These limits do in general not exist.
- One needs *some kind of a regular variation condition* which in general will be weaker than multivariate regular variation of $(X_1, X_2)$, e.g. hidden regular variation, see Resnick (2006).

- If X is regularly varying with index $\alpha$ and limiting measure $\mu$,

$$
\begin{aligned}
P(X_1 > x \mid X_2 > x) &= \frac{P(X_1 > x\,, X_2 > x)}{P(X_1 > x)} \\
&= \frac{P(x^{-1}\mathrm{X} \in (1,\infty)^2)/P(|\mathrm{X}| > x)}{P(x^{-1}\mathrm{X} \in (1,\infty) \times (0,\infty))/P(|\mathrm{X}| > x)} \\
&\to \frac{\mu((1,\infty)^2)}{\mu((1,\infty) \times (0,\infty))} = \lambda_U\,.
\end{aligned}
$$

- Since the upper tail dependence coefficient focuses on the sets $(1,\infty)^2$ it is a rather restricted way of studying extremes.
- The existence of the upper tail index does not ensure multivariate regular variation of X.



- About the use and abuse of copulas, see Mikosch (2006) and the corresponding discussion in the journal *Extremes*.



# 7. Modeling teletraffic

## 7.1. Some facts.

- Since the beginning of the 1990s models have been proposed for large communication networks (Internet, local area networks,...).
- Classical queuing models for waiting and service times fail to explain typical behavior.
- There is general agreement that the process of active sources exhibits long range dependence. This notion makes sense for stationary processes.



- The integrated process (cumulative input) is believed to be well approximated by a self-similar process (such as fractional Brownian motion, stable Lévy motion).
- Although the expected cumulative input is growing roughly linearly through time (such as in classical queuing networks) there are strong deviations from linearity due to erratic behavior.
- Since work by Taqqu, Willinger, Leland, Crovella,... (1993–) and others the assumption of heavy tailed distributions for file sizes, transmission durations, transmission rates,... has been accepted as a reasonable working hypothesis.

- There exists rather convincing evidence that file sizes, transmission durations, transmission rates,... have Pareto like distributions:

$$\overline{F}_X(x) = P(X_t > x) \approx x^{-\alpha}, \quad x \to \infty.$$

- Given the stationarity of the process of active sources, $\alpha$ is often found to be between 1 and 2. (infinite variance)



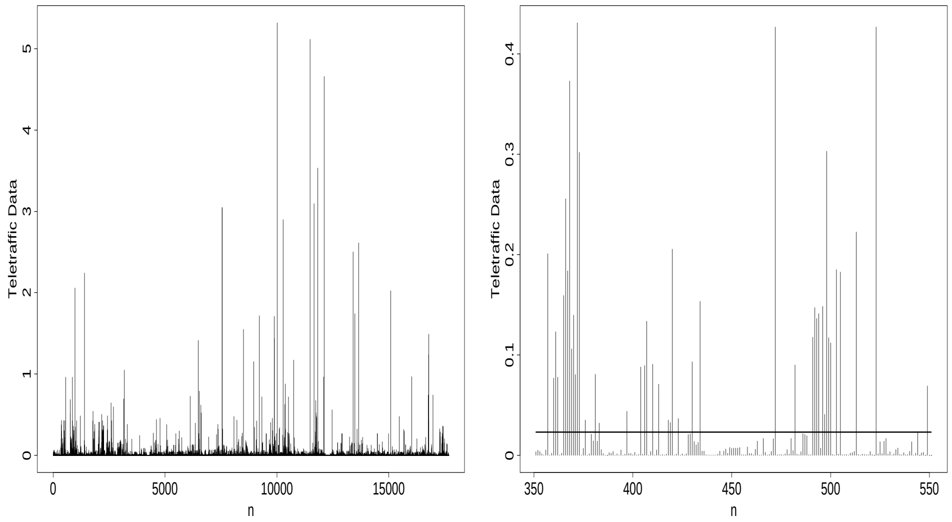


Figure 30. Time series of transmission durations (BU data).



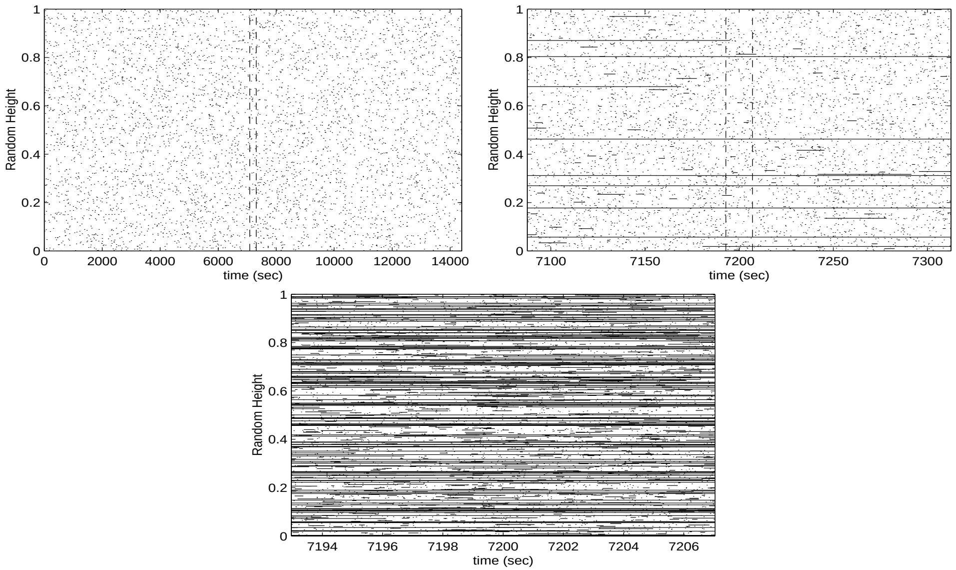


Figure 31.

Mice and elephants plots (S. Marron).



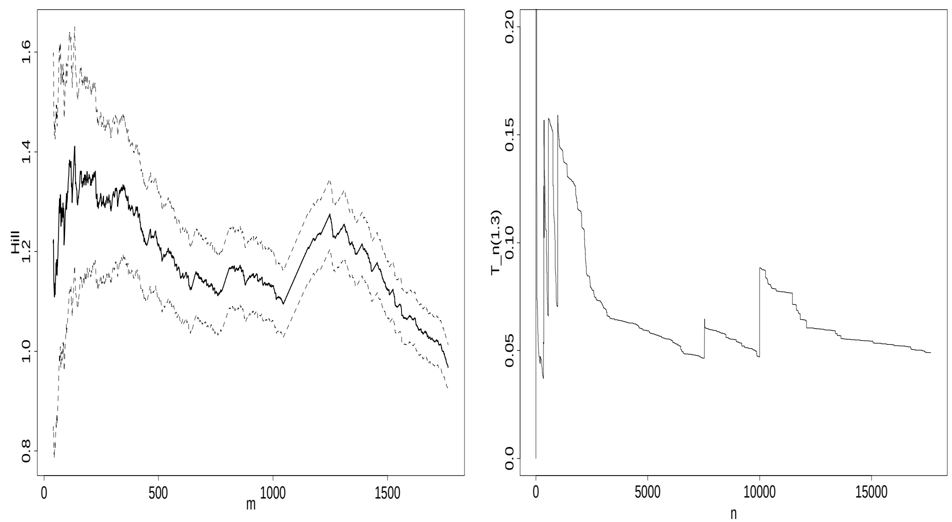


Figure 32.

Methods for determining $\alpha$.

## 7.2. Standard models for the process of active sources.

- Communication networks are too complex to be understood in detail.
- They are run by machines which are very fast (in contrast to human beings) and therefore fail a lot (in contrast to human beings who can use their brains).
- Although we do (perhaps) understand a single machine (car) and we know that the machines' joint behavior (Autobahn) is directed by a protocol (traffic lights, police) we do not understand their interplay (e.g. traffic jam).
- Therefore any model is nothing but a simplistic proxy to reality.



- A "realistic" model should, to some extent, explain the observed phenomena: self-similarity of cumulative input, interplay between long range dependence of activity process and heavy tailed components.



### 7.2.1. The ON/OFF process.

- During a transmission, a source transmits at unit rate. Otherwise, it is silent.
- Lengths of ON and OFF periods are described by two independent iid sequences of positive random variables.
- The ON periods have heavy-tailed distribution.
- The activity of the network is understood as the superposition of a large number of iid ON/OFF sources.
- See Taqqu, Willinger, Leland, Crovella,... (1993-1996), Heath, Resnick, Samorodnitsky (1998), M., Resnick, Rootzén, Stegeman (2002).



### 7.2.2. The infinite source Poisson model.

- Transmission initiations or connections of sources happen at the points of a rate $\lambda$ homogeneous Poisson process

$$\cdots < \Gamma_{-1} < \Gamma_0 < 0 < \Gamma_1 < \Gamma_2 < \cdots .$$

- Transmission durations are iid random variables $Y_i$, independent of $(\Gamma_i)$.
- During a transmission a source transmits at unit rate.
- The stationary process of active sources at time $t$

$$N_t = \sum_{i\in\mathbb{Z}} I\{\Gamma_i \le t < \Gamma_i + Y_i\}, \quad t \ge 0 .$$

- Since the points $(\Gamma_i, Y_i)$ constitute a PRM($\lambda$Leb $\times F_Y$), a simple calculation shows

$$\gamma(h) = \text{cov}(N_0, N_h) = \lambda \int_h^\infty \overline{F}_Y(t)\, dt\,.$$

- If $\overline{F}_Y(t) = P(Y > t) = L(t)t^{-\alpha}$, $\alpha > 1$, for some slowly varying $L$, by Karamata's theorem,

$$\gamma(h) \sim \lambda\,(\alpha-1)^{-1}\, h\, \overline{F}_Y(h) = \lambda\,(\alpha-1)^{-1}\, h^{1-\alpha}\, L(h)\,, \quad h \to \infty\,.$$

- Non-summability of $\gamma$ for $\alpha \in (1,2)$ is interpreted as long range dependence. The Hurst coefficient is $H = (3-\alpha)/2 \in (0.5, 1)$,

See Samorodnitsky and Taqqu (1994), Doukhan et al. (2003)



- The cumulative input process

$$A(t) = \int_0^t N_s\, ds\,, \quad t \geq 0\,,$$

has stationary increments.

7.3. Scaling limits for the cumulative input process.

- For $\alpha \in (1,2)$ scaling limits of $(A(Tt))_{t\geq 0}$ converge to spectrally positive $\alpha$-stable Lévy motion. (infinite variance, independent increments)
- Letting the intensity $\lambda = \lambda_T$ grow sufficiently fast, (fast growth)

$$\lim_{T\to\infty} \lambda_T\, T\, \overline{F}_Y(T) = \infty \quad \Leftrightarrow \quad \lim_{T\to\infty} \text{cov}(N_0, N_T) = \infty\,.$$

scaling limits of $(A(Tt))_{t\geq 0}$ converge to fractional Brownian motion $B_H$ with Hurst index $H = (3-\alpha)/2$ and covariance



structure

$$\text{cov}(B_H(t), B_H(s)) = 0.5(t^{2H} + s^{2H} - |t-s|^{2H})\,.$$

See M., Resnick, Rootzén, Stegeman (2002).

- Fractional Brownian motion $B_H$ with $H \in (0.5, 1)$ inherits long range dependence for the increment process

$$B_H(h) - B_H(h-1), \quad h = 1, 2, \ldots\,,$$

but loses the heavy tails.
- Similar results exist for superpositions of ON/OFF processes given the number $M = M_T$ of superimposed processes grows sufficiently fast with $T$.



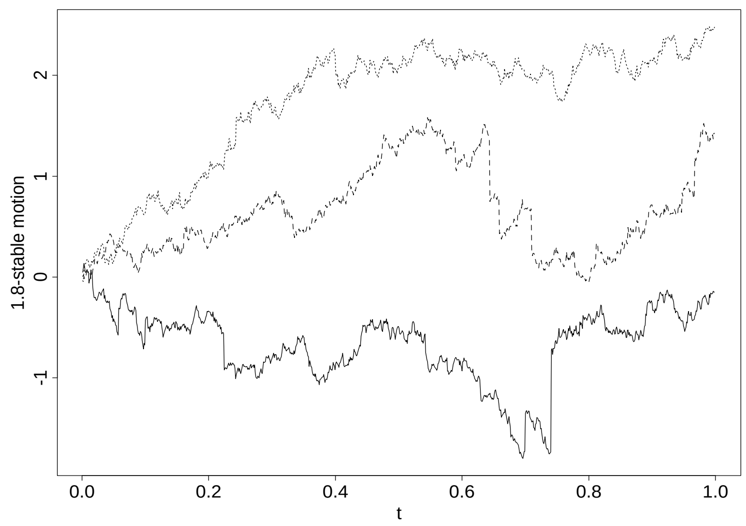


Figure 33. $1.8$-stable sample paths.

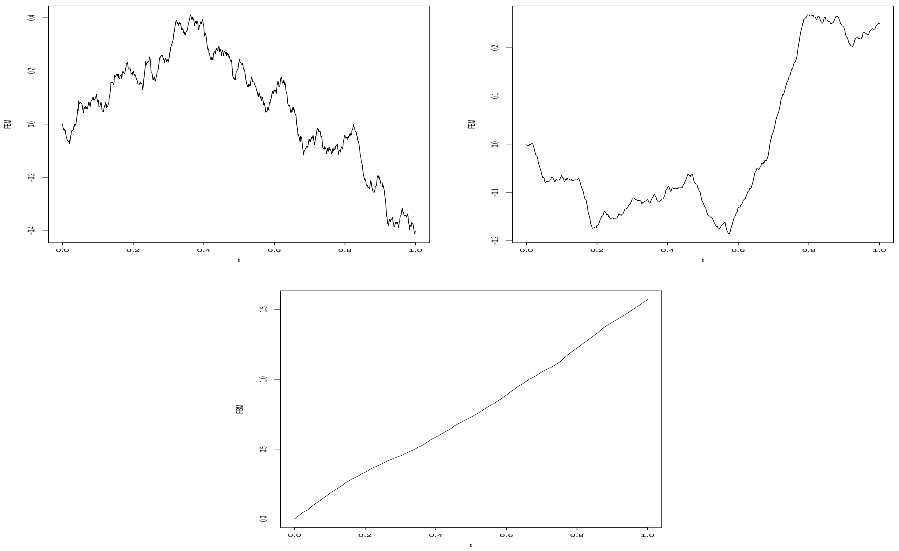

FIGURE 34. Simulation of fractional Brownian sample paths $B_H(t)$ on $[0,1]$. *Top left:* $H = 0.7$. *Top right:* $H = 0.9$ *Bottom:* $H = 0.99$.



- If $\lambda_T$ (or $M_T$) increase too slowly (slow growth):

$$\lim_{T\to\infty} \lambda_T T\, \overline{F}_Y(T) = 0 \quad \Leftrightarrow \quad \lim_{T\to\infty} \operatorname{cov}(N_0, N_T) = 0\,,$$

$\alpha$-stable Lévy motion appears in the limit.

- $\alpha$-stable Lévy motion inherits the heavy tails but loses the long range dependence of the activity processes.
- The case of intermediate growth has been treated in Gaigalas and Kaj (2003).
- In a series of papers, Levy, Pipiras, Taqqu (200...) have extended the scaling results for the ON-OFF model with random rewards. The random reward model goes back to Taqqu and Levy (1986)



## 7.4. The Poisson cluster process Faÿ, González-Arévalo, M., and Samorodnitsky (2006).

- At the points $\Gamma_i$ of a rate $\lambda$ homogeneous Poisson process on $\mathbb{R}$ the first packet of the $i$th flow ($i$th activity) arrives.
- The $i$th flow of packets consists of $K_i$ packets which arrive at times

$$Y_{ik} = \Gamma_i + S_{ik} = \Gamma_i + \sum_{j=1}^{k} X_{ij}\,, \quad 0 \le k \le K_i\,.$$

- $(X_{ik})_{i,k}$ are iid, $(K_i)$ are iid; $(X_{ik})$, $(K_i)$, $(\Gamma_i)$ are independent.
- The counting process

$$N(B) = \#\{(i,k) : i \in \mathbb{Z}\,, 0 \le k \le K_i : Y_{ik} \in B\}$$

is stationary.



- The points $(\Gamma_i, K_i, (X_{ik})_k)$ constitute a PRM($\lambda\,$Leb $\times F_K \times F_X^\infty$), $N^*$, in $\mathbb{R}\times\mathbb{N}_0\times\mathbb{R}^\infty$ and

$$N(a,b] = \int_{\mathbb{R}\times\mathbb{N}_0\times\mathbb{R}^\infty} \textstyle\sum_{j=0}^{k} I\{\gamma + \sum_{i=0}^{j} x_i \in (a,b]\}\, dN^*(\gamma, k, (x_i))\,.$$

- Let

$$\gamma_N(h) = \operatorname{cov}(N(0,1], N(h,h+1])\,.$$

**How can one get long range dependence for the increments $N(h,h+1]$?**

- If $\operatorname{var}(K) < \infty$

$$\int_1^\infty \gamma_N(h)\,dh \;=\; \lambda E\Big[\sum_{k=1}^{K}(K-k+1)\int_0^1 (x\wedge(2-x))\overline{F}_{S_k}(x)\,dx\Big] < \infty\,,$$

for the generic renewal process $S_k = X_1 + \cdots + X_k$.

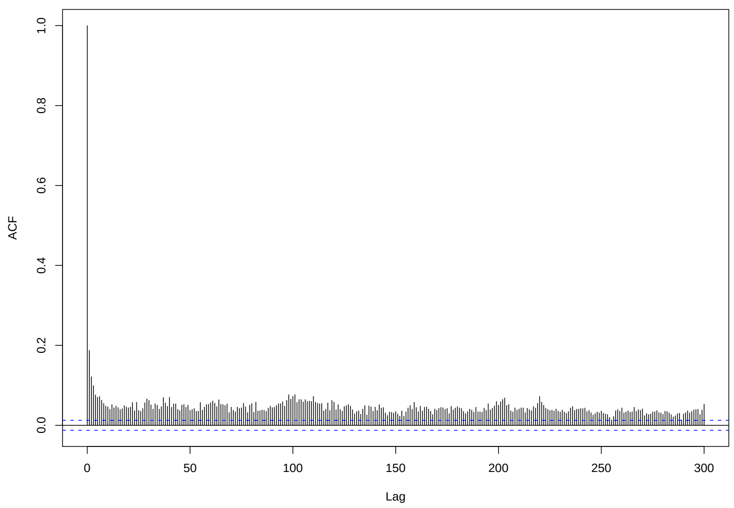


FIGURE 35. Sample ACF of the increments of the UNC packet arrival counting data.



- **Long range dependence is impossible unless var$(K) = \infty$** ***whatever the distribution of $X$.***
- **This is in agreement with technological restrictions: $S_K$ is large due to a large number $K$.**
- **A weighted renewal argument** Alsmeyer (1992) **yields**

$$\begin{aligned}\gamma_N(h) &= \lambda \sum_{k=1}^{\infty} E(K-k+1)_+ \int_{(0,1]} \left(\overline{F}_{S_k}(x+h-1) - \overline{F}_{S_k}(x+h)\right) dx \\ &= \lambda \int_h^{h+1} \left( \sum_{k=1}^{\infty} E(K-k+1)_+ \left(\overline{F}_{S_k}(y-1) - \overline{F}_{S_k}(y)\right) \right) dy\,. \\ &\sim \lambda\,(EX)^{\alpha-2}(\alpha-1)^{-1} h\, P(K>h)\,,\end{aligned}$$

**if $P(K>x) = x^{-\alpha}L(x)$ for some $\alpha \in (1,2)$.**

- **This corresponds to long range dependence.**



## Where do the heavy tails of $S_K$ come from?

- **$S_K$ can be large due to large $K$ or large $X$.**
- **$P(X>x) = x^{-\alpha}L(x)$, $EK < \infty$ and $P(K>x) = o(P(X>x))$. Then**

$$P(S_K > x) \sim EK\,P(X>x)\,.$$

- **$P(K>x) = x^{-\beta}L(x)$ for some $\beta \geq 0$, $EX < \infty$ and $P(X>x) = o(P(K>x))$. Then**

$$P(S_K > x) \sim (EX)^{\beta} P(K>x)\,.$$

- **The assumptions are close to necessity.**



## 7.5. A more general model M. and Samorodnitsky (2007).

- **A stationary marked point process on $\mathbb{R}$:**

$$(T_n, Z_n)\,, \quad n \in \mathbb{Z}\,,$$

**with $\cdots < T_{-1} \leq T_0 \leq 0 \leq T_1 \leq T_2 < \cdots$ arrival times of packets and $Z_n \geq 0$ amount of work brought to the system at time $T_n$.**

- **Active sources at time $t$**

$$N_t = \sum_{n \in \mathbb{Z}} I\{T_n \leq t < T_n + Z_n\}$$

- **Cumulative input process**

$$A(t) = \int_0^t N_s\,ds = \sum_{n\in\mathbb{Z}} \left[Z_n \wedge (t-T_n)_+ - Z_n \wedge (-T_n)_+\right]\,, \quad t \geq 0\,.$$

### 7.5.1. Growth of the variance.

- The variance of the cumulative input $A(t)$ is given by

$$\operatorname{var}(A(t)) = 2\int_0^t (t-x)\,g(x)\,dx\,,$$

where

$$g(x) = \int_{\mathbb{R}^2\times\mathbb{R}_+^2} I\{s_1 \le 0 < s_1+u_1, s_2 \le x < s_2+u_2\}\ \gamma_2(ds_1,ds_2,du_1,du_2)\,,$$

and $\gamma_2$ is the covariance measure of the stationary MPP.

- Assume $g$ is regularly varying with index $\beta \le 0$. If $\beta \in (-1,0]$ then

$$\operatorname{var}(A(t)) \sim \tfrac{2}{(1+\beta)(2+\beta)}\,t^2\,g(t)\,,\quad t\to\infty\,.$$

If $\beta < -1$ then $\operatorname{var}(A(t)) \sim C\,t$.



- The function $g$ describes the memory in the cumulative input process $A$.
- Examples. *ON-OFF*. $g(t) \sim c\,t\,\overline{F}_{\rm ON}(t)$ if $\overline{F}_{\rm ON} \in \mathrm{RV}(-\alpha)$, $\alpha \in (1,2)$. Then

$$\operatorname{var}(A(t)) \sim c_\alpha\, t^3 \overline{F}_{\rm ON}(t) \in \mathrm{RV}(3-\alpha)\,.$$

*Marks independent of the ground point process* $(T_n)$. This case can be difficult in general, but for $(T_n)$ homogeneous Poisson with rate $\lambda$,

$$g(t) \sim \lambda\,E(Z-t)_+ \sim \lambda\,(\alpha-1)^{-1} t\,\overline{F}_Z(t)$$

if $\overline{F}_Z \in \mathrm{RV}(-\alpha)$, $\alpha > 1$, and for $\alpha \in (1,2)$,

$$\operatorname{var}(A(t)) \sim c_\alpha\, t^3\,\overline{F}_Z(t) \in \mathrm{RV}(3-\alpha)\,.$$



### 7.5.2. Scaling limits for the superposition of cumulative input processes.

**Fast growth.**

- Consider the superposition of iid centered copies $A_i$ of the cumulative input $A$:

$$D_{n,T} = \sum_{i=1}^n (A_i(tT) - \mu\,t\,T)\,,\quad t \ge 0\,.$$

- Stationary increments of $A$ and the CLT for iid vectors yield

$$n^{-1/2} D_{n,T}(t) \stackrel{d}{\to} G(t\,T)\,,\quad n\to\infty\,,$$

where $G$ is centered Gaussian with stationary increments and $\operatorname{var}(G(t)) = \operatorname{var}(A(t))$.



- If $g \in \mathrm{RV}(\beta)$, $\beta \in (-1,0]$ (hence $\operatorname{var}(A(t)) \in \mathrm{RV}(2+\beta)$),

$$(\operatorname{var}(A(T)))^{-1/2} G(t\,T) \stackrel{d}{\to} B_H(t)\,,\quad T\to\infty\,,$$

where $H = 1+\beta/2$.

- If $g \in \mathrm{RV}(-\beta)$, $\beta < -1$ (hence $\operatorname{var}(A(t)) \sim c\,t$), $H = 0.5$ (Brownian motion).
- The results remain valid for $\lambda_n \to \infty$ satisfying a fast growth condition $[n\operatorname{var}(A(\lambda_n))]/\lambda_n^2 \to \infty$ (Lyapunov condition):

$$[n\operatorname{var}(A(\lambda_n))]^{-1/2} D_{n,\lambda_n}(t) = [n\operatorname{var}(A(\lambda_n))]^{-1/2} \sum_{i=1}^n (A_i(\lambda_n\,t) - \mu\,t\,\lambda_n) \stackrel{d}{\to} B_H(t)\,.$$

- This follows by Lyapunov's CLT.

**Slow growth.**

- Let $M(t) = \#\{i \in \mathbb{Z} : T_i \in [0,t]\}$ be the ground counting process.
- The cumulative input of a single source can be decomposed as follows if the stationary marks $Z_n$ have a finite first moment $E_0 Z$ under the Palm measure:

$$A(t) - \mu\, t \stackrel{d}{=} \sum_{i=1}^{M(t)} (Z_i - E_0 Z) + E_0 Z\, (M(t) - \lambda\, t) + O_P(1)$$
$$= \sum_{i=1}^{M(t)} (Z_i - \lambda\, E_0 Z\, (T_m - T_{m-1})) + O_P(1)\,.$$



- A variety of limit processes is possible for the workload of a single source with a regularly varying scaling function $a \in \mathrm{RV}(\alpha)$ for some $\alpha > 0$,

$$(a(T))^{-1}\,(A(tT) - \mu\, t\, T) \stackrel{d}{\rightarrow} V(t)\,.$$

- $\alpha$-stable Lévy motion $V$, $\alpha \in (1,2]$, is one out of many possible limits.
- The scaling limit is inherited by the superposition of iid workload processes under slow growth conditions on $\lambda_n$:

$$(a(\lambda_n))^{-1}\, D_{n,\lambda_n}(t) = \sum_{i=1}^{n} (A_i(\lambda_n\, t) - \mu\, \lambda_n\, t) \stackrel{d}{\rightarrow} V(t)\,.$$



- Fractional Brownian motion is a possible limit process as well, e.g. when $(T_n)$ comes from a Poisson cluster process: at each (homogeneous) Poisson point an independent cluster point process of type

$$N_c[0,t] = N_0[0,t] \wedge (K+1)\,, \quad t \geq 0\,,$$

starts, where $\overline{F}_K \in \mathrm{RV}(-\alpha)$, $\alpha \in (1,2)$, independent of the process $N_0$ with increasing arrivals $T_n^{(0)} \geq 0$. If the condition $E(T_n^{(0)}) \leq c\, n^{\alpha}$ fails, fractional Brownian motion may occur as a limit.



## 7.6. Conclusions.

- The general stationary MPP includes the classical models (ON-OFF, infinite source Poisson) and allows for flexible modeling of dependence of the inter-arrival times $T_n - T_{n-1}$, the clustering behavior due to the arrival of an impulse generating a flow of activities, as well as dependence between the arrival sequence $(T_n)$ and the mark sequence $(Z_n)$.
- The memory in the workload depends on a variety of factors such as the tails of the inter-arrival times or the tails of the activities initiated at an arrival $T_n$ or the number of activities started at $T_n$.

- Fractional Brownian motion and $\alpha$-stable Lévy motion are typical scaling limits of one source as well as of the superposition of an increasing number of iid workload processes, but many other limits may occur as well.
- However, fractional Brownian motion seems to be a "more robust" limit than others.



## References


[1] Alsmeyer, G. (1992) On generalized renewal measures and certain first passage times. *Ann. Probab.* **20**, 1229–1247.
[2] Arov, D.Z. and Bobrov, A.A. (1960) The extreme members of a sample and their role in the sum of independent variables. (Russian) *Teor. Verojatnost. i Primenen.* **5**, 415–435.
[3] Barndorff-Nielsen, O.E. and Shephard, N. (2001) Modelling by Lévy processes for financial econometrics. In: Barndorff-Nielsen, O.E., Mikosch, T. and Resnick, S.I. (Eds.) *Lévy Processes: Theory and Applications*, pp. 283–318. Birkhäuser, Boston.
[4] Basrak, B., Davis, R.A. and Mikosch, T. (2002a) A characterization of multivariate regular variation. *Ann. Appl. Probab.* **12**, 908–920.
[5] Basrak, B., Davis, R.A. and Mikosch. T. (2002b) Regular variation of GARCH processes. *Stoch. Proc. Appl.* **99**, 95–116 .
[6] Berkes, I., Horváth, L. and Kokoszka, P. (2003) GARCH processes: structure and estimation. *Bernoulli* **9**, 201–228.
[7] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987) *Regular Variation.* Cambridge University Press, Cambridge.
[8] Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31**, 307–327.
[9] Boman, and Lindskog, F. (2007) Support theorems for the Radon transfomr and Cramér-Wold theorems. Technical report.
[10] Bougerol, P. and Picard, N. (1992a) Strict stationarity of generalized autoregressive processes. *Ann. Probab.* **20**, 1714–1730.
[11] Bougerol, P. and Picard, N. (1992b) Stationarity of GARCH processes and of some nonnegative time series. *J. Econometrics* **52**, 115–127.
[12] Breidt, F.J. and Carriquiry, A.L. (1996) Improved quasi-maximum likelihood estimation for stochastic volatility models. In: *Modelling and Prediction. Proceedings Hsinchu, 1994*, pp. 228–247. Springer, New York.
[13] Breidt, F.J. and Davis, R.A. (1998) Extremes of stochastic volatility models. *Ann. Appl. Probab* **8**, 664–675.
[14] Breiman, L. (1965) On some limit theorems similar to the arc-sin law. *Theory Probab. Appl.* **10**, 323–331.
[15] Brockwell, P.J. and Davis, R.A. (1991) *Time Series: Theory and Methods*, 2nd edition. Springer, New York.
[16] Brockwell, P.J. and Davis, R.A. (1996) *Introduction to Time Series and Forecasting.* Springer, New York.
[17] Crovella, M., Bestavros, A. and Taqqu, M.S. (1996) Heavy-tailed probability distributions in the world wide web. Preprint, 1996.
[18] Davis, R.A. and Hsing, T. (1995) Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Ann. Probab.* **23**, 879–917.
[19] Davis, R.A. and Mikosch, T. (2001a) Point process convergence of stochastic volatility processes with application to sample autocorrelations. *J. Appl. Probab.* Special Volume **38A**, A Festschrift for David Vere-Jones, 93–104.
[20] Davis, R.A. and Mikosch, T. (2001b) The sample autocorrelations of financial time series models. In: Fitzgerald, W.J., Smith, R.L., Walden, A.T. and Young, P.C. (Eds.) *Nonlinear and Nonstationary Signal Processing*, pp. 247–274. Cambridge University Press, Cambridge (U.K.).
[21] Davis, R.A. and Mikosch, T. (2007a) Fundamental properties of stochastic volatility models. In: Andersen, T.G., Davis, R.A., Kreiss, J.-P. and Mikosch, T. (2007) *Handbook of Financial Time Series.* Springer, Berlin. Forthcoming.
[22] Davis, R.A. and Mikosch, T. (2007b) Extremes of stochastic volatility models. In: Andersen, T.G., Davis, R.A., Kreiss, J.-P. and Mikosch, T. (2007) *Handbook of Financial Time Series.* Springer, Berlin. Forthcoming.
[23] Doukhan, P., Oppenheim, G. and Taqqu, M.S. (Eds.) (2003) *Long Range Dependence.* Birkhäuser, Boston.





[24] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin.
[25] Engle, R.F. (1982) Autoregressive conditional heteroscedastic models with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987–1007.
[26] Engle, R.F. (Ed.) (1995) *ARCH Selected Readings.* Oxford University Press, Oxford (UK).
[27] Engle, R.F. and Bollerslev, T. (1986) Modelling the persistence of conditional variances. With comments and a reply by the authors. *Econometric Rev.* **5**, 1–87
[28] Faÿ, G., Gonzalez-Arevalo, B., Mikosch, T. and Samorodnitsky, G. (2006) Modeling teletraffic arrivals by a Poisson cluster process. *Queuing Systems Appl.* **54**, 121–140.
[29] Feller, W. (1971) *An Introduction to Probability Theory and Its Applications.* Vol. II. Second edition. Wiley, New York.
[30] Gaigalas, R. and Kaj, I. (2003) Convergence of scaled renewal processes and a packet arrival model. *Bernoulli* 9:671–703.
[31] Galambos, J. (1978) *The Asymptotic Theory of Order Statistics.* Wiley, New York.
[32] Haan, L. de, Fereira, A. (2006) *Extreme Value Theory: An Introduction.* Springer, Berlin, New York.
[33] Heath, D., Resnick, S. and Samorodnitsky, G. (1998) Heavy tails and long range dependence in ON/OFF processes and associated fluid models. *Math. Oper. Res.* **23**, 145–165.
[34] Hult, H. and Lindskog, F. (2005) Extremal behavior of regularly varying stochastic processes. *Stoch. Proc. Appl.* **115**, 249–274.
[35] Hult, H. and Lindskog, F. (2006a) On regular variation for infinitely divisible random vectors and additive processes. *Adv. in Appl. Probab.* **38**, 134–148
[36] Hult, H. and Lindskog, F. (2006b) On Kesten's counterexample to the Cramér-Wold device for regular variation. *Bernoulli* **12**, 133–142.
[37] Ibragimov, I.A. and Linnik, Yu.V. (1971) *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff, Groningen.
[38] Jessen, A.H., Mikosch, T. (2006) Regularly varying functions. Available under www.math.ku.dk/∼mikosch
[39] Kesten, H. (1973) Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131**, 207–248.
[40] Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983) *Extremes and Related Properties of Random Sequences and Processes.* Springer, Berlin.
[41] Leland, W.E., Taqqu, M.S., Willinger, W. and Wilson, D.V. (1993) On the self-similar nature of Ethernet traffic. *ACM/SIGCOMM Computer Communications Review* 183–193.
[42] Levy, J.B. and Taqqu, M.S. (2001) Dependence structure of a renewal-reward process with infinite variance. *Fractals* **9**, 185–192.
[43] Maller, R.A. and Resnick, S.I. (1984) Limiting behaviour of sums and the term of maximum modulus. *Proc. London Math. Soc.* **49** , 385–422.
[44] Mandelbrot, B. (1963) The variation of certain speculative prices. *J. Busin. Univ. Chicago* **36**, 394–419.
[45] Mandelbrot, B. (1997) *Fractals and Scaling in Finance.* Discontinuity, concentration, risk. Selecta Volume E. Selected works by Benoit B. Mandelbrot. Springer, New York.
[46] Mandelbrot, B. (2001) Scaling in financial prices I. Tails and dependence. *Quantitative Finance* **1**, 113-123.
[47] McNeil, A., Frey, R. and Embrechts, P. (2005) *Quantitative Risk Management.* Princeton University Press.
[48] Mikosch, T. (2003) Modelling dependence and tails of financial time series. In: B. Finkenstädt and H. Rootzén (Eds.) *Extreme Values in Finance, Telecommunications and the Environment.* Chapman and Hall, pp. 185–286.





[49] Mikosch T. (2006) Copulas: tales and facts, with a discussion and a rejoinder. *Extremes* **9**, 3-20.
[50] Mikosch, T., Resnick, S., Rootzén, H. and Stegeman, A. (2002) Is network traffic approximated by stable Lévy motion or fractional Brownian motion? *Ann. Appl. Probab.* **12**, 23–68.
[51] Mikosch, T. and Samorodnitsky, G. (2000) The supremum of a negative drift random walk with dependent heavy-tailed steps. *Ann. Appl. Probab.* **10**, 1025–1064.
[52] Mikosch, T. and Samorodnitsky, G. (2007) Scaling limits for cumulative input processes. *Math. Oper. Res.*, to appear.
[53] Nelson, D.B. (1990) Stationarity and persistence in the GARCH$(\mathbf{1},\mathbf{1})$ model. *Econometric Theory* **6**, 318–334.
[54] O'Brien, G. L. (1980) A limit theorem for sample maxima and heavy branches in Galton-Watson trees. *J. Appl. Probab.* **17**, 539–545.
[55] Petrov, V.V. (1975) *Sums of Independent Random Variables.* Springer, Berlin.
[56] Petrov, V.V. (1995) *Limit Theorems of Probability Theory.* Clarendon Press, Oxford.
[57] Pipiras, V., Taqqu, M.S. and Levy, J.B. (2004) Slow, fast and arbitrary growth conditions for renewal-reward processes when both the renewals and the rewards are heavy-tailed. *Bernoulli* **10**, 121–163.
[58] Resnick, S.I. (1986) Point processes, regular variation and weak convergence. *Adv. Appl. Prob.* **18**, 66–138.
[59] Resnick, S.I. (1987) *Extreme Values, Regular Variation, and Point Processes.* Springer, New York.
[60] Resnick, S.I. (2006) *Heavy Tail Phenomena: Probabilistic and Statistical Modeling.* Springer, New York. To appear.
[61] Rvačeva, E. L. (1962) On domains of attraction of multi-dimensional distributions. *Select. Transl. Math. Statist. and Probability* American Mathematical Society, Providence, R.I. **2**, 183–205.
[62] Samorodnitsky, G. and Taqqu, M.S. (1994) *Stable Non–Gaussian Random Processes. Stochastic Models with Infinite Variance.* Chapman and Hall, London.
[63] Shephard, N. (1996) Statistical aspects of ARCH and stochastic volatility. In: Cox, D.R., Hinkley, D.V. and Barndorff-Nielsen, O.E. (Eds.) *Likelihood, Time Series with Econometric and Other Applications.* Chapman and Hall, London.
[64] Straumann, D. (2005) *Estimation in Conditionally Heteroscedastic Time Series Models.* Lecture Notes in Statistic **181** Springer, Berlin.
[65] Straumann, D. and Mikosch, T. (2006) Quasi-MLE in heteroscedastic time series: a stochastic recurrence equations approach. *Ann. Statist.*, to appear.
[66] Taqqu, M.S. (2002) The modelling of Ethernet data and of signals that are heavy-tailed with infinite variance. *Scand. J. Statist.* **29**, 273–295.
[67] Taylor, S.J. (1986) *Modelling Financial Time Series.* Wiley, Chichester.
[68] Taqqu, M.S. and Levy, J.B. (1986) Using renewal processes to generate long-range dependence and high variability. In: *Dependence in Probability and Statistics (Oberwolfach, 1985)*, pp. 73–89. Birkhäuser, Boston.
[69] Willinger, W., Taqqu, M.S., Sherman, R. and Wilson, D. (1995) Self-similarity through high variability: statistical analysis of ethernet lan traffic at the source level. Proceedings of the ACM/SIGCOMM'95, Cambridge, MA. *Computer Communications Review* **25**, 100–113.