Power law tails in applied probability

Some recent developments[1]

Thomas Mikosch

University of Copenhagen

[1]Prague, September 4, 2015

1. Modeling power law (heavy) tails via regular variation

- **Regularly varying function:** $f(x) = x^{\rho} L(x)$, $x > 0$, $\rho \in \mathbb{R}$, L slowly varying. Bingham, Goldie, Teugels (1987)

- **Regularly varying random variable:** For some $\alpha > 0$,

$$\mathbb{P}(X > x) \sim p \frac{L(x)}{x^{\alpha}} \quad \text{and} \quad \mathbb{P}(X \leq -x) \sim q \frac{L(x)}{x^{\alpha}}. \tag{1.1}$$

- (1.1) appears as domain of attraction condition in limit theory for sums, maxima,... for sequences of iid random variables (X_t).

- **Example.** (X_t) iid **Pareto:**

$$\mathbb{P}(X_t > x) = x^{-\alpha}\,, \qquad x > 1, \quad \alpha > 0\,,$$

and

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

Then

$$\mathbb{P}(M_n/n^{1/\alpha} \le x) = \Big(1 - \frac{x^{-\alpha}}{n}\Big)^n \to \mathrm{e}^{-x^{-\alpha}}\,, \qquad x > 0\,.$$

- The limit is the **Fréchet distribution.**
- It is one of the three **max-stable distributions**; they are the only non-degenerate limit distributions for maxima of an iid sequence.

- **Example.** The **Cauchy distribution** has density and characteristic function, respectively,

$$f_X(x) = \frac{1}{\pi(1+|x|^2)} \qquad \text{and} \qquad \varphi_X(t) = \mathrm{e}^{-|t|}\,.$$

- A Cauchy random variable X is **regularly varying with index 1.**
- IID copies (X_t) of X satisfy

$$n^{-1}S_n = n^{-1}(X_1 + \cdots + X_n) \stackrel{d}{=} X_1\,, \qquad n \geq 1\,.$$

- The Cauchy distribution is one of the infinite variance **sum-stable limit distributions for iid random variables.**

- The regular variation condition

$$\mathbb{P}(X > x) \sim p\,\frac{L(x)}{x^\alpha} \quad \text{and} \quad \mathbb{P}(X \le -x) \sim q\,\frac{L(x)}{x^\alpha}\,.$$

is used as modeling assumption in applied probability,

– in insurance mathematics,

– queuing (e.g. data networks),

– (financial) time series analysis,

– climate and weather research,

– seismology, . . .

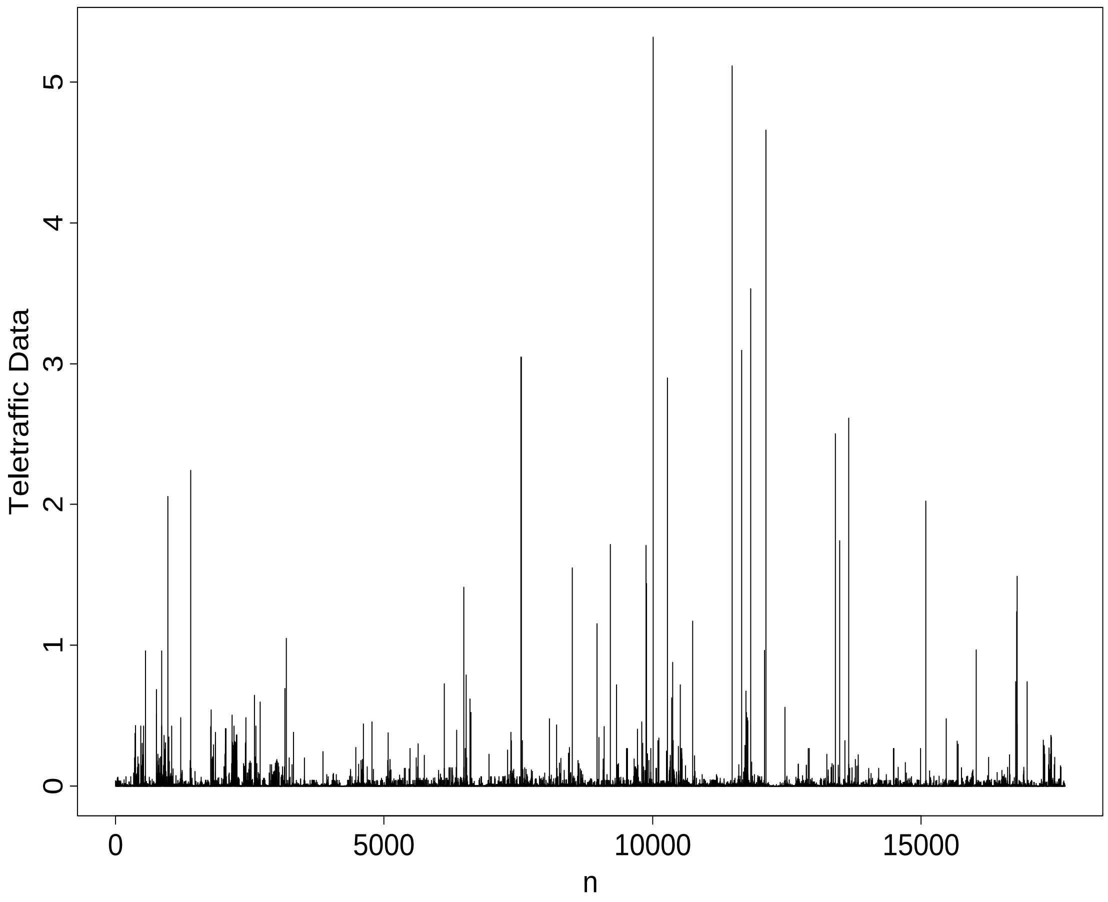

Figure 1. Time series of transmission durations of Ethernet files.

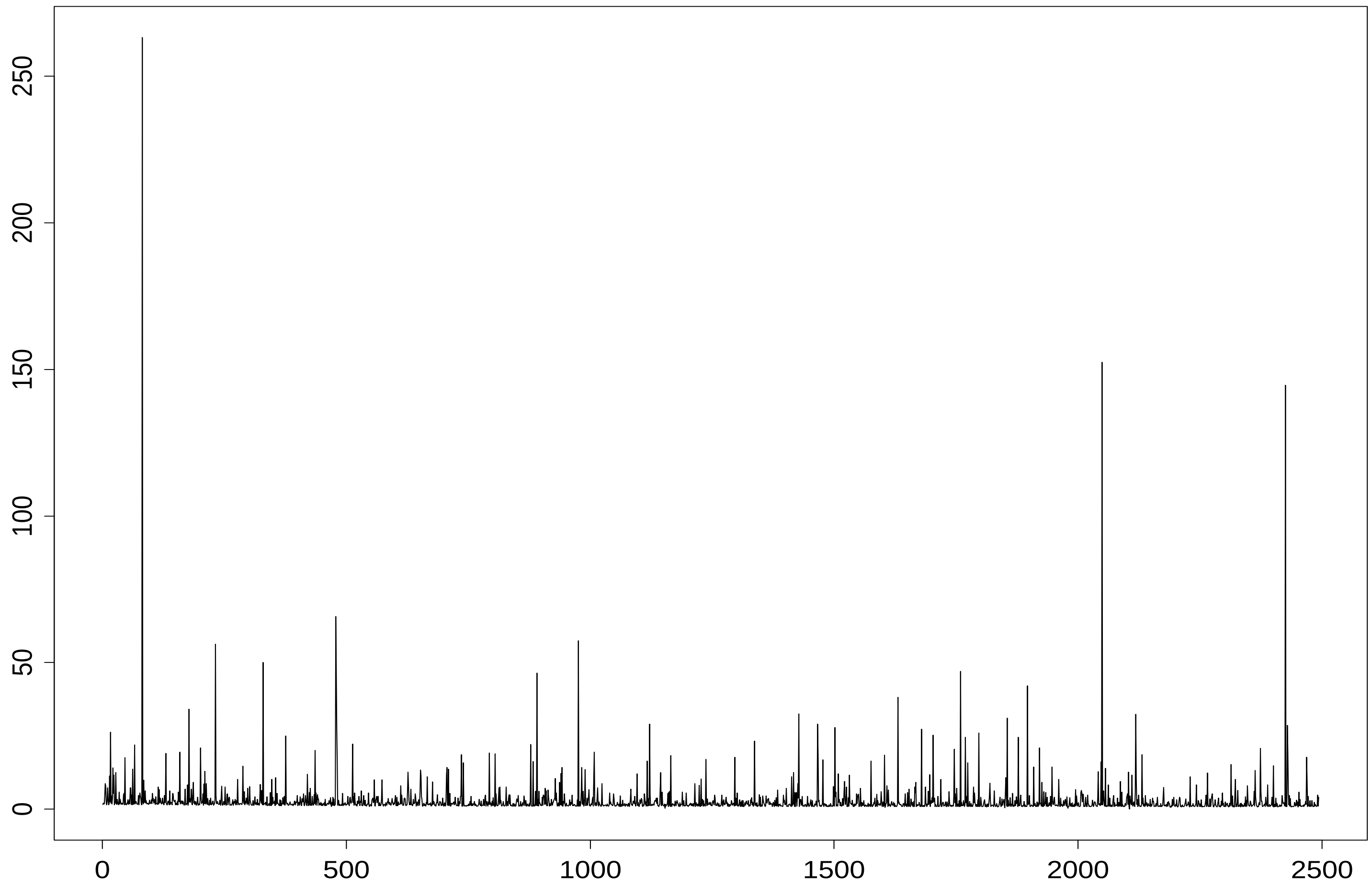

FIGURE 2. Danish fire insurance data.

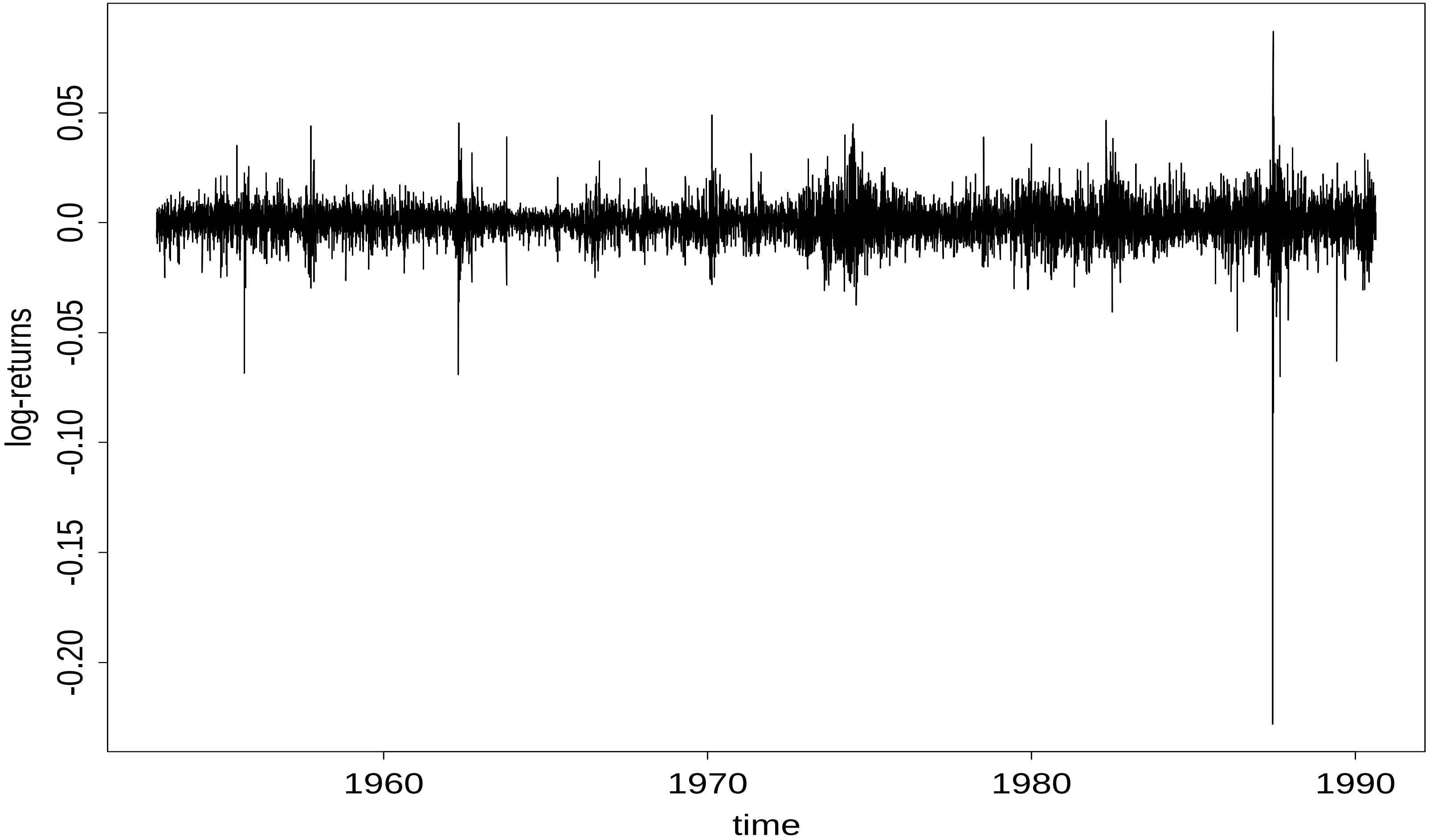

FIGURE 3. Plot of **9558** *S&P500* daily log-returns from January *2, 1953,* to December *31, 1990.* The year marks indicate the beginning of the calendar year.

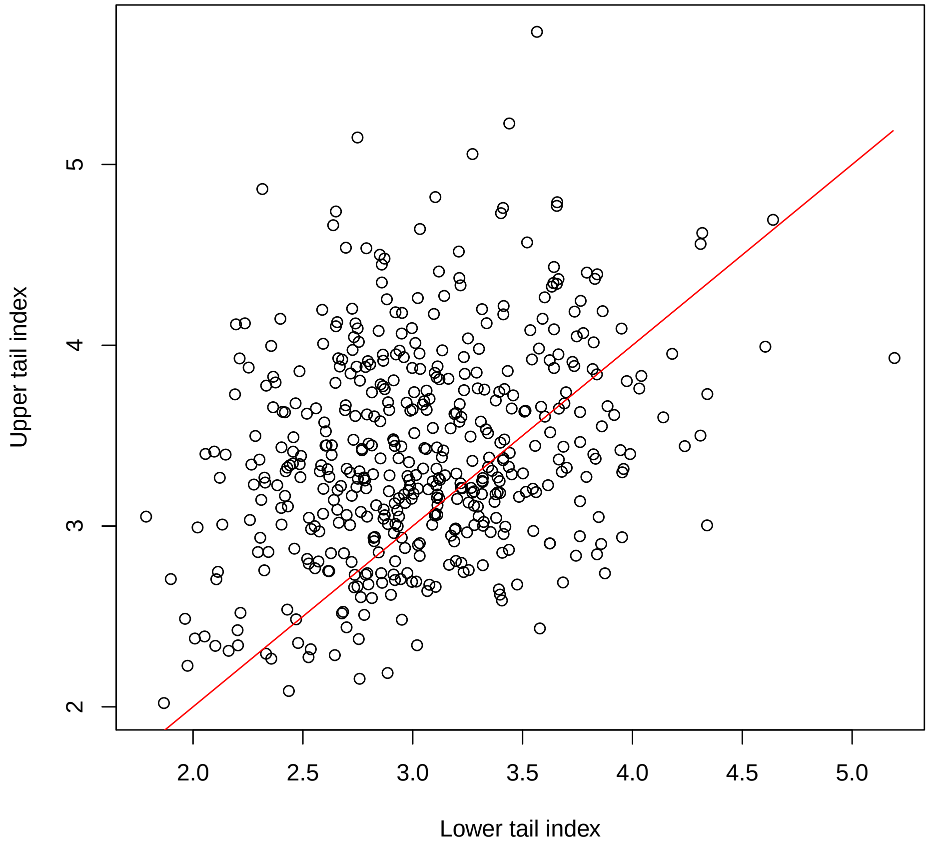

FIGURE 4. Hill estimates of the upper and lower tail indices of log-returns for 420 univariate time series from S&P 500.

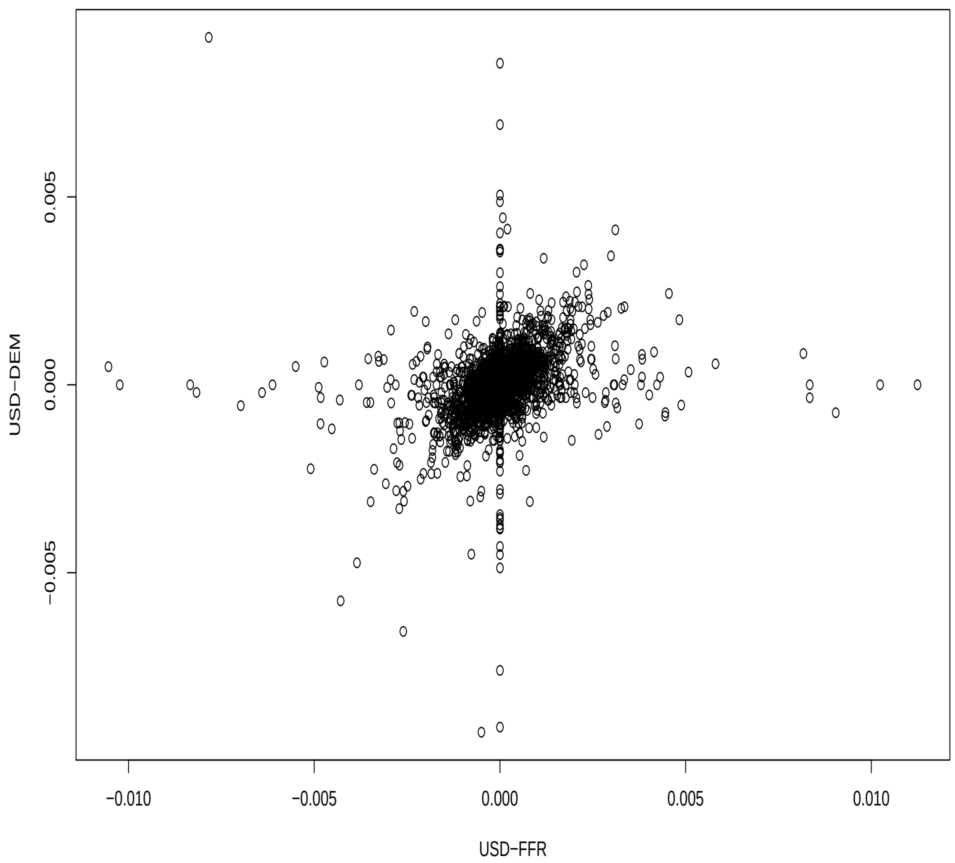

Figure 5. Scatterplot of 5 minute foreign exchange rate log-returns, USD-DEM against USD-FRF.

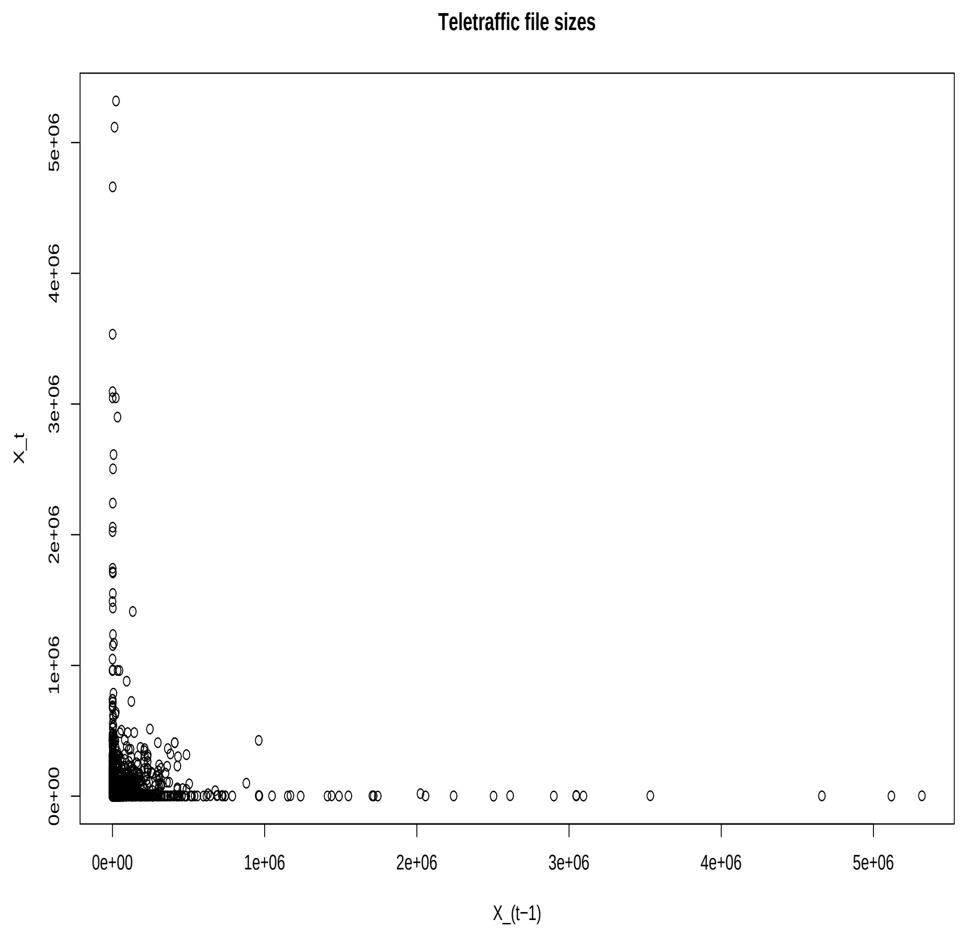

FIGURE 6. Scatterplot of file sizes of teletraffic data.

2. Multivariate regular variation

- An $\mathbb{R}^d$-valued random vector $\mathbf{X} = (X_1, \ldots, X_d)$ is regularly varying if there exists a non-null Radon measure ν_d on $\mathbb{R}_0^d = \mathbb{R}^d \backslash \{\mathbf{0}\}$[2] such that Resnick (1987,2007)

$$\mu_x(\cdot) = \frac{\mathbb{P}(x^{-1}\mathbf{X} \in \cdot)}{\mathbb{P}(|\mathbf{X}| > x)} \xrightarrow{v} \nu_d(\cdot)\,, \qquad x \to \infty\,.$$

- The measure ν_d determines the extremal dependence structure of the vector $\mathbf{X}$. It has the scaling property

$$\nu_d(tA) = t^{-\alpha}\,\nu_d(A), \quad t > 0, \qquad \text{for some } \alpha \geq 0.$$

- α is the index of regular variation or tail index.

[2] The measure is finite on sets bounded away from zero.

- $\mathbf{X}$ is regularly varying with index α if and only if there exist a vector $\boldsymbol{\Theta}_0$ with values in the unit sphere of $\mathbb{R}^d$ and an independent Pareto distributed random variable $\boldsymbol{Y}_0$, i.e., $\mathbb{P}(\boldsymbol{Y}_0 > t) = t^{-\alpha}$, $t > 1$, such that

$$\Big(\frac{\mathbf{X}}{|\mathbf{X}|}, |\mathbf{X}|\Big) \;\Big|\; |\mathbf{X}| > x \xrightarrow{d} (\boldsymbol{\Theta}_0\,, \boldsymbol{Y}_0)\,, \qquad x \to \infty\,.$$

- The radial and the spherical parts of $\mathbf{X}$ are asymptotically independent for large values of $|\mathbf{X}|$.
- This property allows one to estimate probabilities of extreme events for which no or only a few data are available.

- Regular variation is preserved under homogeneous continuous mappings $f : \mathbb{R}^k \to \mathbb{R}^d$.
- For example,

$$\frac{\mathbb{P}(x^{-1}(X_1 + \cdots + X_k) \in \cdot)}{\mathbb{P}(|\mathbf{X}| > x)} \xrightarrow{v} \nu_k(\{\mathbf{x} \in \mathbb{R}^k : x_1 + \cdots + x_k \in \cdot\}) \,.$$

  Sums of regularly varying random variables are regularly varying (possibly degenerate).
- Regular variation can be extended in a straightforward way to abstract spaces.

3. Regularly varying stationary sequences

- **An $\mathbb{R}^d$-valued strictly stationary sequence $(\mathbf{X}_t)$ is regularly varying with index $\alpha > 0$ if its finite-dimensional distributions are regularly varying with index α.** Davis, Hsing (1995)

- **This means: for every $k \geq 1$, there exists a non-null Radon measure μ_k on $(\mathbb{R}_0^d)^k$ such that**

$$\frac{\mathbb{P}(x^{-1}(\mathbf{X}_1, \ldots, \mathbf{X}_k) \in \cdot)}{\mathbb{P}(|\mathbf{X}_0| > x)} \xrightarrow{v} \mu_k(\cdot)\,.$$

- **Alternatively,** Basrak, Segers (2009) **for $\alpha > 0$, $k \geq 0$, there exists a sequence $(Y_t)_{t\geq 0}$ such that**

$$\mathbb{P}(x^{-1}(X_0,\ldots,X_k) \in \cdot \mid |X_0| > x) \xrightarrow{w} \mathbb{P}((Y_0,\ldots,Y_k) \in \cdot)\,,$$

- **$|Y_0|$ is independent of $(Y_0,\ldots,Y_k)/|Y_0|$**
- **and $\mathbb{P}(|Y_0| > y) = y^{-\alpha}$, $y > 1$.**

- **Example. Assume (X_t) iid. Then $Y_t = 0$ for $t \geq 1$.**
- **This means: If $|X_0|$ is large, $|X_1|,\ldots,|X_k|$ cannot be large.**

  Extremes appear close to the axes.

- **Example.** Consider the stochastic recurrence equation

$$X_t = A_t X_{t-1} + B_t \qquad \text{for iid } (A_t, B_t),\ A_t, B_t > 0. \tag{3.2}$$

- If there exists $\alpha > 0$ such that $\mathbb{E}[A_1^\alpha] = 1$ and $\mathbb{E}[B_1^\alpha] < \infty$ then

$$\mathbb{P}(X_1 > x) \sim c\, x^{-\alpha} \qquad x \to \infty\,,$$

and (X_t) is regularly varying with index α. Kesten (1973), Goldie (1991)

- Then

$$x^{-1}(X_0, \ldots, X_k) \Big| X_0 > x \xrightarrow{d} Y_0\, (1, A_1, A_1\, A_2\,, \ldots, A_1 \cdots A_k)$$

- For example, the extremogram of (X_t): as $x \to \infty$,

$$\mathbb{P}(X_k > x \mid X_0 > x) \to \mathbb{P}(Y_0\, A_1 \cdots A_k > 1) = \mathbb{E}[1 \wedge (A_1 \cdots A_k)^\alpha]\,.$$

- **Example.** Consider a GARCH(1, 1) process Bollerslev (1986)

$$X_t = \sigma_t Z_t \,, \qquad (Z_t) \text{ iid mean-zero, unit variance}\,,$$

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

- The GARCH(1, 1) is a major model for returns (relative changes in a given time unit) of speculative prices.
- The process (σ_t^2) satisfies (3.2) with $A_t = \alpha_1 Z_{t-1}^2 + \beta_1$ and $B_t = \alpha_0$.
- The volatility process (σ_t) is regularly varying and the process (X_t) inherits this property.

- **This is by virtue of**

Breiman's lemma:

for non-negative independent random variables ξ and η with $\mathbb{P}(\xi > x) \sim c\, x^{-\alpha} L(x)$ and $\mathbb{E}[\eta^{\alpha+\delta}] < \infty$,

$$\mathbb{P}(\xi\,\eta > x) \sim \mathbb{E}[\eta^{\alpha}]\,\mathbb{P}(\xi > x)\,, \qquad x \to \infty\,.$$

- **For the GARCH$(1,1)$,**

$$\begin{aligned}\mathbb{P}(X_t > x) &= \mathbb{P}(\sigma_t\, Z_t > x)\\ &\sim \mathbb{E}[(Z_t^+)^{\alpha}]\,\mathbb{P}(\sigma_t > x)\\ &\sim \mathbb{E}[(Z_t^+)^{\alpha}]\,c\,x^{-\alpha}\,.\end{aligned}$$

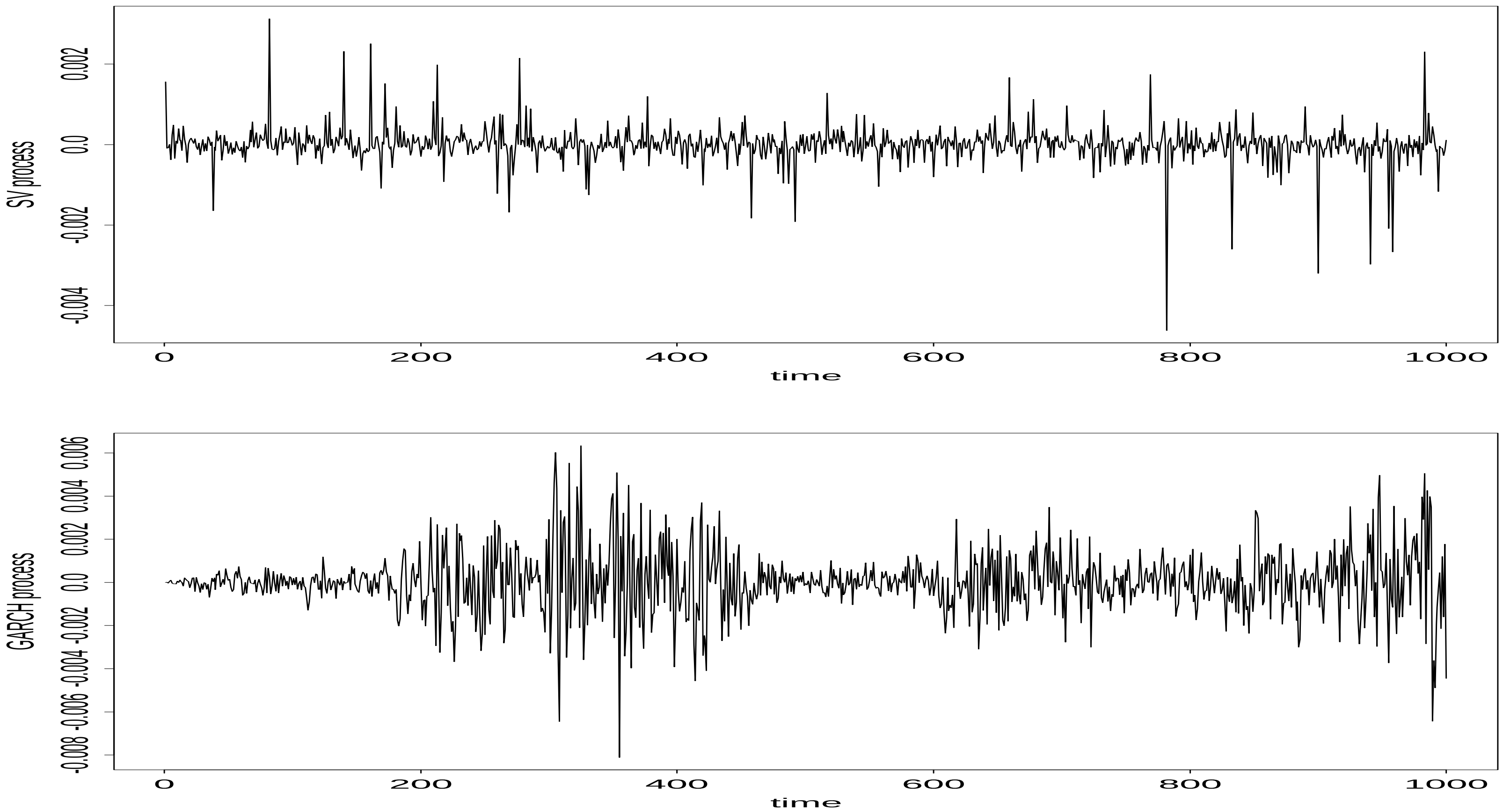

FIGURE 7. *Top:* Stochastic volatility process $X_t = \sigma_t Z_t$ for iid student (Z_t) with 4 degrees of freedom, Gaussian ARMA(1,1) process $\log \sigma_t = 0.5 \log \sigma_{t-1} + 0.3\eta_{t-1} + \eta_t$. *Bottom:* GARCH$(1,1)$ process $X_t = (0.0001 + 0.1X_{t-1}^2 + 0.9\sigma_{t-1}^2)^{0.5} Z_t$ for iid standard normal (Z_t).

Examples of regularly varying stationary sequences.

- **IID sequence (Z_t) with regularly varying Z_0.**
- **Linear processes e.g. ARMA processes with iid regularly varying noise (Z_t).** Rootzén (1978,1983), Davis, Resnick (1985)
- **Solutions to stochastic recurrence equation: $X_t = A_t X_{t-1} + B_t$** Kesten (1973), Goldie (1991)
- **GARCH process.** Bollerslev (1986), M., Stărică (2000), Davis, M. (1998), Basrak, Davis, M. (2000,2002)
- **The simple stochastic volatility model with iid regularly varying noise.** Davis, M. (2001)

- **Infinite variance α-stable stationary processes are regularly varying with index $\alpha \in (0, 2)$.** Samorodnitsky, Taqqu (1994), Rosiński (1995,2000)
- **Max-stable stationary processes with Fréchet (Φ_α) marginals are regularly varying with index $\alpha > 0$.** de Haan (1984), Stoev (2008), Kabluchko (2009)

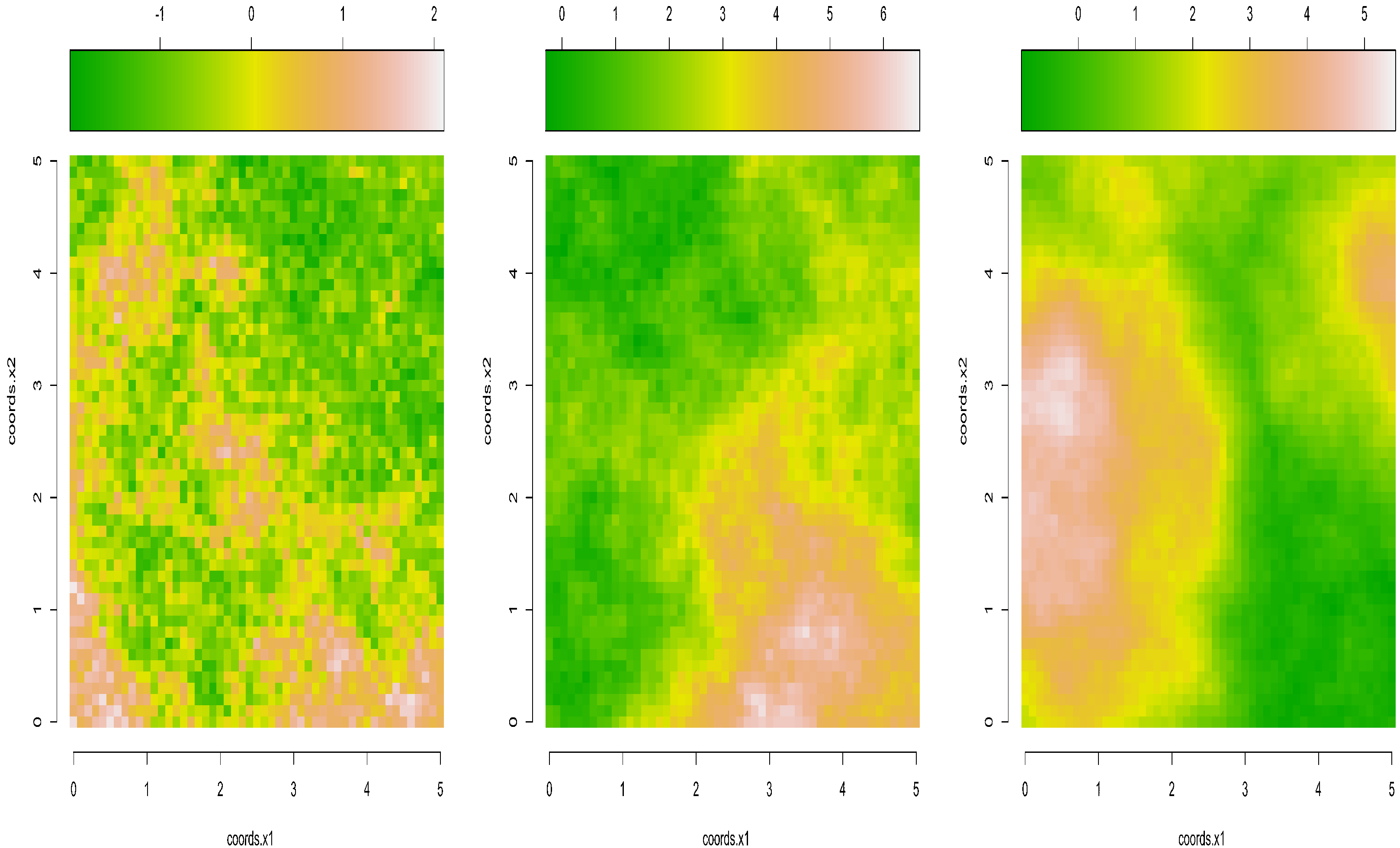

FIGURE 8. Sample of a Brown-Resnick random field on $[0,5]^2$ with variogram $\gamma(t) = |t|^\alpha/2$ for $\alpha = 1/2$, $\alpha = 1$, $\alpha = 3/2$ from left to right, respectively. The grid mesh is 0.1.

4. Limit theory for regularly varying sequences

- **Asymptotic theory for sums**

$$S_n = X_1 + \cdots + X_n,$$

sample covariances, sample autocorrelations with α-stable limit for $\alpha < 2$, or $\alpha/2$-limit for $\alpha < 4$, periodogram,. . . : Rootzén (1983), Davis, Resnick (1985,1986), Jakubowski (1993,1997), Davis, Hsing (1995), M., Küppelberg (1992-1995), Buraczewski, Damek, M., Jakubowski, Wintenberger, Basrak, Segers (2005-2015), in particular Markov chains

- Cluster Poisson limits for point processes

$$N_n = \sum_{t=1}^{n} \varepsilon_{X_t/a_n} \xrightarrow{d} N = \sum_{k=1}^{\infty} \sum_{i=1}^{\infty} \varepsilon_{\Gamma_i^{-1/\alpha} Q_{ik}}$$

where $\mathbb{P}(|X_t| > a_n) \sim n^{-1}$, $\Gamma_i = E_1 + \cdots + E_i$, (E_i) iid exponential, $(Q_{ik})_{k\geq 1}$ iid, $\sup_k |Q_{ik}| \leq 1$ a.s.

- Limit theory for maxima and order statistics and continuous functionals acting on them, extremal index. Davis, Resnick (1985,1986), Davis, Hsing (1995), Basrak, Davis, M. (2000-2002), Basrak, Segers (2009), M., Wintenberger (2013-2015)

- **Extremogram.** **An autocorrelaton function for serial extremal dependence.** Davis, M., Zhao (2009-)
- **For an $\mathbb{R}^d$-valued strictly stationary regularly varying sequence $(\mathbf{X}_t)$ and a Borel set A bounded away from zero the extremogram is the limiting function**

$$
\begin{aligned}
\rho_A(h) &= \lim_{x\to\infty} \mathbb{P}(x^{-1}\mathbf{X}_h \in A \mid x^{-1}\mathbf{X}_0 \in A) \\
&= \lim_{x\to\infty} \frac{\mathbb{P}(x^{-1}\mathbf{X}_0 \in A\,, \quad x^{-1}\mathbf{X}_h \in A)}{\mathbb{P}(x^{-1}\mathbf{X}_0 \in A)} \\
&= \frac{\mu_{h+1}(A \times \overline{\mathbb{R}}_0^{d(h-1)} \times A)}{\mu_{h+1}(A \times \overline{\mathbb{R}}_0^{dh})}\,, \qquad h \geq 0\,.
\end{aligned}
$$

- Since

$$\frac{\mathrm{cov}(I(x^{-1}\mathbf{X}_0 \in A), I(x^{-1}\mathbf{X}_h \in A))}{\mathbb{P}(x^{-1}\mathbf{X}_0 \in A)}$$

$$= \mathbb{P}(x^{-1}\mathbf{X}_h \in A \mid x^{-1}\mathbf{X}_0 \in A) - \mathbb{P}(x^{-1}\mathbf{X}_0 \in A)$$

$$\to \rho_A(h)\,, \quad h \geq 0\,,$$

- $(\rho_A(h))$ **is the autocorrelation function of a stationary process.**

- **One can use the notions of classical time series analysis to describe the extremal dependence structure in a strictly stationary sequence.**

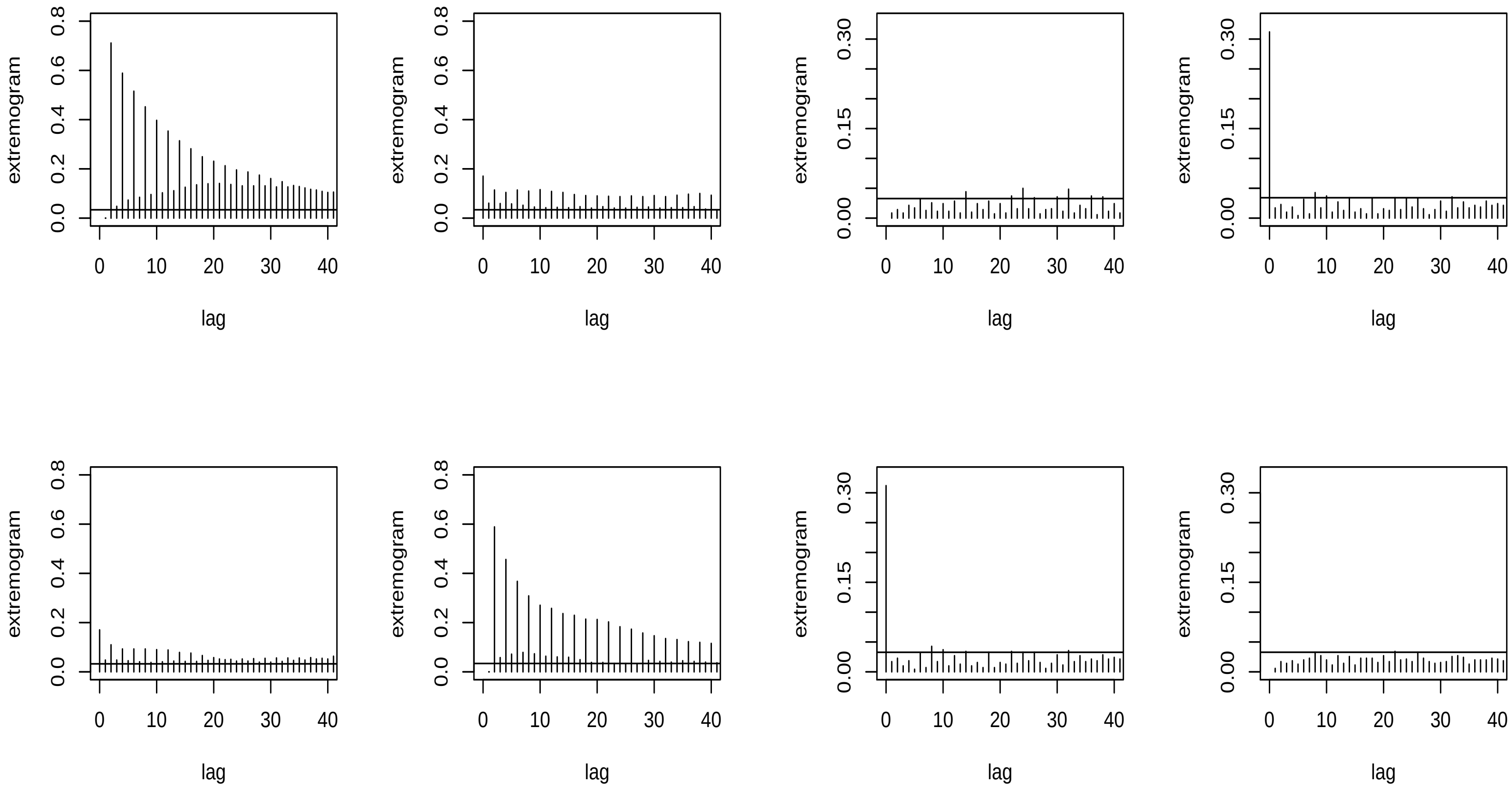

Figure 9. Five-minute returns of USD-DEM and USD-FRF foreign exchange rates. Left: (Cross-) extremograms of the original data. Right: (Cross-) extremograms of the residuals after an AR-GARCH fit.

- **Large deviations.** **Approximations to rare event probabilities such as**

$$\mathbb{P}(b_n^{-1} S_n \in A) \to 0$$

for sets A bounded away from zero. A.V. and S.V. Nagaev (1960-1979), Hult, Lindskog, M., Samorodnitsky (1998-2007), Bartkiewicz, Damek, M., Wintenberger (2007-)

- Tail bounds for subadditive functionals acting on a random walk with negative drift $(S_n - \mu\, n)$, where $\mathbb{E}[X_0] = 0$, $\alpha > 1$ and $\mu > 0$. For example, the ruin probability

$$\psi(u) = \mathbb{P}\Big(\sup_{n\ge 1} (S_n - \mu\, n) > u\Big)\,, \quad u \to \infty\,.$$

- The classical ruin bound for iid (X_t) Embrechts, Veraverbeke (1982)

$$\psi(u) \sim \mathrm{const}\, u\, \mathbb{P}(X_0 > u)\,, \qquad u \to \infty\,.$$

M., Samorodnitsky (1998-2002), with Hult, Lindskog for multivariate random walks (2005), with Buraczewski, Damek, Wintenberger (2007-) for Markov chains and more general structures

5. The principle of a single big jump (heavy-tail heuristics)

- **Consider an iid real-valued sequence (X_t) with partial sums**

$$S_n = X_1 + \cdots + X_n \,.$$

- **A large value of S_n appears in the most natural way, due to a single large summand X_t.**

- **As $x \to \infty$,**

$$\begin{aligned}
\mathbb{P}(S_n > x) &\sim \mathbb{P}\Big(\bigcup_{i=1}^{n} \{X_t > x\}\Big) \\
&\sim \mathbb{P}\Big(\bigcup_{i=1}^{n} \{X_t > x \,, X_j \le x \,, j \neq i\}\Big) \\
&\sim \sum_{i=1}^{n} \mathbb{P}(X_i > x) \,.
\end{aligned}$$

- **These heuristics remain valid for classes of heavy-tailed distributions other than the regularly varying ones, e.g. the subexponential distributions which are standard distribution in insurance mathematics and queuing theory.**

6. Large deviations for a regularly varying sequence

- **Partial sums** of a univariate regularly varying sequence (X_t):

$$S_n = X_1 + \cdots + X_n\,, \quad n \geq 1\,,$$

Assume $\mathbb{E}[X_0] = 0$ if $\mathbb{E}[|X_0|] < \infty$ and $\mathbb{P}(|X_0| > a_n) \sim n^{-1}$.

- Then the following relation holds for $\alpha > 0$ and suitable sequences $b_n \uparrow \infty$ A.V. Nagaev (1969), S.V. Nagaev (1979)

$$\lim_{n\to\infty} \sup_{x \geq b_n} \left| \frac{\mathbb{P}(S_n > x)}{n\,\mathbb{P}(|X_0| > x)} - p \right| = 0\,.$$

For $\alpha \leq 2$, one can choose any (b_n) such that $b_n/a_n \to \infty$, for $\alpha > 2$, $b_n > \sqrt{a n \log n}$, $a > \alpha - 2$.

- **A functional (Donsker) version for multivariate regularly varying summands holds in the iid case and is applied to get bounds for ruin probabilities.** Hult, Lindskog, M., Samorodnitsky (2005)
- **For dependent sequences, the limit in Nagaev's result has to be adjusted for extreme cluster effects.**

7. Concluding remarks

- Over the last 25 years, a (functional) calculus of regular variation for stochastic processes, and functionals acting on them has been developed.
- This calculus has been triggered by problems arising in areas such as time series analysis, data networks, climate research.
- Regular variation focuses on power laws, but not all heavy-tail phenomena are due to power law tails.
- The literature on processes with semi-exponential (subexponential) multivariate tails is relatively sparse and constitute a widely open field. See e.g. Asmussen, Rojas-Nandayapa (2008,2015), Foss, Korshunov, Zachary (2013), Tankov (2013)

- **Problems of simulating rare event probabilities, max-stable processes,... and estimation problems for regularly varying structures are difficult and unsolved.**