Heavy tail phenomena and dependence of extremes [1]

Thomas Mikosch
University of Copenhagen
www.math.ku.dk/∼mikosch/

[1] PhD Course an Extremes in Space and Time, 27-30, May, 2013

1. Heavy tails in real-life data

1.1. Finance.

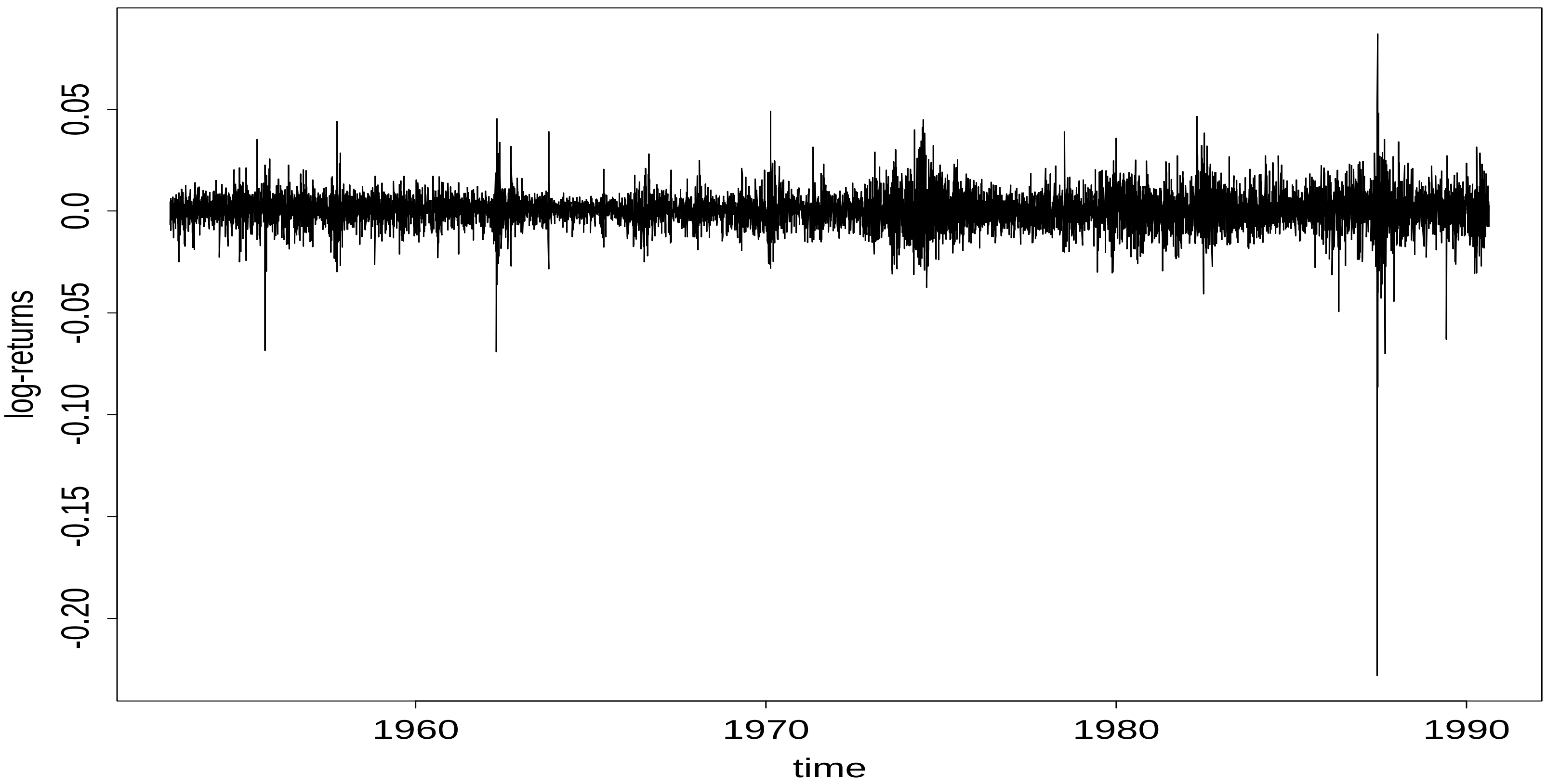
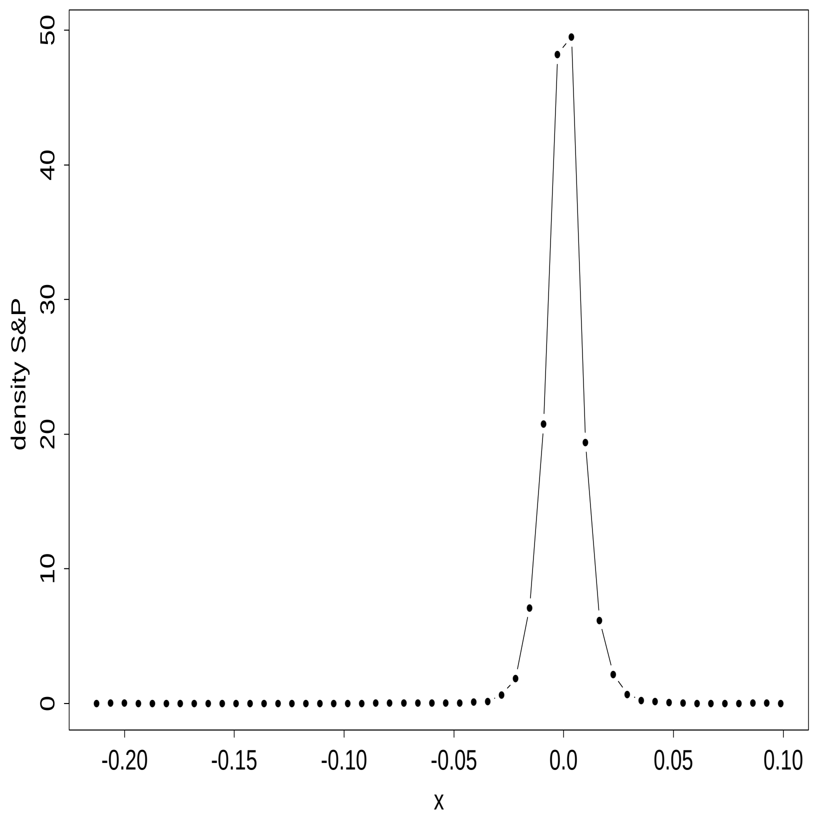


Figure 1. Plot of **9558** *S&P500* daily log-returns from January *2, 1953,* to December *31, 1990.* The year marks indicate the beginning of the calendar year.



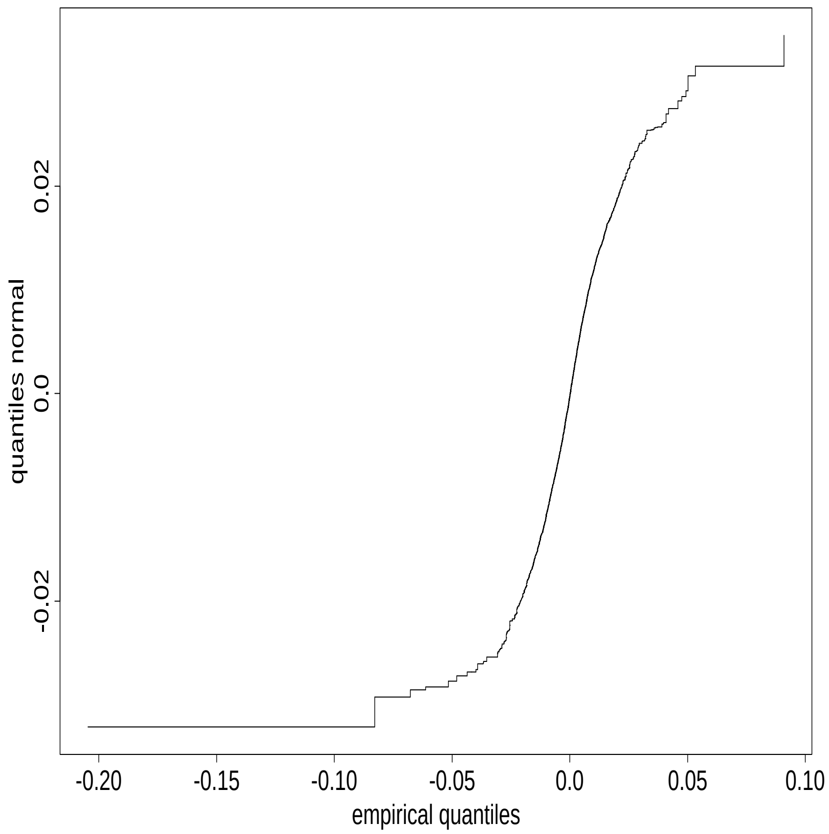


FIGURE 2. Left: Density plot of the *S&P500* data. The limits on the $\boldsymbol{x}$-axis indicate the range of the data. QQ-plot of the *S&P500* data against the normal distribution.

1.2. **Insurance.**

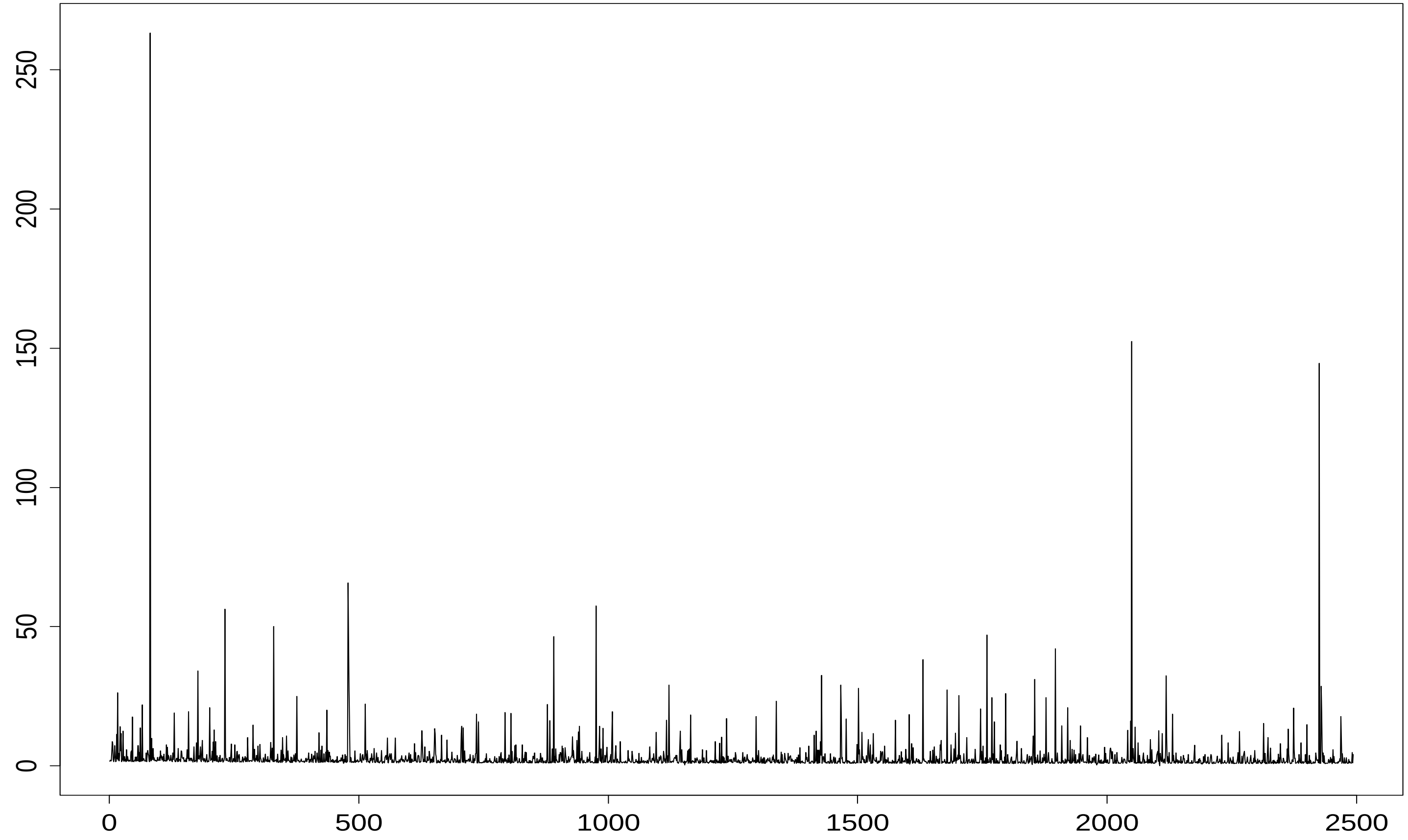


FIGURE 3. Danish fire insurance data.

1.3. **Telecommunications.**

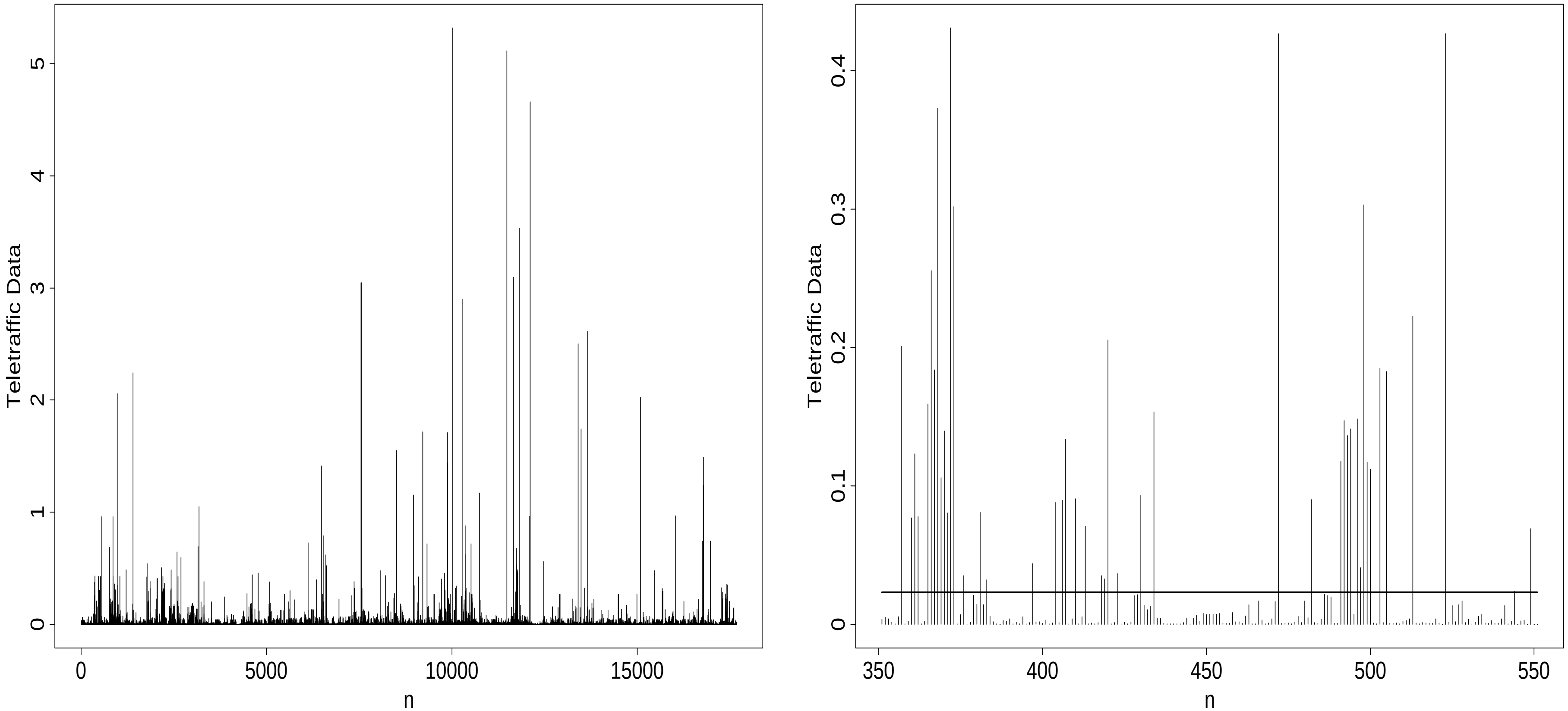


Figure 4. Time series of transmission durations (BU data).

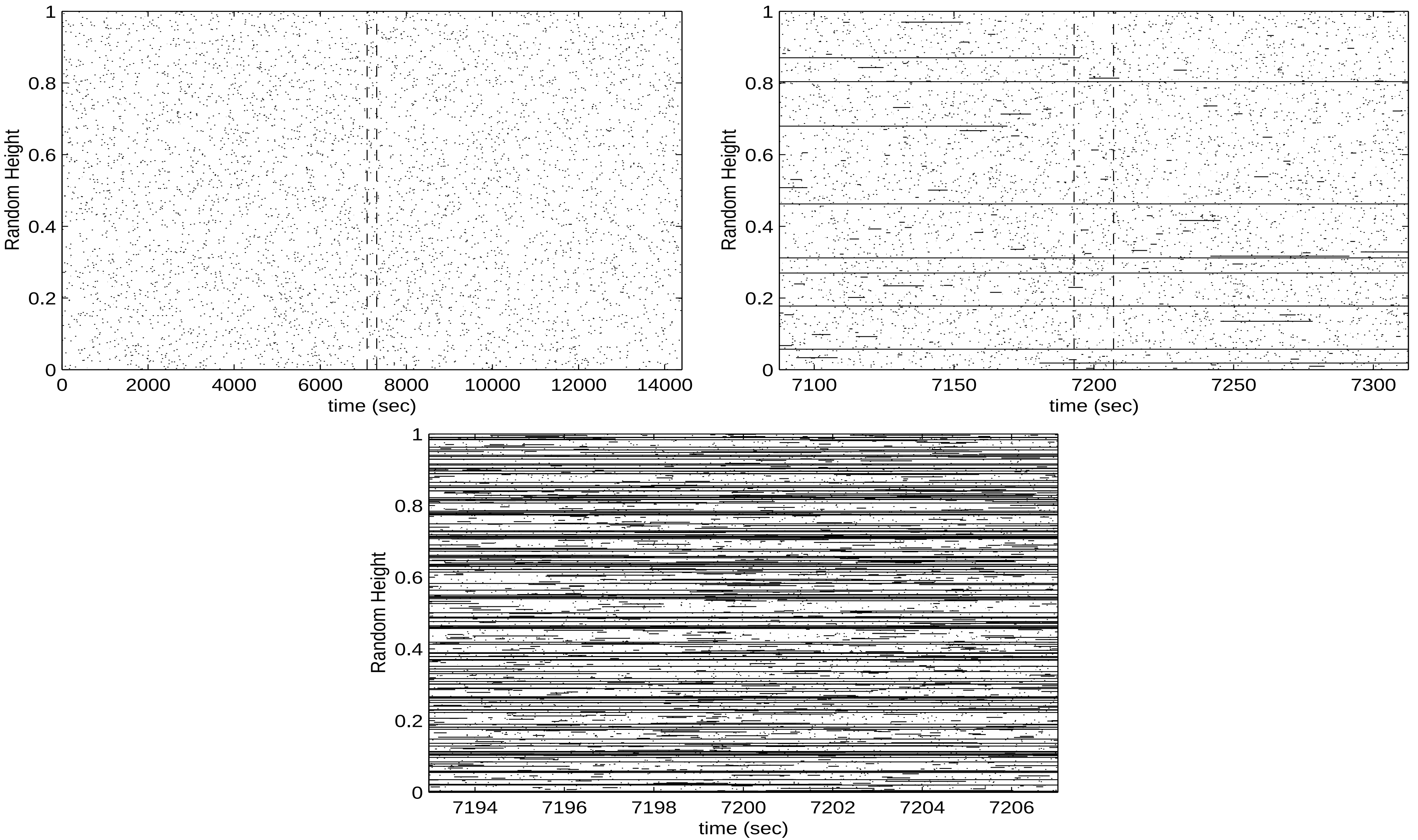


Figure 5. Mice and elephants plots (S. Marron).

2. Extremal dependence/independence in real-life data

2.1. Independence in insurance data.

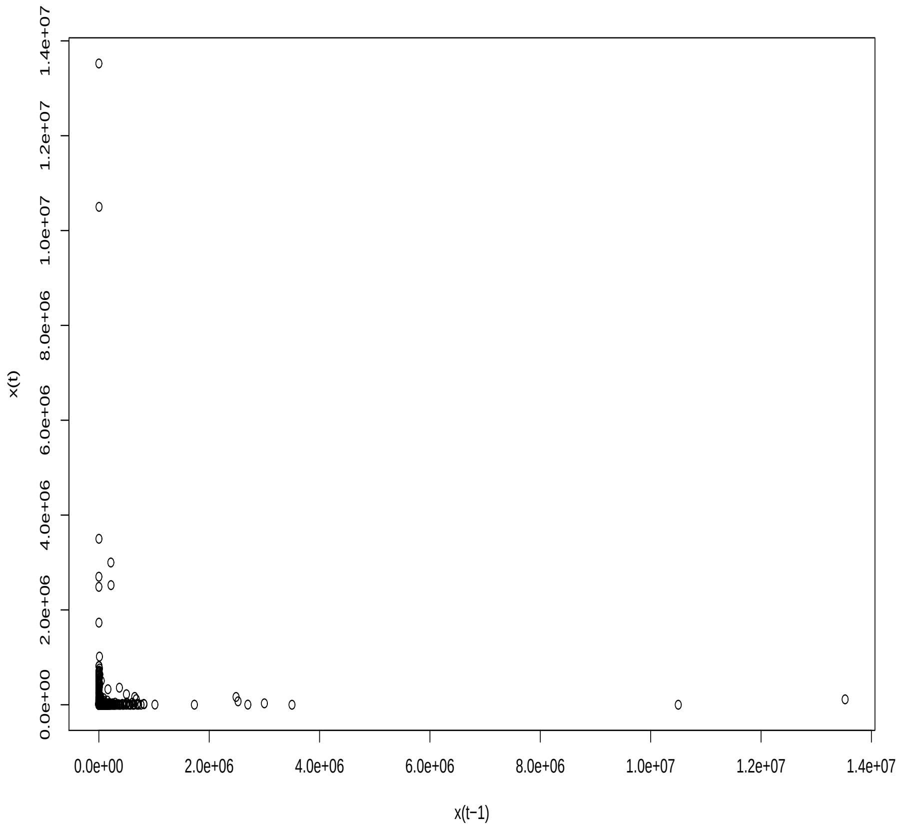


Figure 6. Scatterplot of US fire insurance losses - independence.

2.2. **Extremal independence in telecommunication data.**

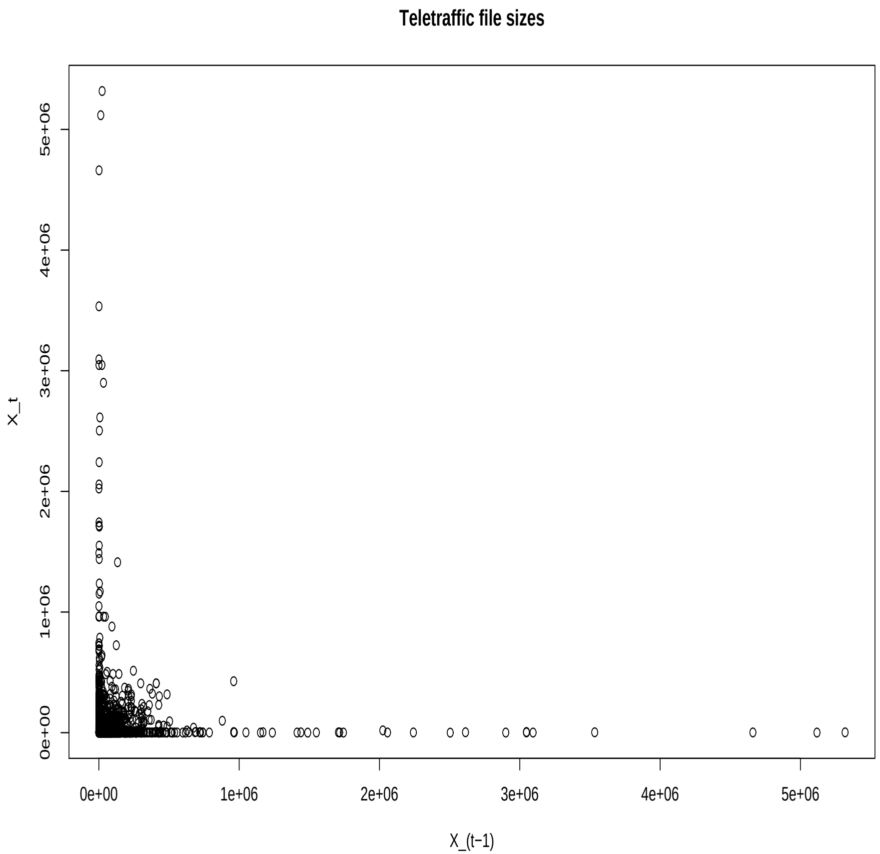


Figure 7. Scatterplot of file sizes of teletraffic data - extremal independence

2.3. **Extremal dependence in financial data.**

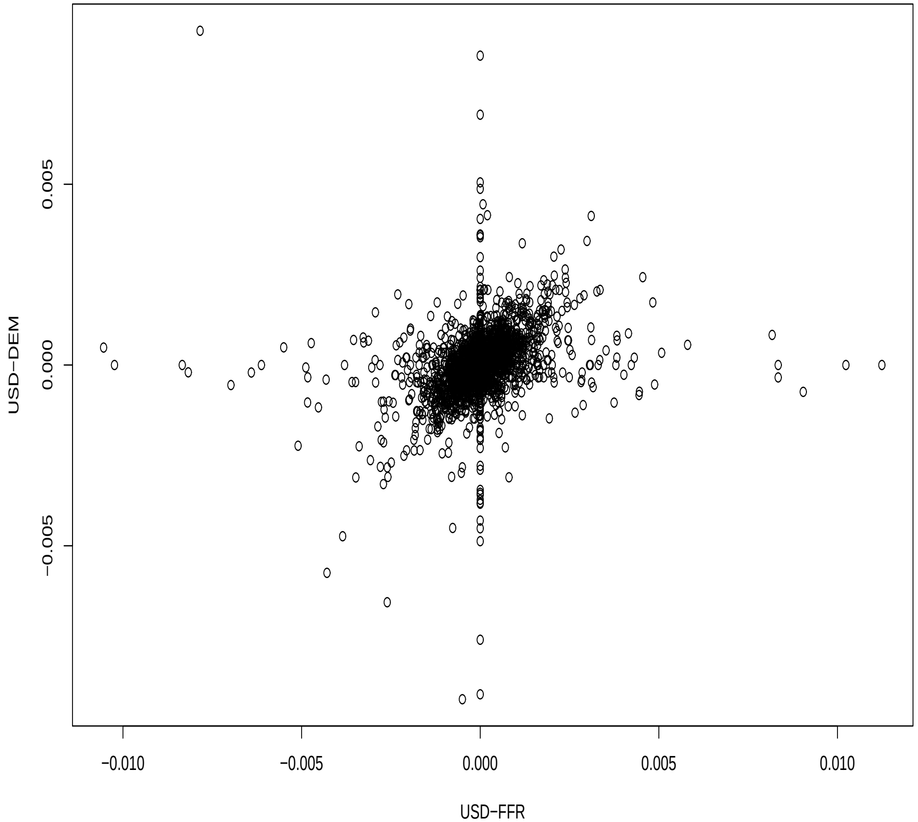


Figure 8. Scatterplot of 5 minute foreign exchange rate log-returns, USD-DEM against USD-FRF.

3. Extreme value theory for iid sequences Leadbetter et al. (1983), Resnick (1987), Embrechts et al. (1997), de Haan and Ferreira (2006)

3.1. Max-stable distributions.

- A random variable X and its distribution F are max-stable if for every $n \geq 2$ there exist $c_n > 0$, $d_n \in \mathbb{R}$, such that for iid copies (X_i) of X,

$$c_n^{-1}(M_n - d_n) = c_n^{-1}\big(\max_{i=1,\ldots,n} X_i - d_n\big) \stackrel{d}{=} X\,.$$

- **Any max-stable distribution belongs to the location/scale family of one of the three standard max-stable distributions (also called extreme value distributions):**

$$\Phi_\alpha(x) = e^{-x^{-\alpha}}, \quad x > 0, \quad \alpha > 0 \quad \text{Fréchet}$$

$$\Psi_\alpha(x) = e^{-|x|^{\alpha}}, \quad x < 0, \quad \alpha > 0 \quad \text{Weibull}$$

$$\Lambda(x) = e^{-e^{-x}}, \quad x \in \mathbb{R}, \quad \text{Gumbel}\,.$$

- **The max-stable distributions are the only possible non-degenerate weak limits for standardized maxima of an iid sequence (Fisher-Tippett Theorem 1928, Gnedenko (1943)).**
- **The 3 max-stable types can be written as one parametric family (generalized extreme value distribution (GEV)).**

3.2. Maximum domains of attraction (MDA).

- The distribution F of X is in the maximum domain of attraction of the max-stable distribution $G \in \{\Phi_\alpha, \Psi_\alpha, \Lambda\}$ ($F \in \mathrm{MDA}(G)$) if there exist constants $a_n > 0$, $b_n \in \mathbb{R}$ such that

$$\lim_{n\to\infty} P(a_n^{-1}(M_n - b_n) \le x) \to G(x)\,, \quad x \in \mathbb{R}\,.$$

- $F \in \mathrm{MDA}(\Phi_\alpha)$: Regular variation of the right tail

$$\overline{F}(x) = 1 - F(x) = P(X > x) = x^{-\alpha} L(x)\,, \quad x > 0\,,$$

for a slowly varying function L.

Then moments of order $\alpha + \delta$, $\delta > 0$, are infinite.

- $F \in \mathrm{MDA}(\Psi_\alpha)$: F has finite right endpoint x_F.
- $F \in \mathrm{MDA}(\Lambda)$: Moderately heavy $\to$ light tails.

- **Examples:**

  MDA(Φ_α): Student with α degrees of freedom,

  Cauchy ($\alpha = 1$),

  infinite variance α-stable distributions,

  Pareto $\overline{F}(x) = x^{-\alpha}$, $x > 1$,

  log-gamma distribution.

  MDA(Ψ_α): uniform, β-distribution.

  MDA(Ψ_α): log-normal distribution,

  Weibull $\overline{F}(x) = \mathrm{e}^{-x^\tau}$, $x > 0$, $\tau > 0$,

  gamma distribution,

  normal distribution.

3.3. The Pickands-Balkema-de Haan Theorem and the Generalized Pareto Distribution (GPD).

- $F \in \mathrm{MDA}(G)$ for a max-stable distribution G if and only if there exists $a(u) > 0$ such that

$$\lim_{u \uparrow x_F} P\Big(\frac{X-u}{a(u)} > x \mid X > u\Big)$$

$$= \begin{cases} (1+\alpha^{-1}x)^{-\alpha} & x > 0, & G = \Phi_\alpha\,, \\ (1-\alpha^{-1}x)^{\alpha} & 0 < x < \alpha, & G = \Psi_\alpha\,, \\ \mathrm{e}^{-x} & x > 0, & G = \Lambda\,. \end{cases}$$

$$= \overline{G}_\xi(x)\,.$$

- G_ξ defines the Generalized Pareto Distribution (GPD) with

$$\xi = \begin{cases} 1/\alpha & \Phi_\alpha\,, \\ -1/\alpha & \Psi_\alpha\,, \\ 0 & \Lambda\,. \end{cases}$$

4. The extremal index – a measure of the extremal cluster size

- **Let $(X_t)_{t\in\mathbb{Z}}$ be a strictly stationary real-valued time series.**
- **Its autocovariance and autocorrelation functions do in general not contain information about extremal dependence.**

The extremal index.

- **The *extremal index* θ_X is a standard measure of extremal dependence in a sequence:[2] for $M_n = \max_{t=1,\dots,n} X_t$ and some sequence $u_n \uparrow x_F$**

$$P(M_n \le u_n) \approx [P(X_1 \le u_n)]^{n\,\theta_X}\,.$$

- **$\theta_X \in [0,1]$ has the interpretation as reciprocal of the expected cluster size above high thresholds.**

[2] See Leadbetter, Lindgren, Rootzén (1983); cf. Embrechts et al. (1997), Section 8.1

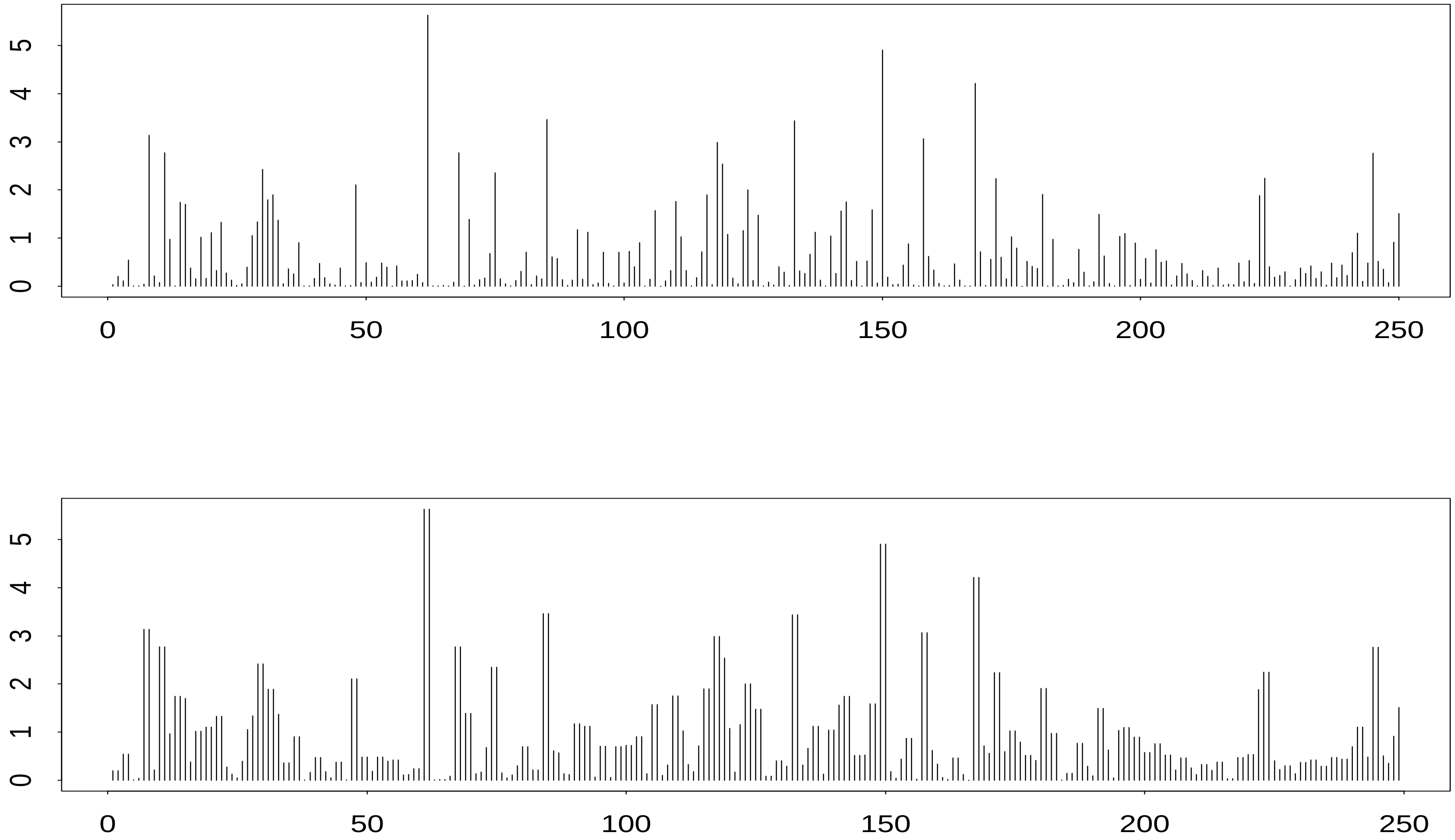


FIGURE 9. A sequence of iid random variables $\boldsymbol{Y_i}$ (Top) with distribution function $\sqrt{\boldsymbol{F}}$, where $\boldsymbol{F}$ is standard exponential. Bottom: the sequence of pairwise maxima $\mathbf{max}(\boldsymbol{Y_i}, \boldsymbol{Y_{i+1}})$ with distribution $\boldsymbol{F}$. By construction, extremes appear in clusters of size 2. The extremal index is $\boldsymbol{\theta = 1/2}$.

Examples.

- A Gaussian stationary sequence (X_t) with autocorrelation function $\rho_X(h) = o(1/\log h)$, $h \to \infty$:

  $\theta_X = 1$. No extremal clustering.

- AR(1) model $X_t = \phi\, X_{t-1} + Z_t$, $\phi \in (-1,1)$, (Z_t) iid student with α degrees of freedom:

  $\theta_X = 1 - |\phi|^\alpha$.

- Models for log-returns $X_t = \log P_t - \log P_{t-1}$:

  $$X_t = \sigma_t\, Z_t, \qquad (Z_t) \text{ iid}, \qquad \sigma_t > 0$$

- The simple stochastic volatility model: $(\log \sigma_t)$ linear Gaussian, independent of iid student (Z_t):

  $\theta_X = 1$ Davis, Mikosch (2001ab,2009ab) No extremal clustering.

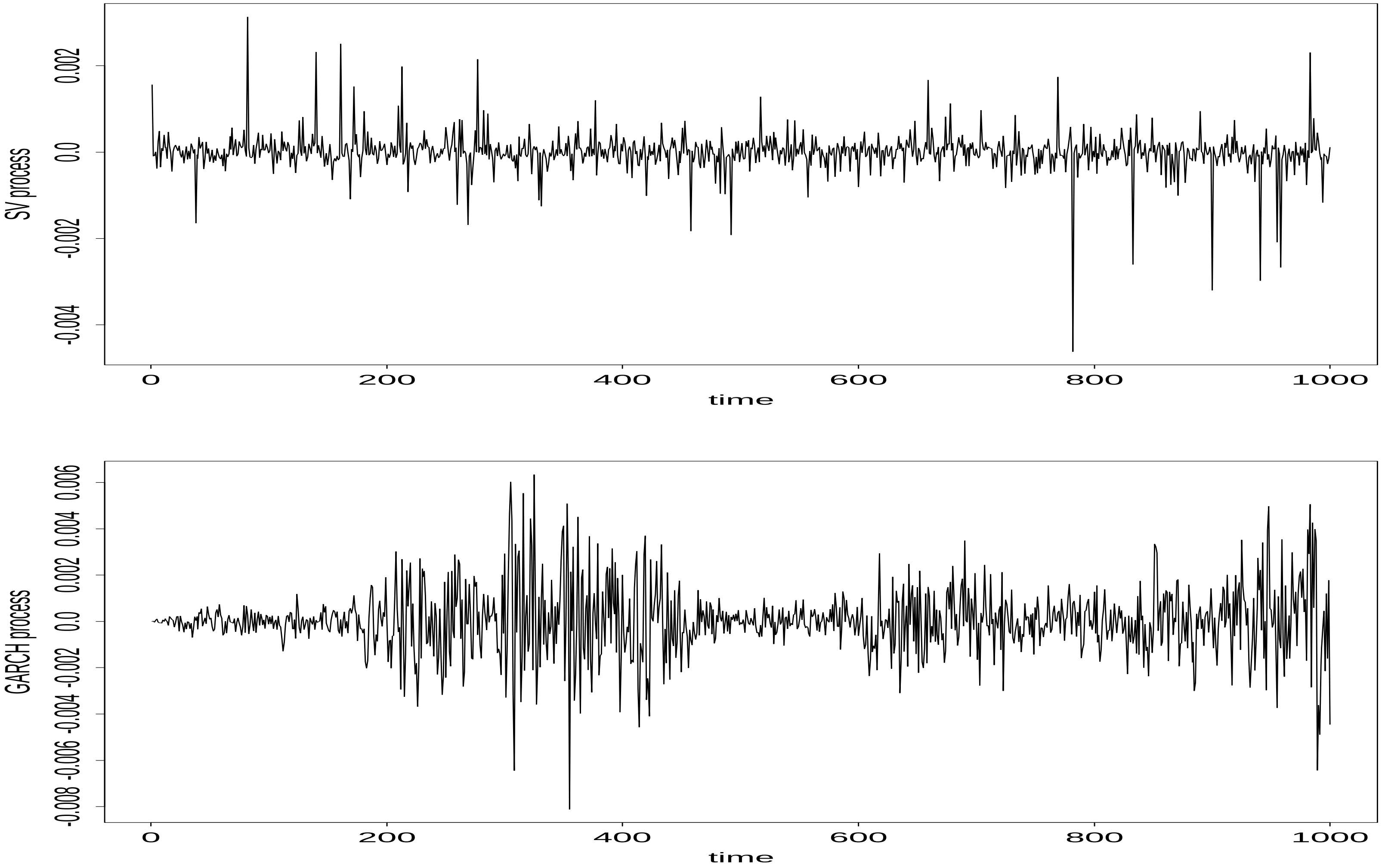


FIGURE 10. *Top:* Stochastic volatility process $\boldsymbol{X_t = \sigma_t \, Z_t}$ for iid student $\boldsymbol{(Z_t)}$ with $\mathbf{4}$ degrees of freedom, Gaussian ARMA(1,1) process $\boldsymbol{\log \sigma_t = 0.5 \log \sigma_{t-1} + 0.3\eta_{t-1} + \eta_t}$. *Bottom:* GARCH$\boldsymbol{(1,1)}$ process $\boldsymbol{X_t = (0.0001 + 0.1X_{t-1}^2 + 0.9\sigma_{t-1}^2)^{0.5} Z_t}$ for iid standard normal $\boldsymbol{(Z_t)}$.

- The GARCH(1,1) model:[3] $X_t = \sigma_t Z_t$,

$$\sigma_t^2 = \alpha_0 + (\alpha_1 Z_{t-1}^2 + \beta_1)\,\sigma_{t-1}^2\,, \quad (Z_t) \text{ iid } N(0,1).$$

There exists $\alpha > 0$ such that $E(\alpha_1 Z_1^2 + \beta_1)^{\alpha/2} = 1$ and[4]

$$\frac{\alpha}{2}\int_1^\infty P\left(\max_{n\geq 1}\prod_{t=1}^n(\alpha_1 Z_t^2 + \beta_1) \leq y^{-1}\right)\, y^{-\frac{\alpha}{2}-1}\,dy = \theta_\sigma \in (0,1)\,.$$

[3] Bollerslev (1986)

[4] de Haan, Resnick, Rootzén, de Vries (1989)

- **The GARCH(1, 1) model:** $X_t = \sigma_t Z_t$,

$$\sigma_t^2 = \alpha_0 + (\alpha_1 Z_{t-1}^2 + \beta_1)\,\sigma_{t-1}^2\,, \quad (Z_t) \text{ iid } N(0,1).$$

There exists $\alpha > 0$ such that $E(\alpha_1 Z_1^2 + \beta_1)^{\alpha/2} = 1$ and

$$\frac{\alpha}{2}\int_1^\infty P\left(\max_{n\geq 1}\prod_{t=1}^n (\alpha_1 Z_t^2 + \beta_1) \leq y^{-1}\right) y^{-\frac{\alpha}{2}-1}\,dy = \theta_\sigma \in (0,1)\,.$$

- **Expressions for the extremal index of a stationary process are often complicated.**

- The GARCH(1, 1) model: $X_t = \sigma_t Z_t$,

$$\sigma_t^2 = \alpha_0 + (\alpha_1 \, Z_{t-1}^2 + \beta_1) \, \sigma_{t-1}^2 \,, \quad (Z_t) \text{ iid } N(0,1).$$

There exists $\alpha > 0$ such that $E(\alpha_1 Z_1^2 + \beta_1)^{\alpha/2} = 1$ and

$$\frac{\alpha}{2} \int_1^\infty P \left(\max_{n \ge 1} \prod_{t=1}^n (\alpha_1 \, Z_t^2 + \beta_1) \le y^{-1} \right) \, y^{-\frac{\alpha}{2} - 1} \, dy = \theta_\sigma \in (0,1) \,.$$

- Expressions for the extremal index of a stationary process are often complicated.

- Monte Carlo simulation is not straightforward.

- The GARCH(1,1) model: $X_t = \sigma_t Z_t$,

$$\sigma_t^2 = \alpha_0 + (\alpha_1 Z_{t-1}^2 + \beta_1)\,\sigma_{t-1}^2\,, \quad (Z_t) \text{ iid } N(0,1).$$

There exists $\alpha > 0$ such that $E(\alpha_1 Z_1^2 + \beta_1)^{\alpha/2} = 1$ and

$$\frac{\alpha}{2}\int_1^\infty P\left(\max_{n\ge 1}\prod_{t=1}^n(\alpha_1 Z_t^2 + \beta_1) \le y^{-1}\right) y^{-\frac{\alpha}{2}-1}\,dy = \theta_\sigma \in (0,1)\,.$$

- Expressions for the extremal index of a stationary process are often complicated.

- Monte Carlo simulation is not straightforward.

- Estimation of the extremal index and extremal cluster size distribution is non-trivial; see C. Y. Robert (2009)

5. Time series models with heavy tails

5.1. Regularly varying distributions.

- Recall that $F \in \mathrm{MDA}(\Phi_\alpha)$ for some $\alpha > 0$ if and only if

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

for some slowly varying function L.

- We call a random variable $X \in \mathbb{R}$ and its distribution F regularly varying with index $\alpha > 0$ if there exist constants $p, q \geq 0$ such that $p + q = 1$ and

$$F(-x) \sim q\, x^{-\alpha} L(x) \quad \text{and} \quad \overline{F}(x) \sim p\, x^{-\alpha} L(x)\,, \quad x \to \infty\,.$$

If e.g. $p = 0$: $\overline{F}(x) = o(x^{-\alpha} L(x))$, $x \to \infty$.

- Equivalently, $|X|$ is regularly varying with index $\alpha > 0$ and

$$\frac{P(X \le -x)}{P(|X| > x)} \to q \quad \text{and} \quad \frac{P(X > x)}{P(|X| > x)} \to p\,, \quad x \to \infty\,.$$

- Examples. Pareto, student, Cauchy, α-stable, $\alpha \in (0,2)$, Burr, log-gamma, Fréchet.

- **Two fundamental operations.**[5]

Convolution. Feller (1971) Let $X_1 > 0$ be regularly varying with $\alpha > 0$. Assume X_2 regularly varying with index α and independent of X_1 OR $P(|X_2| > x) = o(P(X_1 > x))$. Then $X_1 + X_2$ is regularly varying with index α and

$$P(X_1 + X_2 > x) \sim P(X_1 > x) + P(X_2 > x), \quad x \to \infty.$$

Products. Breiman (1965) $\sigma > 0$, $X > 0$ independent and $E\sigma^{\alpha+\delta} < \infty$ for some $\delta > 0$, X regularly varying with index α. Then as $x \to \infty$,

$$P(\sigma X > x) \sim E\sigma^{\alpha} P(X > x).$$

[5] Cf. Resnick (2007)

- **Examples.**

  Stochastic volatility model. $X_t = \sigma_t Z_t$, $t \in \mathbb{Z}$, σ_t log-normal, (Z_t) iid regularly varying with index α, (σ_t) and (Z_t) independent. Then as $x \to \infty$,

$$P(X_t > x) \sim E\sigma_0^\alpha P(Z_0 > x)\,,$$
$$P(X_t \le -x) \sim E\sigma_0^\alpha P(Z_0 \le -x)\,.$$

  Moving average. $X_t = \theta_0 Z_t + \theta_1 Z_{t-1} + \cdots + \theta_m Z_{t-m}$, $t \in \mathbb{Z}$, $m \ge 1$, $Z_t > 0$ iid regularly varying with index α.

$$P(X_t > x) \sim P(Z_0 > x) \sum_{i=0}^{m} |\theta_i|^\alpha \big(I_{\theta_i > 0} + I_{\theta_i < 0}\big)\,, \quad x \to \infty\,.$$

How can one model extremal spatio-temporal dependence and heavy tails?

- **One needs to model both the size and the direction of extremes.**

Asymptotic extremal independence.

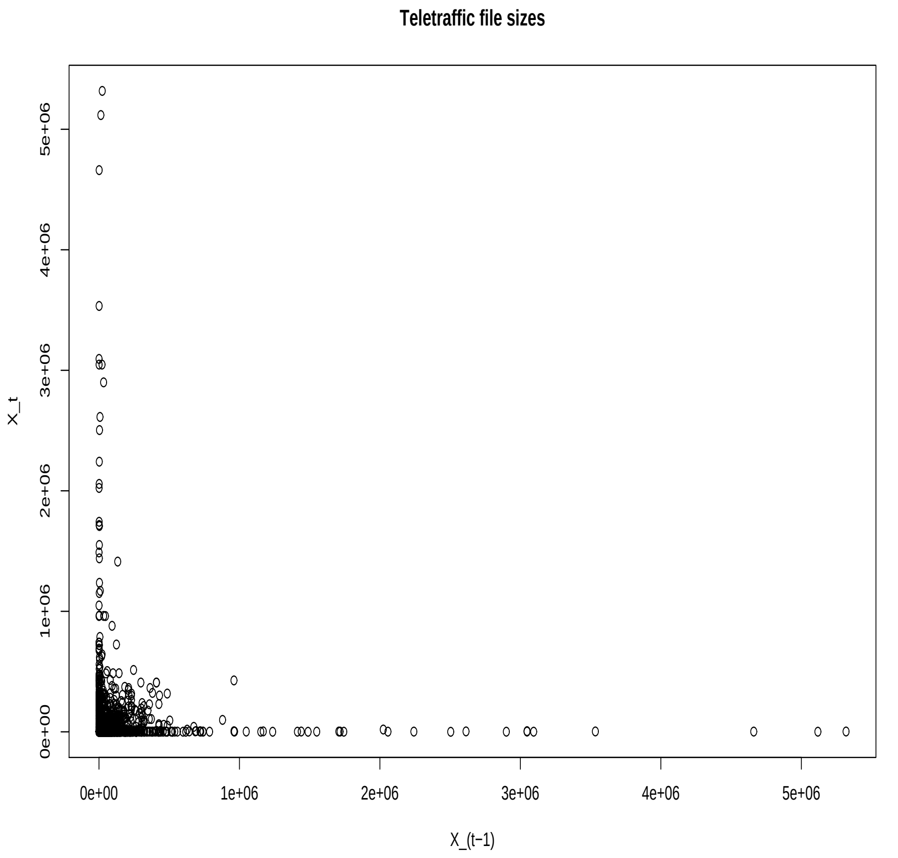


FIGURE 11. Scatterplot of file sizes of teletraffic data.

Asymptotic extremal dependence.

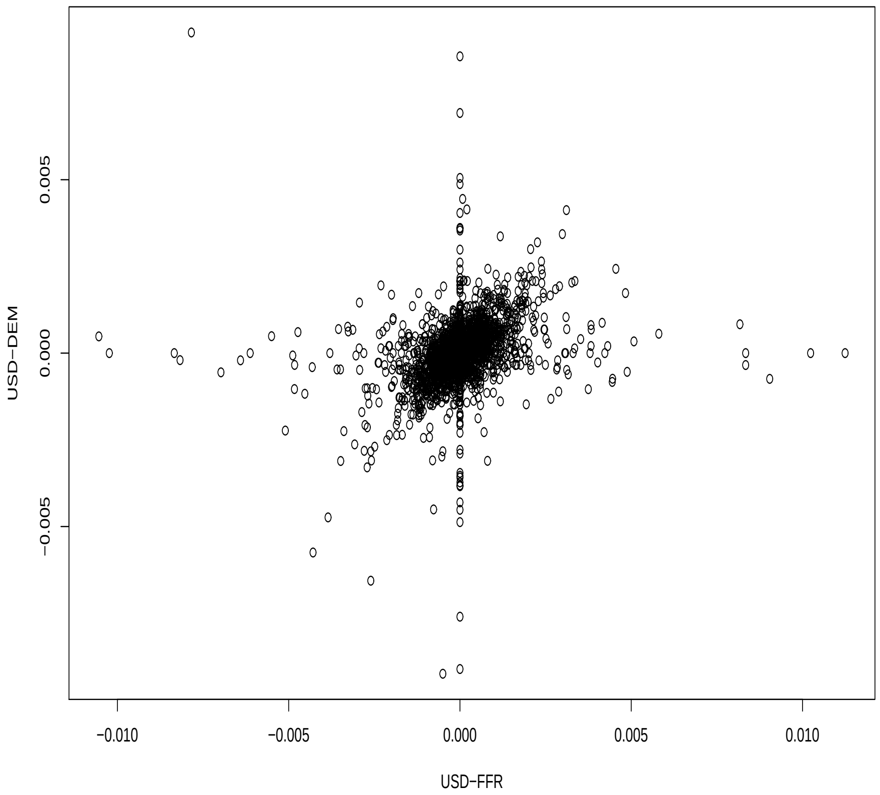


Figure 12. Scatterplot of 5 minute foreign exchange rate log-returns, USD-DEM against USD-FRF.

5.2. Multivariate regular variation Resnick (1987,2007).

- A random vector $\mathbf{X} \in \mathbb{R}^d$ and its distribution are regularly varying with index $\alpha > 0$: there exists a random vector $\mathbf{\Theta} \in \mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : |x| = 1\}$ such that for $t > 0$:

$$\frac{P\left(|\mathbf{X}| > tx\,, \mathbf{X}/|\mathbf{X}| \in \cdot\right)}{P(|\mathbf{X}| > x)} \xrightarrow{w} t^{-\alpha} P(\mathbf{\Theta} \in \cdot)\,, \quad x \to \infty\,.$$

The distribution of $\mathbf{\Theta}$: spectral measure of $\mathbf{X}$.

- Equivalently,

$$\frac{P\left(x^{-1}\mathbf{X} \in \cdot\right)}{P(|\mathbf{X}| > x)} \xrightarrow{v} \mu(\cdot)\,,$$

for a non-null Radon measure μ on the Borel σ-field of $\overline{\mathbb{R}}_0^d = \overline{\mathbb{R}}^d \backslash \{0\}$ with $\mu(tA) = t^{-\alpha}\mu(A)$, $t > 0$.

- Equivalently: as $x \to \infty$,

$$\frac{P(|\mathbf{X}| > tx)}{P(|\mathbf{X}| > x)} \to t^{-\alpha}, \quad t > 0, \quad \text{and}$$

$$P\left(\frac{\mathbf{X}}{|\mathbf{X}|} \in \cdot \mid |\mathbf{X}| > x\right) \xrightarrow{w} P(\Theta \in \cdot)\,.$$

- A toy example. R, θ independent, θ distributed on $[0, 2\pi)$,

$P(R > r) = r^{-\alpha}$, $r > 1$, Pareto.

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

Then

$$|\mathbf{X}| = R \quad \text{and} \quad \Theta = (\cos\theta, \sin\theta)\,.$$

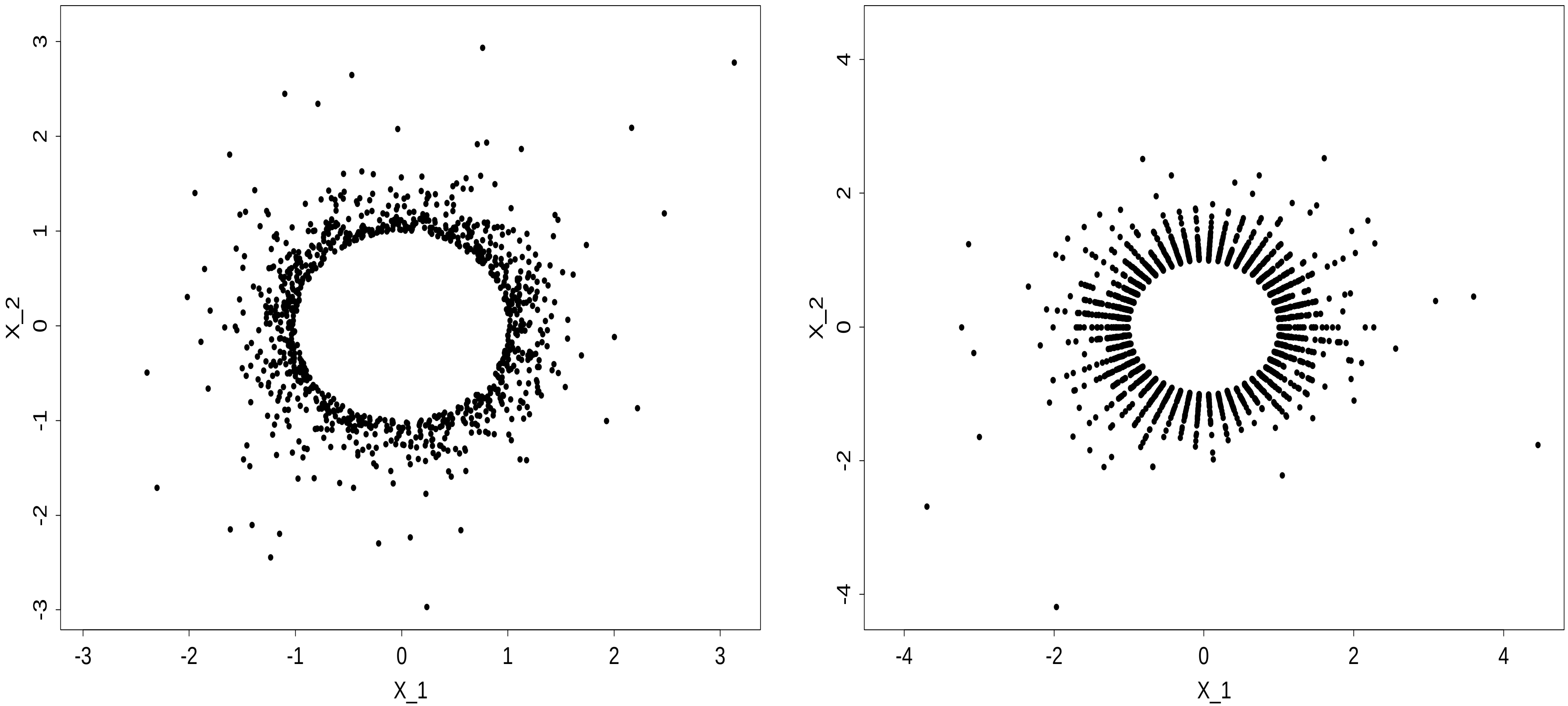


FIGURE 13. IID vectors $\mathbf{X}_i$ from the toy model with tail index $\alpha = 5$. Left: θ is uniform on $[0, 2\pi)$. Right: θ has a discrete uniform distribution on the points $2\pi i/50$.

- **Examples**
  - **X has iid student(α) distributed components. $P(\boldsymbol{\Theta} \in \cdot)$ is concentrated on the intersection of unit ball and axes.**
  - **X has a multivariate student(α) distribution. $P(\boldsymbol{\Theta} \in \cdot)$ is supported on the whole unit ball.**
  - **X is obtained from a Gaussian vector by transforming the marginals to student(α). Then $P(\boldsymbol{\Theta} \in \cdot)$ is concentrated on the intersection of unit ball and axes.**

6. Regularly varying stationary sequences

- **A real-valued stationary sequence (X_t) is regularly varying with index $\alpha > 0$ if its finite-dimensional distributions are regularly varying with index α.**
- **Equivalently, for every $k \geq 1$,**

$$\frac{P(x^{-1}(X_1,\ldots,X_k) \in \cdot)}{P(|X_0| > x)} \stackrel{v}{\rightarrow} \mu_k(\cdot)\,.$$

The measures μ_k determine the extremal dependence structure of the finite-dimensional distributions.

- **Notice: Normalization $P(|X_0| > x)$ does not depend on k.**

Examples of regularly varying stationary sequences

Linear processes.

- Examples of linear processes are ARMA processes with iid noise (Z_t), e.g. the AR(p) and MA(q) processes

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

- A linear process

$$X_t = \sum_j \psi_j Z_{t-j}, \quad t \in \mathbb{Z},$$

is regularly varying with index $\alpha > 0$ if the iid sequence (Z_t) is regularly varying with index α.

- **Under mild conditions on (ψ_j),**[6]

$$\frac{P(X > x)}{P(|Z| > x)} \sim \sum_j |\psi_j|^\alpha (p\, I_{\psi_j>0} + q\, I_{\psi_j<0}) = \|\psi\|_\alpha^\alpha\,, \quad x \to \infty\,.$$

6 See Resnick (1987); cf. Embrechts et al. (1997), Appendix

Solutions to stochastic recurrence equation.

- For an iid sequence $((A_t, B_t))_{t\in\mathbb{Z}}$, $A > 0$, the stochastic recurrence equation

$$X_t = A_t X_{t-1} + B_t\,, \quad t \in \mathbb{Z}\,,$$

has a unique stationary solution

$$X_t = B_t + \sum_{i=-\infty}^{t-1} A_t \cdots A_{i+1} B_i\,, \quad t \in \mathbb{Z},$$

provided $E \log A < 0$, $E|\log|B|| < \infty$.

- The sequence (X_t) is regularly varying with index α which is the unique solution to $EA^\kappa = 1$, $\kappa > 0$, (given this solution exists) Kesten (1973), Goldie (1991) and

$$P(X_0 > x) \sim c_+ x^{-\alpha}\,, \quad P(X_0 \le -x) \sim c_- x^{-\alpha}\,, \quad x \to \infty\,.$$

- The GARCH(1, 1) process[7] satisfies a stochastic recurrence equation: for an iid standard normal sequence (Z_t), positive parameters $\alpha_0, \alpha_1, \beta_1$,

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

The process $X_t = \sigma_t Z_t$ is regularly varying with index α satisfying $E(\alpha_1 Z^2 + \beta_1)^{\alpha/2} = 1$.

Other examples of regularly varying sequences.

- α-stable stationary processes are regularly varying with index α provided $\alpha \in (0, 2)$. Samorodnitsky and Taqqu (1994)
- Max-stable stationary processes with Fréchet marginals are regularly varying.

[7] Bollerslev (1986)

References

[1] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987) *Regular Variation.* Cambridge University Press, Cambridge.

[2] Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31**, 307–327.

[3] Breiman, L. (1965) On some limit theorems similar to the arc-sin law. *Theory Probab. Appl.* **10**, 323–331.

[4] Brockwell, P.J. and Davis, R.A. (1991) *Time Series: Theory and Methods,* 2nd edition. Springer, New York.

[5] Brockwell, P.J. and Davis, R.A. (1996) *Introduction to Time Series and Forecasting.* Springer, New York.

[6] 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.

[7] 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.).

[8] Davis, R.A. and Mikosch, T. (2009a) Fundamental properties of stochastic volatility models. In: Andersen, T.G., Davis, R.A., Kreiss, J.-P. and Mikosch, T. *Handbook of Financial Time Series.* Springer, Berlin.

[9] Davis, R.A. and Mikosch, T. (2009b) 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.

[10] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin.

[11] Feller, W. (1971) *An Introduction to Probability Theory and Its Applications.* Vol. II. Second edition. Wiley, New York.

[12] Goldie, C.M. (1991) Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.* **1**, 126–166.

[13] Haan, L. de, Fereira, A. (2006) *Extreme Value Theory: An Introduction.* Springer, Berlin, New York.

[14] Haan, L. de, Resnick, S.I., Rootzén, H. and Vries, C. de (1989) Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes. *Stoch. Proc. Appl.* **32**, 213–224.

[15] Kesten, H. (1973) Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131**, 207–248.

[16] Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983) *Extremes and Related Properties of Random Sequences and Processes.* Springer, Berlin.

[17] Resnick, S.I. (1987) *Extreme Values, Regular Variation, and Point Processes.* Springer, New York.

[18] Resnick, S.I. (2007) *Heavy Tail Phenomena: Probabilistic and Statistical Modeling.* Springer, New York.

[19] Robert, C.Y. (2009) Inference for the limiting cluster size distribution of extreme values. *Ann. Statist.* **37**, 271–310.

[20] Samorodnitsky, G. and Taqqu, M.S. (1994) *Stable Non–Gaussian Random Processes. Stochastic Models with Infinite Variance.* Chapman and Hall, London.