On the estimation of the heavy-tail exponent and the extremal index in time series

Stilian Stoev
University of Michigan, Ann Arbor
(based on joint works with K. Hamidieh, G. Michailidis, and M. Taqqu)

Departamento de Estadística
Universidad Carlos III de Madrid, May 7, 2010

Outline

- Heavy–tailed time series
- Scaling of maxima: Tail exponent α and extremal index θ
- The max–spectrum: Estimation of α and θ
- Some asymptotic results
- Applications

Preliminaries

A r.v. X has a heavy right tail with tail exponent $\alpha > 0$, if:

$$\mathbb{P}\{X > x\} \sim L(x)x^{-\alpha}, \quad \text{as } x \to \infty,$$

where $L(x)$ is a slowly varying function.

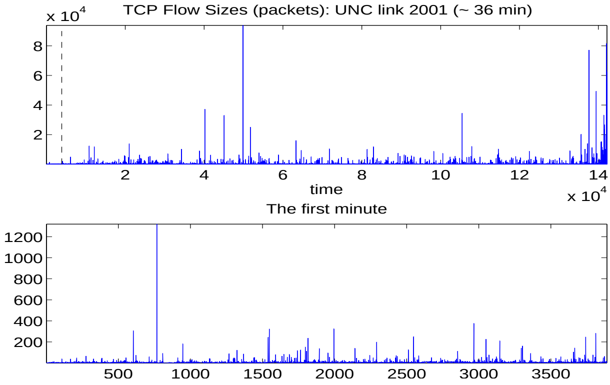
- For simplicity, here we focus on positive X with $L(x) \equiv \mathrm{const}$.
- Thus, for $p > 0$, the moment:

$$\mathbb{E}X^p < \infty \quad \text{if and only if} \quad p < \alpha.$$

- Heavy tailed models are natural in many applications: physics, meteorology, insurance, finance, telecommunications, etc.
- Pareto, $\alpha-$stable, $t-$distribution, Cauchy are all heavy–tailed.

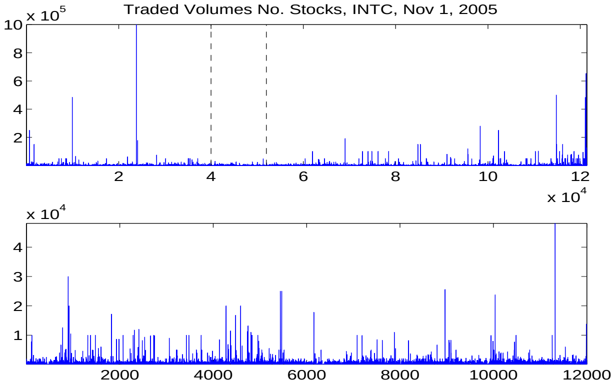
Heavy tails everywhere: Network traffic data

- Transmission Control Protocol (TCP) connection sizes (in # packets)



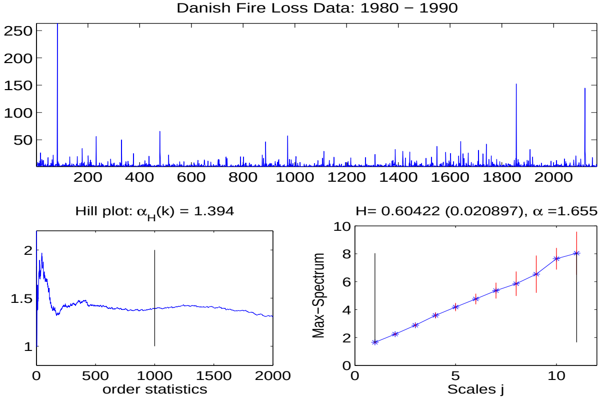
Heavy tails everywhere: Traded volume data

- Traded volume per individual transactions (in # shares)



Heavy tails everywhere: Insurance data

- Insurance claims due to fire–loss (in MLN Danish Krones)



Tail exponent estimation: an old problem

• Hill (1975) derived the MLE in the Pareto model
$\mathbb{P}\{X > x\} = x^{-\alpha},\ x \geq 1.$
∘ He arrived at the so–called *Hill plot*:

$$\widehat{\alpha}_H(k) := \Big(\frac{1}{k}\sum_{i=1}^{k}\log(X_{i,n}) - \log(X_{k+1,n})\Big)^{-1},$$

where $X_{1,n} \geq X_{2,n} \geq \cdots \geq X_{k,n}$ are the *top–k order statistics* of the sample.
• A lot of work for iid data – less for dependent:
∘ Resnick and Stărică (1995) – consistency of Hill–type estimators.
∘ Johnathan Hill (2006?) – asymptotic normality of Hill–type estimators under NED (near epoch dependence) conditions. ∘ ...
• Even for iid data, Hill plots are: *volatile & hard to interpret: "Hill horror plot"*

Scaling of Maxima

Scaling of maxima: Independent data

Consider iid, positive, and heavy–tailed $X_1, X_2, \ldots, X_n, \ldots$. Let

$$M_n = \max_{1\le i\le n} X_i \equiv \bigvee_{1\le i\le n} X_i.$$

If $\mathbb{P}\{X > x\} \sim cx^{-\alpha}$, $x \to \infty$, then

$$\frac{1}{n^{1/\alpha}} M_n \xrightarrow{d} Z, \quad \text{as } n \to \infty,$$

where Z has the α–Fréchet distribution:

$$\mathbb{P}\{Z \le x\} = e^{-cx^{-\alpha}}, \quad (x > 0).$$

- Indeed,

$$\mathbb{P}\{M_n \le n^{1/\alpha}x\} = (1-\mathbb{P}\{X_1 > n^{1/\alpha}x\})^n \sim \Big(1-\frac{cx^{-\alpha}}{n}\Big)^n \sim e^{-cx^{-\alpha}},$$

for other extreme value distributions and their domains of attraction, see e.g. Resnick (1987).

Scaling of maxima: Dependent data

Let now $\{X_k\}_{k\in\mathbb{Z}}$ be a stationary, heavy–tailed time series with tail exponent $\alpha > 0$ and

$$M_n = \max_{1\le i\le n} X_i \equiv \bigvee_{1\le i\le n} X_i.$$

- How do the maxima M_n scale?

◦ By the seminal results of Leadbetter (see e.g. Leadbetter et al (1983)) under mild dependence conditions "$D(u_n)$":

$$\frac{1}{n^{1/\alpha}} M_n \xrightarrow{d} \theta^{1/\alpha} Z, \quad \text{with } \theta \in [0,1],$$

where for iid $\{X_k^*\}_{k\in\mathbb{Z}}$, with $X_i =^d X_i^*$, we have

$$\frac{1}{n^{1/\alpha}} \bigvee_{1\le i\le n} X_i^* \xrightarrow{d} Z, \quad \text{as } n\to\infty.$$

- Here $\theta \in [0,1]$ is called the extremal index of the time series $\{X_k\}_{k\in\mathbb{Z}}$.

More on the extremal index θ

- We always have $0 \leq \theta \leq 1$.
- If $\theta > 0$, then the maxima M_n scale at the same rate "$n^{1/\alpha}$" for dependent and independent data.
- For independent data $\theta = 1$, but not conversely.
- The extremal index is related to the clustering of extremes phenomenon.
- Under general conditions, the exceedence times of $\{X_k\}_{k\in\mathbb{Z}}$ converge to a cluster Poisson point process:

$$\theta = \frac{1}{\text{(expected cluster size)}}.$$

- The estimation of α and θ is important in many applications.

Max–spectrum

Max–spectrum: Definition

Data: $X_i,\ 1 \le i \le n$ (Think: stationary time series $\{X_k\}_{k\in\mathbb{Z}}$)
Dyadic Block maxima: Define

$$C_{j,k} = \bigvee_{1\le i\le 2^j} X_{2^j(k-1)+i}.$$

- Illustration:

$$\underbrace{X_1,\ X_2}_{},\ \underbrace{X_3,\ X_4}_{},\cdots$$
$$\underbrace{C_{1,1},\qquad C_{1,2}}_{C_{2,1}},\ \cdots$$

- Introduce the statistics:

$$Y_j := \frac{1}{n_j}\sum_{k=1}^{n_j}\log_2 C_{j,k}, \quad \text{where } n_j = [n/2^j].$$

Max–spectrum: properties

- the set of statistics:

$$Y_j := \frac{1}{n_j}\sum_{k=1}^{n_j} \log_2 C_{j,k}, \quad \text{where } n_j = [n/2^j], \quad \text{and } 1 \le j \le \log_2 n$$

is said to be the max–spectrum of the data $X_1, \ldots, X_n$.
The scaling of maxima implies:

$$Y_j \overset{P}{\sim} \frac{1}{\alpha} j + \frac{1}{\alpha} \log_2 \theta + \textbf{const}, \qquad (1)$$

as both j and $n \to \infty$.

- the max–spectrum is asymptotically linear even for dependent data, as long as the extremal index θ is positive.
- The tail index α can be readily estimated from the Y_j's.
- With more work, one can also get θ. Note that the **const** in (1) is unknown.

Estimating α

- Consider the max–spectrum $Y_j,\ 1 \le j \le J$ of a heavy–tailed time series.

◦ Define the regression–based estimator

$$\widehat{\alpha}(j_1, j_2) := \Big(\sum_{j_1 \le j \le j_2} w_j Y_j \Big)^{-1},$$

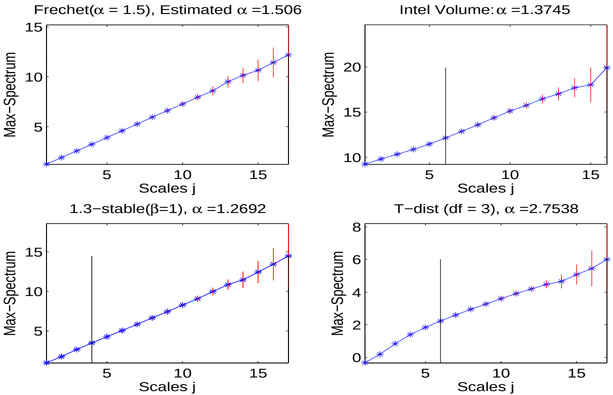
where $\sum_{j_1 \le j \le j_2} w_j = 0$ and $\sum_{j_1 \le j \le j_2} j w_j = 1$.

- Since $1/\alpha$ is the slope of the max-spectrum.

◦ Relation (1) implies the consistency of $\widehat{\alpha}(j_1, j_2)$, provided that j_1, j_2 and n tend to ∞.

◦ More precise details to come...

Max–spectra: an illustration



Estimating θ

For non–iid time series data, we have:

$$Y_j \stackrel{P}{\sim} \frac{1}{\alpha} j + \frac{1}{\alpha} \log_2 \theta + \mathbf{const}, \quad \text{as } j \text{ and } n \to \infty. \qquad (2)$$

How to get rid of the **const**?

- Use resampling, i.e. consider

$$X_1^*, X_2^*, \cdots, X_n^*, \quad \text{which is}$$

either a random permutation or a bootstrap sample of the data.
For iid X_i^*'s with the same marginal as X_i's, the max–spectrum is:

$$Y_j^* \stackrel{P}{\sim} \frac{1}{\alpha} j + \frac{1}{\alpha} + \mathbf{const}, \quad \text{as } j \text{ and } n \to \infty. \qquad (3)$$

- The **const** in (2) and (3) are equal and θ in (3) is gone!

Estimating θ (cont'd)

On scale j, the point estimate of θ is:

$$\widehat{\theta}(j) = 2^{\widehat{\alpha}\times(Y_j^* - Y_j)}, \quad \text{for some } \widehat{\alpha}.$$

• But one can also resample multiple times to get a sample $\widehat{\theta}_i(j),\ i = 1, \dots, m$.

What scale j to choose? Ans: $j = j(n) \to \infty$ should grow with n.

More difficult questions:

◦ What is the optimal rate of $j(n)$?

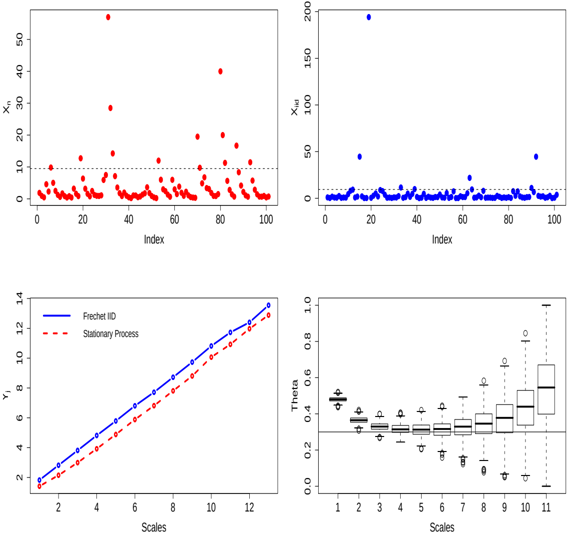
◦ What is the role of $m = m(n)$?

◦ What is the asymptotic distribution of $\widehat{\alpha}(j_1, j_2)$?

◦ What is the asymptotic distribution of $\widehat{\theta}(j)$?

Some answers to come...

Estimating θ: an illustration



Some Asymptotic Results

Asymptotic results: for iid X_i's

One can show the asymptotic normality for $\widehat{\alpha}(j_1, j_2)$.
The key condition is on the second order asymptotics of the tails:

$$|x^{\alpha}\mathbb{P}\{X_i > x\} - C| \le Dx^{-\beta}, \quad \text{for large } x, \tag{4}$$

and some $\beta > 0$.
The asymptotic regime:

- Let $(j_1, j_2) = (1, k) + r(n)$, for fixed k.
- Consider the max–spectrum vector

$$\vec{Y}_r = (Y_{r+j})_{j=1}^{k}, \quad \text{for the range of scales } (j_1(n), j_2(n)).$$

- The next result is a uniform CLT for $\vec{Y}_r$.

CLT for the max–spectrum: iid case

Theorem(S., Michailidis & Taqqu) *Under (4) and another mild condition, for any* $\vec{v} \in \mathbb{R}^k$:

$$\sup_{x\in\mathbb{R}} \Big| \mathbb{P}\{\sqrt{n/2^r}\Big((\vec{v},\vec{Y}_r) - (\vec{v},\vec{\mu}_r)\Big) \le x\} - \Phi(x/\sigma_{\vec{v}})\Big| \le C_{\vec{v}}\Big(1/2^{r\beta/\alpha} + r2^{r/2}/\sqrt{n}\Big), \qquad (5)$$

where Φ *is the standard Normal c.d.f.*

- For the mean $\vec{\mu}_r = (\mu_r(j))_{j=1}^k$ and the variance, we have:

$$\mu_r(j) = (r+j)/\alpha + \mathbf{const}, \quad \text{and} \quad \sigma_{\vec{v}}^2 = (\vec{v}, \Sigma_\alpha \vec{v}).$$

- The asymptotic covariance matrix Σ_α can be identified!
- The proof involves Berry–Esseen and sharp rates for $\mathbb{E}f(n^{-1/\alpha}\max_{1\le i\le n} X_i)$ with $f(\cdot) = \log_2(\cdot)$.
- The uniform CLT is more informative than the usual CLT!

Dependent data

- We have asymptotic normality results for $\widehat{\theta}$ and $\widehat{\alpha}$ for *m*−dependent time series.
- The general case is under investigation.
- We have results on:
  - confidence intervals
  - automatic selection of scales (j_1, j_2).
  - diagnostics

Asymptotics for $\widehat{\alpha}$: Two regimes

- **Intermediate scales:** Fix $j_1 < j_2$ integer and let

$$\widehat{\alpha}_n = \widehat{\alpha}[r(n)+j_1, r(n)+j_2], \quad \text{where} \quad r(n) \to \infty \text{ and } 2^{r(n)}/n \to 0, \quad \text{as } n \to \infty.$$

◦ We expect to get consistency and asymptotic normality for $\widehat{\alpha}_n$.

- **Large scales:** Fix $\ell \in \mathbb{N}$ and focus on the largest $\ell + 1$ scales:

$$\widehat{\alpha}_n = \widehat{\alpha}[\log_2 n - \ell, \log_2 n].$$

◦ We can only get "distributional consistency":

$$\widehat{\alpha}_n \xrightarrow{d} \alpha_Z, \quad \text{as } n \to \infty,$$

with α_Z a random variable.

- Both regimes are useful/interesting in practice.
- More details ...

Intermediate scales asymptotics

The regularity conditions: for $M_n := \max_{1\le k\le n} X_k$

$$\mathbb{P}\{n^{-1/\alpha}M_n \le x\} = \exp\{-c(n,x)x^{-\alpha}\}, \quad x>0, \quad \text{where}$$

$$|c(n,x) - c_X| \le c_1(x)n^{-\beta}, \ \forall x>0, \ \text{ with } c_1(x) = \mathcal{O}(x^{-R}), \ x\downarrow 0. \quad (6)$$

(Plus a technicality at $x \approx 0$.)
∘ **Intuition:** β controls the second order tail behavior of M_n.
∘ **Caveat:** Relation (6) may be hard to verify! We have it for moving maxima.
• We get rates on moments of $f(M_n/n^{1/\alpha})$, in particular:
Thm [S. & Michailidis (2006)] Under the above conditions, for all $k\in\mathbb{N}$,

$$\mathbb{E}|\log^k(M_n/n^{1/\alpha}) - \mathbb{E}\log^k(Z)| = \mathcal{O}(n^{-\beta}), \quad \text{as } n\to\infty,$$

provided $\int_1^\infty c_1(x)x^{-\alpha-1+\delta}dx$, for $\delta>0$.

Intermediate scales: asymptotic normality

Let (X_k) be stationary with tail exponent $\alpha > 0$.
Thm [S. & Michailidis (2006)] *Under the above conditions, and if* (X_k) *is m–dependent, we have*

$$\sqrt{n_{r(n)}}(\widehat{\alpha}_n - \alpha) \stackrel{d}{\longrightarrow} \mathcal{N}(0, \alpha^2 c_w),$$

where $c_w = \vec{w}^t \Sigma_1 \vec{w}$, *and* $\widehat{\alpha}_n = \widehat{\alpha}[r(n) + j_1, r(n) + j_2]$, *provided*

$$2^{r(n)}/n + n/2^{r(n)(1+2\min\{1,\beta\})} \longrightarrow 0, \quad as \quad n \to \infty.$$

Remarks:

- The same asymptotic variance as in the iid case.
- ◦ **Intuition:** The block–maxima $D(j,k)$, $1 \le k \le n_j$ – asymptotically iid!
- β captures: second order tails PLUS dependence.
- Asymptotic confidence intervals available!
- Optimal linear GLS estimators available!

Large scales: distributional consistency

The **regularity conditions** and **m–dependence** are restrictive.

• As in Davis & Resnick (1985), let

$$X_k = \sum_{i=0}^{\infty} c_i \xi_{k-i}, \quad \text{where } \sum_i |c_i|^\delta < \infty, \quad 0 < \delta < \min\{1, \alpha\}.$$

∘ Here (ξ_k) are iid and $\mathbb{P}\{|\xi_1| > x\} \sim Cx^{-\alpha}$, $x \to \infty$, with $\mathbb{P}\{\xi_1 > x\}/\mathbb{P}\{|\xi_1| > x\} \to p \in [0,1]$, as $x \to \infty$.

Lemma *For* $X_k(m) := \max_{1 \le i \le m} X_{m(k-1)+i}$, $k = 1, 2, \ldots,$ *we get*

$$\{m^{-1/\alpha} X_k(m)\}_{k \in \mathbb{N}} \stackrel{fdd}{\longrightarrow} \{Z_k\}_{k \in \mathbb{N}}, \quad \text{as } m \to \infty,$$

where (Z_k) *are iid* $\alpha-$*Fréchet. Provided* $p \max_i c_i > 0$ *or* $(1-p)\max_i(-c_i) > 0$.

• This justifies the "asymptotic independence phenomenon" for the block–maxima $(D(j,k))_k$ as $j \to \infty$!

Thm [S. & Michailidis (2006)] *Under the above conditions, with fixed* ℓ

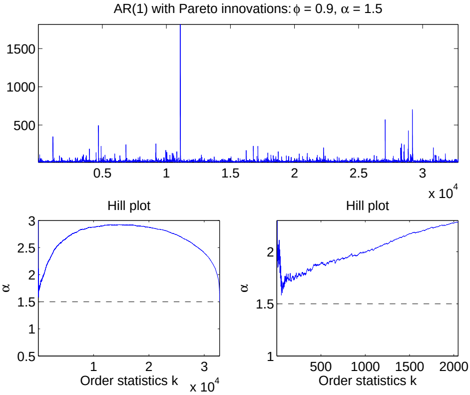
$$\widehat{\alpha}_n \stackrel{d}{\longrightarrow} \widehat{\alpha}_{Z,\ell}, \quad \text{as } n \to \infty,$$

where $\widehat{\alpha}_n = \widehat{\alpha}[$ *top–*ℓ *scales*$]$ *and* $\widehat{\alpha}_Z$ *is based on iid* $\alpha-$*Fréchet data*

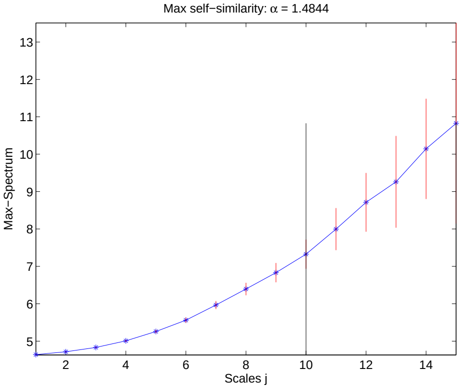
Distributional consistency: implications

- No consistency but confidence intervals!
- Covers more processes!
- The approximation is often valid for "small" n.

AR(1) with Pareto ($\alpha = 1.5$) innovations

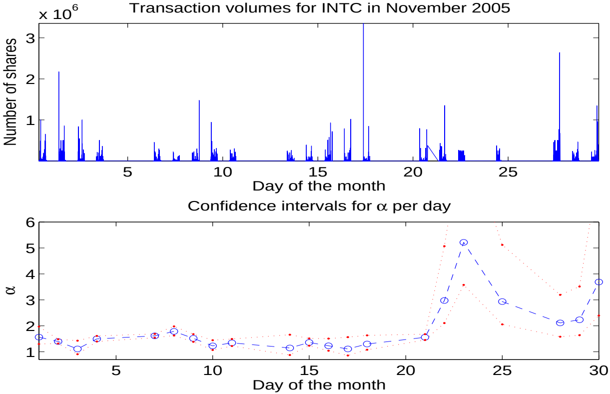


The max–spectrum ...

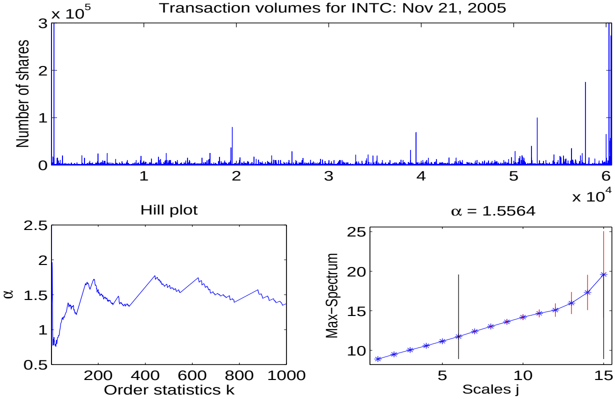


Applications

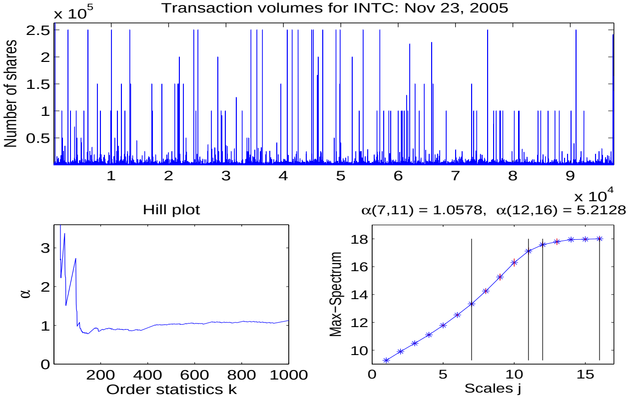
Intel Stock: traded volumes November 2005



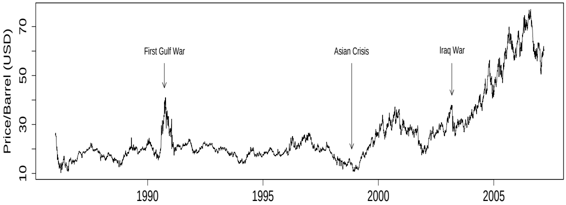
Intel Stock: traded volumes, November 21, 2005

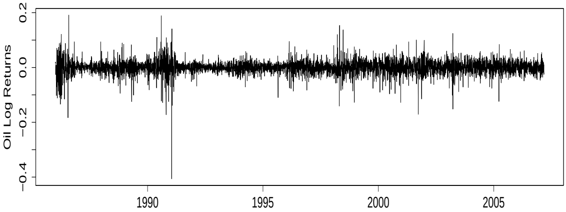


Intel Stock: traded volumes, November 23, 2005

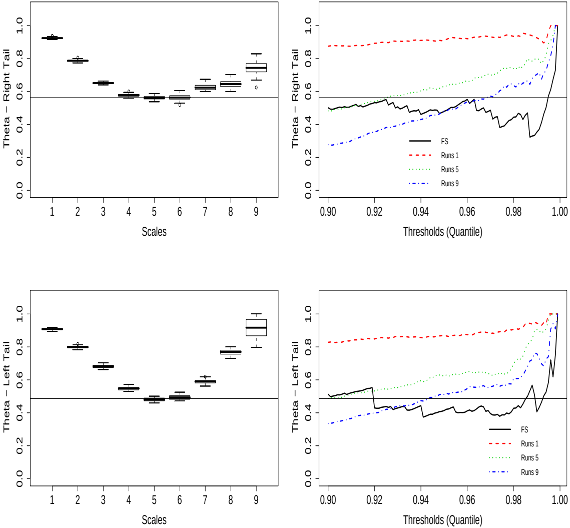


Oil data


Price/Barrel (USD)
10
30
50
70
First Gulf War
Asian Crisis
Iraq War
1990
1995
2000
2005


Oil Log Returns
−0.4
−0.2
0.0
0.2
1990
1995
2000
2005

Oil returns: extremal index estimates



References

Davis, R. A. and Resnick, S.I. (1985) Limit theory for moving averages of random variables with regularly varying tail probabilities. *The Annals of Probability 13*(1), 179–195.

Hall, P. (1982) On some simple estimates of an exponent of regular variation, *J. Roy. Stat. Assoc. (Ser B)*, *44*, 37–42.

Hill, B. M. (1975) A simple general approach to inference about the tail of a distribution. *The Annals of Statistics 3*, 1163–1174.

Leadbetter, M. R. and G. Lindgren, and H. Rootzén (1983) *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag.

Resnick, S. and Stărică, C. (1995) Consistency of Hill's estimator for dependent data. *Journal of Applied Probability 32*, 139–167.

Resnick, S. (1987) *Extreme Values, Regular Variation and Point Processes*, Springer-Verlag.

Stoev, S. and Michailidis, G. (2006) On the estimation of the heavy–tail exponent in time series using the max–spectrum, *Technical Report, University of Michigan.*

Stoev, S., Michailidis, G., and Taqqu, M.S. (2006) Estimating heavy–tail exponents through max self–similarity, *Technical Report, University of Michigan.*

WRDS `https://wrds.wharton.upenn.edu/`. *Wharton School of Management, Universty of Pennsylvania.*