On the estimation of the heavy–tail exponent in time series using the max–spectrum

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

ISBIS, Prague, July 2008

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.

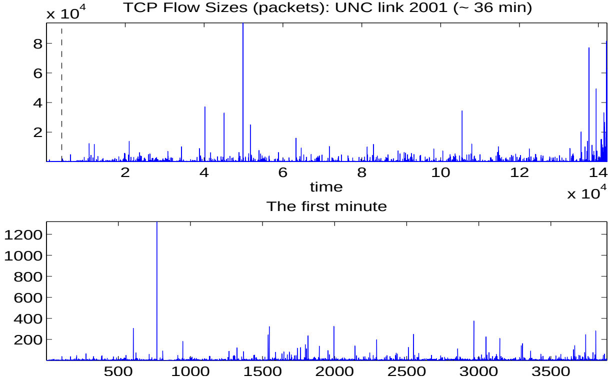
- For simplicity, here we focus on positive X with $L(x) \equiv \mathrm{const}$.
- Thus, for $p > 0$, the moment:

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

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

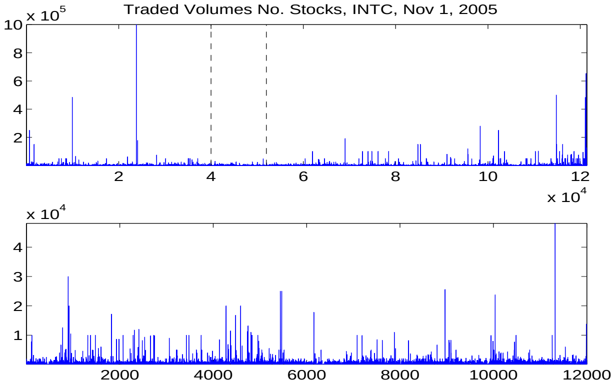
Network traffic data

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



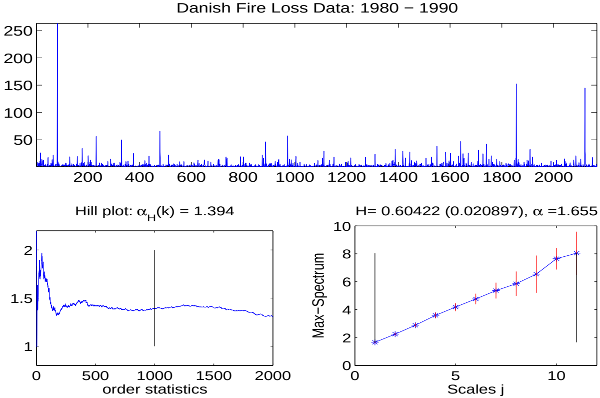
Traded volume data

- Traded volume per individual transactions (in # shares)



Insurance data

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



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 $\alpha-$Fréchet distribution:

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

- 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?

The seminal results of Leadbetter (see e.g. Leadbetter, Lindgren & Rootzén (1983)) imply that under mild dependence conditions "$D(u_n)$":

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

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

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

- 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 $0 \le \theta \le 1$.
- 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: 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:

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

- 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 $X_1, \dots, X_n$.
The scaling of maxima implies:

$$Y_j \overset{P}{\sim} \frac{1}{\alpha} j + \frac{1}{\alpha} \log_2 \theta + \mathbf{const}, \tag{1}$$

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

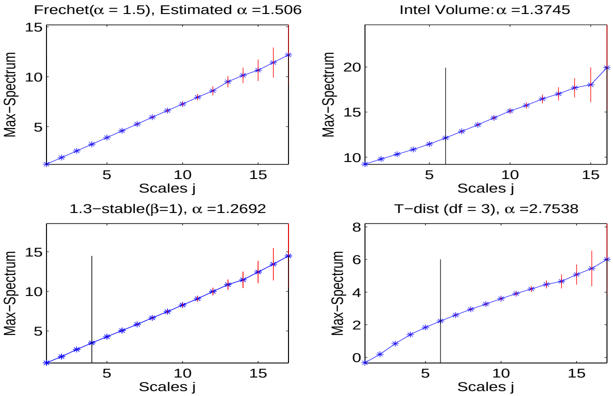
Consider the max–spectrum $Y_j,\ 1 \le j \le J$ of a heavy–tailed time series and define:

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

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

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

Max–spectra: an illustration



Estimating θ

For non–iid time series data, we have:

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

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:

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

- 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 $\widehat{\theta}_i(j),\ \ i = 1, \ldots, m.$

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

More difficult questions:

◦ What is the optimal rate of $j(n)$?

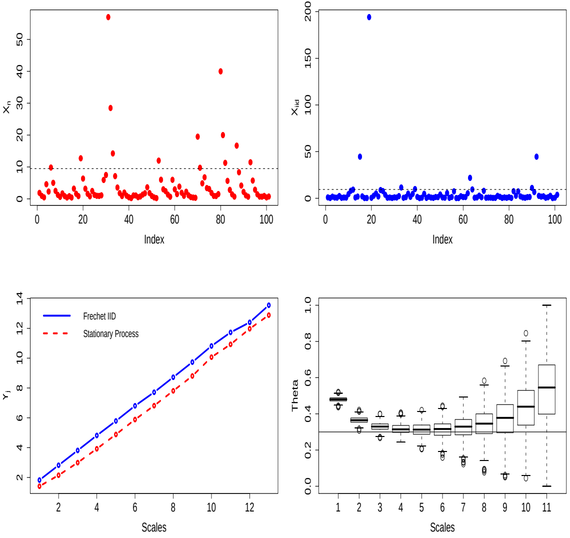
◦ What is the role of $m = m(n)$?

◦ 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



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* $\vec{v} \in \mathbb{R}^k$*:*

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

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:

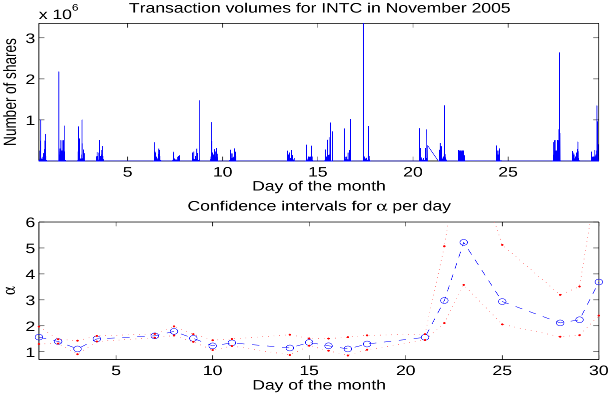
$$\mu_r(j) = (r+j)/\alpha + \mathbf{const}, \quad \text{and} \quad \sigma_{\vec{v}}^2 = (\vec{v}, \Sigma_\alpha \vec{v}).$$

- 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!

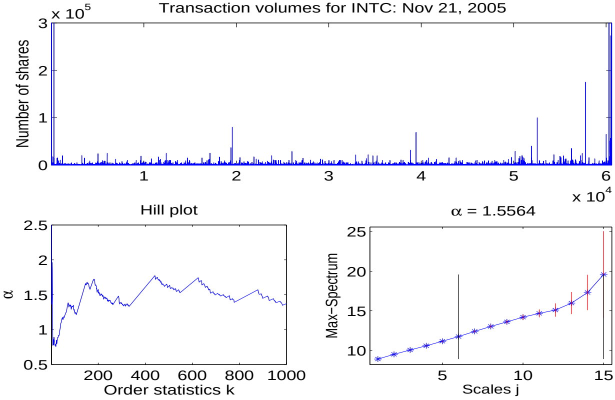
Dependent data

- 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

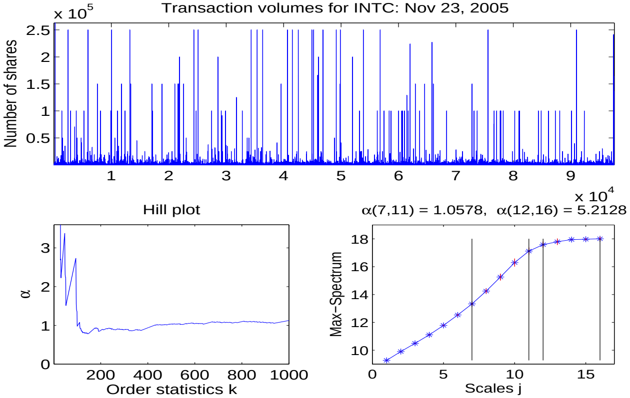
Intel Stock: traded volumes November 2005



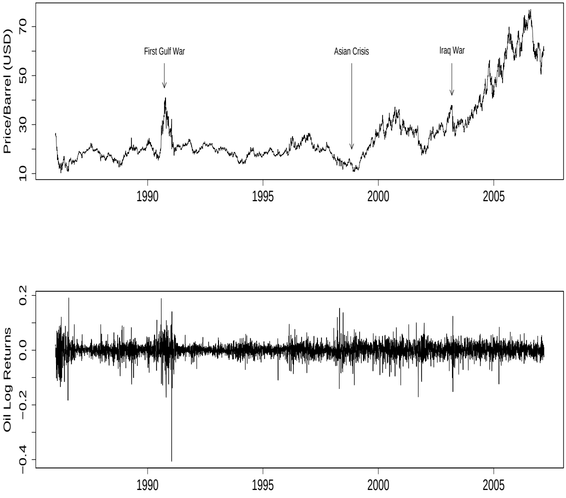
Intel Stock: traded volumes, November 21, 2005



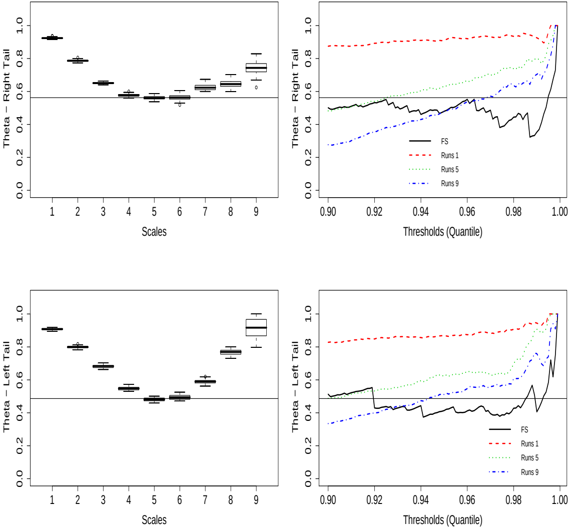
Intel Stock: traded volumes, November 23, 2005



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

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

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