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

Stilian A. Stoev (sstoev@umich.edu)

University of Michigan, Ann Arbor, U.S.A.

JSM, Salt Lake City, 2007

joint work with: George Michailidis (gmichail@umich.edu) *and Murad Taqqu* (murad@math.bu.edu)

Outline

- Heavy tails are ubiquitous
- An old problem
- Max–spectrum
- The estimator
- Asymptotic properties
- Data examples

Heavy tails

• A random variable X is said to be *heavy–tailed* if

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

for some $\alpha > 0$ and a slowly varying function L.

∘ Here we focus on the simpler but important context:

$$X \geq 0, \ \text{ a.s.} \quad \text{and} \quad \mathbb{P}\{X > x\} \sim Cx^{-\alpha}, \quad \text{as } x \to \infty.$$

∘ X (*infinite moments*) For $p > 0$,

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

In particular,

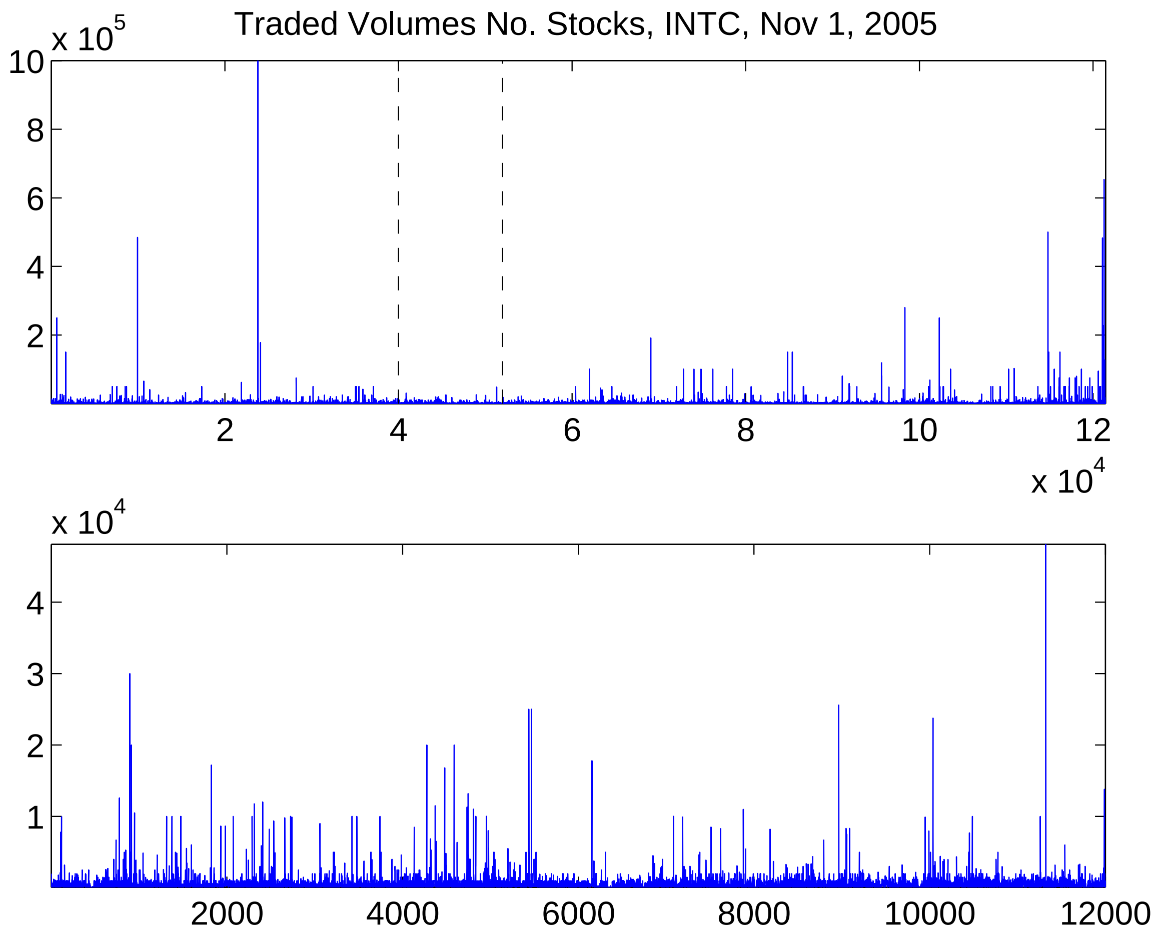
$$0 < \alpha \leq 2 \quad \Rightarrow \quad \mathrm{Var}(X) = \infty$$

and

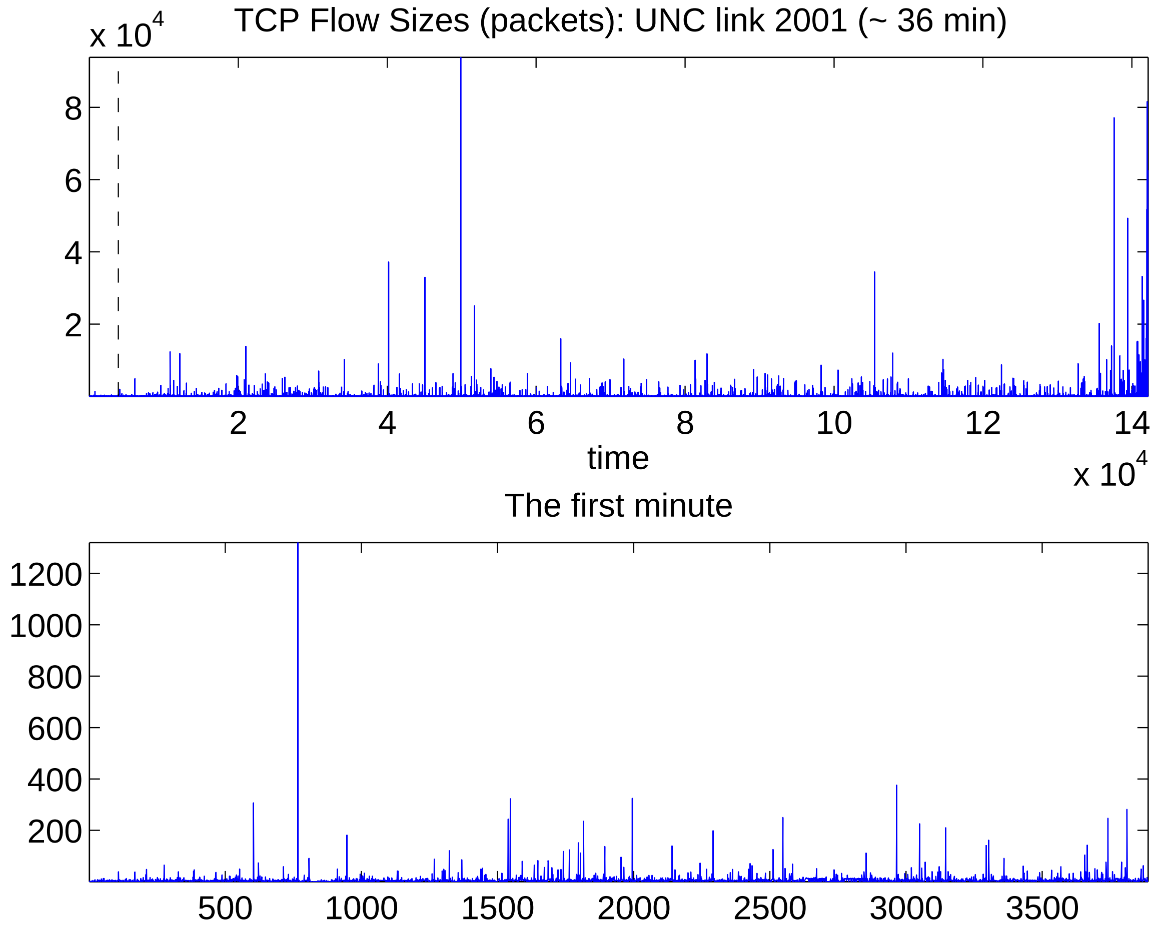
$$0 < \alpha \leq 1 \quad \Rightarrow \quad \mathbb{E}|X| = \infty.$$

• The estimation of the *heavy–tail exponent* α is an important problem with rich history.

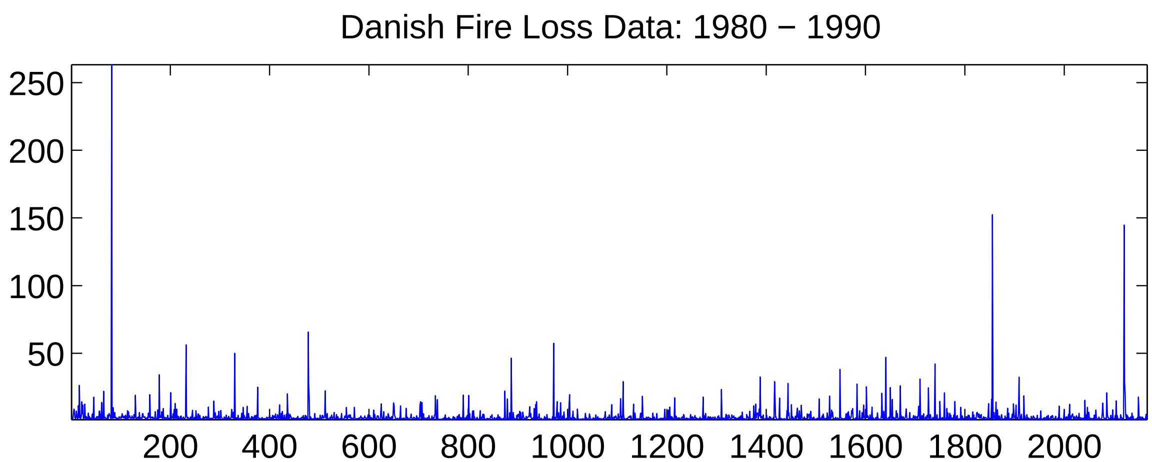
Heavy tails everywhere: Traded volumes

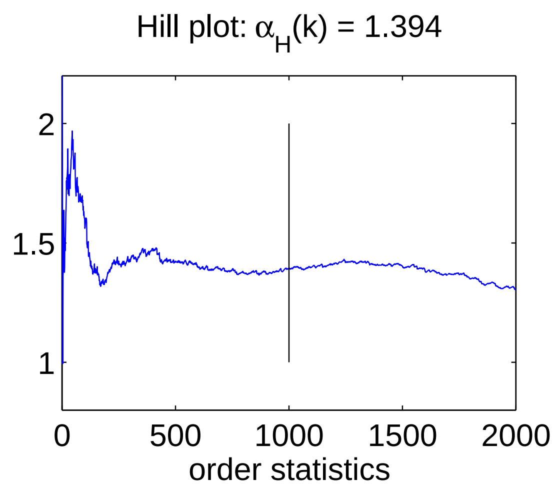


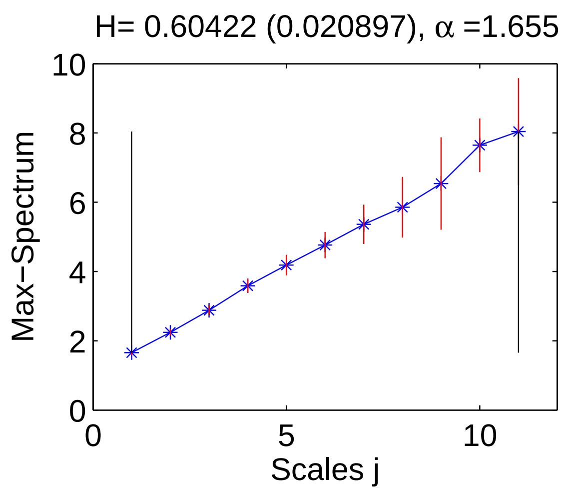
Heavy tails everywhere: TCP durations



Heavy tails everywhere: Insurance claims







Tail exponent estimation: an old problem

• Hill (1975) – the MLE in the Pareto model $\mathbb{P}\{X > x\} = x^{-\alpha},\ x \geq 1$ and introduced the *Hill plot*:

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

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.

◦ J. 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"*

Another approach: max self–similarity

• For iid (X_k) with tail exponent α

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

where $\mathbb{P}\{Z \le x\} = \exp\{-Cx^{-\alpha}\},\ x > 0$.

∘ The above continues to hold for many *dependent* stationary (X_k)!

• Given $X_1, \dots, X_n$, set

$$D(j,k) := \bigvee_{i=1}^{2^j} X_{2^j(k-1)+i}, \quad 1 \le k \le n_j := [n/2^j],\ 1 \le j \le \log_2(n).$$

to be *block–maxima* of dyadic sizes.

∘ Observe that

$$Y_j := \frac{1}{n_j} \sum_{k=1}^{n_j} \log_2 D(j,k) \simeq \mathbb{E} \log_2 2^{j/\alpha} Z = j/\alpha + \mathbb{E} \log_2 Z, \quad \text{as } j \to \infty.$$

The max–spectrum: iid asymptotics

The Y_j's, $1 \le j \le \log_2 n$ is the max–spectrum of the data set $(X_k,\ 1 \le k \le n)$.

• An estimator of α is then derived from Y_j via regression:

$$\widehat{\alpha} = \widehat{\alpha}[j_1, j_2] := \sum_{j=j_1}^{j_2} w_j Y_j, \quad \text{with } \sum_j w_j = 0,\ \sum_j j w_j = 1.$$

• For iid data: The estimator $\widehat{\alpha}[j_1, j_2]$ is consistent and asymptotically normal, as $j_1, j_2 \to \infty$ but so that $n/2^{j_1}, n/2^{j_2} \to \infty$.

Thm [S., Michailidis & Taqqu (2006)] *For iid data under second order tail regularity conditions. Let* $1 \le r(n) \le \log_2 n$ *be such that*

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

then

$$\sup_{x \in \mathbb{R}} |\mathbb{P}\{\sqrt{n_{j_2+r(n)}}(\langle \vec{\theta}, \vec{Y} \rangle - \langle \vec{\theta}, \vec{\mu_r} \rangle) \le x\} - \Phi(x/\sigma_{\vec{\theta}})| \longrightarrow 0, \quad n \to \infty.$$

The max–spectrum: iid asymptotics (cont'd)

Here $\vec{Y} = (Y_{j+r(n)})_{j=j_1}^{j_2}$, $\vec{\theta} = (\theta_j)_{j=j_1}^{j_2}$, *and*

$$\vec{\mu}_r = ((j + r(n))/\alpha + C,\ j_1 \le j \le j_2), \quad \textit{and} \quad \sigma^2_{\vec{\theta}} = \alpha^{-2}\vec{\theta}^t\Sigma_1\vec{\theta}.$$

Remarks:

• The $\beta > 0$ governs the "second order" tail behavior. Roughly:

$$\mathbb{P}\{X > x\} \sim Cx^{-\alpha}(1 + Dx^{-\beta}), \quad \text{as } x \to \infty.$$

• The asymptotic cov matrix Σ_1 is the same as for 1−Fréchet data.

◦ It does not depend on α and $C = \mathbb{E}\log_2 Z$.

• Consistency and asymptotic normality for $\widehat{\alpha}[r(n) + j_1, r(n) + j_2]$ follow.

◦ The rates are the same as for the Hill estimator − Hall (1982).

• The explicit asymptotic cov $\alpha^{-2}\Sigma_1$ of the max–spectrum $\vec{Y}$ yields the optimal linear GLS estimators − important in practice.

The max–spectrum: dependent data

Let $(X_k)_{k\in\mathbb{Z}}$ be stationary, with tail exponent α and extremal index $\theta > 0$.

- Then,

$$\frac{1}{n^{1/\alpha}} \bigvee_{1\le k\le n} X_k \stackrel{d}{\to} \theta^{1/\alpha} Z \quad \text{where} \quad \frac{1}{n^{1/\alpha}} \bigvee_{1\le k\le n} X_k^* \stackrel{d}{\to} Z, \quad (n\to\infty)$$

where (X_k^*) are iid copies of X_1.

◦ Since $\theta > 0$, the max–spectrum (Y_j) for time series scales as for iid data:

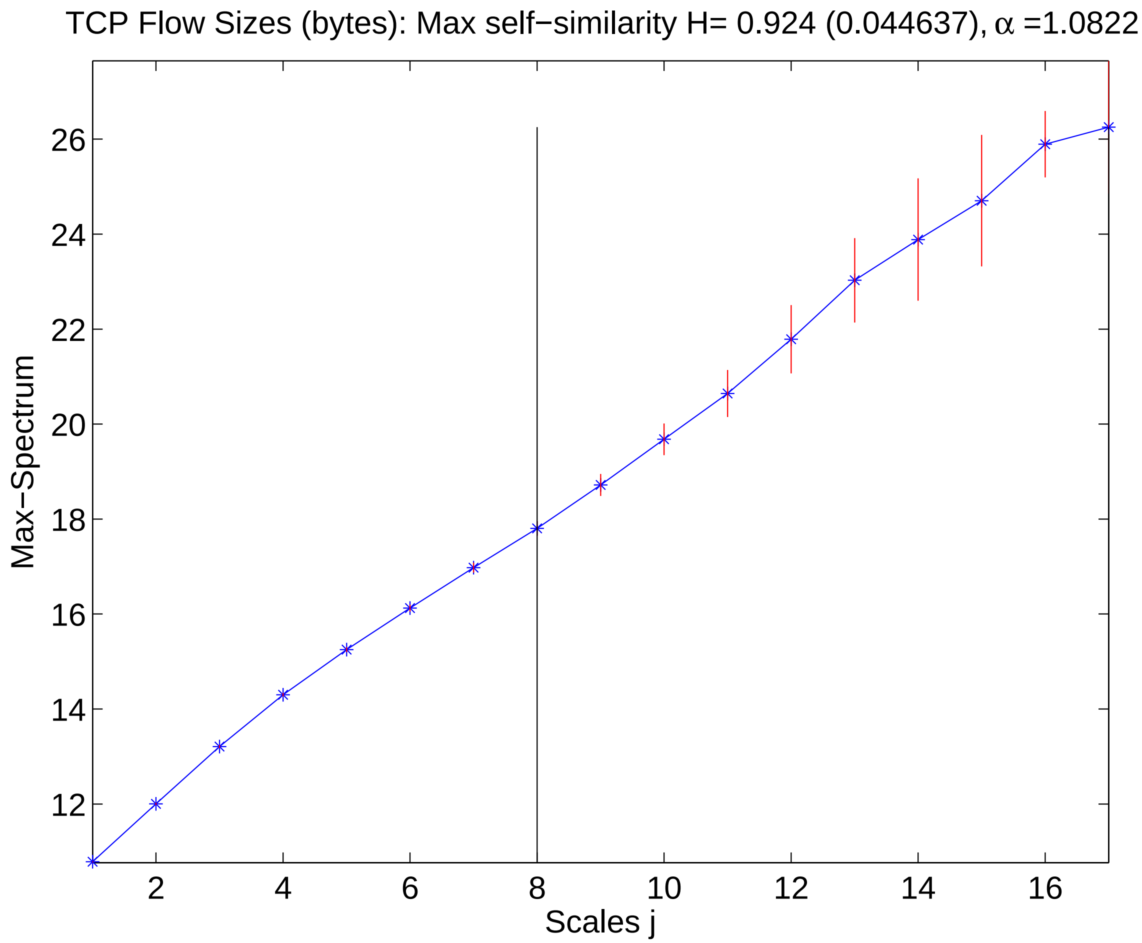
$$Y_j \simeq j/\alpha + C, \qquad \text{as } j\to\infty \;\text{ and } n_j = n/2^j \to \infty.$$

- The same, regression–based, estimators $\widehat{\alpha} = \sum_{j=j_1}^{j_2} w_j Y_j$ work!

- The asymptotics for $\widehat{\alpha}$ are harder (than for iid data)!

◦ Intuition: the block–maxima $D(j,k), 1 \le k \le n_j$ are asymptotically iid, as $j \to \infty$.

Max–spectrum illustration: TCP durations



Two asymptotic regimes

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

$\widehat{\alpha}_n = \widehat{\alpha}[r(n)+j_1, r(n)+j_2]$, where $r(n) \to \infty$ and $2^{r(n)}/n \to 0$, 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 $M_n := \max_{1\le k\le n} X_k$

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

$$|c(n,x) - c_X| \le c_1(x) n^{-\beta}, \ \forall x > 0, \ \ \text{with } c_1(x) = \mathcal{O}(x^{-R}), \ x \downarrow 0. \qquad (1)$$

(Plus a technicality at $x \approx 0$.)

◦ Intuition: β controls the second order tail behavior of M_n.

◦ Caveat: Relation (1) 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*

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

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*

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

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* α*–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

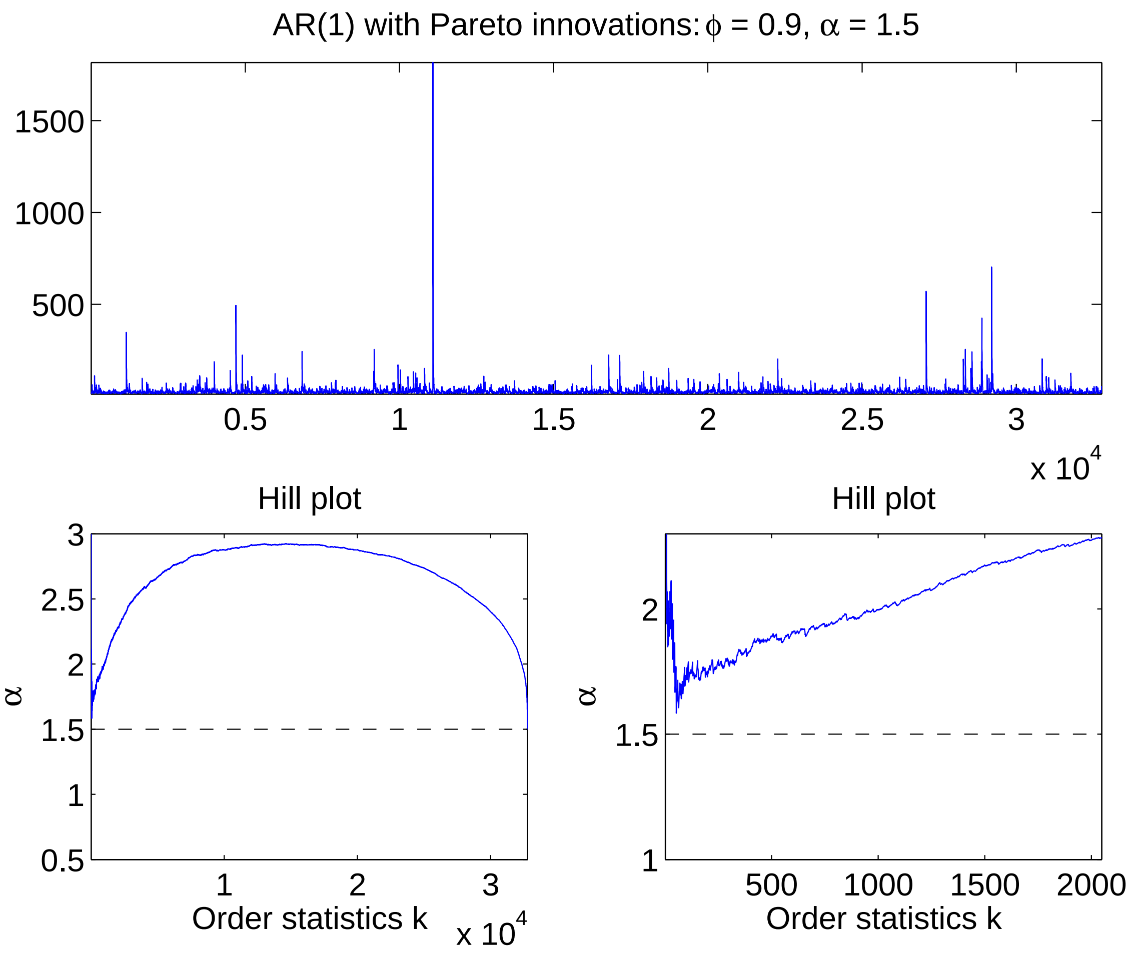
$$\widehat{\alpha}_n \stackrel{d}{\longrightarrow} \widehat{\alpha}_{Z,\ell}, \quad \text{as } n \to \infty,$$

where $\widehat{\alpha}_n = \widehat{\alpha}[$ *top–*ℓ *scales*$]$ *and* $\widehat{\alpha}_Z$ *is based on iid* α*–Fréchet data* $Z_1, \ldots, Z_{2^{\ell+1}}$.

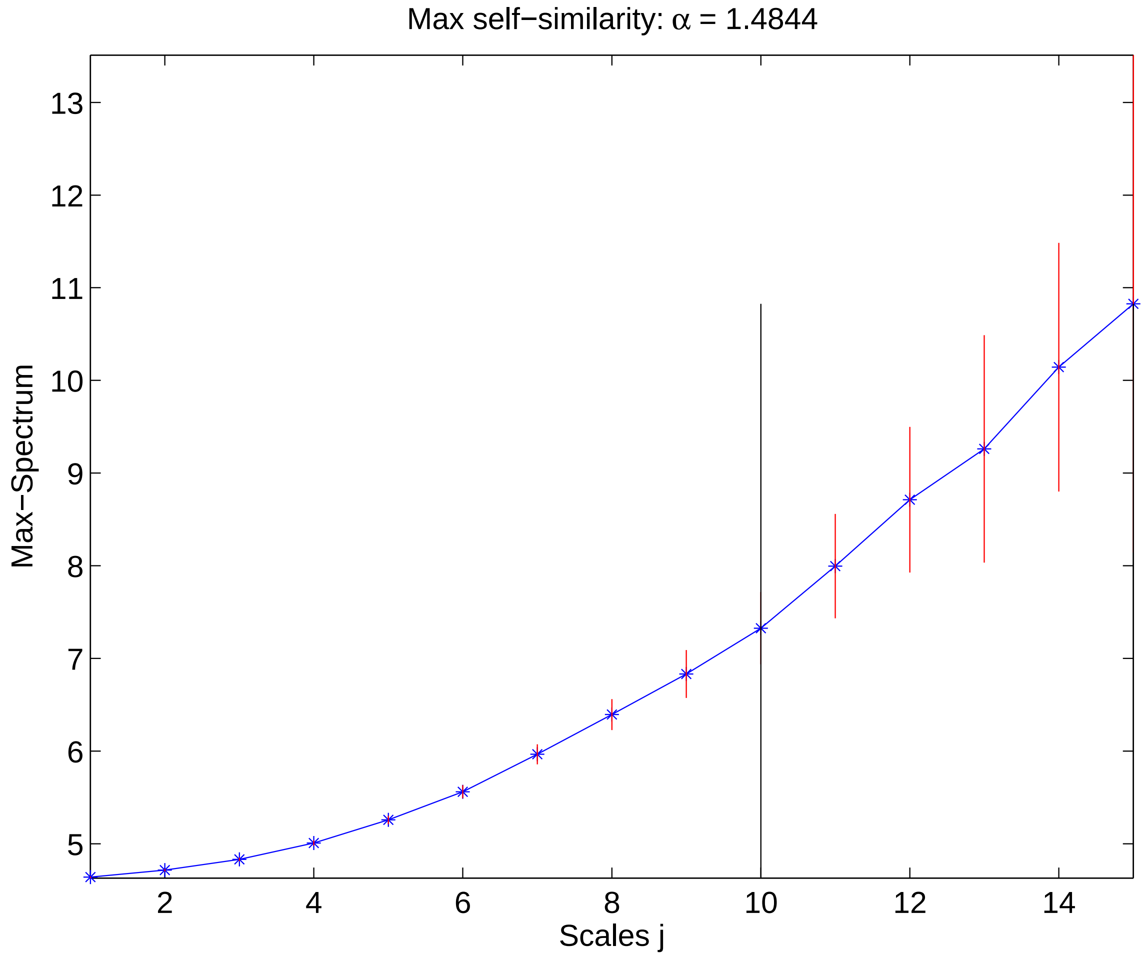
Distributional consistency: implications

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

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

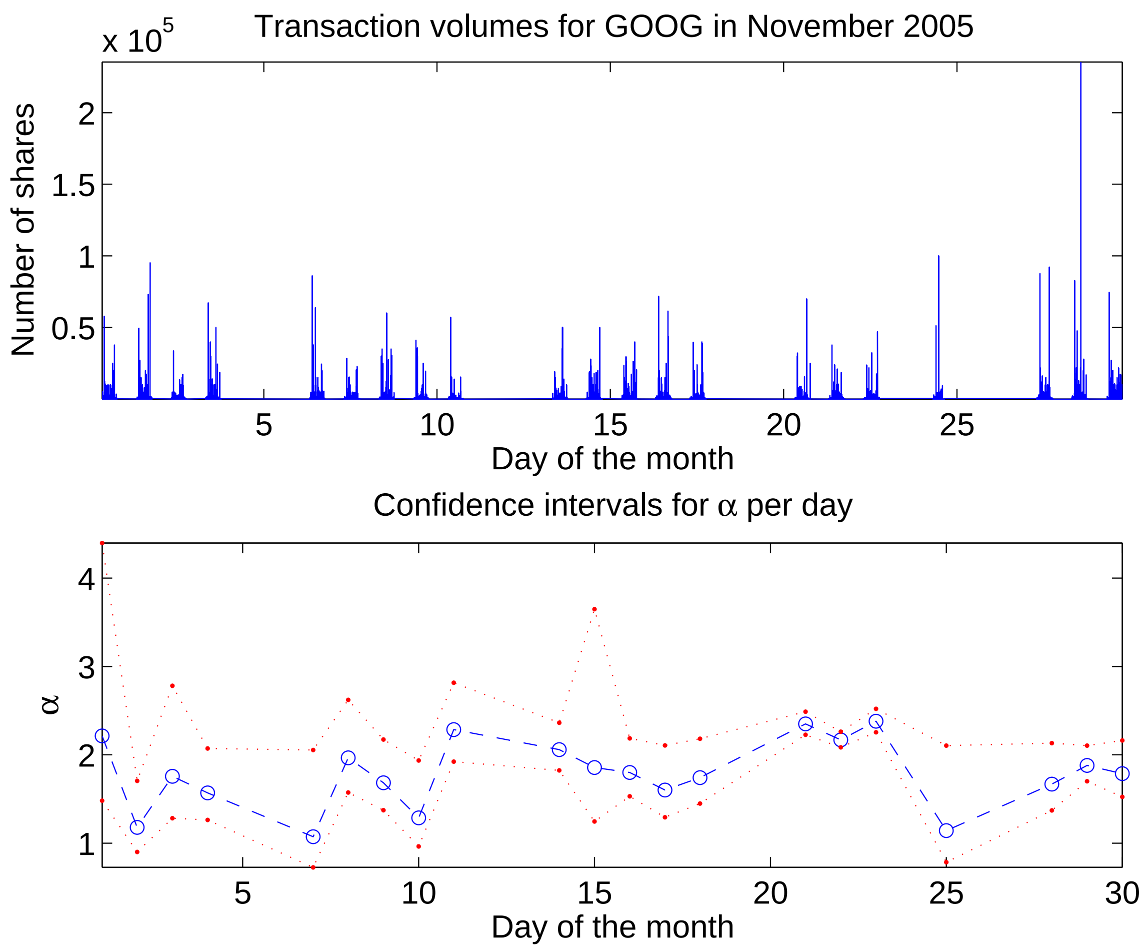


The max–spectrum ...

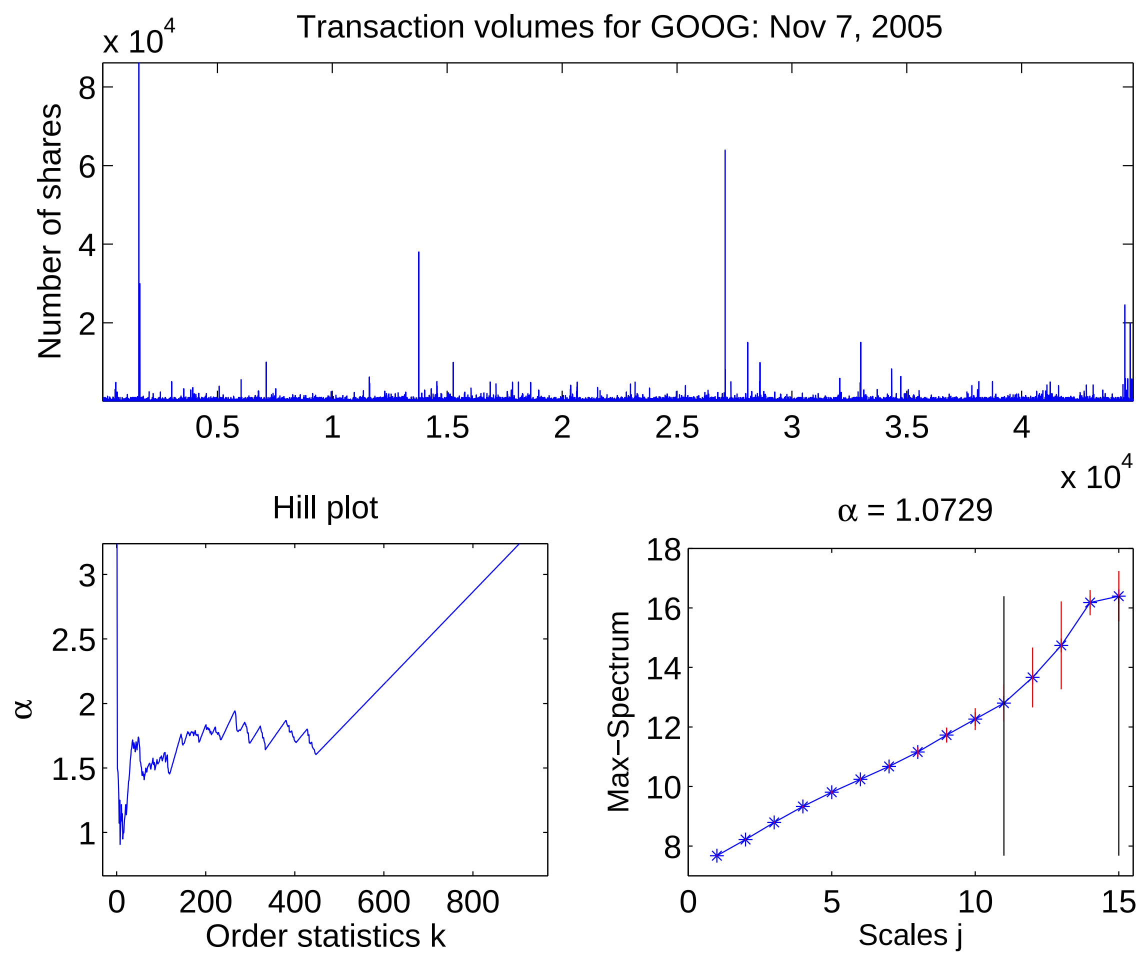


Data examples: the advantage of time scales

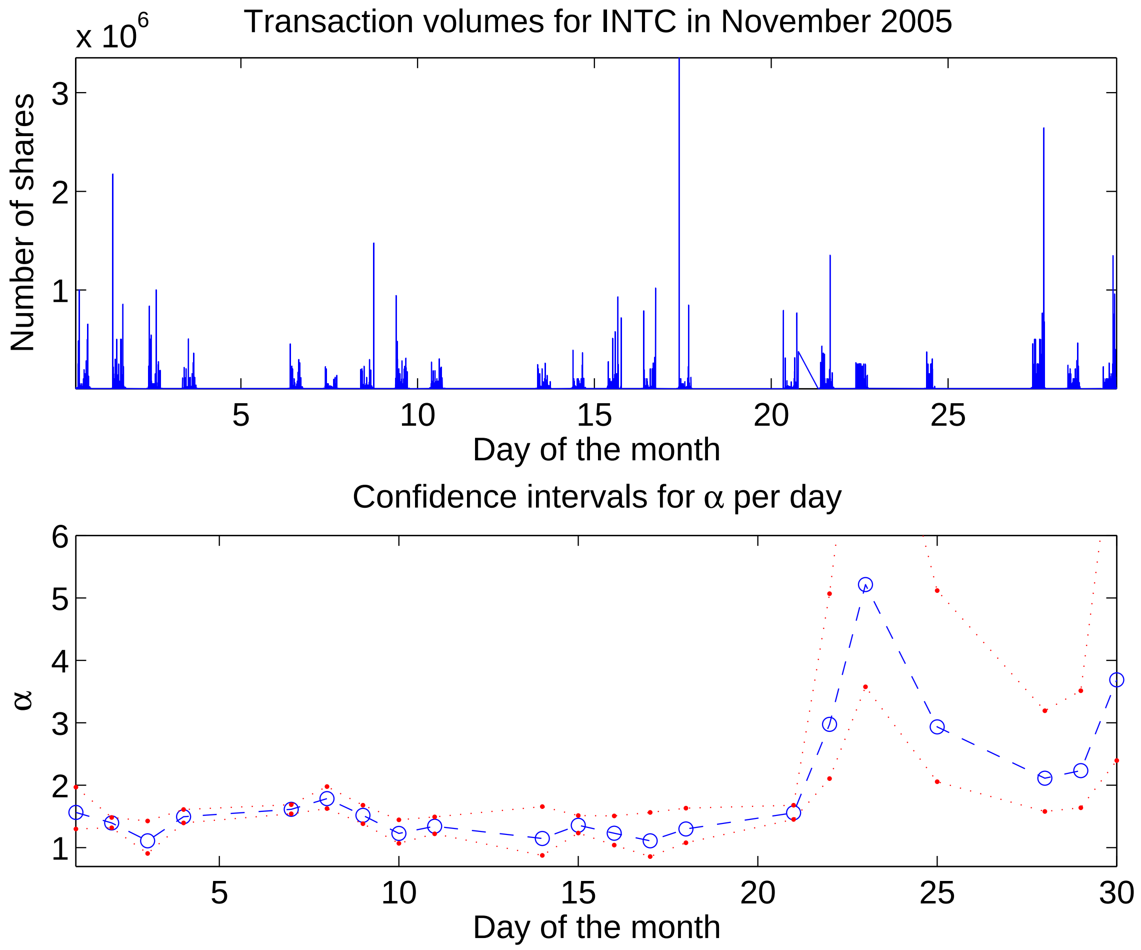
Google: traded volume



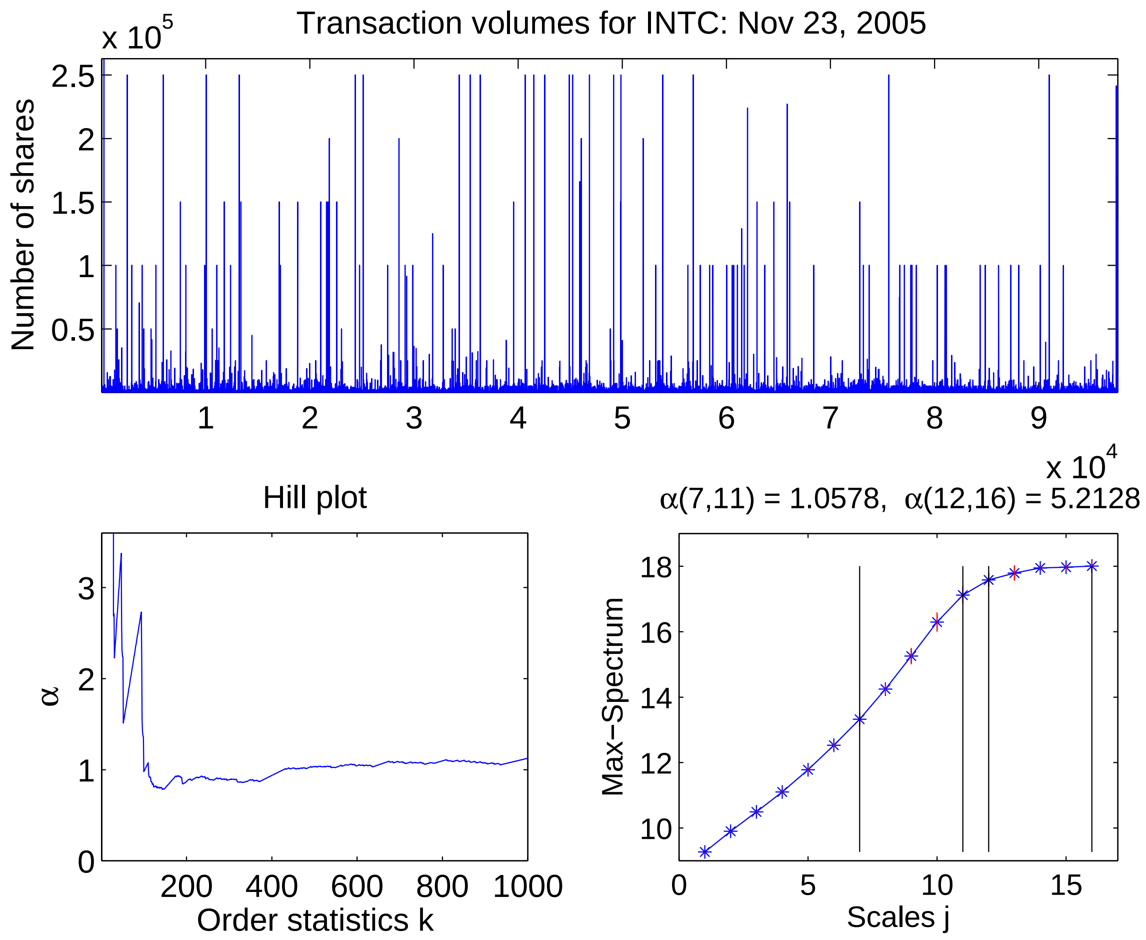
Google: traded volume – the time series



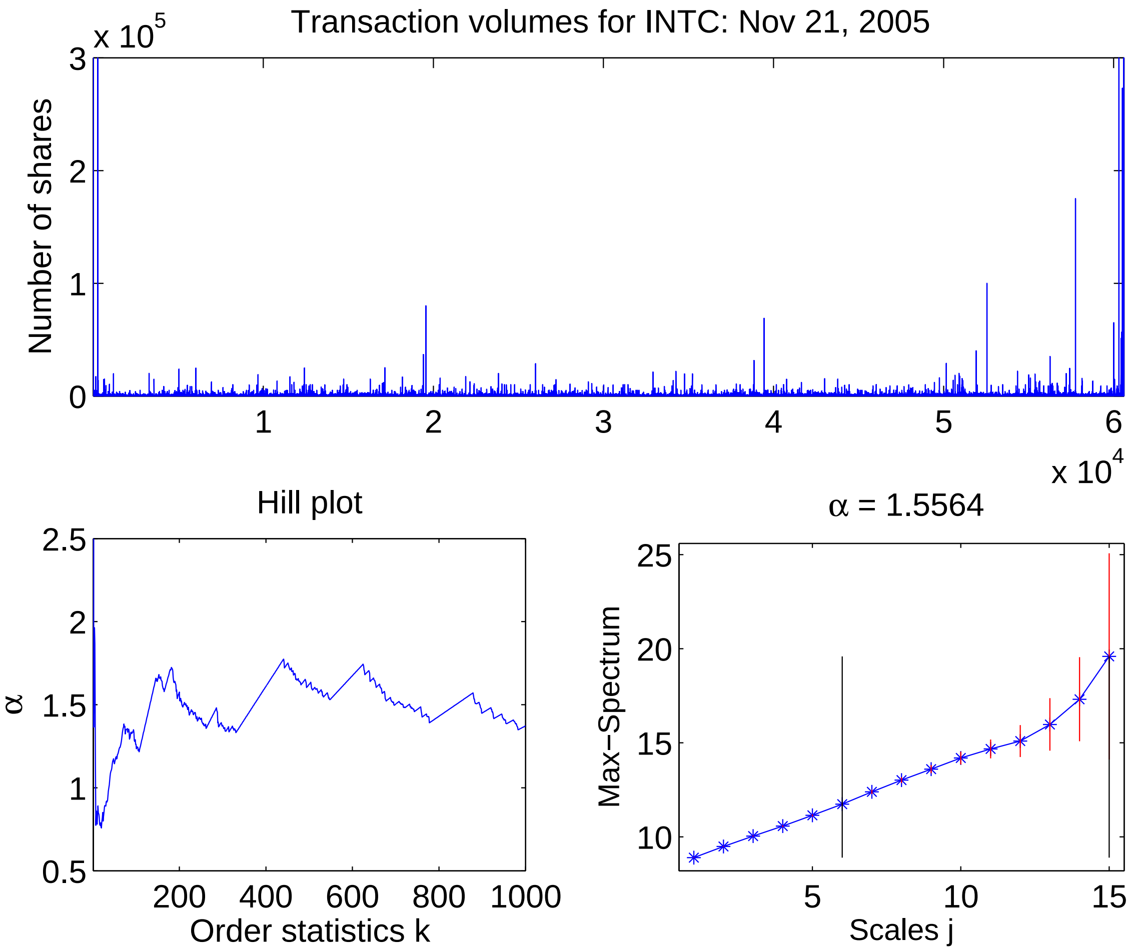
Intel: traded volume



Intel: strange time series



Intel: typical time series



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.

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

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.*