APPLIED STOCHASTIC MODELS IN BUSINESS AND INDUSTRY
Appl. Stochastic Models Bus. Ind. 2010; **26**:224–253
Published online 31 March 2009 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asmb.764

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

Stilian A. Stoev*, † and George Michailidis

Department of Statistics, The University of Michigan, Ann Arbor, U.S.A.

SUMMARY

This paper addresses the problem of estimating the tail index α of distributions with heavy, Pareto-type tails for dependent data, that is of interest in the areas of finance, insurance, environmental monitoring and teletraffic analysis. A novel approach based on the max self-similarity scaling behavior of block maxima is introduced. The method exploits the increasing lack of dependence of maxima over large size blocks, which proves useful for time series data.

We establish the consistency and asymptotic normality of the proposed max-spectrum estimator for a large class of m-dependent time series, in the regime of intermediate block-maxima. In the regime of large block-maxima, we demonstrate the distributional consistency of the estimator for a broad range of time series models including linear processes. The max-spectrum estimator is a robust and computationally efficient tool, which provides a novel time-scale perspective to the estimation of the tail exponents. Its performance is illustrated over synthetic and real data sets. Copyright © 2009 John Wiley & Sons, Ltd.

Received 25 September 2008; Revised 16 January 2009; Accepted 19 January 2009

KEY WORDS: heavy-tail exponent; max-spectrum; block-maxima; heavy-tailed time series; moving maxima; max-stable; Fréchet distribution

1. INTRODUCTION

The problem of estimating the exponent in heavy-tailed data has a long history in statistics, due to its practical importance and the technical challenges it poses. Heavy-tailed distributions are characterized by the slow, hyperbolic decay of their tail. Formally, a real-valued random variable

*Correspondence to: Stilian A. Stoev, Department of Statistics, The University of Michigan, 439 West Hall, 1085 South University, Ann Arbor, MI 48109-1107, U.S.A.
†E-mail: sstoev@umich.edu

Contract/grant sponsor: NSF; contract/grant number: DMS-0806094

Copyright © 2009 John Wiley & Sons, Ltd.

X with cumulative distribution function (c.d.f.) $F(x)=\mathbb{P}\{X\leqslant x\},\ x\in\mathbb{R}$ is (right) heavy-tailed with index $\alpha>0$, if

$$\mathbb{P}\{X>x\}=1-F(x)\sim L(x)x^{-\alpha}\quad \text{as } x\to\infty \tag{1}$$

where $\sim$ means that the ratio of the left-hand side to the right-hand side in (1) tends to one, as $x\to\infty$. Here $L(\cdot)$ is a slowly varying function at infinity, i.e. $L(\lambda x)/L(x)\to 1$, as $x\to\infty$, for all $\lambda>0$. For simplicity purposes, we suppose that X is almost surely positive i.e. $F(0)=0$, and we also focus on the case when $L(\cdot)$ is asymptotically constant, namely

$$L(x)\sim\sigma_0^{\alpha}\quad \text{as } x\to\infty \tag{2}$$

for some $\sigma_0>0$. The case when the slowly varying function $L(\cdot)$ is non-trivial is discussed in the Remarks after Theorem 3, below.

The *tail index* (*exponent*) α controls the rate of decay of the tail of F. The presence of heavy tails in data was originally noted in the work of Zipf [1] on word frequencies in languages, who also introduced a graphical device for their detection [2]. Subsequently, Mandelbrot [3] noted their presence in financial data. Since the early 1970s heavy-tailed behavior has been noted in many other scientific fields, such as hydrology, insurance claims and social and biological networks (see, e.g. [4, 5]. In particular, the emergence of the Internet and the World Wide Web gave a new impetus to the study of heavy-tailed distributions, due to their omnipresence in Internet packet and flow data, the topological structure of the Web, the size of computer files, etc. (see e.g. [6–11]). In fact, heavy-tailed behavior is a characteristic of highly optimized physical systems, as argued in [12].

Heavy tails are also ubiquitous in stock market data. It is well documented that the returns of many stocks measured at high-frequency exhibit non-negligible extreme fluctuations, consistent with a non-Gaussian, heavy-tailed model. The availability of high-frequency tic-by-tic data reveals further pronounced presence of heavy tails in the transaction volumes. Figure 1 shows the volumes associated with all single transactions of the Honeywell Inc. stocks recorded during 4th January 2005 at the New York Stock Exchange (NYSE) and National Association of Securities Dealers Automated Quotations (NASDAQ) (see [13]). The transactions are ordered by their occurrence in time. The presence of large spikes indicates heavy tails, similar, for example, to the moving average with Pareto innovations shown in Figure 2 below.

Some important features of such data are: (i) their large size due to the fine time-scale resolution (high-frequency) at which they are collected (ii) their temporal structure that introduces dependence among observations and (iii) their sequential nature, since observations are added to the data set over time. Traditional methods for estimating the tail index are not well suited for addressing these issues, as discussed below.

The majority of the approaches proposed in the literature focuses on the scaling behavior of the largest order statistics $X_{(1)}\geqslant X_{(2)}\geqslant\cdots\geqslant X_{(n)}$ obtained from an independent and identically distributed (i.i.d.) sample $X(1),\ldots,X(n)$ from F; typical examples include Hill's estimator [14] and its numerous variations [15, 16], kernel-based estimators [17, 18]. Reviews of these methods and their applications are given in [2, 19]. The most widely used in practice is the Hill estimator $\hat{\alpha}_{\mathrm{H}}(k)$ defined as

$$\hat{\alpha}_{\mathrm{H}}(k):=\left(\frac{1}{k}\sum_{i=1}^{k}\ln X_{(i)}-\ln X_{(k+1)}\right)^{-1} \tag{3}$$

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

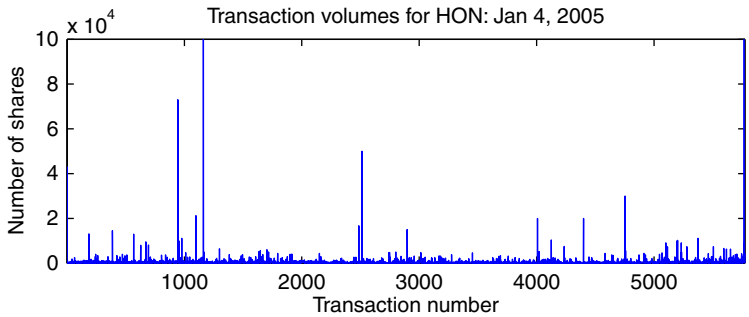

Figure 1. Transaction volumes for *Honeywell Inc.* (HON) from the NYSE and NASDAQ consolidated trades and quotes data base during 4 January 2005. The observations correspond to the volumes (in a number shares) per single transaction. The transactions are listed in the order of their occurrence in time. The observed heavy-tailed behavior of traded volumes is ubiquitous across different trading days and across the entire spectrum of relatively liquid stocks.

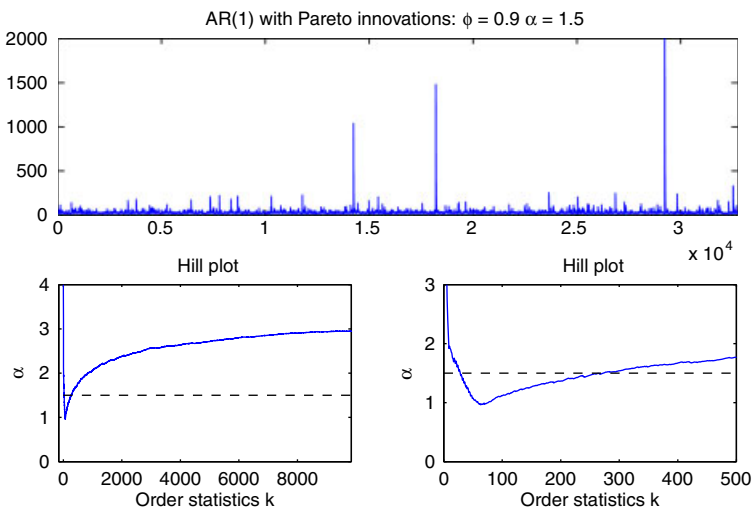

Figure 2. *Top panel*: auto-regressive time series of order one with Pareto innovations of tail exponent $\alpha = 1.5$. *Bottom left and right panels*: the Hill plot for this data set and its zoomed-in version, respectively. The dashed horizontal line indicates the value of $\alpha = 1.5$.

with $k, 1 \leqslant k \leqslant n-1$ being the number of included order statistics. The parameter k is typically selected by examining the plot of the $\hat{\alpha}_{\mathrm{H}}(k)$'s versus k, known as the *Hill plot*. In practice, one chooses a value of k where the Hill plot exhibits a fairly constant behavior (see e.g. [19]). However, the use of order statistics requires *sorting* the data that is computationally expensive (requires at least $\mathcal{O}(n \log(n))$ steps) and destroys the time ordering of the data and hence their temporal structure. Further, as can be seen from the brief review above, most of the emphasis has been placed on point estimation of the tail index and little on constructing confidence intervals. Exceptions can be found in the work of Cheng and Peng [20] and Lu and Peng [21] for the construction

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

of confidence intervals and in the work of Resnick and Stărică [22] on the estimation of α for dependent data.

The purpose of this study is to introduce a method for estimating the tail index that overcomes the above listed shortcomings of other techniques. It is based on the asymptotic *max self-similarity* properties of heavy-tailed maxima. Specifically, the maximum values of data calculated over blocks of size m scale at a rate of $m^{1/\alpha}$. Therefore, by examining a sequence of growing, dyadic block sizes $m=2^j$, $1\leqslant j\leqslant \log_2 n$, $j\in\mathbb{N}$, and subsequently estimating the mean of logarithms of block-maxima (log-block-maxima), one obtains an estimate of the tail index α. Notice that by using blocks of data, the temporal structure of the data is preserved. This procedure requires $\mathcal{O}(n)$ operations, making it particularly useful for large data sets; further, the estimates for α can be updated recursively as new data become available, by using only $\mathcal{O}(\log_2 n)$ memory and without the knowledge of the entire data set, thus making the proposed estimator particularly suitable for streaming data. Estimators based on max-self similarity for the tail index for i.i.d. data were introduced in [23], where their consistency and asymptotic normality were established. In this paper, we extend them to dependent data, prove their consistency, examine and illustrate their performance using synthetic and real data sets and discuss a number of implementation issues.

The remainder of the paper is structured as follows: in Section 2 the max-spectrum estimators are introduced. Their consistency and asymptotic normality are established in Section 3.1, for m-dependent processes. The distributional consistency of the estimators is established in Section 3.2 for a large class of time series models (including linear processes) under a mild asymptotic independence condition. The construction of confidence intervals is further addressed in Section 3.3. The important problem of automatic selection of parameters is addressed in Section 4. Applications to financial time series are discussed in Section 5, whereas most technical proofs are given in the Appendix.

2. MAX SELF-SIMILARITY AND TAIL EXPONENT ESTIMATORS

Here we introduce the max self-similarity estimators for the tail exponent and demonstrate several of their characteristics. We start by reviewing the basic ideas for the case of independent and i.i.d. data. A detailed exposition is given in [23].

Consider the sequence of block-maxima

$$X_m(k) := \max_{1\leqslant i\leqslant m} X(m(k-1)+i) = \bigvee_{i=1}^{m} X(m(k-1)+i), \quad k=1,2,\ldots,m\in\mathbb{N}$$

where $X_m(k)$ denotes the largest observation in the kth block. By (1) and (2) and the Fisher–Tippett–Gnedenko Theorem,

$$\left\{\frac{1}{m^{1/\alpha}}X_m(k)\right\}_{k\in\mathbb{N}} \xrightarrow{d} \{Z(k)\}_{k\in\mathbb{N}} \quad \text{as } m\to\infty \tag{4}$$

where $\xrightarrow{d}$ denotes convergence of the finite-dimensional distributions, with the $Z(k)$'s being independent copies of an α-Fréchet random variable. A random variable Z is said to be α-Fréchet,

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

$\alpha>0$, with scale coefficient $\sigma>0$, if

$$\mathbb{P}\{Z\leqslant x\}=\begin{cases}\exp\{-\sigma^{\alpha}x^{-\alpha}\}, & x>0\\ 0, & x\leqslant 0\end{cases} \tag{5}$$

The Fréchet variable Z is said to be standard if $\sigma=1$.

Thus, for large m's the block-maxima $X_m(k)$'s behave like a sequence of i.i.d. α-Fréchet variables, which suggests the following:

Definition 1
A sequence of random variables $X=\{X(k)\}_{k\in\mathbb{N}}$ is said to be max self-similar with self-similarity parameter $H>0$, if for any $m>0$,

$$\left\{\bigvee_{i=1}^{m} X(m(k-1)+i)\right\}_{k\in\mathbb{N}} \stackrel{d}{=} \{m^H X(k)\}_{k\in\mathbb{N}} \tag{6}$$

with $=^d$ denoting equality of the finite-dimensional distributions.

Relationship (6) holds asymptotically for i.i.d. data and exactly for Fréchet distributed data. Hence, any sequence of i.i.d. heavy-tailed variables can be regarded as *asymptotically max self-similar* with self-similarity parameter $H=1/\alpha$. This feature suggests that an estimator of H and consequently α can be obtained by focusing on the scaling of the maximum values in blocks of growing size. A similar idea applied to block-wise sums was used in [24] for estimating α, in the case $0<\alpha<2$.

For an i.i.d. sample $X(1),\ldots,X(n)$ from F, define

$$\mathbb{D}(j,k):=\max_{1\leqslant i\leqslant 2^j} X(2^j(k-1)+i)=\bigvee_{i=1}^{2^j} X(2^j(k-1)+i), \quad k=1,2,\ldots,n_j \tag{7}$$

for all $j=1,2,\ldots,[\log_2 n]$, where $n_j:=[n/2^j]$ and where $[x]$ denotes the largest integer not greater than $x\in\mathbb{R}$. By analogy to the discrete wavelet transform, we refer to the parameter j as the *scale* and to k as the *location* parameter. We consider dyadic block-sizes because of their algorithmic and computational advantages. Introduce the statistics

$$Y_j:=\frac{1}{n_j}\sum_{k=1}^{n_j}\log_2 D(j,k), \quad j=1,2,\ldots,[\log_2 n] \tag{8}$$

The Law of Large Numbers implies that for fixed j, as $n_j\to\infty$, the Y_j's are consistent and unbiased estimators of $\mathbb{E}Y_j=\mathbb{E}\log_2\mathbb{D}(j,1)$, if finite (see Corollary 3.1 in [23]). On the other hand, the asymptotic max self-similarity (4) of X and (7) suggests that under additional tail regularity conditions (see e.g. Proposition A1 below)

$$\mathbb{E}Y_j=\mathbb{E}\log_2 D(j,1)\simeq jH+C\equiv j/\alpha+C \quad \text{as } j\to\infty \tag{9}$$

where $C=C(\sigma_0,\alpha)=\mathbb{E}\log_2\sigma_0 Z$, and where $\simeq$ means that the difference between the left- and the right-hand side tends to zero, with Z being an α-Fréchet variable with unit scale coefficient.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

Then, a regression-based estimator of $H=1/\alpha$ (and hence α) for a range of scales $1\leqslant j_1\leqslant j\leqslant j_2\leqslant[\log_2 n]$ is given by

$$\hat{H}(j_1,j_2):=\sum_{j=j_1}^{j_2} w_j Y_j \quad \text{and} \quad \hat{\alpha}(j_1,j_2):=1/\hat{H}(j_1,j_2) \tag{10}$$

where the weights w_j's are chosen so that $\sum_{j=j_1}^{j_2} w_j=0$ and $\sum_{j=j_1}^{j_2} j w_j=1$. The optimal weights w_j's can be calculated through *generalized least-squares* (GLS) regression using the asymptotic covariance matrix of the Y_j's. In practice, it is important to at least use *weighted least-squares* (WLS) regression to account for the difference in the variances of the Y_j's (see [23]).

We propose to use the estimator defined in (10) *for dependent time series data.* We first illustrate its usage through a simulated data example. A data set of size $n=2^{15}=32\,768$ was generated from an auto-regressive (AR) time series of order one with Pareto innovations. Specifically,

$$X(k)=\phi X(k-1)+Z(k)=\sum_{i=0}^{\infty}\phi^i Z(k-i), \quad k=1,\ldots,n$$

where $\phi=0.9$ and $\mathbb{P}\{Z(k)>x\}=x^{-\alpha},\ x>1$, with $\alpha=1.5$. The data together with its Hill plot are shown in Figure 2. Notice that even though the Hill estimator works best for Pareto data, the dependence structure in the model leads to a Hill plot, which is substantially different from that for independent Pareto data (see the bottom left panel). The zoomed-in version of the Hill plot (bottom right panel) however indicates that the tail exponent should be in the range between one and two. The choices of k in the range between 200 and 400 in fact lead to estimates around 1.5. This range however is hard to guess if one does not know the true value of $\alpha=1.5$. Resnick and Stărică [16] have shown that the Hill estimator is consistent for such dependent data sets. Nevertheless, as this example indicates, the Hill plot can be difficult to assess in practice.

In Figure 3, the *max-spectrum* plot is shown; i.e. the plot of the statistics Y_j versus the available dyadic scales j, $1\leqslant j\leqslant[\log_2 n](=15)$. The estimated tail exponent over the range of scales $(10,15)$ is 1.4774, which is very close to the nominal value of $\alpha=1.5$. Moreover, the max-spectrum is easy to assess and interpret. One sees a 'knee' in the plot near scale $j=10$, where the max-spectrum curves upwards and thus it is natural to choose the range of scales $(10,15)$ to estimate α. The choice of the scales (j_1,j_2) can be also automated, as briefly discussed in Section 4 below.

Remark
(*On the algorithmic implementation*): The max-spectrum $Y_j,\ j=1,\ldots,[\log_2 n]$ of a data set $X_1,\ldots,X_n$ can be computed efficiently in $\mathcal{O}(n)$ steps, without sorting the data. Indeed, this is evident from the recursive construction of block maxima, since

$$D(j+1,k)=\max\{D(j,2k-1),D(j,2k)\}, \quad k=1,\ldots,[n/2^{j+1}],\ \ 1\leqslant j<[\log_2 n]$$

Moreover, this property can be further used to obtain a *sequential algorithm* for the computation of the Y_j's. Indeed, keeping in addition to the Y_j's, the last block-maximum $D_j:=D(j,n_j),\ n_j=[n/2^j]$ per scale j and also the extra variables $R_j=\max_{2^j n_j<i\leqslant n} X(i)$, which represent the maxima of the 'left-over' X_i's over the range $2^j[n/2^j]<i\leqslant n$. Now, if a new observation X_{n+1} is recorded, one can easily update the R_j's and the Y_j's, with the help of the R_j's, and the $D_{j'}:=D(j',n_{j'})$'s, for $1\leqslant j'<j$. Thus, one recovers the $(Y_j,D_j,R_j)_j$ representation of the data $X_1,\ldots,X_{n+1}$. Since only $\log_2 n$ scales are available, we perform $\mathcal{O}(\log_2 n)$ operations per update and use $\mathcal{O}(\log_2 n)$ memory to store the max-spectrum and the auxiliary data.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

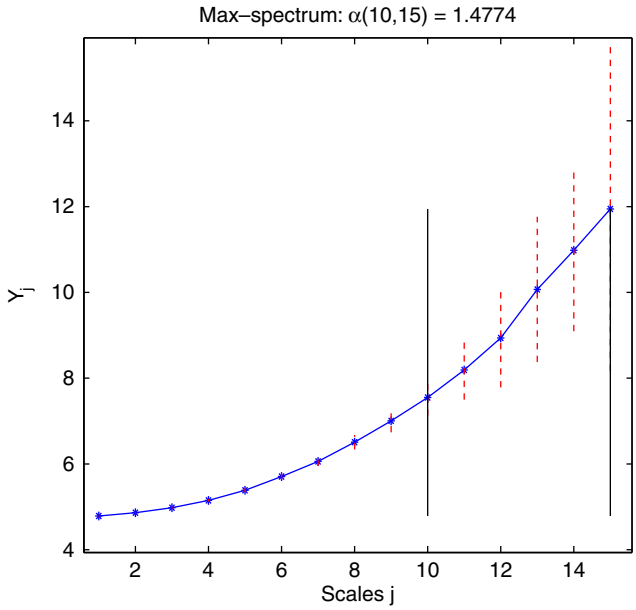

Figure 3. The max-spectrum plot for the data set in Figure 2. The max self-similarity estimator of the tail exponent, obtained from the range of scales $(j_1, j_2) = (10, 15)$, is $\hat{\alpha}(10, 15) = 1.4774$.

This sequential implementation of the max-spectrum is of critical importance in the context of data streams in modern data bases or Internet traffic applications. In such settings, large volumes of data are observed in short amounts of time; they cannot be stored and/or sorted efficiently while at the same time rapid 'queries' need to be answered about various statistics of the data. The proposed max-spectrum estimator provides a unique tool for the estimation of the tail-exponent of such data. Notice that the other available techniques require sorting the data, which is impossible without having to store the entire data set. A sequential implementation of the Hill estimator for example would require $\mathcal{O}(n)$ memory, which is prohibitive in many applications.

3. ASYMPTOTIC PROPERTIES

3.1. Asymptotic normality (in the intermediate scales regime)

The estimators $\hat{H}$ and $\hat{\alpha} = 1/\hat{H}$ in (10) utilize the scaling properties of the max-spectrum statistics Y_j in (8). The discussion in Section 2 suggests that the max self-similarity estimators in (10) will be consistent as both the scale j and n_j tend to infinity. The consistency and asymptotic normality of these estimators were established in [23] for i.i.d. data. This was accomplished by assessing the rate of convergence of moment-type functionals of block-maxima, such as $\mathbb{E} \log_2 D(j, 1)$, under mild conditions on the rate of the tail decay in (1). Here, we focus on the case of dependent data and establish the asymptotic normality of the proposed max self-similarity estimators under analogous conditions on the rate.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

Consider a strictly stationary process (time series) $X=\{X(k)\}_{k\in\mathbb{Z}}$ with heavy-tailed marginal c.d.f. F as in (1) and (2). Further, assume that the $X(i)$'s are positive, almost surely, that is, $F(0)=0$. In many contexts, the block-maxima of X scales at a rate $m^{1/\alpha}$ as the block size m grows even under the presence of strong dependence. This is so, for example, when the time series X has a *positive extremal index* (see, p. 53 in [25]). The following conditions make this more precise by quantifying further the rate of convergence. Let $M_n:=\max_{1\leqslant k\leqslant n} X(k)$ and let

$$F_n(x)\equiv\mathbb{P}\{M_n/n^{1/\alpha}\leqslant x\}=:\exp\{-c(n,x)x^{-\alpha}\},\quad x>0,\ \ n\in\mathbb{N}$$

One can see that $M_n/n^{1/\alpha}\xrightarrow{d} Z$, $n\to\infty$ if and only if $c(n,x)\to c_X\equiv\text{const}$, $n\to\infty$, for all $x>0$, where Z is an α-Fréchet variable with scale $\sigma=c_X^{1/\alpha}$. The following conditions will help us to quantify the rate of the last convergence and also to obtain rates of convergence for moment functionals of block-maxima in Proposition 2 below.

Condition 3.1
There exists $\beta>0$ and $R\in\mathbb{R}$ such that

$$|c(n,x)-c_X|\leqslant c_1(x)n^{-\beta}\quad\text{for all } x>0\quad\text{and}\quad c_1(x)=\mathcal{O}(x^{-R}),\quad x\downarrow 0 \tag{11}$$

for some $c_X>0$.

Condition 3.2
For all $x>0$, we have

$$c(n,x)\geqslant c_2\min\{1,x^{\gamma}\}\quad\text{for some } \gamma\in(0,\alpha) \tag{12}$$

for all sufficiently large $n\in\mathbb{N}$, where $c_2>0$ does not depend on n.

Remarks

1. Conditions (11) and (12) are not very restrictive. They can be shown to hold, for example, for a large class of moving maxima processes (see, Proposition A2).
2. Condition (11) implies in particular that $M_n/n^{1/\alpha}\xrightarrow{d} c_X^{1/\alpha}Z$, as $n\to\infty$, for a standard α-Fréchet variable Z. In view of (1), we also have that $M_n^*/n^{1/\alpha}\xrightarrow{d}\sigma_0 Z$, as $n\to\infty$, with $M_n^*:=\max_{1\leqslant k\leqslant n} X(k)^*$, where the $X(k)^*$'s are i.i.d. random variables with c.d.f. F. This implies that the *extremal index* θ of the time series X is: $\theta:=c_X/\sigma_0^{\alpha}$ (see, e.g. p. 53 in [25]).

Conditions (11) and (12) yield the following important result on the rate of convergence of log-block maxima, similar to Corollary (11) in [23].

Proposition 2
Let $X=\{X(k)\}_{k\in\mathbb{Z}}$ be a strictly stationary time series that satisfies Conditions (11) and (12). Suppose that $\int_1^{\infty} c_1(x)x^{-\alpha-1+\delta}\mathrm{d}x<\infty$, for some $\delta>0$. Then, with $M_n:=\max_{1\leqslant k\leqslant n} X_k$, we have $\mathbb{E}|\ln(M_n)|^p<\infty$, for all $p>0$ and all sufficiently large $n\in\mathbb{N}$. Moreover, for any $p>0$ and $k\in\mathbb{N}$, we have

$$|\mathbb{E}|\ln(M_n/n^{1/\alpha})|^p-\mathbb{E}|\ln(Z)|^p|=\mathcal{O}(n^{-\beta})\quad\text{and}\quad|\mathbb{E}(\ln(M_n/n^{1/\alpha}))^k-\mathbb{E}(\ln(Z))^k|=\mathcal{O}(n^{-\beta})$$

as $n\to\infty$, where Z is an α-Fréchet random variable with scale coefficient $\sigma=c_X^{1/\alpha}$.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

The proof is given in Section 5. Proposition 2 readily implies

$$\mathbb{E}(Y_j - j/\alpha) \equiv \mathbb{E}\log_2(D(j,k)/2^{j/\alpha}) = \mathbb{E}\log_2(c_X^{1/\alpha} Z_1) + \mathcal{O}(1/2^{j\beta}) \tag{13}$$

as $j \to \infty$, where Z_1 is a standard α-Fréchet variable. This result yields an asymptotic bound on the bias of the estimators $\hat{H}(j_1, j_2)$ in (10) above.

Proposition 2 can be further used to establish the asymptotic normality of $\hat{\alpha}(j_1, j_2) = 1/\hat{H}(j_1, j_2)$ in (10). To do so, we focus on a range of scales (j_1, j_2) that grows with the sample size. Namely, we fix $\ell \in \mathbb{N}$, $\ell \geqslant 2$, let $j_1 := 1 + j(n)$ and $j_2 := \ell + j(n)$, and as in (10) define

$$\hat{H}_n := \sum_{i=1}^{\ell} w_i Y_{i+j(n)} \quad \text{and} \quad \hat{\alpha}_n := 1/\hat{H}_n$$

where $\sum_{i=0}^{\ell-1} i^\kappa w_i = \kappa$, $\kappa = 0, 1$. The next theorem is the main result of this section. It establishes the asymptotic normality of the estimator $\hat{\alpha}_n$, as $j(n)$ and n tend to infinity.

Theorem 3
Let $X_1, \ldots, X_n$ be a sample from an m-dependent heavy-tailed process $X = \{X(k)\}_{k \in \mathbb{Z}}$. Suppose that (11) and (12) hold and let $j = j(n) \in \mathbb{N}$ be such that

$$2^{j(n)}/n + n/2^{j(n)(1+2\min\{1,\beta\})} \longrightarrow 0 \quad \text{as } n \to \infty \tag{14}$$

Then, as $n \to \infty$,

$$\sqrt{n_j}(\hat{\alpha}_n - \alpha) \xrightarrow{d} \mathcal{N}(0, \alpha^2 c_w) \quad \text{with } c_w = \sum_{i',i''=1}^{\ell} w_{i'} w_{i''} \Sigma_1(i', i'') \tag{15}$$

where $n_j = [n/2^j]$. Here $\Sigma_1(i', i'') = 2^{\min\{i',i''\}} \operatorname{Cov}(\log_2 Z_1, \log_2(Z_1 \vee (2^{|i'-i''|} - 1) Z_2))$, for $i', i'' = 1, \ldots, \ell$, where Z_1 and Z_2 are independent standard 1-Fréchet random variables.

Proof
By the 'Delta method' (see e.g. Theorem 3.1 in [26]), it suffices to show that

$$\sqrt{n_j}(\hat{H}_n - H) \xrightarrow{d} \mathcal{N}(0, H^2 c_w) \quad \text{as } n \to \infty \tag{16}$$

Indeed, since $\hat{\alpha}_n = f(\hat{H}_n)$, with $f(x) = 1/x$, we have $\hat{\alpha}_n - \alpha = -H^{-2}(\hat{H}_n - H) + o_p(\hat{H}_n - H)$, as $n \to \infty$.

Let now

$$\tilde{D}(j', k) := \bigvee_{i=1}^{2^{j'}-m} X(2^{j'}(k-1) + i) \quad \text{and} \quad \tilde{Y}_{j'} := \frac{1}{n_{j'}} \sum_{k=1}^{n_{j'}} \log_2 \tilde{D}(j', k)$$

for all $j_1 \leqslant j' \leqslant j_2$, and $k = 1, \ldots, n_{j'}$, where $n_{j'} := [n/2^{j'}]$. Observe that since the time series X is m-dependent, the $\tilde{D}(j', k)$'s are now independent in k. Hence, in view of Conditions (11) and (12) and Proposition 2, the results of Theorem 4.1 in [23] readily apply to the max-spectrum $\tilde{Y}_{j'}$, $j' = j_1, \ldots, j_2$, which is based on the independent $\tilde{D}(j', k)$'s. Therefore, by setting $\tilde{H}_n := \sum_{i=1}^{\ell} w_i \tilde{Y}_{i+j(n)}$, we obtain

$$\sqrt{n_j}(\tilde{H}_n - H) \xrightarrow{d} \mathcal{N}(0, H^2 c_w) \quad \text{as } n \to \infty \tag{17}$$

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

In view of (17), to establish (16), it is enough to show that $\hat{H}_n - \tilde{H}_n = o_p(1/\sqrt{n_j})$, $n \to \infty$, or that, for example,

$$\mathbb{E}(\hat{H}_n - \tilde{H}_n)^2 = \mathrm{Var}(\hat{H}_n - \tilde{H}_n) + (\mathbb{E}\hat{H}_n - \mathbb{E}\tilde{H}_n)^2 = o(1/n_j) \quad \text{as } n \to \infty \tag{18}$$

Consider first the term $\mathrm{Var}(\hat{H}_n - \tilde{H}_n)$. Since

$$\hat{H}_n - \tilde{H}_n = \sum_{i=1}^{\ell} w_i (Y_{i+j(n)} - \tilde{Y}_{i+j(n)})$$

we have

$$\mathrm{Var}(\hat{H}_n - \tilde{H}_n) \leqslant C \sum_{i=1}^{\ell} \mathrm{Var}(Y_{i+j(n)} - \tilde{Y}_{i+j(n)}) \leqslant \frac{C'}{n_j} \sum_{i=1}^{\ell} \mathrm{Var}(\log_2 D(i+j(n),1) - \log_2 \tilde{D}(i+j(n),1))$$

for some constants C and C', where the last inequality follows from Lemma A1. Now, Lemmas A2 and A3 imply that $\mathrm{Var}(\log_2 D(i+j(n),1) - \log_2 \tilde{D}(i+j(n),1)) \to 0$, $n \to \infty$, and hence $\mathrm{Var}(\hat{H}_n - \tilde{H}_n) = o(1/n_j)$, as $n \to \infty$.

Now, we focus on the term $(\mathbb{E}\hat{H}_n - \mathbb{E}\tilde{H}_n)^2$ *in* (18). For some constant $C_\ell > 0$, we have

$$(\mathbb{E}\hat{H}_n - \mathbb{E}\tilde{H}_n)^2 \leqslant C_\ell \sum_{i=1}^{\ell} (\mathbb{E}Y_{i+j(n)} - \mathbb{E}\tilde{Y}_{i+j(n)})^2 = C_\ell \sum_{i=1}^{\ell} \left(\mathbb{E} \log_2 \frac{D(i+j(n),1)}{\tilde{D}(i+j(n),1)} \right)^2 \tag{19}$$

where we used the inequality $(\sum_{i=1}^{\ell} x_i)^2 \leqslant \ell \sum_{i=1}^{\ell} x_i^2$ and the stationarity (in k) of the $D(i+j(n),k)$'s and $\tilde{D}(i+j(n),k)$'s. We further have that

$$\begin{aligned} \mathbb{E} \log_2 \frac{D(i+j(n),1)}{\tilde{D}(i+j(n),1)} &= \mathbb{E} \log_2 \left(\frac{D(i+j(n),1)}{2^{j(n)/\alpha}} \right) - \mathbb{E} \log_2 \left(\frac{\tilde{D}(i+j(n),1)}{(2^{j(n)} - m)^{1/\alpha}} \right) \\ &\quad - \frac{1}{\alpha} \log_2 \left(1 - \frac{m}{2^{j(n)}} \right) \end{aligned} \tag{20}$$

Relation (13) implies that the last two expectations are both equal to $\mathbb{E} \log_2(Z) + \mathcal{O}(1/2^{j(n)\beta})$, as $n \to \infty$, where Z is an α-Fréchet variable with scale $c_X^{1/\alpha}$. Therefore, from (19) and (20), we obtain

$$(\mathbb{E}\hat{H}_n - \mathbb{E}\tilde{H}_n)^2 = \mathcal{O}(1/2^{2j(n)\beta}) + \mathcal{O}(1/2^{2j(n)}) = \mathcal{O}(1/2^{2j(n)\min\{1,\beta\}}) \quad \text{as } n \to \infty$$

where in the last relation we used that $\log_2(1-x) = \mathcal{O}(x)$, $x \to 0$.

By combining the above derived bounds on the terms in the right-hand side of (18), we get

$$\hat{H}_n - \tilde{H}_n = o_P(1/\sqrt{n_j}) + o_p(1/2^{j(n)\min\{1,\beta\}}) = o_p(1/\sqrt{n_j}) \quad \text{as } n \to \infty$$

where the last equality follows from (14) since $n/2^{j(n)(1+2\min\{1,\beta\})} = n_j/2^{2j(n)\min\{1,\beta\}} \to 0$, as $n \to \infty$. This implies (16) and completes the proof of the theorem. □

We conclude this section with several important remarks on the scope of validity of the asymptotic results in Theorem 3.

Remarks

1. (*On the role of* β): The parameter $\beta > 0$ in Condition (11) controls the rate of the convergence in distribution of $M_n/n^{1/\alpha}$ to the α-Fréchet limit law. The larger the value of β, the

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

faster the convergence in (11), and in view of (14), the wider the range of scales $j(n)$'s in Theorem 3 that lead to asymptotically normal $\hat{\alpha}_n$'s. In particular, the larger the β, the faster the convergence of the $\hat{\alpha}_n$'s can be made, since one could choose relatively small $j(n)$'s. On the other hand, when the rate of convergence of the law of $M_n/n^{1/\alpha}$ to its limit is relatively slow, then the values of $\beta>0$ can be close to zero. This can lead to arbitrarily slow rates of the convergence of $\hat{\alpha}_n$ since one may have to choose relatively large scales $j(n)$'s to compensate for the rate of the bias in the max-spectrum on smaller scales.

2. (*On the connection with Hill estimators*): As argued in [23], for the case of independent data, Condition (11) corresponds precisely to the second-order condition used in [27], where the asymptotic normality of the Hill estimator was established. The rates of convergence in (15) above are, in the case of independent data, in close correspondence with the rates for the Hill estimator, obtained in [27].
3. (*On data with regularly varying tails*): Consider the case when the $X(k)$'s satisfy (1) where now the slowly varying function $L(\cdot)$ is *non-trivial*. Then, the max-spectrum-based estimators of α will continue to work. Indeed, for the case of i.i.d. data, we have that

$$\frac{M_n}{a_n} \xrightarrow{d} Z \quad \text{as } n \to \infty$$

where Z is a standard α-Fréchet variable, and where $a_n = n^{1/\alpha}\ell(n)$ is such that

$$n a_n^{-\alpha} L(a_n) \equiv \frac{1}{\ell(n)} L(n^{1/\alpha}\ell(n)) \longrightarrow 1 \quad \text{as } n \to \infty$$

Here $\ell(\cdot)$ is another slowly varying function related to L (see e.g. Proposition 1.11 in [28]). If one replaces $n^{1/\alpha}$ by $a_n \equiv n^{1/\alpha}\ell(n)$ and c_X by 1 in Conditions (11) and (12) then Propositions 2 and A1 will continue to hold with $M_n/n^{1/\alpha}$ replaced by M_n/a_n. The proofs are essentially the same. In the case of independent data, one has that

$$\mathrm{Var}(Y_j) = \frac{1}{n_j}\mathrm{Var}(\log_2 D(j,1)/(2^{j/\alpha}\ell(2^j)^{1/\alpha})) = \frac{1}{n_j}(\mathrm{Var}(\log_2 Z) + o(1)) \tag{21}$$

where the remainder term $o(1)$ vanishes, as $j \to \infty$, because of the analog of Relation (13). Also, by the counterpart of (13), one obtains

$$\mathbb{E}(Y_j - j/\alpha - \log_2(\ell(2^j))/\alpha) = \mathbb{E}\log_2(Z) + \mathcal{O}(1/2^{j\beta}) \quad \text{as } j \to \infty \tag{22}$$

Consider now a fixed $m \in \mathbb{N}$, $m \geqslant 2$ and let

$$\hat{H}_n = \sum_{i=1}^{m} w_i Y_{i+j(n)}$$

Relation (21) implies that $\mathrm{Var}(\hat{H}_n) \to 0$, as n and $j(n)$ tend to infinity. On the other hand, Relation (22) shows that

$$\begin{aligned}\mathbb{E}\hat{H}_n &= \frac{1}{\alpha}\sum_{i=1}^{m} i w_i + \mathcal{O}(1/2^{j\beta}) + (j(n) + \mathbb{E}\log_2(Z))\sum_{i=1}^{m} w_i + \frac{1}{\alpha}\sum_{i=1}^{m}\log_2(\ell(2^{i+j})) \\ &= \frac{1}{\alpha} + \mathcal{O}(1/2^{j\beta}) + \frac{1}{\alpha}\sum_{i=1}^{m}\log_2(\ell(2^i \cdot 2^{j(n)})/\ell(2^{j(n)}))\end{aligned} \tag{23}$$

where in the last two relations we used the facts that $\sum_{i=1}^{m} i w_i = 1$ and $\sum_{i=1}^{m} w_i = 0$.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

Now, the fact that $\ell(\cdot)$ is a slowly varying function implies that $\ell(2^i \cdot 2^{j(n)})/\ell(2^{j(n)}) \to 1$, as $j(n) \to \infty$. This shows that the right-hand side of (23) converges to $H \equiv 1/\alpha$, as $j(n) \to \infty$ and hence the estimator $\hat{H}_n$ is consistent, as $n \to \infty$ and as $j(n) \to \infty$. Note that the rate of the bias $(\mathbb{E}\hat{H}_n - H)$ depends not only on the term $\mathcal{O}(1/2^{j\beta})$, but also on the rate of the convergence

$$\ell(\lambda j)/\ell(j) \longrightarrow 1 \quad \text{as } j \to \infty$$

This last rate depends on the structure of the slowly varying function $\ell(\cdot)$ and it may be possible to control in terms of Karamata's integral representation

$$\ell(x) = c_\ell(x) \exp\left\{-\int_{x_0}^{x} \varepsilon(u)/u \, \mathrm{d}u\right\}$$

at the expense, however, of two additional parameters controlling the rates of $c_\ell(\cdot)$ and $\varepsilon(\cdot)$.

This argument shows the consistency of the max-spectrum-based estimator of α for i.i.d. $X(k)$'s with regularly varying tails. In principle, one can establish asymptotic normality of these estimators along similar line, but this would involve technically complicated assumptions on the slowly varying functions considered. Further, as in Theorem 3 one can establish asymptotic normality results for the $\hat{\alpha}_n$'s for m-dependent data. We chose not to pursue the general case of regularly varying tails here since the technical details may obscure the idea behind the estimator. These important theoretical results will be pursued in subsequent work on the subject.

3.2. *Distributional consistency* (*in the large scales regime*)

In Theorem 3, we consider an asymptotic regime where the number of block-maxima n_j on the scale $j = j(n)$ grows, as $n \to \infty$. This is essential for the consistency of the estimators $\hat{\alpha}_n$. In practice, however, the situation where we have a fixed number of block-maxima per scale is also of interest. Namely, for a sample $X(1), \ldots, X(n)$ and a *fixed* number of block-maxima r, we let $j(n) := [\log_2(n/r)]$ and consider the estimator

$$\hat{\alpha}_n := \sum_{i=1}^{\ell} w_i Y_{i+j(n)} \quad \text{where } \ell = [\log_2 r] \tag{24}$$

This estimator corresponds to taking the largest ℓ scales in the max-spectrum, where ℓ is *fixed*. One cannot expect the estimators $\hat{\alpha}_n$ to be consistent (even for independent data) since they involve averages over a fixed number of block-maxima statistics. Nevertheless, the asymptotic distribution of $\hat{\alpha}_n$ is of interest.

The next result establishes the 'distributional consistency' of the estimators $\hat{\alpha}_n$ in the aforementioned regime. We do so under the condition that the block-maxima in (4) are asymptotically independent. This condition is in fact quite mild, as shown in Lemmas 5 and 6 below.

Theorem 4
Suppose that (4) holds where the $Z(k)$'s are i.i.d. α-Fréchet. Then,

$$\hat{\alpha}_n \xrightarrow{d} \hat{\alpha}_Z \quad \text{as } n \to \infty \tag{25}$$

where $\hat{\alpha}_F := 1/(\sum_{i=1}^{\ell} w_i Y_i^Z)$ and where $\{Y_i^Z\}_{i=1}^{\ell}$ is the max-spectrum of a sequence of i.i.d. α-Fréchet variables $Z(1), \ldots, Z(r)$.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

Proof
The result readily follows from the continuous mapping theorem. Indeed, by (4), and in view of (7), we have

$$\{D(j(n),k)/2^{j(n)/\alpha},\ k=1,\dots,r\} \xrightarrow{d} \{Z(k), k=1,\dots,r\} \quad \text{as } n\to\infty$$

and hence

$$\{\log_2 D(j(n),k) - j(n)/\alpha, k=1,\dots,r\} \xrightarrow{d} \{\log_2 Z(k), k=1,\dots,r\} \quad \text{as } n\to\infty$$

Owing to the dyadic structure of the block-maxima $D(j(n),k)$'s, one can recover $D(i+j(n),k)$'s, for $i=1,\dots,\ell$ and $k=1,\dots,[r/2^i]$ from the block-maxima $D(j(n),k)$, $k=1,\dots,r$ through a *continuous* combination of maxima operations. Thus, by applying the continuous mapping theorem again, we obtain

$$\{Y_{i+j(n)} - j(n)/\alpha\}_{i=1}^{\ell} \xrightarrow{d} \{Y_i^Z\}_{i=1}^{\ell} \quad \text{as } n\to\infty$$

which yields the convergence (25) since $\sum_{i=1}^{\ell} w_i\, j(n)/\alpha = 0$. □

Condition (4) appears stringent, but contrary to intuition, it holds in most practical situations. We were unable to find an example of ergodic heavy-tailed time series X (of positive extremal index) with *asymptotically dependent* block-maxima. We next show that (4) holds for the large class of linear processes.

Let ξ_k, $k\in\mathbb{Z}$ be i.i.d. heavy-tailed innovations, such that $\mathbb{P}\{|\xi_k|>x\}\sim\sigma_\xi^\alpha x^{-\alpha}$, $x\to\infty$, where $\mathbb{P}\{\xi_k>x\}/\mathbb{P}\{|\xi_k|>x\}\to p$, $x\to\infty$ for some $p\in[0,1]$. Consider the linear process

$$X(k) := \sum_{i=-\infty}^{\infty} c_i\,\xi_{k-i}, \quad k\in\mathbb{Z} \tag{26}$$

Mikosch and Samorodnitsky [29] provide a recent and comprehensive treatment of the linear processes as in (26) (see also Davis and Resnick [30]). More precisely, by Lemma A.3 in [29], the following conditions on the c_i's guarantee the almost sure convergence of the series in (26):.

$$\sum_{i=-\infty}^{\infty} c_i^2 < \infty \quad (\text{if } \alpha>2) \quad \text{and} \quad \sum_{i=-\infty}^{\infty} |c_i|^{\alpha-\varepsilon} < \infty, \quad (\text{if } \alpha\leqslant 2),\ \text{for some } \varepsilon>0 \tag{27}$$

These conditions are necessary for $\alpha>0$, and nearly optimal for $\alpha\leqslant 2$ (see Lemma A.3 in [29]). By Lemma A.3 in [1], we also have that the tails of the $X(k)$'s are regularly varying with exponent α (see Relation (A.2) therein).

The following result shows that (4) holds for the linear process $X=\{X(k)\}_{k\in\mathbb{Z}}$ under the conditions (27). The proof follows by a simple combination of arguments in [29, 30] and it is given in the Appendix, for completeness.

Lemma 5
Let $c_+ := \max_{i\geqslant 0} c_i$ and $c_- := \max_{i\geqslant 0}(-c_i)$. Suppose that either $pc_+>0$ or $(1-p)c_->0$. Then, the linear process in (26) satisfies (4) where the $Z(k)$'s are i.i.d. α-Fréchet with scale coefficient $\sigma_\xi(pc_+^\alpha + (1-p)c_-^\alpha)^{1/\alpha}$.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

The next result provides some further insight to the observed *independence phenomenon* for block-maxima. Namely, it turns out that the block-maxima of a heavy-tailed time series are always asymptotically independent, provided that they converge to a max-stable process.

Lemma 6
Let $X=\{X(k)\}_{k\in\mathbb{Z}}$ be a heavy-tailed time series with marginal distributions as in (1). Suppose that (4) holds where the $Z(k)$'s are not assumed independent.

If the limit time series $Z=\{Z(k)\}_{k\in\mathbb{N}}$ is multivariate max-stable, then it consists of i.i.d. random variables.

The proof is given in the Appendix.

3.3. On the construction of confidence intervals

In many applications, an uncertainty assessment about the estimated tail exponent is important, which requires the construction of confidence intervals.

The literature is rather sparse for confidence intervals for the heavy-tail exponent even in the case of independent data. We are not aware of any general results on the asymptotic distribution of the Hill or the moment estimator of α for dependent data. Theorem 3 above suggests the following *asymptotic confidence interval* for α of level γ, $0<\gamma<1$:

$$((\hat{H}-\hat{H}z_{(1-\gamma)/2}\sqrt{c_w}/\sqrt{n_j})^{-1},(\hat{H}+\hat{H}z_{(1-\gamma)/2}\sqrt{c_w}/\sqrt{n_j})^{-1}) \tag{28}$$

where $z_{(1-\gamma)/2}$ is $(1+\gamma)/2-$quantile of the standard normal distribution, and where c_w is as in Theorem 3. Here, as recommended in [23], we use the reciprocal of a symmetric confidence interval for H to obtain one for $\alpha=1/H$ (see also (16)).

Tables I and II illustrate coverage probabilities of confidence intervals for α, based on Theorems 3 and 4, respectively. They are based on 1000 independent replications of max-AR(1) time series $X=\{X(k)\}_{k\in\mathbb{Z}}$:

$$X(k):=\phi X(k-1)\vee Z(k)=\bigvee_{i=0}^{\infty}\phi^i Z(k-i),\quad k=1,\ldots,n \tag{29}$$

of size $n=2^{15}=32768$ for different values of ϕ. Here the $Z(k)$'s are i.i.d. and α-Fréchet with $\alpha=1.5$. The coverage probabilities for 90, 95 and 99% levels of confidence are reported in each row, as a function of j_1.

Observe that when the data are closer to independent ($\phi=0.1$), the coverage probabilities match the nominal values even for small j_1's. As the degree of dependence grows, larger values for j_1 are required to achieve accurate coverage probabilities. Nevertheless, even in the most dependent setting ($\phi=0.9$), the value of $j_1=8$ in Table I yields very good results.

Observe that coverage probabilities in Table I deteriorate for very large scales j_1. This is due to the inadequacy of the normal approximation in Theorem 3 in the presence of a limited number of block-maxima. For large j_1's, the regime described in Theorem 4 is more applicable. Table II shows that the coverage probabilities based on (25) are very accurate even for the largest scales $j_1=13$. We obtained these confidence intervals by using a Monte Carlo method. Namely, we approximate the distribution of the statistics $\hat{\alpha}_F$ based on 1000 independent paths of i.i.d. 1-Fréchet variables, multiplied by the estimated $\hat{\alpha}_n$'s. Although these confidence intervals are significantly slower to compute than (28), they exhibit excellent coverage probabilities even for the largest scales j_1.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

Table I. Coverage probabilities of the asymptotic confidence intervals (28) for α for max-AR(1) time series as in (29) of length 2^{15}.

j_1	3	4	5	6	7	8	9	10	11
90% *c.i.* $\phi=0.1$	0.891	0.894	0.912	0.919	0.897	0.903	0.889	0.895	0.875
$\phi=0.3$	0.759	0.888	0.914	0.915	0.899	0.901	0.889	0.895	0.875
$\phi=0.5$	0.229	0.772	0.889	0.915	0.892	0.899	0.888	0.895	0.875
$\phi=0.7$	0.000	0.299	0.801	0.895	0.895	0.899	0.887	0.895	0.875
$\phi=0.9$	0.000	0.000	0.070	0.641	0.843	0.890	0.877	0.890	0.875
95% *c.i.* $\phi=0.1$	0.943	0.952	0.954	0.953	0.949	0.950	0.931	0.931	0.904
$\phi=0.3$	0.844	0.940	0.952	0.953	0.949	0.950	0.931	0.931	0.904
$\phi=0.5$	0.321	0.854	0.950	0.954	0.948	0.950	0.931	0.931	0.904
$\phi=0.7$	0.000	0.395	0.872	0.946	0.944	0.950	0.931	0.931	0.904
$\phi=0.9$	0.000	0.000	0.123	0.738	0.911	0.941	0.927	0.930	0.904
99% *c.i.* $\phi=0.1$	0.990	0.990	0.989	0.991	0.987	0.993	0.975	0.972	0.947
$\phi=0.3$	0.946	0.985	0.990	0.991	0.987	0.992	0.975	0.972	0.947
$\phi=0.5$	0.552	0.953	0.984	0.990	0.987	0.991	0.975	0.972	0.947
$\phi=0.7$	0.000	0.642	0.959	0.981	0.988	0.990	0.974	0.972	0.947
$\phi=0.9$	0.000	0.000	0.276	0.897	0.968	0.984	0.973	0.972	0.947

Max self-similarity estimators $\hat{H}=\hat{H}(j_1, j_2)$ were used with $1\leqslant j_1\leqslant j_2$ and $j_2=15$. Results for three confidence levels: 90, 95 and 99% are shown for different values of j_1.

Table II. Coverage probabilities of empirical confidence intervals based on Theorem 4 for α for max-AR(1) time series as in (29) of length 2^{15}.

j_1	5	6	7	8	9	10	11	12	13
90% *c.i.* $\phi=0.1$	0.884	0.909	0.903	0.907	0.901	0.914	0.887	0.902	0.917
$\phi=0.3$	0.903	0.906	0.915	0.888	0.898	0.910	0.906	0.907	0.916
$\phi=0.5$	0.911	0.908	0.905	0.905	0.898	0.906	0.890	0.902	0.898
$\phi=0.7$	0.837	0.885	0.879	0.898	0.906	0.908	0.907	0.906	0.899
$\phi=0.9$	0.103	0.735	0.863	0.888	0.894	0.909	0.920	0.909	0.915
95% *c.i.* $\phi=0.1$	0.945	0.953	0.950	0.947	0.947	0.951	0.946	0.953	0.959
$\phi=0.3$	0.956	0.944	0.953	0.941	0.942	0.955	0.946	0.963	0.956
$\phi=0.5$	0.949	0.955	0.956	0.947	0.935	0.945	0.947	0.952	0.933
$\phi=0.7$	0.894	0.949	0.939	0.949	0.939	0.956	0.954	0.947	0.947
$\phi=0.9$	0.163	0.820	0.935	0.943	0.934	0.958	0.959	0.959	0.957
99% *c.i.* $\phi=0.1$	0.992	0.993	0.992	0.994	0.984	0.989	0.989	0.991	0.997
$\phi=0.3$	0.995	0.991	0.987	0.993	0.985	0.992	0.992	0.992	0.998
$\phi=0.5$	0.990	0.996	0.991	0.997	0.993	0.986	0.984	0.988	0.980
$\phi=0.7$	0.953	0.990	0.989	0.992	0.990	0.994	0.988	0.994	0.993
$\phi=0.9$	0.337	0.933	0.984	0.990	0.980	0.995	0.984	0.989	0.993

Max self-similarity estimators $\hat{H}=\hat{H}(j_1, j_2)$ were used with $1\leqslant j_1\leqslant j_2$ and $j_2=15$. Results for three confidence levels: 90, 95 and 99% are shown for different values of j_1.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

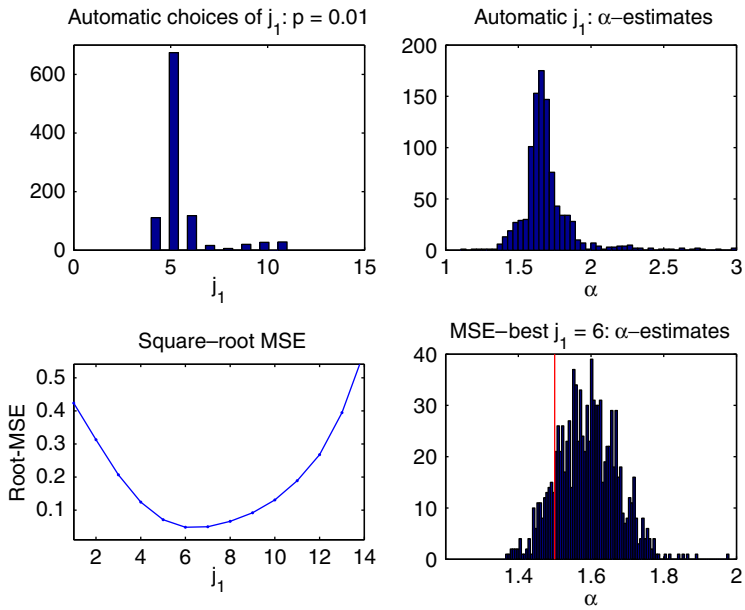

Figure 4. The top-plot shows the histogram of automatically selected j_1 values for 1000 independent samples of size $N=2^{15}$ from an exponential moving maxima α-Fréchet process, $X=\{X(k)\}_{k\in\mathbb{Z}}$, defined as in (29) with $\phi=0.9$ and with i.i.d. 1.5-Fréchet innovations. We used significance level and back-start parameters are $p=0.01$ and $b=4$, respectively. The top-right plot shows the histogram of the resulting $\hat{\alpha}=1/\hat{H}$ estimates. The bottom-left plot shows estimates of the square root of the mean squared error (MSE) $\mathbb{E}(\hat{H}-H)^2$ as a function of j_1. The bottom-right plot contains a histogram of $\hat{\alpha}$ estimates obtained with the MSE-optimal choice of $j_1=11$.

In conclusion, the brief numerical experiments suggest that the confidence intervals in (28) work well in practice, even for dependent data, for judicious choice of scales j_1 and j_2. The confidence intervals based on Theorem 4, on the other hand, work well for all sufficiently large scales, where the asymptotic normality may not apply. Both types of confidence intervals are useful in practice.

4. ON THE AUTOMATIC SELECTION OF THE CUT-OFF SCALE j_1

In the ideal case of α-Fréchet i.i.d. data, the max-spectrum plot of Y_j is linear in j. When the distribution of the data is not Fréchet, or when the data are dependent, then the max-spectrum is asymptotically linear, as the scales j tend to infinity. It is therefore important to select appropriately the range of large scales j for estimation purposes. In view of (9), one can always choose $j_2=[\log_2 n]$ to be the largest available scale and hence, the problem is reduced to choosing the scale j_1, $1\leqslant j_1<j_2$. The estimator of α is then obtained by performing a WLS or GLS linear regression of Y_j versus j, $j_1\leqslant j\leqslant j_2$ (see (10)).

The 'cut-off' parameter j_1 can be selected either by visually inspecting the max-spectrum or through a data-driven procedure. In [23] an automatic procedure for selecting the cut-off parameter was proposed, in the case of independent data, whose main steps are briefly summarized next.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

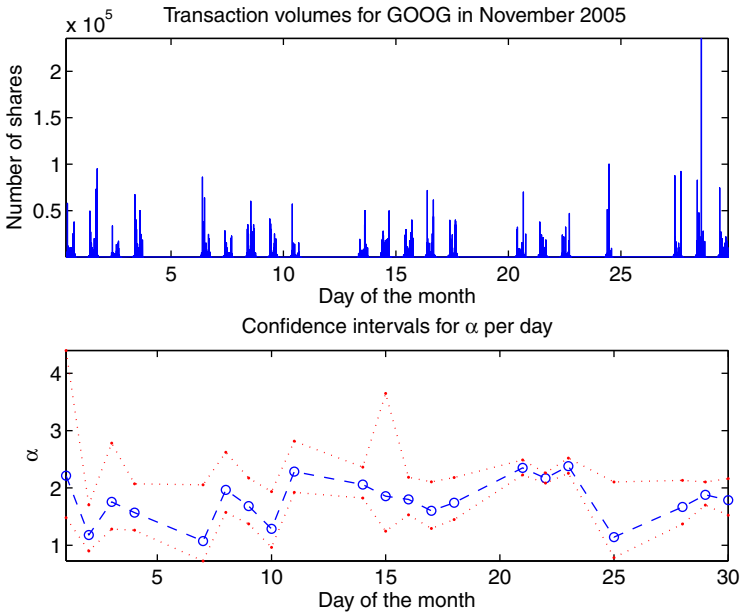

Figure 5. *Top panel:* traded volumes for the Google stock from the TAQ data base of consolidated trades of NYSE and NASDAQ for the month of November 2005. The x-axis and y-axis correspond to time and number of traded shares, respectively. This is a high-frequency data set, where each data point corresponds to the volume of a single transaction and no temporal aggregation is performed. The gaps of zeros in the data correspond to hours of the day with no trading and/or weekends. *Bottom panel:* estimated tail exponents (indicated by circles) from the max-spectrum and their corresponding 95% confidence intervals (indicated by broken lines), based on the asymptotic expression in (28). Automatic selection of the cut-off scale j_1 was done with $p=0.1$ and $b=3$ (see Section 4). Every estimate was computed from a day worth of transaction volumes.

We also demonstrate that it performs satisfactorily for dependent data. The algorithm sets $j_2 := [\log_2 n]$ and $j_1 := \max\{1, j_2 - b\}$, with $b=3$ or 4 in practice for moderate sample sizes. Next, j_1 is iteratively decreased until statistically significant deviations from linearity of Y_j, $j_1 \leqslant j \leqslant j_2$ are detected. Namely, as $j_1 > 1$, at each iteration over the scale j_1 the following two quantities are calculated $\hat{H}_{\text{new}} = \hat{H}(j_1 - 1, j_2)$ and $\hat{H}_{\text{old}} = \hat{H}(j_1, j_2)$. Whenever the value of zero is *not* contained in a confidence interval centered at $(\hat{H}_{\text{new}} - \hat{H}_{\text{old}})$, the algorithm stops and returns the selected j_1 and $\hat{\alpha} = 1/\hat{H}_{\text{old}}$; otherwise, it sets $j_1 := j_1 - 1$ and proceeds accordingly. The construction of the confidence interval about $(\hat{H}_{\text{new}} - \hat{H}_{\text{old}})$ utilizes the covariance matrix Σ_1 in Theorem 3, which is the same as in the i.i.d. case, see [23]. The asymptotic normality result suggests that the methodology in the case of i.i.d. data applies asymptotically to dependent data, for moderately large scales j_1. Alternatively, the results of Theorem 4 may be used to suitably correct the confidence intervals on the largest scales j_1. We did not implement this method, since it is computationally demanding in practice.

Figure 4 demonstrates the performance of the automatic selection procedure in the case of dependent data. Even though the marginal distributions of X are Fréchet, the dependence causes a

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

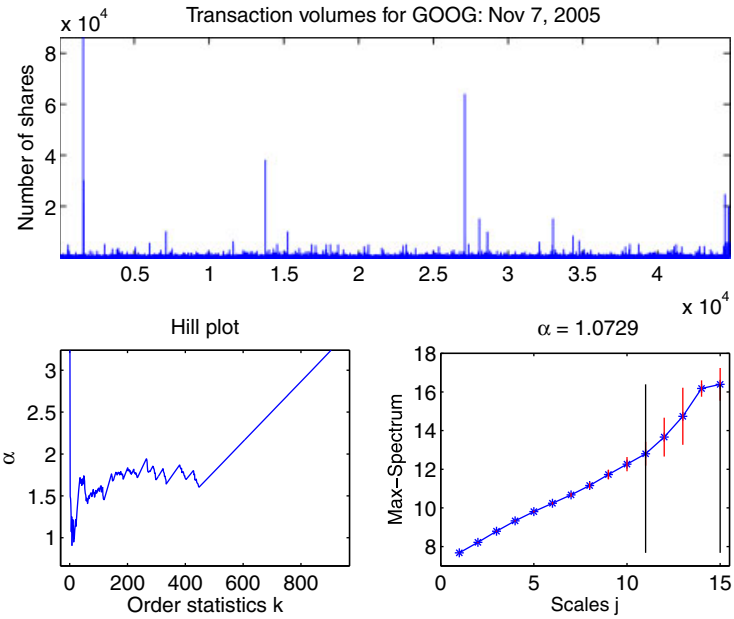

Figure 6. *Top panel*: the transaction volumes during the trading hours of November 7, 2005. The x-axis corresponds to the number of the transaction and the y-axis to number of shares. Note that about 50 000 transactions occurred on this day, which is typical for the Google stock. Observe also the fairly classical heavy-tailed nature of the volume data. *Bottom panels*: the Hill plot (left) and the max-spectrum (right) of the data. The Hill plot is zoomed-in to a range where it is fairly constant and a tail exponent between 1.5 and 2 can be identified. The max-spectrum reveals more: on large scales the plot is steeper than on small scales with the tail exponent about 1 on the range of scales $(11, 15)$ and exponent about 1.7 on scales $(1, 10)$. The presence of a knee in the max-spectrum plot suggests different behavior of the largest volumes on large time scales than on small time scales and can be contributed to the several very large spikes of over 20 000 traded shares (about 5 million US dollars) the top plot.

knee in the max-spectrum plot (see, e.g. Figure 3). The automatic selection procedure picks up this 'knee' and yields reasonably unbiased and precise *automatic estimates of* α (see the top-right panel in Figure 4). Comparing the mean squared error (MSE) plot and the histogram of the selected j_1 values, we see that over 70% of the times the value $j_1 = 5$ was chosen, which is close to the optimal value of $j_1 = 6$. The histogram of the resulting automatic estimates of α (top-right panel) is similar (with the exception of a few outliers) to the histogram of the estimators corresponding to the MSE-optimal $j_1 = 6$ (bottom-right panel).

Recall Table I, and observe that the case $\phi = 0.9$ corresponds to the time series analyzed in Figure 4. The coverage probabilities of the confidence intervals for α essentially match the nominal levels, for $j_1 \geqslant 8$. On the other hand, the MSE-optimal value is $j_2 = 6$ (Figure 4), which is only slightly smaller than $j_1 = 8$. This can be contributed to the fact that the bias involved in the estimators at $j_1 = 6$, although comparable with their standard errors is significant and noticeably *shifts* the confidence interval. As the scale j_1 grows, the bias quickly becomes negligible and the resulting confidence intervals become accurate.

These brief experiments suggest that the automatic procedure is practical and works reasonably well in the case of dependent moving maxima time series. Similar experiments for independent

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

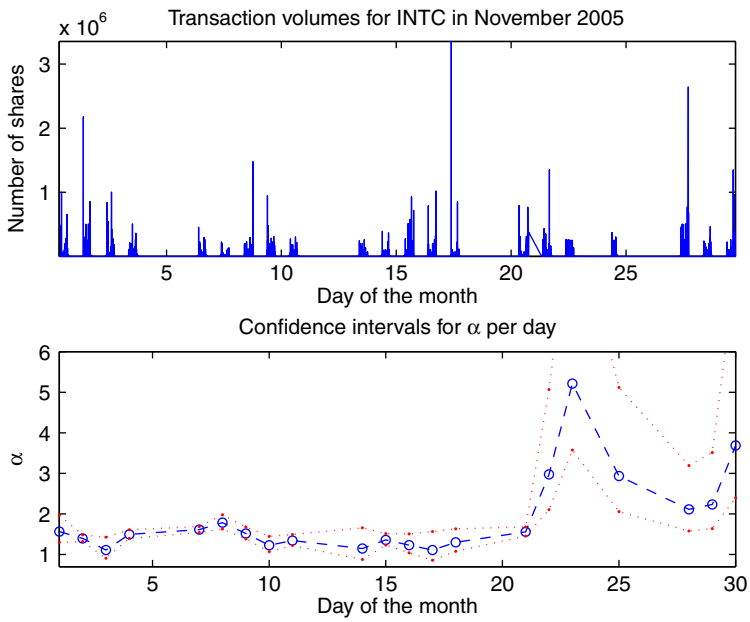

Figure 7. This figure has the same format as Figure 5. On the top panel, the traded volumes of the Intel stock for the month of November 2005 are shown. Observe that the tail exponent estimates on the bottom plot fluctuate between 1.5 and 2 up to 21 November. On and after 22 November, unusually high values of α appear (compare with the case of the Google stock in Figure 5). This is further analyzed in Figures 8 and 9, below.

heavy-tailed data (not shown here) indicate that the automatic selection procedure continues to perform well and chooses values of j_1 close to the MSE-optimal ones, thus making it appropriate for use in empirical work. Nevertheless, a detailed study of its performance under a combination of heavy-tailed distributions and dependence structures, as well as its sensitivity to the choice of the back-start parameter b and the level of significance p, is necessary and is the subject of future work.

5. APPLICATIONS TO FINANCIAL DATA

We analyze market transactions for two stocks—Intel (symbol INTC) and Google (GOOG)—using the max-spectrum. The data sets were obtained from the *trades and quotes* (TAQ) data base of *consolidated transactions* of the NYSE and NASDAQ (see [13]) and include the following information about every single trade of the underlying stock: *time of transaction* (up to seconds), *price* (of the share) and *volume* (in number of shares). In our analysis, we focus on the traded volumes of the two stocks for November 2005, which could provide information about the respective sector's, as well as the market's economic conditions [31].

A ubiquitous feature of the volume data sets is the presence of heavy, Pareto-type tails, as can be seen in Figure 6. Specifically, the top panel shows transaction volumes for the Google stock on 7 November 2005, while the bottom panels show the Hill and the max-spectrum plots, respectively. The tail exponent, estimated from the max-spectrum over the range of scales $(11, 15)$ is $\hat{\alpha} = 1.0729$.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

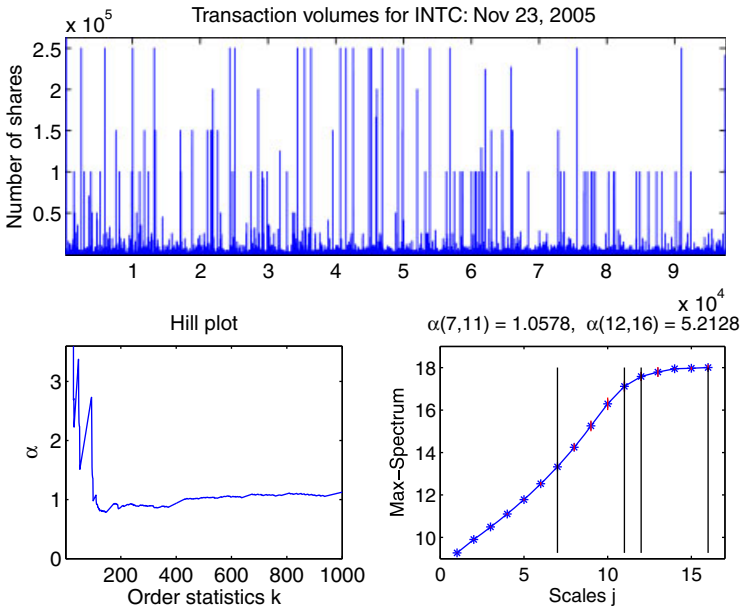

Figure 8. *Top panel*: traded volumes of the Intel stock for 23 November 2005. Observe the regular occurrence of many very large trades of approximately the same sizes: 10 000, 15 000, 25 000 and a few of 20 000 shares. This is a very unusual behavior of the volume data, as compared with a typical trading day (see, e.g. Figure 9). *Bottom panels*: the Hill plot and the max-spectrum of the data. Notice that the Hill plot fails to identify the unusual behavior of the data, whereas the max-spectrum flattens out, on large scales due to the regular non-heavy-tailed behavior of the largest traded volumes. Once identified on the max-spectrum plot, one can perhaps read-off these details from the volatile Hill plot for very small values of k. On small scales, where the regular large transactions are not frequent and do not play a role, the max-spectrum yields tail exponents about one. This is in line with the Hill plot.

The Hill plot indicates heavy-tail exponent estimates between 1.5 and 2, which correspond to the slope of the max-spectrum over the range of scales $(1, 10)$. The small dip in the Hill plot for very large order statistics (small values of k) can be related to the behavior of the max-spectrum for scales $(11, 15)$. Such behavior is typical for almost all liquid stocks, as well as the presence of non-stationarity and dependence. In order to minimize the intricate non-stationarity effects, we focus here on traded volumes within a day. The max-spectrum yields consistent tail exponent estimates even in the presence of dependence. This fact and the robustness of the max-spectrum suggest that it may be safely used in various practical scenarios involving heavy-tailed data. In Figure 5, we show the max self-similarity estimates of the tail exponents, for each of the 21 trading days in November 2005. The max-spectra of these 21 time series (not shown here) of trading volumes are essentially linear. This confirms the validity of a heavy-tailed model for the data, valid over a wide range of time scales—from seconds up to hours and days. Further, at the beginning and end of the trading day, several large volume transactions are observed, as documented in [32]. Nevertheless, the trading activity of Google remains essentially linear over the period under study, with a few bumps at the largest scales due to diurnal effects and other non-stationarities.

In Figure 5, the daily tail exponent estimates are shown for the Google stock, which fluctuate between one and two, along with pointwise confidence intervals (broken lines). These estimates

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

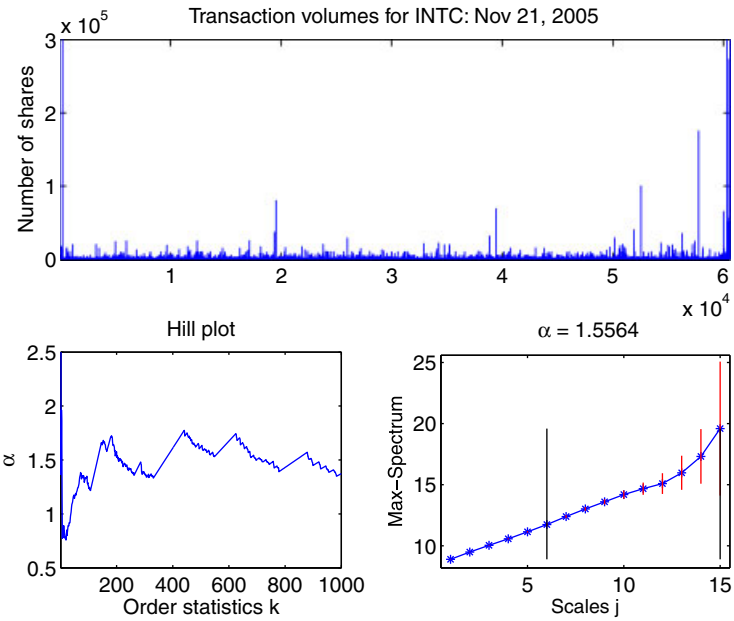

Figure 9. This figure has the same format as Figure 8. The top plot shows the volumes of INTC during 21 November 2005, which as the volumes of GOOG in Figure 6, behave like a classical heavy-tailed sample. The Hill plot and the max-spectrum (bottom left and right panels, respectively) identify tail exponents around 1.5. The cut-off scale in the max-spectrum plot was selected automatically with $p=0.1$ and $b=3$ (as in Figure 7). Notice the volatile, saw-tooth shape of the Hill plot, which is due to its non-robustness to deviations from the Pareto model. The max-spectrum is more robust and fairly linear with a small knee on scale $j=12$, which may be due to a few clusters of large volumes in the beginning and at the end of the trading day.

indicate that the tail exponent exhibits a significant degree of variability over the period of a month, and that an infinite variance model may be most appropriate for modeling trading volumes. For example, on 7 November (see Figure 6), the estimate of α is nearly one, which may be due to the several extremely large peaks in the volume data. The upward knee in the max-spectrum of this data set is likely caused by these peaks. The max-spectra on most other days are much closer to linear than the one in Figure 6. Such correspondence between the presence of large peaks in the data and the behavior of the max-spectrum can be used to identify statistically significant fluctuations in the volume data. Hence, the max-spectrum plot can be used not only to estimate α, but also to detect changes in the market. We illustrate this last point next, by examining an unusual trading pattern in the Intel stock toward the end of November 2005.

Figure 7 shows the max-spectrum estimates of the tail exponents for the traded volumes of the Intel stock for 21 trading days in November 2005. Notice that up to 21 November, the tail exponent is fairly constant, fluctuating between 1.2 and 2. On 22 November (Tue) and 23 November (Wed), before the Thanksgiving holiday on 24 November (Thur), the tail exponent takes values larger than three and five, respectively. This change is quite surprising and it is deemed significant by the corresponding confidence intervals. A closer look at the data from 23 November (Figure 8) shows a changing but persistent pattern of trading as compared with 21 November; see for example Figure 9.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

This behavior proves persistent and continues on 25 November, after the Thanksgiving holiday. Moreover, no such behavior was observed for the Google data on any of the 21 trading days in November 2005. Although trading of extremely large volumes occurs on November 23, as seen in Figure 8, these trades are very regular and hence inconsistent with a heavy-tailed model. Although regular in time, these large transactions occur on a time scale of several minutes, and hence the small scales of the max-spectrum are not affected by these peaks and behave as on a normal trading day (see Figure 9). However, the large peaks dominate the larger scales j and their regularity makes the max-spectrum essentially horizontal. The Hill plot, shown on the bottom-left panel of Figure 8, fails to pick up the unusual behavior, since it suggests values of $\alpha \approx 1$, which correspond only to the smallest portion of the max-spectrum, where $\hat{\alpha}(7, 11) = 1.0578 \approx 1$.

Our best guess is that this change in activity is related to the approval by the board of directors of the Intel Corp. on 10 November of a program for a stock buy-back worth of up to 25 billion U.S. dollars (see, e.g. the Financial Times, London, on Thursday 11 November, p. 27); hence, some of the delayed effects of the announcement of the program and market reaction to it are demonstrated in the volume activity discussed above.

APPENDIX A

A.1. Rates of convergence for moment functionals of dependent maxima

Proposition A1

Suppose that $f:(0,\infty)\to\mathbb{R}$ is an absolutely continuous function on any compact interval $[a,b]\subset(0,\infty)$, and such that $f(x)=f(x_0)+\int_{x_0}^{x} f'(u)\,\mathrm{d}u,\ x>0$ for some (any) $x_0>0$.

Let for some $m\in\mathbb{R}$ and $\delta>0$,

$$x^m|f(x)|+\operatorname{ess\,sup}_{0<y\leqslant x} y^m|f'(y)| \longrightarrow 0 \quad \text{as } x\downarrow 0 \tag{A1}$$

$$x^{-\alpha}|f(x)|+x^{1+\delta}\operatorname{ess\,sup}_{y\geqslant x} y^{-\alpha}|f'(y)| \longrightarrow 0 \quad \text{as } x\to\infty \tag{A2}$$

Suppose also that the time series $X=\{X_n\}_{n\in\mathbb{Z}}$ satisfies Conditions (11) and (12), where $c_1(x)$ is such that

$$\int_1^\infty c_1(x)x^{-\alpha}|f'(x)|\,\mathrm{d}x<\infty \tag{A3}$$

Then, $E|f(M_n)|<\infty$, for all sufficiently large $n\in\mathbb{N}$, and for some $C_f>0$, independent of n,

$$|\mathbb{E}f(M_n/n^{1/\alpha})-\mathbb{E}f(Z)|\leqslant C_f n^{-\beta} \tag{A4}$$

where Z is an α-Fréchet variable with scale coefficient $\sigma:=c_X^{1/\alpha}$.

Proof

The proof is similar to the proof of Theorem 3.1 in [23]. Indeed, as in the above reference, one can show that $\mathbb{E}|f(Z)|<\infty$ and $\mathbb{E}|f(M_n)|<\infty$, for all sufficiently large n. Further, by using the conditions (A1) and (A2) and integration by parts, we have that

$$\mathbb{E}f(M_n/n^{1/\alpha})-\mathbb{E}f(Z)=\int_0^\infty (G(x)-F_n(x))f'(x)\,\mathrm{d}x \tag{A5}$$

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

where $F_n(x) := \mathbb{P}\{M_n/n^{1/\alpha} \leqslant x\}$ and $G(x) = \mathbb{P}\{Z \leqslant x\}$. Since $F_n(x) = \mathrm{e}^{-c(n,x)x^{-\alpha}}$, by the mean value theorem, we have

$$|G(x) - F_n(x)| = |\mathrm{e}^{-c_X x^{-\alpha}} - \mathrm{e}^{-c(n,x)x^{-\alpha}}| \leqslant |c(n,x) - c_X| x^{-\alpha} \mathrm{e}^{-\min\{\theta c_X, c(n,x)\}x^{-\alpha}}$$
$$\leqslant n^{-\beta} c_1(x) x^{-\alpha} (\mathrm{e}^{-c_2 x^{-(\alpha-\gamma)}} + \mathrm{e}^{-\theta c_X x^{-\alpha}})$$

where in the last inequality, we used Relations (11) and (12).

Thus, by (A5), we have that

$$|\mathbb{E} f(M_n/n^{1/\alpha}) - \mathbb{E} f(Z)| \leqslant n^{-\beta} \int_0^\infty c_1(x) x^{-\alpha} |f'(x)| (\mathrm{e}^{-c_2 x^{-(\alpha-\gamma)}} + \mathrm{e}^{-c_X x^{-\alpha}}) \, \mathrm{d}x$$
$$=: n^{-\beta} \left(\int_0^1 + \int_1^\infty \right) \tag{A6}$$

The last integral is finite. Indeed, since the exponential terms above are bounded, Relation (A3) implies that the integral '$\int_1^\infty$' is finite. On the other hand, conditions (11) and (A1) imply that $c_1(x)|f'(x)| = \mathcal{O}(x^{-R})$, $x \downarrow 0$, for some $R \in \mathbb{R}$. However, for all $p > 0$, we have $(\mathrm{e}^{-c_2 x^{-(\alpha-\gamma)}} + \mathrm{e}^{-c_X x^{-\alpha}}) = o(x^p)$, $x \downarrow 0$, since $\alpha - \gamma > 0$. This implies that the integral in '$\int_0^1$' in (A6) is also finite. This completes the proof of (A4). □

Proof of Proposition 2

It is enough to show that the functions $f(x) := |\ln(x)|^p$ and $f(x) := (\ln(x))^k$, $p > 0$, $k \in \mathbb{N}$ satisfy the conditions of Proposition A1. In the first case, for example, $|f'(x)| = px^{-1}|\ln(x)|^{p-1}$, $x > 0$. Therefore, the assumption $\int_1^\infty c_1(x) x^{-\alpha-1+\delta} \, \mathrm{d}x < \infty$ implies (A3), since $|\ln(x)|^{p-1} \leqslant \mathrm{const}\, x^\delta$, for all $x \in [1, \infty)$. The conditions (A1) and (A2) are also fulfilled in this case, and hence Proposition A1 yields the desired order of convergence. The functions $f(x) = (\ln(x))^k$, $k \in \mathbb{N}$ can be treated similarly.

In the rest of this section, we demonstrate that Conditions (11) and (12) apply to a general class of moving maxima processes.

Let $\{Z_n\}_{n \in \mathbb{N}}$ be a sequence of i.i.d. random variables with the c.d.f. $P\{Z \leqslant z\} = F_Z(z)$. As in [23], we suppose that

$$F_Z(z) = \exp\{-c(z) z^{-\alpha}\}, \quad z > 0 \tag{A7}$$

and impose two further conditions, analogous to Conditions (11) and (12).

Condition A1

There exists $\beta' > 0$, such that

$$|c(z) - c_Z| \leqslant K z^{-\beta'} \quad \text{for all } z > 0 \tag{A8}$$

where $c_Z > 0$ and $K \geqslant 0$.

Condition A2

$F_Z(0) = 0$ and for all $x > 0$,

$$c(z) \geqslant c \min\{1, z^\gamma\} \quad \text{for some } \gamma \in (0, \alpha) \tag{A9}$$

with $c > 0$.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

Observe that (A8) implies $c(z) \to c_Z$, $z \to \infty$ and in fact $P\{Z>z\} = 1 - F_Z(z) \sim c_Z z^{-\alpha}$, as $z \to \infty$. Define now the moving maxima process $X = \{X_k\}_{k \in \mathbb{Z}}$:

$$X_k := \max_{1 \leqslant i \leqslant m} a_i Z_{k-i+1}, \quad k \in \mathbb{Z} \tag{A10}$$

with some coefficients $a_i > 0$, $i = 1, \ldots, m$, and $m \geqslant 1$. The following result shows that the process X satisfies Conditions (11) and (12).

Proposition A2
If the Z_n's satisfy Conditions 6.1 and 6.2, then the process $X = \{X_k\}_{k \in \mathbb{Z}}$ in (A10) satisfies (1), Conditions (11) and (12) with γ as in (A9),

$$\sigma_0^\alpha = c_Z \sum_{i=1}^m a_i^\alpha, \quad \beta = \min\{1, \beta'/\alpha\} \quad \text{and} \quad c_1(x) := \text{const}(1 + x^{-\beta'}) \tag{A11}$$

where β' is as in (A8) and where $c_X := c_Z \max_{1 \leqslant i \leqslant m} a_i^\alpha$. In particular, the extremal index of X is $\theta = c_X / \sigma_0^\alpha = \max_{1 \leqslant i \leqslant m} a_i^\alpha / \sum_{i=1}^m a_i^\alpha$.

Proof
We first derive the marginal distribution of the X_k's. By (A7) and (A10), we have

$$\mathbb{P}\{X_k \leqslant x\} = P\{Z_k \leqslant x/a_1, \ldots, Z_{k-m+1} \leqslant x/a_m\} = \exp\left\{-\sum_{i=1}^m c(x/a_i) a_i^\alpha x^{-\alpha}\right\}$$

Thus, in view of (A8), $c(x/a_i) \to c_Z$, $x \to \infty$, and hence, as $x \to \infty$

$$\mathbb{P}\{X_k > x\} \sim \sigma_0^\alpha x^{-\alpha} \quad \text{where } \sigma_0^\alpha := c_Z \sum_{i=1}^m a_i^\alpha \tag{A12}$$

We now focus on the maxima $M_n := \max_{1 \leqslant i \leqslant n} X_i$. For $n > m$, and $x > 0$, we have that $F_n(x) := \mathbb{P}\{M_n / n^{1/\alpha} \leqslant x\}$ equals

$$\begin{aligned} F_n(x) &= \mathbb{P}\{X_1 \leqslant n^{1/\alpha} x, \ldots, X_n \leqslant n^{1/\alpha} x\} \\ &= \mathbb{P}\left\{\bigvee_{j=2-m}^{0} g_{j,m} Z_j \leqslant n^{1/\alpha} x, \bigvee_{j=1}^{n-m+1} a_{(1)} Z_j \leqslant n^{1/\alpha} x, \bigvee_{j=0}^{m-2} h_j Z_{n-j} \leqslant n^{1/\alpha} x\right\} \end{aligned}$$

where

$$a_{(1)} := \bigvee_{k=1}^{m} a_k, \quad g_{j,m} = \bigvee_{k=2-j}^{m} a_k, \quad h_j = \bigvee_{k=1}^{1+j} a_k$$

Therefore, by using the independence of the Z_j's and Relation (A7), we get $F_n(x) = \exp\{-c(n,x) x^{-\alpha}\}$, $x > 0$, where

$$\begin{aligned} c(n,x) = \frac{1}{n} \Bigg(&\sum_{j=2-m}^{0} c(n^{1/\alpha} x / g_{j,m}) g_{j,m}^\alpha + (n-m+1) a_{(1)}^\alpha c(n^{1/\alpha} x / a_{(1)}) \\ &+ \sum_{j=0}^{m-2} c(n^{1/\alpha} x / h_j) h_j^\alpha \Bigg) \end{aligned} \tag{A13}$$

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

We will now show that Relation (11) holds with β and $c_1(\cdot)$ as in (A11). Let $c_X := c_Z a_{(1)} = c_Z \max_{1\leqslant i\leqslant m} a_i^\alpha$. By (A13), we have

$$\begin{aligned}|c(n,x)-c_X| &= |c(n,x)-c_Z a_{(1)}^\alpha| \\ &\leqslant \frac{1}{n}\sum_{j=2-m}^{0} |c(n^{1/\alpha}x/g_{j,m})-c_Z| g_{j,m}^\alpha + \frac{(n-m+1)}{n}|c(n^{1/\alpha}x/a_{(1)})-c_Z|a_{(1)}^\alpha \\ &\quad + \frac{1}{n}\sum_{j=0}^{m-2}|c(n^{1/\alpha}x/h_j)-c_Z|h_j^\alpha + \frac{C}{n} =: A_1+A_2+A_3+\frac{C}{n} \end{aligned} \tag{A14}$$

where the constant C does not depend on x. In the last relation, we add and subtract the finite number of $2(m-1)$ terms of the type $g_{j,m}^\alpha c_Z$ and $h_j^\alpha c_Z$ and apply the triangle inequality.

Now, by applying Relation (A8) to each one of the absolute value terms in A_1, we obtain

$$A_1 \leqslant \frac{K a_{(1)}^\alpha}{n}\sum_{j=2-m}^{0} n^{-\beta'/\alpha}x^{-\beta'} g_{j,m}^{\beta'} \leqslant \frac{m-1}{n^{1+\beta'/\alpha}} K a_{(1)}^{\alpha+\beta'} x^{-\beta'} = \frac{C_1}{n^{1+\beta'/\alpha}} x^{-\beta'} \tag{A15}$$

where the constant C_1 does not depend on n and x and where in the last inequalities we used that $g_{j,m} \leqslant a_{(1)}$. One obtains a similar bound for the term A_3 in (A14):

$$A_3 \leqslant \frac{C_3}{n^{1+\beta'/\alpha}} x^{-\beta'} \tag{A16}$$

where the constant C_3 does not depend on n and x.

Now, for the term A_2 in (A14), we also have by (A8) that

$$A_2 \leqslant \frac{n-m+1}{n} K a_{(1)}^{\alpha+\beta'} x^{-\beta'} n^{-\beta'/\alpha} \leqslant \frac{C_2}{n^{\beta'/\alpha}} x^{-\beta'} \tag{A17}$$

where the constant C_2 does not depend on n and x.

By combining the bounds in (A15)–(A17), for the terms in (A14), we obtain

$$|c(n,x)-c_X| \leqslant \frac{(C_1+C_3)}{n^{1+\beta'/\alpha}} x^{-\beta'} + \frac{C_2}{n^{\beta'/\alpha}} x^{-\beta'} + \frac{C}{n}$$

which shows that (11) holds with $c_1(x) = \text{const}\ (1+x^{-\beta'})$, where $\beta := \beta'/\alpha$.

We now show that (12) holds. Since (12) involves a lower bound, we can ignore the two positive sums in (A13). Recall (A9) and note that $c(n^{1/\alpha}x/a_{(1)}) \geqslant c_2' \min\{1, (n^{1/\alpha}x/a_{(1)})^\gamma\}$. Since, for sufficiently large n, $n^{1/\alpha} > a_{(1)}$, and $(n^{1/\alpha}x/a_{(1)})^\gamma \geqslant x^\gamma$, we obtain $c(n^{1/\alpha}x/a_{(1)}) \geqslant c_2' \min\{1, x^\gamma\}$. Therefore, by (A13), since for all sufficiently large n, $(n-m+1)/n \geqslant 1/2$, we have $c(n,x) \geqslant c_2 \min\{1, x^\gamma\}$, where $c_2 = a_{(1)}^\alpha c_2'/2$. This implies (12) and completes the proof of the proposition. □

A.2. Auxiliary lemmas

The next three lemmas were used in the proof of Theorem 3.

Lemma A1

Under the conditions of Theorem 3, for all $j > \log_2 m$, we have

$$\mathrm{Var}(Y_j - \tilde{Y}_j) \leqslant \frac{3}{n_j}\mathrm{Var}(\log_2(D(j,1)/\tilde{D}(j,1)))$$

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

Proof
For notational simplicity, let $\xi_k := \log_2(D(j,k)/\tilde{D}(j,k))$, $k=1,\ldots,n_j$. We have, by the stationarity of ξ_k in k, that

$$\mathrm{Var}(Y_j-\tilde{Y}_j)=\frac{1}{n_j}\mathrm{Var}(\xi_1)+\frac{2}{n_j^2}\sum_{k=1}^{n_j-1}(n_j-k)\,\mathrm{Cov}(\xi_{k+1},\xi_1)$$

Note that $\xi_{k+1}=\log_2(D(j,1+k)/\tilde{D}(j,1+k))$ and $\xi_1=\log_2(D(j,1)/\tilde{D}(j,1))$ are independent if $k>1$. Indeed, this follows from the fact that the process X is m-dependent, and since ξ_{k+1} and ξ_1 depend on blocks of the data separated by at least $2^j>m$ lags. Therefore, only the lag-1 covariances in the above sum will be non-zero and hence

$$\mathrm{Var}(Y_j-\tilde{Y}_j)\leqslant\frac{1}{n_j}\mathrm{Var}(\xi_1)+\frac{2}{n_j}|\mathrm{Cov}(\xi_2,\xi_1)|\leqslant\frac{3}{n_j}\mathrm{Var}(\xi_1)$$

since by the Cauchy–Schwartz inequality we have $|\mathrm{Cov}(\xi_2,\xi_1)|\leqslant\mathrm{Var}(\xi_2)^{1/2}\mathrm{Var}(\xi_1)^{1/2}=\mathrm{Var}(\xi_1)$. This completes the proof of the lemma. □

Lemma A2
Under the conditions of Theorem 3, for any fixed k, we have $D(j,k)/\tilde{D}(j,k)\xrightarrow{P}1$, as $j\to\infty$.

Proof
Let $\delta\in(0,1/\alpha)$ be arbitrary and observe that

$$\mathbb{P}\{D(j,k)/\tilde{D}(j,k)<1\}=\mathbb{P}\{R>\tilde{D}(j,k)\}\leqslant\mathbb{P}\{R>2^{j\delta}\}+\mathbb{P}\{2^{j\delta}>\tilde{D}(j,k)\} \tag{A18}$$

where $R=\max_{1\leqslant i\leqslant m}X_{2^j(k-i)+1}$. Now, by stationarity,

$$\mathbb{P}\{R>2^{j\delta}\}=\mathbb{P}\left\{\max_{1\leqslant i\leqslant m}X_i>2^{j\delta}\right\}\to 0\quad\text{as } j\to\infty$$

On the other hand, Relation (11) implies that $2^{-j/\alpha}\tilde{D}(j,k)\xrightarrow{d}Z$, as $n\to\infty$, where Z is a non-degenerate α-Fréchet variable. Thus, since $\delta\in(0,1/\alpha)$, we have that

$$\mathbb{P}\{2^{j\delta}>\tilde{D}(j,k)\}\to 0\quad\text{as } j\to\infty$$

The last two convergence and the inequality (A18) imply that $\mathbb{P}\{D(j,k)/\tilde{D}(j,k)<1\}\to 0$, $j\to\infty$. Since trivially $\mathbb{P}\{D(j,k)/\tilde{D}(j,k)>1\}=1$, we obtain $D(j,k)/\tilde{D}(j,k)$ converges in distribution to the constant one, as $j\to\infty$. This completes the proof since convergence in distribution to a constant implies convergence in probability. □

Lemma A3
The set of random variables

$$\left|\log_2\left(\frac{D(j,k)}{\tilde{D}(j,k)}\right)\right|^p,\quad j,k\in\mathbb{N}$$

is uniformly integrable, for all $p>0$, where $D(j,k)$ and $\tilde{D}(j,k)$ are as in Theorem 3.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

Proof
Let $q>p$ be arbitrary. By using the inequality $|x+y|^q \leqslant 2^q(|x|^q+|y|^q)$, $x, y \in \mathbb{R}$, we get

$$\mathbb{E}\left|\log_2 \frac{D(j,k)}{\tilde{D}(j,k)}\right|^q \leqslant 2^q\,\mathbb{E}|\log_2(D(j,k)/2^{j/\alpha})|^q + 2^q\,\mathbb{E}|\log_2(\tilde{D}(j,k)/2^{j/\alpha})|^q$$

In view of Proposition 2, applied to the block-maxima $D(j,k)$ and $\tilde{D}(j,k)$, we obtain

$$\mathbb{E}|\log_2(D(j,k)/2^{j/\alpha})|^q = \mathbb{E}|\log_2(M_{2^j}/2^{j/\alpha})|^q \longrightarrow \text{const} \quad \text{as } j\to\infty$$

Thus the set $\{\mathbb{E}|\log_2(D(j,k)/2^{j/\alpha})|^q,\ j,k\in\mathbb{N}\}$ is bounded. We similarly have that the set $\{\mathbb{E}|\log_2(\tilde{D}(j,k)/2^{j/\alpha})|^q,\ j,k\in\mathbb{N}\}$ is bounded since $\log_2(2^j-m)\sim j,\ j\to\infty$, for any fixed m.

We have thus shown that

$$\sup_{j,k\in\mathbb{N}} \mathbb{E}\left|\log_2 \frac{D(j,k)}{\tilde{D}(j,k)}\right|^q < \infty$$

for $q>p$, which yields the desired uniform integrability. □

We now present the proofs of Lemmas 5 and 6 in Section 3.2.

Proof of Lemma 5
We start by noting that the results of Lemma 2.3, and Theorems 2.4 and 3.1 in [30] continue to hold for the linear process $X=\{X(k)\}_{k\in\mathbb{Z}}$, under the more general conditions in (27). The proofs of Theorems 2.4 and 3.1 in [30] depend on the specific conditions in (27) only through Lemma 2.3 and Relation (2.7) therein. These two results (Lemma 2.3 and Relation (2.7)) are valid, thanks to Lemma A.3 of [29].

Now, following the proof of Theorem 3.1 in [30], introduce the map $T_r: M_p((0,\infty)\times\mathbb{R}\setminus\{0\}) \to \mathbb{R}^r$,

$$T_r\left(\sum_{k=1}^{\infty} \varepsilon_{(u_k,v_k)}\right) := \left(\bigvee_{u_k\in(0,1/r]} v_k, \ldots, \bigvee_{u_k\in(i/r,(i+1)/r]} v_k, \ldots, \bigvee_{u_k\in((r-1)/r,1]} v_k\right) \tag{A19}$$

The map T_r is simpler than the map $T: M_p((0,\infty)\times\mathbb{R}\setminus\{0\}) \to D(0,\infty)$ considered in [30], where $D(0,\infty)$ denotes the Skorkhod space of *càdlág* functions. The space M_p of Radon point measures is equipped with the topology of vague convergence, where a set $K\subset(0,\infty)\times\mathbb{R}\setminus\{0\}$ is compact if it is closed and bounded away from zero. We will argue below that the map T_r is almost surely continuous when applied to suitable Poisson random measures (PRMs).

Proceeding as in the proof of Theorem 3.1 in [30], (by Theorem 2.4 (i) therein) we get

$$\sum_{k=1}^{\infty} \varepsilon_{(k/m, m^{1/\alpha}X(k))} \Longrightarrow \sum_{i=0}^{\infty}\sum_{k=1}^{\infty} \varepsilon_{(t_k, j_k c_i)} \tag{A20}$$

where '$\Rightarrow$' denotes weak convergence of point processes and where $\varepsilon_{(t,j)}$ denotes a point measure with unit mass concentrated at $(t,j)\in(0,\infty)\times\mathbb{R}\setminus\{0\}$. In (A20), $\{(t_k,j_k)\}_{k\geqslant 0}$ are the points of a PRM with intensity measure

$$\mu(\mathrm{d}t,\mathrm{d}x) = \mathrm{d}x\times\lambda(\mathrm{d}x)$$

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

where

$$\lambda(dx)=\alpha p x^{-\alpha-1}1_{(0,\infty)}(x)\,\mathrm{d}x+\alpha(1-p)(-x)^{-\alpha-1}1_{(-\infty,0)}(x)\,\mathrm{d}x$$

(recall the distribution of the ξ_k's above and see (2.1) in [30]).

Let now $m=\sum_{i=0}^{\infty}\sum_{k=1}^{\infty}\varepsilon_{(t_k,j_kc_i)}$ be the PRM in (A20) and observe that $\mathbb{P}\{m(\partial B)=0\}=1$, where $B:=((0,1/r]\times\mathbb{R}\setminus\{0\})\cup\cdots\cup((r-1)/r,1]\times\mathbb{R}\setminus\{0\})$ is the set associated with the map T_r in (A19), and where ∂B denotes the boundary of B. Indeed, this follows from the fact that the intensity measure $\mu(\mathrm{d}t,\mathrm{d}x)$ of the PRM m does not charge with positive mass sets of Lebesgue measure zero. The fact that $\mathbb{P}\{m(\partial B)=0\}=1$ shows that, almost surely, the points $\{(t_k,j_k)\}$ do not lie on the boundary ∂B. Since the points of discontinuity of T_r are at only those measures in M_p with atoms on ∂B, it follows that the map T_r is almost surely continuous when applied to the realizations of the PRM m. Therefore, the continuous mapping theorem (see e.g. Theorem 3.4.3 in [33]) yields

$$T_r\left(\sum_{k=1}^{\infty}\varepsilon_{(k/m,m^{1/\alpha}X(k))}\right)\xrightarrow{d}T_r\left(\sum_{i=0}^{\infty}\sum_{k=1}^{\infty}\varepsilon_{(t_k,j_kc_i)}\right)\quad\text{as } m\to\infty$$

where

$$T_r\left(\sum_{i=0}^{\infty}\sum_{k=1}^{\infty}\varepsilon_{(t_k,j_kc_i)}\right)=\left(\bigvee_{t_k\in(0,1/r]}\bigvee_{i=0}^{\infty}c_ij_k,\ldots,\bigvee_{t_k\in((r-1)/r,1]}\bigvee_{i=0}^{\infty}c_ij_k\right)=:(Z(1),\ldots,Z(r))$$

However, since the intervals $(0,1/r],(1/r,2/r],\ldots,((r-1)/r,1]$ in (A19) do not overlap, the random variables $Z(1),\ldots,Z(r)$ are independent. Moreover, the stationarity (in t) of the intensity of the PRM shows that the $Z(k)$'s are i.i.d. Now, it remains to argue that the $Z(k)$'s have the desired α-Fréchet distribution. This follows as in [30], since for $Z(1)$, for example, we have

$$Z(1)=\bigvee_{t_k\in(0,1/r]}\bigvee_{i=0}^{\infty}c_ij_k=\bigvee_{t_k\in(0,1/r]}(c_+j_k\vee(-c_-)j_k)$$

which in fact equals the extremal process $Y(t)$ therein evaluated at $t=1/r$. □

Proof of Lemma 6

For multivariate max-stable distributions, pairwise independence implies independence (Chapter 5 in [28]). Thus, it suffices to show that $Z(1)$ and $Z(2)$ are independent. The continuous mapping theorem implies that

$$\frac{1}{m^{1/\alpha}}(X_m(1)\vee X_m(2))\xrightarrow{d}Z(1)\vee Z(2)\quad\text{as } m\to\infty$$

We also have that $X_m(1)\vee X_m(2)=X_{2m}(1)$ and since $(2m)^{-1/\alpha}X_{2m}(1)\xrightarrow{d}Z(1)$, as $m\to\infty$, we obtain

$$Z(1)\vee Z(2)\overset{d}{=}2^{1/\alpha}Z(1)\tag{A21}$$

In view of (1), the marginal distributions of Z can only be α-Fréchet. Thus, since $(Z(1),Z(2))$ is a max-stable vector, Proposition 5.11' in [28], implies

$$\mathbb{P}\{Z_1\leqslant x_1,Z_2\leqslant x_2\}=\exp\left\{-\int_0^1\frac{f_1^{\alpha}(u)}{x_1^{\alpha}}\vee\frac{f_2^{\alpha}(u)}{x_2^{\alpha}}\,\mathrm{d}u\right\},\quad x_1,x_2>0\tag{A22}$$

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

for some non-negative functions f_1 and f_2 such that $\int_0^1 f_i^\alpha(u)\,\mathrm{d}u<\infty$, $i=1,2$. Thus, by (A21), for all $x>0$

$$\mathbb{P}\{Z(1)\vee Z(2)\leqslant x\}=\exp\left\{-x^{-\alpha}\int_0^1 f_1^\alpha(u)\vee f_2^\alpha(u)\,\mathrm{d}u\}\mathbb{P}\{Z_1\leqslant 2^{-1/\alpha}x\right\}$$

$$=\exp\left\{-2x^{-\alpha}\int_0^1 f_1^\alpha(u)\,\mathrm{d}u\right\}$$

This, since by stationarity $\int_0^1 f_1^\alpha(u)\,\mathrm{d}u=\int_0^1 f_2^\alpha(u)\,\mathrm{d}u$, yields

$$\int_0^1 f_1^\alpha(u)\vee f_2^\alpha(u)\,\mathrm{d}u=\int_0^1 f_1^\alpha(u)\,\mathrm{d}u+\int_0^1 f_2^\alpha(u)\,\mathrm{d}u$$

The last relation is valid if and only if the non-negative functions $f_1(u)$ and $f_2(u)$ have disjoint supports. This fact, in view of (A22), implies the independence of $Z(1)$ and $Z(2)$. □

ACKNOWLEDGEMENTS

The authors would like to thank an anonymous referee for their detailed comments that led to an improved presentation of the material. We are indebted to the referee for helping us strengthen the result of Lemma 5. The work of S. S. and G. M. was supported in part by NSF grant DMS-0806094.

REFERENCES

1. Zipf G. *Selective Studies and the Principle of Relative Frequency in Language*. Harvard University Press: Cambridge, MA, 1932.
2. de Sousa B, Michailidis G. A diagnostic plot for estimating the tail index of a distribution. *Journal of Computational and Graphical Statistics* 2004; **13**(4):974–995.
3. Mandelbrot BB. The Pareto–Lévy law and the distribution of income. *International Economic Review* 1960; **1**:79–106.
4. Finkenstädt B, Rootzén H (eds). *Extreme Values in Finance*, *Telecommunications*, *and the Environment*. Monographs on Statistics and Applied Probability. vol. 99. Chapman and Hall/CRC: London, New York, 2004.
5. Barabasi A-L. *Linked*: *The New Science of Networks*. Perseus Publishing: Cambridge, MA, U.S.A., 2002.
6. Adler R, Feldman R, Taqqu MS (eds). *A Practical Guide to Heavy Tails*: *Statistical Techniques and Applications*. Birkhäuser: Boston, 1998.
7. Resnick SI. Heavy tail modeling and teletraffic data. *The Annals of Statistics* 1997; **25**:1805–1869. With discussions and rejoinder.
8. Faloutsos M, Faloutsos P, Faloutsos C. On power–law relationships of the Internet topology. *SIGCOMM*. Cambridge, MA, U.S.A., 1999; 251–262.
9. Adamic L, Huberman B. The nature of markets in the world wide web. *Quarterly Journal of Electronic Commerce* 2000; **1**:5–12.
10. Adamic L, Huberman B. Zipf's power law and the Internet. *Glottometrics* 2002; **3**:143–150.
11. Park K, Willinger W (eds.). *Self-similar Network Traffic and Performance Evaluation*. Wiley: New York, 2000.
12. Carlson JM, Doyle J. Highly optimized tolerance: a mechanism for power laws in designed systems. *Physical Review E* 1999; **60**(2):1412–1427.
13. Wharton Research Data Service (url). Wharton School of Management, Universty of Pennsylvania, `https://wrds.wharton.upenn.edu/`.
14. Hill BM. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* 1975; **3**:1163–1174.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb

15. Kratz M, Resnick SI. The qq-estimator and heavy tails. *Stochastic Models* 1996; **12**:699–724.
16. Resnick S, Stărică C. Smoothing the Hill estimator. *Advances in Applied Probability* 1997; **29**(1):271–293.
17. Csörgő S, Deheuvels P, Mason D. Kernel estimates of the tail index of a distribution. *Annals of Statistics* 1985; **13**(3):1050–1077.
18. Feuerverger A, Hall P. Estimating a tail exponent by modeling departure from a Pareto distribution. *Annals of Statistics* 1999; **27**(2):760–781.
19. de Haan L, Drees H, Resnick S. How to make a Hill plot. *Annals of Statistics* 2000; **28**(1):254–274.
20. Cheng S, Peng L. Confidence intervals for the tail index. *Bernoulli* 2001; **7**(5):751–760.
21. Lu J-C, Peng L. Likelihood based confidence intervals for the tail index. *Extremes* 2002; **5**(4):337–352.
22. Resnick S, Stărică C. Consistency of Hill's estimator for dependent data. *Journal of Applied Probability* 1995; **32**:139–167.
23. Stoev S, Michailidis G, Taqqu M. Estimating heavy–tail exponents through max self–similarity. *Technical Report 447*, University of Michigan, 2006.
24. Crovella ME, Taqqu MS. Estimating the heavy tail index from scaling properties. *Methodology and Computing in Applied Probability* 1999; **1**:55–79.
25. Leadbetter MR, Lindgren G, Rootzén H. *Extremes and Related Properties of Random Sequences and Processes*. Springer: New York, 1983.
26. van der Vaart AW. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics, vol. 3. Cambridge University Press: Cambridge, 1998.
27. Hall P. On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society*, *Series B* 1982; **44**:37–42.
28. Resnick SI. *Extreme Values*, *Regular Variation and Point Processes*. Springer: New York, 1987.
29. Mikosch T, Samorodnitsky G. The supremum of a negative drift random walk with dependent heavy-tailed steps. *The Annals of Applied Probability* 2000; **10**(3):1025–1064.
30. Davis RA, Resnick SI. Limit theory for moving averages of random variables with regularly varying tail probabilities. *The Annals of Probability* 1985; **13**(1):179–195.
31. Lo A, Wang J. Trading volume: definitions, data analysis, and implications of portfolio theory. *Review of Financial Studies* 2000; **13**:257–300.
32. Hong H, Wang J. Trading and returns under periodic market closures. *Journal of Finance* 2000; **55**:297–354.
33. Whitt W. *Stochastic–Process Limits*: *An Introduction to Stochastic–Process Limits and Their Application to Queues*. Springer: New York, 2002.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/asmb