Extremes (2016) 19:467–490
DOI 10.1007/s10687-016-0254-4

On the measurement and treatment of extremes in time series

Tucker McElroy[1]

Received: 29 March 2016 / Accepted: 20 April 2016 /
Published online: 14 May 2016

Abstract The paper reviews the topic of extremal time series. The literature documenting the presence of extremes in time series data is first reviewed, followed by a discussion of various probabilistic measures, along with the associated statistical inference problems. The impact of extremes upon statistical analyses is discussed, and the connection to extremal latent components is emphasized. Two data sets illustrate the methods.

Keywords Extremal clustering · Extremal index · Extremogram · Noah effect · Regular variation · Tail index

AMS 2000 Subject Classifications 60G10 · 60G70 · 62M10

1 Introduction

Many time series exhibit anomalies of various species, such as outliers, extremes, and other aberrancies. The frequency and magnitude of such anomalies – as well as the possibility of their clustering – varies depending on the scientific field and the measurement instruments. Some data streams have inherent sources of extremal behavior, such that extreme values are ubiquitous in the observed series, while others may exhibit only isolated outliers, which may have been induced by the vagaries of measurement. Because such aberrancies can greatly distort common statistics – such as sample means and sample autocorrelations – used in time series analysis, there is

✉ Tucker McElroy
tucker.s.mcelroy@census.gov

1 US Census Bureau, Washington, DC, USA

a consensus that detecting, measuring, and accounting for such behavior is vital to obtaining valid statistical results, and that failure to do so can generate spurious conclusions. Moreover, for many applications the frequency and magnitude of extremes themselves is of chief interest (e.g., weather data). These matters are the topic of this review paper: the detection, measurement, and analysis of extremes in time series data.

It is useful to distinguish between phenomena that innately generate extreme observations (e.g., seismic activity, flooding, or economic contagion) and fairly stable phenomena whose record has been contaminated (e.g., by data entry errors, reclassifications, or sampling mechanisms). When it is possible – on the basis of metadata – to distinguish between innate extremes and contaminations, the techniques of detection and measurement can be determined accordingly. However, in some cases the distinction is empirically unclear.

Aberrant values arising from contamination are typically described as *outliers*, and are treated with statistical techniques diverse from those used to analyze proper extremes. The study of *extremal time series* is focused on extreme values that arise as innate facets of a stochastic process. This article focuses on extremal time series, omitting the outlier literature; the scope of this paper is to discuss the analysis of extremal time series data, highlight the best current methods, and direct the interested reader to related literature for deeper study.

2 The presence of extremes in time series data

There is a substantial literature documenting the presence of anomalies in time series data. (We focus on the time series literature, although work on robust estimation goes back at least to Newcomb (1886), as discussed in Stigler (1973)). Foremost was Benoit B. Mandelbrot, who described the presence in hydrological time series – i.e., Nile river flood levels – of both long-range dependence and heavy tails; see Mandelbrot and Wallis (1968, 1969). Here, the term *long-range dependence* refers to non-trivial serial correlation in the time series data, that actually has a sustained and long-term pattern of correlation, while *heavy tails* refers to the slow rate of decay observed in the tail of the marginal probability density function. Mandelbrot also coined the evocative terminology of the *Noah and Joseph* effects, for heavy tails and long-range dependence respectively. This nomenclature was drawn from the Old Testament, referring respectively to a global flood (an extremal event in hydrology) and to periods of persistent famine and abundance in Egypt circa 2000 B.C. (indicative of long-range dependence in agronomy):

> "...all the fountains of the great deep [were] broken up, and the windows of heaven were opened. And the rain was upon the earth forty days and forty nights." (Genesis 7:11-12).

> "Behold, there shall come seven years of great plenty throughout all the land of Egypt: And there shall arise after them seven years of famine..." (Genesis 41:29-30).

Mandelbrot also argued that the Noah and Joseph phenomena were present in communications data (Berger and Mandelbrot 1963; Mandelbrot 1965), economic data (Mandelbrot 1963, 1969), and more broadly were present throughout the natural sciences (Mandelbrot 1974, 1983). Beginning with Fama (1965), which asserted the heavy-tailed nature of stock market prices, there were a few articles published on the topic of heavy tails in financial returns, such as Clark (1973) and Epps and Epps (1976), signaling an increasing interest among economists. While the financial motivation for such investigations has been the control of risk (and minimization of financial losses), in the environmental sciences interest has principally focused on extremes because of the direct impact upon society (e.g., storm damage generated by extremes in wind and precipitation time series).

By the 1990s there was a greatly increased body of literature on time series extremes. Castillo (1988) documented the presence of heavy tails in hydrology; Koedijk et al. (1990) studied extremal data in economics and finance; Duffy et al. (1993, 1994) as well as Meier-Hellstern et al. (1991) considered heavy-tailed teletraffic data; Grigoriu (1995) considered applications of heavy-tailed processes to reliability and structural engineering problems. Willinger et al. (1997) studied the connection between high variability (extreme values) and self-similarity (a stochastic characterization of long-range dependence) in teletraffic data, relating the Noah effect parameters (in traffic waiting times) to Joseph effect parameters (long-range dependence in aggregate network traffic). Also see Beran et al. (1995), Willinger et al. (1995), and Heath et al. (1997) for related work.

In the late 1990s, the Internet was just coming under scientific study, and the empirical connection between Noah and Joseph effects in teletraffic time series then prompted much of the subsequent mathematical research into extremal stochastic processes over the succeeding two decades. Sidney Resnick claimed that "...such phenomena as file lengths, cpu time to complete a job, call holding times, interarrival times between packets in a network and lengths of on/off cycles appear to be generated by distributions which have heavy tails" (Resnick 1997, p.1805). Subsequent articles on teletraffic data include Crovella et al. (1998), Cappe et al. (2002), and D'Auria and Resnick (2006, 2008). The book of Resnick (2007) extends this discussion, and also explores applications of extremal theory to the field of insurance.

Generally speaking, there has been an explosion of articles – since the turn of the century – treating extremal time series in the natural and social sciences; here we mention only a few references that have a statistical orientation. Insurance claims, for example, are directly connected to the uncertainties of life on this planet, and given that insurance is marketed as a buffer against severe calamities, extremes must be accounted for in the stochastic structure of insurance models. The book of Reiss and Thomas (1997) provides illustrations arising in insurance, finance, and hydrology, and discusses the *XTREMES* software; the volume of Coles (2001) provides some examples from the natural sciences. Anderson and Meerschaert (1998) examined river flow with heavy-tailed distributions, and Tesfaye et al. (2006) considered seasonality as well. Weller et al. (2012) applied extreme-value methodology to spatial precipitation data; also, Thibaud et al. (2013) considered an application to hydrology of spatial extremal processes.

Turning to economics, it is natural to expect anomalies in markets, as human behavior is closely dependent upon natural disasters (e.g., drought, hurricane, earthquake) or man-made debacles (e.g., terrorism, war, embargo, piracy). In recent literature, Mikosch (2003) discussed difficulties with the assertion of Mandelbrot (1969) that financial data follow a stable distribution. Farmer et al. (2004) showed that heavy tails are intrinsic to markets in general, and Iglesias (2012) applied extreme value ideas to the study of the foreign exchange market. The presence of extremes – as well as long-range dependence – in trading volume is considered in Rossi and Santucci de Magistris (2013), and extremal values in shortfall is studied in Linton and Xiao (2013). Hill and Shneyerov (2013) considered heavy tails in cross-sectional auction bid data, and Davison et al. (2013) discussed the presence of extremes in temperature (e.g., heatwaves) and precipitation (e.g., flooding), with concomitant effects on food security.

If one is not restricting the viewpoint to time series processses, the amount of literature expands greatly to include insurance and catastrophe data (Beirlant et al. 2004), the classical province of extreme value theory (Embrechts et al. 1997), as well as demographics – infant birth weight data was studied by Chernozhukov and Fernández-Val (2011). Ruin theory and the calculation of Value at Risk (VaR) has a long history, though the early literature focused on techniques for non-serially correlated data (McNeil and Frey 2000, as well as Xiao and Koenker 2009, consider approaches involving extreme value theory for time series). We mention that a few of the above citations – Weller et al. (2012) and Thibaud et al. (2013) – discuss extremal spatial processes; this is a nascent field, but given Mandelbrot's prophetic prescription for the role of extreme value theory in the natural sciences, it seems an entirely plausible arena of application.

As a concrete example of an extremal time series, we display in Fig. 1 a time series obtained by counting the number of packets arriving in one-second intervals. Known as the OctExt Ethernet trace, this is integer-valued data, characterized by frequent

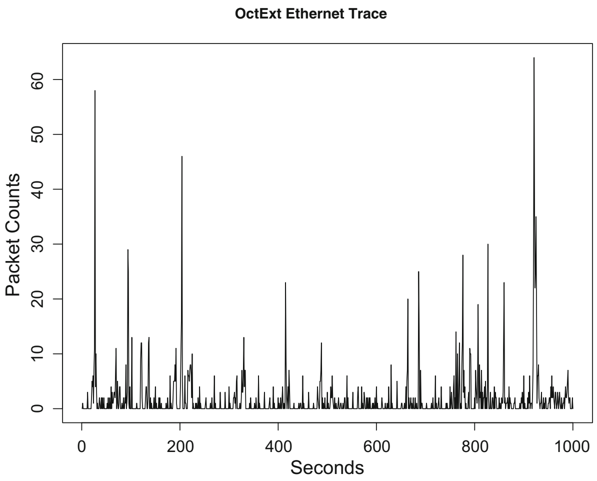

Fig. 1 OctExt Ethernet trace (Source: Bellcore Morristown Research and Engineering facility, 1994)

zero values and large extremes. We'll return to this example in Section 3, to illustrate some of the inferential methods of extremal time series.

As a second illustration, consider Fig. 2, which displays the average quarterly price of the West Texas Intermediate standard blend of crude oil (Source: Department of Energy), from 1970.Q1 through 2005.Q4. This series, which was studied in Trimbur (2010), exhibits extremes in its growth rate (Phillips 1990). The dramatic movements in the series could be explained in terms of realignments in the oil market. If we believe that anomalies are the result of a singular market shock, then they can be treated as a special case treated with outlier methodologies; if instead we believe that such upsets are a natural feature of commodity markets, where human beings are driven by greed and fear, then a stochastic approach is more appropriate. We explore the impact of extremes upon understanding oil price trends in Section 4.

3 Measurement and estimation of extremes

If the data anomalies are viewed as part of the stochastic structure of the time series, then the use of fixed effects to model them will artificially thin the tails of the marginal distribution, and potentially lead to understatements of the true kurtosis – with ramifications to quantiles and forecast uncertainty. Given that extreme events (such as a terrorist attack or natural disaster) may recur in the future, it is misleading to exclude these data from quantifications of extremal dependence. In favor of this view, there is a substantial literature extending the study of extremes to the context of time series. We first review the popular stochastic structures, and then discuss the corresponding statistical inference problems.

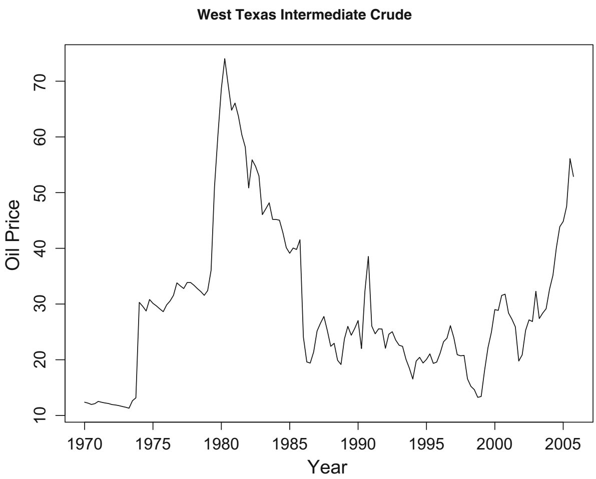

Fig. 2 Price of West Texas Intermediate standard blend of crude oil (Source: Department of Energy), in US Dollars per barrel; adjusted (base year 2000) using the GDP deflator (Source: Bureau of Economic Analysis) as a price index

3.1 Measures of extremal dependence

There are several measures of extremal dependence considered in the literature, which are reviewed and evaluated in Fasen et al. (2010); we begin by discussing the clustering of extremes. The empirical phenomenon of outlier patches (Martin and Yohai 1986) has been formalized probabilistically via the concept of *extremal clustering* – often referred to simply as "extremal dependence" – in Leadbetter (1983). Chapter 8 of Embrechts et al. (1997) provides a historical overview; early work includes Leadbetter (1974), which generalizes the contributions of Sibuya (1959) and Loynes (1965). Extremal clustering means that extreme values are likely to occur in temporal proximity (recall the Joseph effect, discussed in Section 2). The work of Leadbetter et al. (1983) provides a flexible framework (encompassing many stochastic models used for time series data) for defining extremal clustering, using the concept of "exceedances". Also see Leadbetter and Rootzén (1988).

We follow the discussion in Davis et al. (2013), which provides an excellent overview of the topic of extremal clustering. Given a strictly stationary discrete time stochastic process $\{X_t\}$ with marginal cumalitive distribution function (cdf) F, we suppose the existence of a sequence a_n such that

$$n\,(1 - F(a_n)) \to 1 \qquad (1)$$

as $n \to \infty$. For example, if the marginal distribution has a Pareto-like right tail, i.e., $\mathbb{P}[X_0 > x] \sim C\,x^{-\alpha}$ as $x \to \infty$, for some index $\alpha > 0$ and constant $C > 0$, then (1) is satisfied with $a_n = C^{1/\alpha} n^{1/\alpha}$. More generally, a_n can be given (non-uniquely) by the right-tail quantile corresponding to the fraction $1/n$.

One can assess the incidence and clustering of extremes by simply counting extremal phenomena – this is the idea behind the point process of exceedances. With 1_A denoting the indicator function for a set A, given a sample $X_1, X_2, \cdots, X_n$ we can define the point process of exceedances as a random measure N_n on the Borel sets of (0, 1], such that

$$N_n(B) = \sum_{t=1}^{n} 1_{[t/n \in B,\, X_t > a_n]}. \qquad (2)$$

By altering the set B, it is possible to examine subsets of indices t for which extremal phenomena occur. According to Hsing et al. (1988), under a weak dependence condition the sequence of extremal measures N_n converge weakly to a homogeneous compound Poisson process with intensity θ, where the limiting cluster size (a discrete random variable) has distribution π. Necessarily $\theta \geq 0$, but it can also be shown that $\theta \leq 1$ must be true; θ is called the *extremal index*, and is defined in Leadbetter (1983) – although the concept appears in prior literature, such as Loynes (1965).

In the case of an i.i.d. process, the extremal index equals one ($\theta = 1$). Under some mild assumptions (Rootzén 1978, 2009) the reciprocal of the extremal index equals the mean of π, so that greater dependence (small θ) indicates a higher probability of clustering, with $\theta = 0$ corresponding to long-range dependence. However, $\theta = 1$ can also occur with non-i.i.d. data; Leadbetter et al. (1983) showed that Gaussian time series with even a very substantial pattern of serial correlation have unit extremal index.

Markovich (2014) studied various properties of extremal clustering. Davis et al. (2013) discussed the explicit form of θ for a variety of particular stochastic processes $\{X_t\}$, including linear processes, Markov processes, GARCH processes, and Stochastic Volatility (SV) models. Nevertheless, the expressions for the extremal index are typically too complicated to allow for parametric inference.

Of separate interest is the *tail index* α, which quantifies the likelihood of obtaining an anomalous value – separately from the question of clustering. Again with F denoting the cdf of the marginal X_0, a general definition is implicitly given via

$$1 - F(x) = x^{-\alpha} L_+(x) \qquad F(-x) = x^{-\alpha} L_-(x) \tag{3}$$

for $x > 0$, where L_+ and L_- are slowly-varying functions. A slowly-varying function L has the property that $L(tx)/L(x) \to 1$ as $x \to \infty$ for all $t > 0$ (Embrechts et al. 1997). (One can broaden the definition, to allow for different α for the right and left tails.) The parameter α is called the *tail index*.

There is a vast literature on tail index estimation in the i.i.d. case, and many articles investigate the related *extreme-value index* (not to be confused with the extremal index) defined as the reciprocal of α. There is also a substantial literature on tail index estimation from time series data, where it is acknowledged that serial dependence will impact inference. Interest in the tail index stems from its direct connection to the incidence of extremes, which in turn affects VaR (Fasen et al. 2014), forecasting, and moments (e.g., expected claim size in insurance applications). As in the case of the Noah effect, heavy-tailed time series with a small α ($\alpha < 2$) have infinite variance, which interferes with the classical Hilbert space methodology used for extrapolation and signal extraction.

In fact, the autocorrelation function can be quite misleading for extremal time series. In Jach et al. (2012) a Noah-Joseph (NJ) process was studied, which exhibited both infinite variance and long memory (in particular, having finite autocovariances); letting $\{Z_t\}$ be a Gaussian process with arbitrary serial correlation and letting $\{S_t\}$ be an i.i.d. positive process with tail index $\alpha/2$, it follows that

$$X_t = \sqrt{S_t} \cdot Z_t \tag{4}$$

defines a process $\{X_t\}$ satisfying (3) with autocovariance function well-defined at non-zero lags. If $\alpha > 2$ the variance is also finite, but if $\alpha < 2$ the so-called NJ process has infinite variance, implying that the sample autocorrelation function resembles that of white noise. For such a process (which embodies both of the features that Mandelbrot observed in nature) the sample autocorrelations are misleading.

Along these lines, Davis and Mikosch (2009b) in the context of a GARCH process stated that "the autocorrelation function provides little insight into the dependence structure of the process," and the authors go on to discuss difficulties with assessing serial dependence; also see Davis and Mikosch (2009a), as well as Hill (2009) for work on cross-correlations of exceedances. This discussion motivates the *extremogram* of Davis and Mikosch (2009b), defined as

$$\gamma_{A\,B}(h) = \lim_{n\to\infty} n \,\mathrm{Cov}\left(1_{[a_n^{-1} X_0 \in A]}, 1_{[a_n^{-1} X_h \in B]}\right) \tag{5}$$

for integer h, where the sets A and B should be bounded away from zero. Here the sequence a_n must be suitably chosen, and in practice is typically taken to satisfy (1) in absolute values of the random variable, i.e., $n\,\mathbb{P}[|X_0| > a_n] \to 1$. Then it follows that with the choice of $A = B = (1, \infty)$ in Eq. 5 that

$$\rho(h) = \gamma_{(1,\infty)\,(1,\infty)}(h) = \lim_{x\to\infty} \mathbb{P}[X_h > x | X_0 > x], \tag{6}$$

which is a more intuitive expression than (5). The *coefficient of tail dependence* (Ledford and Tawn 2003) is a separate – and in some sense, "orthogonal" – concept from the extremogram, being defined through the expression

$$\mathbb{P}[X_h > x | X_0 > x] \sim x^{1-1/\eta_h}\, L(x), \tag{7}$$

where L is slowly-varying. Ledford and Tawn (2003) define $\Lambda_h = 2\eta_h - 1 \in (-1, 1]$ as their measure of extremal dependence (i.e., the coefficient of tail dependence). While a nonzero $\rho(h)$ in Eq. 6 corresponds to $\Lambda_h = 1$, or full extremal dependence, a value $\Lambda_h < 1$ for all h implies the extremogram resembles the autocorrelation function of white noise.

Other measures of extremal dependence include the *extreme dependence measure* of Resnick (2004), and the *extremal coefficient function* of Fasen et al. (2010). An overview and comparison of these measures is discussed in Larsson and Resnick (2012); also see Hill (2011a). Hill (2011b) makes the point that the extremal index and extremogram only capture linear tail dependence, and the study of nonlinear tail dependence requires other tools.

Many more examples involving the extremogram are provided in Davis et al. (2013); the authors also mentioned that *regular variation* provides a sufficient paradigm for which the extremogram is well-defined. A time series has regular variation if all its finite-dimensional distributions possess the regular variation property of random vectors. This requires that for any random vector $X = [X_1, X_2, \cdots, X_d]'$ sampled from the time series $\{X_t\}$, with $x > 0$

$$\mathbb{P}[\|X\| > x] = x^{-\alpha}\, L(x) \tag{8}$$

for a slowly-varying function L and some vector norm $\|\cdot\|$ on $\mathbb{R}^d$, and also that

$$\mathbb{P}\,[X/\|X\| \in \cdot \,|\, \|X\| > x] \xrightarrow{w} \mathbb{P}[\Theta \in \cdot],$$

where the limit is taken as $x \to \infty$ and is a weak convergence over the Borel sets of the unit sphere (Θ is a random vector supported on the unit sphere in $\mathbb{R}^d$). Resnick (1997) summarized the basic theory of regularly varying processes; as such processes are heavy-tailed, they constitute time series for which tail index, extremal index, and extremogram are all valid concepts. Davis et al. (2013) provided a useful list of commonly studied regularly varying processes, with a discussion of their associated extremogram $\rho(h)$, including linear processes, stochastic recurrence equations, SV models, GARCH processes, α-stable processes, and max-stable processes.

Some of these concepts can be extended to continous-time processes, to the spatial context, or even to space-time processes. Schlather (2002) discussed max-stable random fields, based upon the root idea of a max-stable process – essentially the weak limit of the maxima of a collection of i.i.d. processes (de Haan 1984). Several stationary spatial models were proposed in de Haan and Pereira (2006), and the consistency of estimators was established. Some other constructions were considered in Kabluchko (2009) and Kabluchko et al. (2009). This work has been extended to the case of a space-time process by Davis et al. (2013a); also see Davis et al. (2013b) and Huser and Davison (2014).

3.2 Inference for extremal measures

In the previous subsection we discussed the extremal index, the tail index, and the extremogram. The first two objects are parameters that are well-defined for a wide class of stochastic processes; while both the extremal index and tail index can be defined for some i.i.d. data, the former becomes trivial ($\theta = 1$ always). The tail index is still a meaningful parameter when serial dependence is present, but inference becomes more challenging; moreover, its estimation is affected by extremal clustering. The extremogram, on the other hand, is a broad measure of extremal dependence, which is well-defined for regularly varying processes. We discuss inference for each in order; for brevity we omit discussion of other measures, such as the extreme dependence measure.

Leadbetter and Nandagopalan (1989) introduced a nonparametric estimator of θ, later extended by Nandagopalan (1990), defined as

$$\widehat{\theta} = \frac{\sum_{t=1}^{n-1} 1_{[X_t \leq u_n < X_{t+1}]}}{\sum_{t=1}^{n} 1_{[X_t > u_n]}} \tag{9}$$

for an exceedance threshold u_n satisfying

$$n\,(1 - F(u_n)) \to \infty. \tag{10}$$

This condition can essentially be viewed as taking $u_n = o(a_n)$, with a_n defined via (1); adopting (10) yields a consistent estimator of θ under some weak dependence assumptions. The choice of u_n has been studied in Davision and Smith (1990), which also provides a treatment of modeling, estimation, and diagnosis for size and occurence of exceedances over thresholds. More recently, Gomes et al. (2008) studied a reduced bias version of Eq. 9, and utilized subsampling (Politis et al. 1999) methods to improve performance.

Smith and Weissman (1994) provided an overview of some early literature on estimation, pointing out that the "runs" estimator of Smith (1989) is more general than that of Nandagopalan (1990). Given a tuning parameter r_n (the length of the runs) chosen by the practitioner, the runs estimator is defined via

$$\overline{\theta} = \frac{\sum_{t=1}^{n-r_n} 1_{[X_t > u_n]} \cdot 1_{[X_{t+1} \leq u_n]} \cdots 1_{[X_{t+r_n} \leq u_n]}}{\sum_{t=1}^{n} 1_{[X_t > u_n]}}. \tag{11}$$

Clearly $\overline{\theta}$ with the choice $r_n = 1$ is identical with $\widehat{\theta}$.

Concurrently, Hsing (1991a) considered estimation of certain functionals of the empirical cluster size distribution, and studied applications of these results to conducting inference for the extremal index. Later Hsing (1993) proposed an adaptive inference procedure for the extremal index. Another approach to modeling thresholds is given in Smith et al. (1997), which considers a Markovian time series structure and utilizes bivariate extreme value theory to propose transition distributions.

More recently, Ledford and Tawn (2003) proposed diagnostic tools to evaluate the serial dependence assumptions utilized in the analysis of extremal processes. Also Ferro and Segers (2003) considered the impact of a de-clustering scheme upon clustering characteristics, and proposed an automated de-clustering algorithm; bootstrap methods were advanced for estimating the variability. This work was later extended by Robert (2009a), which established asymptotic results for the interval estimators of Ferro and Segers (2003). Also see Robert et al. (2009) and Robert (2009b).

There has been considerably more literature on tail index inference; methods originally developed for i.i.d. data have been applied to weakly dependent time series. While for short time series (e.g., the oil series of Fig. 2) a parametric specification may be attractive, for longer time series (e.g., the OctExt Ethernet trace of Fig. 1) the parametric approach is less feasible – the number of parameters required to explain the marginal structure becomes intractable. As argued in Drees (2008), the parametric model misspecification can be quite damaging in the context of tail index estimation. Hence, semi-parametric and nonparametric approaches to extreme value estimation have become the norm for longer extremal time series.

Consider (3). This formulation for the marginal distribution describes the tail index parameter α, while allowing for the presence of a nuisance function L. A parametric approach attempts to provide an exact specification for L (many potential distributions, such as Pareto and Student t, are discussed in Embrechts et al. 1997), while a semi-parametric approach might adopt an infinite basis expansion, allowing for more basis functions as sample size expands. Nonparametric approaches do not take the form of L into account in construction of the estimator, but are designed to be generally effective (e.g., consistent) for a variety of nuisance functions.

The Hill estimator (Hill 1975), which has become a benchmark estimator of the tail index by virtue of its early formulation and good statistical properties, is given by

$$\widehat{\alpha}^{(k)}_{HILL} = \left(\frac{1}{k} \sum_{j=1}^{k} \log X_{(j)} - \log X_{(k)} \right)^{-1} \tag{12}$$

in the case of positive random variables (if negative values are present, extensions are possible), where $X_{(j)}$ is the jth upper order statistic of the sample $X_1, X_2, \cdots, X_n$. For theoretical derivations, it is assumed that $1/k + k/n \to 0$ as $n \to \infty$, with the perspective that $k = k(n)$ is a function of sample size. In practice, k is chosen as some small fraction of n; see Drees and Kaufmann (1998) for practical recommendations, and McElroy and Nagaraja (2016) for a discussion. An excellent recent review of the i.i.d. case is Gomes and Guillou (2015), in which many other tail index estimators are discussed. However, there seems to be no systematic investigation of the selection of k when serial dependence is present.

Hsing (1991b) extended the asymptotic normality of the Hill estimator to the case of serially correlated data. This paper seems to be a chronological outlier; interest in extremal time series blossomed at the turn of the century, as more empirical literature appeared that documented the coincidence of heavy tails and serial correlation in teletraffic data (Resnick and Stărică 1995, 1998). The results for the Hill estimator were broadened by Drees (2000), who studied the tail empirical process – a measure of exceedances not unlike the point process measure (2). This work was extended in Drees (2002), and applied to quantile estimators (Csörgö and Yu 1996) – used in VaR applications – in Drees (2003). Hill (2009) established convergence results for tail arrays, thereby broadening the class of time series for which the Hill estimator is effective; also see Hill (2010). Brito and Freitas (2010) considered an estimator similar to Hill's (1975), and established its validity under serial dependence.

A feature of Drees' work is the wide applicability of empirical tail process results to a broad class of tail index estimators. Further theoretical work in this direction includes Rootzén (2009) and Drees and Rootzén (2010), which extended the results on tail empirical processes, and focused upon studying extremal clustering. The impact of long-range dependence on the empirical tail process was considered in Kulik and Soulier (2011), and was extended by McElroy and Jach (2012a) to NJ time series.

The Hill and related estimators are semi-parametric, in that selection of a tuning parameter k is required, where its optimal selection ultimately depends upon the unknown structure of L. (Another semi-parametric approach based on quantiles is given in Holan and McElroy 2010). Turning to nonparametric estimators, Meerschaert and Scheffler (1998) proposed tail index estimation based upon sample moments (and not order statistics), which utilized the fact that for $\alpha \in (0, 2)$ a time series with marginal structure given by Eq. 3 obeys a non-central limit theorem. In particular,

$$\widehat{\alpha}_{MS} = 2\,\frac{\log n}{\log\left(\sum_{t=1}^{n} X_t^2\right)} \tag{13}$$

utilizes the result that the sample second moment for such heavy-tailed processes is bounded in probability of order $n^{-2/\alpha}L^2(n)$. In contrast to Eq. 12, it does not require the choice of a tuning parameter. This idea was extended in Politis (2002) and McElroy and Politis (2007c).

In the case of a GARCH process, a moment-based estimator was proposed by Berkes et al. (2003). The theory of Kesten (1973) on stochastic recurrence relations provided the framework to establish (Mikosch and Stărică 2000) that stochastic volatility can be associated with heavy-tailed marginals – also see Basrak et al. (2002). Such a result indicates that heavy tails may be manifested in time series data via infinite moments – and not by visually salient bursts (cf., Fig. 1) – and hence moment-based methods may be useful, in contrast with methodology based upon order statistics.

Another idea in the detection of extremes is based on the question: how would the data change if the extreme were omitted? This query motivates the outlier detection literature, but the same thinking can be applied to stochastic extremes by windowing the data. McElroy and Politis (2007b) introduced the notion of a *scan*, which allows one to study the growth rate of a simple statistic – such as the sample second moment underlying the MS statistic (13) – as a function of sample size. A related idea is the *max self-similar* concept of Stoev et al. (2011), which is based on the scaling property of block maxima (also see Hamidieh et al. 2009).

Turning to estimation of the extremogram, there is a sparser literature due to the relative novelty of the concept – see Larsson and Resnick (2012). Davis et al.

(2012) remarked on the difficulties attendant on developing credible intervals for the extremogram, and employ the stationary bootstrap to meet these challenges. Davis et al. (2013) establish consistency and asymptotic normality under a condition similar to Eq. 10. Mikosch and Zhao (2014) considered estimation of the Fourier transform of the extremogram (6), viewing this as a spectral quantity yielding frequency information about extremes; their estimator is a periodogram of extreme events. They extended this to integrated functionals of this rare-event periodogram in Mikosch and Zhao (2015).

By placing the regular variation on $\{Z_t\}$ instead of $\{S_t\}$ in Eq. 4, and allowing the volatility $\{S_t\}$ to have short- or long-range dependence, one obtains a SV process that is often utilized to model financial log returns. Kulik and Soulier (2011, 2013) studied such a process, for which the extremal index is $\theta = 1$ and the extremogram is that of white noise (Davis et al. 2013); Kulik and Soulier (2013) showed that measures of conditional extremal events are nontrivial in such a case, and long memory in the volatility has an impact on their asymptotics. These results were further refined in Kulik and Soulier (2015), through their study of the *conditional extreme value*.

In all of these inference problems, limiting distributions may be non-pivotal, and/or may converge at unknown rates. While self-normalization (McElroy and Politis 2002) has been advanced to address the second problem, a variety of resampling methods have been proposed to obtain asymptotic quantiles. Lahiri (2003) provided a comprehensive background for resampling and subsampling techniques, with some discussion of the bootstrap methodology for extremal time series. Ferro and Segers (2003) utilized the bootstrap to measure the variability of their estimates of the extremal index. The work of Doukhan et al. (2011) – with comments and discussion – advances the application of subsampling methodology to extremal time series, with a focus on inference problems. Davis et al. (2012) and Mikosch and Zhao (2015) applied the bootstrap to conduct inference for the extremogram. Applications are considered in Li et al. (2008) – who considered the bootstrap and jack-knife on oceanographic data – and in Kyselý (2008), which gave an empirical study of parametric and nonparametric bootstrap performance on extreme-value data.

Beyond problems of inference, resampling/subsampling methods have been used to determine tuning parameters and assist with extremal estimation. Danielsson et al. (2001) applied the bootstrap to address the problem of estimating the tuning parameter k in the Hill estimator (12). A similar idea is employed in Pandey et al. (2003) for extreme quantile inference. Gomes et al. (2008) applied subsampling and the jack-knife to assist with estimation, while Gomes et al. (2012) utilized resampling methods to obtain an adaptive estimate of the extreme value index.

3.3 Application to OctExt data

Returning to the OctExt Ethernet Data displayed in Fig. 1, we proceed to apply some of the above statistical methods. While the plot shows marked evidence of extremes, it also seems plausible that the extremes are clustered. In order to estimate the extremal index via (9) or (11), it is necessary to specify u_n appropriately. The discussion above indicates that we could take $u_n = o(n^{1/\alpha})$, assuming some estimate of the tail index α. Utilizing Eq. 12 for the Hill estimator, with k chosen to be small relative to n, should yield an estimate. The choices of $k = [n^{2/3}]$, $k = [n^{1/2}]$, and $k = [n^{1/3}]$ with $n = 1000$ yield tuning parameters of 100, 31, and 10 respectively, with

$$\widehat{\alpha}_{HILL}^{(100)} = 1.53, \qquad \widehat{\alpha}_{HILL}^{(31)} = 1.85, \qquad \widehat{\alpha}_{HILL}^{(10)} = 2.71.$$

This demonstrates the general feature that the estimate greatly depends upon the tuning parameter. If we repeat the exercise with a range of tuning parameters, say for k between 9 and 500 (half the sample size), we obtain a plot of Hill estimates that is quite variable – referred to by some as a Hill horror plot; see Fig. 3. Without recourse to any of the more nuanced techniques mentioned above, we obtain merely a vague confirmation that the data is heavy-tailed, with α roughly between 1 and 3.

An alternative to the Hill estimator is to utilize the moment-based estimator (13), which yields the estimate $\widehat{\alpha}_{MS} = 1.35$; this choice yields as an upper bound for u_n the number $\lfloor n^{1/\widehat{\alpha}_{MS}} \rfloor = 107$, which exceeds the maximum of the data – if we computed $\widehat{\theta}$ on this basis, we'd obtain the value one. Clearly, we must take u_n much smaller. With the largest Hill estimate we obtain $\lfloor n^{1/\widehat{\alpha}_{HILL}^{(10)}} \rfloor = 12$, and the choice of $u_n = 12$ yields 25 exceedances. Computing (11) over a range of r_n (from 1 up to 248, after which $\overline{\theta}$ is zero) yields estimates of the extremal index ranging from about .6 down to zero; see Fig. 4. These methods give some rough guesses about the extreme value behavior in the time series, but there is little confidence associated with any single point estimate (without a deeper analysis) due to the high sensitivity of results to tuning parameters.

4 The impact of extremes

Given that extremes have been identified and measured (e.g., via estimation of the tail index parameter), we may wish to account for the impact of extremes. Suppose that we are interested in VaR for financial data, or wish to compute a ruin probability

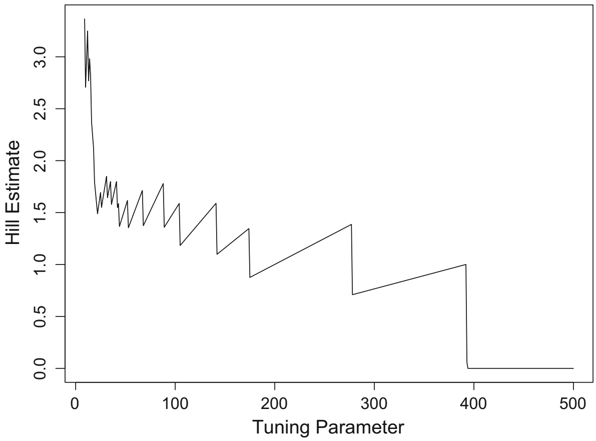

Fig. 3 Hill estimate for OctExt Data, plotted as a function of the tuning parameter k

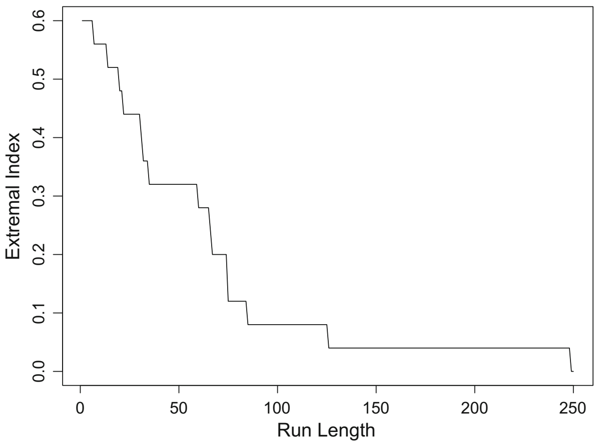

Fig. 4 Extremal Index for OctExt Data, plotted as a function of the run length r_n, with threshold $u_n = 12$

for insurance claim data, or wish to determine the likelihood of a catastrophic event: these scenarios essentially require knowledge of the marginal distribution of the time series, so that a now-cast or fore-cast that accounts for the extremal nature can be accurately computed. For such applications, measurement of the extremal parameters – perhaps followed by a fuller parametric description of the marginal distribution – is immediately connected to the goals of analysis. See Embrechts et al. (1997) for an extensive discussion of actuarial and financial applications.

There is a substantial literature that describes the impact of heavy-tailed marginal distributions on common statistics, such as sample moments, in the context of serially dependent data. Davis and Resnick (1985, 1986) explored sample means and autocovariances for linear process with α-stable errors, and numerous articles have expanded these early efforts. Kokoszka and Taqqu (1996, 1999) extended these early results to the case of long-range dependence, and Basrak et al. (2012) established limit theorems for empirical processes of cluster functionals. Also Kulik and Soulier (2012) provided nice results for sample autocovariances of heavy tailed stochastic volatility models with long memory. Anderson and Meerschaert (1997) considered the seasonal case, while Meerschaert and Scheffler (2001) made extensions to multivariate heavy-tailed time series. Horváth and Kokoszka (2008) further developed inference for sample autocorrelations in extremal time series; similarly, inference problems for NJ processes (4) were considered in McElroy and Politis (2007d), Jach et al. (2012), and McElroy and Jach 2012b. McElroy and Politis (2007a) studied inference for the mean, in the context of infinite variance data observed at irregular spatial observations.

Beyond the work of Kokoszka and Taqqu (1996, 1999), there are a few articles on fitting extremal time series models. Hill (2013a) studied the fitting of infinite variance autoregressions; Hill (2015) considered fitting the GARCH model to financial data, allowing for heavy-tailed (and asymmetric) errors. Hill and Aguilar (2013) considered moment-based score statistics when the time series data has infinite variance, with applications to volatility spillover in the financial literature on market contagion;

also see Aguilar and Hill (2015). In terms of spatial inference, Einmahl et al. (2015) studied the fitting of high-dimensional spatial data with application to the analysis of wind speed.

There are many applications that make direct use of the measures of extremal dependence. For example, the VaR problem for extremal data has been considered in Gencay and Selcuk (2004) and Gilli and Këllezi (2006). Suppose the lower quantile q_p is defined such that $1 - F(q_p) = p \in (0, 1)$; the VaR is defined as $-q_p > 0$, with probability p of a loss $-X_t$ exceeding the threshold $-q_p$. The expected shortfall is defined as the expected loss given it exceeds the VaR; the problem of estimating expected shortfall is approached nonparametrically in Hill (2013b). Also see Linton and Xiao (2013).

Bollerslev et al. (2013) applied extreme value theory to estimate extreme jumps in financial asset prices. There can be systematic or idiosyncratic jumps in such data, and their article shows these jumps to be heavy-tailed, and assesses extreme dependencies in these jump tails. Also see Rossi and Santucci de Magistris (2013). Another interesting application is to the testing of common values in first-price auctions, where the Hill estimator is directly employed (Hill and Shneyerov 2013).

In other contexts, extreme values are merely a nuisance, and it is desirable to remove – or control for – the extreme effects, so that we may proceed to other goals of analysis. For example, we may wish to forecast the trend or cyclical movements in a time series, first accounting for extremes (e.g., consider Fig. 2). Or we may wish to seasonally adjust agricultural yield data, where only the seasonal component exhibits extremes. Or perhaps the goal is to estimate the real volatility of a financial asset, and the series of volatility estimates need to be cleaned of extremes before understanding the serial dependence structure.

For forecasting and signal extraction problems, there are two chief paradigms: first, we might attempt to remove the extreme values and then proceed with a methodology that is optimal for light-tailed time series data. Second, we might leave the extremes and instead utilize a methodology that is robust with respect to the presence of extremes. In the first approach, if the extremes have been treated as outliers – via the device of using regressors – then the regression-adjusted data will generally be lighter-tailed (discussed in Findley et al. 1998). If instead the extremes are viewed as being stochastic (appropriate for more systemic extreme-value behavior) then extreme-value adjustment takes the form of shrinkage. In this case, we might develop forecasting or signal extraction methods that directly account for data or latent components that are heavy-tailed.

We develop this discussion of extremal signal extraction somewhat more deeply. Consider the framework

$$Y_t = X_t + E_t, \tag{14}$$

where $\{X_t\}$ represents an unobserved stochastic trend, while $\{E_t\}$ is i.i.d. noise. The ℓ_1 Hodrick-Prescott (HP) trend of Kim et al. (2009) is obtained by finding the minimizers $X_1, X_2, \cdots, X_n$ of

$$\sum_{t=1}^{n} (Y_t - X_t)^2 + \lambda \sum_{t=2}^{n-1} |X_{t-1} - 2X_t + X_{t+1}| \tag{15}$$

for some $\lambda \geq 0$. The usual Hodrick-Prescott trend is obtained if the absolute value in the second term of Eq. 15 is instead a square; the solution is the conditional expectation of trend given data when $\{X_t\}$ and $\{E_t\}$ are Gaussian and the trend is an integrated random walk (McElroy 2008). Interestingly, Yamada and Jin (2013) showed that the criterion (15) corresponds to Gaussian noise with an integrated random walk trend drive by the double exponential distribution.

More generally, let $W_t = (1-B)^2 X_t$ be i.i.d. of probability density f, potentially corresponding to a regularly varying distribution (3), and suppose that $\{W_t\}$ is independent of X_1 and X_2. With $\{E_t\}$ i.i.d. Gaussian (and independent of the trend) with variance σ^2, the conditional probability density of $X_1, X_2, \cdots, X_n$ given $Y_1, Y_2, \cdots, Y_n$ is proportional to

$$\prod_{t=3}^{n} f(W_t) \cdot (2\pi\sigma^2)^{-n/2} \exp\left\{ -\sum_{t=1}^{n} \frac{(Y_t - X_t)^2}{2\sigma^2} \right\}.$$

After eliminating constants, $2\sigma^2$ times the negative log of this conditional probability density equals (15) in the case that f is double-exponential, identifying λ with $2\sigma^2$. This indicates that the trend estimator corresponds to the mode of the conditional density of trend given data.

It is an interesting open problem to study the features of trend estimates arising from regularly varying trend innovations – or to allocate the heavy tails to the noise $\{E_t\}$ instead. Heavy-tailed trend innovations may be suitable for modeling stochastic level shifts, whereas heavy-tailed noise is appropriate for idiosyncratic extremal events; Trimbur (2010) considered a student t distribution for both stochastic components. For a general f, criterion (15) becomes modified to

$$\sum_{t=1}^{n} (Y_t - X_t)^2 - \lambda \sum_{t=3}^{n} \log f(W_t), \qquad (16)$$

and this can be applied in the case of a heavy-tailed distribution. For example, when f is student t with α degrees of freedom, then $-\log f(x) \propto .5(1+\alpha)\log(1+\alpha^{-1}x^2)$, and $\lambda = \sigma^2(1+\alpha)$.

More generally, there is scope for applications that synthesize the error-structure form of Eq. 14 with the stochastic approach to extremal time series described in Section 3. Rendering time series (and spatial) data analysis robust with respect to anomalies will significantly broaden the impact of extreme value theory. Developing a disciplined scientific procedure for discriminating between deterministic and stochastic approaches to extreme-value treatment is an important (and challenging) objective.

Returning to the Oil Price data of Fig. 2, we proceed to estimate trends in the data. The ℓ_1 HP filter implicitly defined via (15) depends upon a choice of the smoothing parameter λ, and final results are heavily contingent upon this, much like in the case of tuning parameter selection for the tail index estimation of OctExt data. We compute the resulting trend with the somewhat arbitrary choice of $\lambda = 60$, and repeat the exercise with the heavy-tailed criterion (16) based upon a student t with one degree of freedom (and again with $\lambda = 60$). The resulting trend estimates are displayed in the left and right panels, respectively, of Fig. 5.

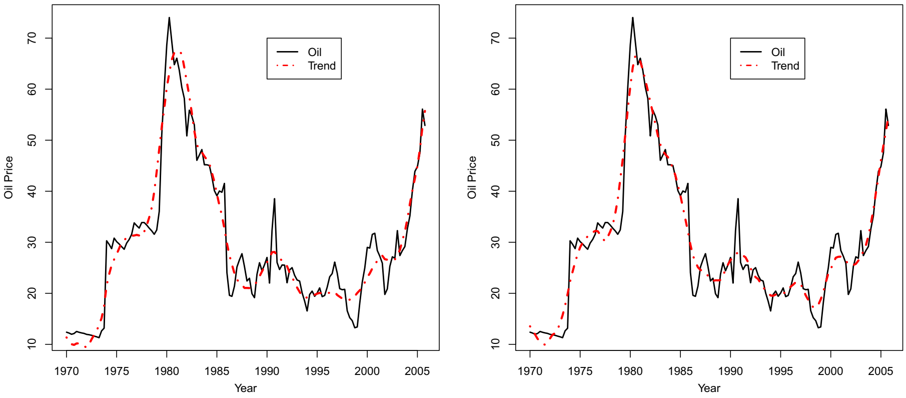

Fig. 5 Oil price data (*black solid*) with estimated trends (*red dot dashed*), from the ℓ_1 HP (*left panel*) and the heavy-tailed HP (*right panel*) based on $\lambda = 60$

Whereas the overall degree of smoothing (i.e., to what extent peaks are retained) is similar for both methods, there are some important differences in the results. The heavy-tailed filter does a superior job of tracking the long-term movements in the series. In contrast, the ℓ_1 HP does more smoothing of kinks, and is itself less smooth than the conventional HP filter. Depending on the goals of analysis, the preservation of kinks and sudden level shifts in the trend could be a desirable feature.

Altering the noise-to-signal ratio λ up to 125 induces additional smoothing, which is displayed in Fig. 6 for both methods. The ℓ_1 HP still has kinks in odd locations, whereas the heavy-tailed HP has a single peak in the early 1980s.

5 Discussion

Stochastic anomalies, or extremes, are prevalent in time series data encountered throughout the social and natural sciences. Some of the chief measures of these

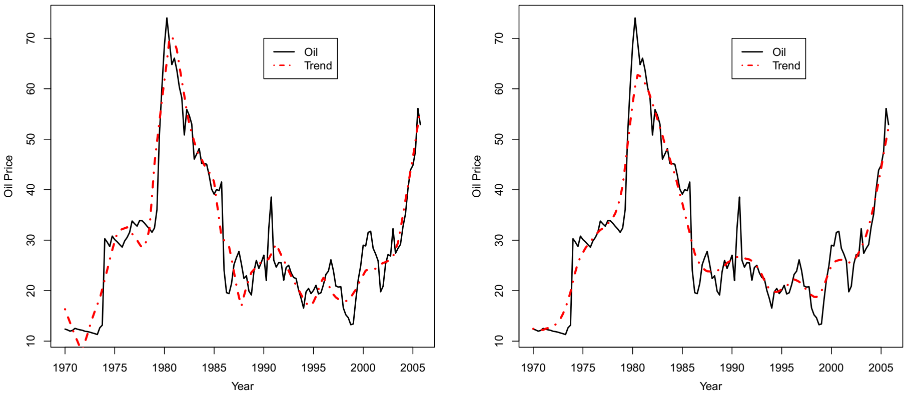

Fig. 6 Oil price data (*black solid*) with estimated trends (*red dot dashed*), from the ℓ_1 HP (*left panel*) and the heavy-tailed HP (*right panel*) based on $\lambda = 125$

anomalies include the tail index, the extremal index, and the extromogram. Some applications require an explicit measurement of these quantities, whereas in other cases anomalies merely reflect a nuisance to be corrected. While many techniques are available – giving the practitioner a breadth of possible tools – there seems to be little consensus in the literature as to which methods are best, or even how to decide between competitors. For example, in tail index estimation alone there are copious estimators, each of which has a carefully developed estimation methodology – but which is best, or most appropriate for a given series? In classic time series model building, there are many tools available for evaluating postulated models, and rigorously comparing their performance on a given series; but the literature on extremal time series more resembles that of nonparametric spectral density estimation, where a variety of tapers and bandwidths are available to the practitioner, each of which has their own partisans.

There lacks an agreed upon set of criteria for determining which tools are appropriate for an analysis, or even how to select the tuning parameters of the method. Why should one favor the Hill estimator over a moment-based estimator, or the extremogram over the extreme dependence measure? (Assuming one is ignorant, as in practice, about the true data process). It seems the field has not yet coalesced about some canonical choice of methods. This situation might be compared to the literature on nonparametric spectral density estimation, which proceeded through a competition of various tapers – with various schema for bandwidth choice – until a recent surge of work finally provided tools for assessing the impact of taper and bandwidth on the finite-sample distribution. An analagous breakthrough in this topic of extremal time series has not yet occurred. As a consequence, it is hard to imagine making a case for any of these techniques being used in official statistics at a national statistical agency or central bank, due to the marked subjectivity of putative methods, and the lack of objective defensible criteria for selecting between contenders.

So what are the key questions that need to be addressed? Given that nonparametric and semi-parametric methods are favored (which seem appropriate for longer series), how can we evaluate the fitness of a particular estimator for a series, versus competitors? This question seems akin to a model selection problem, or to the assessment of the impact of taper choice upon spectral density estimation. Such an evaluation should take into account the tuning of the estimator, where the tuning parameter may have been selected according to some optimality criterion. The tuning parameter (which is featured in tail index estimators, the extremal index estimators, and also the extremogram) is akin to the bandwith parameter in spectral density estimation; can tuning parameters be estimated according to particular user-criteria, analogous to how time series model parameters can be fitted to optimize one-step ahead or multi-step ahead forecasting loss functions, according to the desired application?

For example, suppose that we are interested in trend estimation for the Oil price data, and the trend is generated from Eq. 16 via a choice of α; if we use a Hill estimator, the tuning parameter k could be selected such that the loss function (16) is minimized. Or as another example, in a VaR application an empirical version of the expected shortfall could be devised to obtain a loss function for the tuning parameter of α. Cross-validation studies are in the same vein, but require partitions of the data; it is more appealing to devise a loss function directly tied to the goals of analysis, and select between methods and parameters accordingly. Expanding the extremal time series literature in these directions (e.g., forecasting, signal extraction, VaR minimization) should be, in our opinion, the focus of substantial future research efforts.

Acknowledgments The author thanks the editor for the invitation to write the article, and constructive criticism from two helpful referees.

Disclaimer This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the author and not necessarily those of the U.S. Census Bureau.

References

Aguilar, M., Hill, J.B.: Robust score and portmanteau tests of volatility spillover. J. Econ. **184**, 37–61 (2015)

Anderson, P.L., Meerschaert, M.M.: Periodic moving averages of random variables with regularly varying tails. Ann. Stat. **25**, 771–785 (1997)

Anderson, P.L., Meerschaert, M.M.: Modeling river flows with heavy tails. Water Resour. Res. **34**, 2271–2280 (1998)

Basrak, B., Davis, R.A., Mikosch, T.: Regular variation of GARCH processes. Stoch. Proc. Appl. **99**, 95–115 (2002)

Basrak, B., Krizmanić, D., Segers, J.: A functional limit theorem for dependent sequences with infinite variance stable limits. Ann. Probab. **40**, 2008–2033 (2012)

Beirlant, J., Goegebeur, Y., Teugels, J., Segers, J., de Wall, D., Ferro, C.: Statistics of Extremes: Theory and Applications. Wiley, New York (2004)

Beran, J., Sherman, R., Taqqu, M., Willinger, W.: Long-range dependence in Variable-bit rate video traffic. IEEE Trans. Comm. **43**, 1566–1579 (1995)

Berger, J.M., Mandelbrot, B.B.: A new model for error clustering in telephone circuits. IBM J. Res. Develop. **7**, 224–236 (1963)

Berkes, I., Horváth, L., Kokoszka, P.: Estimation of the maximal moment exponent of a GARCH(1,1) sequence. Econ. Theory **19**, 565–586 (2003)

Bollerslev, T., Todorov, V., Li, S.Z.: Jump tails, extreme dependencies, and the distribution of stock returns. J. Econ. **172**, 307–324 (2013)

Brito, M., Freitas, A.C.M.: Consistent estimation of the tail index for dependent data. Stat. Probab. Lett. **80**, 1835–1843 (2010)

Cappe, O., Moulines, E., Pesquet, J., Petropulu, A., Xueshi, Y.: Long-range dependence and heavy-tail modeling for teletraffic data. Signal Process. Mag. IEEE **19**, 14–27 (2002)

Castillo, E.: Extreme Value Theory in Engineering. Academic Press, San Diego (1988)

Chernozhukov, V., Fernández-Val, I.: Inference for extremal conditional quantile models, with an application to market and birthweight risks. Rev. Econ. Stud. **78**, 559–589 (2011)

Clark, P.K.: A subordinate stochastic process model with finite variance for speculative prices. Econometrica **41**, 135–155 (1973)

Coles, S.: An Introducton to Statistical Modeling of Extreme Values. Springer, London (2001)

Crovella, M., Taqqu, M., Bestavros, A.: Heavy-tailed probability distributions in the World Wide Web. In: Adler, R.J., Feldman, R.E., Taqqu, M.S. (eds.) A Practical Guide to Heavy Tails. Birkhäuser, Boston (1998)

Csörgö, M., Yu, H.: Weak approximations for quantile processes of stationary sequences. Canad. J. Stat. **24**, 403–430 (1996)

Danielsson, J., de Haan, l., Peng, L., de Vries, C.G.: Using a bootstrap method to choose the sample fraction in tail index estimation. J. Multivar. Anal. **76**, 226–248 (2001)

D'Auria, B., Resnick, S.I.: Data network models of burstiness. Adv. Appl. Probab. **38**, 373–404 (2006)

D'Auria, B., Resnick, S.I.: The influence of dependence on data network models. Adv. Appl. Probab. **40**, 60–94 (2008)
Davis, R., Klüppelberg, C., Steinkohl, C.: Statistical inference for max-stable processes in space and time. J. R. Stat. Soc. Series B **75**, 791–819 (2013a)
Davis, R., Klüppelberg, C., Steinkohl, C.: Max-stable processes for modeling extremes observed in space and time. J. Kor. Stat. Soc. **42**, 399–414 (2013b)
Davis, R., Mikosch, T.: Extreme value theory for GARCH processes. In: Mikosch, T., Kreiss, J.-P., Davis, R., Andersen, T.G. (eds.) Handbook of Financial Time Series. Springer-Verlag, Berlin (2009a)
Davis, R., Mikosch, T.: The extremogram: A correlogram for extreme events. Bernoulli **15**, 977–1009 (2009b)
Davis, R., Mikosch, T., Cribben, I.: Towards estimating extremal serial dependence via the bootstrapped extremogram. J. Econ. **170**, 142–152 (2012)
Davis, R., Mikosch, T., Zhao, Y.: Measures of serial extremal dependence and their estimation. Stoch. Process. Appl. **123**, 2575–2602 (2013)
Davis, R.A., Resnick, S.I.: Limit theory for moving averages of random variables with regularly varying tail probabilities. Ann. Probab. **13**, 179–195 (1985)
Davis, R.A., Resnick, S.I.: Limit theory for the sample covariance and correlation functions of moving averages. Ann. Stat. **14**, 533–558 (1986)
Davison, A.C., Huser, R., Thibaud, E.: Geostatistics of dependent and asymptotically independent extremes. Math. Geosci. **45**, 511–529 (2013)
Davision, A.C., Smith, R.L.: Models for exceedances over high thresholds. J. R. Stat. Soc. Ser. B **52**, 393–442 (1990)
de Haan, L.: A spectral representation for max-stable processes. Ann. Probab. **12**, 1194–1204 (1984)
de Haan, L., Pereira, T.T.: Spatial extremes: models for the stationary case. Ann. Stat. **34**, 146–168 (2006)
Doukhan, P., Prohl, S., Robert, C.: Subsampling weakly dependent time series and applications to extremes. Test **20**, 447–479 (2011)
Drees, H.: Weighted approximations of tail processes for β-mixing random variables. Ann. Probab. **10**(4), 1274–1301 (2000)
Drees, H.: Tail empirical processes under mixing conditions. In: Empirical Process Techniques for Dependent Data, pp. 325–342. Birkhäuser, Boston (2002)
Drees, H.: Extreme quantile estimation for dependent data, with applications to finance. Bernoulli **9**(4), 617–657 (2003)
Drees, H.: Some aspects of extreme value statistics under serial dependence. Extremes **11**, 35–53 (2008)
Drees, H., Kaufmann, E.: Selecting the optimal sample fraction in univariate extreme value estimation. Stoch. Process. Appl. **75**, 149–172 (1998)
Drees, H., Rootzén, H.: Limit theorems for empirical processes of cluster functionals. Ann. Stat. **38**, 2145–2186 (2010)
Duffy, D., McIntosh, A., Rosenstein, M., Willinger, W.: Analyzing telecommunications traffic data from working common channel signaling subnetworks. In: Proceedings of the 25th Interface, San Diego. Interface Foundation of North America (1993)
Duffy, D., McIntosh, A., Rosenstein, M., Willinger, W.: Statistical analysis of CCSN/SS7 traffic data from working CCS subnetworks. IEEE J. Select. Areas Commun. **12**, 544–551 (1994)
Einmahl, J.H., Kiriliouk, A., Krajina, A., Segers, J.: An M-estimator of spatial tail dependence. J. R. Stat. Soc. Ser. B. published online (2015)
Embrechts, Mikosh, Kluppelberg: Modeling Extrema Events for Insurance and Finance. Springer-Verlag, Berlin (1997)
Epps, T.W., Epps, M.L.: The stochastic dependence of security price changes and transaction volumes. Econometrica **44**, 305–321 (1976)
Fama, E.F.: The behavior of stock-market prices. J. Business **38**, 34–105 (1965)
Farmer, J.D., Gillemot, L., Lillo, F., Mike, S., Sen, A.: What really causes large price changes? Quant. Finan. **4**, 383–397 (2004)
Fasen, V., Klüppelberg, C., Schlather, M.: High-level dependence in time series models. Extremes **13**, 1–33 (2010)
Fasen, V., Klüppelberg, C., Menzel, A.: Quantifying extreme risks. In: Klüppelberg, C., Straub, D., Welpe, I.M. (eds.) Risk – A Multidisciplinary Introduction, pp. 151–181. Springer (2014)
Ferro, C.A., Segers, J.: Inference for clusters of extreme values. J. R. Stat. Soc. Ser. B **65**, 545–556 (2003)
Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C., Chen, B.: New capabilities and methods for the X-12-ARIMA seasonal adjustment program. J. Bus. Econ. Stat. **16**, 127–152 (1998)

Gencay, R., Selcuk, F.: Extreme value theory and value-at-risk: Relative performance in emerging markets. Int. J. Forecast. **20**, 287–303 (2004)

Gilli, M., Këllezi, E.: An application of extreme value theory for measuring financial risk. Comput. Econ. **27**, 207–228 (2006)

Gomes, M.I., Figueiredo, F., Neves, M.M.: Adaptive estimation of heavy right tails: Resampling-based methods in action. Extremes **15**, 463–489 (2012)

Gomes, M.I.., Guillou, A.: Extreme value theory and statistics of univariate extremes: A review. Int. Stat. Rev. **83**, 263–292 (2015)

Gomes, M.I., Hall, A., Miranda, C.: Subsampling techniques and the Jackknife methodology in the estimation of the extremal index. Comput. Stat. Data Anal. **52**(4), 2022–2041 (2008)

Grigoriu, M.: Applied Non-Gaussian Processes. Prentice Hall, Englewood Cliffs (1995)

Hamidieh, K., Stoev, S., Michailidis, G.: On the estimation of the extremal index based on scaling and resampling. J. Comput. Graph. Stat. **18**, 731–755 (2009)

Heath, D., Resnick, S., Samorodnitsky, G.: Heavy tails and long range dependence in on/off processes and associated fluid models. http://www.orie.cornell.edu/trlist/ (1997)

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

Hill, J.B.: On functional central limit theorems for dependent, heterogeneous arrays with applications to tail index and tail dependence estimation. J. Stat. Plan. Infer. **139**, 2091–2110 (2009)

Hill, J.B.: On tail index estimation for dependent, heterogeneous data. Econometric Theory **26**, 1398–1436 (2010)

Hill, J.B.: Tail and nontail memory with applications to extreme value and robust statistics. Econ. Theory **27**, 844–884 (2011a)

Hill, J.B.: Extremal memory of stochastic volatility with an application to tail shape inference. J. Stat. Plan. Infer. **141**, 663–676 (2011b)

Hill, J.B.: Least tail-trimmed squares for infinite variance autoregressions. J. Time Ser. Anal. **34**, 168–186 (2013a)

Hill, J.B.: Expected shortfall estimation and Gaussian inference for infinite variance time series. J. Finan. Econ. Published online (2013b)

Hill, J.B., Aguilar, M.: Moment condition tests for heavy tail time series. J. Econ. **172**, 255–274 (2013)

Hill, J.B.: Robust estimation and inference for heavy tailed GARCH. Bernoulli **21**, 1629–1669 (2015)

Hill, J.B., Shneyerov, A.: Are there common values in first-price auctions? A tail-index nonparametric test. J. Econ. **174**, 144–164 (2013)

Holan, S.H., McElroy, T.S.: Tail exponent estimation via broadband log density-quantile regression. J. Stat. Plan. Infer. **140**, 3693–3708 (2010)

Horváth, L., Kokoszka, P.S.: Sample autocovariances of long-memory time series. Bernoulli **14**, 405–418 (2008)

Hsing, T.: Estimating the parameters of rare events. Stoch. Proc. Appl. **37**, 117–139 (1991a)

Hsing, T.: On tail index estimation using dependent data. Ann. Stat. **19**, 1547–1569 (1991b)

Hsing, T.: Extremal index estimation for a weakly dependent stationary sequence. Ann. Stat. **21**, 2043–2071 (1993)

Hsing, T., Hüsler, J., Leadbetter, M.R.: On the exceedance point process for a stationary process. Probab. Th. Rel. Fields **78**, 97–112 (1988)

Huser, R., Davison, A.C.: Space-time modelling of extreme events. J. R. Stat. Soc. Ser. B **76**, 439–461 (2014)

Iglesias, E.M.: An analysis of extreme movements of exchange rates of the main currencies traded in the Foreign Exchange market. Appl. Econ. **44**, 4631–4637 (2012)

Jach, A., McElroy, T.S., Politis, D.N.: Subsampling inference for the mean of heavy-tailed long memory time series. J. Time Ser. Anal. **33**, 96–111 (2012)

Kabluchko, Z.: Extremes of space-time Gaussian processes. Stoch. Process. Appl. **119**, 3962–3980 (2009)

Kabluchko, Z., Schlather, M., de Haan, L.: Stationary max-stable fields associated to negative definite functions. Ann. Probab. **37**(5), 2042–2065 (2009)

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

Kim, S.J., Koh, K., Boyd, S., Gorinevsky, D.: ℓ_1 Trend filtering. SIAM Rev. **51**, 339–360 (2009)

Koedijk, K., Schafgans, M., De Vries, C.: The tail index of exchange rate returns. J. Int. Econ. **29**, 93–108 (1990)

Kokoszka, P.S., Taqqu, M.S.: Parameter estimation for infinite variance fractional ARIMA. Ann. Stat. **24**, 1880–1913 (1996)

Kokoszka, P.S., Taqqu, M.S.: Discrete time parametric models with long memory and infinite variance. Math. Comput. Model. **293**, 203–215 (1999)

Kulik, R., Soulier, P.: The tail empirical process of some long memory sequences. Stoch. Process. Appl. **121**, 109–134 (2011)

Kulik, R., Soulier, P.: Limit theorems for long memory stochastic volatility models with infinite variance: Partial sums and sample covariances. Adv. Appl. Probab. **44**(4), 1113–1141 (2012)

Kulik, R., Soulier, P.: Estimation of limiting conditional distibutions for the heavy tailed long memory stochastic volatility process. Extremes **16**, 203–239 (2013)

Kulik, R., Soulier, P.: Heavy tailed time series with extremal independence. Extremes **18**, 273–299 (2015)

Kyselý, J.: A cautionary note on the use of nonparametric bootstrap for estimating uncertainties in extreme-value models. Amer. Metereol. Soc. **47**, 3236–3251 (2008)

Lahiri, S.: Resampling Methods for Dependent Data. Springer, New York (2003)

Larsson, M., Resnick, S.I.: Extremal dependence measure and extremogram: The regularly varying case. Extremes **15**, 231–256 (2012)

Leadbetter, M.R.: On extreme values in stationary sequences. Prob. Th. Rel. Fields **28**, 289–303 (1974)

Leadbetter, M.R.: Extremes and local dependence in stationary sequences. Probab. Theory Related Fields **65**, 291–306 (1983)

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

Leadbetter, M.R., Nandagopalan, S.: On exceedance point processes for stationary sequences under mild oscillation restrictions. In: Hüsler, J., Reiss, R.-D. (eds.) Extreme Value Theory, pp. 69–80. Springer-Verlag, New York (1989)

Leadbetter, M.R., Rootzén, H.: Extremal theory for stochastic processes. Ann. Probab. **16**, 431–478 (1988)

Ledford, A., Tawn, J.: Diagnostics for dependence within time series extremes. J. R. Stat. Soc. Ser. B **65**, 521–543 (2003)

Li, Y., Simmonds, D., Reeve, D.: Quantifying uncertainty in extreme values of design parameters with resampling techniques. Ocean Eng. **35**, 1029–1038 (2008)

Linton, O., Xiao, Z.: Estimation of and inference about the expected shortfall for time series with infinite variance. Econ. Theory **29**, 771–807 (2013)

Loynes, R.M.: Extreme values in uniformly mixing stationary stochastic processes. Ann. Math. Stat. **36**, 993–999 (1965)

Mandelbrot, B.B.: The variation of certain speculative prices. J. Bus. **35**, 394–419 (1963)

Mandelbrot, B.B.: Self-similar error clusters in communications systems and the concept of conditional systems and the concept of conditional stationarity. IEEE Trans. Commun. Technol. **COM-13**, 71–90 (1965)

Mandelbrot, B.B.: Long-run linearity, locally Gaussian processes, H-spectra and infinite variances. Int. Econ. Rev. **10**, 82–113 (1969)

Mandelbrot, B.B.: Intermittent turbulence in self similar cascades; divergence of high moments and dimension of the carrier. J. Fluid Mech. **62**, 331–358 (1974)

Mandelbrot, B.B.: The Fractal Geometry of Nature. W.H.Freeman and Company (1983)

Mandelbrot, B.B., Wallis, J.R.: Noah, Joseph, and operational hydrology. Water Resour. Res. **4**, 909–917 (1968)

Mandelbrot, B.B., Wallis, J.R.: Some long-run properties of geophysical records. Water Resour. Res. **5**, 321–340 (1969)

Markovich, N.M.: Modeling clusters of extreme values. Extremes **17**, 97–125 (2014)

Martin, D.R., Yohai, V.J.: Influence functionals for time series. Ann. Stat. **14**, 781–855 (1986)

McElroy, T.S.: Exact formulas for the Hodrick-Prescott filter. Econ. J. **11**, 209–217 (2008)

McElroy, T.S., Jach, A.: Tail index estimation in the presence of long memory dynamics. Comput. Stat. Data Anal. **56**, 266–282 (2012a)

McElroy, T.S., Jach, A.: Subsampling inference for the autocovariances of heavy-tailed long-memory time series. J. Time Ser. Anal. **33**, 935–953 (2012b)

McElroy, T.S., Nagaraja, C.H.: Tail index estimation with a fixed tuning parameter fraction. J. Stat. Plan. Infer. **170**, 27–45 (2016)
McElroy, T.S., Politis, D.N.: Robust inference for the mean in the presence of serial correlation and heavy-tailed distributions. Econ. Theory **5**, 1019–1039 (2002)
McElroy, T.S., Politis, D.N.: Stable marked point processes. Ann. Stat. **35**, 393–419 (2007a)
McElroy, T.S., Politis, D.N.: Computer-intensive rate estimation, diverging statistics, and scanning. Ann. Stat. **35**, 1827–1848 (2007b)
McElroy, T.S., Politis, D.N.: Moment-based tail index estimation. J. Stat. Plan. Infer. **137**, 1389–1406 (2007c)
McElroy, T.S., Politis, D.N.: Self-normalization for heavy-tailed time series with long memory. Statistica Sinica **17**, 199–220 (2007d)
McNeil, A.J., Frey, R.: Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. J. Empir. Finan. **7**, 271–300 (2000)
Meerschaert, M.M., Scheffler, H.-P.: A simple robust estimator for the thickness of heavy tails. J. Stat. Plan. Infer. **71**, 19–34 (1998)
Meerschaert, M.M., Scheffler, H.-P.: Sample cross-correlations for moving averages with regularly varying tails. J. Time Ser. Anal. **22**, 481–492 (2001)
Meier-Hellstern, K., Wirth, P., Yan, Y., Hoeflin, D.: Traffic models for ISDN data users: Office automation application. In: Jensen, A., Iversen, V.B. (eds.) Teletraffic and Datatraffic in a Period of Change. Proceedings of the 13th ITC, pp. 167–192. North Holland, Amsterdam (1991)
Mikosch, T.: Modeling dependence and tails of financial time series. In: Finkenstädt, B., Rootzén, H. (eds.) Extreme Values in Finance, Telecommunications, and the Environment. Chapman and Hall, Boca Raton (2003)
Mikosch, T., Stărică, C.: Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. Ann. Stat. **28**, 1427–1451 (2000)
Mikosch, T., Zhao, Y.: A Fourier analysis of extreme events. Bernoulli **20**, 803–845 (2014)
Mikosch, T., Zhao, Y.: The integrated periodogram of a dependent extremal event sequence. Stoch. Process. Appl. **125**, 3126–3169 (2015)
Nandagopalan, S.: Multivariate extremes and estimation of the extremal index. Ph.D, Thesis, Univ. North Carolina at Chapel Hill (1990)
Newcomb, S.: A generalized theory of the combination of observations so as to obtain the best result. Amer. J. Math. **8**, 343–366 (1886)
Pandey, M.D., Van Gelder, P.H.A.J.M., Vrijling, J.K.: Bootstrap simulations for evaluating the uncertainty associated with peaks-over-trheshold estimates of extreme wind velocity. Environmetrics **14**, 27–43 (2003)
Phillips, P.C.B.: Time series regression with a unit root and infinite variance errors. Econ. Theory **6**, 44–62 (1990)
Politis, D.N.: A new approach on estimation of the tail index. C.R. Acad. Sci. Paris, Ser. I **335**, 279–282 (2002)
Politis, D.N., Romano, J.P., Wolf, M.: Subsampling. Springer, New York (1999)
Reiss, R.-D., Thomas, M.: Statistical Analysis of Extreme Values. Basel, Birkhäuser (1997)
Resnick, S.: Special invited paper: Heavy tail modeling and teletraffic data. Ann. Stat. **25**(5), 1805–1849 (1997)
Resnick, S.: The extremal dependence measure and asymptotic independence. Stoch. Models **20**, 205–227 (2004)
Resnick, S.: Heavy-tail Phenomena: Probabilisitc and Statistical Modeling. Springer, New York (2007)
Resnick, S., Stărică, C.: Consistency of Hill's estimator for dependent data. J. Appl. Probab. **32**, 139–167 (1995)
Resnick, S., Stărică, C.: Tail index estimation for dependent data. Ann. Appl. Probab. **8**(4), 1156–1183 (1998)
Robert, C.Y., Segers, J., Ferro, C.A.T.: A sliding blocks estimator for the extremal index. Electron. J. Stat. **3**, 993–1020 (2009)
Robert, C.Y.: Asymptotic distributions for the intervals estimators of the extremal index and the cluster-size probabilities. J. Stat. Plann. Infer. **139**, 3288–3309 (2009a)
Robert, C.Y.: Inference for the limiting cluster size distribution of extreme values. Ann. Stat. **37**, 271–310 (2009b)

Rootzén, H.: Extremes of moving averages of stable processes. Ann. Probab. **6**, 847–869 (1978)
Rootzén, H.: Weak convergence of the tail empirical process for dependent sequences. Stoch. Proces. Appl. **119**(2), 468–490 (2009)
Rossi, E., Santucci de Magistris, P.: Long memory and tail dependence in trading volume and volatility. J. Empir. Finan. **22**, 94–112 (2013)
Schlather, M.: Models for stationary max-stable random fields. Extremes **5**(1), 33–44 (2002)
Sibuya, M.: Bivariate extreme statistics, I. Ann. Inst. Stat. Math. **11**, 195–210 (1959)
Smith, R.L.: Extreme value analysis of environmental time series: An example based on ozone data. Stat. Sci. **4**, 367–393 (1989)
Smith, R.L., Tawn, J.A., Coles, S.G.: Markov chain models for threshold exceedances. Biometrika **84**, 249–268 (1997)
Smith, R.L., Weissman, I.: Estimating the extremal index. J. Roy. Stat. Soc. Ser. B **56**, 515–528 (1994)
Stigler, S.M.: Simon Newcomb, Percy Daniell, and the history of robust estimation 1885–1920. J. Amer. Stat. Assoc. **68**, 872–879 (1973)
Stoev, S., Michailidis, G., Taqqu, M.: Estimating heavy-tail exponents through max self-similarity. IEEE Trans. Inf. Theory **57**, 1615–1635 (2011)
Tesfaye, Y.G., Meerschaert, M.M., Anderson, P.L.: Identification of periodic autoregressive moving average models and their application to the modeling of river flows. Water Resour. Res. **42**, 1–11 (2006)
Thibaud, E., Mutzner, R., Davison, A.C.: Threshold modeling of extreme spatial rainfall. Water Resour. Res. **49**, 4633–4644 (2013)
Trimbur, T.M.: Stochastic level shifts and outliers and the dynamics of oil price movements. Int. J. Forecast. **26**, 162–179 (2010)
Weller, G.B., Cooley, D.S., Sain, S.R.: An investigation of the pineapple express phenomenon via bivariate extreme value theory. Environmetrics **23**, 420–439 (2012)
Willinger, W., Taqqu, M., Leland, W., Wilson, D.: Self-similarity in high-speed packet traffic: Analysis and modeling of ethernet traffic measurements. Stat. Sci. **10**, 67–85 (1995)
Willinger, W., Taqqu, M., Sherman, R., Wilson, D.: Self-similarity through high-variability: Statistical analysis of Ethernet LAN traffic at the source level. IEEE/ACM Trans. Network. **5**, 71–96 (1997)
Xiao, Z., Koenker, R.: Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models. J. Amer. Stat. Assoc. **104**, 1696–1712 (2009)
Yamada, H., Jin, L.: Japans output gap estimation and ℓ_1 trend filtering. Empir. Econ. **45**, 81–88 (2013)