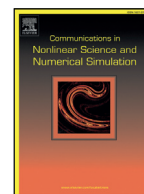




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Invited Article

Invited article by M. Gidea Extreme events and emergency scales[☆]

Veniamin Smirnov, Zhuanzhuan Ma, Dimitri Volchenkov*

Department of Mathematics and Statistics, Texas Tech University, Lubbock, TX 79409, United States

ARTICLE INFO

Article history:

Received 27 November 2019

Revised 9 April 2020

Accepted 19 May 2020

Available online 22 May 2020

Keywords:

Extreme events

Emergency scales

Uncertainty of threshold

ABSTRACT

An event is extreme if its magnitude exceeds the threshold. A choice of a threshold is subject to uncertainty caused by a method, the size of available data, a hypothesis on statistics, etc. We assess the degree of uncertainty by the Shannon's entropy calculated on the probability that the threshold changes at any given time. If the amount of data is not sufficient, an observer is in the state of Lewis Carroll's *Red Queen* who said "When you say hill, I could show you hills, in comparison with which you'd call that a valley". If we have enough data, the uncertainty curve peaks at two values clearly separating the magnitudes of events into three emergency scales: *subcritical*, *critical*, and *extreme*. Our approach to defining the emergency scale is validated by 39 years of Standard and Poor's 500 (S&P500) historical data.

Published by Elsevier B.V.

1. Introduction

Not a single day passes by without hearing about extreme events which surround us almost everywhere. On one hand, climate change results in droughts, heat waves, tornadoes, storms etc; movement of the tectonic plates is responsible for earthquakes and volcano eruptions. On the other hand, certain aspects of human activities may crash stock markets, influence tensions not only among groups of people in a country, but also between countries sometimes leading to military confrontations and mass migrations, etc. While there is no single definition of extreme events [1], they are considered events that cause infrastructure failures, economic and property losses, risk to health and life.

In order to quantify extreme events, practitioners developed several scales. For instance, Modified Mercalli Intensity scale [2] (describes the intensity of visible damage of earthquakes), Beaufort Wind scale [3] (measures speed and observed effects the wind has on the sea and land), Saffir-Simpsons Hurricane scale [4] (measures wind speed), Fujita scale [5] (rates the intensity of a tornado after it has passed), US Homeland Security Terror Alert scale [6] (measures five color-coded terror alert levels), U.S. Climate Extremes Index [7], etc. Rohn Emergency Scale [8] unites emergency scales using three independent dimensions: (i) scope; (ii) topographical change (or lack thereof); and (iii) speed of change. The intersection of the three dimensions provides a detailed scale for defining any emergency [8]. In some papers, the threshold for an extreme event is related to the number of standard deviations from the average amplitude [9,10]. However, existing empirical scales tend to describe the characteristics of the event itself rather than the consequences; such scales are ill-suited to describe emergencies in a way that is meaningful for response [11].

[☆] Work done under contract with the AVX Aircraft Company Contract #W911W6-13-2-0004.

* Corresponding author.

E-mail addresses: veniamin.smirnov@ttu.edu (V. Smirnov), zhuanzhuan.ma@ttu.edu (Z. Ma), dimitri.volchenkov@ttu.edu (D. Volchenkov).

For instance, extreme events of different magnitudes in financial markets range from global recessions (defined by a global annual GDP growth rate of 3.0 percent or less) that happened in 1975, 1982, 1991, 2009 to “flash crashes” (for instance, on May 6, 2010, the S&P500 declined 7% in less than 15 min, and then quickly rebounded). Unlike the Richter magnitude scale, the severity of flash crashes in financial markets is defined by the measures that need to be taken to ease panic such as halting trading. Under 2012 rules, market-wide circuit breakers (or ‘curbs’) kick in when the S&P 500 index drops 7% for Level 1; 13% for Level 2; and 20% for Level 3 from the prior days close. A market decline that triggers a Level 1 or 2 circuit breaker before 3:25 p.m. Eastern Time will halt trading for 15 min, but will not halt trading at or after 3:25 p.m. Circuit breakers can also be imposed on single stocks as opposed to the whole market. Under current rules, a trading halt on an individual security is placed into effect if there is a 10% change in value of a security that is a member of the S&P 500 Index within a 5-minute time frame, 30% change in value of a security whose price is equal or greater than \$1 per share, and 50% change in value of a security whose price is less than \$1 per share [14]. Ironically, in August 2015, single stock circuit breakers produced unprecedented disruption as 327 exchange-traded funds experienced more than 1000 trading halts during a single day.

For the short-term reactions of stock markets, measured in terms of returns, there exist several approaches to defining their severity. For instance, some authors [12] define extreme events of stocks if prices change greater than 2.5%, for both positive and negative returns. However, others [13] suggest a threshold of 10%.

In general, statistics of extreme events show that the extreme events are found in the tails of probability distributions (i.e. the distributions extremities). Inference over tails is usually performed by fitting an appropriate limiting distribution over observations that exceed a fixed threshold. The choice of the threshold over which to fit a probability distribution is hard and arbitrary. Although tools to guide this choice exist, inference can greatly vary for different thresholds. In addition to that, different distributions may admit asymptotic tails of the same appearance (a power law). In our paper we apply two approaches existing in practical extreme value analysis to study S&P 500 time series in the period from January 2, 1980 till December 31, 2018.

The first one relies on deriving block maxima series as a preliminary step. We fit the actual data to the Generalized Extreme Value Distribution (GEVD). Our results (Section 4.1) show that the distribution of block maxima is a composition of several distributions (Fig. 6b).

The second approach relies on extracting the peak values reached for any period during which values exceed a certain threshold (falls below a certain threshold). We will cover major methods of choosing the threshold in Section 4.2. Usually with the second approach the analysis may involve fitting two distributions: one for the number of events in a time period considered and a second for the size of the exceedances. But in practice a suitable threshold is unknown and must be determined for each analysis [15].

We demonstrate several methods of choosing a threshold. An example of an empirical method, the rule of thumb, is studied in Section 4.2.3. A choice of the threshold can be made using statistical analysis tools, such as graphical approaches (Section 4.2) and automatic methods (Section 4.2.2). Rather than characterize extreme events solely in terms of the distribution of weekly maxima, threshold methods take account of all events beyond a given high threshold (Fig. 1).

Every decision we make or approach we choose is made in the face of uncertainty. For instance, as of March 25, 2020, 414, 179 cases and 18, 440 deaths due to coronavirus disease 2019 (COVID-19), caused by the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), had been reported worldwide [16]. In attempts to slow down spread of the disease, the economies of many counties slammed brakes and financial markets plummeted in the light of uncertainty, but the above does not describe severity of this particular extreme event, but rather, a reaction of the mankind to COVID-19. One natural characteristic of severity of pandemic is death rate, which in case of SARS-CoV-2 ranges from 0.66% [17] to 4% [18]. Mortality rate is higher among people with underlying medical conditions, sometimes, deaths partly occur due to coronavirus. Therefore, some believe that this pandemic is poised to be more dramatic than anything we have seen in our lifetimes [19], but other compare the situation to an elephant being attacked by a house cat. Frustrated and trying to avoid the cat, the elephant accidentally jumps off a cliff and dies [20].

An uncertainty depends on amount of data considered, a method used, etc. Moreover, once new data becomes available, we cannot guarantee sustainability of the previously chosen threshold value. In addition to that, assessment of the severity of the extreme event is also ruled by consequences, which may be revealed *ex post facto*. In Section 4.3 we presented statistics of extreme events under threshold uncertainty. We analyzed the degree of uncertainty of the threshold value (chosen with the rule of thumb) based on the probability of its change at any given day. Fig. 16 demonstrates the degree of uncertainty that depends on both the amount of data available and the values of thresholds. We observed several cases:

- (i) If the amount of data is not sufficient (a solid line corresponding to an 18-day window of observation in Fig. 16), then the uncertainty curve forms a skewed profile attaining a single maximum for some value of the threshold. In this situation, an observer's perception of events reminds *Red Queen* from “Through The Looking-Glass and What Alice Found There” by Lewis Carroll [21] who said “When you say hill, I could show you hills, in comparison with which you'd call that a valley”. Our understanding of the events whether they are extreme or not is very limited and uncertainty is blurry. As events become more severe, our uncertainty that the events are extreme decreases. The observer realizes that the events are extreme, but a precise point at which the events turn to be severe cannot be determined. This case is called the *Red Queen State*.

4 VENIAMIN SMIRNOV, ZHUANZHUAN MA, AND DIMITRI VOLCHENKOV

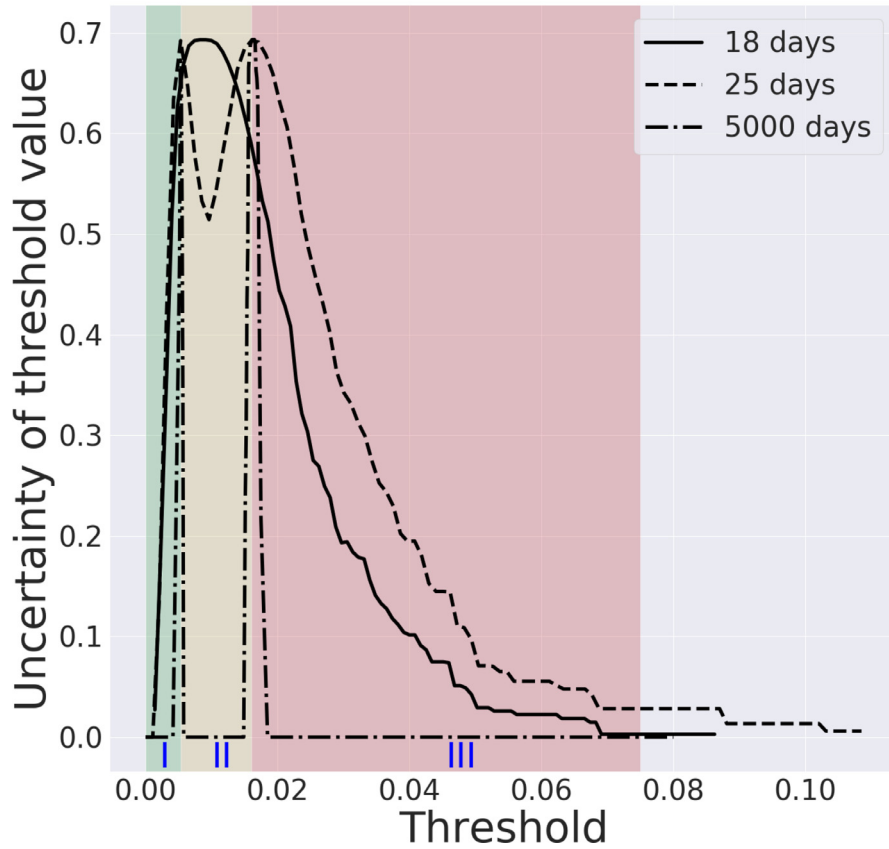


Fig. 1. Degree of uncertainty of different values of the thresholds based on the amount of trading days taken into account. Three shaded regions mark three scales of emergency: I - subcritical, II - critical, III - extreme and the solid curve represents uncertainty of the Red Queen State.

- (ii) As the window of observation becomes larger, the uncertainty curve exhibits two maxima indicating that the amount of data is sufficient. As we further extend the window the curve torrents into sharp peaks (Fig. 16). The latter ones clearly separate the threshold values into three regions: three levels of emergency.
- (I) In the region I (*subcritical*), the threshold values are small to raise concern about extreme events. Then we can observe a spike with the degree of uncertainty attaining its first maximum. This extremum indicates a transition to the next kind of uncertainty.
- (II) In the region II (*critical*), uncertainty is conceptualized with a question whether a magnitude of the event is already critical, extreme or not yet. Further, we see another jump of uncertainty. At this point we certain that events are not regular anymore.
- (III) We consider all events in this region *extreme* with our uncertainty decreasing. If these events are not extreme then what are they?

We consider the Red Queen state and three scales, which are based on the degree of uncertainty, our contribution to the discussion on the extreme events and emergency scales. We conclude in the last Section 5.

2. Data source and description

The analysis of extreme events in the S&P 500 time series of log-return has been performed based on the data collected during 9835 trading days (39 years), in the period from January 2, 1980, till December 31, 2018 (see Fig. 2) acquired using the publicly available source at Yahoo Finance (<https://finance.yahoo.com/quote/%5EGSPC/>).

The data set contains S&P 500 index at market open, highest point reached in the day, lowest point in the day, index of the stock at market close, number of shares traded. We have used index at market close since this information was always present in the data set. Computations were made using Python's numerical libraries, such as NumPy and Pandas, as well as R-language to perform statistical analysis.



Fig. 2. The S&P 500 Index of 500 large-cap U.S. stocks assessing market performance.

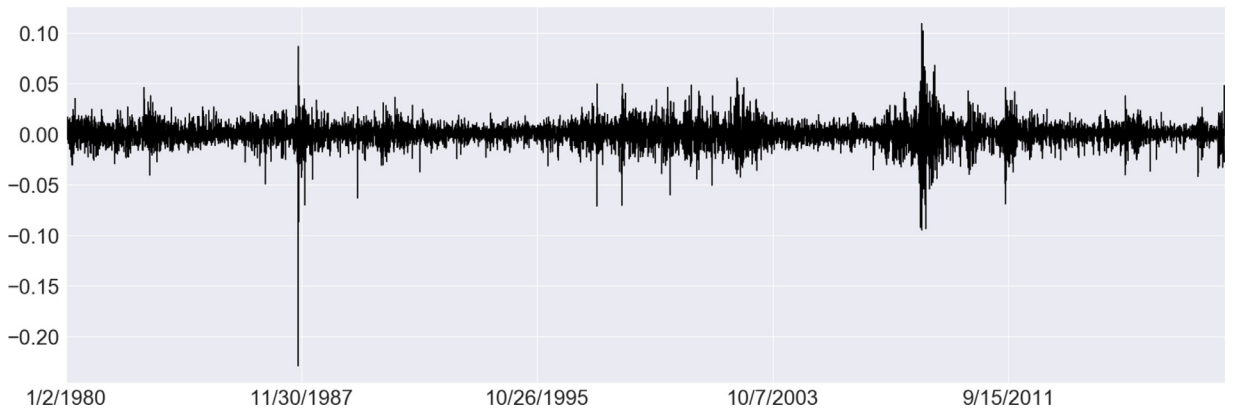


Fig. 3. The S&P 500 index log-return at market close.

3. Log returns: Between noise and random walks

Due to complexity of financial data, in order to analyze dynamics of the given financial time series, we computed *log return*, denoted R_{ln} , according to the following formula [22]:

$$R_{ln}(t) = \ln \left(\frac{S\&PIndex(t)}{S\&PIndex(t-1)} \right), \quad (3.1)$$

where $S\&PIndex(t)$ is the value of S&P 500 at market close (Fig. 3).

The lowest drop of log-return in Fig. 3 happened Monday, October 19, 1987 known as Black Monday when the S&P500 shed value of nearly 23% [23]. Another significant drop occurred in 2008 when S&P 500 fell 38.49%, its worst yearly percentage loss [24]. In September 2008, Lehman Brothers collapsed as the financial crisis spread. However, on Oct 13, 2008: S&P 500 marks its best daily percentage gain, rising 11.58% and registers its largest single-day point increase of 104.13 points [24].

Distribution of R_{ln} values (Fig. 4) is asymmetrically skewed, with fat heterogeneous right and left tails.

The log-return time series has a scale invariant structure when the structure repeats itself on sub-intervals of the signal, $X(ct) = c^H X(t)$, where the Hurst exponent H characterizes the asymptotic behavior of the *auto-correlation function* of the time series [25–27]. The larger Hurst exponent is visually seen as more slow evolving variations (i.e., more persistent structure) of the time series [26,28,29]. Processes with $0 < H < 0.5$ exhibit *antipersistence*, with an increase in the process is likely to be followed by a decrease in the next time interval resulting in sample paths with a very rough structure [26,28,29]. On the contrary, values $0.5 < H < 1$ lead to long-range dependence (“long memory”) in the time series, with more likely the same sign of successive increments (*persistence*) and smoother sample trajectories. Finally, the time series constitutes *random walks* when $H > 1$ that have more apparent slow evolving fluctuations [26,28,29]. The q -order Hurst exponent H_q is only one of several types of scaling exponents used to parameterize the multifractal structure of time series [26,30].

The log-return time series for S&P 500 exhibits local fluctuations with both extreme small and large magnitudes, as well as short- and long-range dependences on different time scales [31,32]; it is not normal distributed and all q -order statistical moments should be considered to describe the spatial and temporal variation that reveals a departure of the log-return

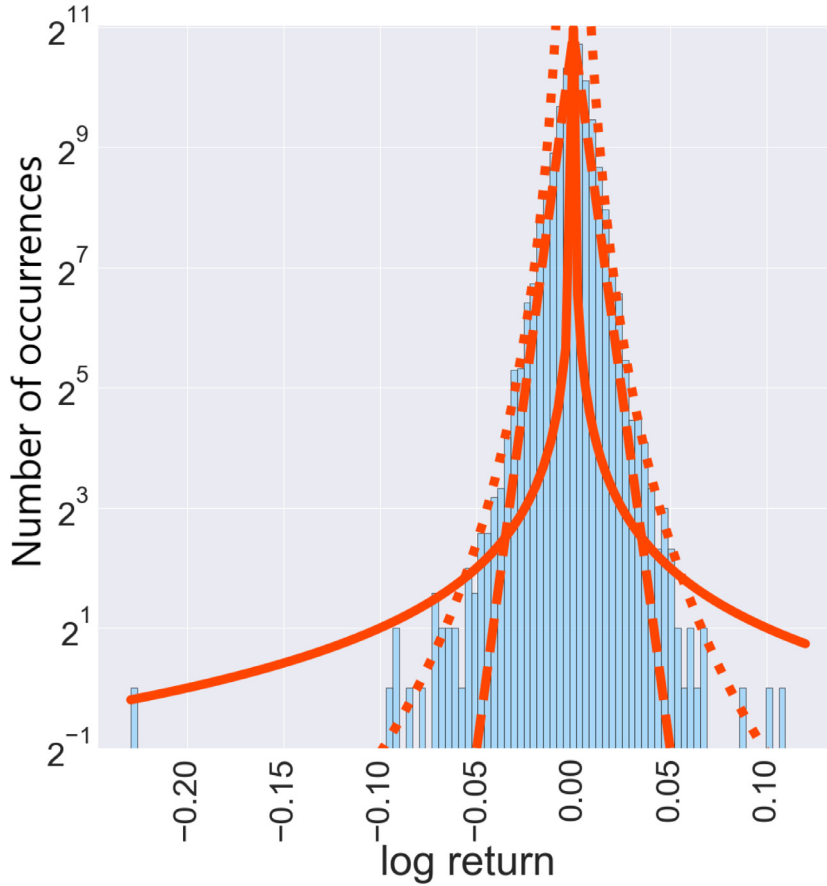


Fig. 4. Distribution of log-return values in the log2-linear scale. Solid lines correspond to a Zipf's Law ($\propto x^{-1}$), dotted lines represent Power Law ($\propto x^{-3.5}$), and dashed lines correlate to Gaussian distribution ($\propto \exp(-x^2)$). Curves are given for reference only.

time series from simple random walk behavior [26,28]. The q -order weights the influence of segments with large and small fluctuations. The negative q 's are influenced by the time series segments with small fluctuations, and large fluctuations influence the time series segments for positive q 's.

In our work, we use the standard *multifractal detrended fluctuation analysis* (MFDFA) algorithm [26,30] for estimating the q -order Hurst exponents and the multifractal spectra directly from the time series:

- (1) The original time series x_k , $k = 1, \dots, N$ is aggregated by computing the cumulative sums $Y(n) = \sum_{k=1}^n (x_k - \langle x \rangle)$, $n = 1, \dots, N$, where $\langle x \rangle$ denotes the sample mean;
- (2) The aggregated data is divided into $\lfloor N/s \rfloor$ non-overlapping segments of length s ;
- (3) The maximum likelihood estimator of the residual variance in segment ν , $\nu = 1, \dots, N_s$,

$$F^2(\nu, s) = \begin{cases} \frac{1}{s} \sum_{i=1}^s \{Y[(\nu-1)s+i] - y_\nu(i)\}^2, & \text{for } \nu = 1, \dots, N_s, \\ \frac{1}{s} \sum_{i=1}^s \{Y[N - (\nu - N_s)s + i] - y_\nu(i)\}^2, & \text{for } \nu = N_s + 1, \dots, 2N_s, \end{cases}$$

where $y_\nu(i)$ is the m -degree polynomial fitted the aggregated observations in the segment;

- (4) For each segment of length s and for each positive or negative values of the moment order q , the q -order fluctuation function, $F_q(s) = \left\{ \frac{1}{N} \sum_{\nu=1}^{N_s} [F^2(\nu, s)]^{q/2} \right\}^{1/q}$, are calculated. The local fluctuations F_q with large and small magnitudes is graded by the magnitude of the negative or positive q -order, respectively;
- (5) A linear regression of $\ln F_q(s)$ on $\ln s$ for all s is performed, and the slope of the linear function $\ln F_q(s) \propto H_q \ln s$ is used as an estimator of the q -order Hurst exponent H_q for each q -order fluctuation function F_q .

The fractal structures of the *positive* and *negative* log-return time series and its deviations within time periods with large and small fluctuations are assessed by the q -order Hurst exponents (see Fig. 5).

The slopes H_q of the regression lines are q -dependent for the multifractal time series of positive (the dashed line) and negative (the bold line) log-returns. (see Fig. 5). Decreasing H_q with the q order indicates that the segments with *small* fluctuations have a random walk like structure whereas segments with *large* fluctuations have a noise like structure.

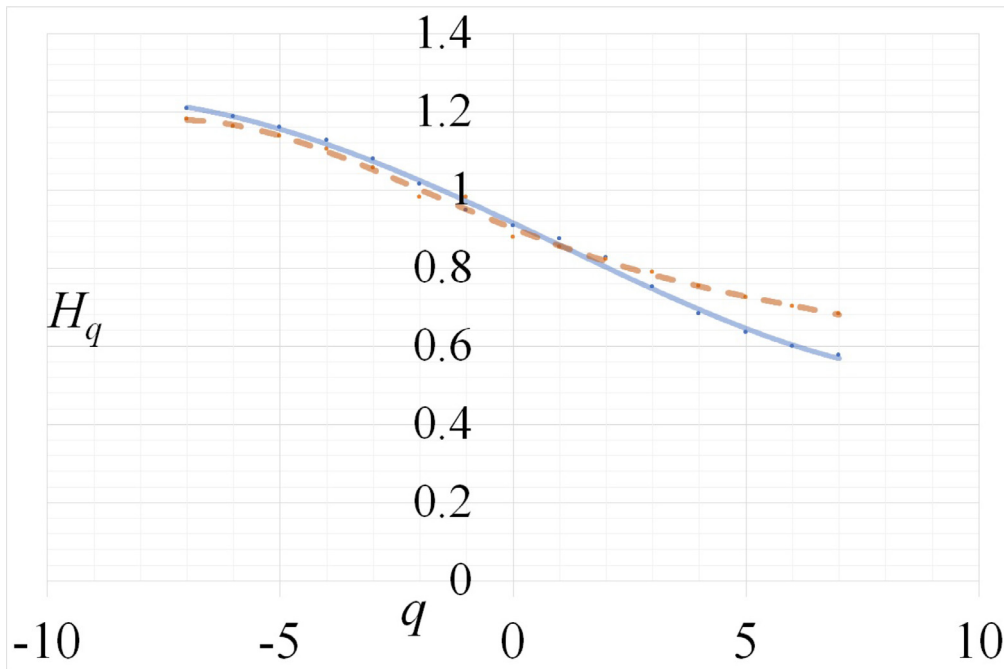


Fig. 5. The q -order Hurst exponents H_q for the time series of positive (the dashed line) and negative (the bold line) log-returns.

4. Tails, thresholds, and extreme events

There are two primary approaches to analyzing extreme values (the extreme deviations from the median of the probability distributions) in data:

- (i) The first and more classical approach reduces the data considerably by taking maxima of long blocks of data, e.g., annual maxima. The GEVD function has theoretical justification for fitting to block maxima of data [33].
- (ii) The second approach is to analyze excesses over a high threshold. For this second approach the generalized Pareto distribution (GPD) function has similar justification for fitting to excesses over a high threshold [33].

4.1. Generalized extreme value distributions

The GEVD is a flexible three-parameter continuous probability distributions that was developed with extreme value theory to combine the Gumbel, Fréchet, and Weibull extreme values distributions into one single distribution [34,35]. The GEV distribution has the following pdf [36]:

$$f(x; \mu, \sigma, \xi) = \frac{1}{\sigma} t(x)^{\xi+1} e^{-t(x)},$$

where

$$t(x) = \begin{cases} (1 + \xi (\frac{x-\mu}{\sigma}))^{-1/\xi} & \xi \neq 0 \\ e^{-(x-\mu)/\sigma} & \xi = 0 \end{cases}$$

and $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter, and $\xi \in \mathbb{R}$ is the shape parameter. When the shape parameter ξ is equal to 0, greater than 0, and lower than 0 [33], the GEV distribution is equivalent to Gumbel [37], Fréchet [38] and “reversed” Weibull distributions [39], respectively.

The Gumbel distribution, also named as the Extreme Value Type I distribution, has the following pdf and cdf:

$$f(x; \mu, \beta) = \frac{1}{\beta} e^{-\left(\frac{x-\mu}{\beta} + e^{-\frac{x-\mu}{\beta}}\right)}, \quad (4.1)$$

$$F(x; \mu, \beta) = e^{-e^{-\frac{x-\mu}{\beta}}} \quad (4.2)$$

where $x \in \mathbb{R}$, μ is the location parameter, $\beta > 0$ is the scale parameter. Specially, when $\mu = 0$ and $\beta = 1$, the distribution becomes the standard Gumbel distribution. Generalizations of the Gumbel distribution, which are of flexible skewness and kurtosis due to the addition of one more shape parameter are widely used for extreme value data as they better fit data

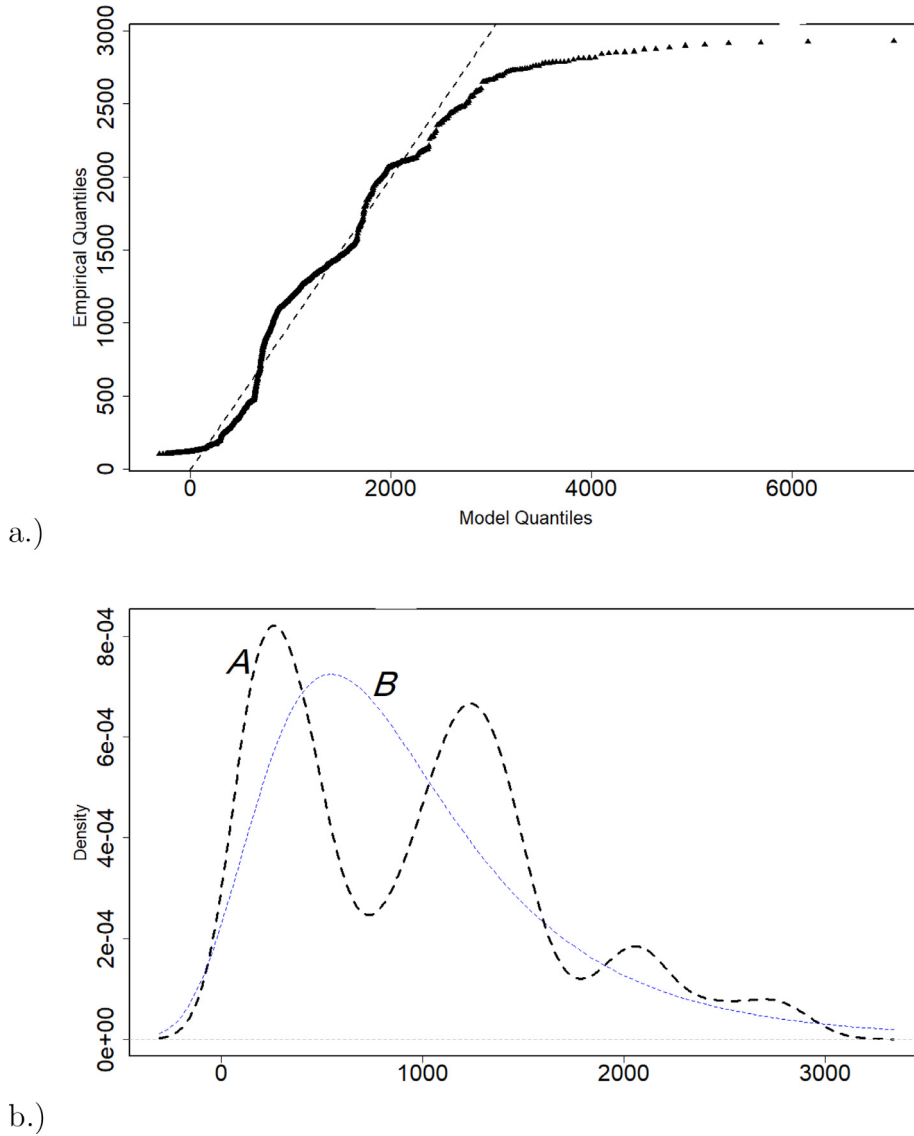


Fig. 6. (a) A general extreme value QQ-plot with maximum likelihood estimation; (b) Density plot of empirical data where a dashed curve *A* is based on the empirical data, and a dashed curve *B* is modeled. $N = 2036$ and bandwidth is 135.9.

[40]. The distribution in (4.1) has been employed as a model for extreme values [41,42]. The distribution has a light right tail, which declines exponentially, since its skewness and kurtosis coefficients are constant.

The *Fréchet distribution*, also known as the Extreme Value Type II distribution, has the following *pdf* and *cdf*, respectively:

$$f(x; \alpha, \beta) = \frac{\alpha}{\beta} \left(\frac{\beta}{x}\right)^{\alpha+1} e^{-\left(\frac{\beta}{x}\right)^{\alpha}},$$

$$F(x; \alpha, \beta) = e^{-\left(\frac{\beta}{x}\right)^{\alpha}}$$

where $\alpha > 0$ is the shape parameter and $\beta > 0$ is the scale parameter.

The *Weibull distribution* is known as the Extreme Value Type III distribution. The *pdf* and *cdf* of a Weibull random variable are shown as follows, respectively:

$$f(x; \lambda, k) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$F(x; \lambda, k) = \begin{cases} 1 - e^{-(x/\lambda)^k} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Table 1

Parameter estimates for the GEV fitted model with maximum likelihood estimator. The 95% confidence intervals for each estimate are included.

	Location $\hat{\mu}$	Scale $\hat{\sigma}$	Shape $\hat{\xi}$
Estimated parameter	606.3260	511.1713	0.1215
95% a lower bound of the confidence interval	576.43	486.42	0.05
95% an upper bound of the confidence interval	636.22	535.92	0.19

where $\lambda > 0$ is the scale parameter and $k > 0$ is the shape parameter.

Further we show the application of the GEV model to the stock market close price using the *weekly-return* data that was calculated by

$$R(t) = \frac{(\text{maximum close price of week } t) - (\text{maximum close price of week } (t - 1))}{(\text{maximum close price of week } (t - 1))}.$$

The results of fitting the GEV distribution to (weekly) block maxima data is presented in Fig. 6 and Table 1 that present the Quantile-quantile plot (QQ-plot), quantiles from a sample drawn from the fitted GEV pdf against the empirical data quantiles with 95% confidence bands. The maximum likelihood estimators of the GEV distribution are the values of the three parameters (μ, σ, ξ) that maximize the log-likelihood. The magnitude along with positive sign of ξ indicates the fat-tailness of the weekly-return data, which is consistent with the quantile plot.

Based on the statistical analysis presented above (Fig. 6a), we see that the distribution of the weekly-return data can be described by a combination of different distributions. The density plot (Fig. 6b) having two humps validates the idea of a mixture of distributions.

4.2. How to choose a threshold

The classical approach for modeling extreme events is based on the GPD. It was proved [43] that if a threshold u is chosen and $X_1, X_2, \dots, X_n$ are observations above u , then the limiting distribution for excess over threshold is indeed GPD. In applications, the GPD is used as a tail approximation [44] of values $x - u$ exceeding the threshold u . The GPD is determined by scale and shape parameters $\sigma_u > 0$ and ξ , respectively, or in terms of threshold excess $x - u$, producing the following formula

$$G(x|u, \sigma_u, \xi) = \Pr(X < x|X > u) = \begin{cases} 1 - [1 + \xi \left(\frac{x-u}{\sigma_u}\right)]_+^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp\left[-\left(\frac{x-u}{\sigma_u}\right)\right]_+, & \xi = 0 \end{cases} \quad (4.3)$$

where $f_+ = \max(f, 0)$. When $\xi > 0$, it takes the form of the ordinary Pareto distribution. This case is the most relevant for financial time series, since it is heavy-tailed. For security returns or high-frequency foreign exchange returns, the estimates of ξ are usually less than 0.5. When $\xi = 0$, the GPD corresponds to the exponential distribution [45].

There are several properties of GPD [43], such as, ‘threshold stability’ property: if X is GPD and $u > 0$, then $X - u$ provided $X > u$ is also GPD. Therefore, a Poisson process of exceedance times with generalized Pareto excess implies the classical extreme value distributions [46]. The above suggests that generalized Pareto distribution is a practical tool for statistical estimation of the extreme values, given a sufficiently high threshold. The rest of this chapter is devoted to a question about how high a threshold should be.

4.2.1. Graphical approaches to estimate threshold

One of the most common ways to determine a suitable threshold is to graphically inspect data. This approach [44] requires substantial expertise, that can be subjective and time consuming. In some cases, when dealing with several data sets, a uniform threshold may be proposed and kept fixed making the entire evaluation even more subjective.

The most common graphical tools are: mean excess plot [47], threshold stability plot [44], QQ-plot [48], Hill plot [49], return level plots [44], etc.

The mean excess plot is a tool widely used in the study of risk, insurance and extreme values. One use is in validating a generalized Pareto model for the excess distribution. The distribution of the excess over a threshold u for a random variable X with distribution function F is defined as

$$F_u(x) = \Pr(X - u \leq x|X > u). \quad (4.4)$$

This excess distribution is the foundation for peaks over threshold modeling which fits appropriate distributions to data on excesses and widespread with many application in hydrology [50,51], actuarial science [52,53], survival analysis [54]. This modeling is based on the GPD that is suitable for describing properties of excesses. The mean excess (ME) function is one of the most common tools to determine a suitable threshold u . The ME function of a random variable X is defined as

$$M(u) = E(X - u|X > u), \quad (4.5)$$

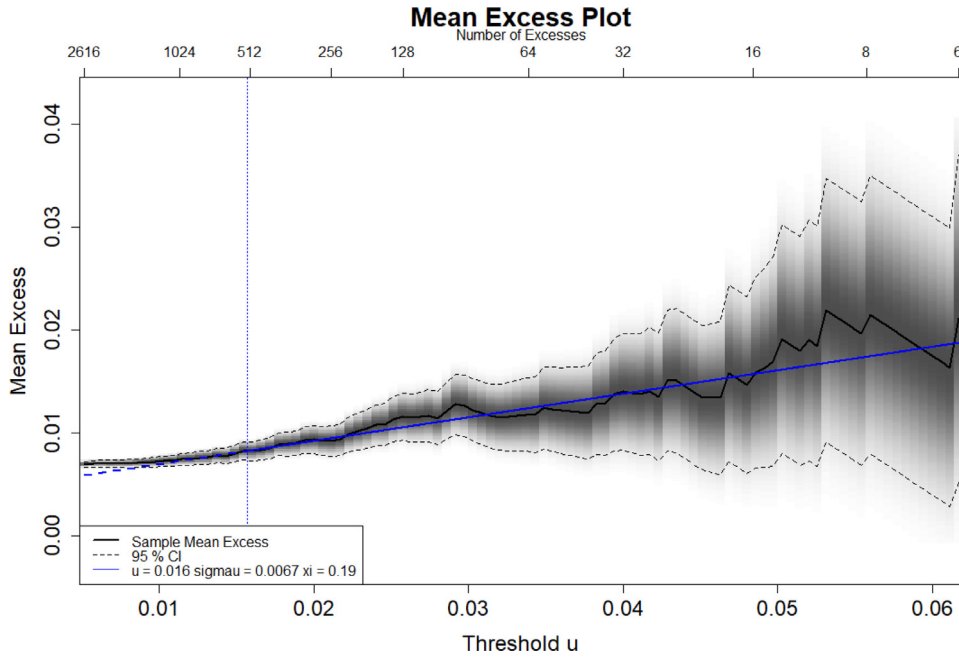


Fig. 7. Mean residual life plot for the S&P 500 positive returns. Solid jagged line is empirical MRL with approximate pointwise Wald 95% confidence intervals as dashed lines. The threshold u is estimated at 0.016. A vertical dashed line marks this threshold.

provided $EX_+ < +\infty$, which is also known as mean residual life function. As Ghosh [47] noted for a random variable $X \approx G_{\xi, \beta}$, $E(X) < +\infty$ if and only if $\xi < 1$ and in this case, the ME function of X is linear in u :

$$M(u) = \frac{\beta}{1-\xi} + \frac{\xi}{1-\xi}u, \quad (4.6)$$

where $0 \leq u < +\infty$ if $\xi \in [0, 1)$ and $u \in [0, -\frac{\beta}{\xi}]$ if $\xi < 0$. The linearity of the ME function characterizes the GPD class [47]. Davison and Smith [46] developed a simple graphical tool that checks data against a GPD model. Let $X_1 \geq X_2 \geq \dots \geq X_n$ be the order statistics of the data, then ME plot depicts the points $\{X_k, \hat{M}(X_k) | 1 < k \leq n\}$, where $\hat{M}$ is the empirical ME function defined as

$$\hat{M}(u) = \frac{\sum_{i=1}^n (X_i - u) I_{[X_i > u]}}{\sum_{i=1}^n I_{[X_i > u]}}, \quad u \geq 0. \quad (4.7)$$

If the ME plot is close to linear for sufficiently large values of the threshold then there is no evidence against use of a GPD model. Another problem is to obtain a natural estimation $\hat{\xi}$ of ξ . There are several methods to estimate $\hat{\xi}$, such as: (i) Least squares [44], (ii) Maximum likelihood estimation [46], (iii) the Hill estimator [49], (iv) the Pickands estimator [55], (v) quantile-quantile plot (QQ-plot) [44], (vi) the moment estimator [56].

For example, the QQ plot depicts the points

$Q_{m,n} = \{(-\log(\frac{i}{m}), \log(\frac{X_i}{X_m})) | 1 \leq i \leq m\}$ where $m < n$ and $\xi > 0$. In case $\xi < 0$, QQ plot is the plot of points $Q'_{m,n} = \{(X_i, G_{\xi,1}^{-1}(\frac{i}{n+1})) | 1 \leq i \leq n\}$, where $\hat{\xi}$ is an estimate of ξ based on m upper order statistics.

Recently, new graphical diagnostic tools have been introduced: a new multiple-threshold GP model with a piece-wise constant shape parameter [57]; plots measuring a surprise at different candidates to select threshold, using results of Bayesian statistics [58,59]; structure of maximum likelihood estimators have been studied to develop diagnostic plots with more direct interpretability [60].

With this choice for a threshold, $u = 0.016$, we get 507 exceedances with the empirical life becoming close to linear above this choice of the threshold (Fig. 7). Similarly, we find a threshold for negative returns. In this case, all computations were repeated for absolute values of negative returns (Fig. 8). With this choice for the threshold $u = 0.017$, there are 462 exceedances. One can see that empirical MRL becomes almost linear above $u = 0.017$.

Ghosh and Resnick [47] noted that despite graphical diagnostic is a tool commonly accepted by practitioners, there are some problems associated with the methods mentioned above, such as: (i) an analyst needs to be convinced that $\xi < 1$ since for $\xi \geq 1$ random sets are the limits for the normalized ME plot. Such random limits lead to wrong impressions. Certain methods described above work with ξ defined on specific intervals; (ii) in this case distributions are not close to GPD can mislead the mean excess diagnostics.

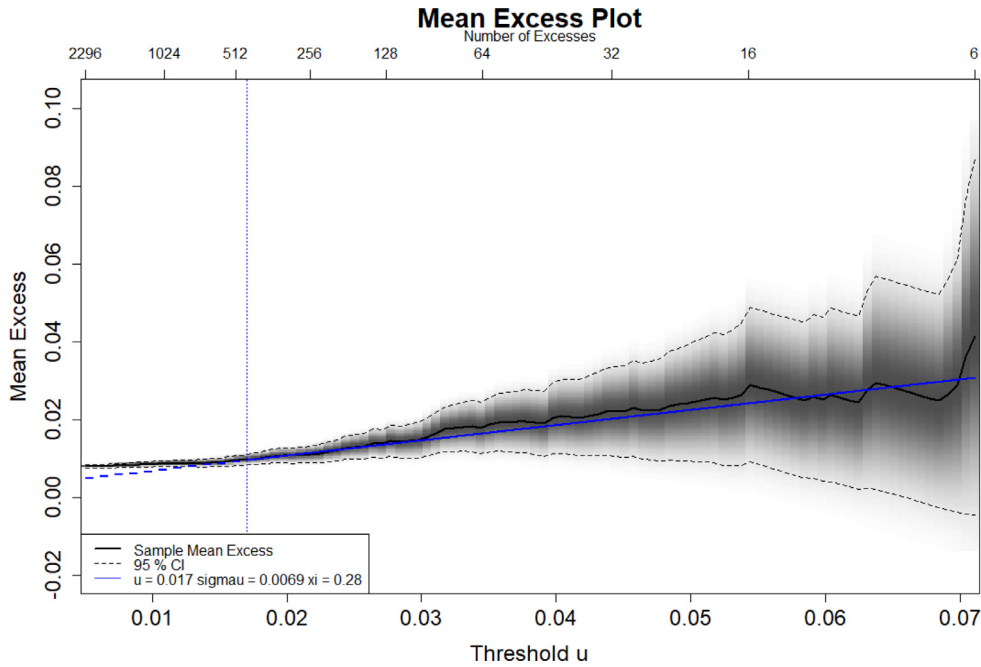


Fig. 8. Mean residual life plot for the S&P 500 negative returns. Solid jagged line is empirical MRL with approximate point-wise Wald 95% confidence intervals as dashed lines. The threshold u is estimated at 0.017. A vertical dashed line marks this threshold.

Based on the graphical approach, the threshold for negative return of S&P 500 was chosen at $u = -0.017$ and for positive return at $u = 0.016$. Distribution of exceedances over respective thresholds are shown in the Fig. 9.

The QQ-plots shown in the Fig. 10 demonstrate that with a few exceptions, exceedances follow the GPD.

4.2.2. Automatic methods to estimate thresholds

As was mentioned above, the graphical approaches as well as rules of thumb can be highly objective, time consuming, and require certain professional background. Thus some authors have proposed automatic selection methods that can treat chunks of Big Data: a pragmatic automated, simple and computationally inexpensive threshold selection method based on the distribution of the difference of parameter estimates when the threshold is changed [61]: it was shown that better performance is demonstrated by graphical methods and Goodness of Fit metrics that rely on pre-asymptotic properties of the GPD [62] using weighted least squares to fit linear models in the traditional mean residue life plot; the recently developed stopping rule ForwardStop [63], which transforms the results of ordered, sequentially tested hypotheses to control the false discovery rate [64] that provides reasonable error control [58].

A particular interest has a method that suggests a way to determine threshold automatically without time consuming and rather subjective visual approaches based on L-moments of GPD that summarize probability distributions, perform estimation of parameters and hypothesis testing [58].

Probability weighted moments, defined by Greenwood [65], are precursors of L-moments. Sample probability weighted moments computed from data values $X_1, X_2, \dots, X_n$ arranged in increasing order, are given by

$$b_0 = \frac{1}{n} \sum_{j=1}^n X_j, \quad b_r = \frac{1}{n} \sum_{j=r+1}^n \frac{(j-1)(j-2) \dots (j-r)}{(n-1)(n-2) \dots (n-r)} X_j.$$

L-moments are certain linear combinations of probability weighted moments that have simple interpretations as measure of location, dispersion and shape of the data sample. The first few L-moments are defined by

$$\begin{aligned} \lambda_1 &= b_0, & \lambda_2 &= 2b_1 - b_0, \\ \lambda_3 &= 6b_2 - 6b_1 + b_0, & \lambda_4 &= 20b_3 - 30b_2 + 12b_1 - b_0. \end{aligned}$$

(the coefficients are those of the shifted Legendre polynomials).

The first L-moment is the sample mean, a measure of location. The second L-moment is (a multiple of) Gini's mean difference statistic, a measure of the dispersion of the data values about their mean. By dividing the higher-order L-moments by the dispersion measure, we obtain the L-moment ratios, $\tau_r = \frac{\lambda_r}{\lambda_2}$, $r = 3, 4, \dots$

These are dimensionless quantities, independent of the units of measurement of the data. τ_3 is a measure of skewness and τ_4 is a measure of kurtosis - these are respectively the L-skewness and L-kurtosis. They take values between -1 and

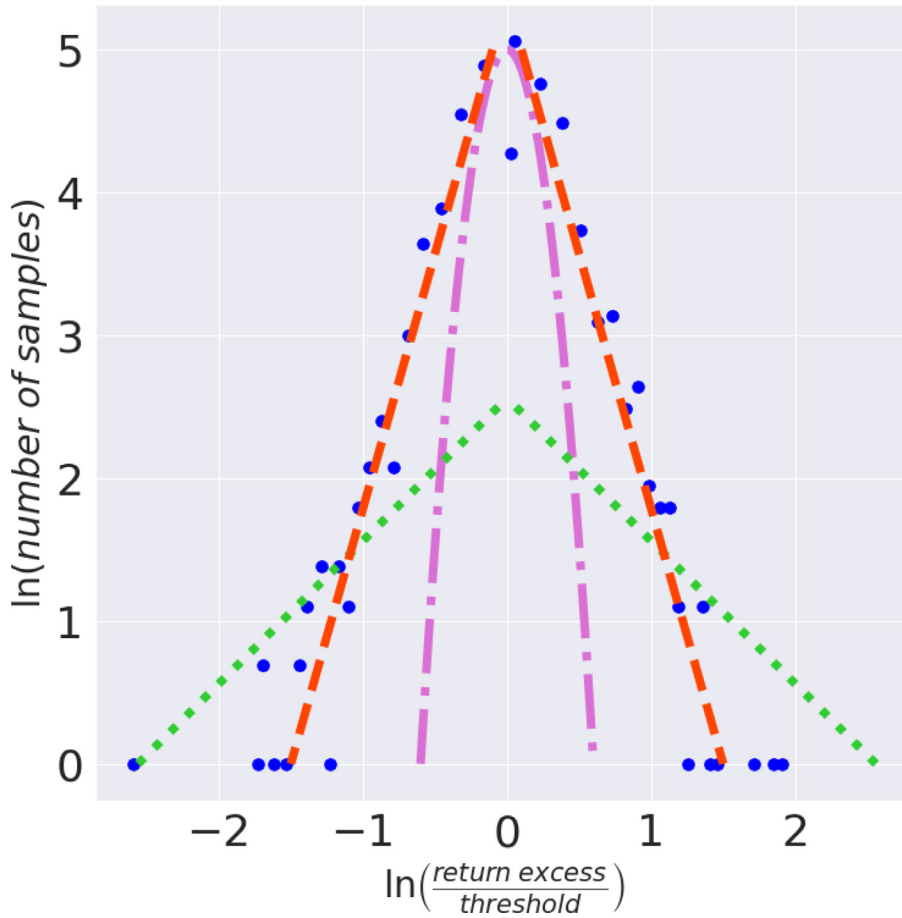


Fig. 9. Distribution of exceedances normalized by thresholds $u = 0.016$ for positive and $u = -0.017$ for negative returns, respectively. Dotted lines represent Zipf's Law ($\propto x^{-1}$), dash dot lines represent Gaussian distribution ($\propto \exp(-x^2)$) and dashed lines represent power law ($\propto x^{-3.3}$). The curves are given for reference only.

+1. For random variable with GPD with $\xi < 1$, the particular relationship between L-skewness and L-kurtosis is defined as

$$\tau_4 = \tau_3 \frac{1 + 5\tau_3}{5 + \tau_3}.$$

Given a sample $x_1, x_2, \dots, x_n$, the Automatic L-moment Ratio Selection Method (ALRSM) works as follows [58]:

1. Define the set of candidate thresholds $\{u_i\}_{i=1}^I$ as $I = 20$ sample quantiles, starting at 25% by steps of 3.7%.
2. Compute the sample L-skewness and L-kurtosis for the excess over each candidate threshold $(\tau_{3,u_i}, \tau_{4,u_i})$ and determine d_{u_i} - the Euclidean distance:

$$d_{u_i} = \sqrt{(\tau_{3,u_i} - \tau_3)^2 + (\tau_{4,u_i} - g(\tau_3))^2},$$

for $i = 1, \dots, I$ with

$$g(\tau_3) = \tau_3 \frac{1 + 5\tau_3}{5 + \tau_3}.$$

3. The threshold after which the behavior of the tail of the underlying distribution can be considered approximately GPD is then automatically selected as

$$u^* = \operatorname{argmin}_{u_i, 1 \leq i \leq I} \{d_{u_i}\},$$

that is, the level above which the corresponding L-statistics fall closest to the curve.

Using the L-moments method, we computed thresholds for S&P 500 log-return depending on an observation period, 100 trading days and 400 trading days. The Fig. 11 indicates that a threshold not only time dependent, but also depends on the size of data set used. Once again, we cannot choose one value of the threshold that can be absolutely accurate.

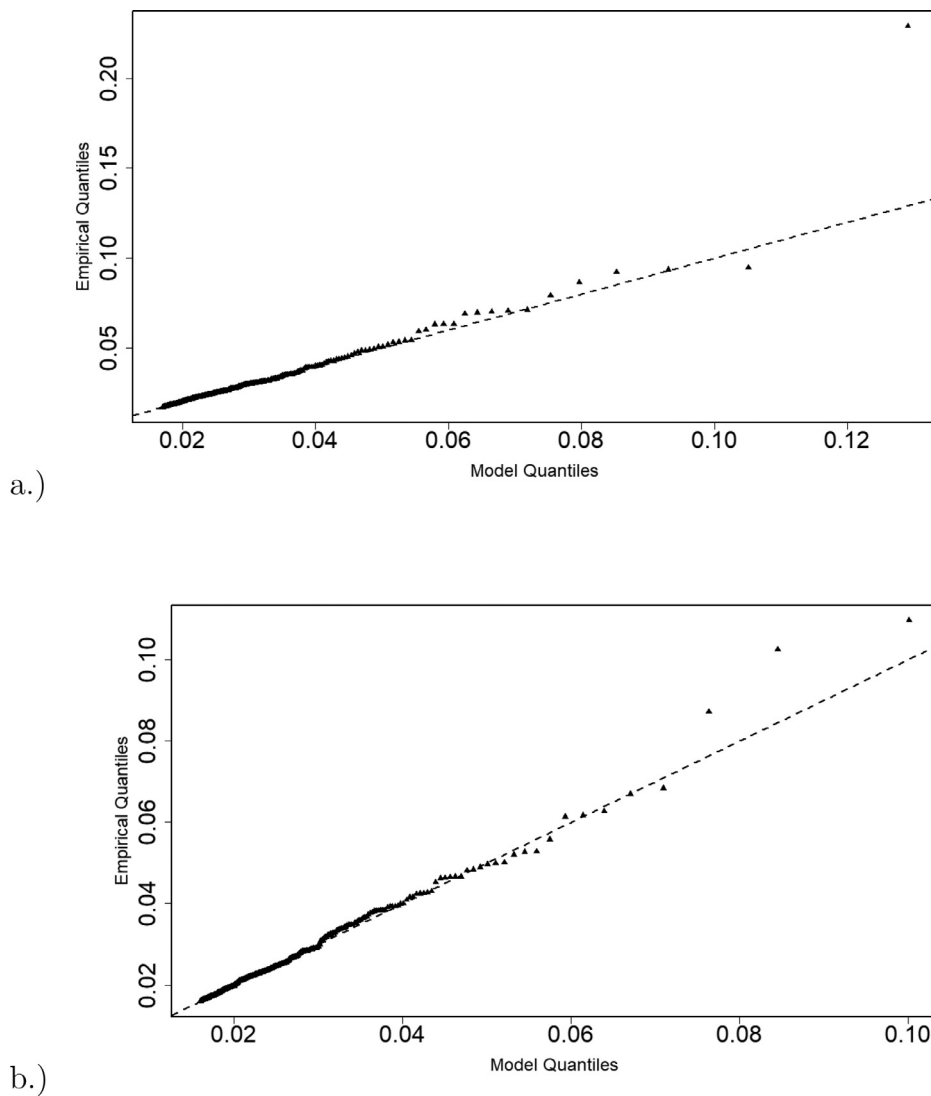


Fig. 10. (a) Quantile-Quantile plot with maximum likelihood estimation for the negative threshold; (b) QQ-plot with maximum likelihood estimation for the positive threshold.

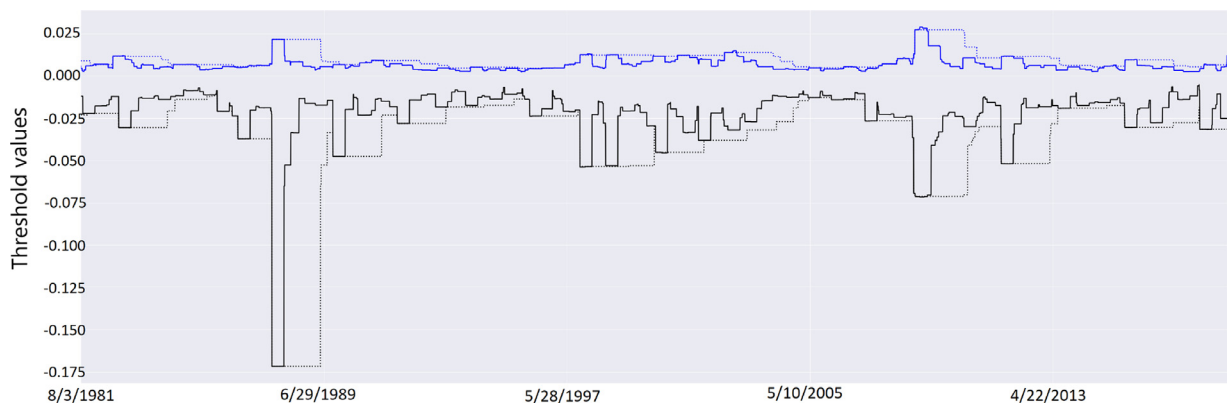


Fig. 11. Value of the thresholds for positive and negative log-return based on L-moments. The solid black and blue lines correspond to negative and positive log return thresholds, respectively, and based on a window of 100 trading days. The dotted black and blue lines correspond to negative and positive log return thresholds, respectively, and based on a window of 400 trading days. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

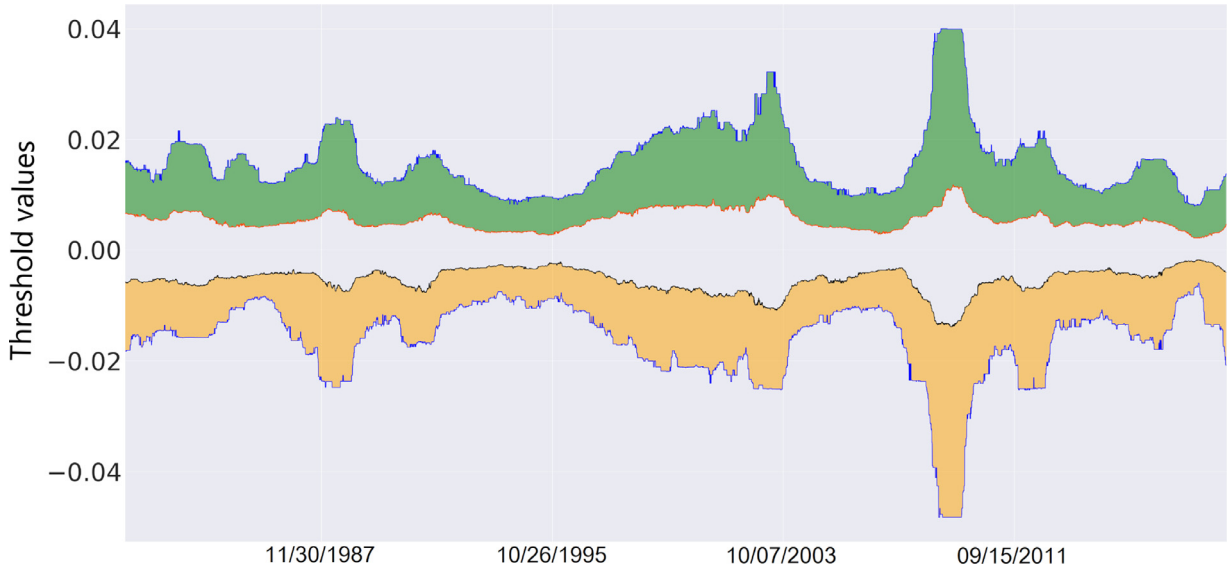


Fig. 12. Possible threshold changing ranges from 03/11/1981 to 12/31/2018 based on 300 preceding trading days. A green strip represents positive log-return and an orange strip shows the threshold domain for negative log-return values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2.3. Rules of thumb to choose a threshold

As earlier noted, the threshold sequence is a function of the properties of the GPD provided that a population is in the domain of attraction of the GPD. In case a distribution function F is known, derivation of the threshold selection is possible, however, in practice if F is unknown then there is no general form for the threshold sequence [44].

Practitioners often use so-called rules of thumb, many of them have little to know theoretical justification, for instance, a fixed quantile rule [66]: the upper 10% rule or its derivative 5% rule; the square root rule $k = \sqrt{n}$ [67]. There is a procedure that tries to find a region of stability among the estimates of the extreme value index [68]. This method depends on a tuning parameter, whose choice is further analysed in Neves and Alves [69]. Unfortunately, no theoretical analysis exists for this approach [70]. More comprehensive reviews of the threshold selection methods can be found in Scarrott and MacDonald [44].

Even though most of methods mentioned above have no theoretical justification for an exact value of a threshold, we can find an approximate location for a threshold given a data set R that has M values of log-return, $R = [x_1, x_2, \dots, x_M]$.

1. Let R_n be a list of n consecutive values of log-return, $R_n = [x_1, x_2, \dots, x_n]$. Split R_n into two parts: $R_{n+} = \{x \in R_n | x \geq 0\}$ and $R_{n-} = \{x \in R_n | x < 0\}$; then sort them in an increasing order such that

$$R_{n+} = [x_1^+, x_2^+, \dots, x_p^+],$$

$$R_{n-} = [x_1^-, x_2^-, \dots, x_q^-].$$

where $p, q \in \mathbb{N}$ such that $p + q = n$ and n is the width of observation or a window of observation.

2. Next, we compute medians of R_{n+}, R_{n-} , call them x_{m+} and x_{m-} . These values will be a lower bound for a positive threshold and an upper bound for a negative threshold, respectively.
3. At this step, an upper bound for a positive threshold x_u and a lower bound for a negative threshold x_l are estimated using the rule of thumb, namely, the fixed quantile rule with upper 10% for positive log-return values and lower 10% for negative log-return values. With this we have

$$R_{n+} = [x_1^+, x_2^+, \dots, x_{m+}, \dots, x_u, \dots, x_p^+],$$

$$R_{n-} = [x_1^-, x_2^-, \dots, x_l, \dots, x_{m-}, \dots, x_q^-].$$

The indices u, l can be found as $u = \lfloor p/1.111 \rfloor$, $l = \lceil q/10 \rceil$.

4. A threshold for negative log-return values ranges from x_l to x_{m-} and a threshold for positive values is within x_{m+} and x_u , based on n observations from R_n .

Further, we shifted R_n to $R_n = [x_2, x_3, \dots, x_{n+1}]$ to estimate new values of thresholds based on previous n observations. The process repeated until we exhausted the entire data set R .

We chose a window of 300 days that was moving over the entire dataset producing a domain for threshold existence as shown in the Fig. 12. It is clear from the Fig. 12 that certain values of thresholds cannot sustain the entire period and must be updated from time to time.

Similarly, the window of 300 trading days, the Fig. 13 demonstrates how ranges of thresholds change as we move a window of 600 trading days across available data. Once again, some values of thresholds can exist for almost entire period, while other can exist a few months and then must be replaced with an updated value.

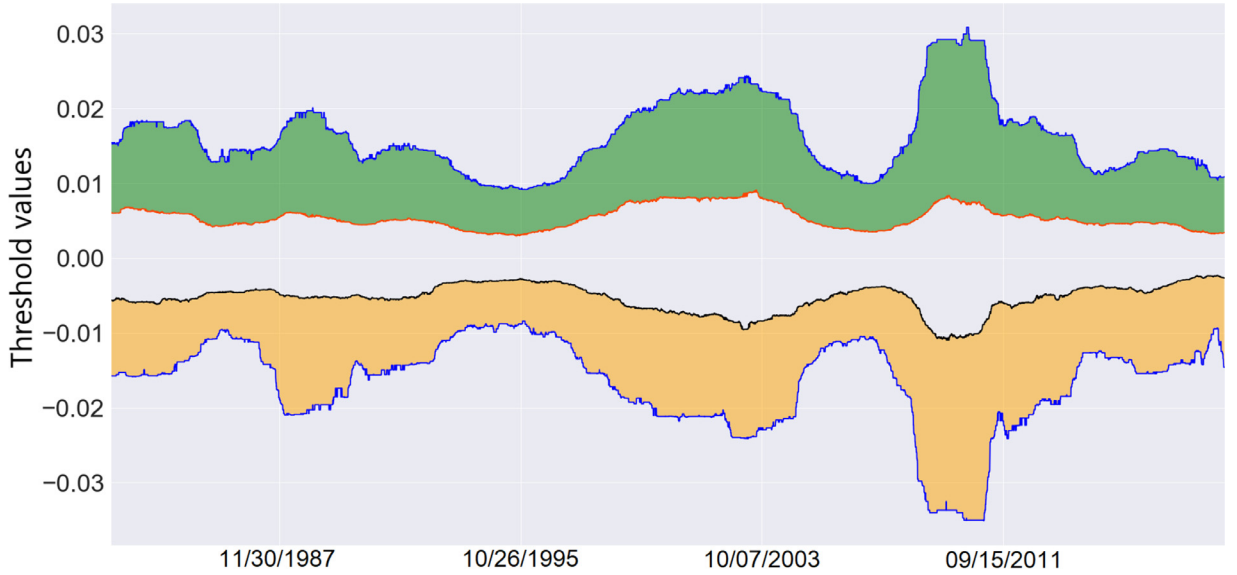


Fig. 13. Possible threshold changing ranges from 05/18/1982 to 12/31/2018 based on 600 preceding trading days. A green strip represents positive log-returns and an orange strip shows the threshold domain for negative log-return values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Based on Figs. 12 and 13 we can see that a choice of the threshold depends on a size of the data set used. Moreover, the rules of thumbs alike graphical approaches require practitioners' involvement in studying data before making a final choice for thresholds.

Equipped with these results, we can compute a validity period $\tau(u)$ for a threshold u . Let n be the window of observation. By moving the window over the data set containing M values of log-returns we obtain four lists with $M - n + 1$ elements in each

$$\begin{aligned} X_{m+} &= [x_{m+,n}, x_{m+,n+1}, \dots, x_{m+,M}] & X_{m-} &= [x_{m-,n}, x_{m-,n+1}, \dots, x_{m-,M}] \\ X_u &= [x_{u,n}, x_{u,n+1}, \dots, x_{u,M}] & X_l &= [x_{l,n}, x_{l,n+1}, \dots, x_{l,M}] \end{aligned}$$

containing medians of positive and negative log-returns, upper 10% cut-offs and lower 10% cut-offs for positive and negative log-returns, respectively, as computed previously.

1. First, we compute the k threshold candidates u_i for positive log-return as $\{u_i^+ \in \mathbb{R} \mid \min(X_{m+}) + \frac{i}{k}(\max(X_u) - \min(X_{m+})), i = 1, \dots, k\}$. In a similar fashion, we compute a set of threshold candidates for negative log-returns, $\{u_i^- \in \mathbb{R} \mid \min(X_l) + \frac{i}{k}(\max(X_{m-}) - \min(X_l)), i = 1, \dots, k\}$.
2. For each threshold candidate u_i found in the previous step we compute its validity duration, i.e. a number of days N_{u_i} a candidate would fall within admissible ranges. Set $N_{u_i} = 0$. For $j = n \dots M$ if $x_{m+,j} < u_i < x_{u,j}$, then $N_{u_i} = N_{u_i} + 1$ for positive threshold candidates, similarly, if $x_{l,j} < u_i < x_{m+,j}$, then $N_{u_i} = N_{u_i} + 1$ for negative threshold candidates.
3. The probability that a threshold candidate will be changed on any given day is $\eta(u_i) = 1 - N_{u_i}/(M - n + 1)$.

The Fig. 14 shows probabilities that a particular choice of the threshold can be changed on any day depending on the size of the dataset used, which is important to know especially when new data becomes available and is included for consideration. The more data we use to estimate a value of a threshold, the more likely the threshold will stay unchanged, however, as we add data, the threshold should be reconsidered. It also brings another issue: in many cases, statistical analysis is performed on a historical dataset that does not reflect a phenomenon we study at the present time.

4.3. Statistics of extreme events under threshold uncertainty

The statistics of extreme events under threshold uncertainty can be described with the help of a simple model, in which the log-return of the index and the value of threshold are treated as random variables that yet can change inconsistently. The model that we are going to adopt and modify had been put forward by us for the first time in Floriani et al. [71] to describe the behavior of systems close to a threshold of instability and used later in Volchenkov [72], [73] to model survival under uncertainty and the events of mass extinction. In the model of extreme events under threshold uncertainty, the current value of the log-return is quantified by a random number $x \in [0, 1]$ drawn accordingly some probability distribution function $\Pr\{x < z\} = F(z)$. The threshold value that might change any time once the new data become available is another random number $y \in [0, 1]$, which is drawn from another probability distribution function, $\Pr\{y < z\} = G(z)$. We assume

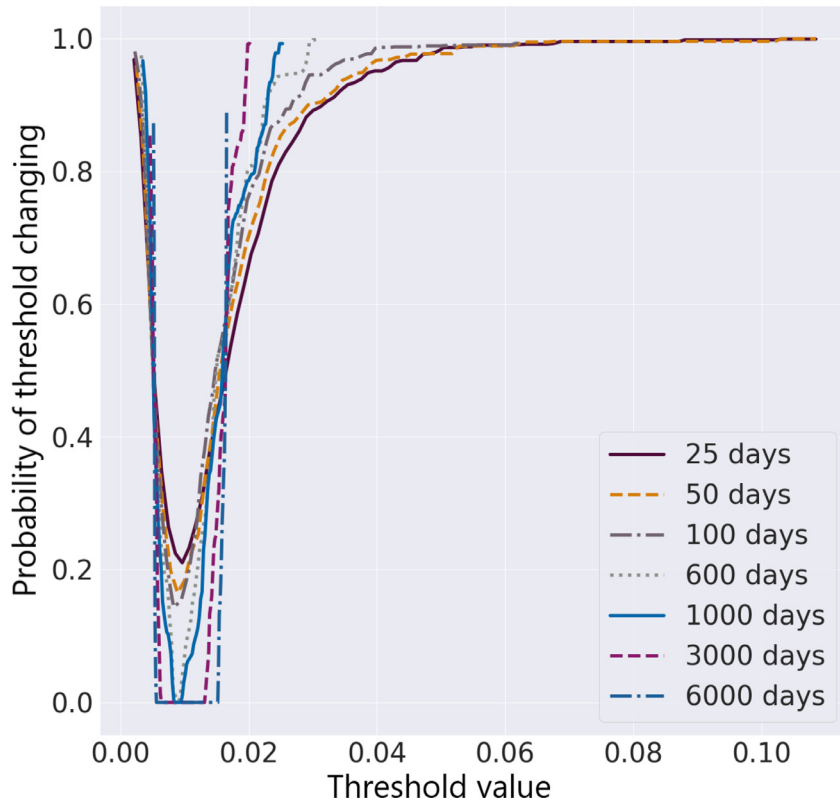


Fig. 14. The probability that a threshold of the log-return values will be changed on any given day calculated over the different data windows ranging from 25 to 6000 trading days.

that the rate of daily variations of the log-return values is greater than or equal to that of the threshold values, ultimately determining whether the current log-return value is extreme or not. In fact, it is the relative rate of random updates of x and y described in our model by the probability of inconsistency $\eta \geq 0$, that actually determines the statistics of extreme events.

At time $t = 0$, the value of log-return x is chosen with respect to the probability distribution function F , and the value of the threshold y is chosen with respect to the probability distribution function G . If $y \geq x$, the event is regular and the process keeps going to time $t = 1$. At time $t \geq 1$, either, with probability $\eta \geq 0$, the value of log-return x is drawn anew from the probability distribution function F , but the threshold keeps the value y it had at time $t - 1$, or with probability $1 - \eta$, the value of log-return x is updated anew from the probability distribution function F , and the level of supply y is updated either with respect to the probability distribution function G . As long as the value of threshold is not exceeded ($x \leq y$), the event is classified as regular, but the event is extreme whenever $x > y$.

The value of probability $\eta > 0$ can be interpreted as the reciprocal characteristic time interval, during which the threshold level remains unchanged, and vice versa the probability that a threshold of the log-return values will stay unchanged on any given day can be calculated as the inverse of the time interval during which the threshold value stays put. The probability that a threshold of the log-return values will stay unchanged on any given day calculated over the different data windows ranging from 25 to 6000 trading days is shown in Fig. 14.

We are interested in the probability distribution $P_\eta(t)$ of the interval duration t between the sequential extreme events for some probability distribution functions F and G and a given value of the probability $\eta \geq 0$. A straightforward computation [71,72], shows that, independently upon the value of η , the initial probability of choosing the level of demand below the supply level (to start the subsistence process) is $\int_0^1 dG(z)F(z)$. The general formulas for $P_\eta(t)$ can be found in Floriani et al. [71], Volchenkov [72].

When $\eta = 0$, the values of log-return and the threshold value are updated coherently the resulting probability function,

$$P_{\eta=0}(t) = \left[\int_0^1 dG(z)F(z) \right]^t \int_0^1 dG(z)(1 - F(z)), \quad (4.8)$$

decays exponentially fast with t , for any choice of the probability distribution functions F and G . In particular, if the threshold level and log-returns are drawn uniformly at random, $dF(z) = dG(z) = dz$, over the interval $[0,1]$, the occurrence of an

extreme event is statistically equivalent to simple flipping a fair coin, for which head and tail come up equiprobably,

$$P_{\eta=0}(t) = \frac{1}{2^{(t+1)}}. \quad (4.9)$$

On the contrary, when the threshold level is kept unchanged, $\eta = 1$, the statistics of intervals between the sequential extreme events,

$$P_{\eta=1}(t) = \int_0^1 dG(z)F(z)^t(1-F(z)), \quad (4.10)$$

decays asymptotically algebraically as $t \gg 1$ [71,72]. For example, in the special case of uniformly random updates of the threshold and log-return values, the probability function (4.10) decays algebraically as

$$P_{\eta=1}(t) = \frac{1}{(t+1)(t+2)} \simeq \frac{1}{t^2}. \quad (4.11)$$

For a general family of invariant measures of a map of the interval $[0,1]$ with a fixed neutral point defined by the probability distributions F and G , absolutely continuous with respect to the Lebesgue measure, i.e.,

$$\begin{aligned} dF(z) &= (1+\alpha)z^\alpha dz, \quad \alpha > -1, \\ dG(z) &= (1+\beta)(1-z)^\beta dz, \quad \beta > -1. \end{aligned} \quad (4.12)$$

Eq. (4.10) gives the probability function that exhibits the power law asymptotic decay for $t > 1$ [71–73]:

$$P_{\eta=1}(t) \simeq \frac{(1+\beta)\Gamma(2+\beta)(1+\alpha)^{-1-\beta}}{t^{2+\beta}} \left(1 + O\left(\frac{1}{t}\right)\right). \quad (4.13)$$

The asymptotic decay of (4.13) seems to be algebraic,

$$P_{\eta=1}(\tau) \simeq \frac{1}{\tau^{2+\beta}}, \quad (4.14)$$

for $\beta > -1$, for any choice of the distributions F and G although it is mainly the character of the probability function G that determines the rate of decay of $P_{\eta=1}(t)$ with time. In the limiting case when the support of the probability distribution $G(x)$ determining the choice of the supply level is concentrated close to $x = 1$, i.e., is zero everywhere in the interval $[0,1]$, except for a small interval of length ε up to 1, the Zipf power law asymptote $\propto t^{-1-\varepsilon}$, $\varepsilon > 0$, follows directly from (4.13) [71–73]. A possible modeling function for such a bountiful probability distribution, forming a thin spike as $x \rightarrow 1$, can be chosen in the form,

$$G_\varepsilon(x) = 1 - (1-x)^\varepsilon, \quad \varepsilon > 0, \quad (4.15)$$

with the probability density in the interval $[0,1]$,

$$dG_\varepsilon(x) = \frac{\varepsilon dx}{(1-x)^{1-\varepsilon}}. \quad (4.16)$$

The straight line shown in Fig. 15 represents the hyperbolic decay of time intervals between the extreme events.

5. Defining emergency scales by thresholds uncertainty.

Once we assume statistics for log-returns, or ranges for threshold values are defined using different methods shown in Section 4.1, we introduce uncertainty to the threshold value that depends on the method itself and the amount of data we use. We measure uncertainty to justify an emergency scale to represent the extreme events.

Using the rule of thumb (Section 4.2.3), we determined the ranges for the threshold values depending on the window of observation (Fig. 12) for both positive and negative log return values. For each admissible positive choice of a threshold u , we determine the probabilities $\eta(u)$ that the threshold u can be changed on any given day (Fig. 14), degrees of uncertainty have been assessed by the means of the Shannon's entropy for each threshold value and the window of observation [73] (5.1). The case with the negative values of the threshold is analogous.

$$H(\eta) = \begin{cases} -\eta \log \eta - (1-\eta) \log(1-\eta), & 0 < \eta < 1 \\ 0, & \eta = 0, 1 \end{cases} \quad (5.1)$$

where $\eta(u)$ is the probability that the threshold will be changed on any given day (Fig. 14).

We observed the Red Queen State and three emergency scales can be readily interpreted.

- (i) If amount of data is not sufficient (a solid line corresponding to the 18-day window of observation in Fig. 16), then the uncertainty curve forms a skewed profile attaining a single maximum of 0.69295 for the threshold value 0.008272. In this situation, an observer's perception of the events reminds *Red Queen* from "Through The Looking-Glass and What Alice Found There" by Lewis Carroll [21] who said "When you say hill, I could show you hills, in comparison with which you'd call that a valley". Our understanding of the events whether they are extreme or not is very limited

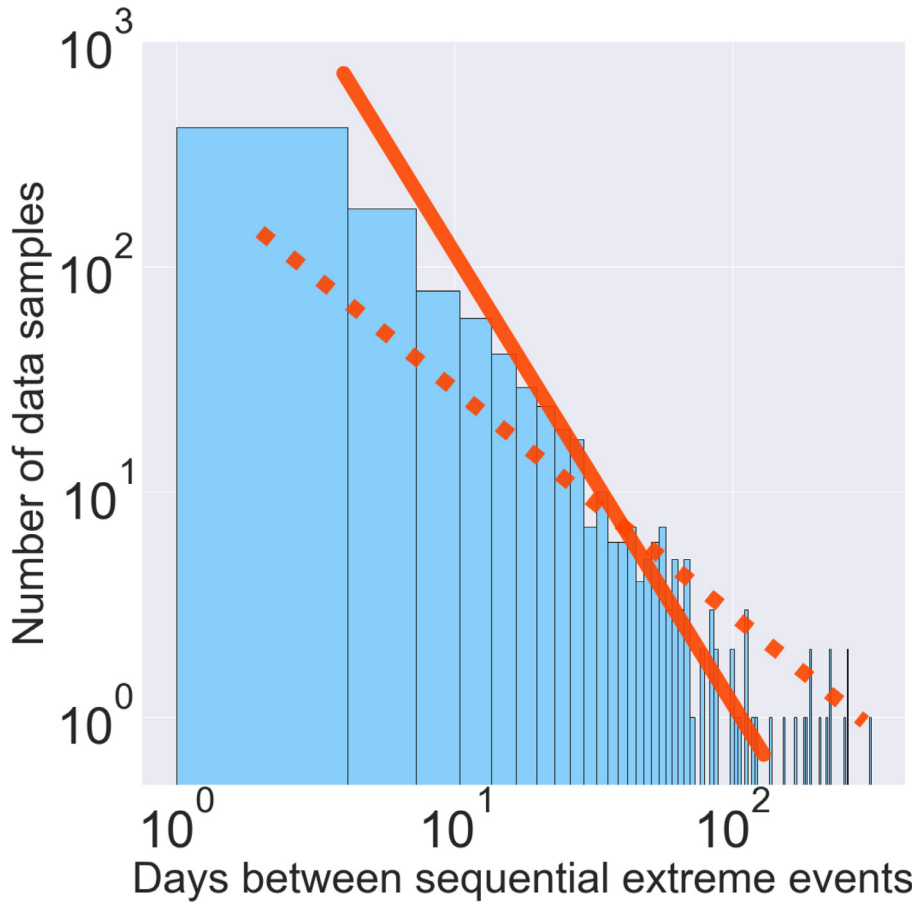


Fig. 15. The statistics of time intervals (in days) between the sequential extreme events for the fixed threshold values, $u = 0.016$ and $u = -0.017$, for positive and negative fluctuations of the log-return respectively. The solid line $\propto t^{-2}$ corresponding to the asymptotic quadratic decay (4.11) is given for reference.

Table 2
Location of extrema points of the uncertainty curves from Fig. 16.

Window	First maximum		Second maximum	
	Threshold	Uncertainty	Threshold	Uncertainty
18 days	0.00827	0.69296	–	–
50 days	0.00461	0.67411	0.01532	0.69290
2000 days	0.00534	0.69315	0.01672	0.69034
4000 days	0.00533	0.68217	0.01580	0.69314

and uncertainty is blurry. As events become more severe, our uncertainty that the events are extreme decreases. The observer realizes that the events are extreme, but a precise point at which the events turn to be severe cannot be determined. This case is called the *Red Queen State*.

- (ii) As the window of observation becomes larger, 25 days, for instance, the uncertainty curve exhibits two maxima indicating that the amount of data is sufficient. This happens because $H(\eta)$ attains its maximum for $\eta = 1/2$ and the η curve admits a value $1/2$ twice (Fig. 14). As we further extend the window the curve torrents into sharp peaks (Fig. 16). The latter ones clearly separate the threshold values into three regions: three levels of emergency. The locations of peaks of the curves are summarized in Table 2. The location of two peaks is not sensitive to the window of observation.

Emergency scales for S&P 500.

- (1) *Subcritical*. In the region I, the threshold values are small to raise concern about extreme events. Then we can observe a spike with the degree of uncertainty attaining its first maximum. This extremum indicates a transition to the next kind of uncertainty. For the window of 2000 days, the region I lies in the interval $[0, 0.00534)$ (see Fig. 16, Table 2).

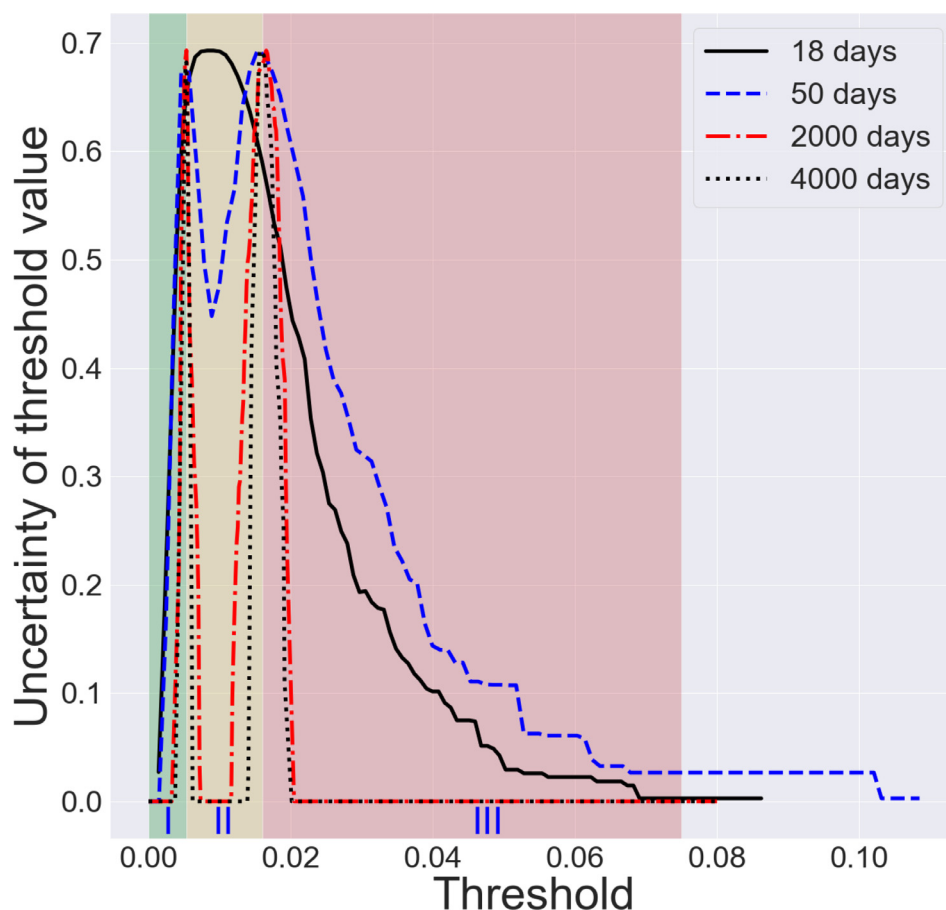


Fig. 16. Degree of uncertainty of different values of the thresholds based on the amount of trading days taken into account. Three shaded regions mark three scales of emergency: I - subcritical, II - critical, III - extreme and the solid curve represents the Red Queen State.

(II) *Critical*. In the region II, the interval $[0.00534, 0.01672)$ for the 2000-day window, uncertainty is conceptualized with a question whether a magnitude of the event is already critical, extreme or not yet. Further, we see another jump of uncertainty reaching 0.69034. At this point we are certain that events are not regular anymore.

(III) *Extreme*. In the case of the window of 2000 days, the interval $[0.01672, \infty)$ constitutes the region III. We consider all events in this region extreme with our uncertainty decreasing as threshold values increase.

With the analysis presented above, we define an emergency scale of three levels based on the three regions of the threshold values corresponding to three peaks of the uncertainty curve. This emergency scale is not sensitive to the size of the window of sufficient amount of data considered.

6. Conclusion

The S&P500 times series in the period from January 2, 1980 till December 31, 2018 exhibits an asymmetrical skewness of the distribution with the right and left power law tails. Multifractal detrended fluctuation analysis of log return time series for S&P 500 index reveals a scale invariant structure for the fluctuations on both small- and large scale magnitudes, as well as its short- and long-range dependence on different time scales. Moreover, the segments with small fluctuations have a random walk like structure whereas segments with large fluctuations have a noise like structure.

We have reviewed different methods of threshold selection and studied the extreme events presented in the time series using different statistical approaches. We found that the distribution of the weekly-return data can be described by a combination of different distributions. Based on a graphical approach for threshold selection, we chose separate thresholds for the positive and negative values of the log return, 0.016 and -0.017 , respectively. With this choice, we registered 507 instances of extreme events corresponding to raise of market and 462 extreme events related to market declines. With a few exceptions, exceedances over (under for negative log return values) the threshold follow the GPD.

The rule of thumb showed that a threshold value depends on the width of observation window, and the threshold can change at any moment, once new data become available. Uncertainty of the threshold values can be determined by the probability of changing the threshold on any given day.

The moment we assume statistics of distributions or the dataset is fixed, it leads to uncertainty of the threshold value which can be resolved by the emergency scales rigid to variation on the size of the dataset.

We suggested a statistical model that describes registration frequency of extreme events under threshold uncertainty. Our model fits well the statistics of occurrence of the extreme values in the S&P 500 time series.

Declaration of Competing Interest

The authors declare no conflict of interests.

Acknowledgments

VS acknowledges the support of Department of Mathematics and Statistics of Texas Tech University provided him a temporary Research Assistant position. We are grateful to Dr. M. Toda for her support. Work was done under a contract #W911W6-13-2-0004 with the AVX Aircraft Company.

References

- [1] Kaplan S. Tolley's handbook of disaster and emergency management: principles and best practices; disaster and emergency management systems. UK: Lexis Nexis; 2004.
- [2] Scale R.. Mercalli scale. <https://www.usgs.gov/media/images/modified-mercalli-intensity-mmi-scale-assigns-intensities>.
- [3] Beaufort. Wind scale. <https://www.spc.noaa.gov/faq/tornado/beaufort.html>.
- [4] Saffir-Simpson. Hurricane wind scale. <https://www.nhc.noaa.gov/aboutsshws.php>.
- [5] Tornado F.. Damage scale. <https://www.spc.noaa.gov/faq/tornado/f-scale.html>.
- [6] Security H.. Advisory system. https://en.wikipedia.org/wiki/Homeland_Security_Advisory_System.
- [7] Climate U.S.. Extremes index (CEI). <https://www.ncdc.noaa.gov/extremes/cei/introduction>.
- [8] Rohn E, Blackmore D. A unified localizable emergency events scale. Int J Inf Syst Crisis Response Manag (IJISCRAM) 2009;1(4):1–14.
- [9] Pisarchik AN, Jaimes-Reátegui R, Sevilla-Escoboza R, Huerta-Cuellar G, Taki M. Rogue waves in a multistable system. Phys Rev Lett 2011;107:274101.
- [10] Pisarchik AN, Grubov VV, Maksimenko VA. Extreme events in epileptic EEG of rodents after ischemic stroke. Eur Phys J Spec Top 2018;227:921–32.
- [11] Plotnick L, Gomez E., White C., Turoff M.. Furthering Development of a Unified Emergency Scale Using Thurstones Law Of comparative Judgment: A Progress Report ABSTRACT. 2007. https://www.dhs.gov/xlibrary/assets/hsas_unified_scale_feedback.pdf.
- [12] Brown K, Harlow WV, Tinic S. Risk aversion, uncertain information and market aptitude. J Financ Econ 1988;22:355–85.
- [13] Corrado C, Jordan BD. Risk aversion, uncertain information and market aptitude reexamining the evidence. Rev Quan Finance Account 1997;8:51–68.
- [14] Halt T. Definition. James Chen - <https://www.investopedia.com/terms/t/tradinghalt.asp>.
- [15] Lee J., Fany Y., Sisson S.A.. Bayesian threshold selection for extremal models using measures of surprise. 2014. arXiv:1311.2994v2[stat.ME].
- [16] WHO. Coronavirus disease 2019 (COVID-19) situation report-43.. <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200303-sitrep-43-covid-19.pdf>.
- [17] Verity R., Okell L., Dorigatti I., Winskill P., Whittaker C., Imai N., et al., "Estimates of the severity of coronavirus disease 2019: a model-based analysis", Lancet Infect Dis. S1473-3099(20)30243-7 (2020) doi: 10.1016/S1473-3099(20)30243-7.
- [18] WHO. Coronavirus disease 2019 (COVID-19) situation report-46.. <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200306-sitrep-46-covid-19.pdf>.
- [19] CFR. Tackling COVID-19: A problem so big, you can see it from space. 2020. <https://www.cfr.org/blog/tackling-covid-19-problem-so-big-you-can-see-it-space>.
- [20] Ioannidis J.. A fiasco in the making? as the coronavirus pandemic takes hold, we are making decisions without reliable data. <https://www.statnews.com/2020/03/17/a-fiasco-in-the-making-as-the-coronavirus-pandemic-takes-hold-we-are-making-decisions-without-reliable-data/>.
- [21] Carroll L. Through the looking-glass and what Alice found there. Chicago: W.B. Conkey Co.; 1900.
- [22] Hudson R, Gregoriou A. Calculating and comparing security returns is harder than you think: a comparison between logarithmic and simple returns. Int Rev Financ Anal 2015;38:151–62.
- [23] Onnela J-P, Chakraborti A, Kaski K, Kertsz J. Dynamic asset trees and black monday. Phys A 2003;1–2:247–52.
- [24] Birru J, Figlewski S. Anatomy of a meltdown: the risk neutral density for the s&p 500 in the fall of 2008. J Financ Mark 2012;15(2):151–80.
- [25] Beran J. Statistics for long-memory processes. Boca Raton, FL: Chapman & Hall/CRC; 1994.
- [26] Ihlen EA. Introduction to multifractal detrended fluctuation analysis in matlab. Front Phys 2012;3:141.
- [27] Thompson JR, Wilson JR. Multifractal detrended fluctuation analysis: practical applications to financial time series. Math Comput Simul 2016;126:63–88.
- [28] Harte D. Multifractals. London: Chapman & Hall; 2001.
- [29] Calvet L, Fisher A. Multifractality in asset returns: theory and evidence. Rev Econ Stat 2002;84:381–406.
- [30] Kantelhardt JW, Zschiegner SA, Koschielny-Bunde E, Havlin S, Bunde A, Stanley HE. Multifractal detrended fluctuation analysis of nonstationary time series. Phys A 2002;316:87–114.
- [31] Smirnov V, Volchenkov D. Five years of phase space dynamics of the standard & poors 500. Appl Math Nonlinear Sci 2019;4(1):203–16.
- [32] Jiang Z-Q, Xie W-J, Zhou W-X, Sornette D. Multifractal analysis of financial markets: a review. Reports on Progress in Physics 2019. <https://doi.org/10.1088/1361-6633/ab42fb>.
- [33] Coles S. An introduction to statistical modeling of extreme values. London: Springer; 2001.
- [34] Fisher RA, LHC T. Limiting forms of the frequency distribution of the largest and smallest member of a sample. Proc Camb Philos Soc 1928;24:180–90.
- [35] Gnedenko BV. Sur la distribution limite du terme maximum d'une série aléatoire. Ann Math 1943;44:423–53.
- [36] Zhang X, Lian J. Analysis of extreme value at risk to amazon stocks. Int J Eng Res Dev 2018;14(2):62–71.
- [37] Nadarajah S, Kotz S. The beta Gumbel distribution. Math Probl Eng 2007;4:323–32.
- [38] Abbas K, Yincai T. Comparison of estimation methods for Frechet distribution with known shape. Casp J Appl Sci Res 2012;1(10):58–64.
- [39] Rinne H. The Weibull distribution: a handbook. Chapman and Hall/CRC; 2008.
- [40] Pinheiro EC, Ferrari SLP. A comparative review of generalizations of the Gumbel extreme value distribution with an application to wind speed data. J Stat Comput Simul 2015;86(11):2241–61.
- [41] Castillo E, Hadi AS, Balakrishnan N, Sarabia JM. Extreme value and related models with applications in engineering and science. New Jersey: John Wiley & Sons; 2005.

- [42] Ferrari SLP, Pinheiro EC. Small-sample one-sided testing in extreme value regression models. *AStA Adv Stat Anal* 2016;100(1):79–97.
- [43] Pickands J. Statistical inference using extreme order statistics. *Ann Stat* 1975;3:119–31.
- [44] Scarrott C, MacDonald A. A review of extreme value threshold estimation and uncertainty quantification. *RevStat - Stat J* 2012;10(1):33–60.
- [45] Gencay R, Selcuk F, Ulugulyagci A. EVIM: a software package for extreme value analysis in MATLAB. *Stud Nonlinear Dyn Econom* 2001;5(3):1–29.
- [46] Davison AC, Smith RL. Models for exceedance over high thresholds (with discussion). *JR Stat Soc B* 1990;52:237–54.
- [47] Ghosh S, Resnick SI. A discussion on mean excess plots. *Stoch Process Appl* 2010;120.
- [48] Kratz M, Resnick SI. The QQ-estimator and heavy tails. *Stoch Models* 1996;12(4):699–724.
- [49] Drees H, Haan L, DeResnick S. How to make a hill plot. *Ann Stat* 2000;28(1):254–74.
- [50] Todorovic P, Rousselle J. Some problems of flood analysis. *Water Resour Res* 1971;7(5):1144–50.
- [51] Todorovic P, Zelenhasic E. A stochastic model for flood analysis. *Water Resour Res* 1970;6(6):1641–8.
- [52] Hogg R, Klugman S. Loss distribution. New York: Wiley; 1984.
- [53] Embrechts P, McNeil A, Frey R. Quantitative risk management: concepts, techniques, and tools. New Jersey: Princeton University Press; 2005.
- [54] Guesf F, Proschan F. Mean residual life: theory and applications. Defence Technical Information Center; 1985.
- [55] Drees H. Refined estimators of the extreme value index. *Ann Stat* 1995;23(6):2059–80.
- [56] Dekkers AL, Einmahl JH, De Haan L. A moment estimator for the index of an extreme-value distribution. *Ann Stat* 1989;17(4):1833–55.
- [57] Northrop PJ, Coleman CL. Improved threshold diagnostic plots for extreme value analyses. *Extremes* 2014;17:289–303.
- [58] Lomba JS, Alves ML. L-moments for automatic threshold selection in extreme value analysis. 2019. [arXiv:1905.08726v1\[stat.ME\]](https://arxiv.org/abs/1905.08726v1).
- [59] Manurung A, Wigena AH, Djuraidah A. GPD Threshold estimation using measure of surprise. *Int J Sci* 2018;42(3):16–25.
- [60] Wadsworth J. Exploiting structure of maximum likelihood estimators for extreme value threshold selection. *Technometrics* 2016;58(1):116–26.
- [61] Thompson P, Cai Y, Reeve D, Stander J. Automated threshold selection methods for extreme wave analysis. *Coastal Eng* 2009;56:1013–21.
- [62] Langousis A, Mamalakis A, Puliga M, Deidda R. Threshold detection for the generalized Pareto distribution: review of representative methods and application to the NOAA NCDC daily rainfall database. *Water Resour Res* 2016;52:2659–81.
- [63] G'Sell M, Wager S, Chouldechova A, Tibshirani R. Sequential selection procedures and false discovery rate control. *J R Stat Soc* 2016;78(2):423–44.
- [64] Bader B, Yan J, Zhang X. Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate. *Ann Appl Stat* 2018;12(1):310–29.
- [65] Greenwood JA. Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resour Res* 1979;15(5):1049–54.
- [66] DuMouchel WH. Estimating the stable index α in the order to measure tail thickness: a critique. *Ann Stat* 1983;11:1019–31.
- [67] Ferreira A, De Haan L, Peng L. On optimising the estimation of high quantiles of a probability distribution. *Statistics* 2003;37:401–34.
- [68] Reiss RD, Thomas M. Statistical analysis of extreme values: with applications to insurance, finance, hydrology and other fields. Boston: Birkhauser; 2017.
- [69] Neves C, Alves MIF. Reiss and thomas' automatic selection of the number of extremes. *Comp Stat Data Anal* 2004;47:689–704.
- [70] Schneider LF, Krajina A, Krivobokova T. Threshold selection in univariate extreme value analysis. 2019. [arXiv:1903.02517v1\[stat.ME\]](https://arxiv.org/abs/1903.02517v1).
- [71] Floriani E, Volchenkov D, Lima R. A system close to a threshold of instability. *J Phys A* 2003;36:4771–83.
- [72] Volchenkov D. Survival under uncertainty an introduction to probability models of social structure and evolution. Springer Series: Understanding Complex Systems. 2016.
- [73] Volchenkov D. Grammar of complexity: from mathematics to a sustainable world. Nonlinear Physical Science. World Scientific Series; 2018.